UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA

CAIO ROBLEDO D’ ANGIOLI COSTA QUAIO

Achados moleculares das doenças raras

São Paulo
2023

CAIO ROBLEDO D' ANGIOLI COSTA QUAIO

Achados moleculares das doenças raras

Versão Corrigida

(Versão original encontra-se na unidade que aloja o Programa de Pós-graduação e na Biblioteca Digital de Teses e Dissertações da USP)

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Programa de Pediatria

Orientadora: Profa. Dra. Chong Ae Kim

São Paulo

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Quaio, Caio Robledo D'Angioli Costa
Achados moleculares das doenças raras / Caio Robledo D'Angioli Costa Quaio. -- São Paulo, 2023.
Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Pediatria.
Orientadora: Chong Ae Kim.

Descritores: 1.Doenças raras 2.Doenças genéticas 3.Estudo molecular 4.Sequenciamento do exoma 5.Diagnóstico etiológico 6.Análise genômica

USP/FM/DBD-274/23

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Nome: QUAIO, Caio Robledo D' Angioli Costa

Título: Achados moleculares das doenças raras

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para a obtenção do título de Doutor em Ciências.

Aprovado em:

Banca Examinadora

Prof(a). Dr(a). ____

Instituição: ____

Julgamento: ____

Prof(a). Dr(a). ____

Instituição: ____

Julgamento: ____

Prof(a). Dr(a). _______________________________________

Instituição: ____

Julgamento: ____

DEDICATÓRIA

Aos muitos raros e muito raros: que a esperança possa guiá-los em suas jornadas e que a ciência ilumine a busca pela verdade.

AGRADECIMENTOS

Neste momento de gratidão profunda, expresso meus sinceros reconhecimento e apreciação à minha amada esposa Helena pelo amor e apoio incondicionais. À minha mãe, Maria Tereza, meu irmão, Caio Ragrício, e toda minha família por serem o porto seguro e fonte de inspiração para a jornada da vida.

Minha profunda gratidão à Profa. Chong pela sabedoria, dedicação e orientação durante o percurso do doutorado. O suporte inabalável e a confiança depositados em mim ao longo dessa jornada moldaram este trabalho.

Aos meus valorosos amigos, colegas e colaboradores da Faculdade de Medicina da Universidade de São Paulo, do Grupo Fleury e do Hospital Israelita Albert Einstein, agradeço por compartilharem o fardo dos trabalhos e, também, as valorosas conquistas científicas ao longo desses quatro anos. A amizade sempre torna o desafio mais leve e a vitória mais saborosa!

Também temos de reconhecer o toque divino presente em cada passo dado, em cada obstáculo superado e em cada bênção recebida. Agradeço a Deus pela inspiração celestial.

Aos portadores de doenças raras, razão deste estudo: sou grato por todas as oportunidades de aprendizado. Que os conhecimentos gerados por esta pesquisa possam beneficiá-los.

RESUMO

Quaio CRDC. Achados moleculares das doenças raras [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2023.

As doenças raras são um grupo amplo e diverso de entidades clínicas, 80% das quais apresenta etiologia genética. A elucidação etiológica é fundamental para o manejo clínico adequado, mas o processo diagnóstico pode ser desafiador. O Sequenciamento do Exoma, método de estudo genômico em larga escala, revolucionou o diagnóstico das doenças raras. Este estudo busca entender como o Sequenciamento do Exoma pode auxiliar no diagnóstico de doenças raras e quais são seus impactos clínicos. Os objetivos secundários incluem comparar o rendimento diagnóstico desta técnica com outras metodologias, listar as principais indicações clínicas para investigação etiológica e expandir o conhecimento sobre variantes genéticas raras no Brasil e seus impactos populacionais. Para tanto, foram analisados os dados clínicos e moleculares retrospectivos de 500 indivíduos com doenças raras que realizaram sequenciamento do exoma no laboratório Fleury entre os anos de 2014 e 2020 para investigação etiológica. Observou-se rendimento diagnóstico geral de 31,6% (158/500) com o uso desta metodologia e 15,6% (78/500) desta coorte teve potencial direto de redirecionamento dos cuidados clínicos (terapia-alvo, rastreio para neoplasias, adoção de protocolo de seguimento padronizado, ajuste medicamentoso ou monitorização de complicações específicas). As doenças mais frequentes foram: síndrome de Rett, doenças associadas ao gene *POLG*, doença de Niemann–Pick tipo C, síndrome KBG, doenças associadas ao gene *DEAF1*, síndrome de Cowden e síndrome de Noonan. Foram também observados achados secundários com impacto clínico relevante em 7,4% (n = 37) da coorte. Observou-se, ainda, que 71,9% (230/320) dos indivíduos eram portadores em heterozigose de variantes associadas a doenças autossômicas recessivas. Considerando-se essas frequências de portadores (2pq), a equação de Hardy-Weinberg foi utilizada para estimar de forma aproximada a frequência populacional de doenças recessivas (q^2) em 26,39/10.000. Ademais, usando-se dados moleculares, populacionais e estudo de segregação nos genitores, 51 variantes raras foram classificadas como benignas e 211 como provavelmente benignas. Este estudo, pioneiro no Brasil, revelou detalhes importantes sobre achados moleculares de pacientes com doenças raras em uma população pouco representada na literatura científica. O diagnóstico preciso é fundamental para o acolhimento adequado desses indivíduos. O aprofundamento sobre o conhecimento das variantes genéticas raras no Brasil e seus impactos na saúde são importantes alicerces para o planejamento de estratégias para as doenças raras.

Palavras-chave: Doenças raras. Doenças genéticas. Estudo molecular. Sequenciamento do exoma. Diagnóstico etiológico. Análise genômica.

ABSTRACT

Quaio CRDC. Molecular findings of rare diseases [thesis]. São Paulo: “Faculdade de Medicina, Universidade de São Paulo”; 2023.

Rare diseases are a broad and diverse group of clinical entities, 80% of which present a genetic etiology. Etiological elucidation is essential for proper clinical management, but the diagnostic process can be challenging. Exome Sequencing, a method of large-scale genomic study, has revolutionized the diagnosis of rare diseases. This study aims to understand how Exome Sequencing can assist in the diagnosis of rare diseases and its clinical impacts. Secondary objectives include comparing the diagnostic yield of this technique with other methodologies, listing the main clinical indications for etiological investigation, and expanding knowledge about rare genetic variants in Brazil and their population impacts. For this purpose, clinical and molecular data from 500 individuals with rare diseases who underwent exome sequencing between 2014 and 2020 for etiological investigation were analyzed. An overall diagnostic yield of 31.6% (158/500) was observed using this methodology, and a significant portion of this cohort (15.6%; 78/500) had direct potential for redirection of clinical care (targeted therapy, screening for neoplasms, adoption of standardized follow-up protocol, medication adjustment, or monitoring of specific complications). The most frequent diseases were Rett syndrome, *POLG*-associated diseases, Niemann-Pick disease type C, KBG syndrome, *DEAF1*-associated diseases, Cowden syndrome, and Noonan syndrome. Secondary findings with relevant clinical impact were also observed in 7.4% (37/500) of the cohort. Furthermore, 71.9% (230/320) of individuals were heterozygous carriers of variants associated with autosomal recessive diseases. Considering these carrier frequencies (2pq), the Hardy-Weinberg equation was used to estimate the approximate population frequency of recessive diseases (q^2) at 26.39/10,000. Moreover, using molecular and population data, as well as parental segregation studies, 51 rare variants were classified as benign, and 211 as probably benign. This study revealed important details about molecular findings in patients with rare diseases in a population that is poorly represented in the scientific literature. Accurate diagnosis is essential for the proper care of these individuals. Further understanding of rare genetic variants in Brazil and their impacts on health are important foundations for planning strategies for rare diseases.

Keywords: Rare diseases. Genetic diseases. Molecular study. Exome sequencing. Etiologic diagnosis. Genomic analysis.

LISTA DE ABREVIAÇÕES E SIGLAS

a:	anos
ACMG:	"*The American College of Medical Genetics and Genomics*"
AD:	autossômico(a) dominante
AR:	autossômico(a) recessiva
BWA:	"*Burrows–Wheeler Aligner*"
CNVs:	variações do número de cópias
DP:	desvio-padrão
F:	Feminino(a)
GATK:	"*Genome Analysis ToolKit*"
LX:	ligada ao cromossomo X
M:	masculino(a)
m:	meses
mtDNA:	DNA mitocondrial
N:	número
NGS:	"*next-generation sequencing*"
OMIM:	"*Online Mendelian Inheritance in Man*"
P:	patogênicos(as)
P/PP:	patogênicos(as) ou provavelmente patogênicos(as)
PP:	provavelmente patogênicos(as)
VEP:	"*Variant Effect Predictor*"

SUMÁRIO

RESUMO
ABSTRACT
LISTA DE ABREVIAÇÕES E SIGLAS
1. INTRODUÇÃO 1
1.1. Diagnóstico de doenças monogênicas 1
1.2. Sequenciamento de nova geração (NGS) 2
1.3. Achados Secundários (Incidentais) 3
1.4. Estudo de doenças raras por sequenciamento do exoma na América Latina e justificativa do estudo 3
2. OBJETIVOS 5
3. METODOLOGIA 6
3.1. Seleção dos casos e aquisição de dados clínicos 6
3.2. Considerações éticas 7
3.3. Análise molecular, bioinformática e limitações metodológicas 7
3.4. Achados moleculares: Primários e Secundários 8
3.5. Estimativa de impacto clínico dos Achados Primários 9
3.6. Detecção de portadores e estimativa populacional da frequência de doenças recessivas 9
3.7. Variantes benignas e provavelmente benignas 11
3.8. Comparação simulada do rendimento diagnóstico entre o sequenciamento do exoma e painéis de genes-alvos 11
4. RESULTADOS 13
4.1. Achados Primários 13
4.2. Estimativa de impacto clínico dos Achados Primários 17
4.3. Achados secundários 17
4.4. Detecção de portadores e estimativa populacional da frequência de doenças recessivas 18
4.5. Variantes benignas e provavelmente benignas 20
4.6. Comparação simulada do rendimento diagnóstico entre o sequenciamento do exoma e painéis de genes-alvos 20
5. DISCUSSÃO 22
6. CONCLUSÃO 25
7. CONSIDERAÇÕES FINAIS E SUGESTÕES PARA TRABALHOS FUTUROS 26
8. REFERÊNCIAS 27
ANEXO I – Achados Secundários 32
ANEXO II – Listas de publicações científicas 34
ANEXO III – Premiações e honrarias 36

ANEXO IV – Apresentações em congressos e eventos científicos 37
ANEXO V – Íntegras dos artigos científicos 38

1. INTRODUÇÃO

As doenças raras correspondem a grupo amplo e diverso de aproximadamente 7.000 entidades clínicas, das quais cerca de 80% apresentam etiologia genética e acometem, individualmente, até 65 pessoas em cada 100 mil indivíduos (1, 2). A elucidação etiológica pode trazer benefícios no manejo clínico, prevenção de eventuais complicações de longo prazo, melhora do prognóstico e mais apropriado aconselhamento genético; entretanto, este processo diagnóstico pode ser desafiador e muitos desses pacientes enfrentam uma verdadeira “odisseia diagnóstica”. Esse desafio decorre de alguns fatores, entre os quais:

a) o diagnóstico guiado somente pela manifestação clínica é dificultado pelo fato de muitas doenças raras apresentarem sinais e sintomas comuns e, muitas vezes, inespecíficos;

b) o diagnóstico correto por ora depende da elucidação do mecanismo molecular;

c) os mecanismos moleculares podem ser complexos;

d) o número de doenças raras e, consequentemente, os mecanismos moleculares e genes a elas associados crescem continuamente no célere ritmo do avanço científico.

O Sequenciamento Completo do Exoma (referido neste texto como Sequenciamento do Exoma) consiste em método de estudo das regiões codificantes do Genoma Humano em larga escala. Enquanto que o primeiro estudo genômico humano teve custo total da ordem de US$1 bilhão em 2003 e diminuição para em torno de US$20-25 milhões até 2006, foi o advento da técnica de sequenciamento de nova geração que propiciou queda dramática no custo dos estudos genômicos para menos de US$1.000,00 e a adoção destas técnicas na rotina diagnóstica (3, 4). O estudo genômico de larga escala revolucionou o diagnóstico das doenças mendelianas raras e é a base do conceito de medicina personalizada.

1.1. Diagnóstico de doenças monogênicas

Antes do advento dos estudos genômicos na prática clínica, o diagnóstico das doenças monogênicas dependia, primeiramente, do reconhecimento de padrão de acometimento clínico. Neste contexto, o diagnóstico clínico era passo imprescindível

para guiar o diagnóstico molecular (5). Se para diversas doenças geneticamente determinadas o padrão clínico é facilmente reconhecível por geneticista experiente (p.ex., síndrome de Noonan, mucopolissacaridose, entre outras) possibilitando o diagnóstico clínico e teste genético específico considerando o mecanismo molecular da condição reconhecida, esta abordagem não é prontamente conseguida na grande maioria das doenças genéticas que não apresentam padrão clínico tão facilmente reconhecido (5-9). As técnicas baseadas em sequenciamento de nova geração, das quais o Sequenciamento do Exoma faz parte, trouxeram vantagens no processo de busca diagnóstica e potencialmente propiciam a elucidação de etiologia genética em até cerca de 30% dos casos em estudos internacionais (10, 11). Entretanto, dados brasileiros sólidos sobre o tema não foram encontrados.

1.2. Sequenciamento de nova geração (NGS)

O sequenciamento de nova geração (sigla em inglês NGS, de "*next-generation sequencing*"), também referido como sequenciamento paralelo em massa (do inglês "*massively parallel or deep sequencing*"), é a técnica que ascendeu em meados dos anos 2.000 e revolucionou o estudo genômico. Esta técnica permite tanto:

a) o estudo do genoma completo humano (i.e., todos os ~22.000 genes) em um único teste, como

b) selecionar múltiplas áreas restritas do genoma que possibilitem o estudo específico de regiões de interesse.

Desta forma, destacam-se dois dos principais produtos NGS rotineiramente utilizados:

a) Sequenciamento do Exoma (representa o sequenciamento de toda região codificante dos 20.000 genes do genoma) e

b) painéis de genes-alvos (são estudos customizáveis, menos abrangentes que o Sequenciamento do Exoma, apresentam menor custo e estudam número restrito de genes, variando de um único gene a algumas centenas de genes).

As plataformas de sequenciamento NGS utilizam-se do seguinte princípio: o DNA é fragmentado, é realizada uma etapa de captura de segmentos de interesse (que pode compreender desde todos os 20.000 genes até alguns poucos), é realizado o sequenciamento (i.e., a sequência molecular é lida e transformada em código alfanumérico) de milhões de fragmentos de DNA de modo concomitante (12-15). Toda

a informação gerada é analisada por protocolos de bioinformática para alinhar todas as sequências em um genoma referência e apontar as possíveis diferenças (12, 15, 16). Esta abordagem abre espaço para uma série de aplicações clínicas que podem contribuir muito para o diagnóstico e tratamento de doenças raras ou desconhecidas. Hoje, o Sequenciamento do Exoma custa menos de R$5.000,00 no Brasil, fato que ampliou o acesso a este método diagnóstico.

1.3. Achados Secundários (Incidentais)

Achados secundários, outrora referidos como "incidentais", representam os achados moleculares que não estão associados ao quadro clínico que indicou a realização do teste diagnóstico, mas apresentam relevância para os cuidados médicos do indivíduo (17). O Colégio Americano de Genética Médica e Genômica (ACMG, do inglês "*The American College of Medical Genetics and Genomics*") publicou diretrizes para identificação e incorporação no resultado de achados relevantes para, originalmente, 62 doenças clínicas que apresentam conduta padronizada e cujo diagnóstico modifica o acompanhamento clínico do paciente. Estas doenças estão assinaladas na tabela do Anexo I. Esta lista de doenças passa por atualizações constantes. A frequência de achados secundários varia de acordo com a população estudada, de cerca de 1,5% em norte-americanos (18), 0,59% na península arábica (19), 7% em coreanos (20) e 0,7% em europeus (21). Não foram encontrados dados acerca deste tema no Brasil e em outros países da América Latina. Os achados secundários, apesar de não estarem diretamente associados ao quadro que indicou a investigação por teste genético, apresentam impacto relevante para a saúde, prognóstico e aconselhamento genético do indivíduo e sua família.

1.4. Estudo de doenças raras por sequenciamento do exoma na América Latina e justificativa do estudo

Não foram encontrados dados relevantes acerca do estudo de doenças raras por Sequenciamento do Exoma no Brasil e em outros países da América Latina. Ademais, não são também encontrados dados consolidados acerca da epidemiologia das doenças raras diagnosticadas por Sequenciamento do Exoma no Brasil. De fato, não foram encontrados estudos acerca do impacto diagnóstico e clínico desta técnica em coortes brasileiras.

Também não foram encontradas diretrizes ou recomendações de nenhuma sociedade médica no Brasil sobre políticas laboratoriais de reportar achados secundários, tampouco discussão ética abrangente sobre o tema ou padronização de termos de consentimento.

Desta forma, o estudo mais aprofundado sobre os achados moleculares do Sequenciamento do Exoma pode trazer luz:

a) ao impacto desta técnica no diagnóstico de doenças raras no Brasil;

b) ao estudo epidemiológico das doenças raras em nossa população;

c) à relevância dos achados secundários na população brasileira;

d) a ser potencialmente o embrião de discussão ético-legal inicial acerca do tema;

e) ao embasamento de decisões de sociedades médicas sobre o tema;

f) ao estabelecimento de diretrizes para termos de consentimento de testes genômicos;

g) à concepção de políticas públicas, no âmbito do Sistema Único de Saúde, que contemplem o acesso amplo a testes genéticos.

2. OBJETIVOS

Objetivo Primário:

- Identificar, caracterizar e comparar com a literatura internacional os Achados Primários e Secundários observados nos exames de Sequenciamento de Exoma para estudo de doenças raras, bem como dimensionar o impacto clínico desses achados.

Objetivos Secundários:

- Estudar o rendimento diagnóstico desta técnica nos diferentes grupos de doenças investigadas e comparar com outras metodologias NGS e dados internacionais;
- Elencar as principais indicações clínicas para investigação diagnóstica com sequenciamento completo do exoma;
- Ampliar o conhecimento sobre variantes genéticas brasileiras raras e seus potenciais impactos populacionais.

3. METODOLOGIA

Foi realizado estudo retrospectivo de todos os resultados de testes de Sequenciamento de Exoma já realizados pelo laboratório Fleury e que objetivavam a elucidação de doenças geneticamente determinadas entre os anos de 2014 e 2020 (22). O total de casos analisados compreendeu 500 pacientes. Exames que foram realizados com outros propósitos (ex.: participaram em outros protocolos de pesquisa) não foram avaliados. De modo sumarizado, temos que:

a) os dados clínicos e moleculares foram revisados e tabulados;

b) os achados primários e secundários foram revisados e seus dados tabulados;

c) os dados de segregação de variantes foram revisados e tabulados;

d) dados adicionais foram também revisados e tabulados.

3.1. Seleção dos casos e aquisição de dados clínicos

Os achados moleculares dos primeiros 500 Sequenciamentos do Exoma de pacientes adultos, pediátricos e fetais foram revisados. Todas as amostras foram coletadas entre os anos de 2014 a 2020. O estudo incluiu pacientes sintomáticos que realizaram o teste molecular para investigação diagnóstica de possível etiologia genética; 31 pacientes haviam sido previamente excluídos por terem se recusado a compartilhar seus dados genômicos ou por terem realizado o teste para fins não-diagnósticos (participaram de outros protocolos de pesquisa ou o teste tinha o propósito exclusivo de aconselhamento genético reprodutivo).

Os dados clínicos dos 500 pacientes incluídos no estudo foram coletados por questionário abrangente aplicado previamente à realização do teste. Também foram revisadas as informações eventualmente presentes nos relatórios médicos ou pedidos de exames disponibilizados.

As manifestações clínicas foram categorizadas em:

a) transtornos do neurodesenvolvimento (incluindo transtorno do espectro autista, deficiência intelectual e distúrbios do comportamento);

b) epilepsia (incluindo crises convulsivas, espasmos e encefalopatia);

c) outras alterações neurológicas (grupo amplo que inclui hipotonia, hipertonia, doenças neuromusculares, perda auditiva, alterações da visão, ataxia);

d) quadros sindrômicos/malformativos (inclui microcefalia, macrocefalia, alterações esqueléticas, más-formações congênitas);

e) distúrbio do crescimento (incluindo restrição de crescimento intrauterino, baixa estatura, alta estatura, déficit pôndero-estatural, obesidade);

f) alterações imunológicas ou hematológicas (imunodeficiências primárias, infecções recorrentes, autoimunidade, alergias, citopenias e defeitos da coagulação);

g) câncer/tumores;

h) doenças cardiovasculares (incluindo cardiomiopatias, arritmias, más-formações cardiovasculares, acidentes vasculares);

i) doenças gastrointestinais ou hepáticas;

j) consanguinidade;

k) alterações metabólicas ou hormonais;

l) outras manifestações

Nota: O mesmo paciente pode apresentar múltiplas manifestações de múltiplos grupos clínicos.

3.2. Considerações éticas

O estudo foi conduzido de acordo com princípios éticos seguindo as orientações contidas na declaração de Helsinki a foi aprovado pelas Comissões de Ética em Pesquisa de todas as instituições envolvidas (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury#3.372.339). Foi pré-requisito obrigatório para realização do Sequenciamento do Exoma o preenchimento prévio de termo de consentimento.

3.3. Análise molecular, bioinformática e limitações metodológicas

A análise molecular consistiu na extração e fragmentação de DNA genômico a partir de amostra sanguínea, salivar ou pré-natal (biópsia de vilosidade coriônica ou líquido amniótico) seguida de identificação e captura com kit específico *Agilent Clinical Research Exome* v1® (Agilent, Santa Clara-CA, EUA), seguindo as normativas de boas práticas laboratoriais e recomendações dos fabricantes. O sequenciamento foi realizado utilizando-se a plataforma illumina NextSeq® (illumina Inc., San Diego-CA, EUA) (22).

Os dados gerados pelo sequenciamento foram alinhados ao genoma referência GRCh37.75/hg19 utilizando-se "*Burrows–Wheeler Aligner*" (BWA; versão 0.7. 17-r1188). A seguir, foram manipulados por processos customizados internamente com "*Genome Analysis ToolKit"* (GATK, versão 3. 8-0-ge9d806836) para a identificação de variantes de sequência e "*Variant Effect Predictor*" (VEP, versão 88.14) para anotação. Todos os exames atingiram uma cobertura mínima de bases-alvos de 95% para 10 leituras independentes (10x).

Ao menos duas análises em paralelo foram realizadas para cada pacientes com o intuito de pré-selecionar variantes genômicas considerando-se frequência populacional menor que 1%, impacto funcional, relevância clínica do gene com a respectiva alteração, relatos nos bancos de dados (p.ex., ClinVar, HGMD) e literatura científica. As variantes pré-selecionadas eram, então, discutidas em reunião com três membros da equipe multidisciplinar para avaliar a relevância clínica e, finalmente, selecionar as variantes relevantes para etapa confirmatória com sequenciamento Sanger, conforme descrita a seguir.

Desta forma, as variantes foram interpretadas considerando-se o quadro clínico do paciente e protocolo de classificação de variantes do ACMG (16). Todas as variantes clinicamente relevantes foram confirmadas pelo método Sanger no probando e investigadas nos genitores para estudo de segregação, se as amostras dos genitores estivessem disponíveis.

Este teste não avaliou regiões não-codificantes do genoma, variações do número de cópias (como microdeleções ou microduplicações), tampouco doenças genéticas associadas a expansões de polinucleotídeos ou derivadas de alterações do DNA mitocondrial (mtDNA) (22).

3.4. Achados moleculares: Primários e Secundários

Os Achados Primários foram definidos como variantes genômicas patogênicas ou provavelmente patogênicas observadas em genes associados a doenças com manifestações clínicas com sobreposição ao quadro apresentado pelo probando e padrão de herança compatível. Desta forma, os achados primários são as variantes que definem o diagnóstico molecular do quadro que originalmente suscitou a investigação genética.

Os Achados Secundários referiram-se às variantes patogênicas ou provavelmente patogênicas em genes associados a doenças com conduta clínica padronizada por diretrizes internacionais e cujo diagnóstico precoce traz benefício relevante em termos de redução da morbimortalidade ou melhora da qualidade de vida. Desta forma, os Achados Secundários não apresentam relação direta com o quadro clínico que motivou a investigação do probando, mas podem impactar o manejo clínico e aconselhamento genético do mesmo e de sua família. O ACMG apresentava recomendações sobre a lista de doenças e respectivos genes de relato preconizado, que foi, neste estudo, utilizada como norteador para relato dos Achados Secundários (17). Entretanto e de modo excepcional, alterações em genes fora desta lista e que, por juízo clínico subjetivo, foram consideradas potencialmente impactantes para a saúde do paciente também foram relatadas.

3.5. Estimativa de impacto clínico dos Achados Primários

Objetivando-se estimar de modo mais preciso o impacto clínico de um diagnóstico molecular e inspirando-se em estudo pregresso de Meng *et al.* (23), formou-se equipe multidisciplinar com diversos especialistas de distintas áreas do conhecimento médico (p.ex., neurologia, reumatologia, imunologia, ginecologia) para avaliar potenciais mudanças na condução do tratamento do paciente frente a um diagnóstico assertivo. Assim, experiência clínica dos especialistas, dados contidos no banco "*Online Mendelian Inheritance in Man*" (OMIM) e busca na literatura científica foram utilizados para estimar o potencial impacto clínico dos Achados Primários nos seguintes quesitos: terapia-alvo, rastreio de neoplasias, adoção de diretrizes (*guidelines*) internacionais, ajuste medicamentoso e monitorização de complicações específicas (os detalhes acerca desta abordagem encontram-se publicados em Quaio *et al.*, 2020 (22)).

3.6. Detecção de portadores e estimativa populacional da frequência de doenças recessivas

Exclusivamente para esta análise de portadores de variantes associadas a doenças recessivas, foi utilizado subgrupo de 320 indivíduos pertencentes à coorte original de 500 pessoas. Desta forma e a fim de se caracterizarem as frequências de portadores de alelos patogênicos ou provavelmente patogênicos (P/PP) de condições

autossômicas recessivas raras, foram revisados os achados moleculares de sequenciamento completo do exoma de 320 pacientes da coorte (o detalhamento metodológico desta etapa encontra-se pormenorizado em Quaio *et al*., 2021 (24) e Quaio *et al*., 2022 (25)). Em todos estes 320 indivíduos, o status de portador para doenças autossômicas recessivas foi ativamente pesquisado seguindo a abordagem de duas análises em paralelo anteriormente descrita no item 3.3. Foram excluídas de nossa análise:

a) variantes associadas a doenças autossômicas recessivas com alta frequência alélica e baixa penetrância (por exemplo, genes *HFE* e *SERPINA1*);
b) variantes associadas a quadros sem nenhum impacto clínico (traços fenotípicos sem impacto à saúde);
c) variantes associadas a condições em que o diagnóstico depende de outros testes (por exemplo, variantes em *BTD* e *G6PD*);
d) genes com pseudogenes com alta homologia (como *CYP21A2*).

Para o propósito desta análise, os genes foram divididos em categorias com base em seu impacto clínico. De acordo com nosologias, revisões e diretrizes profissionais disponíveis na literatura, todos os genes foram classificados nas seguintes categorias clínicas:

a) deficiência intelectual (26);
b) câncer (27);
c) doenças musculares (28);
d) ciliopatias (29);
e) distúrbios esqueléticos (30);
f) doenças imunológicas (31);
g) epilepsia (32);
h) perda auditiva (33);
i) retinite pigmentosa (34);
j) distúrbios metabólicos (35);
k) outros (genes não classificados em nenhuma outra categoria clínica).

O mesmo gene pôde pertencer a múltiplas categorias clínicas, exceto por "outros", que incluiu apenas genes que não fazem parte de outra categoria clínica.

Neste estudo, utilizamos a equação de Hardy-Weinberg para estimar as frequências de doenças (q^2) com base nas respectivas frequências de portadores (2pq) observadas. Para isso, consideramos o acasalamento aleatório e a aproximação p~1. Essa análise foi realizada de acordo com o método padronizado previamente (36).

3.7. Variantes benignas e provavelmente benignas

Para o propósito de reclassificação de diversas variantes que foram encontradas em genes associados a doenças autossômicas dominantes e de alta penetrância, utilizou-se o seguinte princípio: para diversas doenças de herança dominante, um evento mutacional *de novo* é o mecanismo molecular patogênico esperado; de modo oposto, se a variante é herdada de um genitor assintomático, este dado de segregação sugere sua benignidade. Nesta abordagem, a análise de segregação e outros dados moleculares de diversas variantes raras foram utilizados para reclassificar como benignas ou provavelmente benignas diversas variantes raras encontradas, nesta coorte de 500 pacientes, em genes associados a doenças de herança autossômica dominante e de alta penetrância (detalhes sobre esta etapa metodológica, o escopo de genes em que foi aplicada, bem como suas limitações estão detalhados em Quaio *et al.*, 2022 (37)).

3.8. Comparação simulada do rendimento diagnóstico entre o sequenciamento do exoma e painéis de genes-alvos

Foi realizada uma comparação simulada entre o rendimento diagnóstico potencial de diferentes painéis de genes-alvos e o Sequenciamento do Exoma. O objetivo desta etapa é responder ao questionamento: “se houvéssemos utilizado painéis de genes-alvos guiados pela manifestação clínica do paciente ao invés do estudo do exoma, qual seria nosso rendimento diagnóstico?”.

Assim, para estimar a fração de possíveis diagnósticos moleculares perdidos por diferentes painéis de genes-alvos, consideraram-se:

a) os casos positivos (i.e., casos com achados primários; n=158) identificados em nossa coorte;
b) a manifestação clínica de cada paciente com teste positivo, de modo que os pacientes com determinada manifestação foram utilizados para a simulação de rendimento diagnóstico exclusivamente dos painéis de

genes-alvos destinados à investigação da mesma manifestação (p.ex., somente os casos com manifestação clínica de epilepsia participaram na simulação dos painéis de genes-alvos destinados à investigação de epilepsia);

c) a correspondente lista de genes estudados por cada painel.

Para este fim, foram selecionados 53 painéis de genes-alvos distintos de oito laboratórios comerciais. As listas de genes abrangidos por cada painel foram providas dos *websites* dos laboratórios e os dados sobre cobertura dos painéis foram acessados entre outubro e novembro de 2020 (o detalhamento completo desta simulação, incluindo seus princípios lógicos, dados acerca de todos os painéis e listas de genes estão disponíveis em Quaio *et al.*, 2021 (38)).

4. RESULTADOS

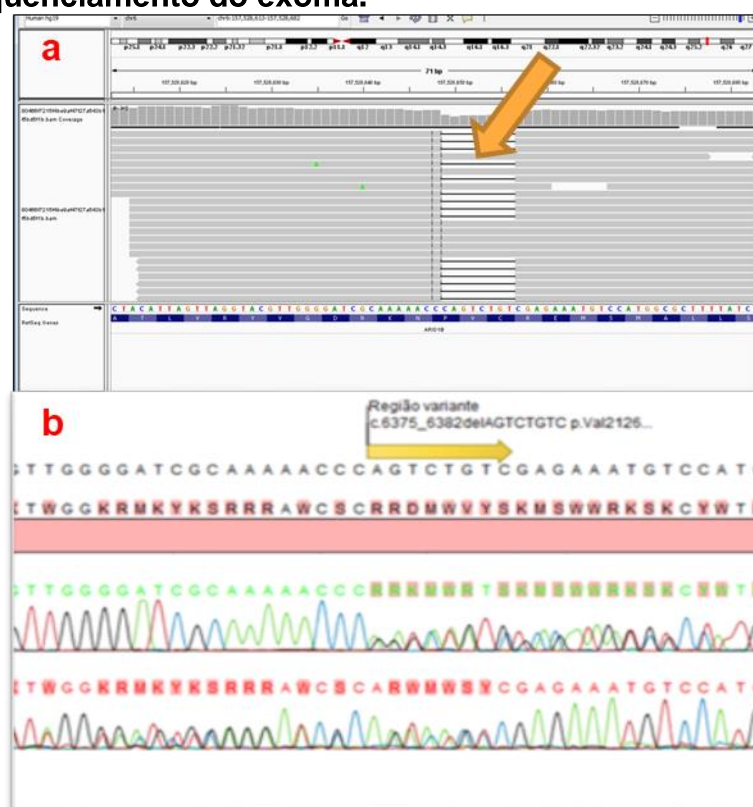
Este estudo revisou achados clínicos e moleculares de 500 pacientes com suspeita de doenças raras de origem genética e que foram investigados por Sequenciamento do Exoma. Os principais achados moleculares desta coorte foram publicados em série de sete artigos científicos cujos detalhes se encontram no Anexo II (22, 24, 25, 37-40). Apesar de as manifestações clínicas que motivaram a investigação terem sido variáveis, observamos um predomínio relevante de indivíduos com algum tipo de manifestação neurológica (67,2%). Os grupos de manifestações clínicas mais relevantes desta coorte e suas correspondentes frequências são: transtornos do neurodesenvolvimento (47%), epilepsia (21%), outras manifestações neurológicas (50%), sindrômicos/malformativos (35%), doenças imunológicas/hematológicas (19%), alteração de crescimento (15%), consanguinidade (1%).

4.1. Achados Primários

Os resultados correspondentes aos Achados Primários foram detalhadamente publicados (22). De modo geral, achados primários foram relatados em 158 pacientes, o que representa um rendimento diagnóstico global de 31,6%.

A Figura 1, abaixo, demonstra, visualmente, como é o alinhamento genômico utilizando-se a técnica em investigação de criança do sexo masculino de quatro anos que apresentava atraso do desenvolvimento neuropsicomotor, face grosseira, boca grande e epilepsia. O Sequenciamento do Exoma revelou a deleção de oito pares de bases em heterozigose no gene *ARID1B* (anotação oficial da variante: *ARID1B*(ENST00000346085): c.6375_6382del8: p.Val2126Argfs*63), associado à síndrome Coffin-Siris. Neste caso, o estudo genômico possibilitou o diagnóstico específico.

Figura 1 – Sequenciamento do exoma.



Deleção em heterozigose de oito pares de bases no gene *ARID1B* (anotação oficial da mutação: *ARID1B*(ENST00000346085): c.6375_6382del8: p.Val2126Argfs*63, heterozigose), associado à síndrome Coffin-Siris. Em **a**, seta demonstra os espaços vazios do segmento faltante do gene no alinhamento do sequenciamento NGS e **b** demonstra a mesma alteração no estudo por sequenciamento Sanger. Fonte: elaborado pelo autor.

As Tabelas 1 e 2 detalham os rendimentos diagnósticos por faixa etária e por manifestação clínica, respectivamente. No total, 101 doenças autossômicas dominantes (AD) foram observadas em 98 indivíduos; 42 doenças autossômicas recessivas (AR) foram relatadas em 42 indivíduos e 21 doenças de herança ligada ao cromossomo X (LX) foram observadas em 21 indivíduos. A presença concomitante de duas doenças genéticas ocorreu em seis pacientes.

Nestes 158 pacientes com achados primários, foram identificadas 195 variantes. Não foram encontrados relatos na literatura para 43,6% (n = 85) das variantes e em 37,9% (n = 74) não havia menção no banco de dados público ClinVar. Destas 195 variantes, 55,9% foram classificadas como patogênicas (n = 109), 43,1% como provavelmente patogênicas (n = 84) e 0,4% como variantes de significado incerto (n = 2). Estas duas variantes de significado incerto foram relatadas como achados primários porque ambas foram encontradas em heterozigose composta (i.e., em

"*trans*") com segunda variante claramente patogênica em genes associados a doenças autossômicas recessivas compatíveis com a manifestação clínica.

Considerando-se as 101 doenças AD, 68,3% (n = 69) decorrem de variantes em heterozigose que ocorreram de evento mutacional *de novo*, enquanto que em 13,9% (n = 14) a variante foi herdada de um dos genitores. Outras seis variantes correspondem a haplótipo recorrente no gene *POLG* e foram observadas em três pacientes (3%). O estudo de segregação de variantes não foi realizado em 14,9% (n = 15) destes indivíduos por falta de amostra dos genitores.

Em relação às 42 doenças AR, 33,3% (n = 14) decorreram de variantes em homozigose e 54,8% (n = 23) de duas variantes em heterozigose composta. Em 11,9% (n = 5), foram encontradas duas variantes em heterozigose, mas a diferenciação de fase (i.e., *cis* ou *trans*) não foi possível por falta de amostra dos genitores.

Considerando-se as 21 doenças LX, 47,6% (n = 10) foram associadas a variantes em hemizigose e 52,4% (n = 11) a variantes em heterozigose. Em relação à segregação destas variantes, 47,6% (n = 10) decorreram de evento mutacional *de novo*, 28,6% (n = 6) foram herdadas das mães e em 23,8% (n = 5) a segregação não foi determinada.

As doenças mais frequentemente observadas neste estudo e seus respectivos genes são:

a) síndrome de Rett (*MECP2*, OMIM:300005): sete pacientes;
b) doenças associadas ao gene *POLG* (*POLG*, OMIM: 174763): quatro pacientes;
c) doença de Niemann–Pick tipo C (*NPC1*, OMIM: 607623): três pacientes;
d) síndrome KBG (*ANKRD11*, OMIM: 611192): três pacientes;
e) doenças associadas ao gene *DEAF1* (*DEAF1*, OMIM: 602635): três pacientes;
f) síndrome de Cowden (*PTEN*, OMIM: 158350): três pacientes;
g) síndrome de Noonan (*PTPN11*, OMIM: 176876): três pacientes.

Tabela 1 – Rendimento diagnóstico por faixa etária.

Faixa etária	N =	Gênero	Idade média	Rendimento diagnóstico
Pré-natal	6	-	-	67%
<1 ano	55	M=58,2%	5m	44%
1-5 anos	161	M=61,5%	2a9m	32%
5-10 anos	78	M=66,7%	7a	32%
10-18 anos	63	M=54%	13a6m	30%
18-25 anos	26	F=57,7%	21a2m	27%
25-50 anos	66	F=59%	36a10m	32%
>50 anos	45	F=60%	61a6m	13%
Todas idades	500	M=55,3%	15a5m	32%

As colunas informam, respectivamente, a divisão de faixas etárias, o número (N) de pacientes por respectiva faixa etária, o gênero predominante para cada faixa etária (masculino [M] ou feminino [F]) com a correspondente frequência, a idade média por faixa etária (ano [a] e mês [m]) e, finalmente, o rendimento diagnóstico obtida pelo sequenciamento do exoma para cada faixa etária. A última linha congrega todos os dados da coorte de 500 indivíduos com doenças raras. Fonte: Adaptado de Quaio *et al.*, 2020 (22).

Tabela 2 – Rendimento diagnóstico por manifestação clínica.

Indicações clínicas	Rendimento diagnóstico
Consanguinidade	50%
Doença Gastrointestinal/Hepática	44%
Sindrômico/Malformativo	41%
Transtornos do neurodesenvolvimento	35%
Outras anomalias neurológicas	35%
Alteração de crescimento	35%
Epilepsia	34%
Alteração Metabólica/hormonal	33%
Câncer/tumores	32%
Doença Cardiovascular	31%
Doenças imunológicas/hematológicas	19%
Outras	13%

A segunda coluna informa o rendimento diagnóstico, em ordem decrescente, para cada grupo de manifestação clínica (primeira coluna). Fonte: Adaptado de Quaio *et al.*, 2020 (22).

4.2. Estimativa de impacto clínico dos Achados Primários

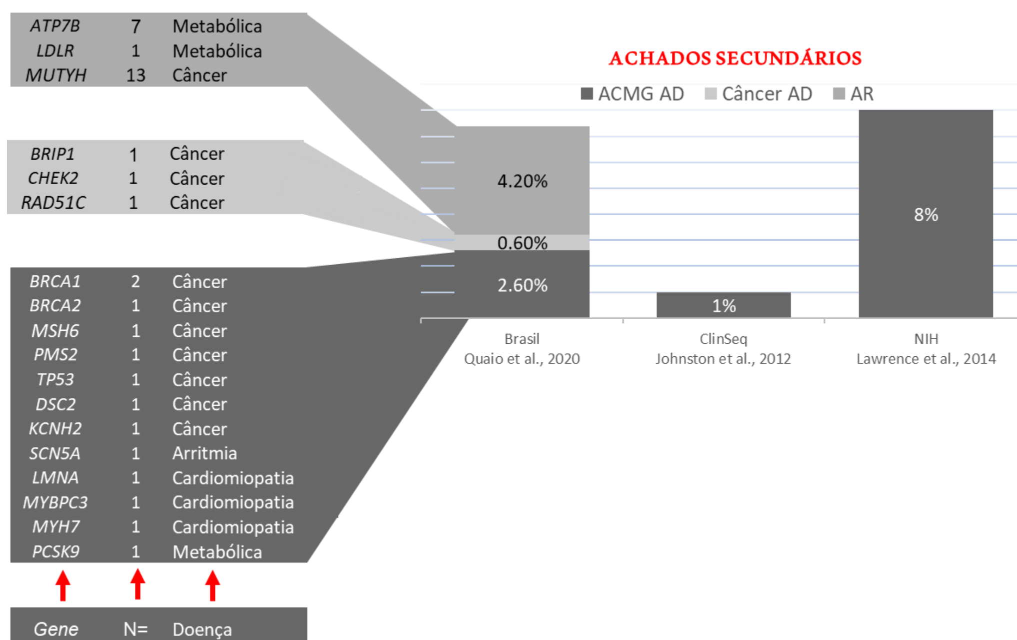
A estimativa de impacto clínico dos Achados Primários foi pormenorizada em publicações anteriores (22, 39, 40). Foi estimado que 78 pacientes (15,6% de toda a coorte, ou 49,4% dos pacientes com Achados Primários) tiveram potencial benefício clínico direto com a elucidação etiológica. Nós observamos indicação de terapia-alvo em 3,2% da coorte (n = 16), rastreio específico para doenças com aumento de risco de neoplasias em 1,6% (n = 8), adoção de protocolo de seguimento padronizado em 0,8% (n = 4), ajuste medicamentoso em 4,4% (n = 22) e monitorização de complicações específicas das doenças para 5,6% (n = 28).

4.3. Achados secundários

Os resultados correspondentes aos Achados Secundários, incluindo estimativas de impacto clínico e na expectativa de vida, foram detalhadamente publicados (22, 39, 40). Achados secundários foram observados em 7,4% (n = 37) dos indivíduos e englobaram 23 variantes em heterozigose em 18 genes distintos.

Do total de pacientes da coorte, 2,6% (n = 13) apresentaram variantes em heterozigose associadas a doenças AD (arritmias hereditárias, cardiomiopatias hereditárias, câncer hereditário e doenças metabólicas) pertencentes à lista do ACMG, outros 0,6% (n = 3) apresentaram variantes em genes associados a outras formas de câncer hereditário AD fora da lista do ACMG (genes *CHEK2*, *RAD51C* e *BRIP1*) e 4,2% (n = 21) eram portadores de uma única variante monoalélica em genes associados a doenças AR. A Figura 2, abaixo, sumariza os achados secundários identificados no presente estudo e compara as frequências com outros dois estudos internacionais.

Figura 2 – Achados Secundários.



A Figura 2 sumariza as doenças clínicas com conduta padronizada que foram observadas na coorte com 500 pacientes com doenças raras que realizaram Sequenciamento do Exoma com propósito diagnóstico (primeira coluna, Brasil), juntamente com a frequência de achados secundários em outros dois grandes estudos internacionais (segunda e terceira colunas). Os achados encontrados no presente estudo estão pormenorizados no diagrama da esquerda: ACMG AD representa o grupo de doenças de herança autossômica dominante de relato obrigatório de acordo com as recomendações do ACMG; Câncer AD representa genes associados a outras formas de câncer hereditário fora da lista do ACMG e que foram julgados clinicamente relevantes para vigilância tumoral e aconselhamento genético e AR representa o grupo portador de variantes monoalélicas em genes associados a doenças autossômicas recessivas. Fonte: Adaptado de Quaio *et al.* (22, 39, 40).

4.4. Detecção de portadores e estimativa populacional da frequência de doenças recessivas

Os resultados correspondentes à detecção de portadores e estimativa populacional da frequência de doenças recessivas juntamente com a íntegra dos dados foram detalhadamente publicados (24, 25). Um total de 425 ocorrências de 351 variantes diferentes (47 variantes recorreram mais de uma vez) em 278 genes diferentes foram relatadas para 230 pacientes (71,9%) em nossa amostra de 320 pacientes que tiveram o status de portador estudado. Em 90 indivíduos (28,1%), nenhuma variante foi relatada. Observamos uma média de 1,32 variantes/indivíduo (mediana = 1; DP = 1,188; máximo = 5), sendo que 218 variantes diferentes foram classificadas como provavelmente patogênicas e 133 como patogênicas.

A maioria das variantes patogênicas ou provavelmente patogênicas já havia sido observada na literatura (n = 257; 73,2%) e estava presente no Banco de Dados

ClinVar (n = 267; 76%). Por outro lado, 72 variantes (20,5%) não foram encontradas na literatura, tampouco no banco de dados ClinVar. Todas essas 72 variantes apresentaram frequência consideravelmente baixa nos bancos de dados populacionais (por exemplo, frequência inferior a 0,01%) junto com um mecanismo bem estabelecido de perda de função e, com base em uma combinação dos critérios das diretrizes ACMG, foram classificadas como provavelmente patogênicas.

Das 425 ocorrências totais, aquelas ocorridas em genes associados a distúrbios metabólicos compreenderam quase metade de todas as variantes (48,2%). Este grupo foi seguido por epilepsia (22,8%) e deficiência intelectual (18,8%). As variantes presentes em genes associados a doenças musculares corresponderam à categoria menos comum (2,4%).

Esta abordagem também propiciou o estudo do número de portadores e o cálculo da frequência de portadores por grupo de doenças, demonstrada na tabela abaixo.

Tabela 3 – Frequência e número de heterozigotos por grupo de doenças.

Grupo de doenças	Heterozigotos	
	Número de heterozigotos	Frequência de heterozigotos (%)
Doenças metabólicas	156	48,80%
Epilepsia	83	25,90%
Deficiência intelectual	58	18,10%
Outras	66	20,60%
Displasias esqueléticas	50	15,60%
Doenças imunológicas	35	10,90%
Perda auditiva	29	9,10%
Retinite pigmentosa	24	7,50%
Ciliopatias	24	7,50%
Câncer	14	4,40%
Doenças musculares	10	3,10%
Total	228	71,30%

As segunda e terceira colunas informam, respectivamente, o número e frequência de heterozigotos por Grupo de Doenças (primeira coluna). O total de indivíduos avaliado para este quesito foi 320 (100%). Fonte: Adaptado de Quaio *et al.* (24, 25).

Utilizando-se a equação de Hardy-Weinberg, estimamos de forma aproximada as frequências populacionais de doenças recessivas (q^2) com base nas respectivas frequências de portadores (2pq) observadas neste estudo e na aproximação de p~1.

De acordo com este método, estimamos as seguintes sete frequências de doenças mais elevadas por 10.000 pessoas:

a) perda auditiva não sindrômica associada ao gene *GJB2*: 3,52;
b) polipose relacionada ao gene *MUTYH*: 1,98;
c) retinose pigmentar associada ao gene *ABCA4*: 1,20;
d) doença de Gaucher: 0,88;
e) fenilcetonúria: 0,88;
f) síndrome de Aicardi-Goutieres: 0,88;
g) paraplegia espástica hereditária 7: 0,61.

A frequência combinada dessas sete doenças mais frequentes é estimada em 10,72 por 10.000 pessoas, enquanto que a frequência combinada de todas as doenças recessivas analisados por este estudo é de 26,39/10.000.

4.5. Variantes benignas e provavelmente benignas

Os resultados correspondentes à classificação de variantes benignas e provavelmente benignas juntamente com a íntegra dos dados foram detalhadamente publicados (37). Usando-se dados moleculares, populacionais e estudo de segregação nos genitores, 51 variantes raras foram classificadas como benignas e 211 variantes raras foram classificadas como provavelmente benignas. Destas 262 variantes, parte considerável não apresentava correspondência nos bancos de dados: 76 variantes (29%) sem correspondência no dbSNP e 189 variantes (72,1%) não encontradas entradas no ClinVar.

4.6. Comparação simulada do rendimento diagnóstico entre o sequenciamento do exoma e painéis de genes-alvos

O estudo acerca da comparação simulada do rendimento diagnóstico entre o sequenciamento do exoma e painéis de genes-alvos juntamente com a íntegra dos dados foi publicado (38). Os resultados dos 158 indivíduos com achados primários (resultados positivos) foram utilizados para simular a fração de diagnósticos que seria perdida por painéis de genes-alvos guiados pelas manifestações clínicas. A fração de diagnósticos moleculares perdidos, considerando-se todos os 53 painéis, foi estimada em média em 64%, variando de 14% a 100%, dependendo do laboratório e da

indicação clínica. O grupo de Doenças Metabólicas representa os painéis com a maior média de diagnósticos perdidos (86%), caso a investigação dos pacientes com suspeita de doenças metabólicas fosse conduzida por painéis de genes-alvos ao invés do Sequenciamento do Exoma. Por outro lado, Epilepsia foi o grupo com a menor média de diagnósticos potencialmente perdidos (46%). A tabela abaixo mostra a simulação para cinco grupos de doenças de doenças, a fração média de diagnósticos potencialmente perdidos e a simulação do rendimento diagnóstico dos painéis de genes-alvos para cada grupo de doenças.

Tabela 4 – Rendimento diagnóstico simulado dos painéis de genes-alvos por grupo de doenças.

Grupos de doenças (número de indivíduos)	Número de painéis	Fração média de diagnósticos potencialmente perdidos	Rendimento diagnóstico do Sequenciamento do Exoma	Simulação do rendimento diagnóstico dos painéis de genes-alvos
Transtornos do neurodesenvolvimento (n=54)	9	57%	35%	15%
Epilepsia (n=36)	11	46%	34%	18%
Sindrômico/malformativo (n=72)	4	59%	41%	17%
Doenças imunológicas (n=18)	9	50%	19%	10%
Doenças cardiovasculares (n=16)	10	78%	31%	7%
Doenças metabólicas (n=17)	10	84%	33%	5%

A primeira coluna demonstra os diferentes grupos de doenças avaliados nesta simulação e os correspondentes números de pacientes com achados primários utilizados para a comparação simulada; a segunda coluna informa o número de painéis avaliados para cada grupo de doença; a terceira coluna informa a fração média de diagnósticos potencialmente perdidos pelos correspondentes painéis de genes-alvos; a quarta consta o rendimento diagnóstico do Sequenciamento do Exoma observado nesta coorte e, finalmente, a última coluna informa o rendimento diagnóstico simulado dos painéis de genes-alvos para cada grupo de doenças. Fonte: Adaptado de Quaio *et al.* (38).

5. DISCUSSÃO

Este estudo, pioneiro no Brasil e que avaliou a utilização da técnica de sequenciamento do exoma em coorte de indivíduos com doenças raras e manifestação clínica heterogênea, evidenciou elucidação etiológica em 31,6% dos indivíduos. Ademais, uma parcela significativa desta coorte (15,6%) tem potencial pós-teste de redirecionamento dos cuidados clínicos, fato que sugere benefício da investigação molecular das doenças raras.

Rendimentos diagnósticos entre 22% e 30% já foram relatados em outros estudos que avaliaram a utilização desta técnica molecular na investigação de doenças raras (10, 41-46). Vários fatores podem impactar na variação das taxas diagnósticas, incluindo a avaliação clínica pré-teste e os passos complexos da técnica de NGS (captura das regiões de interesse, cobertura dos genes, protocolos de bioinformática e a análise médica). Em tratando-se de amostra de conveniência de um laboratório de análises clínicas com fins comerciais, a disponibilização dos dados clínicos (eventuais restrições de quantidade e qualidade de dados fornecidos) é um fator que pode ter influenciado nossos resultados.

Mesmo assim, o rendimento diagnóstico observado neste estudo é congruente com as taxas encontradas em outros estudos que avaliaram doenças raras em distintos países e etnias. As mais elevadas taxas diagnósticas foram encontradas para amostras pré-natais, crianças com menos de um ano de idade, indivíduos com genitores consanguíneos e quadros sindrômicos/malformativos (Tabelas 1 e 2). Estes mesmos grupos também apresentaram as taxas mais elevadas em outros estudos (23, 42, 44, 45).

De modo geral, o rendimento diagnóstico do sequenciamento do exoma mostra-se superior à de outras técnicas utilizadas no estudo de doenças raras, como o estudo por microarranjos (do inglês "*Microarray*"), técnicas tradicionais de citogenética ou mesmo painéis de genes-alvos (47). Nós utilizamos os dados de nossa coorte para estimar a fração potencial de diagnósticos moleculares perdidos por diferentes painéis de genes-alvos de diferentes laboratórios comerciais e observamos que os painéis de genes-alvos poderiam perder cerca de 64% dos diagnósticos de nossa coorte. Estas constatações em conjunção com o advento de análise de variantes de número de cópias (p.ex., microdeleções ou microduplicações) por protocolos de bioinformática

acoplados à análise por NGS tornaram o sequenciamento do exoma a primeira linha investigativa das doenças raras em diversos serviços de genética.

Nosso estudo identificou uma frequência relevante (7,4%) de achados secundários, que são alterações deletérias em genes associados a doenças de importante repercussão clínica, como câncer hereditário, arritmias, cardiomiopatias e doenças metabólicas (Figura 2). Existem desafios na determinação de quais desses achados moleculares com condutas clínicas padronizadas devem ser relatados de modo obrigatório. Em geral, os laboratórios de análises moleculares seguem as diretrizes norte-americanas ou europeias, uma vez que não existem regulações ou legislação nacional que abordem o tema (17, 22, 39, 40). É imperativo que as sociedades médicas brasileiras liderem a discussão nacional sobre os achados moleculares secundários de relato obrigatório no contexto brasileiro.

Este trabalho também propiciou o pioneiro estudo da frequência de portadores em heterozigose de variantes associadas a doenças autossômicas recessivas e, ainda, possibilitou estimar a frequência combinada de doenças recessivas raras na população brasileira. Considerando-se nossos dados e a equação de Hardy-Weinberg, nós estimamos a frequência de doenças recessivas em nossa população ser ao redor de 26,39/10.000 indivíduos, ou 0,26% (24, 25). Surpreendentemente, estimativas muito próximas (~30/10.000 indivíduos) foram encontradas por diferentes centros de estudo do mundo, utilizando-se, muitas vezes, metodologias distintas (48-50). Esta congruência com a literatura valida, ao menos em parte, os nossos achados e nossa metodologia.

Os dados populacionais sobre as doenças recessivas são de suma importância, uma vez que muitas delas apresentam tratamento específico e são de investigação obrigatória na triagem neonatal (p.ex., fenilcetonúria). A literatura sobre frequências de portadores para doenças autossômicas recessivas é limitada na América Latina e em outras populações fora da América do Norte e Europa. Os estudos disponíveis geralmente estão restritos a doenças únicas ou grupos de doenças. Como não encontramos estudos abrangentes que abordem o status de portador para múltiplas doenças recessivas no Brasil ou em outros países da América Latina, esperamos que nosso estudo estimule outras pesquisas que visem entender melhor o impacto das doenças recessivas em nossos países em desenvolvimento.

Utilizamos, ainda, análise de segregação e outros dados moleculares para reclassificar um grupo de variantes raras como benignas (n = 51) ou provavelmente

benignas (n = 211) (37). A interpretação adequada de variantes raras é um passo crucial para um diagnóstico molecular adequado e, consequentemente, para o manejo clínico e aconselhamento genético. Portanto, os esforços para distinguir variantes patogênicas de variantes benignas raras são um passo fundamental no diagnóstico molecular, mas determinar quais variantes raras têm impacto clínico continua sendo um grande desafio. A literatura apresenta uma limitação importante quanto à descrição de variantes benignas ou provavelmente benignas, porque os estudos e revistas científicas geralmente apresentam um forte viés para resultados positivos. Acreditamos que esta abordagem é valiosa para identificar variantes benignas raras e melhorar a interpretação correta dos resultados genéticos.

Não foram encontrados estudos abrangentes que abordem a análise de segregação e classificação de variantes benignas ou provavelmente benignas no Brasil ou em outros países da América Latina. Esses países são geralmente sub-representados em bancos de dados internacionais e apresentam populações geneticamente heterogêneas, com importantes contribuições genéticas de ameríndios, afro-americanos e europeus ocidentais. Quanto mais estudos de sequenciamento em larga escala e de análise de segregação estiverem disponíveis em tais nações, maior será o conhecimento sobre as variantes benignas raras regionais.

6. CONCLUSÃO

Este estudo, pioneiro no Brasil, permitiu avaliar o impacto diagnóstico da utilização da técnica molecular de Sequenciamento de Exoma e vislumbrar detalhes importantes sobre os achados moleculares primários e secundários em pacientes com doenças raras em uma população muito sub-representada na literatura científica.

Primeiramente, observaram-se Achados Primários em 158 indivíduos, o que indica um rendimento diagnóstico geral de 31,6% (158/500) com o uso do sequenciamento do exoma no estudo de doenças raras. Ademais, 15,6% (78/500) desta coorte teve potencial direto de redirecionamento dos cuidados clínicos. As doenças mais frequentemente observadas foram: síndrome de Rett (n = 7); doenças associadas ao gene *POLG* (n = 4); doença de Niemann–Pick tipo C (n = 3); síndrome KBG (n = 3); doenças associadas ao gene *DEAF1* (n = 3); síndrome de Cowden (n = 3); síndrome de Noonan (n = 3). Já os Achados Secundários com impacto clínico relevante foram observados em 7,4% (n = 37) da coorte.

Evidenciou-se que a grande maioria indivíduos desta coorte apresentaram algum tipo de manifestação neurológica (67,2%) e os grupos de manifestações clínicas mais relevantes e suas correspondentes frequências são: transtornos do neurodesenvolvimento (47%), epilepsia (21%), outras manifestações neurológicas (50%), sindrômicos/malformativos (35%), doenças imunológicas/hematológicas (19%), alteração de crescimento (15%), consanguinidade (1%).

Observou-se, ainda, que 71,9% (230/320) dos indivíduos eram portadores em heterozigose de variantes associadas a doenças autossômicas recessivas. Considerando-se essas frequências de portadores (2pq), a equação de Hardy-Weinberg foi utilizada para estimar de forma aproximada a frequência populacional de doenças recessivas (q2) em 26,39/10.000. Ademais, usando-se dados moleculares, populacionais e estudo de segregação nos genitores, 51 variantes raras foram classificadas como benignas e 211 como provavelmente benignas.

O diagnóstico preciso das doenças raras é o primeiro passo para o acolhimento desses indivíduos no sistema de saúde e a adoção de terapêuticas mais precisas. Nosso desafio atual é aprofundar o conhecimento sobre variantes genéticas raras no Brasil, dimensionar melhor seus impactos na saúde e qualidade de vida e propor estratégias para disseminar este conhecimento entre profissionais da saúde.

7. CONSIDERAÇÕES FINAIS E SUGESTÕES PARA TRABALHOS FUTUROS

Apesar de o Sequenciamento do Exoma propiciar o diagnóstico preciso em número significativo dos indivíduos com doenças raras, a indefinição etiológica ainda persiste para a maioria. A incorporação de abordagens mais abrangentes e avançadas, como o Sequenciamento do Genoma, pode complementar a investigação do estudo do exoma, uma vez que o Sequenciamento do Genoma permite a análise de praticamente todo o conjunto de genes humanos e inclui regiões não codificantes, de modo a proporcionar uma visão mais completa das variantes genéticas associadas às doenças raras e permitir melhor precisão no estudo de alterações estruturais.

Outro tema relevante para futuros estudos é a integração de informações genômicas, clínicas e ambientais para personalizar o diagnóstico, tratamento e cuidados de saúde no contexto da medicina de precisão. Esta abordagem pode fornecer benefícios importantes para o manejo de doenças raras.

Ademais, a incorporação de estudos funcionais à rotina diagnóstica pode trazer importantes detalhes que auxiliem à compreensão do impacto biológico das variantes raras. Esses futuros estudos podem envolver técnicas como análise da expressão gênica, de proteômica e metabolômica. A integração dessas informações “multiômicas” pode fornecer uma compreensão mais abrangente das alterações moleculares subjacentes às doenças genéticas raras, melhorar a precisão diagnóstica, a seleção de terapias e o cuidado individualizado dos pacientes.

8. REFERÊNCIAS

1- Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013 Oct;14(10):681-91.

2- McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet. 2007 Apr;80(4):588-604.

3- Davies D. Annual Report of the Chief Medical Officer 2016 [Internet]. Langenberg C, editor. 2016 [Acesso em 1º de maio de 2023]. Disponível em: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/631043/CMO_annual_report_generation_genome.pdf

4- The cost of sequencing a human genome [Internet]. Genome.gov. [Acesso em 29 de setembro de 2021]. Disponível em: https://www.genome.gov/sequencingcosts

5- Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, *et al*. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med. 2014 Feb;16(2):176-82.

6- Quaio CR, Carvalho JF, da Silva CA, Bueno C, Brasil AS, Pereira AC, *et al*. Autoimmune disease and multiple autoantibodies in 42 patients with RASopathies. Am J Med Genet A. 2012 May;158A(5):1077-82.

7- Grinberg H, Quaio CR, Avila MS, Ferreira SM, Vieira ML, Benvenuti LA, *et al*. The first cardiac transplant experience in a patient with mucopolysaccharidosis. Cardiovasc Pathol. 2012;21(4):358-60.

8- Quaio CR, Grinberg H, Vieira ML, Paula AC, Leal GN, Gomy I, *et al*. Report of a Large Brazilian Family With a Very Attenuated Form of Hunter Syndrome (MPS II). JIMD Rep. 2012;4:125-8.

9- Quaio CR, Koda YK, Bertola DR, Sukalo M, Zenker M, Kim CA. Case report. Johanson-Blizzard syndrome: a report of gender-discordant twins with a novel UBR1 mutation. Genet Mol Res. 2014 Jun 9;13(2):4159-64.

10- Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, *et al*. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015 Jul;47(7):717-26.

11- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, *et al*. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013 Oct 17;369(16):1502-11.

12-Behjati S, Tarpey PS. What is next generation sequencing. Arch Dis Child Educ Pract Ed. 2013 Dec;98(6):236-8.

13-Vassy JL, Christensen KD, Schonman EF, Blout CL, Robinson JO, Krier JB, *et al*. The Impact of Whole-Genome Sequencing on the Primary Care and Outcomes of Healthy Adult Patients: A Pilot Randomized Trial. Ann Intern Med. 2017 Aug 1;167(3):159-69.

14-Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, *et al*. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol. 2011 Jul 25;12(7):R68.

15-Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.

16-Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24.

17-Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, *et al*. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013 Jul;15(7):565-74.

18-Sapp JC, Johnston JJ, Driscoll K, Heidlebaugh AR, Miren Sagardia A, Dogbe DN, *et al*. Evaluation of Recipients of Positive and Negative Secondary Findings Evaluations in a Hybrid CLIA-Research Sequencing Pilot. Am J Hum Genet. 2018 Sep 6;103(3):358-66.

19-Jain A, Gandhi S, Koshy R, Scaria V. Incidental and clinically actionable genetic variants in 1005 whole exomes and genomes from Qatar. Mol Genet Genomics. 2018 Aug;293(4):919-29.

20-Jang MA, Lee SH, Kim N, Ki CS. Frequency and spectrum of actionable pathogenic secondary findings in 196 Korean exomes. Genet Med. 2015 Dec;17(12):1007-11.

21-Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, *et al*. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 2015 Mar;25(3):305-15.

22-Quaio CRDC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Groenner Penna M, Perazzio SF, *et al*. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. Am J Med Genet C Semin Med Genet. 2020 Dec;184(4):955-64.

23-Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Vetrini F, *et al*. Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of

Severe Single-Gene Disorders and Effect on Medical Management. JAMA Pediatr. 2017 Dec 4;171(12):e173438.

24-Quaio CRDC, Chung CH, Perazzio SF, Dutra AP, Moreira CM, Filho GMN, *et al*. Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. Am J Med Genet C Semin Med Genet. 2021 Sep;187(3):364-72.

25-Quaio CRAC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, Kim CA. Frequency of carriers for rare metabolic diseases in a Brazilian cohort of 320 patients. Mol Biol Rep. 2022 May;49(5):3911-8.

26-Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, *et al*. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am J Hum Genet. 2016 Jan 7;98(1):149-64.

27-LaDuca H, Polley EC, Yussuf A, Hoang L, Gutierrez S, Hart SN, *et al*. A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. Genet Med. 2020 Feb;22(2):407-15.

28-Schorling DC, Kirschner J, Bönnemann CG. Congenital Muscular Dystrophies and Myopathies: An Overview and Update. Neuropediatrics. 2017 Aug;48(4):247-261. doi: 10.1055/s-0037-1604154. Epub 2017 Jul 2. PMID: 28669131.

29-Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol. 2017 Sep;18(9):533-47.

30-Mortier GR, Cohn DH, Cormier-Daire V, Hall C, Krakow D, Mundlos S, *et al*. Nosology and classification of genetic skeletal disorders: 2019 revision. Am J Med Genet A. 2019 Dec;179(12):2393-419.

31-Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, *et al*. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol. 2020 Jan;40(1):24-64.

32-Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, *et al*. Epilepsy-associated genes. Seizure. 2017 Jan;44:11-20.

33-Shearer AE, Hildebrand MS, Smith RJH. Hereditary Hearing Loss and Deafness Overview. 1999 Feb 14 [Updated 2017 Jul 27]. In: Adam MP, Ardinger HH, Pagon RA, *et al*., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2021. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK1434/

34-Fahim AT, Daiger SP, Weleber RG. Nonsyndromic Retinitis Pigmentosa Overview. 2000 Aug 4 [Updated 2017 Jan 19]. In: Adam MP, Ardinger HH, Pagon RA, *et al*., editors. GeneReviews® [Internet]. Seattle (WA): University of

Washington, Seattle; 1993-2021. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK1417/

35- Ferreira CR, van Karnebeek CDM, Vockley J, Blau N. A proposed nosology of inborn errors of metabolism. Genet Med. 2019 Jan;21(1):102-6.

36- Edwards AW. G. H. Hardy (1908) and Hardy-Weinberg equilibrium. Genetics. 2008 Jul;179(3):1143-50.

37- Quaio CRD'C, Ceroni JRM, Cervato MC, Thurow HS, Moreira CM, Trindade ACG, *et al*. Parental segregation study reveals rare benign and likely benign variants in a Brazilian cohort of rare diseases. Sci Rep. 2022 May 11;12(1):7764.

38- Quaio CRDC, Obando MJR, Perazzio SF, Dutra AP, Chung CH, Moreira CM, *et al*. Exome sequencing and targeted gene panels: a simulated comparison of diagnostic yield using data from 158 patients with rare diseases. Genet Mol Biol. 2021;44(4):20210061.

39- Quaio CRDC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, Kim CA. Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned. Sao Paulo Med J. 2022;140(5):734-6.

40- Quaio CRDC, Kim CA. Diagnóstico Molecular de Doenças Raras: Abrindo Portas para o Tratamento. Revista Bioética. Forthcoming 2023.

41- Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, *et al*. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am J Hum Genet. 2014 Jun 5;94(6):809-17.

42- Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, *et al*. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med. 2015 Jul;17(7):578-86.

43- Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, *et al*. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014 Nov 12;312(18):1880-7.

44- Tan TY, Dillon OJ, Stark Z, Schofield D, Alam K, Shrestha R, *et al*. Diagnostic Impact and Cost-effectiveness of Whole-Exome Sequencing for Ambulant Children With Suspected Monogenic Conditions. JAMA Pediatr. 2017 Sep 1;171(9):855-62.

45- Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Köster J, Marais A, *et al*. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet. 2017 Feb;25(2):176-82.

46- Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014 Nov 12;312(18):1870-9.

47- Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. 2010 May 14;86(5):749-64.

48- Henneman L, Borry P, Chokoshvili D, Cornel MC, van El CG, Forzano F, *et al.* Responsible implementation of expanded carrier screening. Eur J Hum Genet. 2016 Jun;24(6):e1-e12.

49- Sankaranarayanan K. Ionizing radiation and genetic risks IX. Estimates of the frequencies of mendelian diseases and spontaneous mutation rates in human populations: a 1998 perspective. Mutat Res. 1998 Sep;411(2):129-78.

50- Ropers HH. On the future of genetic risk assessment. J Community Genet. 2012 Jul;3(3):229-36.

ANEXO I – Achados Secundários

Tabela 5 – Achados Secundários do ACMG.

Doença e Número OMIM	Gene
Polipose adenomatosa familiar (OMIM 175100)	*APC*
Aneurisma aórtico familiar 4 (OMIM 132900)	*MYH11*
Aneurisma aórtico familiar 6 (OMIM 611788)	*ACTA2*
Cardiomiopatia ventricular direita arritmogênica 5 (OMIM 604400)	*TMEM43*
Cardiomiopatia ventricular direita arritmogênica 8 (OMIM 607450)	*DSP*
Cardiomiopatia ventricular direita arritmogênica 9 (OMIM 609040)	*PKP2*
Cardiomiopatia ventricular direita arritmogênica 10 (OMIM 610193)	*DSG2*
Cardiomiopatia ventricular direita arritmogênica 11 (OMIM 610476)	*DSC2*
Câncer de mama e ovário hereditário 1 (OMIM 604370)	*BRCA1*
Câncer de mama e ovário hereditário 2 (OMIM 612555)	*BRCA2*
Síndrome de Brugada 1 (OMIM 601144)	*SCN5A*
Taquicardia ventricular polimórfica catecolaminérgica (OMIM 604772)	*RYR2*
Cardiomiopatia dilatada 1A (OMIM 115200)	*LMNA* *MYBPC3*
Síndrome de Ehlers-Danlos tipo 4 (OMIM 130050)	*COL3A1*
Doença de Fabry (OMIM 301500)	*GLA*
Hipercolesterolemia familiar (OMIM 143890)	*APOB* *LDLR*
Cardiomioatia hipertrófica familiar 1 (OMIM 192600)	*MYH7*
Cardiomioatia hipertrófica familiar 3 (OMIM 115196)	*TPM1*
Cardiomioatia hipertrófica familiar 4 (OMIM 115197)	*MYBPC3*
Cardiomioatia hipertrófica familiar 6 (OMIM 600858)	*PRKAG2*
Cardiomioatia hipertrófica familiar 7 (OMIM 613690)	*TNNI3*
Cardiomioatia hipertrófica familiar 8 (OMIM 608751)	*MYL3*
Cardiomioatia hipertrófica familiar 10 (OMIM 608758)	*MYL2*
Cardiomioatia hipertrófica familiar 11 (OMIM 612098)	*ACTC1*
Carcinoma medular de tireoide familiar (OMIM 155240)	*RET*
Hipercolesterolemia autossômica domonante 3 (OMIM 603776)	*PCSK9*
Polipose juvenil (OMIM 174900)	*BMPR1A* *SMAD4*
Não-compactação ventricular esquerda 6 (OMIM 601494)	*TNNT2*
Síndrome Li-Fraumeni 1 (OMIM 151623)	*TP53*
Síndrome de Loeys-Dietz 1A (OMIM 609192)	*TGFBR1*
Síndrome de Loeys-Dietz 1B (OMIM 610168)	*TGFBR2*
Síndrome de Loeys-Dietz 2A (OMIM 608967)	*TGFBR1*
Síndrome de Loeys-Dietz 2B (OMIM 610380)	*TGFBR2*
Síndrome de Loeys-Dietz 3 (OMIM 613795)	*SMAD3*
Síndrome do QT longo 1 (OMIM 192500)	*KCNQ1*
Síndrome do QT longo 2 (OMIM 613688)	*KCNH2*
Síndrome do QT longo 3 (OMIM 603830)	*SCN5A*
Síndrome de Lynch (OMIM 120435)	*MLH1* *MSH2*

	MSH6 *PMS2*
Hipertermia maligna (OMIM 145600)	*RYR1* *CACNA1S*
Síndrome de Marfan (OMIM 154700)	*FBN1* *TGFBR1*
Neoplasia endócrina múltipla 1 (OMIM 131100)	*MEN1*
Neoplasia endócrina múltipla 2a (OMIM 171400)	*RET*
Neoplasia endócrina múltipla 2b (OMIM 162300)	*RET*
Polipose associada ao gene *MUTYH* (OMIM 608456)	*MUTYH*
Neurofibromatose tipo 2 (OMIM 101000)	*NF2*
Deficiência de ornitina carbamoiltransferase (OMIM 311250)	*OTC*
Paragangliomas 1 (OMIM 168000)	*SDHD*
Paragangliomas 2 (OMIM 601650)	*SDHAF2*
Paragangliomas 3 (OMIM 605373)	*SDHC*
Paragangliomas 4 (OMIM 115310)	*SDHB*
Síndrome Peutz-Jeghers (OMIM 175200)	*STK11*
Pilomatrixoma (OMIM 132600)	*MUTYH*
Síndrome de tumor-hamartoma associada ao *PTEN* (OMIM 153480)	*PTEN*
Retinoblastoma (OMIM 180200)	*RB1*
Esclerose tuberosa 1 (OMIM 191100)	*TSC1*
Esclerose tuberosa 2 (OMIM 613254)	*TSC2*
Síndrome de Von Hippel-Lindau (OMIM 193300)	*VHL*
Tumor de Wilms (OMIM 194070)	*WT1*
Doença de Wilson (OMIM 277900)	*ATP7B*

A tabela sumariza as doenças clínicas com conduta padronizada e que devem ser reportadas de acordo com o ACMG. A primeira coluna traz o nome da doença e seu número OMIM (“*Online Mendelian Inheritance in Man*”, catálogo online de traços e doenças mendelianas e acessível no endereço www.omim.org) e a segunda coluna traz o nome do gene associado à respectiva doença. Fonte: adaptado de Green RC, *et al.*, 2013 (17).

ANEXO II – Listas de publicações científicas

As tabelas, a seguir, enumeram as publicações produzidas durante o período do doutorado a partir dos dados obtidos diretamente desta tese (Tabela 6) e a partir de pesquisas realizadas em paralelo (Tabela 7).

Tabela 6 – Lista de publicações de artigos científicos produzidos com os dados obtidos diretamente da tese e durante o período do doutorado.

Publicação	PubMed ID	Fator de Impacto	Citações (Google Scholar)
Quaio CRDC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Groenner Penna M, Perazzio SF, et al. **Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases.** Am J Med Genet C Semin Med Genet. 2020 Dec;184(4):955-64.	33258288	3,359	26
Quaio CRDC, Chung CH, Perazzio SF, Dutra AP, Moreira CM, Filho GMN, et al. **Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients.** Am J Med Genet C Semin Med Genet. 2021 Sep;187(3):364-72.	34269512	3,359	7
Quaio CRDC, Obando MJR, Perazzio SF, Dutra AP, Chung CH, Moreira CM, et al. **Exome sequencing and targeted gene panels: a simulated comparison of diagnostic yield using data from 158 patients with rare diseases.** Genet Mol Biol. 2021;44(4):20210061.	34609444	2,635	3
Quaio CRAC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, Kim CA. **Frequency of carriers for rare metabolic diseases in a Brazilian cohort of 320 patients.** Mol Biol Rep. 2022 May;49(5):3911-8.	35229241	2,742	2
Quaio CRD'C, Ceroni JRM, Cervato MC, Thurow HS, Moreira CM, Trindade ACG, et al. **Parental segregation study reveals rare benign and likely benign variants in a Brazilian cohort of rare diseases.** Sci Rep. 2022 May 11;12(1):7764.	35546177	4,996	1
Quaio CRDC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, Kim CA. **Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned.** Sao Paulo Med J. 2022;140(5):734-6.	36102462	0,949	
Quaio CRDC, Kim CA. **Diagnóstico Molecular de Doenças Raras: Abrindo Portas para o Tratamento.** Forthcoming 2023.	-	-	

Fonte: Elaborado pelo autor.

Tabela 7 – Lista de publicações de artigos científicos produzidos a partir de pesquisas em paralelo ao desenvolvimento da tese e durante o período do doutorado.

Publicação	PubMed ID	Fator de Impacto	Citações
Montenegro MM, Quaio CR, Palmeira P, Gasparini Y, Rangel-Santos A, Damasceno J, et al. **Gene expression profile suggesting immunological dysregulation in two Brazilian Bloom's syndrome cases.** Mol Genet Genomic Med. 2020 Apr;8(4):e1133.	32073752	2,473	11
Tonholo Silva TY, Rosa ABR, Quaio CR, Verbeek D, Pedroso JL, Barsottini O. **Does SCA45 Cause Very Late-Onset Pure Cerebellar Ataxia**. Neurol Genet. 2021 Jun;7(3):e581.	33884300	3,663	3
Linnenkamp B, Girardi R, Rocha L, Yamamoto G, Ceroni JR, Mendes AEC, et al. **Vertebral segmentation defects in a Brazilian cohort: Clinical and molecular analysis focused on spondylocostal dysostosis.** Clin Genet. 2022 Apr;101(4):476-8.	35137400	4,296	
Stephan BO, Quaio CR, Spolador GM, de Paula AC, Curiati MA, Martins AM, et al. **Impact of ERT and follow-up of 17 patients from the same family with a mild form of MPS II.** Clinics (Sao Paulo). 2022;77:100082.	35882106	2,898	4
McGee SR, Rajamanickam S, Adhikari S, Falayi OC, Wilson TA, Shayota BJ, et al. **Expansion and mechanistic insights into de novo DEAF1 variants in DEAF1-associated neurodevelopmental disorders.** Hum Mol Genet. 2023 Jan 13;32(3):386-401.	35981081	5,121	
Montenegro MM, Camilotti D, Quaio CRDC, Gasparini Y, Zanardo ÉA, Rangel-Santos A, et al. **Expanding the Phenotype of 8p23.1 Deletion Syndrome: Eight New Cases Resembling the Clinical Spectrum of 22q11.2 Microdeletion.** J Pediatr. 2023 Jan;252:56-60.e2.	36067875	2,99	
Quaio CRAC, Coelho AVC, Moura LMS, Guedes RLM, Chen K, Ceroni JRM, et al. **Genomic study of nonsyndromic hearing loss in unaffected individuals: Frequency of pathogenic and likely pathogenic variants in a Brazilian cohort of 2,097 genomes.** Front Genet. 2022;13:921324.	36147510	4,772	
Quaio CRDC, Torres MA Torres, Centurião NF, Muto NH, Maciel GAR, Souza PK, et al. **Sex chromosome-gender phenotype discordance in a phenotypically-normal woman with normal gonadal function**. Mol Biol Rep. Forthcoming 2023.		2,800	
Quaio CRDC, Ceroni JRM, Pereira MA, Teixeira ACBT, Yamada RY, Cintra VP, et al. **The Hospital Israelita Albert Einstein Standards for Constitutional Sequence Variants Classification - version 2023**. Hum Genomics. Forthcoming 2023.		4,500	

Fonte: Elaborado pelo autor.

ANEXO III – Premiações e honrarias

A Tabela 8 abaixo enumera as premiações e honrarias recebidas durante o período do doutorado.

Tabela 8 – Lista de premiações e honrarias recebidas durante o período de desenvolvimento do doutorado.

Premiação	Evento/Instituição	Ano
Conference fellowship	European Society of Human Genetics 2021 Conference	2021
Top Cited Article 2020-2021	AMERICAN JOURNAL OF MEDICAL GENETICS PART C: SEMINARS IN MEDICAL GENETICS	2022
Prêmio Genética para Todos CBGM 2021: 3º Lugar	XXXII Congresso Brasileiro de Genética Médica	2021
Trabalho Científico Original – Pôster: 2º lugar	III CONGRESSO BRASILEIRO DE NEUROGENÉTICA	2023
Top Cited Article 2021-2022	AMERICAN JOURNAL OF MEDICAL GENETICS PART C: SEMINARS IN MEDICAL GENETICS	2023
Concurso de Artigos Científicos: 1º lugar	V Fórum de Doenças Raras do Conselho Federal de Medicina	2023

Fonte: Elaborado pelo autor.

ANEXO IV – Apresentações em congressos e eventos científicos

A Tabela 9 abaixo enumera as apresentações em congressos e eventos científicos durante o período do doutorado.

Tabela 9 – Lista de apresentações em congressos e eventos científicos durante o período do doutorado.

Título do trabalho	Categoria	Evento	Ano
Reports of incidental findings in a cohort of 500 exomes of Brazilian patients were common and show potential impact in life-expectancy	Pôster	European Human Genetics Conference 2020	2020
Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases	Pôster	2021 ACMG Annual Clinical Genetics Meeting Anais: *Molecular Genetics and Metabolism 132S1 (2021) S59–S198*	2021
Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned	Pôster	ASHG 2021 Virtual Meeting	2021
Reports of incidental findings in a cohort of 500 exomes of Brazilian patients were common and show potential impact in life-expectancy	Apresentação oral	II CONGRESSO BRASILEIRO DE NEUROGENÉTICA	2021
Diagnostic Power and Clinical Impact of Exome Sequencing in a Cohort of 500 Patients with Rare Diseases	Pôster	European Human Genetics Conference 2021	2021
Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients	Apresentação oral	14th International Congress of Inborn Errors of Metabolism	2021
Exome sequencing of 500 brazilian patients with rare diseases: what we have learned.	Pôster	66th Brazilian Congress of Genetics	2021
frequency of carriers of rare recessive mendelian diseases in a brazilian cohort of rare diseases	Apresentação oral	XXXII Congresso Brasileiro de Genética Médica	2021
Evaluation of rare variants of patients with neurodevelopmental disorders using a machine learning method reveals candidate digenic and oligogenic combinations	Pôster	XXIX Congresso Brasileiro de Neurologia	2021
Genomic study of non-syndromic hearing loss in unaffected individuals: frequency of pathogenic and likely pathogenic variants in a Brazilian cohort of 2097 genomes	Pôster	European Human Genetics Conference 2022	2022
Estudo genômico de perda auditiva não-sindrômica em indivíduos não afetados: frequência de variantes patogênicas e provavelmente patogênicas em coorte brasileira com 2.097 genomas	Pôster	III CONGRESSO BRASILEIRO DE NEUROGENÉTICA	2023

Fonte: Elaborado pelo autor.

ANEXO V – Íntegras dos artigos científicos

Seguem, a seguir, as íntegras das publicações produzidas durante o período do doutorado a partir dos dados obtidos diretamente desta tese seguindo a ordem da Tabela 6 (Anexo II).

Received: 9 October 2020 | Revised: 3 November 2020 | Accepted: 9 November 2020
DOI: 10.1002/ajmg.c.31860

RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART C Seminars in Medical Genetics WILEY

Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases

Caio Robledo D'Angioli Costa Quaio[1,2] | Caroline Monaco Moreira[1] | Gil Monteiro Novo-Filho[1] | Patricia Rossi Sacramento-Bobotis[1] | Michele Groenner Penna[1] | Sandro Felix Perazzio[1,3] | Aurelio Pimenta Dutra[1] | Rafael Alves da Silva[1,4] | Monize Nakamoto Provisor Santos[1] | Vanessa Yurie Nozaki de Arruda[1] | Vanessa Galdeno Freitas[1,5] | Vinícius Ceola Pereira[1] | Maria Carolina Pintao[1] | Alexandre Ricardo dos Santos Fornari[1] | Ana Lígia Buzolin[1] | Andre Yuji Oku[1] | Matheus Burger[1] | Rodrigo Fernandes Ramalho[1] | David Santos Marco Antonio[1] | Elisa Napolitano e Ferreira[1] | Otavio Jose Eulalio Pereira[1] | Vanessa Dionisio Cantagalli[1] | Ana Carolina Gomes Trindade[1] | Rafaela Rogerio Floriano de Sousa[1] | Cintia Reys Furuzawa[1] | Fernanda Verzini[1] | Shirley Dezan Matalhana[1] | Naiade Romano[1] | Daniele Paixão[1] | Caroline Olivati[1] | Gustavo Marquezani Spolador[1] | Gustavo Arantes Rosa Maciel[1,6] | Viviane Zorzanelli Rocha[1,7] | Javier Miguelez[1] | Mario Henrique Burlacchini de Carvalho[1,8] | Alexandre Wagner Silva de Souza[1,3] | Luis Eduardo Coelho Andrade[1,3] | Maria de Lourdes Chauffaille[1] | Aline dos Santos Borgo Perazzio[1] | Ana Lucia Pereira Monteiro Catelani[1] | Miguel Mitne-Neto[1] | Chong Ae Kim[2] | Wagner Antonio da Rosa Baratela[1]

[1]Fleury Medicina e Saúde, Grupo Fleury, São Paulo, SP, Brazil

[2]Instituto da Crianca (Children's Hospital), Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, Brazil

[3]Rheumatology Division, Department of Medicine, Escola Paulista de Medicina, Universidade Federal de Sao Paulo, São Paulo, SP, Brazil

[4]Escola Paulista de Medicina, Laboratório de Hepatologia Molecular Aplicada (LHeMA), Universidade Federal de São Paulo, São Paulo, SP, Brazil

[5]Instituto de Matemática e Estatística da Universidade de São Paulo e Instituto de Ensino e Pesquisa do Hospital Sírio Libanês, São Paulo, SP, Brazil

[6]Discipline of Gynecology, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, São Paulo, SP, Brazil

[7]Heart Institute (InCor), Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, São Paulo, SP, Brazil

[8]Disciplina de Obstetrícia, Departamento de Obstetrícia e Ginecologia, Faculdade de Medicina FMUSP, Universidade de Sao Paulo, São Paulo, SP, Brazil

Correspondence
Caio RDAC Quaio, Av. Gen. Valdomiro de Lima, 508 – Jabaquara. São Paulo, SP, 04344-070. Brazil.
Email: caio.quaio@grupofleury.com.br

Abstract

Rare diseases comprise a diverse group of conditions, most of which involve genetic causes. We describe the variable spectrum of findings and clinical impacts of exome sequencing (ES) in a cohort of 500 patients with rare diseases. In total, 164 primary findings were reported in 158 patients, representing an overall diagnostic yield of 31.6%. Most of the findings (61.6%) corresponded to autosomal dominant conditions,

 © 2020 Wiley Periodicals LLC

followed by autosomal recessive (25.6%) and X-linked (12.8%) conditions. These patients harbored 195 variants, among which 43.6% are novel in the literature. The rate of molecular diagnosis was considerably higher for prenatal samples (67%; 4/6), younger children (44%; 24/55), consanguinity (50%; 3/6), gastrointestinal/liver disease (44%; 16/36) and syndromic/malformative conditions (41%; 72/175). For 15.6% of the cohort patients, we observed a direct potential for the redirection of care with targeted therapy, tumor screening, medication adjustment and monitoring for disease-specific complications. Secondary findings were reported in 37 patients (7.4%). Based on cost-effectiveness studies in the literature, we speculate that the reports of secondary findings may influence an increase of 123.2 years in the life expectancy for our cohort, or 0.246 years/cohort patient. ES is a powerful method to identify the molecular bases of monogenic disorders and redirect clinical care.

KEYWORDS

diagnostic yield, exome sequencing, incidental findings, rare diseases, secondary findings

1 | INTRODUCTION

Rare diseases comprise a large and diverse group of approximately 7,000 different conditions (Boycott et al., 2013; McKusick, 2007), most of which involve genetic causes. Etiology determination may avoid further diagnostic procedures in the known "diagnostic odyssey" context of these patients and impact the clinical approach, health system costs, anticipatory guidance, prognosis and genetic counseling of individuals and their families. However, a precise diagnosis may be a challenge because of the following: (a) clinical-based identification is very difficult if most of these conditions present common, imprecise or even overlapping characteristics, (b) diagnosis primarily depends on understanding the underlying molecular mechanisms, (c) molecular mechanisms present high complexity, and (d) the number of genes associated with newly discovered conditions grows continuously.

Genomic studies with exome sequencing (ES), which, in turn, is based on next-generation sequencing (NGS) technology, is a cost-effective alternative to determine the molecular diagnosis of patients with rare diseases because massively parallel sequencing of almost all coding regions of the human genome allows the concomitant study of all known genes associated with genetic conditions. Nevertheless, this approach may also unintentionally unravel occult deleterious genetic variants besides the primary objective of analysis, although associated with clinically actionable conditions and defined as "incidental and secondary findings". The American College of Medical Genetics and Genomics (ACMG) has published recommendations for reporting secondary findings in selective genes (Green et al., 2013; Kalia et al., 2017).

Herein, we describe the variable spectrum of molecular findings, diagnostic yield and potential clinical impact of genomic testing with ES in a heterogeneous cohort of 500 patients with rare diseases. This study aimed to investigate the clinical usefulness of this approach in diagnosis and estimate its impact on the outcome of patients with rare diseases.

2 | METHOD

2.1 | Selection of cases

The exome sequencing (ES) findings of the first 500 samples of adult, pediatric and fetal patients obtained from our service were reviewed. All the samples were collected from 2016 to 2020 in facilities of the Fleury Group, a tertiary private general diagnostic laboratory with subsidiaries in nine states of Brazil and based in São Paulo, Brazil. This study was granted ethics committee approval from both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

The inclusion criteria were restricted to symptomatic patients who had undergone molecular investigation for suspected diseases of genetic etiology. All patients or their legal guardians provided consent before exome analysis. Patients who refused to share genomic data or receive information on medically actionable findings were excluded from this cohort. Additionally, exome analysis performed for other reasons than clinical diagnostic (such as research protocols or prenatal genetic counseling) were excluded from our review. Thirty-one cases that did not meet the inclusion criteria were excluded.

The clinical data were collected through a comprehensive pretest form completed by the attending physician or family, medical reports and clinical notes provided to the laboratory. Patients were not evaluated by specialists from our team. The clinical features were divided as follows: (a) neurodevelopmental disorders (including autism, intellectual disability, and behavior anomaly), (b) seizure (including epilepsy, spasms, and encephalopathy), (c) other neurological anomaly (including hypotonia, hypertonia, neuromuscular disease, hearing loss, visual impairment, and ataxia), (d) syndromic/malformative conditions (including microcephaly, macrocephaly, skeletal anomalies, and malformations), (e) growth anomaly (including intrauterine growth restriction, short stature, tall stature, failure to thrive, and obesity), (f)

immune or hematological anomalies (including primary immunodeficiency, recurrent infections, severe multisystemic autoimmunity, allergy, cytopenias, and coagulation defects), (g) cancer/tumors, (h) cardiovascular disease (including cardiomyopathies, arrhythmia, cardiovascular malformations, stroke, and myocardial infarction), (i) gastrointestinal/liver disease, (j) consanguinity, (k) metabolic/hormonal anomaly, and (l) other. The same patient may be included in more than one group.

2.2 | Molecular analysis, bioinformatics and variant classification

DNA from the proband and both parents were extracted in a clinical setting from peripheral blood leukocytes, saliva or prenatal samples of villus biopsy or amniotic fluid after appropriate cell culture. Exome capture was also conducted in a clinical setting using Agilent Clinical Research Exome v1 according to the manufacturer's instructions. Sequencing was performed using an Illumina NextSeq platform. Exome data were aligned to the GRCh37.75/hg19 reference genome using the Burrows–Wheeler Aligner (BWA; version 0.7. 17-r1188). Variants (single-nucleotide variants [SNVs] and indels) were identified following the best practices of the Broad Institute using the Genome Analysis ToolKit (GATK, version 3. 8-0-ge9d806836) software and annotated using Variant Effect Predictor (VEP, version 88.14). All the exomes met a minimum of 95% of target bases covered at >10×. In-house bioinformatic pipelines were developed by a dedicated team of specialists. The mitochondrial genome and copy number variants were not studied.

At least two parallel analyses were performed to preselect variants considering an allele frequency < 1%, functional impact, clinical relevance of gene harboring the variant, relevant reports from databases (e.g., ClinVar and HGMD) and the literature. These preselected variants were then discussed in a board comprising three experts. After this first board meeting, selected variants would undergo Sanger confirmation in the proband and inheritance pattern determination when the parental samples were available. Paternity and maternity were not confirmed by any specific test in our cohort, although non-paternity/nonmaternity could be disclosed by genotyping multiple rare variants. Variants were classified according to ACMG guidelines(Richards et al., 2015) with assistance of a third-party ACMG calculator by Saphetor SA (Lausanne, Switzerland) (www.varsome.com). Internal adaptations of ACMG guidelines used in this work may be found in Supplementary Material.

Variants that did not match the expected segregation pattern were reported as benign or VUS in a separate table of "Additional Findings". The carrier status for pathogenic or likely pathogenic variants of autosomal recessive conditions was also reported. Nevertheless, additional findings and carrier status reports were not evaluated in this study and will not be discussed.

Primary findings were reported when pathogenic or likely pathogenic variants were observed in a gene that was associated with the patient's phenotype, with compatible zygosity and an adequate inheritance pattern; these cases were considered positive. Two variants of unknown significance (VUS) found in "*trans*" with a second pathogenic variant were exceptionally reported as primary findings in two patients considered to harbor autosomal recessive conditions (cases 196 and 222) because they were unanimously judged to explain the clinical phenotype by three experts in the area.

Secondary findings (conditions unrelated to the indication for testing that might impact the health or quality of life of patients or other family members) were reported for pathogenic and likely pathogenic variants in the ACMG gene list and exceptionally in other genes determined to be clinically relevant by our team of specialists (Green et al., 2013; Kalia et al., 2017). These findings were neither confirmed by Sanger sequencing nor had determined inheritance patterns. Our laboratory also reported the carrier status for genes in the ACMG list because we judged it to be important for parental investigation and genetic counseling.

The samples were analyzed by three distinct protocols as follows: Protocol 1 (cases 1–289; n = 289), in which the samples were processed in our laboratory as described above and variants were interpreted using an in-house web interface (GTAC); Protocol 2 (cases 290–387; n = 98), in which the samples were processed in our laboratory as described above and variants were analyzed using a commercially available diagnostic decision support platform by Emedgene Technologies LTD (Tel-Aviv, Israel) (www.emedgene.com); Protocol 3 (cases 388–500; n = 113), in which the samples were sent to Centogene AG (Rostock, Germany) (www.centogene.com) and processed according to the protocols of this third-party laboratory, including DNA extraction, bioinformatic pipeline, CNV analysis, and variant analysis. For Protocol 3, the primary findings reported by Centogene were confirmed by in-house Sanger sequencing of the proband and genitors to determine the inheritance pattern.

2.3 | Estimation of the clinical impact of ES

Inspired by the previous work of Meng et al., (2017) in a cohort of infants in intensive care units and after reviewing our first 500 exomes, we gathered a meeting of several specialists from different areas of medical knowledge to estimate more objectively the clinical impact of the ES primary findings on medical management. Clinical expertise, OMIM entrance and literature search were used to estimate the potential clinical impact in the following areas: targeted therapy, tumor screening, adoption of standardized clinical guidelines, medication adjustment and monitoring for disease-specific complications.

2.4 | Statistical analysis

Chi square test was used to clarify whether a difference exists in the proportions of diagnoses and clinical features among the three groups of different protocols.

TABLE 1 Distribution of demographic data, clinical features and diagnostic yield (columns) per age group (rows)

Age	Number of cases	Reported sex (male/female)	Mean age	Clinical feature: Neurodevelopmental disorders	Seizure	Other neurological anomaly	Syndromic/ malformative	Growth anomaly	Immune/ hematological diseases	Cancer/ tumors	Cardiovascular disease	Gastrointestinal/ liver disease	Consanguinity	Metabolic/ hormonal anomaly	Other	Diagnostic yield per age group
Prenatal samples	6	n.d.	n.d.	0%	0%	0%	100%	0%	0%	0%	17%	0%	0%	0%	0%	67%
<1 y	55	58.2% / 41.8%	5 m	38%	36%	56%	47%	22%	7%	2%	18%	20%	2%	15%	2%	44%
1–5 y	161	61.5% / 38.5%	2y9m	70%	23%	55%	49%	19%	16%	1%	9%	6%	1%	9%	1%	32%
5–10 y	78	66.7% / 33.3%	7y	72%	29%	49%	28%	13%	10%	1%	10%	5%	0%	6%	3%	32%
10-18 y	63	54% / 46%	13y6m	43%	24%	48%	37%	19%	21%	3%	11%	11%	2%	17%	3%	30%
18-25 y	26	57.7% / 42.3%	21y2m	23%	15%	46%	23%	15%	31%	4%	4%	8%	0%	12%	12%	27%
25–50 y	66	59% / 41%	36y10m	17%	6%	45%	15%	5%	38%	15%	6%	5%	3%	14%	5%	32%
>50 y	45	60% / 40%	61y6m	0%	4%	49%	7%	7%	22%	18%	13%	0%	0%	9%	7%	13%
All age groups	500	55.3% / 44.7%	15y5 m	47%	21%	50%	35%	15%	19%	5%	10%	7%	1%	11%	3%	32%
Diagnostic yield per clinical feature				35%	34%	35%	41%	35%	19%	32%	31%	44%	50%	33%	13%	

Note: The data from all 500 patients are shown in the row "All age groups". The first column "Age" contains the age groups in which the patients were divided; group "<1 y" represents samples collected after birth to less than 1 year of age (not inclusive); the group ">50 y" includes patients 50 years of age and older; the other groups include the first number of the range and does not include the last—e.g., "1–5 y" includes children from 1 year to younger than 5 years. A predominance of pediatric patients (71.4% under 18 y) was noted. The second column "Number of cases" presents the absolute number of individuals per age group. The third column shows the distribution of gender according to age group. Note the predominance of males ("M") in all the pediatric groups (<18 y), while females ("F") are predominant in all the adult groups. The last column presents the diagnostic yield per age group. Note that the highest yield was observed for prenatal cases (67%), followed by children less than 1 year of age (44%), while the lowest yield was observed for adults 50 years or older (13%). All other age groups performed closely from 27% through 32%. The last row presents the diagnostic yield per group of clinical features. Note that the group with the highest diagnostic rate was consanguinity (50%), followed by gastrointestinal/liver disease (44%) and syndromic/malformative conditions (41%), while the groups with the lowest yield were other (13%) and immune/hematological diseases (19%); all other groups varied from 31% to 35% yield.

3 | RESULTS

3.1 | Cohort characteristics

Our cohort comprised 500 symptomatic patients who had undergone exome sequencing for diagnostic purposes. The demographic data, clinical features and their distribution across age groups are shown in Table 1. For the prenatal samples (cases 179, 232, 260, 354, 427 and 443), gender was not determined and they were not included in the calculation of the average age. Individual data for each case (case overview), along with the molecular data (primary findings and secondary findings), can be accessed in the Supplementary Material. Although neurological cases were predominant (67.2%), the clinical indications for exome sequencing included a wide range of clinical concerns and their overall frequencies and distribution across age groups are shown in Table 1.

The distribution of clinical features in Protocols 1, 2 and 3 showed a significant difference for two categories: the frequencies of other neurological anomalies were 49.8%, 40.8% and 59.3%, respectively (X^2 [1, N = 500] = 7.2044, p .027264), while those for cancer/tumors were 4.2%, 10.2% and 2.7%, respectively (X^2 [1, N = 500] = 7.3331, p .025564). No differences for other clinical features were observed (Supplementary Material; Comparison of Protocols).

3.2 | Primary findings and diagnostic yield

Primary findings were reported in 158 patients, with an overall diagnostic yield of 31.6%. Table 1 shows the diagnostic rate per age group (last column) and clinical feature (last row). The diagnostic yield differed nominally for Protocols 1, 2, and 3:28.7% (83/289), 40.8% (40/98) and 31% (35/113), respectively. However, Protocol 2 showed a significant difference compared with Protocol 1 (X^2 [1, N = 387] = 2.615, p = .105) and Protocol 3 (X^2 [1, N = 211] = 1.0513, p = .305).

In total, 101 autosomal dominant (AD) conditions were observed in 98 individuals; 42 autosomal recessive (AR) conditions were observed in 42 individuals, and 21 X-linked (XL) conditions were observed in 21 individuals. Additionally, a dual molecular diagnosis was found in six patients: three presented two AD conditions (cases 167, 382 and 420), one presented AD and AR (case 126), one AD and XL (case 226), and one AR and XL (case 139).

These 158 positive cases harbored 195 variants reported as primary findings, among which 109 were classified as pathogenic, 84 as likely pathogenic and two as variants of unknown significance (VUS). Both VUSs were reported as primary findings because they were judged to explain the clinical phenotype once both were found in "*trans*" with a second pathogenic variant in two patients (cases 196 and 222) with rare autosomal recessive conditions.

The 158 positive cases harbored 195 variants reported as primary findings and were classified as follows: 55.9% were pathogenic (n = 109), 43.1% likely pathogenic (n = 84) and 0.4% VUS (n = 2). The same two variants observed in *cis* in *POLG* recurred in three patients (cases 117, 120 and 446), and this haplotype was considered associated with the AD form of involvement of this gene. One variant in *MEFV* recurred in two cases (cases 168 and 233), and one variant in *TNFRSF13B* recurred in three patients (cases 247, 451 and 464). The other 184 variants were seen only once.

We found no reports in the literature for 43.6% (n = 85) variants, with 37.9% (n = 74) not reported in the ClinVar Database. Regarding the 101 AD diseases, 68.3% (n = 69) were associated with heterozygous variants that occurred de novo (maternity and paternity not confirmed), while 13.9% (n = 14) with heterozygous variants inherited from parents. Six variants corresponded to the *POLG* haplotype observed in three patients (3%), and segregation was not studied for 15 (14.9%) variants because the parental sample was not available. Considering the 42 AR diseases, 14 (33.3%) cases presented one variant in homozygosity, while 23 (54.8%) presented two heterozygous variants in *trans* and five patients (11.9%) presented two heterozygous variants in an unknown phase (segregation was not studied).

For 21 XL diseases, 10 (47.6%) variants were found in hemizygosity and 11 (52.4%) in heterozygosity, revealing a slight predominance of female patients. Segregation of this group revealed the following: 10 (47.6%) were de novo, six (28.6%) were inherited from the mother, and segregation was not determined for five (23.8%).

Of the 158 patients with a positive diagnosis, 109 had genetic disorders seen only once in this study. However, 49 patients had a recurrence of diagnosis. Of the diagnosed cases, X-linked *MECP2*-related Rett syndrome and correlated disorders (*MECP2*, OMIM: 300005) were observed in seven patients (cases 416, 121, 22, 78, 136, 140, and 226) and comprised the most frequent single-gene disorder in our cohort. The following conditions recurred in three or more patients: four patients (117, 120, 446, and 211) presented with *POLG*-associated conditions—three with an autosomal dominant form and one with an autosomal recessive biallelic loss of function (*POLG*, OMIM: 174763); three (380, 489, and 323) with autosomal recessive Niemann–Pick disease type C (*NPC1*, OMIM: 607623); three patients (244, 294, and 376) with autosomal dominant KBG syndrome (*ANKRD11*, OMIM: 611192); three patients (99, 405, and 454) with *DEAF1*-associated conditions—two with autosomal dominant gain of function/dominant negative and one with autosomal recessive biallelic loss of function (*DEAF1*, OMIM: 602635); three (356, 375, and 194) with autosomal dominant Cowden syndrome (*PTEN*, OMIM: 158350); and three (206, 207, 427) with autosomal dominant *PTPN11*-related Noonan syndrome (*PTPN11*, OMIM: 176876).

Some examples of atypical presentation of rare genetic disorders are as follows: 1) four pediatric patients with very similar clinical presentation of rapidly progressive, severe hepatic failure were confirmed to have very different genetic conditions: progressive familial intrahepatic cholestasis (PFIC) type 3 was confirmed in two cases (cases 64 and 330), congenital bile acid synthesis defect in one (case 94) and mitochondrial DNA depletion syndrome in another one (case 424); 2) we identified a rare congenital defect of glycosylation (CDG type IId) in a young girl with almost exclusive myopathic manifestation (case 222); 3) we confirmed X-linked Borjeson–Forssman–Lehmann syndrome presenting with neurodevelopmental delay, typical

dysmorphic features (observed only after molecular diagnosis, with a reverse phenotyping approach) and spots following Blaschko lines suggesting skewed X inactivation in a young girl (case 110); 4) two rare immune deficiencies caused by genes of the STAT family were observed: *STAT1* (case 199) and *STAT3* (case 218); 5) two male infants with severe encephalopathy harboring hemizygous *MECP2* variants (cases 121 and 416); 6) a patient with central nervous system and skin manifestation indicating asparagine synthetase deficiency (case 460); among others.

3.3 | Estimation of the clinical impact of ES

We estimate that 78 patients, representing 15.6% of cohort patients or 49.4% of positive cases, would benefit from the redirection of care in at least one area of medical management. Individual data are available in Supplementary Material (Primary Findings).

We observed the eligibility for targeted therapies or diets in 3.2% (n = 16) of cases: riboflavin supplementation for Brown–Vialetto–Van Laere syndrome (case 4), a ketogenic diet for GLUT1-deficiency (cases 87 and 114), the use of oral cholic acid for bile acid synthesis defect (case 94), anti-IL-1 and colchicine for autoinflammatory diseases (cases 168 and 233), Jak1/2 inhibitor for *STAT1* gain-of-function (case 199), anti-IL-6 (tocilizumab) for *STAT3* gain-of-function (case 218), enzyme replacement therapy or substrate reduction therapy for Niemann–Pick type C (cases 323, 380, and 489), CTLA-4-Ig (abatacept) for *CTLA4* haploinsufficiency (case 422), immunoglobulin replacement for agammaglobulinemia (case 415) and common variable immunodeficiency (cases 247, 451 and 464).

We noticed a potential benefit from the established guidelines of tumor screening for 1.6% (n = 8) of cases: Bloom's syndrome (case 41), Simpson–Golabi–Behmel syndrome (case 248), Cowden syndrome (cases 194, 356, and 375), Lynch syndrome (case 325), Von Hippel–Lindau syndrome (case 332) and hereditary paraganglioma-pheochromocytoma (case 349). Additionally, well-established guidelines for standards of care were indicated for 0.8% (n = 4) of cases: Marfan syndrome (case 129), Loeys–Dietz syndrome (case 119) and glycogen storage diseases (cases 106 and 150).

We also observed a benefit from medication optimization and adjustments in 4.4% (n = 22) of cases: anticoagulation titration for hereditary thrombophilia (case 167), tetrabenazine use for *NKX2*-1-related chorea (case 109), fresh frozen plasma or purified apotransferrin infusions for atransferrinemia (case 77), prophylaxis of alternating hemiplegia of childhood (case 148), acetazolamide for *CACNA1A*-related conditions (case 420), use of sodium channel blockers and avoidance of muscle relaxants in myotonia congenita (case 35), anticonvulsants for *PRRT2*-associated paroxysmal movement disorders (case 344), specific management for hereditary hemophagocytic lymphohistiocytosis including hematopoietic stem-cell transplantation (cases 280 and 320), therapy guidance for some epileptic encephalopathies (8 cases) and five patients with neuromuscular diseases would benefit from avoidance of specific medications (such as statins and neuromuscular blockers) that may worsen symptoms: dysferlinopathy (case 126), *ANO5*-related myopathy (case 263), *COL6A1*-related myopathy (case 155) and central core disease (cases 146 and 212).

For 5.6% (n = 28) of cases, we judged that the diagnosis would be crucial to monitor several disease-specific complications: calcium anomalies in Kenny–Caffey syndrome (case 37), hepatic and renal alterations in familial visceral amyloidosis (case 316), neurological deterioration in GM1 gangliosidosis (case 359), ophthalmologic care for incontinentia pigmenti/NEMO deficiency (case 391), apnea and arrhythmia in Rett syndrome and *ASCL1*-related condition (cases 22, 78, 121, 136, 140, 226, and 416), heart complications for Ogden syndrome (case 462), neurological complications in asparagine synthetase deficiency (case 460), immune complications in Okur–Chung syndrome (case 8), multisystemic complications for mitochondrial diseases (cases 44, 95, 117, 120, 134, 175, 211, 424, 425, 446, and 482) and orthopedic complications for skeletal dysplasias, including chondrodysplasia punctata (case 386) and *COL2A1*-related skeletal dysplasias (cases 426 and 439).

3.4 | Secondary findings

Twenty-eight different heterozygous variants occurring in 18 genes were reported as secondary findings in 37 patients (7.4%). Individual data for secondary findings are available in Supplementary Material (Secondary Findings). All 37 patients presented only one heterozygous variant. Two *ATP7B* variants and two *MUTYH* variants recurred in five and eight patients, respectively. Overall, 13 patients (2.6%) harbored pathogenic or likely pathogenic variants associated with autosomal dominant conditions on the ACMG list (associated with diseases involving the categories arrhythmia, cardiomyopathy, familiar cancer, and metabolic disease), three (0.6%) harbored variants in genes associated with autosomal dominant cancer-predisposing conditions (familial cancer) not on the ACMG list (*CHEK2*, *RAD51C* and *BRIP1*), and twenty-one patients (4.2%) were carriers for autosomal recessive conditions on the ACMG list.

4 | DISCUSSION

Our findings show evidence of a clinical benefit of ES as a diagnostic tool for rare diseases because they reveal an overall diagnostic yield of 31.6% and the potential to redirect care in 15.6% among a heterogeneous cohort of 500 symptomatic Brazilian patients. Diagnostic rates between 22% and 30% have been reported previously in other studies with heterogeneous indications, showing that our overall rate agrees with that in the literature (Beaulieu et al., 2014; Farwell et al., 2015; Lee et al., 2014; Tan et al., 2017; Taylor et al., 2015; Trujillano et al., 2017; Yang et al., 2013; Yang et al., 2014). The diagnostic yield for ES varies for several reasons, including pretest evaluation (clinical evaluation, phenotyping, and proper indication for ES) and the complex steps of NGS analysis (exome capture, gene coverage, bioinformatics protocols, and medical review). As a clinical

diagnostic laboratory with a restricted clinician-laboratory partnership, pretest evaluation may have played an important role in our results.

We observed a nominally higher, but not statistically relevant, rate of diagnosis (40.8%) of patients who had undergone Protocol 2, in which we used a third-party analysis tool. We believe that improvements in technology, data mining, machine learning and automation of medical literature search will make NGS analysis less dependent on human bias and will eventually increase its diagnostic power. An increase in complexity for developing such tools will make NGS diagnostic services more dependent on integrated analysis solutions developed by specialized biotech companies. Despite the shorter turnaround time in Protocol 3, the discouraging results did not demonstrate advantages in the diagnostic rate over the earliest in-house Protocol 1.

The rate of positive findings for ES in our study and similar studies was shown to be superior to that of other molecular diagnostic approaches for rare diseases compared with the diagnostic rates of chromosomal microarray and classic cytogenetics (Miller et al., 2010). This fact, in addition to upcoming new protocols to study copy-number variants by NGS, has made ES become the first-line molecular investigation for patients with suspected genetic disorders in several services. Higher rates of molecular diagnosis in prenatal samples and children aged <1 y have also been observed in other NGS studies in pediatric populations (Farwell et al., 2015; Meng et al., 2017; Tan et al., 2017). Additionally, higher rates for patients with consanguinity and syndromic/malformative features (41%) have also been confirmed in other studies (Farwell et al., 2015; Meng et al., 2017; Tan et al., 2017; Trujillano et al., 2017).

Among autosomal dominant disorders, most cases (68.3%) resulted from de novo mutations. Without segregation studies, at least 13 of these variants would be relegated from probably pathogenic to VUS according to ACMG criteria, negatively impacting our overall diagnostic rate from 31.6% to 29%. This finding emphasizes the importance of segregational studies to confirm the molecular diagnosis of relevant numbers of individuals with AD conditions.

For autosomal recessive diseases, exactly one-third comprised homozygous variants and two-thirds harbored two heterozygous variants. Even with a relatively high frequency of homozygous variants, only 1% of our patients were reported to be consanguineous. We wondered whether consanguinity may have been underreported or whether the Brazilian ethnic background may have played a role, particularly if we consider that some homozygous variants do not have representative frequencies in population databases.

We believe that most of the patients and their relatives who had their etiologic diagnosis determined by ES had a direct impact on prognostic expectations, genetic counseling and reproductive decisions. Moreover, although imprecise, we objectively estimated the overall ES clinical impact rate as 15.6% (n = 78) by quantifying the potential benefit of specific clinical strata (targeted therapy, tumor screening, adoption of standardized clinical guidelines, medication adjustment and disease-specific complication monitoring) and excluding cases in which clinical management would not be overtly changed by uncovering genetic etiology (e.g., most neurodevelopmental, syndromic and malformative conditions). Noteworthy, we decided to obtain conservative but reliable data, an approach that may have underestimated the clinical benefit for some groups.

Recent advances in science have brought rising possibilities of targeted therapies, including pathophysiologically driven medications or dietary management with a direct impact on clinical follow up and the improvement of prognosis. In 3.2% of our cohort patients, we judged that the adoption of targeted therapies would critically impact the disease course. Although this group represents a small subset of patients who had undergone ES, we believe that this group has the potential to increase with upcoming years.

We found evidence of a benefit for tumor screening for 1.6% of cases. Interestingly, only half of these patients presented hereditary cancer investigation as the primary indication for ES. The others represented mainly syndromic/neurodevelopmental disorders (Bloom's, Simpson–Golabi–Behmel, and Cowden syndromes) with an increased risk for tumors, except for case 349 (hereditary pheochromocytoma), in which a cardiovascular event was the primary indication for ES.

An additional 0.8% would benefit from the adoption of well-established clinical guidelines found in the literature. This group presented a more classical diagnosis of dysmorphology (Marfan and Loeys–Dietz syndrome) and metabolic diseases (glycogen storage diseases) in which the diagnosis could be established solely by clinical findings. Because we do not have information about clinical diagnosis before ES, we assume that our molecular diagnosis may be a path for the adoption of adequate follow up.

We also estimate a benefit from medication optimization or adjustments for 4.4% of cases. This group is very heterogeneous and includes several conditions in which specific medications tailored by molecular diagnosis may bring clinical benefit and neuromuscular disorders that may worsen following the use of specific medications.

For 5.6% of cases, we judged that the diagnosis would be crucial to monitor several disease-specific complications. This is also a very heterogeneous group of diagnoses and includes some life-threatening complications that could cause sudden death or life-long complications, such as hypoventilation, arrhythmia and deformities.

Secondary findings occurring in 18 genes were reported in 37 patients (7.4%). This frequency is considerably higher than that in other studies in the literature, particularly compared with the 1% rate of ClinSeq participants (Johnston et al., 2012) and 1.2–3.4% rate for the NHLBI-ESP cohort (Dorschner et al., 2013), but slightly lower than the 8% rate reported by The NIH Undiagnosed Diseases Program (Lawrence et al., 2014). Although we cannot rule out that Brazilian ethnic background might have influenced this divergence, we believe that two main factors may explain this difference: our laboratory considered the carrier status for secondary findings to be reportable (4.2% were carriers for autosomal recessive conditions on the ACMG list). Additionally, we reported some conditions not enlisted by ACMG because they were judged to critically impact clinical care (0.6% harbored variants in genes associated with autosomal dominant cancer-predisposing conditions not on the ACMG list: *CHEK2*, *RAD51C* and *BRIP1*). The remaining (2.6%) are those who harbored pathogenic or likely pathogenic variants associated with autosomal dominant

conditions on the ACMG list, and this number agrees with the NHLBI-ESP cohort estimate.

Regarding carriers for autosomal recessive conditions on the ACMG list, seven patients presented five different heterozygous pathogenic variants in *ATP7B*, two of which recurred in more than one patient. Additionally, one patient presented a pathogenic heterozygous variant in *LDLR* and 13 presented seven different heterozygous variants in *MUTYH*, two of which recurred more than once. Although among these 21 patients the presence of a sole pathogenic variant is not sufficient to confirm the diagnosis of its associated autosomal recessive condition, we cannot rule out the presence of a second pathogenic variant not found by ES, such as intronic or regulatory-site variants. Because these conditions are relatively common, these findings might also uncover the possible diagnosis of a parent that might carry a second pathogenic variant, bringing a benefit beyond the individual tested. These reports may also impact genetic counseling, particularly for consanguineous couples.

Despite ACMG recommendations for secondary findings (Green et al., 2013; Kalia et al., 2017), it remains controversial and quite a challenge to determine the reportable variants. Some important limitations of the secondary findings in our study must be discussed. We did not confirm the presence or segregation pattern of these variants by Sanger sequencing, although the alignments of their reads were checked using Integrative Genome Viewer (IGV). Additionally, we did not perform a thorough clinical evaluation to determine whether patients or their relatives presented any clinical evidence of disease or were asymptomatic carriers once several conditions (e.g., long QT, cardiomyopathies) presented incomplete penetrance. Additionally, and most importantly, it is difficult to measure the correct extent of the health benefit of these reports considering the increased anxiety, additional medical workup and living with a diagnosis that might not even manifest through life.

We could roughly estimate the possible impact of the secondary findings in life expectancy from the decision-analytic models of cost-effectiveness studies based on European and North American populations. Some of these studies have estimated the impact on health benefits measured as quality-adjusted life years (QALYs) or increase in life expectancy of population-based screening strategies of some genetic conditions. Based on these studies, we estimate a potential increase of 21 years in life expectancy for the diagnosis of congenital cardiomyopathy genes (Wordsworth et al., 2010), 7.6 years for long QT (Quaglini et al., 2006) and 6.0 years in highly penetrant breast cancer genes (Long & Ganz, 2015; Petelin et al., 2018). Because two patients harbored variants in long QT genes (*DSC2* and *SCN5A*), four in cardiomyopathy (*KCNH2*, *LMNA*, *MYBPC3* and *MYH7*) and four in high-penetrance breast cancer (*BRCA1* (n = 2), *BRCA2* and *TP53*), we may speculate an increase of 123.2 years in the life expectancy for our cohort, or 0.246 years/cohort patient. This estimate poorly reproduces reality because the studies on which it is based were not designed for this purpose and comprised different ethnicities and clinical management strategies. However, these numbers are promising and warrant further studies to clarify more objective benefits of reporting secondary findings.

Several limitations are raised for this study. Because our service is limited to a diagnostic laboratory, a prospective follow up was not performed by our team. Additionally, several cases lacked thorough phenotyping and the clinical features relied solely on information provided to the laboratory, restricting the quality and quantity of phenotype information available for clinical-molecular correlation at the time of interpretation. The selection criteria for clinical cases were broad and may have included patients without a proper referral for ES, particularly considering that geneticists did not evaluate several cases; some were referred by general practitioners and other specialists without training in genetics or dysmorphology. Information about other genetic investigations (e.g., chromosomal microarray) prior to exome sequencing was not available.

Negative cases comprised most of our patients (68.4%), with a great impact on life and health because, for those, the diagnostic odyssey has not ended. A possible next step for negative results would be genome sequencing, although some studies have recently demonstrated limited superiority in the diagnostic rate to ES (Alfares et al., 2018). The approach that seems to increase the diagnostic yield of ES with the best cost-effectiveness is regular reanalysis of nonpositive results with updated bioinformatic protocols and databases and reverse phenotyping, which may be very helpful in atypical presentations of rare diseases but requires a strong interaction between the laboratory and ordering physician (Alfares et al., 2018; Nambot et al., 2018). For our cohort, updated bioinformatics protocols that include CNV analysis may add extra diagnostic power because pathogenic microdeletions and microduplications were only analyzed in patients that underwent Protocol 3. Reverse phenotyping has already helped to diagnose some patients in this study because some clinical characteristics not primarily considered by the ordering clinician may have gained relevance in the presence of a molecular clue. Therefore, the literature points to our next steps for negative results: strengthening our contact with patients and their ordering clinicians and reanalyzing the ES data regularly. Translational research also poses a mandatory approach under universal conditions where the border between clinical diagnosis and research becomes increasingly attenuated. Because data sharing between institutions is the key for translational research, diagnostic laboratories are facing a dilemma because they will benefit from sharing internal data with their competitors.

5 | CONCLUSIONS

Exome sequencing is a powerful method to identify the molecular bases of monogenic disorders and unravel the diagnosis of rare diseases. This tool showed an overall diagnostic yield of 31.6% and potential clinical benefit for patients and their families, including tailoring targeted therapy, tumor screening, adoption of standardized clinical guidelines, medication adjustment and monitoring for disease-specific complications. Additionally, 7.4% of our patients presented a reportable secondary finding of a medically actionable condition that may impact life expectancy with a potential health benefit.

CONFLICTS OF INTEREST

The following authors are employees (received salary and other bonuses) of Fleury Medicina e Saude: Caio Robledo D'Angioli Costa Quaio; Caroline Monaco Moreira; Gil Monteiro Novo Filho; Patricia Rossi Sacramento-Bobotis; Michele Groenner Penna; Sandro F Perazzio; Aurelio Pimenta Dutra; Rafael Alves da Silva; Monize Nakamoto Provisor Santos; Vanessa Yurie Nozaki de Arruda; Vanessa Galdeno Freitas; Vinícius Ceola Pereira; Maria Carolina Pintao; Alexandre Ricardo dos Santos Fornari; Ana Lígia Buzolin; Andre Yuji Oku; Matheus Burger; Rodrigo Fernandes Ramalho; David Santos Marco Antonio; Elisa Napolitano e Ferreira; Otavio Jose Eulalio Pereira; Vanessa Dionisio Cantagalli; Ana Carolina Gomes Trindade; Rafaela Rogerio Floriano de Souza; Cintia Reys Furuzawa; Fernanda Verzini; Shirley Dezan Matalhana; Naiade Romano; Daniele Paixão; Caroline Olivati; Gustavo Marquezani Spolador; Gustavo Arantes Rosa Maciel; Viviane Zorzanelli Rocha; Javier Miguelez; Mario Henrique Burlacchini de Carvalho; Alexandre Wagner Silva de Souza; Luis Eduardo Coelho Andrade; Maria de Lourdes Chauffaille; Aline dos Santos Borgo Perazzio; Ana Lucia Pereira Monteiro Catelani; Miguel Mitne-Neto; Wagner Antonio da Rosa Baratela.

ETHICAL STATEMENT

This study was granted ethics committee approval from both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339)

AUTHOR CONTRIBUTIONS

CRDC Quaio; CM Moreira; GM Novo Filho; PR Sacramento-Bobotis; MG Penna; SF Perazzio; AP Dutra; WAR Baratela: provided the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content and final approval of the version to be submitted.

RA da Silva; MNP Santos; VYN de Arruda; VG Freitas; VC Pereira; MC Pintao; ARS Fornari; AL Buzolin; AY Oku; M Burger; RF Ramalho; DSM Antonio; EN Ferreira; OJE Pereira; VD Cantagalli; ACG Trindade; RRF Souza; CR Furuzawa; F Verzini; SDM; N Romano: provided acquisition of data, analysis and interpretation of data and final approval of the version to be submitted.

D Paixão; C Olivati; GM Spolador; GAR Maciel; VZ Rocha; J Miguelez; MHB Carvalho; AWS Souza; LEC Andrade; ML Chauffaille; ASB Perazzio; ALPM Catelani; M Mitne-Neto: provided analysis and interpretation of data, revised it critically for important intellectual content and final approval of the version to be submitted.

CONSENT TO PARTICIPATE

All patients or their legal guardians provided written consent before exome analysis

CLINVAR SUBMISSION

All variants will be submitted to ClinVar prior to the publication of this article.

DATA AVAILABILITY STATEMENT

All data are provided as "Supplementary Material".

ORCID

Caio Robledo D'Angioli Costa Quaio https://orcid.org/0000-0003-2873-2820

REFERENCES

Alfares, A., Aloraini, T., Alissa, A., Qudsi, A. A., & Alahmad, A. (2018). Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genetics in Medicine*, *20*, 1328–1333. https://doi.org/10.1038/gim.2018.41

Beaulieu, C. L., Majewski, J., Schwartzentruber, J., Samuels, M. E., Fernandez, B. A., Bernier, F. P., … Boycott, K. M. (2014). FORGE Canada consortium: Outcomes of a 2-year national rare-disease gene-discovery project. *American Journal of Human Genetics*, *94*, 809–817. https://doi.org/10.1016/j.ajhg.2014.05.003

Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: Discovery to translation. *Nature Reviews. Genetics*, *14*, 681–691. https://doi.org/10.1038/nrg3555

Dorschner, M. O., Amendola, L. M., Turner, E. H., Robertson, P. D., Shirts, B. H., Gallego, C. J., … Jarvik, G. P. (2013). Actionable, pathogenic incidental findings in 1,000 participants' exomes. *American Journal of Human Genetics*, *93*, 631–640. https://doi.org/10.1016/j.ajhg.2013.08.006

Farwell, K., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., & Tippin Davis, B. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model–based analysis: Results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine*, *17*, 578–586. https://doi.org/10.1038/gim.2014.154

Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., … American College of Medical Genetics and Genomics. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing [published correction appears in genet med. 2017;19:606]. *Genetics in Medicine*, *15*, 565–574. https://doi.org/10.1038/gim.2013.73

Johnston, J. J., Rubinstein, W. S., Facio, F. M., Ng, D., Singh, L. N., Teer, J. K., … Biesecker, L. G. (2012). Secondary variants in individuals undergoing exome sequencing: Screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. *American Journal of Human Genetics*, *91*, 97–108. https://doi.org/10.1016/j.ajhg.2012.05.021

Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., … Evans, J. P. (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. *Genetics in Medicine*, *19*, 249–255.

Lawrence, L., Sincan, M., Markello, T., Adams, D. R., Gill, F., Godfrey, R., … Boerkoel, C. F. (2014). The implications of familial incidental findings from exome sequencing: The NIH undiagnosed diseases program experience. *Genetics in Medicine*, *16*, 741–750. https://doi.org/10.1038/gim.2014.29

Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., … Nelson, S. F. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Journal of the American Medical Association*, *312*, 1880–1887. https://doi.org/10.1001/jama.2014.14604

Long, E. F., & Ganz, P. A. (2015). Cost-effectiveness of universal BRCA1/2 screening: Evidence-based decision making. *JAMA Oncology*, *1*, 1217–1218. https://doi.org/10.1001/jamaoncol.2015.2340

McKusick, V. A. (2007). Mendelian inheritance in man and its online version, OMIM. *American Journal of Human Genetics*, *80*, 588–604. https://doi.org/10.1086/514346

Meng, L., Pammi, M., Saronwala, A., Magoulas, P., Ghazi, A. R., Vetrini, F., … Lalani, S. R. (2017). Use of exome sequencing for infants in intensive care units: Ascertainment of severe single-gene disorders and effect

on medical management. *JAMA Pediatrics*, *171*, e173438. https://doi.org/10.1001/jamapediatrics.2017.3438

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., … Ledbetter, D. H. (2010). Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American Journal of Human Genetics*, *86*, 749–764. https://doi.org/10.1016/j.ajhg.2010.04.006

Nambot, S., Thevenon, J., Kuentz, P., Duffourd, Y., Tisserant, E., & Bruel, A. L. (2018). Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: Substantial interest of prospective annual reanalysis. *Genetics in Medicine*, *20*, 645–654. https://doi.org/10.1038/gim.2017.162

Petelin, L., Trainer, A. H., Mitchell, G., Liew, D., & James, P. A. (2018). Cost-effectiveness and comparative effectiveness of cancer risk management strategies in BRCA1/2 mutation carriers: A systematic review. *Genetics in Medicine*, *20*, 1145–1156. https://doi.org/10.1038/gim.2017.255

Quaglini, S., Rognoni, C., Spazzolini, C., Priori, S. G., Mannarino, S., & Schwartz, P. J. (2006). Cost-effectiveness of neonatal ECG screening for the long QT syndrome. *European Heart Journal*, *27*, 1824–1832. https://doi.org/10.1093/eurheartj/ehl115

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … ACMG Laboratory Quality Assurance Committee. (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. *Genetics in Medicine*, *17*, 405–424.

Tan, T. Y., Dillon, O. J., Stark, Z., Schofield, D., Alam, K., Shrestha, R., … White, S. M. (2017). Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. *JAMA Pediatrics*, *171*, 855–862. https://doi.org/10.1001/jamapediatrics.2017.1755

Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J. B., Rimmer, A., … McVean, G. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature Genetics*, *47*, 717–726. https://doi.org/10.1038/ng.3304

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., Weiss, M. E. R., Köster, J., Marais, A., … Abou Jamra, R. (2017). Clinical exome sequencing: Results from 2819 samples reflecting 1000 families. *European Journal of Human Genetics*, *25*, 176–182. https://doi.org/10.1038/ejhg.2016.146

Wordsworth, S., Leal, J., Blair, E., Legood, R., Thomson, K., Seller, A., … Watkins, H. (2010). DNA testing for hypertrophic cardiomyopathy: A cost-effectiveness model. *European Heart Journal*, *31*, 926–935. https://doi.org/10.1093/eurheartj/ehq067

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., … Eng, C. M. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England Journal of Medicine*, *369*, 1502–1511. https://doi.org/10.1056/NEJMoa1306555

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., … Eng, C. M. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Journal of the American Medical Association*, *312*, 1870–1879. https://doi.org/10.1001/jama.2014.14601

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Quaio CRDC, Moreira CM, Novo-Filho GM, et al. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. *Am J Med Genet Part C*. 2020;1–10. https://doi.org/10.1002/ajmg.c.31860

Received: 27 February 2021 | Revised: 13 May 2021 | Accepted: 30 June 2021

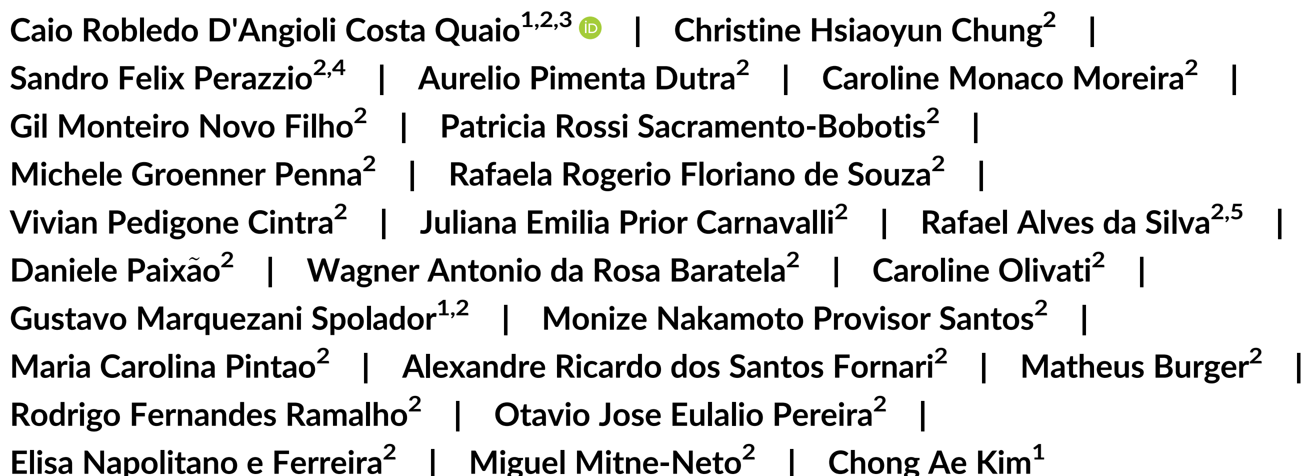
DOI: 10.1002/ajmg.c.31932

AMERICAN JOURNAL OF medical genetics PART C Seminars in Medical Genetics

WILEY

RESEARCH ARTICLE

Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients



Caio Robledo D'Angioli Costa Quaio[1,2,3] | Christine Hsiaoyun Chung[2] | Sandro Felix Perazzio[2,4] | Aurelio Pimenta Dutra[2] | Caroline Monaco Moreira[2] | Gil Monteiro Novo Filho[2] | Patricia Rossi Sacramento-Bobotis[2] | Michele Groenner Penna[2] | Rafaela Rogerio Floriano de Souza[2] | Vivian Pedigone Cintra[2] | Juliana Emilia Prior Carnavalli[2] | Rafael Alves da Silva[2,5] | Daniele Paixão[2] | Wagner Antonio da Rosa Baratela[2] | Caroline Olivati[2] | Gustavo Marquezani Spolador[1,2] | Monize Nakamoto Provisor Santos[2] | Maria Carolina Pintao[2] | Alexandre Ricardo dos Santos Fornari[2] | Matheus Burger[2] | Rodrigo Fernandes Ramalho[2] | Otavio Jose Eulalio Pereira[2] | Elisa Napolitano e Ferreira[2] | Miguel Mitne-Neto[2] | Chong Ae Kim[1]

[1]Instituto da Crianca (Children's Hospital), Hospital das Clinicas HCFMUSP, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

[2]Fleury Medicina e Saúde, São Paulo, Brazil

[3]Laboratório Clínico, Hospital Israelita Albert Einstein, São Paulo, Brazil

[4]Division of Rheumatology, Universidade Federal de Sao Paulo, Sao Paulo, Brazil

[5]Laboratório de Hepatologia Molecular Aplicada (LHeMA), Escola Paulista de Medicina, Universidade Federal de São Paulo, Sao Paulo, Brazil

Correspondence
Caio Robledo D'Angioli Costa Quaio, Instituto da Criança do Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo, Av. Gen. Valdomiro de Lima, 508 - Jabaquara, São Paulo 04344-070, Brazil.
Email: cquaio@gmail.com

Abstract

Several Mendelian disorders follow an autosomal recessive inheritance pattern. Epidemiological information on many inherited disorders may be useful to guide health policies for rare diseases, but it is often inadequate, particularly in developing countries. We aimed to calculate the carrier frequencies of rare autosomal recessive Mendelian diseases in a cohort of Brazilian patients using whole exome sequencing (WES). We reviewed the molecular findings of WES from 320 symptomatic patients who had carrier status for recessive diseases. Using the Hardy–Weinberg equation, we estimated recessive disease frequencies (q^2) considering the respective carrier frequencies (2pq) observed in our study. We calculated the sensitivity of carrier screening tests based on lists of genes from five different clinical laboratories that offer them in Brazil. A total of 425 occurrences of 351 rare variants were reported in 278 different genes from 230 patients (71.9%). Almost half (48.8%) were carriers of at least one heterozygous pathogenic/likely pathogenic variant for rare metabolic disorders, while 25.9% of epilepsy, 18.1% of intellectual disabilities, 15.6% of skeletal disorders, 10.9% immune disorders, and 9.1% of hearing loss. We estimated that an average of 67% of the variants would not have been detected by carrier screening panels. The combined frequencies of autosomal recessive diseases were estimated to be 26.39/10,000 (or ~0.26%). This study shows the potential research utility of WES to determine carrier status, which may be a possible strategy to evaluate the clinical and social burden of recessive diseases at the population level and guide the optimization of carrier screening panels.

 © 2021 Wiley Periodicals LLC.

KEYWORDS
carrier, carrier frequency, next-generation sequencing, rare diseases, recessive Mendelian diseases, whole exome sequencing

1 | INTRODUCTION

Several Mendelian disorders follow a recessive inheritance pattern, and almost 1,875 known protein-coding genes have been linked to recessive diseases; however, this number of known recessive diseases may represent only ∼20% of the projected total, which means that the majority of recessive diseases have yet to be described (Antonarakis, 2019). Epidemiological information on many inherited disorders may be useful to guide health policies for rare diseases, but this information is often inadequate, particularly in developing countries. One important clinical utility for such studies is to adequately plan carrier screening strategies that ultimately have the objective of (a) identifying couples at risk of having children with recessive diseases and (b) subsequently providing genetic counseling, information on reproductive risks and possible options for reproductive intervention.

Recent developments in next-generation sequencing (NGS) techniques, which have brought less expensive and faster DNA studies, have allowed broader access for the precise diagnosis for rare diseases of genetic etiology, including recessive conditions. Whole exome sequencing (WES) is an NGS-based method that allows for studying the majority of known protein-coding human genes. WES can be considered a broad, unbiased diagnostic method for monogenic diseases because it is not restricted to a subset of genes. On the other hand, focused gene panel analysis, also based on NGS methodology, is restricted to select sets of genes implicated in specific diseases.

Carrier screening is defined as the detection of carrier status of recessive diseases in couples or persons who do not have an a priori increased risk of being a carrier based on their or their partners' personal or family history, with the primary goal of facilitating informed reproductive decisions (Henneman et al., 2016). Expanded carrier screening offers carrier screening for multiple recessive disorders. Several private clinical laboratories have been offering carrier screening tests through focused NGS panels, though laboratories usually have different strategies for selecting the sets of genes for their panels, and information about the clinical utility of this approach for the general population is usually very limited in locations other than the United States or Europe (Edwards et al., 2015; Fares et al., 2008; Godard, Ten Kate, Evers-Kiebooms, & Ayme, 2003; Henneman et al., 2016). Additionally, many countries have no clear guidelines or policies for carrier screening.

In this study, we aim to (a) assess carrier frequencies of rare autosomal recessive Mendelian diseases in a cohort of Brazilian patients using WES, (b) estimate the overall frequency of recessive diseases in this population using the Hardy–Weinberg equation, and (c) estimate the sensitivity of carrier screening panels from five different clinical laboratories that offer carrier screening tests in Brazil.

2 | METHOD

2.1 | Selection of cases, molecular analysis, and bioinformatics

To characterize the frequencies of carriers for rare pathogenic or likely pathogenic (P/LP) alleles of autosomal recessive conditions in a cohort of Brazilian patients, we reviewed the molecular findings of WES from 320 symptomatic patients who had carrier status for recessive diseases that we actively searched. These patients were selected from an original cohort of 500 symptomatic patients who had undergone molecular analysis for suspected diseases with genetic etiology from 2016 to 2020 in facilities of the Fleury Group. Full details of the clinical features of patients, molecular analysis, bioinformatics protocols, clinical data, and molecular data for primary and secondary findings were previously published (Quaio et al., 2020). A clinical summary is available in Data S1, Supporting Information (section "Case overview").

All patients or their legal guardians provided consent before exome analysis, and this study was granted ethics committee approval from both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

2.2 | Carrier status identification, variant classification, and gene categories

For all 320 cases, carrier status for autosomal recessive diseases was actively searched, and P/LP variants for recessive conditions were reported according to the same internal criteria (Cases #66 through #386, excluding case #67). For this purpose, at least two parallel analyses were performed to preselect variants considering the following: (a) relevant reports from databases (e.g., ClinVar and HGMD) and the literature and (b) variants that were not reported in ClinVar that had an allele frequency < 1%, had an in silico prediction of a functional impact and was located in a clinically relevant gene. These preselected variants were then discussed by a board comprising three experts. Variants were classified according to ACMG guidelines (Richards et al., 2015) with the assistance of a third-party ACMG calculator by Saphetor SA (Lausanne, Switzerland) (www.varsome.com). Internal adaptations of the ACMG guidelines used in this work may be found in Data S1 (section "Variant classification protocol").

All variants were reviewed and reclassified for the purpose of this article. The complete details, nomenclature, and classification information are available in Data S1 (section "Data"). In this step, a total of 17 variants found in 18 patients were reclassified as variants of unknown significance (VUS) and were excluded from our analysis (Data S1, section "Excluded variants"). Additionally, a likely

pathogenic variant found in the *DMD* gene was excluded because this gene is associated with X-linked inheritance, and the purpose of this article is to study only autosomal recessive diseases.

We excluded the following from our analysis: (a) variants for common autosomal recessive diseases with high allele frequencies and low penetrance; (b) variants associated with no clinical impact (such as traits); (c) variants associated with conditions in which the diagnosis depends on other tests (e.g., variants in *BTD*, *G6PD*, *HFE* and *SERPINA1*); and (d) genes with pseudogenes with high homology (such as *CYP21A2*). A complete list of variants excluded from our analysis is available in Data S1 (section "Variants excluded from analysis").

Genes analyzed in this study were divided into categories based on their clinical impact. According to nosologies, reviews and professional guidelines available in the literature, we classified all genes into the following clinical groups: (a) intellectual disability (Kochinke et al., 2016), (b) cancer (LaDuca et al., 2020), (c) muscular diseases (Schorling, Kirschner, & Bönnemann, 2017), (d) ciliopathies (Reiter & Leroux, 2017), (e) skeletal disorders (Mortier et al., 2019), (f) immune disorders (Tangye et al., 2020), (g) epilepsy (Wang et al., 2017), (h) hearing loss (Shearer, Hildebrand, & Smith, 1999), (i) retinitis pigmentosa (Fahim, Daiger, & Weleber, 2000), (j) metabolic disorders (Ferreira, van Karnebeek, Vockley, & Blau, 2019), and (k) others (genes not classified in any other clinical groups). The same gene may belong to multiple clinical groups, except for others, which only includes genes that are not part of another clinical category. The complete list of genes for every category is available in Data S1 (section "List of genes").

2.3 | Recessive disease frequency estimation

We used the Hardy–Weinberg equation to estimate disease frequencies (q^2) based on the respective carrier frequencies (2pq) observed in this study (Edwards, 2008). For this purpose, we considered random mating and the approximation $p \sim 1$.

2.4 | Sensitivity of carrier screening panels: Selection of laboratories and panels

We also estimated the sensitivity of carrier screening panels from different laboratories. For this purpose, we checked whether every gene that harbored P/LP variants observed in this study was part of the list of genes of the respective panel.

Our team selected five different carrier screening NGS panels from five genetic clinical laboratories based on their national and international recognition of excellence. Gene coverage details for every panel and for each patient are available in Data S1 (sections "Laboratories" and "Laboratories 2"). Two of these laboratories are national laboratories, and the other three are international laboratories, of which two are from the United States and one is from Germany. The panels and respective gene lists were selected according to the information publicly available on the laboratory websites. The names of the laboratories will not be disclosed because the objective of our study is not the comparison between different services but to discuss the sensitivity of this approach in identifying carrier status under the circumstances of this study.

2.5 | Statistical analysis

Unpaired Student's *t* test was used to determine significantly different means.

3 | RESULTS

3.1 | Variant characteristics and detection of heterozygotes

A total of 425 occurrences of 351 different variants (47 variants were recurrent more than once) were reported in 278 different genes for 230 patients (71.9%) in our cohort of 320 patients. For 90 cases (28.1%), no variants were reported. Overall, we observed an average of 1.32 reported variants/patient (median = 1, *SD* = 1.188, maximum = 5). Some patients presented more than one variant for the same clinical category (Data S1, section "Cases & Genes"), while others presented variants for different clinical categories (Data S1, section "Cases per category"). Regarding the ACMG classification, 218 different variants were classified as likely pathogenic and 133 as pathogenic. Regarding the 351 different variants, the majority of them had been previously observed in the literature ($n = 257$, 73.2%) and are present in the ClinVar Database ($n = 267$, 76%). On the other hand, 72 variants (20.5%) were not found in either the literature or the ClinVar database. Among the latter group, all 72 variants presented at a considerably low frequency in population databases (e.g., frequency lower than 0.01%) along with a well-established loss-of-function mechanism, and based on a combination of the PVS1 and PM2 criteria from ACMG guidelines, they were classified as likely pathogenic. Of the total 425 occurrences, those harbored by genes associated with metabolic disorders comprised almost half of all variants (48.2%). This group was followed by epilepsy (22.8%) and intellectual disabilities (18.8%). Variants harbored by genes associated with muscular diseases comprise the least common category (2.4%). The distribution of variants into clinical groups is depicted in Table 1-A.

Based on these findings, we were able to calculate the frequency of heterozygous carriers for each specific clinical category from the data of our 320 patients (Table 1-B). Among all participants in our cohort, almost half (48.8%, $n = 156$) were carriers of at least one heterozygous variant for rare metabolic disorders, representing the most frequent clinical category. Heterozygote frequencies for the other categories were 25.9% for epilepsy, 18.1% for intellectual disability, 15.6% for skeletal disorders, and only 3.1% were carriers for muscular diseases. Data on individual cases are available in Data S1 (section "Cases per category").

The highest carrier frequency for single variants and their respective recessive conditions was 2.5% for the known *RBM8A*:c.-21G>A, which is associated with thrombocytopenia-absent radius syndrome (OMIM#274000). This second highest frequency was 1.9% for

RNASEH2B:p.Ala177Thr, which is associated with interferonopathies, mainly within the spectrum of Aicardi-Goutieres syndrome (OMIM#610181). The third highest carrier frequency was 1.6% for *GJB2*:c.35del, which is associated with nonsyndromic hearing loss (OMIM#220290). Additionally, we observed a frequency of 1.3% for the following variants and their respective associated diseases: *CPT2*:p. Ser113Leu (CPT2 deficiency, OMIM#600649), *ABCA4*:p.Gly1961Glu (retinitis pigmentosa, OMIM#601718), *SPG7*:p.Ala510Val (spastic paraplegia, OMIM#607259), and *EIF2B5*:p.Arg113His (leukoencephalopathy with vanishing white matter, OMIM#603896).

A frequency of 0.9% was observed for the following variants: *MUTYH*:p.Gly396Asp, (associated with *MUTYH*-related polyposis, OMIM#608456), *GBA*:p.Asn409Ser (Gaucher disease, OMIM#230800), *ATP7B*:p.Gly869Arg (Wilson disease, OMIM#277900), *TNFRSF13B*: p.Cys104Arg (common variable immunodeficiency, OMIM#240500), *TCAP*:p.Gln53* (limb-girdle muscular dystrophy, OMIM#601954), *MPO*: p.Arg569Trp (myeloperoxidase deficiency, OMIM#254600), and *ALDOB*: p.Ala150Pro (hereditary fructose intolerance, OMIM#229600). The complete list of variants and details on nomenclature and respective frequencies are available in Data S1 (section "Data").

3.2 | Gene characteristics

Regarding the 278 different genes, almost half (49.6%) were associated with metabolic diseases, followed by epilepsy (25.5%) and intellectual disabilities (20.1%). On the other hand, less than 1% of the genes (0.7%) were associated with cancer. Table 1-C shows the complete distribution of genes among the clinical categories in which they were included.

The combined carrier frequency for all the different P/LP variants for the top eight most frequent genes are (a) 3.75% for *GJB2* (nonsyndromic hearing loss, OMIM#220290), (b) 2.8% for *MUTYH* (*MUTYH*-related polyposis, OMIM#608456), (c) 2.8% for *RBM8A* (thrombocytopenia-absent radius syndrome, OMIM#274000), (d) 2.2% for *ABCA4* (retinitis pigmentosa, OMIM#601718), (e) 1.9% for *GBA* (Gaucher disease, OMIM#230800), (f) 1.9% for *PAH* (phenylketonuria, OMIM#261600), (g) 1.9% for *RNASEH2B* (Aicardi-Goutieres syndrome, OMIM#610181), and (h) 1.6% for *SPG7* (spastic paraplegia, OMIM#607259).

3.3 | Recessive diseases frequency estimation

Using the Hardy–Weinberg equation, we roughly estimated disease frequencies (q^2) based on the respective carrier frequencies (2pq) observed above and the approximation of $p \sim 1$. According to this method, we estimated the following top seven frequencies of diseases per 10,000 people: *GJB2*-associated nonsyndromic hearing loss: 3.52, *MUTYH*-related polyposis: 1.98, *ABCA4*-associated retinitis pigmentosa: 1.20, Gaucher disease: 0.88, phenylketonuria: 0.88, Aicardi-Goutieres syndrome: 0.88, and spastic paraplegia: 0.61. The combined frequency of these seven diseases is estimated to be 10.72

TABLE 1 Distribution of variants, heterozygotes, and genes among clinical groups

	A-variants		B-heterozygotes		C-genes	
Clinical groups	Number of variants	Frequency (%)	Number of heterozygotes	Frequency of heterozygotes (%)	Number of genes	Frequency (%)
Metabolic disorders	205	48.24	156	48.8	138	49.6
Epilepsy	97	22.82	83	25.9	71	25.5
Intellectual disability	80	18.82	58	18.1	56	20.1
Other	71	16.71	66	20.6	52	18.7
Skeletal disorders	56	13.18	50	15.6	38	13.7
Immune disorders	39	9.18	35	10.9	22	7.9
Hearing loss	29	6.82	29	9.1	12	4.3
Retinitis pigmentosa	26	6.12	24	7.5	12	4.3
Ciliopathies	24	5.65	24	7.5	17	6.1
Cancer	14	3.29	14	4.4	2	0.7
Muscular diseases	10	2.35	10	3.1	8	2.9
Total	425	100	230	71.9	278	100.0

Note: A scale of color was used to compare the distribution of frequencies. A, B, C show the absolute number of variants, heterozygotes, and genes, respectively, and their frequency among clinical groups of conditions. Note that the same gene (1-C) or variant (1-A) may be part of multiple clinical groups. Heterozygotes (1-B) represent the carriers of at least one heterozygous variant for the respective clinical category; several individuals were carriers for different variants within the same clinical category and for different clinical categories (individual data available in Data S1).

per 10,000 people, while the combined frequency considering the whole cohort is 26.39/10,000 (Data S1, section "Cases & Genes"). We excluded the *RBM8A* gene (associated with thrombocytopenia-absent radius syndrome) from this approach because its related condition is caused by compound heterozygosity for a rare null mutation and the c.-21G>A variant, which in a homozygous state does not cause disease.

3.4 | Sensitivity of carrier screening panels

Overall, the mean sensitivity of P/LP variants (determined by the number of carriers per gene) from all five laboratories was estimated to be 33%, varying from 23% (Lab1) to 44% (Lab3). Based on these data, we estimate that an average of 67% of the variants found in this study are not harbored by genes included in the carrier screening panels from the five laboratories and, eventually, would be missed by these tests. The estimation of sensitivities per clinical group and per laboratory is summarized in Table 2, where sensitivities were calculated for variants and genes. The sensitivity of genes harboring P/LP variants (not weighted by the number of carriers) varied from 22 to 36%.

3.5 | Comparison of carrier frequencies between patients with and without primary findings

We compared the average of variants between two groups of patients. The first group consisted of patients with primary findings ($n = 106$). These are patients who harbor P/LP variants in a gene that was associated with the patient's phenotype and had compatible zygosity and an adequate inheritance pattern. All primary findings can be found in our previous study (Quaio et al., 2020). The second group consisted of patients without primary findings ($n = 214$). The frequency of heterozygous carriers for P/LP variants was 1.28 for patients with primary findings ($SD = 1.26$) and 1.34 for patients without primary findings ($SD = 1.15$); this difference showed no statistical significance (unpaired t test, 0.6639; 95% confidence interval of this difference: from −0.34 to 0.22).

4 | DISCUSSION

We studied the frequencies of heterozygotes for rare recessive monogenic diseases in 320 Brazilian patients. We found a frequency of

TABLE 2 Sensitivity of the carrier screening panels of five laboratories for each clinical group and the overall sensitivity of each laboratory (last row)

Clinical groups	Total variants	Sensitivity of variants (%) per clinical group, per laboratory					Total genes	Sensitivity of genes (%) per clinical group, per laboratory				
		Lab1 (Brazil) (%)	Lab2 (Brazil) (%)	Lab3 (US) (%)	Lab4 (US) (%)	Lab5 (Germany) (%)		Lab1 (Brazil) (%)	Lab2 (Brazil) (%)	Lab3 (US) (%)	Lab4 (US) (%)	Lab5 (Germany) (%)
Intellectual disability	80	26	28	46	35	34	56	30	21	43	30	30
Cancer	14	0	7	7	7	7	2	0	50	50	50	50
Muscular diseases	10	40	50	60	50	40	8	38	38	50	38	25
Ciliopathies	24	8	25	42	33	21	17	12	24	47	35	18
Skeletal disorders	56	29	13	25	20	18	38	34	11	26	21	18
Immune disorders	39	28	23	38	26	31	22	27	23	36	27	23
Epilepsy	97	15	27	47	34	42	71	20	20	39	28	37
Hearing loss	29	24	66	69	69	69	12	25	25	33	33	33
Retinitis pigmentosa	26	19	31	46	38	31	12	17	25	58	42	25
Metabolic disorders	205	31	38	58	46	48	138	28	28	47	38	41
Overall sensitivity per laboratory	425	23	27	44	35	36	278	22	19	36	28	29

Note: Sensitivities were calculated per variant (left) and gene (right). Sensitivity per variant is determined by the number of carriers per gene; therefore, genes harboring a higher number of variants are overrepresented. Sensitivity per gene is not weighted by different frequencies of carriers per gene, and all genes present the same weight in the calculation.

1.32 reported variants/patient, and at least one P/LP variant was reported in the majority (71.9%) of them. It is important to note, however, that several variants were classified as likely pathogenic based solely on ACMG criteria but were not definitively proven to be pathogenic (disease-causing), as discussed below. Based on our data and the Hardy–Weinberg equation, we estimate the overall frequency of recessive diseases to be 26.39/10,000 people, or approximately 0.26%.

Using different methodologies, other studies have already estimated the overall frequency of recessive diseases to be approximately 30 per 10,000 individuals in Western populations (Henneman et al., 2016; Sankaranarayanan, 1998). These reports are consistent with our findings. Considering these numbers, it has been estimated that approximately 1–2% of couples may be at risk of having a child affected with an autosomal recessive genetic disorder (Ropers, 2012). It is noteworthy that different studies, including ours, with different methodologies reported very close estimations for the frequency of recessive diseases in different populations.

These estimations, however, are challenging because the frequency of carriers may vary widely in different population groups due to founder effects, endogamy, nonrandom mating, and cultural, religious, social, and/or geographical isolation (Antonarakis, 2019). For example, in endogamous societies, the frequency of recessive diseases is estimated to be much higher, with up to 70 per 10,000 individuals (Abouelhoda et al., 2016). Additionally, current estimations may represent only the "tip of the iceberg" because statistical projections indicate that four-fifths of recessive diseases have yet to be described (Antonarakis, 2019).

Other studies using different approaches presented different results. A population-based registry study in Canada with 1 million consecutive live births estimated the frequency of autosomal recessive disorders to be 17:10,000 for congenital diseases or conditions with onset before age 25 years (Baird, Anderson, Newcombe, & Lowry, 1988). It is important to note that the scope of diseases in this study was limited to those with onset prior to 25 years of age, while our study also included those with later onset. Several diseases in our study, including some of our top highest estimated frequencies (*GJB2*-associated nonsyndromic hearing loss, *MUTYH*-related polyposis, *ABCA4*-associated retinitis pigmentosa, and others), may present as forms with a later onset. To compare our estimates with the Canadian study using clinical information available in OMIM (Clinical Synopsis), we calculated the combined estimated frequencies exclusively from the diseases with onset before adulthood (Data S1, section "Cases & Genes") to be 17.7 per 10,000, which is close to the Canadian estimation.

Other authors have demonstrated that, after screening for 448 genes in 104 unrelated samples, the average carrier burden of severe childhood-onset recessive disorders was 2.8 (range 0–7) (Bell et al., 2011), which was higher than we observed in our cohort. Although the frequency of recessive diseases is generally low, the frequency of carriers may be relatively high and can reach 1:17 individuals in certain populations (Fares et al., 2008).

Literature regarding carrier frequencies for autosomal recessive conditions is limited in Latin America and in other populations outside North America and Europe. Available studies are generally limited to sole diseases or restricted groups of diseases. As we did not find comprehensive studies that address carrier status for multiple recessive diseases in Brazil or any other Latin American countries, we hope that our study encourages other studies that aim to better understand the burden of recessive diseases in our developing countries.

The majority of our patients (71.9%) were heterozygous for at least one P/LP variant. It is difficult to compare these numbers because we used a broader approach with WES, while most studies rely on data from focused carrier screening panels that are restricted to different subsets of genes and present low sensitivity for our set of variants (average sensitivity of carrier screening panels that we calculated: 33%). For example, a screening panel of 108 disorders and 417 disease-causing variants identified 24% of individuals as carriers for at least one variant (Lazarin et al., 2013). Considering the low sensitivity that we observed for carrier panels, we believe that WES may be an alternative method to clarify carrier risk, particularly in diverse populations, if used with caution and by experienced genetic counseling teams. However, our study is not sufficient to support WES as a first-line investigation method for carrier status; several other facts must be considered.

A Joint Statement of the American College of Medical Genetics and Genomics, American College of Obstetricians and Gynecologists, National Society of Genetic Counselors, Perinatal Quality Foundation, and Society for Maternal-Fetal Medicine addresses an approach for health care providers and laboratories that wish to or are currently offering expanded carrier screening to their patients (Edwards et al., 2015). This statement highlights the importance of genetic counseling, the principle of confidentiality, and the limitation of completely eliminating risk to offspring when the test is negative, among others. Regarding the latter subject, we have shown that, indeed, an average of 67% of the L/LP variants found in this study would not have been detected by carrier screening panels, and this information may be relevant for genetic counseling, especially for consanguineous couples. The statement also recommends that the condition being screened for should be a health problem that encompasses clinically relevant outcomes, exclude adult-onset conditions and, when molecular testing cannot distinguish between childhood or adult onset, exclude variants with high allele frequencies and low penetrance, among other principles. It is important to note that some of the recommendations in this statement can be subjective, for example, criteria used to define clinically relevant outcomes.

The European Society of Human Genetics also enumerates criteria and principles for screening programs, such as a well-defined target population, stringent quality control of laboratories, protection of the confidentiality of the information, protection of privacy of the individual and respective family, voluntary participation, access to educational programs and genetic counseling, and evaluation of long-term outcomes (Godard et al., 2003). Therefore, although our study shows the potential research utility of WES to unravel carrier status in

population studies, especially considering the limitations in sensitivity of panels developed for this purpose, we do not advocate for the implementation of WES as a strategy for carrier screening in clinical settings.

Broader research studies of carriers may guide the optimization and personalization of carrier screening panels by adapting the subsets of genes based on the characteristics of each population. For example, heterozygous carriers for variants in the *RNASEH2B*, *EARS2*, and *CC2D2A* genes were relevantly recurrent in our cohort, and these genes are not part of any carrier screening panel that we analyzed. Our team subjectively judged the conditions associated with those genes (Aicardi-Goutieres syndrome, combined oxidative phosphorylation deficiency and Joubert syndrome) to be eligible for carrier screening by considering the relevant clinical outcomes and high frequency of P/LP variants.

On the other hand, adult-onset conditions, such as *MUTYH*-related polyposis, produced an expectedly low sensitivity, as most genes associated with late-onset conditions are not covered in carrier screening panels. As *MUTYH* is not covered by any laboratory, cancer variants presented a very low sensitivity among all screening tests, but our cohort was enriched in heterozygous pathogenic variants in the *MUTYH* gene (2.8%).

The limited number of patients is probably the highest bias in our cohort and compromises the direct estimation of single monogenic disease frequencies, though this bias can be diluted by the estimation of frequencies of disease groups. This is the reason why we adopted a strategy to calculate the frequency by groups of diseases. Another limitation is that our estimations are restricted to known P/LP variants, and thus, we did not include variants that are not yet recognized to cause disease or variants that are caused by molecular alterations not studied by NGS. This is an important factor that could lead to underestimation of the frequencies, especially for hypomorphic alleles (such as missense variants) that do not result in clear loss of function, such as nonsense and frameshift mutations. On the other hand, we may have overestimated the carrier frequencies by falsely classifying some variants as P/LP since we relied solely on presumptively assuming the loss-of-function mechanism without proper functional studies, even though we followed the ACMG criteria for variant classification.

We eliminated P/LP variants associated with autosomal recessive disorders reported in homozygosity or compound heterozygosity (primary findings) from this analysis because our objective was to study carriers for recessive diseases; therefore, all patients in this study had only monoallelic variants, several of whom presented with monoallelic variants in more than one gene. Nevertheless, we cannot rule out that some of the variants studied in this work may have contributed to the clinical phenotype of the patient, for example, if the disease presents an unknown codominant, digenic, or even oligogenic mechanism or if the other allele harbored a pathogenic or likely pathogenic variant not detected by the NGS methodology (such as intronic or regulatory site variants). Another issue regarding heterozygous carriers is that some monoallelic variants may have clinical impacts and health complications. For instance, monoallelic variants in some cancer-predisposition genes (such as *ATM* and *MUTYH*) are associated with an increased risk for cancer (Slavin et al., 2017). Other genes (including *POLG*, *TNFRSF13B*, genes of hearing loss, and retinitis pigmentosa) may be associated with autosomal dominant forms that have a generally milder clinical impact.

We believe, however, that this possible bias may be restricted because (a) we did not notice differences in carrier frequencies between patients with primary findings and those without primary findings and (b) carrier frequencies and estimation of recessive disease frequencies from other studies are consistent with our findings. We hope that future studies in Brazil and other Latin American countries that include only asymptomatic adult individuals (the ideal group for carrier studies) may help clarify this impact.

Recessive disorders represent a relevant fraction of Mendelian diseases. National estimates of the impact of recessive diseases are an important and feasible strategy to determine the clinical and social burden at the population level and address health policies, which is especially true if we consider that recessive diseases are relatively frequent in aggregate and often have a high clinical impact, and early management of recessive diseases can impact in morbidity and mortality (Azimi et al., 2016; Esquivel-Sada & Nguyen, 2018; Saunders et al., 2012). Considering the limitations of treatments and proper management, prevention is probably the best health-care policy to avoid them (Abouelhoda et al., 2016).

We have previously demonstrated that WES determined the molecular diagnosis in 31.6% of 500 Brazilian patients with diseases suspected of genetic etiology (Quaio et al., 2020). In this study, we explored novel data on the frequencies of heterozygous carriers for rare recessive monogenic diseases obtained with the same molecular approach and observed the potential research utility of WES to unravel carrier status. This approach may be a possible strategy to evaluate the clinical and social burden of recessive diseases at the population level and guide the optimization of carrier screening panels.

5 | CONCLUSIONS

We studied the frequencies of heterozygotes for rare pathogenic/likely pathogenic alleles of recessive conditions using WES and observed an average of 1.32 reported variants/patient. Almost three-quarters of our cohort (71.9%) were carriers of at least one heterozygous pathogenic/likely pathogenic variant for recessive disorders. We estimated that an average of 67% of the variants found in this study would not have been detected by carrier screening panels. Using the Hardy–Weinberg equation, we roughly estimated the overall frequency for rare recessive disease to be 26.39/10,000 (or ~0.26%). This study shows the potential research utility of WES to determine carrier status, which may be a possible strategy to evaluate the clinical and social burden of recessive diseases at the population level and guide the optimization of carrier screening panels.

CONFLICT OF INTEREST

The following authors are employees (received salary and other bonuses) of Fleury Medicina e Saude: Caio Robledo D'Angioli Costa Quaio, Sandro Felix Perazzio, Aurelio Pimenta Dutra, Christine Hsiaoyun Chung, Caroline Monaco Moreira, Gil Monteiro Novo Filho, Patricia Rossi Sacramento-Bobotis, Michele Groenner Penna, Rafaela Rogerio Floriano de Sousa, Vivian Pedigone Cintra, Juliana Emilia Prior Carnavalli, Rafael Alves da Silva, Monize Nakamoto Provisor Santos, Maria Carolina Pintao, Alexandre Ricardo dos Santos Fornari, Matheus Burger, Rodrigo Fernandes Ramalho, Elisa Napolitano e Ferreira, Otavio Jose Eulalio Pereira, Daniele Paixão, Wagner Antonio da Rosa Baratela, Caroline Olivati, Gustavo Marquezani Spolador, and Miguel Mitne-Neto.

AUTHOR CONTRIBUTIONS

Caio Robledo D'Angioli Costa Quaio, Christine Hsiaoyun Chung, Sandro Felix Perazzio, Aurelio Pimenta Dutra, and Chong Ae Kim provided the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content and final approval of the version to be submitted. Caroline Monaco Moreira, Gil Monteiro Novo Filho, Patricia Rossi Sacramento-Bobotis, Michele Groenner Penna, Rafaela Rogerio Floriano de Sousa, Vivian Pedigone Cintra, Juliana Emilia Prior Carnavalli, Maria Carolina Pintao, and Miguel Mitne-Neto provided acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content and final approval of the version to be submitted. Rafael Alves da Silva, Monize Nakamoto Provisor Santos, Alexandre Ricardo dos Santos Fornari, Ana Lígia Buzolin, Matheus Burger, Rodrigo Fernandes Ramalho, Elisa Napolitano e Ferreira, Otavio Jose Eulalio Pereira, Daniele Paixão, Caroline Olivati, Gustavo Marquezani Spolador, and Wagner Antonio da Rosa Baratela revised the article critically for important intellectual content and final approval of the version to be submitted.

ETHICS STATEMENT

This study was granted ethics committee approval from both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339). All patients or their legal guardians provided written consent before exome analysis.

DATA AVAILABILITY STATEMENT

All data are provided as "Supplementary Material."

ORCID

Caio Robledo D'Angioli Costa Quaio https://orcid.org/0000-0003-2873-2820

REFERENCES

Abouelhoda, M., Sobahy, T., El-Kalioby, M., Patel, N., Shamseldin, H., Monies, D., … Alkuraya, F. S. (2016). Clinical genomics can facilitate countrywide estimation of autosomal recessive disease burden. *Genetics in Medicine*, *18*(12), 1244–1249. https://doi.org/10.1038/gim.2016.37

Antonarakis, S. E. (2019). Carrier screening for recessive disorders. *Nature Reviews. Genetics*, *20*(9), 549–561. https://doi.org/10.1038/s41576-019-0134-2

Azimi, M., Schmaus, K., Greger, V., Neitzel, D., Rochelle, R., & Dinh, T. (2016). Carrier screening by next-generation sequencing: Health benefits and cost effectiveness. *Molecular Genetics & Genomic Medicine*, *4*, 292–302. https://doi.org/10.1002/mgg3.204

Baird, P. A., Anderson, T. W., Newcombe, H. B., & Lowry, R. B. (1988). Genetic disorders in children and young adults: A population study. *American Journal of Human Genetics*, *42*(5), 677–693.

Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., … Kingsmore, S. F. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Science Translational Medicine*, *3*, 65ra4.

Edwards, A. W. (2008). G. H. Hardy (1908) and Hardy-Weinberg equilibrium. *Genetics*, *179*(3), 1143–1150. https://doi.org/10.1534/genetics.104.92940

Edwards, J. G., Feldman, G., Goldberg, J., Gregg, A. R., Norton, M. E., Rose, N. C., … Watson, M. S. (2015). Expanded carrier screening in reproductive medicine-points to consider: A joint statement of the American College of Medical Genetics and Genomics, American College of Obstetricians and Gynecologists, National Society of Genetic Counselors, Perinatal Quality Foundation, and Society for Maternal-Fetal Medicine. *Obstetrics & Gynecology*, *125*(3), 653–662. https://doi.org/10.1097/AOG.0000000000000666

Esquivel-Sada, D., & Nguyen, M. T. (2018). Diagnosis of rare diseases under focus: Impacts for Canadian patients. *Journal of Community Genetics*, *9*(1), 37–50. https://doi.org/10.1007/s12687-017-0320-x

Fahim, A. T., Daiger, S. P., & Weleber, R. G. (2000). Nonsyndromic retinitis pigmentosa overview. In M. P. Adam, H. H. Ardinger, & R. A. Pagon (Eds.), *GeneReviews*. Seattle, WA: University of Washington. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK1417/

Fares, F., Badarneh, K., Abosaleh, M., Harari-Shaham, A., Diukman, R., & David, M. (2008). Carrier frequency of autosomal-recessive disorders in the Ashkenazi Jewish population: Should the rationale for mutation choice for screening be reevaluated? *Prenatal Diagnosis*, *28*(3), 236–241. https://doi.org/10.1002/pd.1943

Ferreira, C. R., van Karnebeek, C. D. M., Vockley, J., & Blau, N. (2019). A proposed nosology of inborn errors of metabolism. *Genetics in Medicine*, *21*(1), 102–106. https://doi.org/10.1038/s41436-018-0022-8

Godard, B., Ten Kate, L., Evers-Kiebooms, G., & Ayme, S. (2003). Population genetic screening programmes: Principles, techniques, practices, and policies. *European Journal of Human Genetics*, *11*(Suppl 2), S49–S87.

Henneman, L., Borry, P., Chokoshvili, D., Cornel, M. C., van El, C. G., Forzano, F., … Peterlin, B. (2016). Responsible implementation of expanded carrier screening. *European Journal of Human Genetics*, *24*(6), e1–e12. https://doi.org/10.1038/ejhg.2015.271

Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., … Schenck, A. (2016). Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *American Journal of Human Genetics*, *98*(1), 149–164. https://doi.org/10.1016/j.ajhg.2015.11.024

LaDuca, H., Polley, E. C., Yussuf, A., Hoang, L., Gutierrez, S., Hart, S. N., … Dolinsky, J. S. (2020). A clinical guide to hereditary cancer panel testing: Evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. *Genetics in Medicine*, *22*(2), 407–415. https://doi.org/10.1038/s41436-019-0633-8

Lazarin, G. A., Haque, I. S., Nazareth, S., Iori, K., Patterson, A. S., Jacobson, J. L., … Srinivasan, B. S. (2013). An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: Results from an

ethnically diverse clinical sample of 23,453 individuals. *Genetics in Medicine*, *15*, 178–186.
Mortier, G. R., Cohn, D. H., Cormier-Daire, V., Hall, C., Krakow, D., Mundlos, S., … Warman, M. L. (2019). Nosology and classification of genetic skeletal disorders: 2019 revision. *American Journal of Medical Genetics. Part A*, *179*(12), 2393–2419. https://doi.org/10.1002/ajmg.a.61366
Quaio, C. R. D. C., Moreira, C. M., Novo-Filho, G. M., Sacramento-Bobotis, P. R., Groenner Penna, M., Perazzio, S. F., … Baratela, W. A. D. R. (2020). Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *184*(4), 955–964. https://doi.org/10.1002/ajmg.c.31860
Reiter, J. F., & Leroux, M. R. (2017). Genes and molecular pathways underpinning ciliopathies. *Nature Reviews Molecular Cell Biology*, *18*(9), 533–547. https://doi.org/10.1038/nrm.2017.60
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … ACMG Laboratory Quality Assurance Committee. (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. *Genetics in Medicine*, *17*, 405–424.
Ropers, H. H. (2012). On the future of genetic risk assessment. *Journal of Community Genetics*, *3*, 229–236.
Sankaranarayanan, K. (1998). Ionizing radiation and genetic risks IX. Estimates of the frequencies of mendelian diseases and spontaneous mutation rates in human populations: A 1998 perspective. *Mutation Research*, *411*, 129–178.
Saunders, C. J., Miller, N. A., Soden, S. E., Dinwiddie, D. L., Noll, A., Alnadi, N. A., … Kingsmore, S. F. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Science Translational Medicine*, *4*, 154ra135. https://doi.org/10.1126/scitranslmed.3004041
Schorling, D. C., Kirschner, J., & Bönnemann, C. G. (2017). Congenital muscular dystrophies and myopathies: An overview and update. *Neuropediatrics*, *48*(4), 247–261. https://doi.org/10.1055/s-0037-1604154
Shearer, A. E., Hildebrand, M. S., & Smith, R. J. H. (1999). Hereditary hearing loss and deafness overview. In M. P. Adam, H. H. Ardinger, & R. A. Pagon (Eds.), *GeneReviews*. Seattle, WA: University of Washington. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK1434/
Slavin, T. P., Maxwell, K. N., Lilyquist, J., Vijai, J., Neuhausen, S. L., Hart, S. N., … Couch, F. J. (2017). The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ Breast Cancer*, *3*, 22. https://doi.org/10.1038/s41523-017-0024-8
Tangye, S. G., Al-Herz, W., Bousfiha, A., Chatila, T., Cunningham-Rundles, C., Etzioni, A., … Sullivan, K. E. (2020). Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *Journal of Clinical Immunology*, *40*(1), 24–64. https://doi.org/10.1007/s10875-019-00737-x
Wang, J., Lin, Z. J., Liu, L., Xu, H. Q., Shi, Y. W., Yi, Y. H., … Liao, W. P. (2017). Epilepsy-associated genes. *Seizure*, *44*, 11–20. https://doi.org/10.1016/j.seizure.2016.11.030

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Quaio, C. R. D. C., Chung, C. H., Perazzio, S. F., Dutra, A. P., Moreira, C. M., Filho, G. M. N., Sacramento-Bobotis, P. R., Penna, M. G., de Souza, R. R. F., Cintra, V. P., Carnavalli, J. E. P., da Silva, R. A., Paixão, D., Baratela, W. A. R., Olivati, C., Spolador, G. M., Santos, M. N. P., Pintao, M. C., Fornari, A. R. S., … Kim, C. A. (2021). Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 1–9. https://doi.org/10.1002/ajmg.c.31932

Molecular Biology Reports
https://doi.org/10.1007/s11033-022-07241-3

ORIGINAL ARTICLE

Frequency of carriers for rare metabolic diseases in a Brazilian cohort of 320 patients

Caio Robledo D' Angioli Costa Quaio[1,2,3,5] · Caroline Monaco Moreira[2] · Christine Hsiaoyun Chung[2] · Sandro Felix Perazzio[2,4] · Aurelio Pimenta Dutra[2] · Chong Ae Kim[1]

Received: 19 October 2021 / Accepted: 8 February 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Background Several metabolic disorders follow an autosomal recessive inheritance pattern. Epidemiological information on these disorders is usually limited in developing countries. Our objective is to assess carrier frequencies of rare autosomal recessive metabolic diseases in a cohort of Brazilian patients that underwent molecular investigation with exome sequencing and estimate the overall frequency of these diseases using the Hardy–Weinberg equation.

Methods and results We reviewed the molecular findings of 320 symptomatic patients who had carrier status for recessive diseases actively searched. A total of 205 rare variants were reported in 138 different genes associated with metabolic diseases from 156 patients, which represents that almost half (48.8%) of the patients were carriers of at least one heterozygous pathogenic/likely pathogenic (P/LP) variant for rare metabolic disorders. Most of these variants are harbored by genes associated with multisystemic involvement. We estimated the overall frequency for rare recessive metabolic diseases to be 10.96/10,000 people, while the frequency of metabolic diseases potentially identified by newborn screening was estimated to be 2.93/10,000.

Conclusions This study shows the potential research utility of exome sequencing to determine carrier status for rare metabolic diseases, which may be a possible strategy to evaluate the clinical and social burden of these conditions at the population level and guide the optimization of health policies and newborn screening programs.

Keywords Whole exome sequencing · Rare diseases · Carrier frequency · Recessive Mendelian diseases · Metabolic diseases · Inborn errors of metabolism

Background

Inborn metabolic diseases (in this article referred to as "metabolic diseases") are a complex group of disorders of genetic etiology that affect breakdown or biosynthesis of substances within specific pathways, recognizable by specific biochemical tests [1]. The majority of these disorders present monogenic etiology, autosomal recessive inheritance and multisystemic involvement. The determination of exact etiology may be challenging and involve biochemical studies searching for abnormal metabolites or enzymatic assays, genomic approaches with next generations sequencing (NGS) or even other growing methods, such as glycomics, proteomics or lipidomics [1].

A rapid diagnosis is key for favorable long-term prognosis for several of these disorders because (1) some affected individuals can die during an acute metabolic crisis and (2) several of these disorders are treatable by vitamins, special

✉ Caio Robledo D' Angioli Costa Quaio
cquaio@gmail.com

1 Instituto da Criança (Children's Hospital), Hospital das Clínicas (HCFMUSP), Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil

2 Fleury Medicina e Saúde, São Paulo, SP, Brazil

3 Laboratório Clínico, Hospital Israelita Albert Einstein, São Paulo, SP, Brazil

4 Division of Rheumatology, Universidade Federal de São Paulo, São Paulo, SP, Brazil

5 Instituto da Criança do Hospital das Clínicas da FMUSP – Unidade de Genética, Av. Dr. Enéas Carvalho de Aguiar, 647. Cerqueira César, São Paulo, SP CEP 05403-900, Brazil

diets, cleansing drugs or other newly-developed treatments [2]. For this reason, the majority of countries have created universal newborn screening for several treatable metabolic disorders that are potentially recognized through this early approach. These screening programs, though, may vary widely. The Newborn Screening ACT Sheets and Algorithms (ACMG ACT Sheets) [3] are a good resource for mapping metabolic diseases that may be identified through newborn screening.

Considering the clinical relevance of these conditions, the possibility of early identification of several diseases through newborn screening and the potential to dramatically change the outcomes with early treatment, algorithms that allow standardized recognition of these diseases become crucial. A recent thorough nosology has mapped 1015 metabolic diseases in 130 groups and an additional 111 diseases lacking definite classification or poorly characterized [4]. These 130 groups of metabolic diseases are divided into nine major groups: A. disorders of nitrogen-containing compounds; B. disorders of vitamins, cofactors, metals and minerals; C. disorders of carbohydrates; D. mitochondrial disorders of energy metabolism; E. disorders of lipids; F. Disorders of tetrapyrroles; G. Storage disorders; H. Disorders of peroxisomes and oxalate and I. Congenital disorders of glycosylation.

Epidemiological information on metabolic diseases may also be useful to guide health policies and newborn screening programs, though this information is often inadequate, particularly in developing countries. Our objective is to evaluate carrier frequencies of rare autosomal recessive metabolic diseases in a cohort of Brazilian patients that underwent molecular investigation with exome sequencing and estimate the overall frequency of these diseases using the Hardy–Weinberg equation.

Methods

Selection of cases, molecular analysis, and bioinformatics

We reviewed the molecular findings of exome sequencing from 320 symptomatic patients who had carrier status for recessive diseases actively searched. These patients represent a convenience sample and were selected from an original cohort of 500 symptomatic patients who had undergone molecular analysis for suspected diseases with genetic etiology from 2016 to 2020 in facilities of the Fleury Group. Full details of the clinical features of patients, molecular analysis, bioinformatics protocols, clinical data, and molecular data for primary and secondary findings were previously published [5]; a clinical summary is available at Supplementary Table 1. Details regarding molecular analysis, bioinformatics protocols and quality parameters are available in Supplementary Table 10.

All patients or their legal guardians provided consent before exome analysis, and this study was granted ethics committee approval from both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

Carrier status identification, variant classification and gene categories

Carrier status for autosomal recessive diseases was actively searched for all 320 patients according to the following protocol of analysis: at least two parallel analyses were performed to preselect variants considering (1) relevant reports from databases (e.g., ClinVar and HGMD) and the literature and (2) variants that were not reported in ClinVar that had an allele frequency $< 1\%$, had relevant molecular impact (e.g., potentially leading to loss of function) or an in silico prediction of a functional impact and was located in a clinically relevant gene; these preselected variants were then discussed by a board comprising three experts. Variants were classified according to ACMG guidelines [6].

We excluded the following from our analysis: (1) variants for common autosomal recessive diseases with high allele frequencies and low penetrance; (2) variants associated with no clinical impact (such as traits); (3) variants associated with conditions in which the diagnosis depends on other tests (e.g., variants in *BTD*, *G6PD*, *HFE* and *SERPINA1*); and (4) genes with pseudogenes with high homology (such as *CYP21A2*).

We have used nosologies, reviews and professional guidelines available in the literature to divide all genes into clinical categories. The overall frequency of pathogenic or likely pathogenic variants associated with all clinical categories initially studied (intellectual disability, cancer, muscular diseases, ciliopathies, skeletal disorders, immune disorders, epilepsy, hearing loss, retinitis pigmentosa and others) along with overall estimations of frequencies of autosomal recessive diseases were previously published [7].

In this article, we explore novel data regarding the rare variants associated with metabolic disorders; the group of genes part of metabolic disorders was selected based on the nosology by Ferreira et al. [4]. The complete list of variants and genes for metabolic disorders along with nomenclature and classification information of all variants are available in the Supplementary Tables 2 and 3.

The nosology divides all metabolic diseases into the following categories: A. Disorders of nitrogen-containing compounds (Groups 1–24); B. Disorders of vitamins, cofactors, metals and minerals (Groups 25–45); C. Disorders of carbohydrates (Groups 26–53); D. Mitochondrial disorders of energy metabolism (Groups 54–81); E. Disorders of

lipids (Groups 82–97); F. Disorders of tetrapyrroles (Groups 98–99); G. Storage disorders (Groups 100–108); H. Disorders of peroxisomes and oxalate (Groups 109–114); I. Congenital disorders of glycosylation (Groups 115–130); and finally, Disorders with individual families described or poorly characterized. We used this division to attribute each individual gene to a specific group of metabolic disorders.

Recessive disease frequency estimation

We used the Hardy–Weinberg equation to estimate disease frequencies (q^2) based on the respective carrier frequencies (2pq) observed in this study [8]. For this purpose, we considered random mating and the approximation $p \sim 1$.

Clinical impact

We used the Human Phenotype Ontology (HPO) database to map possible phenotypic abnormalities encountered in genes associated with metabolic diseases [9]. We chose the following HPO terms that correspond to common clinical findings in metabolic diseases: Intellectual disability (HP:0001249), Encephalopathy (HP:0001298), Seizure (HP:0001250), Acidosis (HP:0001941), Hypoglycemia (HP:0001943), Facial dysmorphism (HP:0001999), Growth abnormality (HP:0001507), Liver disease (HP:0001392), Muscle disease (HP:0003011), Kidney disease (HP:0000077) and Eye disease (HP:0000478).

Newborn screening diseases/genes

We used the Newborn Screening ACT Sheets and Algorithms (ACMG ACT Sheets) [3] for mapping metabolic diseases that may be identified through newborn screening. The main classes of metabolic diseases catalogued by these documents are: amino acidemias, organic acidemias, fatty acid oxidation disorders, galactosemias and lysosomal storage diseases. We used the IEMbase platform [10] and OMIM to assess the list of genes associated with these diseases. The complete list of metabolic diseases potentially identified through newborn screening along with the genes associated with these diseases are available in the Supplementary Table 8.

Results

Variant characteristics and frequency of heterozygotes

We observed a total of 205 occurrences of 172 different variants harbored by 138 different genes associated with metabolic diseases in 156 patients; 23 variants recurred more than once, ranging from two to six times. This result implies that almost half (48.8%) of the patients were carriers of at least one heterozygous P/LP variant for rare metabolic disorders. On the other hand, P/LP variants associated with metabolic disorders were not found in 164 (51.3%) patients. Individual variant data are available in Supplementary Table 2.

In total, 79 different variants were classified as pathogenic and 93 as likely pathogenic according to ACMG criteria. The majority of the 172 different variants was found in the literature (n = 124; 72.1%) and ClinVar Database (n = 128; 74.4%). In contrast, 39 variants (22.7%) were exclusively found in our work: these non-reported variants were classified as likely pathogenic because all of them were considered rare in population databases (frequency lower than 0.01%; ACMG PM2 criteria) and presented a predicted loss-of-function mechanism (such as stop codon, frameshift; ACMG PVS1 criteria). Table 1 shows the distribution of genes, variants and the estimated frequency of diseases (q^2; estimation per 10,000 individuals) per group of metabolic disorders.

Genes and their potential clinical impact

Considering all 138 different genes, 33 (23.9%) belong to group A (disorders of nitrogen-containing compounds), 21 (15.2%) to group D (mitochondrial disorders of energy metabolism) and 19 (13.8%) to group B (disorders of vitamins, cofactors, metals and minerals).

We summed the frequencies of all different P/LP variants per gene and the combined frequency of variants for the most frequent diseases and their related genes are as follows: (1) 1.9% for Gaucher disease (*GBA*), Phenylketonuria (*PAH*) and Aicardi-Goutieres syndrome (*RNASEH2B*); (2) 1.6% for Spastic paraplegia 7 (*SPG7*); (3) 1.3% for Hereditary fructose intolerance (*ALDOB*), Wilson disease (*ATP7B*), CPT II deficiency (*CPT2*), Tay-Sachs disease (*HEXA*), and *PMM2*-CDG [congenital disorder of glycosylation] (*PMM2*).

Considering HPO terms used to estimate the clinical impact and relevance of the genes identified in this study, we observed that the majority (n = 125; 90.6%) presented two or more clinical manifestations; more than half (n = 81; 58.7%) presented five or more manifestations; eleven genes (8%) presented a sole manifestation and two genes (1.4%; *CUBN*, *FMO3*) did not present any of the 11 clinical manifestation in HPO database. The complete list of genes and their corresponding HPOs are available in the Supplementary Tables 5 and 6. Table 2 shows the frequency of 11 clinical manifestations from the HPO database in the 138 genes identified in our study.

Table 1 Distribution of genes, occurrences of pathogenic or likely pathogenic (P/LP) variants, combined allele frequency and predicted frequency of affected individuals per group of metabolic disorders

Groups of metabolic disorders	Number of different genes (%)	Occurrences of P/LP variants (%)	Combined allele frequencies	Predicted frequency of affected individuals (q^2; per 10,000 people)
A. Disorders of nitrogen-containing compounds Groups 1–24	33 (23.9%)	51 (24.9%)	0.159	3.10
B. Disorders of vitamins, cofactors, metals and minerals Groups 25–45	19 (13.8%)	27 (13.2%)	0.084	1.25
C. Disorders of carbohydrates Groups 26–53	11 (8%)	15 (7.3%)	0.047	0.71
D. Mitochondrial disorders of energy metabolism Groups 54–81	21 (15.2%)	29 (14.1%)	0.091	1.44
E. Disorders of lipids Groups 82–97	14 (10.1%)	18 (8.8%)	0.056	0.78
F. Disorders of tetrapyrroles Groups 98–99	4 (2.9%)	6 (2.9%)	0.019	0.24
G. Storage disorders Groups 100–108	17 (12.3%)	31 (15.1%)	0.097	2.08
H. Disorders of peroxisomes and oxalate Groups 109–114	4 (2.9%)	7 (3.4%)	0.022	0.37
I. Congenital disorders of glycosylation Groups 115–130	15 (10.9%)	21 (10.2%)	0.066	1.00
Disorders with individual families described or poorly characterized	–	–	–	–
Total	138 (100%)	205 (100%)	0.641	10.96

The first column represents the groups of metabolic disorders according to the nosology by Ferreira et al. (2019); the second column represents the number of different genes per group of disorders; the third column congregates all occurrences of variants per group per group of diseases; the fourth column represents the combined allele frequencies per group of metabolic disorders in the cohort of 320 patients; the last column represents the predicted frequency of affected individuals (q^2; $p \sim 1$) per 10,000 individuals (a scale of color was used to compare the distribution of frequencies). The last row congregates all groups

Variants

Variants harbored by genes that belong to group A (disorders of nitrogen-containing compounds) were most frequent occurrences (n = 51; 24.9%); this group is followed by group G (storage disorders; n = 31; 15.1%) and group D (mitochondrial disorders of energy metabolism; n = 29; 14.1%). On the other hand, variants of group F (disorders of tetrapyrroles) were the least frequent (n = 6; 2,9%) and we did not find variants harbored by genes associated with disorders with individual families described or poorly characterized (n = 0). Table 1 (third column) shows the distribution of occurrences of pathogenic or likely pathogenic (P/LP) variants per group of metabolic disorders.

Metabolic diseases frequency estimation

As stated before, we used the Hardy–Weinberg equation to estimate disease frequencies (q^2) based on the respective carrier frequencies observed in this study. The combined frequency considering all metabolic diseases was estimated to be 10.96/10,000 (Supplementary Table 3).

Considering only the top nine frequencies of recessive metabolic diseases, their combined frequency is estimated to be 5.2 per 10,000 people. The individual frequencies for these nine diseases per 10,000 individuals are as follow: Gaucher disease (0.88), phenylketonuria (0.88), Aicardi-Goutieres syndrome (0.88), spastic paraplegia (0.61), hereditary fructose intolerance (0.39), Wilson disease (0.39), CPT II deficiency (0.39), Tay-Sachs disease (0.39), *PMM2*-CDG

Table 2 Clinical manifestations associated with metabolic diseases genes

Clinical manifestation	Number of genes	Percentage (total genes = 138)
Muscle disease HP:0003011	111	80%
Eye disease HP:0000478	96	70%
Growth abnormality HP:0001507	91	66%
Seizure HP:0001250	88	64%
Intellectual disability HP:0001249	76	55%
Liver disease HP:0001392	64	46%
Kidney disease HP:0000077	45	33%
Acidosis HP:0001941	44	32%
Hypoglycemia HP:0001943	27	20%
Facial dysmorphism HP:0001999	25	18%
Encephalopathy HP:0001298	19	14%

The table shows the number of genes presenting each manifestation (second column) and the corresponding frequency in percentage (last column) of 11 clinical phenotypes studied (first column). The clinical phenotypes and the complete list of genes associated with them was based on the Human Phenotype Ontology (HPO) database. A scale of color was used to compare the distribution of frequencies

Table 3 Distribution of genes, occurrences of pathogenic or likely pathogenic (P/LP) variants, combined allele frequency and predicted frequency of affected individuals per group of metabolic disorders potentially identified by newborn screening

Metabolic disorders potentially identified by newborn screening	Number of different genes	Occurrences of P/LP variants	Combined allele frequencies	Predicted frequency of affected individuals (q^2; per 10,000 people)
Amino acidemias	5	10	0.016	0.98
Fatty acid oxidation disorders	5	8	0.013	0.49
Galactosemias	1	1	0.002	0.02
Organic acidemias	10	12	0.019	0.44
Lysosomal storage diseases	3	9	0.014	1.00
Total	24	40	0.063	2.929

The first column represents the groups of metabolic disorders potentially identified by newborn screening according to the Newborn Screening ACT Sheets and Algorithms (ACMG ACT Sheets); the second column represents the number of different genes per group of disorders; the third column congregates all occurrences of variants per group; the fourth column represents the combined allele frequencies in the cohort of 320 patients; the last column represents the predicted frequency of affected individuals (q^2; $p \sim 1$) per 10,000 individuals (a scale of color was used to compare the distribution of frequencies). The last row congregates all groups

(0.39). Table 1 (last column) contains the predicted frequency of diseases per group of metabolic disorders.

Newborn screening diseases

A total of 40 variant occurrences (19.5% of total occurrences) are harbored by 24 different genes (17.4% of total genes) that are associated with diseases potentially identified by newborn screening. The combined allele frequency of these diseases is 0.0625, while the predicted frequency of affected individuals (q^2) is 2.93 per 10,000 people. Table 3 shows the distribution of genes, variants and the estimated frequency of diseases (q^2; estimation per 10,000 individuals) per group of metabolic disorders potentially identified by newborn screening.

Discussion

We studied the frequencies of carriers for rare metabolic recessive diseases in a cohort of 320 Brazilian individuals. We identified 205 occurrences (average of 0.64 occurrence per patient) of 172 different variants harbored by 138 different genes in 156 patients. We used carrier frequency data and the Hardy–Weinberg and estimated an overall frequency of recessive metabolic diseases to be approximately 0.11% (10.96/10,000 people).

The majority of the variants are harbored by genes associated with multisystemic involvement, which widens clinical and social burden of these conditions. Several of these conditions may manifest with acute crises that often require critical care and may involve seizures (64%), acidosis (32%), hypoglycemia (20%) or encephalopathy (14%). The majority of the diseases may also present with long-term complications, such as growth abnormality (66%), intellectual disability (55%) and other complications addressed in Table 2.

It is noteworthy that almost one fifth (18%) of these genes may present with facial dysmorphisms, which is a clinical characteristic generally regarded as non-suggestive of metabolic diseases by clinicians. We also found unexpectedly high frequency of muscle (80%) and ophthalmologic involvement (70%) in the genes harboring P/LP variants. These latter involvements may require specific clinical measurements and should not be neglected in the follow-up of these patients.

We have used the HPO database to estimate the clinical impact of the genes found in our study and discussed above. This approach, though, presents some limitations: HPO tends to be highly inclusive even for rare manifestations of diseases; this fact may have overestimated the multisystemic impact of the diseases. On the other hand, some diseases may be neglected; for instance, *CUBN* gene was not associated with any HPO term studied, though it is known that the conditions associated with this gene may present with growth anomaly, intellectual disability, renal involvement, etc. (OMIM#261100).

We have estimated the frequency of rare recessive metabolic diseases to be 10.96 per 10,000 people. Other studies using different approaches have estimated variable frequencies of metabolic disorders: 4.0 (40/100,000) [11], 12,76 (1/784) [12] or 15.6 (66/42,257) [13] per 10,000 people. These estimates, though, are usually challenging especially if they are based on biochemical-based newborn screening programs because several factors may influence the results, including the scope of diseases screened, deaths occurring before a diagnosis is made, ease of access for screening programs, diagnosis confirmation methods, specific regional clinical interests [12]. Additionally, the frequency of these diseases may vary widely in different population groups due to founder effects, endogamy, nonrandom mating, and cultural, religious, social and/or geographical isolation [14]. For example, in endogamous societies this rate can be as high as 8.4% in highly consanguineous populations [15].

Considering the relevant clinical impact of several metabolic diseases, the relatively common combined frequencies and the possibility to improve clinical outcome with early diagnoses, several countries have established newborn screening programs aiming rapid diagnosis and treatment

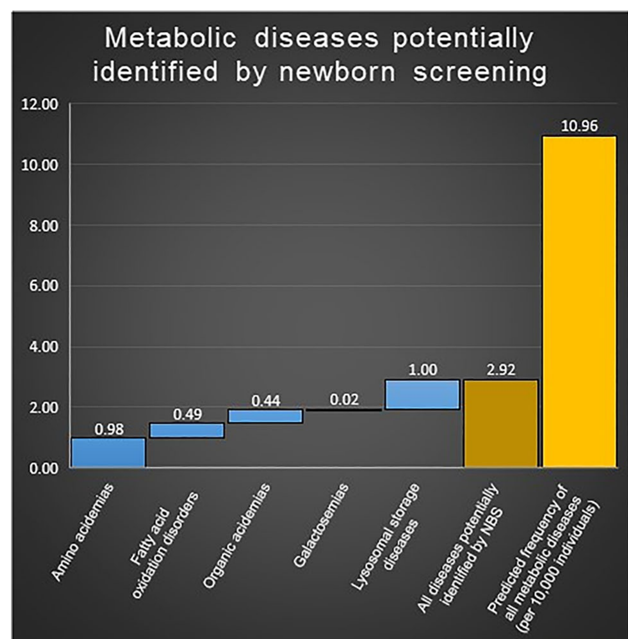


Fig. 1 Estimated frequency of rare recessive metabolic diseases potentially identified by newborn screening. The graphic compares the magnitude of estimated frequencies (q^2; numbers per 10,000 individuals) of all metabolic recessive disorders (light yellow) with the subgroups of diseases potentially identified by newborn screening (blue) and the consolidated frequency of all diseases potentially identified by newborn screening (dark yellow)

for clinically-actionable metabolic diseases. We have used the Newborn Screening ACT Sheets and Algorithms, from ACMG, to map the metabolic diseases potentially identified by newborn screening programs and we estimate the frequency of these diseases to be 2.93 per 10,000 people. Figure 1 helps to understand the magnitude of the estimated frequency of diseases potentially identified by newborn screening in the universe of predicted frequency of all metabolic disorders. If our estimates are accurate, we estimate that 26.6% (2.92/10.96) of patients with metabolic disorders may potentially be detected by neonatal screening. Amino acid studies would have the potential to identify 33.6% (0.98/2.92) of potentially detectable diseases, while acylcarnitine analysis the potential to identify 31.8% (fatty acid oxidation disorders: 0.49/2.92 + organic acidemias: 0.44/2.92) and a small fraction (0.7%; 0.02/2.92) would be detected by assays for galactosemia. An important number of detectable diseases rely on specific tests for lysosomal storage disorders (34.2%; 1.0/2.92) that are generally not widely available in neonatal screening programs.

Our study presents several limitations. The first refers to variant classification and pathogenicity presumption: several variants were classified as likely pathogenic based solely on ACMG criteria (39 variants exclusively found in this work received solely PVS1 and PM2 criteria) but were not definitively proven to be pathogenic (disease-causing); while this fact that may have overestimated the carrier frequencies by falsely classifying some variants as P/LP, we took precautions to standardize variant classification by strictly using ACMG classification criteria uniformly for all sequence variants. On the other hand, we may have excluded variants not detected by NGS, missense hypomorphic alleles that do not result in a clear and recognizable loss of function and several other variants that are not yet recognized to cause disease.

Another important bias that compromises the direct estimation of single variant frequencies or even single monogenic disease frequencies is the limited number of patients. However, we avoided estimations for single diseases and rather used the strategy of studying groups of diseases to estimate frequencies, then diluting the small sample of patients.

Our study did not consist of asymptomatic patients, which would be an ideal group for carrier screening studies (Supplementary Table 1). Precautions were taken, though, to minimize this bias by eliminating P/LP variants associated with autosomal recessive disorders reported in homozygosity or compound heterozygosity (primary findings) from this analysis because our objective was to study carriers for recessive diseases. Therefore, all patients in this study had only monoallelic variants. Nevertheless, we cannot rule out that some of the monoallelic variants studied in this work may have contributed to clinical phenotypes of some patients, for example, if the disease presents digenic or even oligogenic mechanism, an unknown codominant inheritance, or if the other allele harbored a pathogenic or likely pathogenic variant not detected by the NGS methodology (such as intronic or regulatory site variants). Another issue regarding heterozygous carriers is that some monoallelic variants may have clinical impacts and health complications, especially for some genes (including *POLG*) that may be associated with autosomal dominant forms that have a generally milder clinical impact.

Literature regarding carrier frequencies for autosomal recessive metabolic diseases is limited in Latin America. Most of previously published studies are limited to sole or restricted groups of diseases. Additionally, the majority rely on newborn screening findings without molecular confirmation. We hope that our study encourages other studies that aim to better understand the burden of recessive diseases in our developing countries.

Conclusions

We studied the frequencies of heterozygotes for rare pathogenic/likely pathogenic alleles of recessive metabolic diseases exome sequencing and observed a total of 205 occurrences of 172 different variants in almost half (48.8%) of the 320 patients. Most of these variants are harbored by genes associated with multisystemic involvement, most frequently muscle disease, eye disease and growth anomalies. Using the Hardy–Weinberg equation, we estimated the overall frequency for rare recessive metabolic diseases to be 10.96/10,000 people, while the frequency of metabolic diseases potentially identified by newborn screening was estimated to be 2.93/10,000 people. This study shows the potential utility to evaluate clinical and social burden of metabolic diseases and guide health policies for these conditions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-022-07241-3.

Author contributions All authors provided the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content and final approval of the version to be submitted.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Availability of data and material All data are provided as “Supplementary Material”.

Declarations

Conflict of interest The following authors are employees (received salary and other bonuses) of Fleury Medicina e Saude: CRDCQ, CHC, SFP, APD, CMM. Author CAK declares no financial interests.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. This study was granted ethics committee

approval from Fleury Group and Faculdade de Medicina da Universidade de São Paulo (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

Informed consent Informed consent was obtained from all individual participants included in the study. All authors are aware, consented and approved this publication.

References

1. Morava E, Rahman S, Peters V, Baumgartner MR, Patterson M, Zschocke J (2015) Quo vadis: the re-definition of "inborn metabolic diseases." J Inherit Metab Dis 38:1003–1006. https://doi.org/10.1007/s10545-015-9893-x
2. Saudubray JM, Garcia-Cazorla À (2018) Inborn errors of metabolism overview: pathophysiology, manifestations, evaluation, and management. Pediatr Clin N Am 65:179–208. https://doi.org/10.1016/j.pcl.2017.11.002
3. ACMG ACT Sheets and Algorithms [Internet]. Bethesda (MD): American College of Medical Genetics and Genomics; 2001-. Newborn Screening ACT Sheets and Algorithms. https://www.ncbi.nlm.nih.gov/books/NBK55827/
4. Ferreira CR, van Karnebeek CDM, Vockley J, Blau N (2019) A proposed nosology of inborn errors of metabolism. Genet Med 21:102–106. https://doi.org/10.1038/s41436-018-0022-8
5. Quaio CRDC, Moreira CM, Novo-Filho GM et al (2020) Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. Am J Med Genet C Semin Med Genet 184:955–964. https://doi.org/10.1002/ajmg.c.31860
6. Richards S, Aziz N, Bale S 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 17:405–424. https://doi.org/10.1038/gim.2015.30
7. Quaio CRDC, Chung CH, Perazzio SF et al (2021) Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. Am J Med Genet C Semin Med Genet 187:364–372. https://doi.org/10.1002/ajmg.c.31932
8. Edwards AW (2008) G. H. Hardy (1908) and Hardy-Weinberg equilibrium. Genetics 179:1143–1150. https://doi.org/10.1534/genetics.104.92940
9. Köhler S, Gargano M, Matentzoglu N et al (2021) The human phenotype ontology in 2021. Nucleic Acids Res 49:D1207–D1217. https://doi.org/10.1093/nar/gkaa1043
10. Lee JJY, Wasserman WW, Hoffmann GF, van Karnebeek CDM, Blau N (2018) Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. Genet Med 20:151–158. https://doi.org/10.1038/gim.2017.108
11. Applegarth DA, Toone JR, Lowry RB (2000) Incidence of inborn errors of metabolism in British Columbia, 1969–1996. Pediatrics 105:e10. https://doi.org/10.1542/peds.105.1.e10
12. Sanderson S, Green A, Preece MA, Burton H (2006) The incidence of inherited metabolic disorders in the West Midlands, UK. Arch Dis Child 91:896–899. https://doi.org/10.1136/adc.2005.091637
13. Zhang W, Yang Y, Peng W et al (2020) A 7-year report of spectrum of inborn errors of metabolism on full-term and premature infants in a Chinese Neonatal Intensive Care Unit. Front Genet 10:1302. https://doi.org/10.3389/fgene.2019.01302
14. Antonarakis SE (2019) Carrier screening for recessive disorders. Nat Rev Genet 20:549–561. https://doi.org/10.1038/s41576-019-0134-2
15. Keyfi F, Nasseri M, Nayerabadi S, Alaei A, Mokhtariye A, Varasteh A (2018) Frequency of inborn errors of metabolism in a Northeastern Iranian sample with high consanguinity rates. Hum Hered 83:71–78. https://doi.org/10.1159/000488876

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

scientific reports

OPEN

Parental segregation study reveals rare benign and likely benign variants in a Brazilian cohort of rare diseases

Caio Robledo D.'Angioli Costa Quaio[1,2,3,5], Jose Ricardo Magliocco Ceroni[1,3], Murilo Castro Cervato[3], Helena Strelow Thurow[3], Caroline Monaco Moreira[2], Ana Carolina Gomes Trindade[2], Cintia Reys Furuzawa[2], Rafaela Rogerio Floriano de Souza[2], Sandro Felix Perazzio[2,4], Aurelio Pimenta Dutra[2], Christine Hsiaoyun Chung[2] & Chong Ae Kim[1]

Genomic studies may generate massive amounts of data, bringing interpretation challenges. Efforts for the differentiation of benign and pathogenic variants gain importance. In this article, we used segregation analysis and other molecular data to reclassify to benign or likely benign several rare clinically curated variants of autosomal dominant inheritance from a cohort of 500 Brazilian patients with rare diseases. This study included only symptomatic patients who had undergone molecular investigation with exome sequencing for suspected diseases of genetic etiology. Variants clinically suspected as the causative etiology and harbored by genes associated with highly-penetrant conditions of autosomal dominant inheritance underwent Sanger confirmation in the proband and inheritance pattern determination because a "de novo" event was expected. Among all 327 variants studied, 321 variants were inherited from asymptomatic parents. Considering segregation analysis, we have reclassified 51 rare variants as benign and 211 as likely benign. In our study, the inheritance of a highly penetrant variant expected to be de novo for pathogenicity assumption was considered as a non-segregation and, therefore, a key step for benign or likely benign classification. Studies like ours may help to identify rare benign variants and improve the correct interpretation of genetic findings.

Genomic studies based on next-generation sequencing (NGS) technology are cost-effective alternatives to determine the molecular diagnosis of patients with rare diseases of monogenic etiology because they allow the concomitant study of several known genes associated with genetic conditions. On the other hand, these techniques may generate massive amounts of data and unravel several rare genomic variants, which brings challenges in the interpretation of these findings.

Variant classification is a systematic process that gathers different pieces of evidence from multiple sources, including scientific literature, control databases, in-silico predictors, among others, and aims a correct interpretation of genetic findings. The American College of Medical Genetics and Genomics in conjunction with the Association for Molecular Pathology (ACMG-AMP) created guidelines for the interpretation of sequence variants, becoming an important step toward a more uniform classification process. Among the recommendations, we highlight (1) the use of specific standard terminology: 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign' to describe variants identified in Mendelian disorders and (2) a standardized process for classification of variants based on criteria using typical types of variant evidence[1]. Among these types of evidence, segregation of a variant in an affected family may be evidence for pathogenicity; for several diseases, though, the non-inheritance (e.g., de novo event) is the expected mechanism for pathogenicity. On the other hand, lack of segregation of a variant with a phenotype may provide strong evidence against pathogenicity.

[1]Instituto da Criança (Children's Hospital), Hospital das Clínicas HCFMUSP, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, Brazil. [2]Fleury Medicina E Saúde, São Paulo, SP, Brazil. [3]Laboratório Clínico, Hospital Israelita Albert Einstein, São Paulo, SP, Brazil. [4]Division of Rheumatology, Universidade Federal de Sao Paulo, Sao Paulo, Brazil. [5]Instituto da Criança do Hospital das Clínicas da FMUSP – Unidade de Genética, Av. Dr. Enéas Carvalho de Aguiar, 647. Cerqueira César, São Paulo, SP CEP: 05403-900, Brazil. email: cquaio@gmail.com

Another category recommended by ACMG-AMP is in silico analysis. Many in silico algorithms have been developed to predict the impact of missense variants and some of these tools have demonstrated a superior performance, such as REVEL [2. 3].

The finding of a variant of unknown significance (VUS) can be problematic for patients and clinicians working in clinical genetics setting[4–8]. Although definitive re-classification of a VUS as pathogenic or benign may eventually occur, the timeline is typically many years, and may be indefinite for rare VUS, especially if the disease is uncommon[5,9,10]. Therefore, efforts for the differentiation of benign and pathogenic variants gain importance.

In this article, we used segregation analysis and other molecular data to reclassify to benign or likely benign several rare clinically curated variants of autosomal dominant inheritance from a cohort of 500 Brazilian patients with rare diseases.

Materials and methods

Selection of cases. This study is focused on novel data ("Additional Findings") from molecular findings obtained from exome sequencing (ES) analysis of 500 samples of adult, pediatric and fetal patients. Full details of the clinical features of patients, molecular analysis, bioinformatics protocols, clinical data, and molecular data for primary findings, secondary findings and carrier status for recessive diseases were previously published[10,11]. These samples were collected from 2016 to 2020 in facilities of the Fleury Group, which is one of the largest private diagnostic laboratories in Brazil.

This study has been performed in accordance with the Declaration of Helsinki, followed the best standards for scientific research and strictly followed Brazilian law for research involving human subjects. Written informed consent was obtained from all the participants. This study was granted ethics committee approval from Grupo Fleury and Faculdade de Medicina da Universidade de São Paulo (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

Primary findings were reported when pathogenic or likely pathogenic variants were observed in a gene that was associated with the patient's phenotype, with compatible zygosity and an adequate inheritance pattern; these cases were considered positive. Secondary findings (conditions unrelated to the indication for testing that might impact the health or quality of life of patients or other family members) were reported for pathogenic and likely pathogenic variants in the ACMG gene list and exceptionally in other genes determined to be clinically relevant by our team of specialists[12]. Full details of the primary and secondary findings can be found elsewhere[10]. The carrier status for pathogenic or likely pathogenic variants of autosomal recessive conditions has also been published elsewhere[11].

The samples were collected from exclusively symptomatic patients who had undergone molecular investigation for suspected diseases of genetic etiology. The clinical data were collected through a comprehensive pretest form completed by the attending physician or family, medical reports and clinical notes provided to the laboratory.

Molecular analysis and bioinformatics. Molecular analysis and bioinformatics protocols followed the exact workflow as published in previous works by our group[10,11]. In summary, DNA from the proband and both parents were extracted from peripheral blood leukocytes, saliva or prenatal samples of villus biopsy or amniotic fluid after appropriate cell culture. Exome capture used Agilent Clinical Research Exome v1 according to the manufacturer's instructions and sequencing was performed using an Illumina NextSeq platform. Exome data were aligned to the GRCh37.75/hg19 reference genome using the Burrows–Wheeler Aligner (BWA; version 0.7.17-r1188). Variants (single-nucleotide variants [SNVs] and indels) were identified following the best practices of the Broad Institute using the Genome Analysis ToolKit (GATK, version 3.8-0-ge9d806836) software and annotated using Variant Effect Predictor (VEP, version 88.14). All exomes met a minimum of 95% of target bases covered at > 10 ×. In-house bioinformatic pipelines were developed by a dedicated team of specialists. The mitochondrial genome and copy number variants were not studied.

Variant selection and segregation study. The variant analysis workflow has also been performed according to our previous works[10,11]. In summary, at least two parallel analyses were performed to preselect variants considering (1) clinical relevance and overlap of gene harboring the variant and the provand's manifestation, (2) allele frequency < 1%, (3) relevant functional impact, (4) relevant reports from databases (e.g., ClinVar and HGMD) and (5) previous reports from the literature. The first step of variant preselection consisted of filtering those already reported in ClinVar or HGMD as disease-causing and reclassifying them according to ACMG guidelines. Then, the same filtering-reclassifying approach would include (1) variants of predicted loss-of-function harbored by genes with evident sensitivity for haploinsufficiency and (2) rare missense variants harbored by genes with missense constraint. The last step in variant preselection included variants with allele frequency < 1% harbored by genes with clinical overlap with the proband's phenotype.

These preselected variants were then discussed in a board comprising three experts. After this first board meeting, selected variants would undergo Sanger confirmation in the proband and inheritance pattern determination when the parental samples were available. Paternity and maternity were not confirmed by any specific test in our cohort, although nonpaternity/nonmaternity could be inferred by genotyping multiple rare variants. Parents were considered non-affected, unless otherwise reported in the clinical notes provided to the laboratory.

The samples were analyzed by three distinct protocols as follows[10,11]: Protocol 1 (cases 1–289; n = 289), in which the samples were processed in our laboratory as described above and variants were interpreted using an in-house web interface (GTAC); Protocol 2 (cases 290–387; n = 98), in which the samples were processed in our laboratory as described above and variants were analyzed using a commercially available diagnostic decision support platform by Emedgene Technologies LTD (Tel-Aviv, Israel) (www.emedgene.com); Protocol 3 (cases

388–500; n = 113), in which the samples were sent to Centogene AG (Rostock, Germany) (www.centogene.com) and processed according to the protocols of this third-party laboratory, including DNA extraction, bioinformatic pipeline, CNV analysis, and variant analysis. For Protocol 3, the primary findings reported by Centogene were confirmed by in-house Sanger sequencing of the proband and genitors to determine the inheritance pattern.

Additional findings—rare clinically curated variants. Variants preselected as described above in the board of experts were preselected because they presented relevantly low allele frequencies (< 1%) and potential relevance to the clinical manifestation of the corresponding patient. All variants were clinically suspected as the causative etiology for all patients and were harbored by genes associated with highly penetrant conditions of autosomal dominant inheritance. As a de novo event was expected for pathogenicity assumptions, these variants underwent Sanger confirmation in the proband and inheritance pattern determination when the parental samples were available. Individual data for each case, along with the molecular data and segregation studies for all variants can be accessed in the Supplementary Material.

Variant classification protocol. All variants were reviewed and reclassified for the purpose of this article. The complete details, nomenclature and classification information are available in the Supplementary Material. Variants were classified according to ACMG guidelines[1] with assistance of the third-party ACMG calculator Varstation ® (VarsOMICS®, São Paulo, Brazil; www.varsomics.com). Internal adaptations of ACMG guidelines used in this work may be found below.

Internal adaptations of ACMG guidelines. ACMG presents specific rules to classify variants[1]. Each benign criterion is weighted as stand-alone (BA1), strong (BS1–4), or supporting (BP1–6). Benign variants need either a stand-alone (BA1) criterion or two strong (BS1–4), while likely benign variants need one strong and one supporting criterion or at least two supporting criteria. Although variant classification followed the ACMG guidelines, some adaptations were adopted especially for benign criteria BS4, BS1, BS2, BP4 and BP5:

BS4—Variant does not segregate with the disease: considering that the lack of segregation of a variant with a phenotype provides evidence against pathogenicity, this criterion was used if a variant expected to be de novo was inherited from unaffected parents in genotype-positive and phenotype-negative scenarios. If the gene harboring the variant is associated only with early-onset conditions, the non-segregation was considered a strong evidence of non-pathogenicity (BS4_strong), while if the gene is associated with later-onset manifestations the non-segregation was considered a supporting evidence (BS4_supporting). All conditions analyzed in this study were considered enough penetrant for the application of segregation criteria.

BP5 —Variant found in a case with an alternate molecular cause for the disease. When a variant was observed in a case with a clear alternate genetic cause of disease (i.e., variant in a case with an alternate "Primary Finding") for which the reported variant is unlikely to contribute or modulate expressivity of the primary finding, this was considered supporting evidence to classify the variant as benign (BP5_supporting). Primary finding has already been defined above.

BS1—Allele frequency is greater than expected for disorder: we have used a conservative approach to use this criterion: variants received a BS1_strong only when presented with a frequency greater than 0.1% for autosomal dominant diseases in the gnomAD database.

BS2—Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age: since the objective of our study is to analyze the impact of variants exclusively in the autosomal dominant inheritance pattern, we have used a threshold to apply BS2_strong of 5 individuals harboring the same variant in the gnomAD database.

PM2_supporting- Absent from controls for AD in gnomAD. We adjusted all PM2 criteria to PM2_supporting in agreement with recent recommendations from the Clingen working group.

PP3-pathogenic for REVEL score > 0.7, according to the recommendations for best practices of United Kingdom Association for Clinical Genomic Science[13].

BP4- REVEL score < 0.4, according to the recommendations for best practices of United Kingdom Association for Clinical Genomic Science[13].

Genes, diseases, inheritance and penetrance. Information about genes and their corresponding diseases along with inheritance mechanism and penetrance was recalled from The Clinical Genome Database website (https://research.nhgri.nih.gov/CGD/search; date of access: June, 2021) of the National Human Genome Research Institute (NHGRI), based in the United States[14].

Ethics approval. This study has been performed in accordance with the Declaration of Helsinki. This study was granted ethics committee approval from Grupo Fleury and Faculdade de Medicina da Universidade de São Paulo (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

Consent to participate. Written informed consent was obtained from all the participants.

Consent for publication. All authors are aware, consented and approved this publication.

Results

We have analyzed a total of 334 occurrences of 327 unique heterozygous variants (seven variants recurred twice) in 210 different genes.

Regarding the 210 different genes, 171 (81.4%) are solely associated with autosomal dominant diseases and this group of genes harbors 273 occurrences of 267 unique variants (81.7%). Another 37 genes (17.6%) are associated with autosomal dominant and autosomal recessive forms and this group harbors 59 occurrences of 58 unique variants (17.7%). Two genes (*SCNN1A* and *GPR98*; 1%) are associated with either autosomal dominant, autosomal recessive or digenic inheritances and harbor two occurrences of two unique variants (0.6%).

Still regarding the 210 different genes, 90 genes (42.9%) are associated with exclusive pediatric-onset conditions and harbor 145 occurrences of 141 unique variants (43.1%): 138 variants were inherited from asymptomatic parents and received BS4 (strong) criterion, while two variants (*GRIN2A* and *SCN1B* genes, both variants harbored by the same patient 274) were not found in both parents and in one case paternal sample was not available (*SCN1A* gene, case 47). Six genes (2.9%) that are associated with adult-onset conditions harbor eight occurrences of seven unique variants (2.1%), to which a BS4_supporting criterion was attributed. Finally, 114 genes (54.3%) that harbor 181 occurrences of 179 unique variants (54.7%) were not classified by NHGRI regarding the onset of the correspondent diseases: 176 variants were inherited from asymptomatic parents and were attributed the BS4_supporting criterion; two variants were not found in both parents (*GATAD2B* and *CIC* genes, cases 38 and 174 respectively) and in one case paternal sample was not available (*KCNC3* gene, case 109).

All 334 occurrences of 327 unique heterozygous variants were confirmed by Sanger sequencing in probands. Additionally, parental segregation studies were performed for both parents in 332 cases and paternal samples were not available for two cases (cases 47 and 109). Variants were inherited from mothers in 168 cases (50.3%); inherited from fathers in 158 (47.3%); were not found in both parents in four (1.2%); paternal sample was not available in 2 cases (0.6%) and in two other cases both parents were heterozygous for the same variant (0.6%).

Regarding the 327 unique variants, 139 (42.5%) were not found in the gnomAD database (gnomad211) and were attributed the PM2_supporting criterion; 87 variants (26.6%) presented from one and less than five controls in gnomAD. Another 100 variants (30.6%) were present in five or more controls in gnomAD and were attributed the BS2 criterion. Among these 100 variants, 11 (3.4%) presented a frequency higher than 0.1% in gnomAD and also attributed the BS1 criterion.

For 81 unique variants (24.8%), a clear alternate genetic cause for the patients' conditions was unraveled by ES. For these variants, BP5 criterion was used. All alternate causes for diseases (primary findings) were published elsewhere[10].

Figure 1 resumes the workflow of classification and shows the key benign criteria used to classify all 327 unique rare variants in the autosomal dominant model of inheritance. Regarding the ACMG classification, 51 unique variants (15.6%) were classified as benign, among which 47 are missense, one nonsense (stop gain) and three in-frame indels. Additionally, 211 unique variants (64.5%) were classified as likely benign: four stop gains, one start loss, one frameshift-predicted indel, four in-frame indels, two canonical splice-site alterations and the remaining 199 missense variants. Figure 2 contains Venn diagrams of benign and likely benign variants showing the distribution of all benign criteria used in this work.

Considering all variants classified as benign or likely benign (n = 262; 80.1%), we did not find dbSNP correspondence for 76 variants (29%) and ClinVar entrances were not found for 189 variants (72.1%). ClinVar entrances were available for 73 variants (27.9%), among which 15 were classified as benign or likely benign, 40 as VUS, one as likely pathogenic and for 17 variants ClinVar classification was conflicting among participating laboratories.

The remaining 65 unique variants (19.9%) were classified as variants of unknown significance. Two thirds of these variants (n = 44) were not found in the gnomAD database.

For unique variants harbored by genes associated with pediatric-onset conditions, 73% (103/141) were classified as benign or likely benign, while for variants harbored by genes associated with adult-onset conditions, 71.4% (5/7) were classified as likely benign. As for variants harbored by genes non-classified by NHGRI, 86% (154/179) were classified as benign or likely benign.

Discussion

We have used segregation analysis and other molecular data to reclassify rare variants to benign (n = 51) or likely benign (n = 211) in a Brazilian cohort of rare diseases. All variants are harbored by genes associated with autosomal dominant disorders. Proper interpretation of rare variants is a crucial step for adequate molecular diagnosis and, consequently, clinical management and genetic counseling. Therefore, efforts to distinguish pathogenic variants from rare benign variants are a key step in molecular diagnosis, but ascertaining which rare variants have clinical/health impact remains a major challenge.

We have used a systematic approach to preselect the variants in this study. Besides a low allele frequency in control databases, reports from literature and functional impact, all variants were clinically curated: they were harbored by genes associated with autosomal dominant disorders similar to all patients' manifestations. In other words, all variants were associated with autosomal dominant disorders and were judged highly suspicious to explain the final diagnosis that had led to exome sequencing.

Segregation study was assessed to investigate the clinical relevance of the rare variants found in our cohort because a de novo event was expected for all 327 suspicious variants. Segregation is important to assess the pathogenicity of a variant. Indeed, variant segregation in patients and corresponding families are widely used for novel disease-associated genes and pathogenic/likely pathogenic variants discoveries. In this study, we have used the same approach for the opposite reason: uncover benign/likely benign variants. The inheritance of a variant expected to be de novo for pathogenicity assumption was considered in our study as a non-segregation and, therefore, a key step for benign or likely benign classification. However, this was not the sole step in this process, as discussed below.

Figure 1. Workflow of classification of all 327 variants. The figure shows the workflow with the key benign criteria used to classify all 327 unique rare variants in the autosomal dominant model of inheritance. Green diamonds represent the main benign criteria used for variant classification; grey squares represent the number of variants flowing in the decision tree; blue squares congregate all variants classified as benign or likely benign and red squares represent the variants that did not meet minimum requirements for classification as benign or likely benign and kept a final classification of variants of unknown significance (VUS). Note that the 327 variants are initially divided into two groups: 81 variants received the BP5 (supporting) criterion because these variants occurred in patients with a clear molecular alternate cause of disease; 79 out of these 81 variants were inherited from predictively asymptomatic parents, received either BS4 or BS4_supporting criterion and were classified as benign (13 variants) or likely benign (66 variants); two out of these 79 were not inherited (one variant was not found in both parents and for the other one paternal sample was not available): one of them received the BP4 criterion and the other received both BS2 and BP4 criteria and both were classified as likely benign. The remaining 246 variants did not receive BP5 and followed the classification flow. Among the remaining 246 variants, 76 received BS2 criterion because they were found in at least five controls from gnomAD database: 75 were inherited from parents and received an additional BS4 or BS4_supporting and were classified as benign (38 variants) or likely benign (37 variants); one variant was not inherited (not found in parental samples), but as it received the BP4 criterion it was classified as likely benign. Among the remaining 170 that did not receive either BP5 and BS2 criteria, 106 were predicted as having benign effect by REVEL and received BP4 criterion: 104 variants were inherited from parents, received additional BS4 or BS4_supporting criteria and were all classified as likely benign; two variants were not found in parental samples and were classified as VUS since they received only BP4 criterion. Finally, 64 variants did not receive BP5, BS2 nor BP4 criteria: one of them was an inframe indel in non-functional domain (BP3), inherited from the mother (BS4_supporting) and was classified as likely benign; 63 variants did not meet minimum requirements for reclassification (62 were inherited and received only the BS4 or BS4_suporting criteria and one variant was not inherited) and were considered VUS.

Figure 1 resumes the key workflow steps of the variant classification. A total of 81 variants occurred in patients with a clear molecular alternate cause of disease and received the BP5 (supporting) criterion. All these 81 variants were classified as benign (n = 13) or likely benign (n = 66) because they also received, at least, another benign criterion either because were inherited from predictively asymptomatic parents (n = 79) receiving BS4 or BS4_supporting criterion or REVEL score suggested no impact (n = 2). Note that two variants were not inherited—non-inherited variants will be discussed below. In this branch of cases shown in Fig. 1, the combination of BP5 and BS4/BS4_supporting was key for non-pathogenicity assumption. Figure 2 demonstrates that this

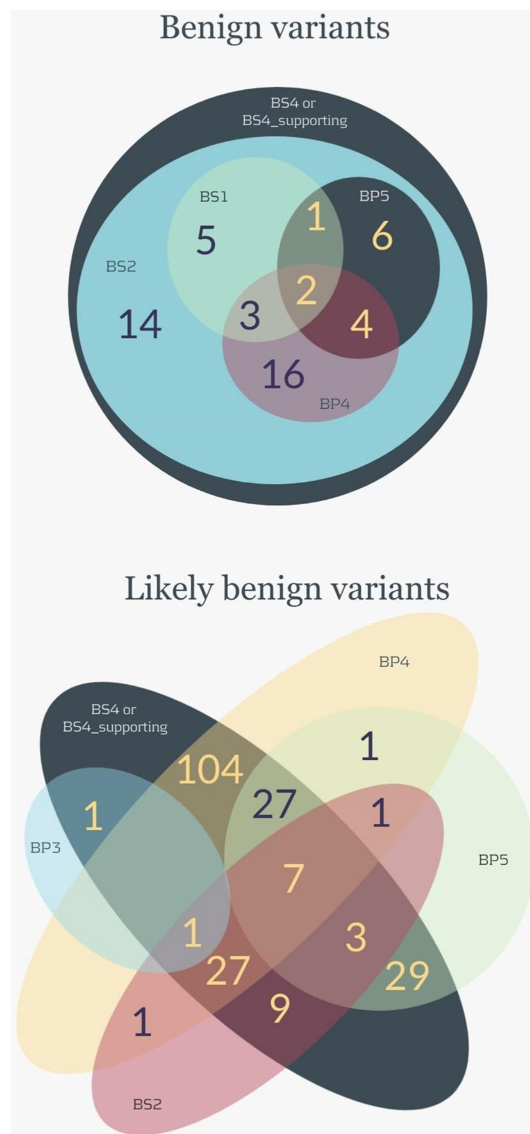


Figure 2. Venn diagrams of benign and likely benign variants. The figure shows the distribution of all benign criteria used in this work to classify 51 variants as benign (upper diagram) and 211 variants as likely benign (lower diagram).

combination of criteria happened for 13 benign variants and 66 likely benign variants: 29 likely benign variants received exclusively these two criteria and this combination was critical for final classification, while the remaining variants received other benign criteria.

Another key step in the classification is the frequency in the control database. In this study, we used only the gnomAD database, a fact that may have underestimated the "benign weight" of BS2 criterion because other databases might suggest higher frequencies especially for ethnicities not covered by gnomAD; on the other hand, the quality of variant calls and integrity of database might be an uncontrolled bias. All variants found in five or more controls in gnomAD were eventually classified as benign or likely benign because they received other benign criteria. Regarding the benign variants, 14 received exclusively the combination BS4 (variants inherited from parents and associated with conditions curated as pediatric onset) and BS2. As for the likely benign variants, nine variants received exclusively BS4_suporting (variants inherited from parents and associated with conditions not curated by NHGRI) and BS2.

A fourth key step was in-silico prediction. In this study, we have used REVEL, which is an ensemble method for predicting the pathogenicity of rare missense variants, as our sole in-silico predictor. REVEL has shown better performance when compared to other methods especially for rare neutral variants[2,3,15]. These new pieces of evidence demonstrate adequate strength of REVEL as a benignity predictor and make this tool fit adequately in our study, which aims to unravel rare neutral variants. The limited use for missense variants, though, is a limitation of REVEL approach.

Among the 170 that did not receive either BP5 and BS2 criteria, 106 were predicted as having benign effect by REVEL (score < 0.4) and received BP4 criterion; 104 of these variants were inherited from parents (received BS4 or BS4_supporting) and the combination of BP4 + BS4/BS4_supporting was critical for the final classification as likely benign.

Finally, 65 variants did not meet minimum requirements for reclassification as benign or likely benign and were considered VUS, among which 63 were inherited from a parent (received solely BS4 or BS4_supporting).

A total of six unique variants were not confirmed to be inherited from either parent in our study, among which four were not found in parental samples (a supposed de novo event) and for the remaining two, the paternal samples were not available, though these two variants were not found in the maternal samples. Half of these variants (n = 3) were classified as likely benign considering other molecular data. Although it is widely known that several de novo genome changes can be associated with a multitude of pathological conditions, a broad range of de novo molecular events is also characteristic for individuals obtained from a general human population[16].

For the majority of variants classified as benign or likely benign, we did not find a ClinVar classification (n = 189; 72.1%). Even for variants with ClinVar entrances, we found a concordance of classification with our study only for 15 variants.

Our study approach focuses on variant segregation with phenotype and makes several simplifying assumptions, including (1) parents were asymptomatic: we recognize that several dominant diseases may present varying degrees of severity and it may be difficult for clinicians to identify oligosymptomatic patients; (2) non-segregation was considered a strong benign criterion for child-onset diseases and a supporting criterion for later-onset diseases or onsets non-classified solely by NHGRI: NHGRI database lacks information for several genes, including some widely known early-onset genes (e.g., *ASH1L, KMT2A, DEAF1, TRPV4, CHD2, SRCAP, ANKRD11, RAI1, ZEB2, SHANK3, ARID1B*, and others) and this fact may have underestimated non-segregation (unclassified genes received a BS4_supporting); (3) variants associated with adult-onset conditions were a minority (seven unique variants) and segregation analysis for these cases may be more complex to understand: we opted to apply a BS4_supporting criterion considering that parents were asymptomatic and we did not have information of other affected family members, even though we may have overestimated the non-segregation effect for these seven variants; (4) our method applies for genetic sequence variants of high penetrance and dominant inheritance: for this model, we considered all conditions sufficiently penetrant for segregation/non-segregation purposes.

The model that we have used presents limitations for more complex inheritance, including unknown codominant, digenic or even oligogenic mechanisms or for recessive inheritance. It is noteworthy that some variants are harbored by genes associated with both autosomal dominant and autosomal recessive forms. However, all variants were clinically selected because only the autosomal dominant form of the corresponding disease was similar to patients' phenotypes and a de novo event was expected for pathogenicity assumption. Therefore, the non-segregation of these variants for the autosomal dominant model was considered relevant for non-pathogenicity.

It is important to note, though, that there are growing pieces of evidence that rare variants make important contributions to human phenotypic variation and disease susceptibility, though detecting the effects of rare variants in complex traits is challenging because 1) it generally requires very large sample sizes to achieve statistical power and 2) rare SNVs are population-specific, which implies difficulties for replication of disease associations across different populations[17]. Therefore, it is possible (or even likely) that a rare variant classified as B/LB for autosomal dominant Mendelian trait may have health/disease implications for complex traits, for recessive forms or even for unknown recessive forms of a gene.

Our study presents important limitations. First, it analyzed the impact of single variants in monogenic-based models of autosomal dominant diseases and did not consider more complex interactions that might modulate phenotypes. Another limitation is that several variants were classified as likely benign/benign based solely on ACMG criteria but were not definitively proven to be benign. We took several precautions to apply some ACMG benign criteria that may have underestimated their effects: BS2 required at least five controls from gnomAD, BS1 required a frequency of at least 0.1% in gnomAD (a frequency considered high for many rare conditions). Even though we followed the ACMG criteria for variant classification, we may have falsely classified some variants as B/LB since we relied solely on presumptively assuming mechanisms without proper functional studies.

Even taking extra precautions that underestimate the weight of several benign criteria (e.g., BS2, BS1) and underestimating the non-segregation for several conditions non-classified by NHGRI, the majority of variants (n = 262; 80.1%) were reclassified as benign or likely benign. One main reason for this is that it is a lot easier for a variant to be classified as B/LB than pathogenic or likely pathogenic because only two supporting benign criteria are enough for LB classification. We used the term "easier" for two reasons: easier because it requires less criteria for B/LB classification and easier because they do not have up- or -downgrades such as several pathogenic criteria. These assumptions have made us raise concerns regarding the wide application of benign criteria. They also made us unilaterally downgrade BS4 criterion for later-onset or non-classified-onset conditions. On the other hand, the addition of extra supporting benign criteria does not influence in the final classification: for instance, a variant receiving a combination of BS2 and BP4 will have the same likely benign classification of a variant receiving BS2, BS4_supporting, BP4 and BP5, though the likelihood for this later variant to be benign is greater than the former.

We did not find comprehensive studies that address segregation analysis and classification of B/LB variants in Brazil or any other Latin American countries. These countries are generally underrepresented populations in international databases and genetically heterogeneous with important genetic contributions from Amerindians, African-Americans and Western-Europeans. The more NGS and segregation studies become available in such nations, the more will be known of the rare regional benign variants.

Literature presents an important limitation regarding the description of benign/likely benign variants because studies and research journals generally present a strong bias for positive results. We believe that studies like

ours are valuable for identifying rare benign variants and this strategy may improve the correct interpretation of genetic findings.

Data availability

All data are provided as "Supplementary Material".

Received: 8 December 2021; Accepted: 3 May 2022
Published online: 11 May 2022

References

1. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424. https://doi.org/10.1038/gim.2015.30 (2015).
2. Tian, Y. *et al.* REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Sci. Rep.* **9**, 12752. https://doi.org/10.1038/s41598-019-49224-8 (2019).
3. Cubuk, C. *et al.* Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet. Med.* **23**, 2096–2104. https://doi.org/10.1038/s41436-021-01265-z (2021).
4. Culver, J. O. *et al.* Variants of uncertain significance in BRCA testing: evaluation of surgical decisions, risk perception, and cancer distress. *Clin. Genet.* **84**, 464–472. https://doi.org/10.1111/cge.12097 (2013).
5. O'Neill, S. C. *et al.* Tolerance for uncertainty and perceived risk among women receiving uninformative BRCA1/2 test results. *Am. J. Med. Genet. C Semin. Med. Genet.* **142C**, 251–259. https://doi.org/10.1002/ajmg.c.30104 (2006).
6. van Dijk, S. *et al.* Clinical characteristics affect the impact of an uninformative DNA test result: the course of worry and distress experienced by women who apply for genetic testing for breast cancer. *J. Clin. Oncol.* **24**, 3672–3677. https://doi.org/10.1200/JCO.2005.03.7259 (2006).
7. Vos, J. *et al.* The counsellees' view of an unclassified variant in BRCA1/2: Recall, interpretation, and impact on life. *Psychooncology.* **17**, 822–830. https://doi.org/10.1002/pon.1311 (2008).
8. Garrett, L. T. *et al.* Family studies for classification of variants of uncertain classification: Current laboratory clinical practice and a new web-based educational tool. *J. Genet. Couns.* **25**, 1146–1156. https://doi.org/10.1007/s10897-016-9993-2 (2016).
9. Lindor, N. M. *et al.* A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum. Mutat.* **33**, 8–21. https://doi.org/10.1002/humu.21627 (2012).
10. Quaio, C. R. D. C. *et al.* Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 955–964. https://doi.org/10.1002/ajmg.c.31860 (2020).
11. Quaio, C. R. D. C. *et al.* Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. *Am. J. Med. Genet. C Semin. Med. Genet.* **187**, 364–372. https://doi.org/10.1002/ajmg.c.31932 (2021).
12. Kalia, S. S. *et al.* 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. https://doi.org/10.1038/gim.2016.190 (2017).
13. Ellard, S., *et al. ACGS Best Practice Guidelines for Variant Classification 2020*. (Association for Clinical Genetics Science—ACGS). https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf (2020).
14. Solomon, B. D., Nguyen, A. D., Bear, K. A. & Wolfsberg, T. G. Clinical genomic database. *Proc. Natl. Acad. Sci. USA* **110**, 9851–9855. https://doi.org/10.1073/pnas.1302575110 (2013).
15. Ioannidis, N. M. *et al.* REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016 (2016).
16. Kloosterman, W. P. *et al.* Characteristics of de novo structural changes in the human genome. *Genome Res.* **25**, 792–801. https://doi.org/10.1101/gr.185041.114 (2015).
17. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69. https://doi.org/10.1126/science.1219240 (2012).

Author contributions

C.R.D.C.Q., J.R.M.C., M.C.C.; H.S.T.: provided the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content and final approval of the version to be submitted. C.M.M.; A.C.G.T.; C.P.R.; R.R.F.S.; S.F.P.; A.P.D.; C.H.C.; C.A.K.: provided acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content and final approval of the version to be submitted. All authors have reviewed and approved the final version of the manuscript.

Competing interests

The following authors are employees (received salary and other bonuses) of Fleury Medicina e Saude: Caio Robledo D'Angioli Costa Quaio; Caroline Monaco Moreira; Ana Carolina Gomes Trindade; Cintia Pereira Reys; Rafaela Rogerio Floriano de Souza; Sandro Felix Perazzio; Aurelio Pimenta Dutra; Christine Hsiaoyun Chung. All other authors have no competing/conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-11932-z.

Correspondence and requests for materials should be addressed to C.R.D.C.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

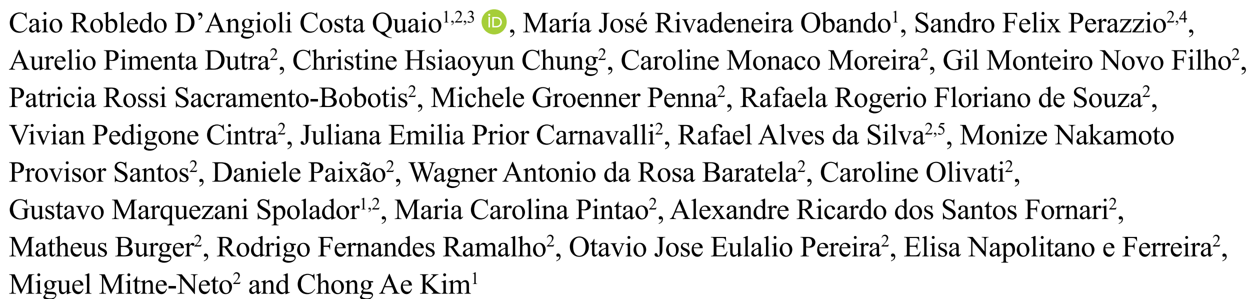
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

Genetics and Molecular Biology, 44, 4, e20210061 (2021)
Copyright © Sociedade Brasileira de Genética.
DOI: https://doi.org/10.1590/1678-4685-GMB-2021-0061

Research Article
Human and Medical Genetics

Exome sequencing and targeted gene panels: a simulated comparison of diagnostic yield using data from 158 patients with rare diseases



Caio Robledo D'Angioli Costa Quaio[1,2,3], María José Rivadeneira Obando[1], Sandro Felix Perazzio[2,4], Aurelio Pimenta Dutra[2], Christine Hsiaoyun Chung[2], Caroline Monaco Moreira[2], Gil Monteiro Novo Filho[2], Patricia Rossi Sacramento-Bobotis[2], Michele Groenner Penna[2], Rafaela Rogerio Floriano de Souza[2], Vivian Pedigone Cintra[2], Juliana Emilia Prior Carnavalli[2], Rafael Alves da Silva[2,5], Monize Nakamoto Provisor Santos[2], Daniele Paixão[2], Wagner Antonio da Rosa Baratela[2], Caroline Olivati[2], Gustavo Marquezani Spolador[1,2], Maria Carolina Pintao[2], Alexandre Ricardo dos Santos Fornari[2], Matheus Burger[2], Rodrigo Fernandes Ramalho[2], Otavio Jose Eulalio Pereira[2], Elisa Napolitano e Ferreira[2], Miguel Mitne-Neto[2] and Chong Ae Kim[1]

[1]*Universidade de São Paulo, Faculdade de Medicina, Hospital das Clínicas (HCFMUSP), Instituto da Criança, São Paulo, SP, Brazil.*
[2]*Fleury Medicina e Saúde, São Paulo, SP, Brazil.*
[3]*Hospital Israelita Albert Einstein, Laboratório Clínico, São Paulo, SP, Brazil.*
[4]*Universidade Federal de São Paulo, Divisão de Reumatologia, São Paulo, SP, Brazil.*
[5]*Universidade Federal de São Paulo, Escola Paulista de Medicina, Laboratório de Hepatologia Molecular Aplicada (LHeMA), São Paulo, SP, Brazil.*

Abstract

Next-generation sequencing (NGS) has altered clinical genetic testing by widening the access to molecular diagnosis of genetically determined rare diseases. However, physicians may face difficulties selecting the best diagnostic approach. Our goal is to estimate the rate of possible molecular diagnoses missed by different targeted gene panels using data from a cohort of patients with rare genetic diseases diagnosed with exome sequencing (ES). For this purpose, we simulated a comparison between different targeted gene panels and ES: the list of genes harboring clinically relevant variants from 158 patients was used to estimate the theoretical rate of diagnoses missed by NGS panels from 53 different NGS panels from eight different laboratories. Panels presented a mean rate of missed diagnoses of 64% (range 14%-100%) compared to ES, representing an average predicted sensitivity of 36%. Metabolic abnormalities represented the group with highest mean of missed diagnoses (86%), while seizure represented the group with lowest mean (46%). Focused gene panels are restricted in covering select sets of genes implicated in specific diseases and they may miss molecular diagnoses of rare diseases compared to ES. However, their role in genetic diagnosis remains important especially for well-known genetic diseases with established genetic locus heterogeneity.

Keywords: Exome sequencing, Next-generation sequencing, targeted gene panels, diagnostic yield, NGS panel.

Received: February 24, 2021; Accepted: August 09, 2021.

Introduction

Next-generation sequencing (NGS) has altered dramatically clinical genetic testing by reducing the cost and widening the access to molecular diagnosis of rare diseases with genetic etiology. Genomic capture is a step in the NGS methodology that can target specific areas in the genome, including subsets of genes of interest or even all known human genes, but this limits the study to what is already known in one hand and reduces the cost on the other (Teer and Mullikin, 2010). As there are numerous molecular tests available, including exome sequencing (ES) and vast amounts of gene panels, physicians may face difficulties selecting the best diagnostic approach for their patients with suspected genetic conditions.

Send correspondence to Caio Robledo D'Angioli Costa Quaio. Universidade de São Paulo, Faculdade de Medicina, Hospital das Clínicas (HCFMUSP), Instituto da Criança, Av. Dr. Enéas Carvalho de Aguiar, 647, 05403-000, São Paulo, SP, Brazil. E-mail: cquaio@gmail.com.

Focused gene panels are restricted in covering select sets of genes implicated in specific diseases. Therefore, they demand some sort of prior clinical diagnostic hypotheses driven by clinical examination and previous exams, which is difficult in several instances when dealing with rare diseases, especially when they present atypical presentations or clinical overlap with other conditions. Clinical laboratories have been providing this genetic diagnostic approach for more than a decade, though laboratories usually show different strategies in selecting the sets of genes for their panels. This makes the lists of genes, thus the sensitivity of the panels, vary widely between different panels from different labs and, sometimes, different panels within the same lab. Another remaining challenge for gene panels is that, given the vast number of genes potentially involved in atypical presentations of rare disorders, and our fast growing understanding of rare diseases, choosing a gene panel which is both specific enough to detect the cause of the suspected genetic disease and broad enough to

include all known genes associated with the condition and all differential diagnoses is often a challenging task (Xue *et al*., 2015). On the other hand, there are advantages in choosing a focused panel: greater coverage of targeted regions and higher depth per base, relatively easier interpretation of the report results and avoiding secondary findings, which are occult deleterious genetic variants associated with clinically relevant conditions besides the primary objective of analysis (Klitzman *et al*., 2013).

ES can be considered a more unbiased diagnostic method for monogenic diseases because it allows a much broader net for molecular diagnoses. In this method, the great majority of known protein-coding human genes are captured. Although this approach is not limited to specific sets of genes and theoretically includes the great majority of human genes, it still relies on clinical data for the interpretation of genomic information (Xue *et al*., 2015).

In this article, we simulate a comparison to estimate diagnostic yields of targeted gene panels from different laboratories using data from a cohort of patients with rare genetic diseases diagnosed with ES. For this purpose, we tested whether our patients diagnosed with ES would have been diagnosed by NGS panels by checking if the gene that harbored pathogenic variants was part of the list of genes of the respective panel.

Material and Methods

We reviewed the clinical indications and primary findings of ES analysis from our cohort of 500 symptomatic patients who had undergone molecular investigation for suspected diseases of genetic etiology from 2016 to 2020 in facilities of the Fleury Group. All patients or their legal guardians provided consent before exome analysis and this study was granted ethics committee approval from both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339).

Selection of patients, molecular analysis, bioinformatics protocols

A total of 500 patients had been referred for exome sequencing in our original cohort (cases #1 through #500; Quaio *et al*., 2020, 2021). All the samples were collected from 2016 to 2020 in facilities of the Fleury Group, a tertiary private general diagnostic laboratory with subsidiaries in nine states of Brazil and based in São Paulo, Brazil. This cohort is restricted to symptomatic patients who had undergone molecular investigation for suspected diseases of genetic etiology. All patients or their legal guardians provided consent before exome analysis. Patients who refused to share genomic data or receive information on medically actionable findings were excluded from this cohort. Additionally, exome analysis performed for other reasons than clinical diagnostic (such as research protocols or prenatal genetic counseling) were excluded from our review. Thirty-one cases that did not meet these inclusion criteria were excluded.

DNA from the proband and both parents were extracted in a clinical setting from peripheral blood leukocytes or saliva. Exome capture was also conducted in a clinical setting using Agilent Clinical Research Exome v1 according to the manufacturer's instructions. Sequencing was performed using an Illumina NextSeq platform. Exome data were aligned to the GRCh37.75/hg19 reference genome using the Burrows–Wheeler Aligner (BWA; version 0.7.17-r1188). Variants (single-nucleotide variants [SNVs] and indels) were identified following the best practices of the Broad Institute using the Genome Analysis ToolKit (GATK, version 3.8-0-ge9d806836) software and annotated using Variant Effect Predictor (VEP, version 88.14). All the exomes met a minimum of 95% of target bases covered at >10×. In-house bioinformatic pipelines were developed by a dedicated team of specialists. The mitochondrial genome and copy number variants were not studied.

Variant classification, clinical data, and molecular findings

At least two independent NGS analysts were responsible to preselect variants considering an allele frequency <1%, variant functional impact, clinical relevance of gene harboring the variant, relevant reports from databases (e.g., ClinVar and HGMD) and the literature. These preselected variants were then discussed in a board comprising three experts. After this first board meeting, selected variants would undergo through confirmation and segregation studies by Sanger sequencing when the parental samples were available. Paternity and maternity were not confirmed by any specific test in our cohort, although nonpaternity/nonmaternity could be disclosed by genotyping multiple rare variants. Variants were classified according to ACMG guidelines with assistance of a third-party ACMG calculator by Saphetor SA (Lausanne, Switzerland) (www.varsome.com).

The clinical data were collected through a comprehensive pretest form completed by the attending physician or family, medical reports and clinical notes provided to the laboratory. Patients were not evaluated by specialists from our team. For the purpose of this article, the clinical features provided to the laboratory were divided as follows: 1) Neurodevelopmental disorders, 2) Seizure, 3) Syndromic/malformative, 4) Immune/hematological diseases, 5) Cardiovascular disease and 6) Metabolic abnormality. One patient may be part of different groups, depending on his/her clinical manifestations. This clinical division was used to select NGS panels, as stated below. Clinical details of all patients are available in Table S1.

Primary findings were reported in 158 patients. We defined as primary findings the variants considered as clinically relevant, observed in a gene that was associated with the patient's phenotype, with compatible zygosity and an adequate inheritance pattern; these cases were considered positive. The other 342 patients had a negative exome analysis.

Considering all 158 patients to whom primary findings were reported, we have the following distribution among clinical groups: Neurodevelopmental disorders (n=82), Seizure (N=36), Syndromic/malformative (n=72), Immune/hematological diseases (n=18), Cardiovascular disease (n=16) and Metabolic abnormality (n=17). One patient may be part of different groups. This clinical division was used to select NGS panels, as stated below. Clinical details of all patients are available in Table S1.

ES was considered in this work as the gold-standard for molecular diagnosis of monogenic, rare diseases because it is

the most comprehensive genetic exam clinically available in Brazil nowadays. The list of genes harboring clinically relevant variants of the 158 patients from our cohort of rare diseases that underwent ES was used to estimate the theoretical rate of diagnoses missed by NGS panels from different laboratories.

Our team selected eight genetic clinical laboratories based on their national and international recognition of excellence, comprehensiveness of portfolios and availability in Brazil. Five of them are national laboratories, among which one is a public institution and the others are private providers. The other three are international laboratories: two from the United States and one from Germany. The panels and respective gene lists were selected according to the clinical indication stated above and information publicly available in laboratory websites and direct contact by email or call center. The names of the laboratories will not be disclosed because the objective of our study is not the comparison between different services, but the comparison of diagnostic methods: exome x targeted gene panels.

We selected 53 different NGS panels from eight different laboratories to simulate the estimated rate of possible missed diagnoses based on the list of genes available studied by these panels. Commercial provider websites were accessed between October and November 2020 to obtain gene coverage. The panels were required to sufficiently cover the differential diagnoses provided by clinical notes; if there were more than one commercial option available for each panel, we tried to identify the most comprehensive options considering gene lists. In some cases, we identified the availability of two options for the same laboratory and both options were included in the analysis. Gene coverage details for every panel and for each patient are available in Tables S2-S7. We took the conservative approach and assumed that the analytical sensitivity and specificity of all panels were 100% based on a previous study (Dillon *et al*., 2018).

Unpaired Student's *t* test was used to determine significantly different means.

Results

Our cohort originally consisted of 500 symptomatic patients who had undergone ES for diagnostic purpose. At least one molecular diagnosis of a genetic disorder was reported for 31.6% (n=158). These 158 cases with positive findings were used to simulate the rate of estimated diagnoses missed by NGS panels by calculating how many of them would not have had a definitive molecular diagnosis if they had been investigated by NGS panels instead of ES. Individual data for each case (brief case overview) are available in Table S1.

The estimation of missed diagnosis per Clinical manifestation and also per laboratory is summarized in Table 1. Overall, the mean rate of missed molecular diagnoses considering all 53 panels was 64%, varying from 14% (Lab H, Seizure panel; panel predicted sensitivity of 86%) to 100% (Lab A, Immune/hematological diseases panel; panel predicted sensitivity of 0%). Metabolic abnormalities represent the group of conditions with highest mean of missed diagnoses (86%; predicted average sensitivity of 14%), while seizure represents the group with lowest mean (46%; predicted average sensitivity of 54%).

Based on each laboratory individually, the mean of missed diagnoses considering all panels per laboratory varied from 51% (predicted laboratory sensitivity of 49%) to 81% (predicted laboratory sensitivity of 19%). Panels from Brazilian laboratories missed more diagnoses (68%, corresponding to a sensitivity of 32%) compared to international laboratories (58%, corresponding to a sensitivity of 42%); this difference showed statistical significance (Unpaired t test, $p<0.0001$; 95% confidence interval of this difference: from 9.29933 to 10.70067).

We have estimated that the rate of missed molecular diagnosis of 158 patients with monogenic diseases would be on average 64% if they had been studied by NGS panels instead of ES, which represents a sensitivity of 36%. This rate ranged widely from 14% to 100%, depending on the clinical indication for testing and the laboratory where the test was performed.

The rate of missed diagnoses varied not only within different labs, but also when comparing different clinical indications and surprisingly different panels for the same diseases within the same laboratory. The biggest variation within the same lab comprises two different panels for Seizure in Lab E, with missed diagnoses rates of 25% and 83%. Patients with metabolic abnormalities and cardiovascular diseases were estimated with rates of missed diagnoses above 70%, which represented the highest rates, while the only group with missed diagnoses rate less than 50% corresponded to patients with seizure. Based on these rates, we have estimated the diagnostic yield of NGS panels, shown in Table 2.

Discussion

We used data from a cohort of patients with rare genetic diseases diagnosed with ES to estimate possible molecular diagnoses missed by different NGS panels from different laboratories and observed that targeted gene panels might miss an average of 64% of the diagnoses of our cohort. Although these high rates of missed diagnoses are remarkable, our methodology presents several limitations that does not allow to a definite conclusion towards the absolute advantage of exome sequencing over targeted gene panels, as we will discuss below.

Literature regarding comparison of NGS diagnostic methods is limited in Latin America and in other populations outside North America and Europe. As we did not find comprehensive studies that address this subject in Brazil or any other Latin American countries, we hope that our data may enrich this discussion and encourage further studies from other groups.

Our objective was to simulate the proportion of ES-diagnosed patients that would not have been covered by targeted gene panels offered by commercial-testing laboratories available in Brazil. For this purpose, we selected 53 different targeted gene panels from eight different clinical laboratories. We observed that the selection of genes for similar clinical indications varied significantly among panels from different clinical laboratories, possibly widening the range of missed diagnosis rates for each laboratory individually. These differences in the number of genes and also in the types of genes included in each panel rely basically on the criteria of inclusion that may reflect not only evidence from the literature, but also commercial purposes.

Table 1 – Rate of molecular diagnoses missed by NGS panels from the eight laboratories per clinical manifestation. In the first column, "n" represents the number of patients in the group; Labs A through E are Brazilian national laboratories, while F through H are international; column "Mean" represents the mean of missed diagnoses per group of Clinical Manifestation.

Clinical manifestation	Number of panels	Rate of missed diagnoses of NGS panels by Laboratory									
		Lab A	Lab B	Lab C	Lab D	Lab E	Lab F	Lab G	Lab H	Mean rate of missed diagnoses	Predicted average sensibility
Neurodevelopmental disorders (n=54)	9	74%	N/A	69%	22-87%	65%	22-39%	87%	48%	57%	43%
Seizure (n=36)	11	58-61%	64%	42%	25%	25-83%	47%	42%	14-47%	46%	54%
Syndromic/ malformative (n=72)	4	N/A	63%	61%	60%	N/A	53%	N/A	N/A	59%	41%
Immune/ hematological diseases (n=18)	9	89-100%	28%	28-78%	N/A	22%	33%	33%	39%	50%	50%
Cardiovascular disease (n=16)	10	81-88%	88%	63%	88%	69%	69%	88%	69-75%	78%	22%
Metabolic anomaly (n=17)	10	94%	94%	88%	59%	94%	53-88%	82-94%	94%	84%	16%

Table 2 – Diagnostic yield of exome sequencing in our cohort of 500 patients with rare diseases (Quaio CRDC *et al*., 2020) and projected diagnostic yield of NGS panels per clinical indication in the same cohort considering the rates of missed molecular diagnoses shown in Table 1.

Clinical manifestation	Diagnostic yield	
	Exome sequencing	Projected yield for NGS panels
Neurodevelopmental disorders	35%	15%
Seizure	34%	18%
Syndromic/malformative	41%	17%
Immune/hematological diseases	19%	10%
Cardiovascular disease	31%	7%
Metabolic anomaly	33%	5%

The discrepancies between the gene panels might also be a result of a lack of consensus in the process of determining the genes clinically relevant to be included in the test as the inclusion criteria may be somewhat subjective.

Another important issue regarding targeted gene panels is that they may become outdated very quickly; some studies have demonstrated that an important number of genetic diagnoses encompass newly discovered diseases: around one fourth are within genes characterized within the last two years and 7% are novel gene discoveries (Farwell *et al.*, 2015; LaDuca *et al*., 2017). Additionally, another important advantage of ES is the possibility of regular reanalysis of negative cases, which has been demonstrating as beneficial in the individual, familial and health system perspectives and improve diagnostic rates over time (Fung *et al*., 2020). Considering these data, we expect that targeted gene panels dating more than two years are more likely to miss an average of almost one third of molecular diagnoses compared to ES.

Several laboratories offer different panels for the same clinical purpose, which is an important limitation not only to the selection of panels that we have made in this study, but also for the clinicians in the clinical setting. Choosing the most appropriate set of genes is sometimes a challenge, as we will try to demonstrate through examples faced in this research.

Patient 126 was referred for the investigation of suspected metabolic disease because he presented weakness and hepatosplenomegaly. The clinical report also mentioned delayed tooth eruption, but it did not mention any details about myopathy or skeletal anomalies. Based on the clinical description provided, this case was attributed to the clinical group "Metabolic abnormality" and correspondent panels were to investigate whether this diagnosed would be missed. This patient was found with a dual molecular diagnosis (*DYSF*-related muscular dystrophy and *PTH1R*-related tooth agenesis) genes that were not included in any metabolic panel considered for this study. In this case, we believe that the clinical evaluation prior to molecular exam may have contributed for a negative targeted panel as gene panels are usually designed including genes associated with a specific phenotype spectrum and, in this example, our patient showed clinical features from two different phenotypic spectrum.

Here are some other examples of restricted clinical information provided to the laboratory: case 95 (clinical report mentioned hypoglycemia and clinical hypothesis of inborn error of metabolism), 211 (report of hypoglycemia and hepatomegaly) and 424 (report of hepatomegaly, hypoglycemia and jaundice: clinical suspicion of glycogen storage disease). These three cases were diagnosed as mitochondriopathies

by ES (respectively *MPV17*, *POLG* and *MPV17*-related mitochondrial dysfunctions) and, probably would have been missed by most of the targeted gene panels studied here because several laboratories did not include mitochondrial genes in their most comprehensive metabolic panels.

These examples above raise an important factor that does not rely on the quality of gene selection criteria, but the importance of a comprehensive clinical evaluation in order to select the most sui genetic testing approach in order to confirm the clinical diagnosis. Some key factor may compromise this pre-test step for targeted gene panels: 1) several ordering physicians are not specialists in rare diseases and may lack a thorough knowledge in this field of expertise; 2) atypical manifestations of rare diseases may skew clinical evaluation towards a false clinical diagnosis and, therefore, an inadequate selection of gene panel.

Many studies have reported on the clinical utility of epilepsy gene panel testing, although they may present substantial variability in their diagnostic rates, ranging from as low as 10% to as high as 50% (Mercimek-Mahmutoglu *et al*., 2015; Butler *et al*., 2017; Rim *et al*., 2018; Jang *et al*., 2019). We have predicted diagnostic rate for epilepsy panels to be an average of 18%, ranging from 5.8% (Laboratory E) to 29.2% (Laboratory H).

Some studies have also compared ES to targeted gene panels for neurodevelopmental disorders. One study compared gene panels from 21 laboratories and observed that the number of genes ranged from 11 to 2,562 and ES diagnoses were made in genes not included in at least one commercial panel in 42% of cases (Dillon *et al*., 2018). In our study, we have predicted a mean rate of missed diagnoses slightly higher (57%), ranging from 22% (Laboratories D and F) to 87% (Laboratories D and G).

Our evaluation of metabolic diseases panels presents the most discrepant estimative for diagnostic rate. While we have estimated an average sensibility of 15% (ranging from 6% [Laboratories A, B, E, G, H] to 44% [Laboratory F]), studies have shown discrepant results varying widely, from 9.8% up to 50% (Ghosh *et al*., 2017; Lee *et al*., 2020). Some studies, though, have obtained a low diagnostic rate (around 15%) for cases with non-specific markers for metabolic diseases (Yubero *et al*., 2016). This is probably one of the most challenging group of diseases, because it encompasses hundreds of different clinical entities and several factors may interfere in diagnostic rates, including symptom onset age, gender, clinical course, organ systems involved, developmental history, family history and prior genetic testing (Lee *et al*., 2020).

Although ES offers a much broader range of genes compared to any panel, and studies have found that ES can detect more than 98% of pathogenic variants identified on different subset of gene panels (LaDuca *et al*., 2017), it is important to highlight that for several well-known genetic diseases with established genetic locus heterogeneity for the great amount of cases, a gene panel approach remains appropriate for an efficient and timely molecular diagnosis (Xue *et al*., 2015). In fact, targeted gene panels may offer advantages over ES in some instances, such as reduced cost, shorter turn-around time, lack of secondary findings and higher coverage of targeted genes. Therefore, more genes are not necessarily better. We believe that two main clinical categories may have more benefits from targeted gene panels compared to exome sequencing: cancer susceptibility genes, to which panels generally present more efficient enrichment allowing the study of recurring atypical, noncoding molecular mechanisms (e.g., recurring known alterations in promoter or intronic regions), and X-linked intellectual disability, in which ES presents worse coverage compared to targeted panels (LaDuca *et al*., 2017).

Even though ES can bypass the need of a previous selection of genes driven by clinical diagnostic hypotheses, variant interpretation still remains dependent on clinical data. When clinical information of the patient is incomplete, the analysis and interpretation of ES might be more challenging when compared to panels due to the greater amount of data obtained. Another important issue to consider in ES is that this diagnostic method may unintentionally unravel occult deleterious genetic variants not associated with the primary objective of analysis, although associated with clinically actionable conditions, defined as “incidental” or “secondary findings” (Klitzman *et al*., 2013). On one hand, secondary findings may impact on individuals’ psychological health negatively by unfolding previously unknown serious risks of future life-threatening events, on the other hand they may positively impact life expectancy (Quaio *et al*., 2020).

Our work presents a strong selection bias because our cohort consists mostly of patients with nonspecific clinical manifestations. It is not difficult to suppose that patients with clinical manifestations highly specific for known clinical conditions caused by a restricted subset of genes tend to undergo investigation through NGS panels. This fact may have substantially underrepresented this group of genetic diseases in our cohort, thus overestimating the rate of missed diagnosis by NGS panels.

Another limitation is that gene selection of targeted gene panels is dynamic and changes over time. As our work is based on the October/November 2020 versions of gene panels, our results may overestimate the rate of missed diagnosis of more up-to-date panels; on the other hand, we may underestimate the rate of missed diagnosis of patients that underwent older versions of gene panels, especially if we consider that NGS panels may become outdated rather quickly (LaDuca *et al*., 2017).

Information about other genetic investigations (e.g., NGS panel) prior to ES was not available. Additionally, several cases lacked thorough phenotyping and the clinical features relied solely on information provided to the laboratory, restricting the quality and quantity of phenotype information available for clinical-molecular correlation at the time of interpretation. The inclusion criteria for clinical cases were broad and may have included patients without a proper referral for ES, particularly considering that not all the patients were evaluated by clinical geneticists, some were referred by general practitioners, and other specialists without any training/experience in clinical genetics or dysmorphology.

The patients submitted to ES that remained without a molecular diagnosis continue to be a big challenge, especially if we consider that they comprised most of the patients (68.4%) from our original cohort with 500 individuals. A feasible,

ongoing approach that is possible for patients submitted to ES with good cost-effectiveness is regular re-evaluation of ES data of nonpositive results with updated databases that include newly discovered disease-associated genes (Alfares *et al*., 2018). Another rising possibility is to study more complex interaction of rare variants in digenic or oligogenic predictors by studying combinations in gene pairs or networks, considering that some rare diseases may present digenic or even oligogenic inheritances (Renaux *et al*., 2019). Both approaches are more difficult and more limited using restricted data from panels compared to ES.

In conclusion, focused gene panels are restricted in covering select sets of genes implicated in specific diseases and they may miss an average of 64% of diagnoses compared to ES for cases of rare diseases without a clear clinical manifestation, though this rate varies widely depending on the clinical indication and laboratory. However, their role in genetic diagnosis remains important especially for well-known genetic diseases with established genetic locus heterogeneity because they still present some advantages, such as reduced cost, shorter turn-around time, and higher coverage of targeted genes.

Conflict of Interest

The following authors are employees (received salary and other bonuses) of Fleury Medicina e Saúde: Caio Robledo D'Angioli Costa Quaio; Sandro Felix Perazzio; Aurelio Pimenta Dutra; Christine Hsiaoyun Chung; Caroline Monaco Moreira; Gil Monteiro Novo Filho; Patricia Rossi Sacramento-Bobotis; Michele Groenner Penna; Rafaela Rogerio Floriano de Sousa; Vivian Pedigone Cintra; Juliana Emilia Prior Carnavalli; Rafael Alves da Silva; Monize Nakamoto Provisor Santos; Maria Carolina Pintao; Alexandre Ricardo dos Santos Fornari; Matheus Burger; Rodrigo Fernandes Ramalho; Elisa Napolitano e Ferreira; Otavio Jose Eulalio Pereira; Daniele Paixão; Caroline Olivati; Gustavo Marquezani Spolador; Miguel Mitne-Neto; Wagner Antonio da Rosa Baratela.

Author Contributions

CRDACQ; MJRO; SFP; APD; CHC; CAK: provided the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content. CMM; GMNF; PRSB; MGP; RRFS; VPC; JEPC; MCP; MMN: provided acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content. RAS; MNPS; ARSF; ALB; MB; RFR; ENF; OJEP; DP; CO; GMS; WARB: revised the article critically for important intellectual content. All authors read and approved the final version.

References

Alfares A, Aloraini T, Subaie LA, Alissa A, Qudsi AA, Alahmad A, Mutairi FA, Alswaid A, Alothaim A, Eyaid W *et al* (2018) Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. Genet Med 20:1328-1333.

Butler KM, da Silva C, Alexander JJ, Hegde M and Escayg A (2017) Diagnostic yield from 339 epilepsy patients screened on a clinical gene panel. Pediatr Neurol 77:61-66.

Dillon OJ, Lunke S, Stark Z, Yeung A, Thorne N, Melbourne Genomics Health Alliance, Gaff C, White SM and Tan TY (2018) Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. Eur J Hum Genet 26:644-651.

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.

Fung JLF, Yu MHC, Huang S, Chung CCY, Chan MCY, Pajusalu S, Mak CCY, Hui VCC, Tsang MHY, Yeung KS, *et al* (2020) A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. NPJ Genom Med 5:37.

Ghosh A, Schlecht H, Heptinstall LE, Bassett JK, Cartwright E, Bhaskar SS, Urquhart J, Broomfield A, Morris AA, Jameson E, *et al* (2017) Diagnosing childhood-onset inborn errors of metabolism by next-generation sequencing. Arch Dis Child 102:1019-1029.

Jang SS, Kim SY, Kim H, Hwang H, Chae JH, Kim KJ, Kim JI and Lim BC (2019) Diagnostic yield of epilepsy panel testing in patients with seizure onset within the first year of life. Front Neurol 10:988.

Klitzman R, Appelbaum PS, Chung W. Return of secondary genomic findings vs patient autonomy: implications for medical care. JAMA. 2013 Jul 24;310(4):369-70. doi: 10.1001/jama.2013.41709. PMID: 23917282; PMCID: PMC3800693.

LaDuca H, Farwell KD, Vuong H, Lu HM, Mu W, Shahmirzadi L, Tang S, Chen J, Bhide S and Chao EC (2017) Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. PLoS One 12:e0170843.

Lee KN, Uhlmann W, Hipp L and Quinonez SC (2020) The diagnosis of inborn errors of metabolism in previously undiagnosed adults referred for medical genetics evaluation. Mol Genet Metab Rep 25:100653.

Mercimek-Mahmutoglu S, Patel J, Cordeiro D, Hewson S, Callen D, Donner EJ, Hahn CD, Kannu P, Kobayashi J, Minassian BA *et al* (2015) Diagnostic yield of genetic testing in epileptic encephalopathy in childhood. Epilepsia 56:707-716.

Quaio CRDC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Penna MG, Perazzio SF, Dutra AP, da Silva RA, Santos MNP, de Arruda VYN *et al* (2020) Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. J Med Genet C Semin Med Genet 184:955-964.

Quaio CRDC, Chung CH, Perazzio SF, Dutra AP, Moreira CM, Novo Filho GM, Sacramento-Bobotis PR, Penna MG, de Souza RRF, Cintra VP *et al* (2021) Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. Am J Med Genet C Semin Med Genet. DOI: 10.1002/ajmg.c.31932.

Renaux A, Papadimitriou S, Versbraegen N, Nachtegael C, Boutry S, Nowé A, Smits G and Lenaerts T (2019) ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. Nucleic Acids Res 77:W93-W98.

Rim JH, Kim SH, Hwang IS, Kwon SS, Kim J, Kim HW, Cho MJ, Ko A, Youn SE, Kim J *et al* (2018) Efficient strategy for the molecular diagnosis of intractable early-onset epilepsy using targeted gene sequencing. BMC Med Genomics 11:6.

Teer JK and Mullikin JC (2010) Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet 19:R145-51.

Xue Y, Ankala A, Wilcox WR and Hegde MR (2015) Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med 17:444-451.

Yubero D, Brandi N, Ormazabal A, Garcia-Cazorla À, Pérez-Dueñas B, Campistol J, Ribes A, Palau F, Artuch R, Armstrong J and Working Group (2016) Targeted next generation sequencing in patients with inborn errors of metabolism. PLoS One 11:e0156359.

Supplementary material

The following online material is available for this article:

Table S1 – Overview of patients.

Table S2 – Neurodevelopmental disorders panels.
Table S3 – Seizure panels.
Table S4 – Syndromic/malformative panels.
Table S5 – Immune/hematological diseases panels.
Table S6 – Cardiovascular disease panels.
Table S7 – Metabolic abnormality panels.

Associate Editor: Roberto Giugliani

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.

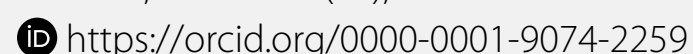
https://doi.org/10.1590/1516-3180.2022.0076.R1.21072022

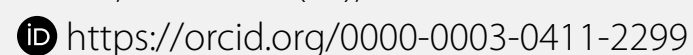
Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned

Caio Robledo D'Angioli Costa Quaio[I], Caroline Monaco Moreira[II], Christine Hsiaoyun Chung[III], Sandro Felix Perazzio[IV], Aurelio Pimenta Dutra[V], Chong Ae Kim[VI]

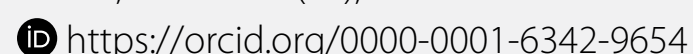
Instituto da Criança (ICr), Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, BR

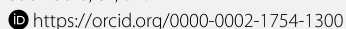
[I]MD. Researcher and PhD Candidate, Instituto da Criança (ICr), Hospital das Clinicas HCFMUSP, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, BR; Medical Geneticist, Fleury Medicina e Saúde, São Paulo (SP), Brazil; Medical Geneticist, Laboratório Clínico, Hospital Israelita Albert Einstein (HIAE), São Paulo (SP), Brazil.
https://orcid.org/0000-0003-2873-2820

[II]MSc. Molecular Biology Leader, Fleury Medicina e Saúde, São Paulo (SP), Brazil.
https://orcid.org/0000-0001-9074-2259

[III]MD. Medical Geneticist, Fleury Medicina e Saúde, São Paulo (SP), Brazil.
https://orcid.org/0000-0003-0411-2299

[IV]MD, PhD. Immunology Division, Fleury Medicina e Saúde, São Paulo (SP), Brazil; Division of Rheumatology, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; Central Laboratory, Hospital das Clinicas HCFMUSP, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, Brazil.
https://orcid.org/0000-0003-1203-8287

[V]MD, PhD. Neurology Division, Fleury Medicina e Saúde, São Paulo (SP), Brazil.
https://orcid.org/0000-0001-6342-9654

[VI]MD, PhD. Head of Genetics Unit and Associate Professor, Instituto da Criança (ICr), Hospital das Clinicas HCFMUSP, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, BR.
https://orcid.org/0000-0002-1754-1300

Dear Editor,

Rare diseases comprise a large and diverse group of an estimated 7,000 different conditions that collectively affect millions of people worldwide. We recently studied the genomic findings of 500 Brazilian patients with suspected rare diseases of genetic etiology who have undergone exome sequencing (ES) for diagnostic purposes.[1]

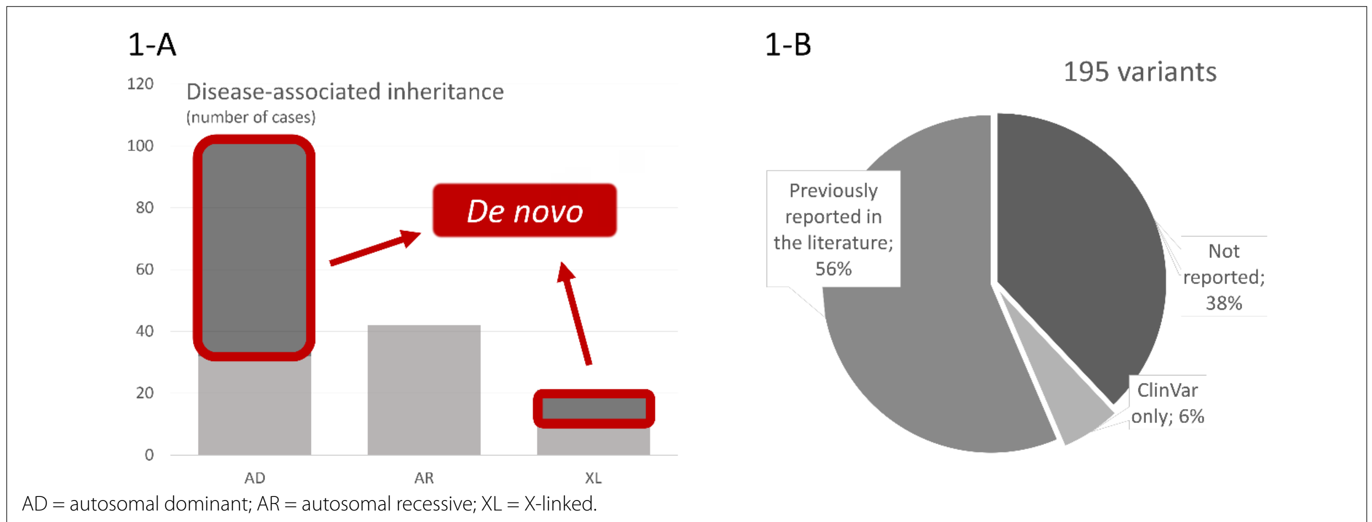
We observed an overall diagnostic yield of 31.6% in our cohort. Figure 1–A shows the inheritance patterns of the genetic diseases. These diagnoses were associated with 195 sequence variants, among which 38% were rare variants that have not been previously published in the literature (Figure 1–B). The diagnostic rate varied widely depending on age, and we observed higher diagnostic rates in prenatal samples (67%) and children younger than one year (44%) and lower rates for adults older than 50 years (13%). Undiagnosed patients still comprise the majority of patients in our cohort (68.4%) and remain a challenge in genomics. Even with the advances in genomic technology, for many patients with rare diseases, the diagnostic odyssey has not come to an end. Valuable techniques such as trio exome analysis (testing of samples from a proband and both parents) or genome sequencing may increase the genetic diagnosis of rare diseases.

We also found additional genetic alterations that may directly affect the morbidity and mortality of individuals. In 37 patients (7.4%), we found deleterious genetic variants associated with clinically actionable conditions, such as hereditary cancer, arrhythmia, metabolic diseases, and cardiomyopathies. These secondary findings were previously referred to as "incidental findings". (Figure 2).

Determining reportable secondary findings remains controversial and challenging.[2-4] Discussions on this subject are prevalent in North American, European, and some Asian countries but have yet to take place in Brazil and other Latin American countries. Indeed, there are no regulatory documents, legislation, or policies from scientific societies in Brazil regarding the protocols for reporting secondary findings in genomic studies. We urge our medical societies to adopt specific policies for reporting these conditions, and more importantly, consider the Brazilian frequencies of rare diseases. We believe that our study has made an important contribution to the knowledge of rare diseases of genetic etiology in Brazil, a country very underrepresented on this subject.

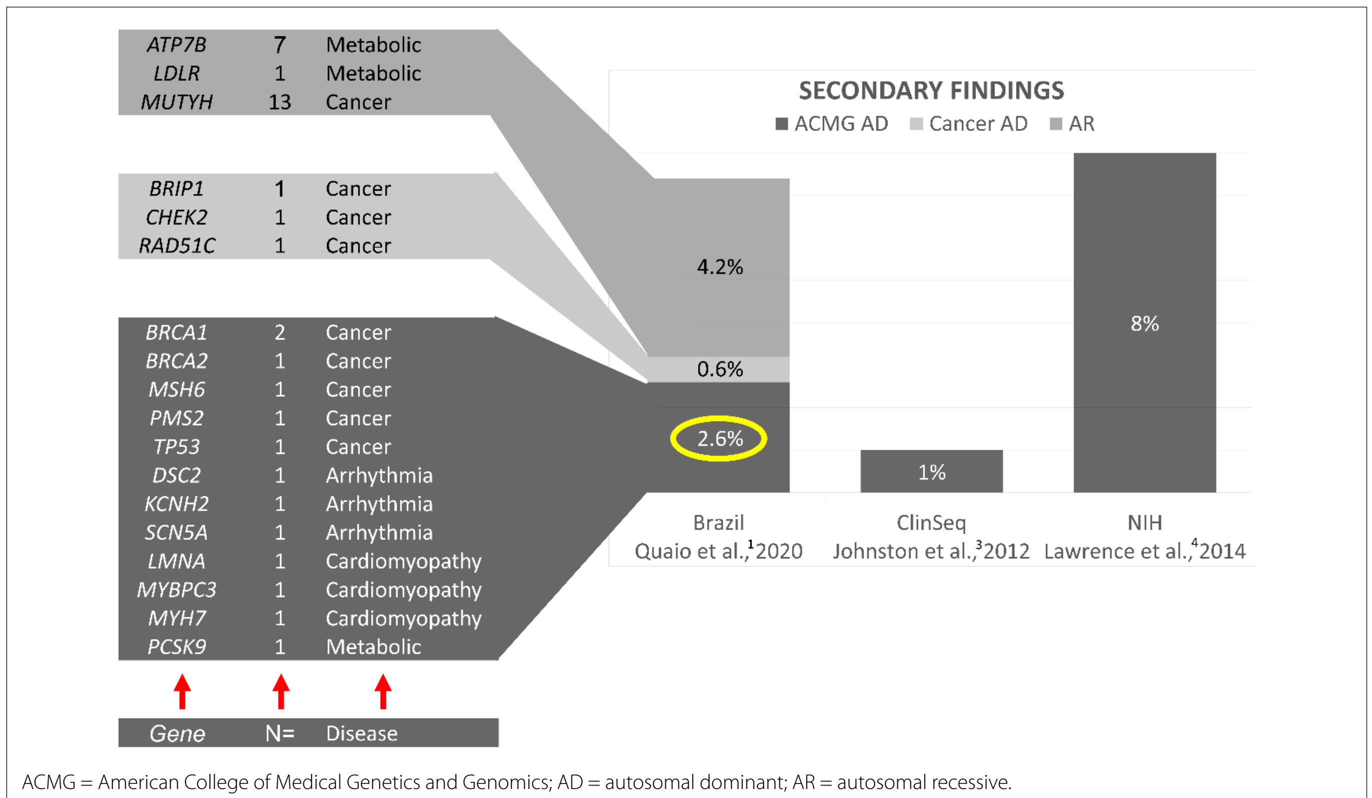
We also studied the carrier status for recessive diseases in 320 symptomatic patients in this cohort.[5,6] We found at least one pathogenic or likely pathogenic heterozygous variant associated with recessive diseases in the majority of individuals (71.9%). We believe that population studies of recessive diseases are important because recessive diseases are relatively frequent in aggregate, have a high clinical impact, early management can impact clinical outcomes, and some of them can be detected by neonatal screening (e.g., phenylketonuria).

In one of the largest cohorts of rare diseases in Latin America, we observed that ES was a powerful method for identifying the molecular basis of monogenic disorders, redirecting clinical care, and guiding health policies for rare diseases. We hope that our study will encourage others to better understand the clinical and social burden of rare genetic diseases in developing countries.



AD = autosomal dominant; AR = autosomal recessive; XL = X-linked.

Figure 1. Inheritance pattern and variant characteristics. **A** shows the distribution of inheritance patterns of the 164 diagnoses and highlights the proportion of *de novo* events; exome sequencing was able to unravel the diagnosis in 158 patients (six patients presented dual molecular diagnoses), which represents an overall diagnostic yield of 31.6%. The 164 diagnoses comprised 101 autosomal dominant, 42 autosomal recessive, and 21 X-linked conditions. **B** shows the distribution of variants according to the literature (HGMD Professional Database) and ClinVar database.



ACMG = American College of Medical Genetics and Genomics; AD = autosomal dominant; AR = autosomal recessive.

Figure 2. Secondary findings observed in our cohort of Brazilian patients compared with other studies in the literature. On the left, we observe the distribution of secondary findings: 37 patients (7.4%) presented a reportable secondary finding, among which 2.6% are those who harbored pathogenic or likely pathogenic variants associated with autosomal dominant diseases of obligatory report according to the American College of Medical Genetics and Genomics recommendations, 0.6% presented other autosomal dominant forms of hereditary cancer and 4.2% were carriers for variants associated with autosomal recessive diseases in American College of Medical Genetics and Genomics list of reportable secondary findings.

REFERENCES

1. Quaio CRDC, Moreira CM, Novo-Filho GM, et al. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. Am J Med Genet C Semin Med Genet. 2020;184(4):955-64. PMID: 33258288; https://doi.org/10.1002/ajmg.c.31860.
2. Kalia SS, Adelman K, Bale SJ, et al. 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. 2017;19(4):484. Genet Med. 2017;19(2):249-255. PMID: 27854360; https://doi.org/10.1038/gim.2016.190.
3. Johnston JJ, Rubinstein WS, Facio FM, et al. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. Am J Hum Genet. 2012;91(1):97-108. PMID: 22703879; https://doi.org/10.1016/j.ajhg.2012.05.021.
4. Lawrence L, Sincan M, Markello T, et al. The implications of familial incidental findings from exome sequencing: the NIH Undiagnosed Diseases Program experience. Genet Med. 2014;16(10):741-50. PMID: 24784157; https://doi.org/10.1038/gim.2014.29.
5. Quaio CRDC, Chung CH, Perazzio SF, et al. Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. Am J Med Genet C Semin Med Genet. 2021;187(3):364-72. PMID: 34269512; https://doi.org/10.1002/ajmg.c.31932.
6. Quaio CRAC, Moreira CM, Chung CH, et al. Frequency of carriers for rare metabolic diseases in a Brazilian cohort of 320 patients. Mol Biol Rep. 2022;49(5):3911-3918. PMID: 35229241; https://doi.org/10.1007/s11033-022-07241-3.

Authors' contributions: Quaio CRDC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, and Kim CA: provided the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, revised it critically for important intellectual content, and final approval of the version to be submitted

Ethics approval: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. This study was granted approval from the ethics committee of both institutions involved (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury# 3.372.339)

Consent to participate: Informed consent was obtained from all patients for inclusion in the study

Sources of funding: None

Conflict of interest: The following authors are employees (received salary and] other bonuses) of Fleury Medicina e Saúde: Caio Robledo D'Angioli Costa Quaio; Caroline Monaco Moreira; Christine Hsiaoyun Chung; Sandro Felix Perazzio; Aurelio Pimenta Dutra. Chong Ae Kim declares no competing interests

Date of first submission: January 29, 2022
Last received: July 7, 2022
Accepted: July 21, 2022

Address for correspondence:
Caio Robledo D'Angioli Costa Quaio
Instituto da Criança do Hospital das Clínicas da FMUSP – Unidade de Genética
Av. Dr. Enéas Carvalho de Aguiar, 647
Cerqueira César — São Paulo (SP) — Brasil
CEP 05403-900
Tel. (+55 11) 2661-8671 — Fax. (+55 11) 2661-8503
E-mail: cquaio@gmail.com

© 2022 by Associação Paulista de Medicina
This is an open access article distributed under the terms of the Creative Commons license.

Diagnóstico Molecular de Doenças Raras: Abrindo Portas para o Tratamento

Diagnóstico Molecular de Doenças Raras

Autores: Caio Robledo D'Angioli Costa Quaio [1]; Chong Ae Kim [2].

1- Médico Geneticista e Doutorando do Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo; Médico Geneticista do Laboratório Clínico do Hospital Israelita Albert Einstein, São Paulo, SP, Brasil. ORCID: 0000-0003-2873-2820

2- Professora Associada e Livre-Docente do Departamento de Pediatria da Faculdade de Medicina da Universidade de São Paulo; Coordenadora da Unidade de Genética do Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brasil. ORCID: 0000-0002-1754-1300

Endereço de correspondência:

Caio Robledo Quaio

Instituto da Criança do Hospital das Clínicas da FMUSP – Unidade de Genética

Av. Dr. Enéas Carvalho de Aguiar, 647. Cerqueira César. São Paulo - SP. Brasil. CEP: 05403-900

Telefone: +55(11)2661-8671; FAX: +55(11)2661-8503; E-mail: caio.quaio@usp.br

Diagnóstico Molecular de Doenças Raras: Abrindo Portas para o Tratamento

Resumo

As doenças raras englobam milhares de entidades clínicas e a elucidação etiológica pode trazer benefícios à saúde. Este estudo revisa os achados moleculares obtidos pelo sequenciamento do exoma e avalia potenciais impactos clínicos desta investigação molecular em coorte de 500 pacientes com doenças raras. Foi observada taxa diagnóstica de 31,6% com o uso desta metodologia e uma parcela significativa desta coorte (15,6%) teve potencial direto de redirecionamento dos cuidados clínicos (terapia alvo, rastreio para neoplasias, adoção de protocolo de seguimento padronizado, ajuste medicamentoso ou monitorização de complicações específicas). Foram também observados achados secundários com impacto clínico relevante em 7,4% da coorte. Observou-se, ainda, que 71,9% dos indivíduos eram portadores em heterozigose de variantes associadas a doenças autossômicas recessivas. O diagnóstico preciso das doenças raras e o conhecimento de suas fisiopatologias moleculares são o primeiro passo para o acolhimento desses indivíduos no sistema de saúde.

Palavras-chave

Doenças raras; Sequenciamento do exoma; Genômica; Sequenciamento de Nova Geração.

Diagnóstico molecular de enfermedades raras: abriendo puertas al tratamiento

Resumen

Las enfermedades raras abarcan miles de entidades clínicas y el diagnóstico puede aportar beneficios clínicos. Este estudio revisa los hallazgos moleculares y el impacto clínico observado en el estudio molecular de cohorte de enfermedades raras. Por lo tanto, se revisaron los hallazgos de la secuenciación del exoma de 500 pacientes. Se observó una tasa diagnóstica del 31,6% con el uso de esta metodología y una parte significativa de esta cohorte (15,6%) tenía potencial directo para redirigir la atención clínica (terapia dirigida, tamizaje de neoplasias, adopción de un protocolo de seguimiento estandarizado, ajuste de medicamentos o monitoreo de complicaciones específicas). También se observaron hallazgos secundarios con un impacto clínico relevante en el 7,4%. También se observó que 71,9% de los individuos eran portadores heterocigotos de variantes asociadas a enfermedades autosómicas recesivas. El diagnóstico preciso de las enfermedades raras y el conocimiento de su fisiopatología molecular son el primer paso para acoger a estas personas en el sistema de salud.

Descriptores

Enfermedades raras; Secuenciación del exoma; Genómico; Secuenciación de próxima generación.

Molecular Diagnosis of Rare Diseases: Opening Doors to Treatment

Abstract

This study aims to evaluate the clinical benefits of diagnosing rare diseases through molecular findings. The study reviews the results of exome sequencing in a cohort of 500 patients with rare diseases, with a diagnostic rate of 31.6%. Notably, 15.6% of the cohort had findings that directly impacted clinical care, such as targeted therapy, screening for neoplasms, adoption of a standardized follow-up protocol, drug adjustment, or monitoring of specific complications. Additionally, 7.4% of patients had secondary findings with clinical implications. The study also found that 71.9% of patients were heterozygous carriers of variants associated with autosomal recessive diseases. Accurate diagnosis and understanding of the molecular pathophysiology of rare diseases are crucial steps in providing comprehensive care for affected individuals. The findings of this study underscore the importance of utilizing molecular techniques in the diagnosis of rare diseases to improve clinical outcomes.

Keywords

Rare diseases; Exome sequencing; Genomics; Next Generation Sequencing.

Diagnóstico Molecular de Doenças Raras: Abrindo Portas para o Tratamento

Introdução

As doenças raras correspondem a grupo amplo e diverso de aproximadamente 7.000 entidades clínicas, muitas das quais apresentam etiologia genética [1,2]. Grande parte das doenças raras apresentam envolvimento clínico multissistêmico, fato que traz grande impacto aos indivíduos, suas famílias e sociedade [1-3]. Desta forma, o seguimento clínico de muitos indivíduos com doenças raras geralmente necessita da atenção de grandes centros terciários de assistência à saúde e os serviços públicos formam a espinha dorsal do acompanhamento desses pacientes.

A elucidação etiológica das doenças raras pode trazer benefícios no manejo clínico, prevenção de eventuais complicações de longo prazo, melhora do prognóstico e mais apropriado aconselhamento genético [3-5]. Entretanto, este processo diagnóstico pode ser desafiador e muitos desses pacientes enfrentam uma verdadeira “odisseia diagnóstica”. Antes do advento dos estudos genômicos na prática clínica, o diagnóstico das doenças mendelianas dependia, primeiramente, do reconhecimento de padrão de acometimento clínico [6].

O Sequenciamento do Exoma, que consiste em método de estudo das regiões codificantes do Genoma Humano em larga escala e é baseado na técnica de sequenciamento de nova geração (NGS), trouxe vantagens no processo de busca diagnóstica e potencialmente propicia elucidação de etiologia genética em até cerca de 30% dos casos em estudos internacionais [6,7]. Entretanto, essas técnicas podem também trazer à luz outras doenças genéticas ocultas nas famílias: estes são os achados secundários. Os dados brasileiros sólidos sobre os achados primários (associados à doença que motivou a investigação molecular) e achados secundários são muito escassos.

O objetivo deste manuscrito é revisar os principais achados moleculares observados, pelo nosso grupo, no estudo de doenças raras com a técnica de sequenciamento do exoma e avaliar os seus potenciais impactos clínicos.

Metodologia

O estudo investigativo das doenças raras foi conduzido de acordo com princípios éticos seguindo as orientações contidas na *Declaração de Helsinki* e foi aprovado pelas Comissões de Ética em Pesquisa de todas as instituições envolvidas (Plataforma Brasil; CAAE# 02617018.3.0000.5474; Fleury#3.372.339; Data: 05/06/2019). Foi pré-requisito obrigatório para realização do Sequenciamento do Exoma o preenchimento prévio de termo de consentimento.

Foi realizado estudo retrospectivo de todos os resultados de testes de Sequenciamento de Exoma já realizados pelo laboratório Fleury e que objetivavam a elucidação de doenças geneticamente determinadas entre os anos de 2014 e 2020 [3-5]. O total de casos analisados compreendeu amostras de 500 pacientes. Os dados clínicos foram coletados por questionário abrangente aplicado previamente à realização do teste. Também foram revisadas as informações eventualmente presentes nos relatórios médicos ou pedidos de exames disponibilizados. As manifestações clínicas foram categorizadas em: a) Transtornos do neurodesenvolvimento; b) Epilepsia; c) Outras alterações neurológicas; d) Quadros sindrômicos/malformativos; e) Distúrbio do crescimento; f) Alterações imunológicas ou hematológicas; g) Câncer/tumores; h) Doenças cardiovasculares; i) Doenças gastrointestinais ou hepáticas; j) Consanguinidade; k) Alterações metabólicas ou hormonais; l) Outras manifestações.

Métodos de biologia molecular, bioinformática e análise de variantes

A análise molecular consistiu na extração e fragmentação de DNA genômico a partir de amostra sanguínea, salivar ou pré-natal (biópsia de vilosidade coriônica ou líquido amniótico) seguida de identificação e captura com kit específico (*Agilent Clinical Research Exome* v1®), seguindo as normativas de boas práticas laboratoriais e recomendações dos fabricantes. O sequenciamento foi realizado utilizando-se a plataforma *illumina NextSeq*®.

Os dados gerados pelo sequenciamento foram alinhados ao genoma referência GRCh37.75/hg19 utilizando-se *Burrows–Wheeler Aligne*r (BWA; versão 0.7. 17-r1188). A seguir, foram manipulados por processos customizados internamente com *Genome Analysis ToolKit* (GATK, versão 3. 8-0-ge9d806836) para a identificação de variantes de sequência e *Variant Effect Predictor* (VEP, versão 88.14) para anotação. Todos os exames atingiram uma cobertura mínima de bases-alvos de 95% para 10 leituras independentes (10x).

Ao menos duas análises em paralelo foram realizadas para cada pacientes com o intuito de pré-selecionar variantes genômicas considerando-se frequência populacional menor que 1%, impacto funcional, relevância clínica do gene com a respectiva alteração, relatos nos bancos de dados (p.ex., ClinVar, HGMD) e literatura científica. As variantes pré-selecionadas eram, então, discutidas em reunião com três membros da equipe multidisciplinar para avaliar a relevância clínica e, finalmente, selecionar as variantes relevantes para etapa confirmatória com sequenciamento Sanger e estudo de segregação nos genitores (quando as amostras eram disponíveis).

As variantes foram interpretadas considerando-se o quadro clínico do paciente e protocolo de classificação de variantes do Colégio Americano de Genética Médica e Genômica [8]. Todas as

variantes patogênicas são confirmadas pelo método Sanger e investigadas nos genitores, se as amostras estiverem disponíveis.

Achados primários e achados secundários

Os Achados Primários foram definidos como variantes genômicas patogênicas ou provavelmente patogênicas observadas em genes associados a doenças com manifestações clínicas com sobreposição ao quadro apresentado pelo probando e padrão de herança compatível. Desta forma, os achados primários são as variantes que definem o diagnóstico molecular do quadro que originalmente suscitou a investigação genética.

Os Achados Secundários referem-se às variantes patogênicas ou provavelmente patogênicas que ocorrem em genes associados a doenças com conduta clínica padronizada por diretrizes internacionais e cujo diagnóstico precoce traz benefício relevante em termos de redução da morbimortalidade ou melhora da qualidade de vida. Desta forma, os Achados Secundários não apresentam relação direta com o quadro clínico que motivou a investigação do probando, mas podem impactar o manejo clínico e aconselhamento genético do mesmo e de sua família. O Colégio Americano de Genética Médica e Genômica (ACMG) apresenta recomendações sobre a lista de doenças e respectivos genes de relato preconizado, que foi neste estudo utilizada como norteador para relato dos Achados Secundários [9]. Entretanto e de modo excepcional, alterações em genes fora desta lista e que foram consideradas potencialmente impactantes para a saúde do paciente também foram relatadas.

Objetivando-se estimar de modo mais preciso o impacto clínico de um diagnóstico molecular e inspirando-se em estudo pregresso de Meng et al., (2017) [10], formou-se equipe multidisciplinar com diversos especialistas de distintas áreas do conhecimento médico para avaliar potenciais mudanças na condução do tratamento do paciente frente a um diagnóstico assertivo. Assim, experiência clínica dos especialistas, dados contidos no banco OMIM e busca na literatura científica foram utilizados para estimar o potencial impacto clínico dos Achados Primários nos seguintes quesitos: terapia alvo, rastreio de neoplasias, adoção de diretrizes (*guidelines*) internacionais, ajuste medicamentoso e monitorização de complicações específicas.

Estudo de portadores para doenças recessivas

Adicionalmente, foram também estudadas as variantes patogênicas ou provavelmente patogênicas em heterozigose em genes associadas a condições autossômicas recessivas em 320 pacientes desta coorte. Indivíduos com alteração em somente um dos alelos de genes associados

a doenças autossômicas recessivas são definidos portadores para a correspondente doença. Para este propósito, os genes que apresentaram alterações foram divididos em categorias baseadas no impacto clínico das doenças a que estão associados de acordo com nosologias, revisões e protocolos de especialistas.

Resultados

Este estudo revisou achados clínicos e moleculares de 500 pacientes com suspeita de doenças raras de origem genética e que foram investigados por Sequenciamento do Exoma. Os principais achados moleculares desta coorte foram publicados em série de seis artigos científicos recentes [3-5,11-13]. Abaixo, explanaremos sobre os principais achados moleculares previamente identificados e, ainda, sobre dados adicionais inéditos acerca da coorte.

Manifestações clínicas

Trata-se de uma coorte clinicamente heterogênea, mas com predomínio de manifestações neurológicas: 67,2% apresentaram alguma alteração neurológica, sendo que 47% com transtornos do neurodesenvolvimento, 21% com epilepsia e 50% com ocorrência de outras manifestações neurológicas. Ainda, 35% dos pacientes apresentaram quadros sindrômicos/malformativos, 19% doenças imunológicas/hematológicas, 15% alteração de crescimento, 11% alterações metabólicas, 10% doenças cardiovasculares, 7% doenças hepáticas/gastrointestinais, 1% presença de consanguinidade dos genitores e 3% outras manifestações.

Achados primários e taxa diagnóstica do sequenciamento do exoma

De modo geral, achados primários foram relatados em 158 pacientes, o que representa uma taxa diagnóstica de 31,6%. As Tabelas 1 e 2 detalham as taxas de diagnóstico por faixa etária e por manifestação clínica, respectivamente. No total, 101 doenças autossômicas dominantes (AD) foram observadas em 98 indivíduos; 42 doenças autossômicas recessivas (AR) foram relatadas em 42 indivíduos e 21 doenças de herança ligada ao cromossomo X (LX) foram observadas em 21 indivíduos. A presença concomitante de duas doenças genéticas ocorreu em seis pacientes. Nestes 158 pacientes com achados primários, foram identificadas 195 variantes. Não foram encontrados relatos na literatura para 43,6% (n = 85) das variantes e em 37,9% (n = 74) não havia menção no banco de dados público ClinVar.

Taxa diagnóstica por faixa etária				
Faixa etária	N=	Gênero	Idade média	Taxa
Pré-natal	**6**	**n.d.**	**n.d.**	67%
<1 ano	55	M=58,2%	5m	**44%**
1-5 anos	161	M=61,5%	2a9m	**32%**
5-10 anos	78	M=66,7%	7a	**32%**
10-18 anos	63	M=54%	13a6m	**30%**
18-25 anos	26	F=57,7%	21a2m	**27%**
25-50 anos	66	F=59%	36a10m	**32%**
>50 anos	45	F=60%	61a6m	**13%**
Todas idades	500	M=55,3%	15a5m	**32%**

Tabela 1: Taxa Diagnóstica Por Faixa Etária. As colunas informam, respectivamente, a divisão de faixas etárias, o número (N) de pacientes por respectiva faixa etária, o gênero predominante para cada faixa etária (masculino [M] ou feminino [F]) com a correspondente frequência, a idade média por faixa etária (ano [a] e mês [m]) e, finalmente, a taxa diagnóstica obtida pelo sequenciamento do exoma para cada faixa etária. A última linha congrega todos os dados da coorte de 500 indivíduos com doenças raras.

Taxa Diagnóstica por Manifestação Clínica	
Indicações clínicas	Taxa
Consanguinidade	50%
Doença Gastrointestinal/Hepática	44%
Sindrômico/Malformativo	41%
Transtornos do neurodesenvolvimento	35%
Outras anomalias neurológicas	35%
Alteração de crescimento	35%
Epilepsia	34%
Alteração Metabólica/hormonal	33%
Câncer/tumores	32%
Doença Cardiovascular	31%
Doenças imunológicas/hematológicas	19%
Outras	13%

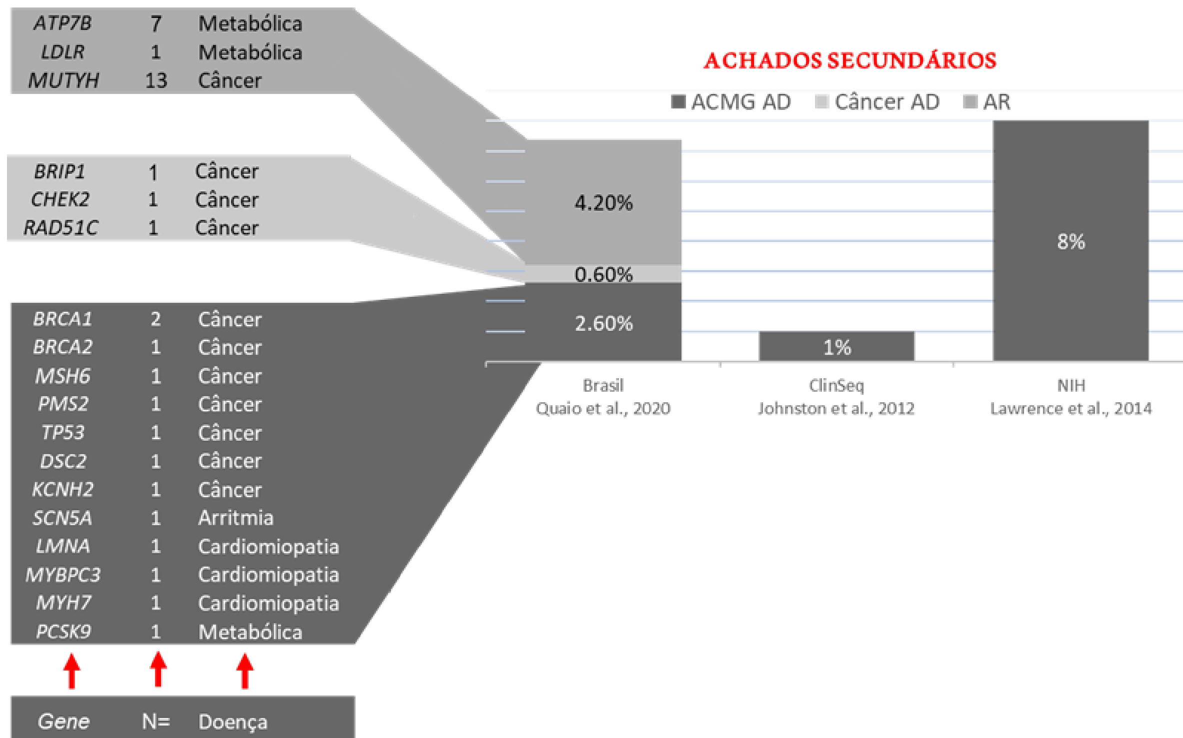
Tabela 2: Taxa Diagnóstica Por Manifestação Clínica. A segunda coluna informa a taxa diagnóstica, em ordem decrescente, para cada grupo de manifestação clínica (primeira coluna).

As doenças mais frequentemente observadas neste estudo e seus respectivos genes são: 1) síndrome de Rett (*MECP2*, OMIM:300005): sete pacientes; 2) doenças associadas ao gene *POLG* (*POLG*, OMIM: 174763): quatro pacientes; 3) doença de Niemann–Pick tipo C (*NPC1*, OMIM: 607623): três pacientes; 4) síndrome KBG (*ANKRD11*, OMIM: 611192): três pacientes; 5) doenças associadas ao gene *DEAF1* (*DEAF1*, OMIM: 602635): três pacientes; 6) síndrome de Cowden (*PTEN*, OMIM: 158350): três pacientes; 7) síndrome de Noonan (*PTPN11*, OMIM: 176876): três pacientes.

Foi estimado que 78 pacientes (15,6% de toda a coorte, ou 49,4% dos pacientes com Achados Primários) tiveram potencial benefício clínico direto com a elucidação etiológica. Nós observamos indicação de terapia alvo em 3,2% da coorte (n = 16), rastreio específico para doenças com aumento de risco de neoplasias em 1,6% (n = 8), adoção de protocolo de seguimento padronizado em 0,8% (n = 4), ajuste medicamentoso e, 4,4% (n = 22) e monitorização de complicações específicas das doenças para 5,6% (n = 28).

Achados secundários

Os achados secundários foram observados em 7,4% (n = 37) dos indivíduos. Detalhes moleculares bem como estimativas de impacto clínico e na expectativa de vida foram publicados [3]. Os achados secundários englobaram 23 variantes em heterozigose em 18 genes distintos. Do total de pacientes, 2,6% (n = 13) apresentaram variantes em heterozigose associadas a doenças AD (arritmias hereditárias, cardiomiopatias hereditárias, câncer hereditário e doenças metabólicas). Outros 0,6% (n = 3) apresentaram variantes em genes associados a outras formas de câncer hereditário fora da lista do ACMG (genes *CHEK2*, *RAD51C* e *BRIP1*) e 4,2% (n = 21) eram portadores de uma única variante monoalélica em genes associados a doenças AR. O Quadro 1, abaixo, sumariza os achados secundários identificados no presente estudo e compara as frequências com outros dois estudos internacionais.



Quadro 1: Achados Secundários. O quadro sumariza as doenças clínicas com conduta padronizada que foram observadas na coorte com 500 pacientes com doenças raras que realizaram Sequenciamento do Exoma com propósito diagnóstico (primeira coluna, Brasil), juntamente com a frequência de achados secundários em outros dois grandes estudos internacionais (segunda e terceira colunas). Os achados encontrados no presente estudo estão pormenorizados no diagrama da esquerda: ACMG AD representa o grupo de doenças de herança autossômica dominante de relato obrigatório de acordo com as recomendações do ACMG; Câncer AD representa genes associados a outras formas de câncer hereditário fora da lista do ACMG e que foram julgados clinicamente relevantes para vigilância tumoral e aconselhamento genético e AR representa o grupo portador de variantes monoalélicas em genes associados a doenças autossômicas recessivas. Adaptado de Quaio et al., 2022 [11].

Estudo de portadores de doenças autossômicas recessivas

Variantes monoalélicas (i.e., que ocorrem em heterozigose em um único alelo) associadas a doenças recessivas raras foram identificadas em 71,9% dos 320 pacientes investigados. Com base nesses dados, a frequência populacional de doenças autossômicas recessivas raras estimada através da equação de equilíbrio de Hardy-Weinberg é de 26,39/10.000 indivíduos (ou ~0.26%). Detalhes sobre estes achados bem como o raciocínio utilizado para a extrapolação populacional dos resultados foram recentemente publicados na literatura científica [4,12].

Os genes que mais frequentemente apresentaram alterações monoalélicas, suas correspondentes frequências de heterozigotos e as doenças associadas são: 1) *GJB2* (perda auditiva não sindrômica): 4,4%; 2) *MUTYH* (polipose intestinal): 4,1%; 3) *ABCA4* (retinite pigmentosa): 2,8%; 4) *RBM8A* (síndrome TAR): 2,8%; 5) *ATP7B* (doença de Wilson): 2,2%; 6) *GBA* (doença de Gaucher): 1,9%; 7) *PAH* (fenilcetonúria): 1,9%; 8) *RNASEH2B* (síndrome de Aicardi-Goutieres): 1,9%; 9) *SPG7* (paraplegia espástica): 1,6%. A Tabela 3 mostra número e a frequência global de heterozigotos por grupo de doenças.

Grupo de doenças	**Heterozigotos**	
	Número de heterozigotos	**Frequência de heterozigotos (%)**
Doenças metabólicas	156	48,80%
Epilepsia	83	25,90%
Deficiência intelectual	58	18,10%
Outras	66	20,60%
Displasias esqueléticas	50	15,60%
Doenças imunológicas	35	10,90%
Perda auditiva	29	9,10%
Retinite pigmentosa	24	7,50%
Ciliopatias	24	7,50%
Câncer	14	4,40%
Doenças musculares	10	3,10%
Total	228	71,30%

Tabela 3: Frequência e Número de Heterozigotos por Grupo de Doenças. As segunda e terceira colunas informam, respectivamente, o número e frequência de heterozigotos por Grupo de Doenças (primeira coluna). O total de indivíduos avaliado para este quesito foi 320 (100%). Adaptado de Quaio et al., 2021 [4].

Discussão

Este estudo, pioneiro no Brasil, avaliou a utilização da técnica de sequenciamento do exoma em coorte de indivíduos com doenças raras. Trata-se de coorte heterogênea, mas com predomínio de manifestações neurológicas e quadros clínicos complexos. A avaliação molecular desta coorte propiciou a elucidação etiológica em 31,6% dos indivíduos e uma parcela significativa desta coorte (15,6%) teve potencial direto pós-teste de redirecionamento dos cuidados clínicos, quer seja com

terapia alvo, rastreio específico para doenças com aumento de risco de neoplasias, adoção de protocolo de seguimento padronizado, ajuste medicamentoso ou monitorização de complicações específicas das doenças. Este fato sugere que a investigação molecular das doenças raras pode trazer benefício clínico aos indivíduos.

As doenças raras de origem genética são complexas e grande parte desses pacientes dependem dos serviços clínicos prestados pelos centros terciários públicos de excelência. A Unidade de Genética do Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo tem papel chave no acolhimento das crianças com doenças raras do Estado de São Paulo e também de outros estados há 45 anos. Nosso serviço, em seus primórdios de 1978, prestava assistência a cerca de 300 pacientes anualmente e, atualmente, houve grande ampliação para cerca de 1.800 atendimentos anuais. Este grande incremento foi propiciado, em parte, pela implantação do programa de Residência Médica em Genética no ano de 2.009, com o intuito de propagar o conhecimento e a expertise técnica no manejo dos pacientes com doenças raras. Atualmente, a unidade conta com três médicos assistentes e 15 médicos residentes. O número de atendimentos e a prestação de serviço de excelência poderia ser ainda maior, se houvesse incremento da estrutura física disponível. Ademais, as pesquisas conduzidas ao longo destas décadas auxiliam no entendimento da situação das doenças raras no Brasil, inclusive na descrição de novas doenças e novos mecanismos fisiopatológicos.

Esta Unidade de Genética presenciou a revolução diagnóstica ocorrida nos últimos anos. Primeiramente, o diagnóstico de uma síndrome rara dependia primariamente do reconhecimento clínico do conjunto de manifestações do paciente. As técnicas baseadas em sequenciamento de nova geração, como o sequenciamento do exoma, aumentaram de maneira drástica a taxa de elucidação etiológica e, assim, facilitaram o acesso a novas terapêuticas a muitos indivíduos. Entretanto, o acesso a esta poderosa ferramenta diagnóstica é restrito no âmbito do sistema público de saúde. Neste artigo, apresentamos resultados promissores, mas de uma coorte que foi investigada em laboratório privado. O próximo passo é estender esta mesma abordagem para ampliarmos o conhecimento acerca das doenças raras na esfera do sistema público de saúde.

A taxa diagnóstica obtida pelo sequenciamento do exoma de nossa coorte de 500 pacientes é congruente com outros estudos. Taxas diagnósticas entre 22% e 30% já foram relatadas em outros estudos que avaliaram a utilização desta técnica molecular na investigação de doenças raras e estas taxas se mostraram muito superiores às taxas de outras metodologias diagnósticas em genéticas, como cariótipo, estudo por microarranjos (*Microarray*), entre outras técnicas [6,7,14-16]. As mais elevadas taxas diagnósticas foram encontradas para amostras pré-natais, crianças com menos de um ano de idade, indivíduos com genitores consanguíneos e quadros sindrômicos/malformativos (Tabelas 1 e 2). Estes mesmos grupos também apresentaram as taxas mais elevadas em outros estudos [10,15,17].

Nosso estudo identificou uma frequência relevante (7,4%) de achados secundários, que são alterações deletérias em genes associados a doenças de importante repercussão clínica, como câncer hereditário, arritmias, cardiomiopatias e doenças metabólicas (Quadro 1) [3]. Existem desafios na determinação de quais desses achados moleculares com condutas clínicas padronizadas devem ser relatados de modo obrigatório. Em geral, os laboratórios de análises seguem as diretrizes norte-americanas ou europeias, uma vez que não existem regulações ou legislação nacional que abordem o tema [9]. É imperativo que as sociedades médicas brasileiras liderem a discussão nacional sobre os achados moleculares secundários de relato obrigatório no contexto brasileiro.

Este trabalho também propiciou o pioneiro estudo da frequência de portadores em heterozigose de variantes associadas a doenças autossômicas recessivas e, ainda, possibilitou estimar a frequência combinada de doenças recessivas raras na população brasileira. Considerando-se nossos dados e a equação de Hardy-Weinberg, nós estimamos a frequência de doenças recessivas em nossa população ser ao redor de 26,39/10.000 indivíduos, ou 0,26% [4,12]. Surpreendentemente, estimativas muito próximas (~30/10.000 indivíduos) foram encontradas por diferentes centros de estudo do mundo, utilizando-se, muitas vezes, metodologias distintas [18-21]. Esta congruência com a literatura valida, ao menos em parte, os nossos achados e nossa metodologia. Os dados populacionais sobre as doenças recessivas são de suma importância, uma vez que muitas delas apresenta tratamento específico e são de investigação obrigatória na triagem neonatal (p.ex., fenilcetonúria).

Em suma, este estudo permitiu vislumbrar detalhes importantes sobre os achados moleculares em pacientes com doenças raras em uma população muito sub-representada na literatura científica. São muito poucos os trabalhos brasileiros e de outros países da américa latina que abordam o tema. O diagnóstico preciso das doenças raras é o primeiro passo para o acolhimento desses indivíduos no sistema de saúde e a adoção de terapêuticas mais precisas. Nosso desafio atual é aprofundar o conhecimento sobre variantes genéticas raras no Brasil, dimensionar melhor seus impactos na saúde e qualidade de vida e propor estratégias para disseminar este conhecimento entre profissionais da saúde.

Referências

1- Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013 Oct;14(10):681-91.

2- McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet. 2007 Apr;80(4):588-604.

3- Quaio CRDC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Groenner Penna M, Perazzio SF, et al. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. Am J Med Genet C Semin Med Genet. 2020 Dec;184(4):955-64.

4- Quaio CRDC, Chung CH, Perazzio SF, Dutra AP, Moreira CM, Filho GMN, et al. Frequency of carriers for rare recessive Mendelian diseases in a Brazilian cohort of 320 patients. Am J Med Genet C Semin Med Genet. 2021 Sep;187(3):364-72.

5- Quaio CRD'C, Ceroni JRM, Cervato MC, Thurow HS, Moreira CM, Trindade ACG, et al. Parental segregation study reveals rare benign and likely benign variants in a Brazilian cohort of rare diseases. Sci Rep. 2022 May 11;12(1):7764.

6- Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015 Jul;47(7):717-26.

7- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013 Oct 17;369(16):1502-11.

8- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24.

9- Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013 Jul;15(7):565-74.

10- Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Vetrini F, et al. Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. JAMA Pediatr. 2017 Dec 4;171(12):e173438.

11- Quaio CRDC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, Kim CA. Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned. Sao Paulo Med J. 2022;140(5):734-6.

12- Quaio CRAC, Moreira CM, Chung CH, Perazzio SF, Dutra AP, Kim CA. Frequency of carriers for rare metabolic diseases in a Brazilian cohort of 320 patients. Mol Biol Rep. 2022 May;49(5):3911-8.

13- Quaio CRDC, Obando MJR, Perazzio SF, Dutra AP, Chung CH, Moreira CM, et al. Exome sequencing and targeted gene panels: a simulated comparison of diagnostic yield using data from 158 patients with rare diseases. Genet Mol Biol. 2021;44(4):20210061.

14- Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am J Hum Genet. 2014 Jun 5;94(6):809-17.

15- Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med. 2015 Jul;17(7):578-86.

16- Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014 Nov 12;312(18):1880-7.

17- Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Köster J, Marais A, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet. 2017 Feb;25(2):176-82.

18- Henneman L, Borry P, Chokoshvili D, Cornel MC, van El CG, Forzano F, et al. Responsible implementation of expanded carrier screening. Eur J Hum Genet. 2016 Jun;24(6):e1-e12.

19- Sankaranarayanan K. Ionizing radiation and genetic risks IX. Estimates of the frequencies of mendelian diseases and spontaneous mutation rates in human populations: a 1998 perspective. Mutat Res. 1998 Sep;411(2):129-78.

20- Ropers HH. On the future of genetic risk assessment. J Community Genet. 2012 Jul;3(3):229-36.

21- Quaio CRAC, Coelho AVC, Moura LMS, Guedes RLM, Chen K, Ceroni JRM, et al. Genomic study of nonsyndromic hearing loss in unaffected individuals: Frequency of pathogenic and likely pathogenic variants in a Brazilian cohort of 2,097 genomes. Front Genet. 2022;13:921324.