

RESEARCH

Open Access



# Lessons learned from additional research analyses of unsolved clinical exome cases

Mohammad K. Eldomery<sup>1,18†</sup>, Zeynep Coban-Akdemir<sup>1†</sup>, Tamar Harel<sup>1†</sup>, Jill A. Rosenfeld<sup>1</sup>, Tomasz Gambin<sup>1,2</sup>, Asbjørg Stray-Pedersen<sup>3</sup>, Sébastien Küry<sup>4</sup>, Sandra Mercier<sup>4,5</sup>, Davor Lessel<sup>6</sup>, Jonas Denecke<sup>7</sup>, Wojciech Wiszniewski<sup>1,8</sup>, Samantha Penney<sup>1</sup>, Pengfei Liu<sup>1,9</sup>, Weimin Bi<sup>1,9</sup>, Seema R. Lalani<sup>1,8</sup>, Christian P. Schaaf<sup>1,8,10</sup>, Michael F. Wangler<sup>1,8</sup>, Carlos A. Bacino<sup>1,8</sup>, Richard Alan Lewis<sup>1,10</sup>, Lorraine Potocki<sup>1,8</sup>, Brett H. Graham<sup>1,8</sup>, John W. Belmont<sup>1,8</sup>, Fernando Scaglia<sup>1,8</sup>, Jordan S. Orange<sup>11,12</sup>, Shalini N. Jhangiani<sup>13</sup>, Theodore Chiang<sup>13</sup>, Harsha Doddapaneni<sup>13</sup>, Jianhong Hu<sup>13</sup>, Donna M. Muzny<sup>13</sup>, Fan Xia<sup>1,9</sup>, Arthur L. Beaudet<sup>1,9</sup>, Eric Boerwinkle<sup>13,14</sup>, Christine M. Eng<sup>1,9</sup>, Sharon E. Plon<sup>1,8,11,15</sup>, V. Reid Sutton<sup>1,8</sup>, Richard A. Gibbs<sup>1,13,16</sup>, Jennifer E. Posey<sup>1</sup>, Yaping Yang<sup>1,9</sup> and James R. Lupski<sup>1,8,11,13,17\*</sup>

## Abstract

**Background:** Given the rarity of most single-gene Mendelian disorders, concerted efforts of data exchange between clinical and scientific communities are critical to optimize molecular diagnosis and novel disease gene discovery.

**Methods:** We designed and implemented protocols for the study of cases for which a plausible molecular diagnosis was not achieved in a clinical genomics diagnostic laboratory (i.e. unsolved clinical exomes). Such cases were recruited to a research laboratory for further analyses, in order to potentially: (1) accelerate novel disease gene discovery; (2) increase the molecular diagnostic yield of whole exome sequencing (WES); and (3) gain insight into the genetic mechanisms of disease. Pilot project data included 74 families, consisting mostly of parent–offspring trios. Analyses performed on a research basis employed both WES from additional family members and complementary bioinformatics approaches and protocols.

**Results:** Analysis of all possible modes of Mendelian inheritance, focusing on both single nucleotide variants (SNV) and copy number variant (CNV) alleles, yielded a likely contributory variant in 36% (27/74) of cases. If one includes candidate genes with variants identified within a single family, a potential contributory variant was identified in a total of ~51% (38/74) of cases enrolled in this pilot study. The molecular diagnosis was achieved in 30/63 trios (47.6%). Besides this, the analysis workflow yielded evidence for pathogenic variants in disease-associated genes in 4/6 singleton cases (66.6%), 1/1 multiplex family involving three affected siblings, and 3/4 (75%) quartet families. Both the analytical pipeline and the collaborative efforts between the diagnostic and research laboratories provided insights that allowed recent disease gene discoveries (*PURA*, *TANGO2*, *EMC1*, *GNB5*, *ATAD3A*, and *MIPEP*) and increased the number of novel genes, defined in this study as genes identified in more than one family (*DHX30* and *EBF3*).

(Continued on next page)

\* Correspondence: jlupski@bcm.edu

†Equal contributors

<sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

<sup>8</sup>Texas Children's Hospital, Houston, TX 77030, USA

Full list of author information is available at the end of the article



(Continued from previous page)

**Conclusion:** An efficient genomics pipeline in which clinical sequencing in a diagnostic laboratory is followed by the detailed reanalysis of unsolved cases in a research environment, supplemented with WES data from additional family members, and subject to adjuvant bioinformatics analyses including relaxed variant filtering parameters in informatics pipelines, can enhance the molecular diagnostic yield and provide mechanistic insights into Mendelian disorders. Implementing these approaches requires collaborative clinical molecular diagnostic and research efforts.

## Background

Applications to clinical practice of whole exome sequencing (WES) and whole genome sequencing (WGS) technologies and the computational interpretation of rare variants in genome data have been revolutionary, allowing conclusions to diagnostic odysseys and enabling molecular diagnoses for thousands of patients [1–7]. Moreover, such genome-wide assays have enabled insights into multi-locus contributions to disease [8]. Recent reports document an initial ~25–30% rate of molecular diagnosis in known disease genes for patients referred for exome sequencing and interpretation [3, 5, 9–12]. The remaining undiagnosed individuals may represent: (1) limitations in concluding a molecular diagnosis using the current experimental and analytical methods of clinical genomics practice; or (2) our limited understanding of the genetics of human disease. Collaboration between the clinical, clinical molecular diagnostic, and research communities may optimize discovery of disease genes, considering the rarity of specific genetic disorders [13–17].

We performed a pilot study for systematic transfer of molecularly “unsolved” exomes from the clinical environment to a research setting, in order to potentially fuel human genetic disease gene discovery. The WES data from 74 probands for whom clinical singleton WES did not reveal a secure molecular diagnosis were augmented with WES from additional family members, where available. Additional bioinformatics filters, database resources, and interpretive analyses were implemented, leveraging systematic studies emerging from the research laboratory. A likely disease contributory gene and potential molecular diagnosis (i.e. known disease gene or novel gene identified in more than one family) was identified in 36% of the probands, and a candidate gene finding (i.e. identified in a single family) was identified in 15% of patients. This experience and resulting findings offer the opportunity to systematically compare different but complementary approaches to optimize molecular diagnostic yield. Several novel gene discoveries (*PURA*, *TANGO2*, *EMC1*, *GNB5*, *ATAD3A*, *MIPEP*) [18–23] were facilitated by this collaborative and systematic clinical/research laboratory approach, and additional novel disease genes were found in multiple families (*DHX30*,

*EBF3*), together highlighting different genetic contributions to pathogenicity [24–26].

## Methods

### Recruitment of non-diagnostic clinical exome cases into research

Probands, whose DNA had been analyzed at Baylor Genetics (BG) laboratory for clinical diagnostic WES, and for whom a molecular diagnosis (defined as a pathogenic or likely pathogenic variant according to American College of Medical Genetics and Genomics [ACMG] guidelines) was not achieved at the time of initial reporting, were classified as “unsolved” and enrolled into the study [3, 27]. For this pilot study, a total of 74 unsolved clinical WES cases, analyzed by the clinical genomics laboratory between April 2012 and April 2014, were enrolled in research between July 2013 and March 2015. Enrollment proceeded in serial order with no specific inclusion or exclusion criteria, other than the inability to achieve a molecular diagnosis in the clinical genomics laboratory and parents consenting to research analyses. Depending on the clinical situation and individual availability, additional family members (e.g. parents or affected siblings) were also enrolled.

Our pilot study consisted of 74 cases including 63 trios, four quartets, one multiplex family involving three affected siblings, and six singleton cases for which parental samples were unavailable. Prior diagnostic work-up was variable from case to case based on the referring physician’s differential diagnosis (e.g. single gene testing, enzyme assays, array comparative-genome hybridization) and included a proband-only WES in all cases. Mitochondrial DNA sequencing was performed for all cases undergoing clinical WES through December 2014. Detailed clinical phenotype data were collected and entered into PhenoDB [14, 28], after contacting the families/patients to obtain informed consent, and did not influence the choice of cases. However, we retrospectively analyzed the phenotypic features of this cohort and found that developmental delay/intellectual disability (DD/ID) was the most prevalent phenotype in our pilot study, consistent with the nature of cases referred for clinical WES [3, 5]. Our phenotypic analysis revealed 59 cases with syndromic DD/ID and one with non-

syndromic DD/ID. Additionally, 14 other phenotypes encountered in the 74 cases of the pilot study were metabolic, gastrointestinal, and mitochondrial abnormalities (Additional file 1: Table S1).

### Whole exome sequencing and annotation

Exome capture was performed with Nimblegen reagents using the Baylor College of Medicine (BCM) Human Genome Sequencing Center (HGSC) custom-designed capture reagent VCRome 2.1 for both clinical and research laboratory exomes. This capture reagent contains more than 196K targets and 42 Mbp of genomic regions and includes predicted coding exons from Vega, CCDS, and RefSeq. Samples are multiplexed (six-plex format) for both capture and sequencing and full-length blocking oligos were employed for hybridization to enhance on-target specificity [3, 5, 29]. Clinical WES targets the coding exons of ~20,000 genes with 130X average depth of coverage and greater than 95% of the targeted bases having >20 reads [3, 5, 29, 30]. Research WES (for additional family members) had an average depth of coverage of 95X, with >92% of the targeted bases having >20 reads. The raw sequence data were post-processed using the Mercury pipeline [31]. First, the raw sequencing data (bcl files) were converted to fastq files using Casava. Then, the Burrows-Wheeler Alignment (BWA) tool was utilized to map short reads to the human genome reference sequence (GRCh37). Finally, the recalibration and variant calling were performed using GATK [30] and the Atlas2 suite, respectively [32]. The Mercury pipeline is available in the cloud via DNANexus (<http://blog.dnanexus.com/2013-10-22-run-mercury-variant-calling-pipeline/>). For research cases, exome variant analyses were then independently performed in the Baylor Hopkins Center for Mendelian Genomics (BHCMG) [15] under a Research Protocol approved by the Institutional Review Board (IRB) for Human Subjects Research at BCM; this protocol enables bi-directional transfer of samples and data between the clinical and research laboratories.

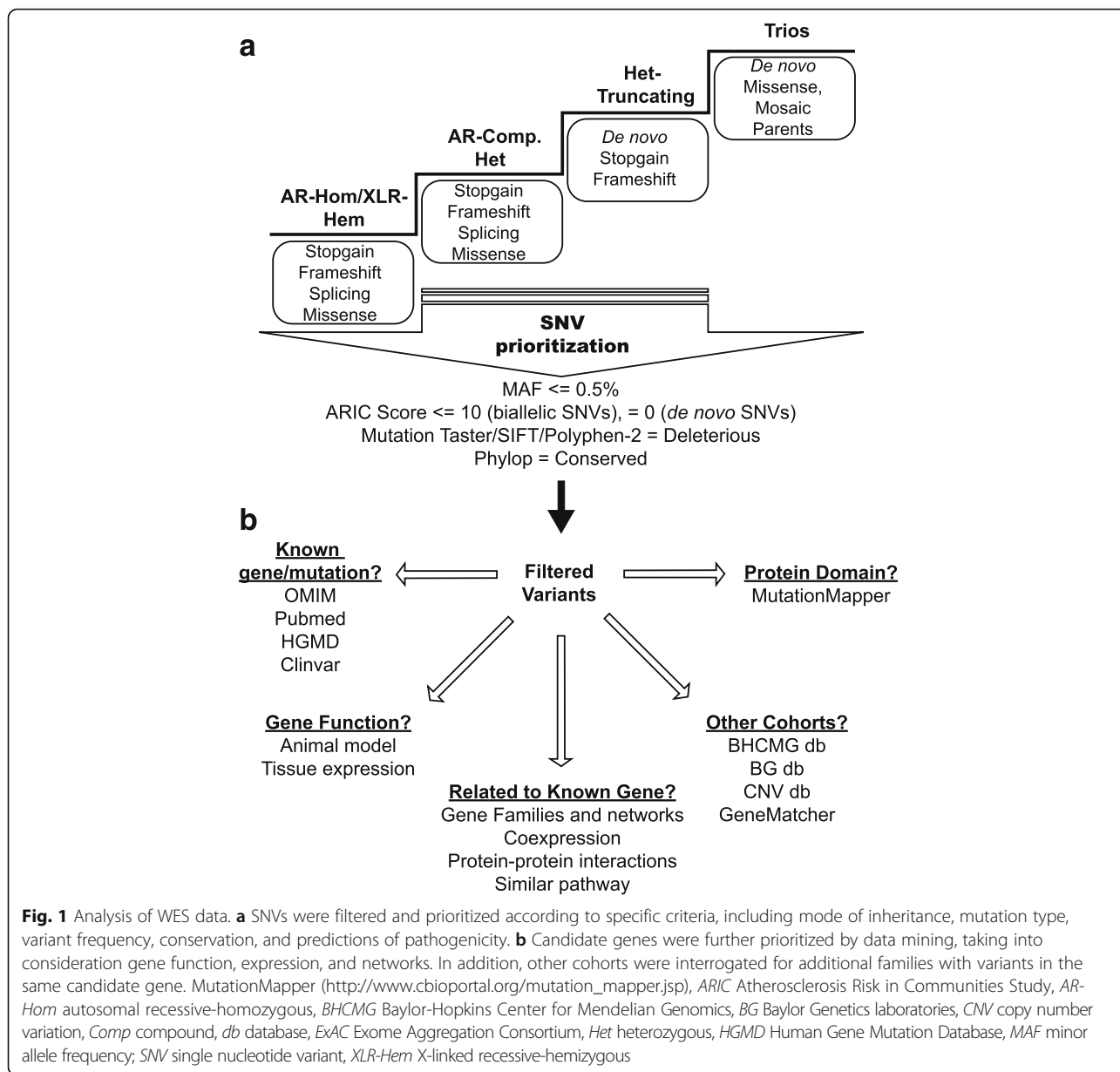
### Identification of de novo single nucleotide variants (SNVs) in BCHMG

De novo variants were identified by an in-house developed software called DNM (de novo mutation)-Finder (<https://github.com/BCM-Lupskilab/DNM-Finder>), available upon request. Parental variants were subtracted in silico from the proband's variants in vcf files, while incorporating read number information extracted from BAM files. Filtering was then implemented using the following criteria: (1) an alternative variant read count greater than 5 in the proband; (2) ratio of alternative variant read count to reference variant read count greater than 30% in the proband; (3) reference variant read count greater than 10 in both parents; and (4) ratio

of alternative variant read count to reference variant read count less than 5% in both parents.

### SNV prioritization and filtering workflow

We designed a stepwise analysis workflow. This workflow consists of four major steps of analysis that examined: (1) recessive homozygous predicted loss-of-function (LOF) variants (stop-gain, frameshift indels, and splice site variants) and/or missense variants; (2) compound heterozygous LOF and/or missense variants; (3) heterozygous LOF variants, which facilitated the detection of potential truncating de novo mutations in proband-only WES cases; and (4) de novo variants and potential parental mosaic variants using trio-WES (Fig. 1a). Variants were prioritized for further study based on their minor allele frequencies (MAF; <0.5%) and the output from several prediction software algorithms that identified cross-species conservation or variant effects on protein function (Fig. 1a) [33]. Subsequently, each variant was further filtered based on additional gene and variant-level information from publicly available databases: Online Mendelian Inheritance in Man (OMIM: <http://www.omim.org>); PubMed; the Human Gene Mutation Database (HGMD: <http://www.hgmd.cf.ac.uk>); and ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>). This sequential analysis eliminates the bias towards a particular mode of inheritance, mutation type, and the prior knowledge of the role of a specific gene in human disease. Additionally, this approach also integrates the fundamentals of WES analysis based on previously described SNV workflows and is compatible with ACMG guidelines for SNV filtering and classification of variant pathogenicity [3, 9, 27]. Analysis of interactomes, gene families, protein-protein interactions, co-expression data, and available literature further prioritized candidate variants (Fig. 1b). The patterns of segregation of the remaining candidate variants were analyzed by Sanger sequencing in available family members. Finally, the most promising candidate genes/variants were used to interrogate local databases and GeneMatcher [13, 14] to find additional affected cases with damaging variants in the same gene and with similar clinical phenotypes. When the BCHMG identified suspected causative variants in either newly discovered or recently published known disease genes, it was communicated to the clinical exome laboratory. Sanger confirmation of the suspected pathogenic variant was performed for all of the identified variants (confirmation rate of 100%) under the auspices of the CAP/CLIA-certified diagnostic laboratory and a formal updated clinical report was issued. Similarly, identified variants both in novel genes and in recent reports from the clinical exome laboratory were reported back to BCHMG, creating a bidirectional channel of intercommunication between research and clinical laboratories.



We also designed a parallel computational analysis of “bulk data” in order to accelerate our WES data analyses by scanning the BCHMG database for rare homozygous/heterozygous stop-gain variants culled from WES data of ~5000 research participants including our pilot study cases. In this analysis, we targeted the rare (MAF <0.5%) homozygous stop-gain variants from ~5000 research participants (including the 74 families) for each gene. Genes were then sorted according to the number of homozygous rare stop-gain variants existing in our database. Second, the genes were analyzed further according to SNV prioritization and filtering workflow as described above (Fig. 1). This approach was designed to accelerate the discovery of novel

disease genes that exhibit phenotypic consequences through a loss-of-function mechanism.

**Detection of CNVs**

CNV detection from WES data has been employed by different clinical laboratories to improve the molecular diagnostic rate [34, 35]. We applied several computational algorithms (CoNVex, Sanger Centre [<ftp://ftp.sanger.ac.uk/pub/users/pv1/CoNVex/Docs/CoNVex.pdf>], CoNIFER, and XHMM) to WES data to identify potential disease associated CNVs; these tools detect a clinically relevant intra-genic CNV when at least three contiguous exons are deleted [36, 37]. Therefore, in addition to these algorithms, we developed an in-house pipeline HMZDelFinder

(<https://github.com/BCM-Lupskilab/HMZDelFinder>) [38] to detect potential homozygous and hemizygous small intragenic deletions from WES data, including single exon “dropout alleles” that may be less robustly identified by current software.

## Results

The findings were sorted into three major categories: (1) known disease genes; (2) novel genes: genes with predicted pathogenic variants in two or more cases with similar phenotypes; and (3) potential candidate genes: genes with variants in a single case (Table 1). Predicted pathogenic variants in known or novel disease genes were identified in 27/74 (36%) families studied. In addition, predicted pathogenic variants in candidate genes were identified in 11/74 (15%) of cases, yielding an overall potential solved rate of 38/74 (51%) of cases. Our findings are represented by six “lessons learned” clusters: (1) de novo changes in known genes; (2) de novo changes in novel and candidate genes; (3) potential mosaicism in parents; (4) biallelic or hemizygous variants in genes known to convey recessive disease traits; (5) biallelic variants in novel and candidate recessive disease genes; and (6) blended phenotypes resulting from dual molecular diagnoses (Fig. 2a, b). Examples for each of these scenarios are presented below.

### De novo changes in known genes

Analysis of WES data from sets of parents and offspring initially yielded ~50–200 putative de novo variants per trio. Further prioritization, based upon read coverage, MAF and mutation type (non-synonymous, stop-gain, frameshift indels, and splicing variants), reduced the number of potential pathogenic de novo variants per family to ~0–5 variants per proband (Additional file 2: Figure S1). We detected de novo variants in five recently published genes (Table 1, Additional file 3: Table S2): *ZBTB20* associated with Primrose syndrome [39] (MIM

259050); *NR2F1* causing the Bosch-Boonstra-Schaaf optic atrophy syndrome [40] (MIM 615722); *DDX3X* associated with X-linked intellectual disability [41] (MIM 300958); *CACNA1A* implicated in non-fluctuating ataxia [42]; and *NALCN* associated with congenital contractures of the limbs and face, hypotonia, and developmental delay [43] (MIM 616266).

### De novo changes in novel and candidate genes

We further identified de novo variants in five novel and ten candidate genes (Table 1, Additional file 3: Table S2). A de novo SNV in *DHX30* c.2344C>T (p.Arg782Trp) was found in a patient with microcephaly, developmental delay/intellectual disability (DD/ID), mild cerebral volume loss, hypotonia, seizures, short stature, failure to thrive, and generalized hirsutism. Through data exchange [3, 5] with the clinical exome laboratory (Baylor Genetics) and GeneMatcher [13, 14], we found three additional unrelated participants with overlapping features, each of whom harbored a de novo SNV in *DHX30*. Interestingly, two of these participants shared the same de novo SNV c.2344C>T, occurring at a CpG dinucleotide (Fig. 3a) [44]. DEAH (Asp-Glu-Ala-His) box helicase 30 (DHX30, MIM 616423) is a member of the RNA helicase family including DHX and DDX proteins involved in DNA transcription, splicing and translation [45–47], and is highly expressed in the brain during neurogenesis [45]. Homozygous knockout of *Dhx30* in mice is lethal [45]. *DDX3X*, another member of this family, has been recognized recently to play a major role in DD/ID [41]. Additionally, *DHX37* has been proposed as a candidate gene for brain malformations [48], highlighting the importance of RNA helicases in development of the central nervous system.

*GSPT2* and *EBF3* each harbored one de novo variant in the genomes from two separate probands (Table 1, Additional file 3: Table S2). Each of these genes is located within CNV intervals identified previously

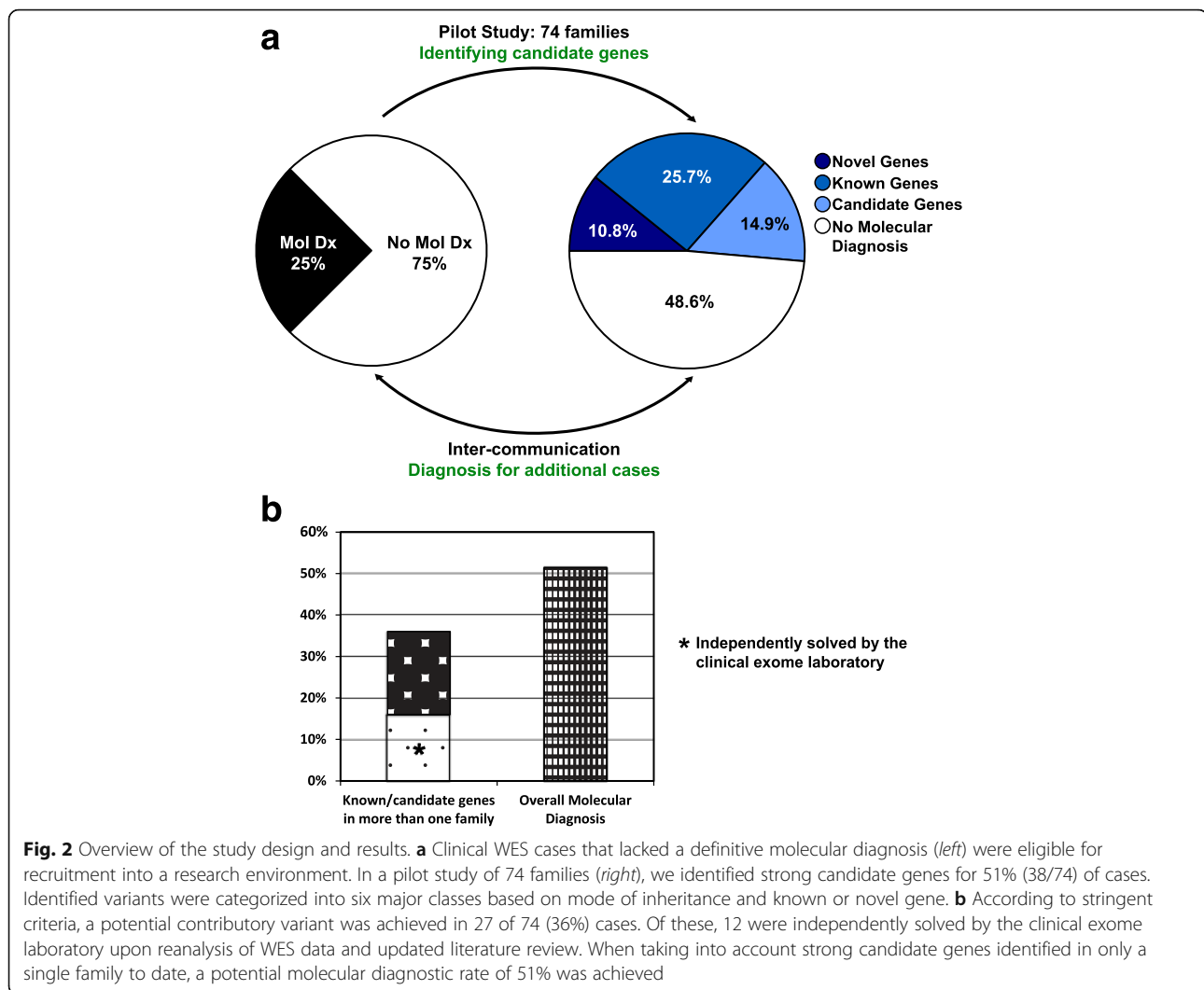
**Table 1** Molecular diagnoses in 74 cases are represented as three major categories: known genes, novel genes and candidate genes

Inheritance	Known genes	Novel genes	Candidate genes
De novo	<i>CACNA1A</i> , <i>DDX3X</i> (X2) <sup>a</sup> , <i>NALCN</i> (X2), <i>NR2F1</i> <sup>a</sup> , <i>ZBTB20</i>	<i>ATAD3A</i> , <i>DHX30</i> , <i>EBF3</i> , <i>EMC1</i> , <i>PURA</i> <sup>a</sup>	<i>CDK20</i> + <i>HIVEP1</i> , <i>DNAH7</i> , <i>GSPT2</i> , <i>GUCY2C</i> , <i>MICALL2</i> + <i>SLC30A7</i> , <i>MPP4</i> , <i>SYN3</i> , <i>SYTL2</i>
Autosomal/X-linked Recessive	<i>ABCA4</i> , <i>DDX3X</i> , <i>FBXL4</i> <sup>a</sup> , <i>NAA10</i> , <i>SLC13A5</i> <sup>a</sup> (X2), <i>TRAPP11</i> , <i>ZNF335</i> <sup>a</sup>	<i>GNB5</i> , <i>MIPEP</i> , <i>TANGO2</i> <sup>a</sup>	<i>ACOT1</i> <sup>a</sup> , <i>NRXN3</i> , <i>USP19</i>
Other		NA	NA
UPD	<i>SLC1A4</i> <sup>a</sup>		
Mosaic	<i>PIK3CD</i>		
Dual molecular diagnosis	<i>PMPCA</i> + <i>KCND3</i> <sup>a</sup> <i>POLRIC</i> + <i>SCN1B</i> <sup>a</sup>		

<sup>a</sup>Cases independently solved by the clinical exome laboratory re-analysis

Three major categories of identified genes include 13 candidate genes identified in 11 families and 26 known or novel genes in 27 families. Note that some families had more than one molecular diagnosis (indicated by *PMPCA* + *KCND3*, *POLRIC* + *SCN1B*, *CDK20* + *HIVEP1*, *MICALL2* + *SLC30A7*) and some genes were identified in more than one family (indicated by “(X2)” in the table)

NA non-applicable



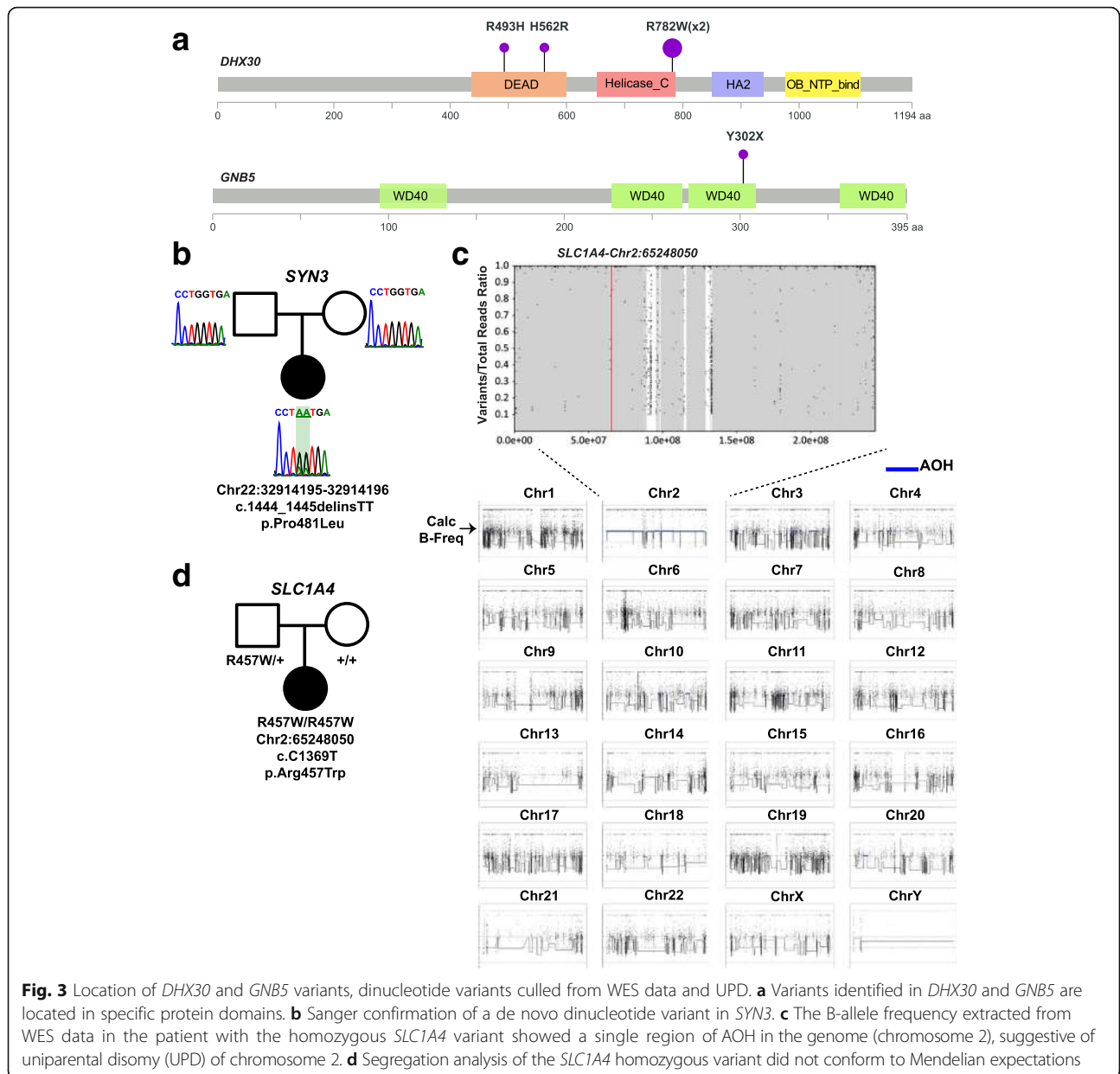
in patients with DD/ID, providing additional evidence for their roles in neurodevelopmental phenotypes (Additional file 4: Supplemental text) [49–51]. Through international collaborative efforts, additional families with de novo *EBF3* variants were identified with similar phenotypes [24–26].

A de novo double nucleotide substitution in *SYN3* (c.1444\_1445delinsTT; p.Pro481Leu), encoding synapsin III, was found in a proband with DD, seizures, atrophy of the cerebellar vermis, hypotonia, and a movement disorder (Fig. 3b). *SYN3* is a member of the synapsin gene family, which includes *SYN1* and *SYN2*, and plays a major role in dopamine regulation [52]. *MECP2* overexpression has been shown to result in upregulation of *SYN3* expression [53]. Taken together with the role of dopamine signaling in epileptogenesis [54], we hypothesize that the observed features of epilepsy and DD in this patient, might be associated with the de novo dinucleotide variant in *SYN3*. Finally, we identified de

novo missense variants in *ATAD3A* (MIM 617183) and *EMC1* (MIM 616875) and additional collaborative efforts supported these as disease genes with both monoallelic and biallelic pathogenic variants [20, 22].

#### Potential mosaicism in parents

In a multiplex family of Arab descent, in which three of five siblings have immunodeficiency, the clinical laboratory sequenced the first sibling without identifying a molecular diagnosis. Therefore, we transferred the first sibling's WES data to BHCMBG and offered WES to the other two affected family members and parents. This approach facilitated the examination of rare shared variants among three affected siblings and led to the detection of potential mosaic SNVs in the father. We identified a heterozygous variant c.1573G > A (p.Glu525Lys) in phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta (*PIK3CD*, MIM 602839) shared by the three affected siblings and inherited from their apparently



healthy father. This variant was reported previously to cause an autosomal dominant form of immunodeficiency [55]. To assess possible mosaicism in the father, we calculated the ratio of variant to total reads: 6/38 (15%). This ratio significantly deviates from the expected 50% of variant to total reads ( $p$  value 0.003). The same calculation in the three affected siblings yielded ratios of 20/53 (37%), 37/68 (54%), and 87/204 (42%), all of which were not significantly different from the expected 50% of variant to total reads ( $p$  value 0.241, 0.731, and 0.164, respectively). This suggested possible mosaicism in the father, whose post-WES clinical work-up revealed mild laboratory signs of immunodeficiency (with mild reduction of NK cell cytotoxicity) compared

to his severely affected children (Table 1, Additional file 3: Table S2) [56].

**Biallelic or hemizygous variants in genes known for recessive disease traits**

Homozygous or compound heterozygous variants were detected in six recently described or well established disease genes: *SLC13A5* (epileptic encephalopathy, MIM 615905); *FBXL4* (mitochondrial DNA depletion syndrome 13, MIM 615471); *ZNF335* (microcephaly in a single family, MIM 615095); *SLC1A4* (spastic tetraplegia, thin corpus callosum, and progressive microcephaly, MIM 616657); *TRAPPC11* (limb girdle muscular dystrophy type 2S, MIM 615356); and *ABCA4* (cone-rod



dystrophy, MIM 604116). Hemizygous variants were identified in two known X-linked disease genes: *NAA10* (Ogden syndrome, ID and long QT, MIM 300855); and *DDX3X* previously associated with X-linked ID (MIM 300958) (Table 1, Additional file 3: Table S2) [41, 57–65].

The homozygous missense variant identified in *SLC1A4*, encoding solute carrier family 1 (glutamate/neutral amino acid transporter), member 4, did not follow the expected Mendelian pattern of inheritance, since only the father carried the variant. The cSNP array data, which are performed as part of the clinical exome analysis, were integrated with calculated B allele frequency information from WES data [38] and showed absence of heterozygosity (AOH) limited to chromosome 2 and encompassing almost the entire chromosome, indicating paternal uniparental disomy with reduction of the variant to homozygosity as the responsible molecular mechanism (Fig. 3c, d).

#### Biallelic variants in novel and candidate recessive disease genes

The parallel computational algorithm for detecting rare homozygous stop-gain mutations from “bulk data” (i.e. ~5000 research exomes in the BHCMG database, including this pilot study), identified a rare homozygous stop-gain mutation in *GNB5* (MIM 617182) in a study participant with DD, hypotonia, retinopathy, and Mobitz type I atrioventricular block (Fig. 3a). A homozygous splice site variant (c.249+3 A>G) in *GNB5* was detected in two affected siblings from the clinical diagnostic laboratory exhibiting DD/ID, nystagmus, and sinus node dysfunction. *GNB5* encodes the Gβ5 protein, a guanine nucleotide-binding protein and a member of the signal-transducing G protein β subunit family. Additional participants with pathogenic variants in this disease gene were subsequently identified [21].

Recessive variants were found in novel (*MIPEP* and *TANGO2*) [19, 23, 66] as well as candidate genes: *USP19*, *NRXN3*, and *ACOT1* (Table 1; Additional file 3: Table S2). Each of these was encountered in only one family. *USP19*, found in a proband with epileptic encephalopathy, DD, and hypotonia, encodes ubiquitin-specific protease 19, which is involved in the regulation of ataxin 3 (spinocerebellar ataxia, MIM 109150) and interacts with *DOCK7* (epileptic encephalopathy, MIM 615859) and *USP7* (autism spectrum disorder) (Additional file 2: Figure S2) [67–69]. *NRXN3* encodes a member of the neurexin family of which *NRXN1* is associated with Pitt-Hopkins-like syndrome 2 (MIM 614325) [70–72] and *ACOT1* is involved in lipid metabolism (Additional file 4: Supplemental text). Finally, we identified compound heterozygous SNVs in *MIPEP* (MIM 617228) and homozygous SNVs in *TANGO2* (MIM 616878) and found additional evidence for pathogenicity

through collaboration [19, 23]. Combinations of SNVs and CNVs on alleles inherited in *trans* have been encountered in both of these genes, underscoring the potential utility of WES in identifying CNVs (including single exon deletions) as well as UPD.

#### Dual molecular diagnoses in known genes

A dual molecular diagnosis [8] was established in two probands. The analysis of trio WES data revealed both compound heterozygous variants in *PMPCA* and a de novo variant in *KCND3* in an individual with DD, ataxia, epilepsy, Hirschsprung disease, and abnormal mitochondrial function (Complex I and III deficiency). In another family, quad WES analysis identified a paternally inherited heterozygous missense variant of unknown significance in *SCN1B*, the gene for Brugada syndrome 5 (MIM 612838) [73], and compound heterozygous changes in *POLR1C*, segregating in two affected siblings (Table 1, Additional file 3: Table S2, Additional file 4: Supplemental text). At the time of the initial clinical exome analysis, neither *PMPCA* nor *POLR1C* were associated with spinocerebellar ataxia and leukodystrophy, respectively. On the other hand, the clinical exome laboratory did report both *KCND3* (MIM 607346) [74] and *SCN1B* (MIM 612838) as potential contributory variants for the patients’ manifested phenotypes. Albeit, the clinical features of the patients’ phenotypes could not be completely explained by the changes in the *KCND3* and *SCN1B* genes, and both cases were subsequently enrolled into our pilot study. Our research analysis prioritized both the *PMPCA* (MIM 613036) and *POLR1C* (MIM 610060) genes along with *KCND3* and *SCN1B* as potential contributory variants prompted by recent publications, which uncovered their roles in human disease [75, 76]. Of note, *POLR1C* was described in 2011 to cause recessive Treacher Collins syndrome (MIM 248390), however the patient’s clinical features were not matching those of Treacher Collins syndrome, potentially complicated by the blended phenotype due to the dual molecular diagnosis [77].

#### Discussion

We developed and implemented a workflow for: (1) the collection of new WES cases; (2) generation of additional data and analyses resources; and (3) application of further data filtering and adjuvant analysis methods to discover novel and candidate disease genes. This workflow and the data generated were used to optimize potential molecular diagnoses of Mendelian disease traits in clinical exome cases for which an initial molecular diagnosis was not achieved. This pilot study investigated 74 families and in 27 families (36%) identified a predicted damaging variant in a known or novel gene. If one considers damaging variants in candidate genes that



were observed in a single family to date (11/74; 15%), the cumulative rate of potential molecular diagnoses in this pilot cohort of unsolved clinical exomes would be 51% (38/74). A particular phenotype, such as the presence or absence of ID, did not significantly influence the rate of diagnosis; i.e. a potential molecular diagnosis was achieved in 54.2% (32/59) of cases with DD/ID and 42.9% (6/14) cases without DD/ID ( $p$  value = 0.05; Additional file 5: Table S3).

Beyond the impact on diagnosis, potential prognostic information, and genetic counseling, several of the newly established molecular diagnoses had implications for medical management. Examples include acetazolamide treatment for *CACNA1A* mutation, surveillance for arrhythmias in *GNB5* and *TANGO2*, and mitochondrial-specific surveillance for cardiac, renal, and liver involvement in *FBXL4*. For patients with the *SCN1B* variant, current practice guidelines suggest treatment with isoproterenol as a first-line agent for electrical storm (ventricular fibrillation) and consideration of an implantable cardioverter defibrillator (ICD) to prevent sudden death [78]. In a recent study applying WES and CNV testing in clinically diagnosed primary immune deficiency diseases (PIDD), a molecular diagnosis was achieved in about 40% of unrelated probands; clinical diagnosis was revised in about half (60/110) and management was directly altered in nearly one-quarter of families based on molecular findings [56].

In about half (12/27, 44.4%) of the families, diagnosed by a known or novel disease gene, the same diagnosis was achieved independently of the research analyses as part of the routine reanalysis of WES data by the clinical exome laboratory based on recent gene discoveries (Table 1). The combined efforts of the clinical and research laboratories have led to multiple reports of novel disease genes, most of which are described in much greater phenotypic and molecular detail in independent publications (*PURA*, *TANGO2*, *EMC1*, *GNB5*, *ATAD3A*, *MIPEP*, and *EBF3*) [18–26]. Since all of these discoveries included cases from the 74 in our pilot study and can be attributed to combined efforts between the clinical and research laboratories, they are included in the overall stated molecular diagnostic yield of 36%.

Our study demonstrates that an increase in the molecular diagnostic yield can be achieved through systematic and comprehensive reanalysis of clinical exome data. Of 74 cases with non-diagnostic clinical exomes, a molecular diagnosis was achieved in 30/63 trios (47.6%), 4/6 singleton cases (66.7%), 1/1 multiplex family involving three affected siblings, and 3/4 (75%) quartet families (Additional file 6: Table S4). This increased diagnostic yield beyond clinical WES may be attributed to: (1) the rapid pace of Mendelian gene discovery; (2) the use of trio sequencing; and (3) the extensive research reanalysis

and implementation of novel tools for identification of de novo and CNV variants. Vigorous collaboration between clinical and research efforts can enhance diagnostic yield and fuel novel disease gene discovery. For instance, the clinical exome laboratories' standard operating procedure for exome analysis and variant interpretation adheres to current ACMG guidelines, which include recommendations for the reporting of pathogenic variants, likely pathogenic variants, and variants of unknown significance (VUS) in known (established) disease genes [27]. These stringent criteria do not provide for reporting of novel disease genes at the time of initial discovery, as by definition they cannot be considered pathogenic or likely pathogenic until a causal relationship between variation at a particular locus and disease has been firmly established. Thus, novel variation and novel disease genes are best studied in a research environment. In our study, the pursuit of novel disease genes for rare disease benefits from a tremendous resource: the combined exome variant dataset from over 15,000 cases referred to either the clinical exome laboratory or the BHCMG research laboratory. Research studies also support reporting of novel disease genes at the time of initial discovery, an important step toward gaining sufficient evidence to meet ACMG guidelines for reporting by a diagnostic laboratory. Additionally, as certain tools are developed and validated in a research setting and are honed to better efficiency, they are frequently translated into the clinical pipeline; GeneMatcher, developed by the BHCMG, is one such example [13, 14].

In this study, trio-WES analysis combined with de novo variant and CNV detection were invaluable for achieving an improved molecular diagnostic rate. The addition of parental samples (i.e. trio analysis) led to molecular diagnoses in 47.6% (30/63) of cases for which proband-WES was non-diagnostic. Analysis of trio-WES data supports the efficient identification of de novo and compound heterozygous variants, leading to improved analysis efficiency across all inheritance models. In an attempt to limit false-positive calls, we required that both parents have at least ten reference reads. Thus, a limitation to our analysis is that regions with poor coverage in either or both parents were parsed during variant filtering. Taking these regions/genes into account by flagging them in the bioinformatics pipeline and performing Sanger confirmation of any candidate variants at these loci may further optimize identification of candidate de novo variants. As the cost of WES continues to fall, we anticipate that trio-WES may ultimately be favored over proband-WES in the clinical setting of a sporadic suspected genetic disease.

Trio analysis also supported the use of a newly developed tool, DNM-Finder (<https://github.com/BCM-Lups/kilab/DNM-Finder>). In our study, 27% of the pathogenic

or candidate variants were de novo variants. Trio analysis can also increase the molecular diagnostic yield over that attained by studies of the proband alone. This is due, in part, to being able to more readily detect by computational filtering de novo variants (Additional file 2: Figure S1). All DNA was extracted from blood or saliva samples in this study; however, detection of mosaicism may be enhanced by including concurrent analyses of more than one lineage: ectodermal (buccal swab, hair), mesodermal, and/or endodermal (blood, saliva). The degree of mosaicism correlates with the timing of mutation, cell migration and designation during development, and tissue-specific growth profile [79]. Taking into account such considerations is important to guide the detection of mosaicism. Additionally, compound heterozygosity is more readily detected.

Analysis of all modes of inheritance and comprehensive review beyond the first identified potential pathogenic variant allowed for identification of dual molecular diagnoses in two cases (e.g. *PMPCA* and *KCND3*; *POLR1C* and *SCN1B*), consistent with previous reports of diagnoses in 5–7% of molecularly diagnosed cases [3, 5, 8, 9, 56, 80, 81]. These cases underscore the need for a systematic and comprehensive approach to exome variant analysis for all possible modes of inheritance as well as an updated

literature review [12]. This is especially true of cases in which we observe apparent “phenotypic expansion” (clinical features previously unreported in association with the gene) or unexpected clinical severity [8].

Ancillary approaches to “bulk data” reanalysis were shown to enhance the molecular diagnostic rate. These include the examination of rare homozygous stop-gain SNVs, extraction of copy number information from the WES data to search for regions of absence of heterozygosity (AOH) (i.e. *ABCA4* and *FBXL4*) and assessing UPD (i.e. *SLC1A4*) and small deletion CNVs that might escape detection by CMA (Fig. 3c, Table 2, Additional file 2: Figure S3) [19, 38]. We also underscore the importance of capturing single-exon genes on WES platforms (*PURA*, *GSPT2*); these can be challenging as they are often rich in GC content, similar to the first exon of multiexonic genes [82].

Our study highlights the limitations of different WES data variant calling pipelines. The research laboratory opts for less stringent filtering of variants, which allows identification of intronic variants located farther from the splice site (i.e. *ABCA4*); albeit, this increases the number of variants to be investigated and consequently time taken for analysis. For instance, clinical WES did not reveal a conclusive molecular variant in a proband who presented with cone-rod dystrophy from

**Table 2** Lessons learned that may increase the molecular diagnostic yield from unsolved clinical exomes

Lesson Learned	Examples
A) Collaboration between research and clinical laboratories	Sharing data, open communication of findings, access to additional patients with damaging variants
B) Facilitating research collaborations including local and international efforts	GeneMatcher for the identification of unrelated affected individuals with the same novel disease
C) Ancillary approaches to enhance molecular diagnostic rate:	
1) Detection of AOH and CNVs from WES	1) <i>ABCA4</i> , <i>SLC1A4</i> , <i>TANGO2</i> , <i>FBXL4</i>
2) Annotation of single exon genes	2) <i>PURA</i> , <i>GSPT2</i>
3) Analyze intronic variants (including those in the unfiltered vcf files)	3) <i>TRAPPC11</i>
4) Optimization of the bioinformatic filters	4) <i>ABCA4</i> , <i>NRXN3</i>
5) Look for <i>in trans</i> inheritance of SNVs and CNVs	5) <i>TANGO2</i> <sup>a</sup> , <i>MIPEP</i>
6) Increase fidelity of calling dinucleotide substitutions	6) <i>SYN3</i>
D) Ancillary approaches for non-conclusive WES:	
1) Consider dual molecular diagnoses	1) <i>PMPCA</i> and <i>KCND3</i> ; <i>POLR1C</i> and <i>SCN1B</i> <sup>a</sup>
2) Analyze additional family members	2) <i>SLC13A5</i> , <i>NAA10</i>
3) Look for parental mosaicism	3) <i>PIK3CD</i>
4) Identify homozygous stop-gain SNVs from bulk data to identify additional affected individuals	4) <i>GNB5</i> <sup>a</sup>
E) Consideration of different inheritance patterns and variant types at a single locus	
1) AR, AD	1) <i>ATAD3A</i> , <i>EBF3</i> , <i>EMC1</i> , <i>GSPT2</i> , <i>NALCN</i> , <i>GUCY2C</i>
2) AD, CNV del	2) <i>EBF3</i> , <i>PURA</i> , <i>ZBTB20</i>
3) AR, CNV del	3) <i>NRXN3</i>
4) XLR, CNV dup	4) <i>GSPT2</i>
5) AD, Tandem repeats	5) <i>CACNA1A</i> <sup>a</sup>
6) AR, AR	6) <i>POLR1C</i>

<sup>a</sup>The molecular diagnosis in these genes had direct implications for clinical management

AD autosomal dominant, AOH absence of heterozygosity, AR autosomal recessive, del deletion, dup duplication, CNV copy number variation, SNV single nucleotide variants, WES whole exome sequencing, WGS whole genome sequencing, XLR X-linked recessive

Indian-Asian ancestry and consanguineous parents. Our research bioinformatics pipeline detected a homozygous intronic frameshift deletion, which possibly could explain the phenotype. The inherited deletion was more common in the South Asian population (seen 14 times on ExAC filtered variants). We also identified shared limitations between both pipelines, i.e. dinucleotide substitutions called as two separate variants. Our pilot study reflects the strength of collaborative efforts and iterative analyses between clinical and research teams for the optimization of computational pipelines to analyze WES data and further understand the genetic architecture underlying disease, both Mendelian disease and common/complex traits (Table 2).

This study confirms and extends our understanding of the relationship between variation at a locus and disease expression. We identified examples of genes for which different variant alleles are associated with either a dominant or a recessive disease trait (e.g. *EMC1*, *ATAD3A*, *NALCN*, and *GUCY2C*) [20, 22, 43, 83–86]. Other genes presented with recessive changes or de novo SNVs/CNVs underlying different phenotypes (e.g. *NRXN3*) for which heterozygous CNVs have previously been reported with disease [87]. Several genes were found to have de novo SNV and CNV alleles: *EBF3*, *GSPT2*, *PURA*, and *ZBTB20* [18, 23, 26, 39, 50, 51, 88–90]. This phenomenon can be exemplified by recent reports of glutamate receptor, ionotropic, delta 2 (*GRID2*, MIM 616204), which can contribute to neurodevelopmental disorders and ataxia through recessive and de novo SNVs as well as homozygous and de novo partial CNV deletions (Table 2) [91–95]. Furthermore, *CACNA1A* gene function can be altered either by de novo SNVs or trinucleotide repeat expansion (CAG) to manifest with a broad range of neurological disorders [42, 96, 97]. Yet another class of genes can present with distinct disorders due to allelic heterogeneity, e.g. *POLR1C* associated with recessive hypomyelinating leukodystrophy 11 (MIM 616494) and recessive Treacher Collins syndrome 3 (MIM 248390) [76] and illustrate allelic affinity wherein different clinical disease phenotypes are due to different alleles at the same locus [98]. In addition, compound heterozygous SNV and CNV alleles have been exemplified in other patients exhibiting *MIPEP* and *TANGO2* associated phenotypes (Table 2) [19, 23, 66]. These findings of different variant allele types and combinations thereof underscore the notion that defining potential causative alleles necessitates consideration of all variant types (SNV, indels, and CNV) and a multitude of potential genetic mechanisms (e.g. alleles causing dominant traits while other alleles cause recessive disease traits, *EMC1* and *ATAD3A* [20, 22]) and inheritance patterns while seeking answers for the genetic basis of Mendelian phenotypes.

These data dramatically illustrate how genetic disease can be driven by rare recent variants introduced into a family, further supporting the clan genomics hypothesis [99]. The clan genomics model could explain how the same gene may contribute to disease by either de novo and/or recessive SNVs and/or CNVs. De novo events arise in each generation from the failure of DNA repair or replication errors [100]. Rare de novo events with strong mutation effects (i.e. rare variants, predicted to be deleterious) may manifest as disease in the first generation [100, 101], in contrast to weaker variant alleles which require a second pathogenic allele or reduction to homozygosity in order to manifest as a trait or disease in subsequent generations (i.e. *ATAD3A*, *EMC1*, *GUCY2C*, and *NALCN* genes) [20, 22, 43, 83–86]. Remarkably, some carrier states for recessive disease may raise the susceptibility for a common, complex trait as age progresses, as demonstrated by heterozygous SNVs observed in *ABCA4*, also known as *ABCR*, (MIM 153800), *CFTR* (MIM 167800), and *LDLR* (MIM 143890) genes leading to age-related macular degeneration, pancreatitis, and familial hypercholesterolemia, respectively [63, 99, 102–107]. Taken together, the heterogeneity of phenotypes, different inheritance patterns, and different kinds of variants (SNVs or CNVs) presented in our study may lead to an enhanced understanding of a unified genetic model for human disease [99].

As the price of WES falls, trio analysis will be more efficient for sporadic traits than singleton analysis, in that it allows for detection of de novo variants and phasing of compound heterozygous variants prior to Sanger validation. This enhances detection of relevant variants in known genes and serves to catalyze novel gene discovery. Based on our study, we propose a workflow for the management of non-conclusive singleton clinical WES (Table 2, Additional file 2: Figure S4). This workflow intends to investigate the many possible scenarios that could be encountered when clinical exomes are non-productive for a specific molecular diagnosis. These include: (1) de novo missense SNVs; (2) dual molecular diagnoses; (3) multiple affected family members; (4) potential parental mosaicism; (5) in *trans* inheritance of SNVs and CNVs; (6) loosening the default parameters of bioinformatics filters for parsing variants; and (7) potential intronic unfiltered variants on variant calling files.

A study that applied WGS to 50 clinical cases that remained unsolved by genomic studies, i.e. in which probands initially had non-conclusive clinical microarrays and WES, revealed an additional molecular diagnostic yield of 42%. This 42% molecular diagnostic yield was driven mainly by de novo SNVs and CNVs impacting the coding regions [108]. In the current research reanalysis of 74 clinical exomes, by implementing WES reanalysis and augmentation with additional family

members, WES achieved a similarly increased molecular diagnosis yield (36%) of unsolved exomes and even higher increased rate (51%) if potential candidate genes are considered. These pilot studies data (50 WGS versus 74 WES + reanalysis) suggest that currently WGS offers no significant advantage to WES and reanalysis when it comes to increasing molecular diagnostic yield from unsolved clinical exomes. Nevertheless, the diagnostic yield by WGS may potentially be increased further by the development of new bioinformatic algorithms to detect intronic variants or variants in regulatory regions as well as structural variations followed by further functional characterization. Future studies comparing WGS versus WES performances for diagnostic yield should use both of these techniques in parallel while also taking other parameters (coverage and cost) into account.

## Conclusion

We have demonstrated that systematic study of “unsolved clinical exomes” can provide a rich resource for Mendelian gene discovery and that reanalysis of data coupled with incorporation of additional family member WES data can improve the molecular diagnostic rate. These research studies can, in turn, provide the basis for improving interpretive algorithms for clinical WES analyses (Additional file 2: Figure S4). Our data additionally highlight the remarkable contribution of new mutation to disease including blended phenotypes resulting from dual molecular diagnoses [3, 5, 8]. Speculation based on these pilot study data suggests that if one considers the 25–30% molecular diagnostic rate achieved by initial clinical exome analyses, in combination with the 51% rate found in these pilot research studies of unsolved clinical exomes, genomic analyses by WES have the potential to identify a rare variant and gene that implicate a molecular diagnosis which could impact clinical decisions in ~ 63 (25% + [75%\*51%]) to 66% (30% + [70%\*51%]) or the majority of cases.

## Additional files

**Additional file 1: Table S1.** Summary of phenotypic features for each case. (XLSX 1227 kb)

**Additional file 2: Figures S1, S2, S3, and S4.** Computational *de novo* variant detection from WES data, interactome analysis of *USP19*, AOH maps for probands with *ABCA4* and *FBXL4* variants and clinical whole exome management. (PDF 3462 kb)

**Additional file 3: Table S2.** *De novo* and recessive variants in candidate or recently published genes identified by collaboration between BG and BHCMG. (XLSX 28 kb)

**Additional file 4:** Supplemental text. (DOCX 118 kb)

**Additional file 5: Table S3.** Diagnostic rate differs among groups of patients with different phenotypic features. (XLSX 8 kb)

**Additional file 6: Table S4.** Diagnostic rate in singleton cases, trios, quartets and families with multiple affected individuals. (XLSX 28 kb)

## Abbreviations

AOH: Absence of heterozygosity; ARIC: Atherosclerosis Risk in Communities; BG: Baylor Genetics; BHCMG: Baylor-Hopkins Center for Mendelian Genomics; CNV: Copy number variant; DD/ID: Developmental delay/intellectual disability; ExAC: The Exome Aggregation Consortium; MAFs: Minor allele frequencies; SNV: Single nucleotide variant; UPD: Uniparental disomy; WES: Whole exome sequencing; WGS: Whole genome sequencing

## Acknowledgements

We thank all patients and their families and the referring physicians who submitted samples for testing. No additional compensation was received for these contributions.

## Funding

This work was funded in part by the US National Human Genome Research Institute (NHGRI)/National Heart Lung and Blood Institute (NHLBI) grant number UM1HG006542 to the Baylor-Hopkins Center for Mendelian Genomics (BHCMG). TH and JEP are supported by the NIH T32 GM07526 Medical Genetics Research Fellowship Program. JEP is supported by a Chao Physician-Scientist Award through the Ting Tsung and Wei Fong Chao Foundation. WW was supported by Career Development Award K23NS078056 from the US National Institute of Neurological Disease and Stroke (NINDS). The Western France consortium HUGODIMS, was supported by a grant from the French Ministry of Health and from the Health Regional Agency from Poitou-Charentes (HUGODIMS, 2013, RC14\_0107). RAL is supported in part by the Genetics Resource Association of Texas (GREAT), Houston, Texas.

## Availability of data and materials

All reported disease associated variants in our pilot study have been deposited into ClinVar in agreement with institutional review board approval and patient consent. All reported variants not published elsewhere have been deposited in ClinVar, accession numbers SCV000494150 through SCV000494199.

## Authors' contributions

MKE, ZCA, TH, JEP, and JRL analyzed the data and wrote the manuscript. JRL supervised the study. JAR, TG, ASP, SK, SM, DL, JEP, WW, SP, PL, WB, SRL, CPS, MFW, CAB, RAL, LP, BHG, JWB, FS, JSO, SNJ, TC, HD, JH, DMM, FX, ALB, EB CME, SEP, VRS, RAG, YY, and JRL generated and advised on data analysis. WW, SK, SM, DL, WW SRL, CPS, MFW, CAB, RAL, LP, BHG, JWB, FS, and JSO identified and collected patients. All authors have read and approved the final manuscript.

## Competing interests

Baylor College of Medicine (BCM) and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of the Baylor Genetics (BG), which performs clinical exome sequencing. JAR, FX, MW, JEP, CME, SEP, ALB, YY, RAG, and JRL are employees of BCM and derive support through a professional services agreement with the BG. SEP and JRL serve on the Scientific Advisory Board of the BG. CME serves as Chief Medical Officer and Chief Quality Officer of the BG. JAR reports personal fees from Signature Genomic Laboratories, PerkinElmer, Inc., in the past 36 months. JRL has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, has stock options in Lasergen, Inc., and is a coinventor of US and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. The remaining authors declare that they have no competing interests.

## Consent for publication

A written consent was obtained to publish the details of all patients from the parents/legal guardians.

## Ethics approval and consent to participate

This research study was approved by the Baylor College of Medicine Institutional Review Board (protocol H-29697). The Baylor College of Medicine IRB (IORG number 0000055) is recognized by the United States Office of Human Research Protections (OHRP) and Food and Drug Administration (FDA) under the federal wide assurance program. The Baylor College of Medicine IRB is also fully accredited by the Association for the Accreditation of Human Research Protection Programs (AAHRPP).

For individuals who were alive at the time the research began, written informed consent was obtained from them or their legally authorized representative/parent. For those who were deceased at the time of the initiation of the study (and where existing specimens or data were utilized in our analysis) parents were notified of the study and agreed verbally to the study. Under United States federal regulations, it is impossible to obtain consent for a deceased individual.

#### Author details

<sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. <sup>2</sup>Institute of Computer Science, Warsaw University of Technology, 00-665 Warsaw, Poland. <sup>3</sup>Norwegian National Unit for Newborn Screening, Women and Children's Division, Oslo University Hospital, 0424 Oslo, Norway. <sup>4</sup>CHU Nantes, Service de Génétique Médicale, 9 quai Moncoussu, 44093 Nantes, CEDEX 1, France. <sup>5</sup>Atlantic Gene Therapies, UMR1089, Nantes, France. <sup>6</sup>Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. <sup>7</sup>Department of Pediatrics, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. <sup>8</sup>Texas Children's Hospital, Houston, TX 77030, USA. <sup>9</sup>Baylor Genetics, Baylor College of Medicine, Houston, TX 77030, USA. <sup>10</sup>Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA. <sup>11</sup>Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA. <sup>12</sup>Texas Children's Hospital Center for Human Immuno-Biology, Houston, TX, USA. <sup>13</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. <sup>14</sup>Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77030, USA. <sup>15</sup>Texas Children's Cancer Center, Texas Children's Hospital, Houston, TX 77030, USA. <sup>16</sup>Baylor-Hopkins Center for Mendelian Genomics, Baltimore, MD, USA. <sup>17</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 604B, Houston, TX 77030-3498, USA. <sup>18</sup>Present Address: Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, 350 W. 11th Street, Indianapolis, IN 46202, USA.

Received: 5 July 2016 Accepted: 8 February 2017

Published online: 21 March 2017

#### References

- Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med*. 2010;362(13):1181–91.
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12(11):745–55.
- 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;369(16):1502–11.
- Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med*. 2011;13(3):255–62.
- 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; 312(18):1870–9.
- de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*. 2012;367(20):1921–9.
- Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annu Rev Med*. 2012;63:35–61.
- Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med*. 2017;376(1):21–31.
- 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;17(7): 578–86.
- 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;312(18):1880–7.
- Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med*. 2016;18(7):696–704.
- Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med*. 2017;19(2):209–14.
- Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*. 2015;36(10):928–30.
- Sobreira N, Schiettecatte F, Boehm C, Valle D, Hamosh A. New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Hum Mutat*. 2015;36(4):425–31.
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet*. 2015;97(2):199–215.
- Bamshad MJ, Shendure JA, Valle D, Hamosh A, Lupski JR, Gibbs RA, et al. The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. *Am J Med Genet A*. 2012; 158A(7):1523–5.
- Lupski JR. Clinical genomics: from a truly personal genome viewpoint. *Hum Genet*. 2016;135(6):591–601.
- Lalani SR, Zhang J, Schaaf CP, Brown CW, Magoulas P, Tsai AC, et al. Mutations in *PURA* cause profound neonatal hypotonia, seizures, and encephalopathy in 5q31.3 microdeletion syndrome. *Am J Hum Genet*. 2014; 95(5):579–83.
- Lalani SR, Liu P, Rosenfeld JA, Watkin LB, Chiang T, Leduc MS, et al. Recurrent muscle weakness with rhabdomyolysis, metabolic crises, and cardiac arrhythmia due to bi-allelic *TANGO2* mutations. *Am J Hum Genet*. 2016;98(2):347–57.
- Harel T, Yesil G, Bayram Y, Coban-Akdemir Z, Charng WL, Karaca E, et al. Monoallelic and biallelic variants in *EMC1* identified in individuals with global developmental delay, hypotonia, scoliosis, and cerebellar atrophy. *Am J Hum Genet*. 2016;98(3):562–70.
- Lodder EM, De Nittis P, Koopman CD, Wiszniewski W, de Souza CF M, Lahrouchi N, et al. *GNB5* mutations cause an autosomal-recessive multisystem syndrome with sinus bradycardia and cognitive disability. *Am J Hum Genet*. 2016;99(2013):786. *Am J Hum Genet* 2016, 99(3):704–710. Erratum.
- Harel T, Yoon WH, Garone C, Gu S, Coban-Akdemir Z, Eldomery MK, et al. Recurrent *de novo* and biallelic variation of *ATAD3A*, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. *Am J Hum Genet*. 2016;99(4):831–45.
- Eldomery MK, Akdemir ZC, Vogtle FN, Charng WL, Mulica P, Rosenfeld JA, et al. *MIPEP* recessive variants cause a syndrome of left ventricular non-compaction, hypotonia, and infantile death. *Genome Med*. 2016;8(1):106.
- Harms FL, Girisha KM, Hardigan AA, Kortum F, Shukla A, Alawi M, et al. Mutations in *EBF3* disturb transcriptional profiles and cause intellectual disability, ataxia, and facial dysmorphism. *Am J Hum Genet*. 2017;100(1): 117–27.
- Chao HT, Davids M, Burke E, Pappas JG, Rosenfeld JA, McCarty AJ, et al. A syndromic neurodevelopmental disorder caused by *de novo* variants in *EBF3*. *Am J Hum Genet*. 2017;100(1):128–37.
- Sleven H, Welsh SJ, Yu J, Churchill ME, Wright CF, Henderson A, et al. *De novo* mutations in *EBF3* cause a neurodevelopmental syndrome. *Am J Hum Genet*. 2017;100(1):138–50.
- 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; 17(5):405–24.
- Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, et al. PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum Mutat*. 2013;34(4):566–71.
- 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;12(7):R68.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.

31. Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundquist A, English A, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. *BMC Bioinformatics*. 2014;15:30.
32. Challis D, Yu J, Evani US, Jackson AR, Paithankar S, Coarfa C, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics*. 2012;13:8.
33. Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061–73.
34. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res*. 2012;22(8):1525–32.
35. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet*. 2012;91(4):597–607.
36. de Ligt J, Boone PM, Pfundt R, Vissers LE, Richmond T, Geoghegan J, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat*. 2013;34(10):1439–48.
37. de Ligt J, Boone PM, Pfundt R, Vissers LE, de Leeuw N, Shaw C, et al. Platform comparison of detecting copy number variants with microarrays and whole-exome sequencing. *Genom Data*. 2014;2:144–6.
38. Gambin T, Akdemir ZC, Yuan B, Gu S, Chiang T, Carvalho CM, Shaw C, Jhangiani S, Boone PM, Eldomery MK, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res* 2017;45(4):1633–48.
39. Cordeddu V, Redeker B, Stellacci E, Jongejan A, Fragale A, Bradley TE, et al. Mutations in *ZBTB20* cause Primrose syndrome. *Nat Genet*. 2014;46(8):815–7.
40. Bosch DG, Boonstra FN, Gonzaga-Jauregui C, Xu M, de Ligt J, Jhangiani S, et al. *NR2F1* mutations cause optic atrophy with intellectual disability. *Am J Hum Genet*. 2014;94(2):303–9.
41. Snijders Blok L, Madsen E, Juusola J, Gilissen C, Baralle D, Reijnders MR, et al. Mutations in *DDX3X* are a common cause of unexplained intellectual disability with gender-specific effects on Wnt signaling. *Am J Hum Genet*. 2015;97(2):343–52.
42. Tonelli A, D'Angelo MG, Salati R, Villa L, Germinasi C, Frattini T, et al. Early onset, non fluctuating spinocerebellar ataxia and a novel missense mutation in *CACNA1A* gene. *J Neurol Sci*. 2006;241(1-2):13–7.
43. Chong JX, McMillin MJ, Shively KM, Beck AE, Marvin CT, Armenteros JR, et al. *De novo* mutations in *NALCN* cause a syndrome characterized by congenital contractures of the limbs and face, hypotonia, and developmental delay. *Am J Hum Genet*. 2015;96(3):462–73.
44. de Beer TA, Laskowski RA, Parks SL, Sipos B, Goldman N, Thornton JM. Amino acid changes in disease-associated variants differ radically from variants observed in the 1000 genomes project dataset. *PLoS Comput Biol*. 2013;9(12), e1003382.
45. Zheng HJ, Tsukahara M, Liu E, Ye L, Xiong H, Noguchi S, et al. The novel helicase helG (DHX30) is expressed during gastrulation in mice and has a structure similar to a human DExH box helicase. *Stem Cells Dev*. 2015;24(3):372–83.
46. Pause A, Methot N, Sonenberg N. The HRIGRXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol*. 1993;13(11):6789–98.
47. Abdelhaleem M. RNA helicases: regulators of differentiation. *Clin Biochem*. 2005;38(6):499–503.
48. Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, et al. Genes that affect brain structure and function identified by rare variant analyses of Mendelian neurologic disease. *Neuron*. 2015; 88(3):499–513.
49. Hoshino S, Imai M, Mizutani M, Kikuchi Y, Hanaoka F, Ui M, et al. Molecular cloning of a novel member of the eukaryotic polypeptide chain-releasing factors (eRF). Its identification as eRF3 interacting with eRF1. *J Biol Chem*. 1998;273(35):22254–9.
50. Whibley AC, Plagnol V, Tarpey PS, Abidi F, Fullston T, Choma MK, et al. Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability. *Am J Hum Genet*. 2010;87(2):173–88.
51. Faria AC, Rabbi-Bortolini E, Reboucas MR, de S Thiago Pereira AL, Frasson MG, Atique R, et al. Craniosynostosis in 10q26 deletion patients: A consequence of brain underdevelopment or altered suture biology? *Am J Med Genet A*. 2016;170(2):403–9.
52. Zaltieri M, Grigoletto J, Longhena F, Navarria L, Favero G, Castrezzi S, et al. alpha-synuclein and synapsin III cooperatively regulate synaptic function in dopamine neurons. *J Cell Sci*. 2015;128(13):2231–43.
53. Orlic-Milacic M, Kaufman L, Mikhailov A, Cheung AY, Mahmood H, Ellis J, et al. Over-expression of either *MECP2\_e1* or *MECP2\_e2* in neuronally differentiated cells results in different patterns of gene expression. *PLoS One*. 2014;9(4), e91742.
54. Bozzi Y, Borrelli E. The role of dopamine signaling in epileptogenesis. *Front Cell Neurosci*. 2013;7:157.
55. Lucas CL, Kuehn HS, Zhao F, Niemela JE, Deenick EK, Palendira U, et al. Dominant-activating germline mutations in the gene encoding the P(3)K catalytic subunit p110delta result in T cell senescence and human immunodeficiency. *Nat Immunol*. 2014;15(1):88–97.
56. Stray-Pedersen A, Sorte HS, Samarakoon P, Gambin T, Chinn IK, Coban Akdemir ZH, et al. Primary immunodeficiency diseases: Genomic approaches delineate heterogeneous Mendelian disorders. *J Allergy Clin Immunol*. 2017;139(1):232–45.
57. Thevenon J, Milh M, Feillet F, St-Onge J, Duffourd Y, Juge C, et al. Mutations in *SLC13A5* cause autosomal-recessive epileptic encephalopathy with seizure onset in the first days of life. *Am J Hum Genet*. 2014;95(1): 113–20.
58. Bonnen PE, Yarham JW, Besse A, Wu P, Faqeh EA, Al-Asmari AM, et al. Mutations in *FBXL4* cause mitochondrial encephalopathy and a disorder of mitochondrial DNA maintenance. *Am J Hum Genet*. 2013;93(3):471–81.
59. Yang YJ, Baltus AE, Mathew RS, Murphy EA, Evrony GD, Gonzalez DM, et al. Microcephaly gene links trithorax and *REST/NRSF* to control neural stem cell proliferation and differentiation. *Cell*. 2012;151(5):1097–112.
60. Damseh N, Simonin A, J alas C, Picoraro JA, Shaag A, Cho MT, et al. Mutations in *SLC1A4*, encoding the brain serine transporter, are associated with developmental delay, microcephaly and hypomyelination. *J Med Genet*. 2015;52(8):541–7.
61. Rope AF, Wang K, Evjenth R, Xing J, Johnston JJ, Swensen JJ, et al. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet*. 2011;89(1):28–43.
62. Bogershausen N, Shahrzad N, Chong JX, von Kleist-Retzow JC, Stanga D, Li Y, et al. Recessive *TRAPPC11* mutations cause a disease spectrum of limb girdle muscular dystrophy and myopathy with movement disorder and intellectual disability. *Am J Hum Genet*. 2013;93(1):181–90.
63. Shroyer NF, Lewis RA, Allikmets R, Singh N, Dean M, Leppert M, et al. The rod photoreceptor ATP-binding cassette transporter gene, *ABCR*, and retinal disease: from monogenic to multifactorial. *Vision Res*. 1999;39(15):2537–44.
64. Casey JP, Stove SI, McGorrian C, Galvin J, Blenski M, Dunne A, et al. *NAA10* mutation causing a novel intellectual disability syndrome with Long QT due to N-terminal acetyltransferase impairment. *Sci Rep*. 2015;5:16022.
65. Popp B, Stove SI, Ende S, Myklebust LM, Hoyer J, Sticht H, et al. *De novo* missense mutations in the *NAA10* gene cause severe non-syndromic developmental delay in males and females. *Eur J Hum Genet*. 2015;23(5):602–9.
66. Kremer LS, Distelmaier F, Alhaddad B, Hempel M, Iuso A, Kupper C, et al. Bi-allelic truncating mutations in *TANGO2* cause infancy-onset recurrent metabolic crises with encephalocardiomyopathy. *Am J Hum Genet*. 2016; 98(2):358–62.
67. He WT, Zheng XM, Zhang YH, Gao YG, Song AX, van der Goot FG, et al. Cytoplasmic ubiquitin-specific protease 19 (USP19) modulates aggregation of polyglutamine-expanded Ataxin-3 and Huntingtin through the HSP90 chaperone. *PLoS One*. 2016;11(1), e0147515.
68. Hao YH, Fountain Jr MD, Fon Tacer K, Xia F, Bi W, Kang SH, et al. USP7 Acts as a molecular rheostat to promote WASH-dependent endosomal protein recycling and is mutated in a human neurodevelopmental disorder. *Mol Cell*. 2015;59(6):956–69.
69. Perrault I, Hamdan FF, Rio M, Capo-Chichi JM, Bodaert N, Decarie JC, et al. Mutations in *DOCK7* in individuals with epileptic encephalopathy and cortical blindness. *Am J Hum Genet*. 2014;94(6):891–7.
70. Borcel E, Palczynska M, Krzisch M, Dimitrov M, Ulrich G, Toni N, et al. Shedding of neurexin 3beta ectodomain by ADAM10 releases a soluble fragment that affects the development of newborn neurons. *Sci Rep*. 2016;6:39310.
71. Nguyen TM, Schreiner D, Xiao L, Traunmuller L, Bornmann C, Scheiffele P. An alternative splicing switch shapes neurexin repertoires in principal neurons versus interneurons in the mouse hippocampus. *Elife*. 2016;5, e22757.
72. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, et al. *CNTNAP2* and *NRXN1* are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in *Drosophila*. *Am J Hum Genet*. 2009;85(5):655–66.
73. Watanabe H, Koopmann TT, Le Scouarnec S, Yang T, Ingram CR, Schott JJ, et al. Sodium channel beta1 subunit mutations associated with Brugada

- syndrome and cardiac conduction disease in humans. *J Clin Invest*. 2008; 118(6):2260–8.
74. Duarri A, Jezierska J, Fokkens M, Meijer M, Schelhaas HJ, den Dunnen WF, et al. Mutations in potassium channel *KCND3* cause spinocerebellar ataxia type 19. *Ann Neurol*. 2012;72(6):870–80.
  75. Jobling RK, Assoum M, Gakh O, Blaser S, Raiman JA, Mignot C, et al. *PMPCA* mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. *Brain*. 2015;138(Pt 6):1505–17.
  76. Thiffault I, Wolf NI, Forget D, Guerrero K, Tran LT, Choquet K, et al. Recessive mutations in *POLR1C* cause a leukodystrophy by impairing biogenesis of RNA polymerase III. *Nat Commun*. 2015;6:7623.
  77. Dauwerse JG, Dixon J, Seland S, Ruivenkamp CA, van Haeringen A, Hoefsloot LH, et al. Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome. *Nat Genet*. 2011; 43(1):20–2.
  78. Sarquella-Brugada G, Campuzano O, Arbelo E, Brugada J, Brugada R. Brugada syndrome: clinical and genetic findings. *Genet Med*. 2016;18(1):3–12.
  79. Campbell IM, Shaw CA, Stankiewicz P, Lupski JR. Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet*. 2015;31(7):382–92.
  80. Boycott KM, Innes AM. When one diagnosis is not enough. *N Engl J Med*. 2017;376:83–5.
  81. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet Med*. 2016;18(7):678–85.
  82. Kalari KR, Casavant M, Bair TB, Keen HL, Cameron JM, Casavant TL, et al. First exons and introns—a survey of GC content and gene structure in the human genome. *In Silico Biol*. 2006;6(3):237–42.
  83. Al-Sayed MD, Al-Zaidan H, Albakheet A, Hakami H, Kenana R, Al-Yafee Y, et al. Mutations in *NALCN* cause an autosomal-recessive syndrome with severe hypotonia, speech impairment, and cognitive delay. *Am J Hum Genet*. 2013;93(4):721–6.
  84. Fiskerstrand T, Arshad N, Haukanes BI, Tronstad RR, Pham KD, Johansson S, et al. Familial diarrhea syndrome caused by an activating *GUCY2C* mutation. *N Engl J Med*. 2012;366(17):1586–95.
  85. Muller T, Rasool I, Heinz-Erian P, Mildnerberger E, Hulstrunk C, Muller A, et al. Congenital secretory diarrhoea caused by activating germline mutations in *GUCY2C*. *Gut*. 2016;65(8):1306–13.
  86. Romi H, Cohen I, Landau D, Alkrinawi S, Yerushalmi B, Hershkovitz R, et al. Meconium ileus caused by mutations in *GUCY2C*, encoding the CFTR-activating guanylate cyclase 2C. *Am J Hum Genet*. 2012;90(5):893–9.
  87. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S, et al. Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am J Hum Genet*. 2012;90(1):133–41.
  88. Brown N, Burgess T, Forbes R, McGillivray G, Kornberg A, Mandelstam S, et al. 5q31.3 Microdeletion syndrome: clinical and molecular characterization of two further cases. *Am J Med Genet A*. 2013;161A(10):2604–8.
  89. Shimojima K, Isidor B, Le Caignec C, Kondo A, Sakata S, Ohno K, et al. A new microdeletion syndrome of 5q31.3 characterized by severe developmental delays, distinctive facial features, and delayed myelination. *Am J Med Genet A*. 2011;155A(4):732–6.
  90. Molin AM, Andrieux J, Koolen DA, Malan V, Carella M, Colleaux L, et al. A novel microdeletion syndrome at 3q13.31 characterised by developmental delay, postnatal overgrowth, hypoplastic male genitals, and characteristic facial features. *J Med Genet*. 2012;49(2):104–9.
  91. Coutelier M, Burglen L, Mundwiller E, Abada-Bendib M, Rodriguez D, Chantot-Bastaraud S, et al. *GRID2* mutations span from congenital to mild adult-onset cerebellar ataxia. *Neurology*. 2015;84(17):1751–9.
  92. Hills LB, Masri A, Konno K, Kakegawa W, Lam AT, Lim-Melia E, et al. Deletions in *GRID2* lead to a recessive syndrome of cerebellar ataxia and tonic upgaze in humans. *Neurology*. 2013;81(16):1378–86.
  93. Van Schil K, Meire F, Karlstetter M, Bauwens M, Verdin H, Coppieters F, et al. Early-onset autosomal recessive cerebellar ataxia associated with retinal dystrophy: new human hotfoot phenotype caused by homozygous *GRID2* deletion. *Genet Med*. 2015;17(4):291–9.
  94. Maier A, Klopocki E, Horn D, Tzschach A, Holm T, Meyer R, et al. De novo partial deletion in *GRID2* presenting with complicated spastic paraplegia. *Muscle Nerve*. 2014;49(2):289–92.
  95. Charng WL, Karaca E, Coban Akdemir Z, Gambin T, Atik MM, Gu S, et al. Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Med Genomics*. 2016;9(1):42.
  96. Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat Genet*. 1997;15(1):62–9.
  97. Damaj L, Lupien-Meilleur A, Lortie A, Riou E, Ospina LH, Gagnon L, et al. *CACNA1A* haploinsufficiency causes cognitive impairment, autism and epileptic encephalopathy with mild cerebellar symptoms. *Eur J Hum Genet*. 2015;23(11):1505–12.
  98. Inoue K, Khajavi M, Ohyama T, Hirabayashi S, Wilson J, Reggin JD, Mancias P, Butler JJ, Wilkinson MF, Wegner M, et al. Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat Genet*. 2004;36(4):361–69.
  99. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. *Cell*. 2011;147(1):32–43.
  100. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. *Nat Genet*. 2016; 48(2):126–33.
  101. Veltman JA, Brunner HG. *De novo* mutations in human genetic disease. *Nat Rev Genet*. 2012;13(8):565–75.
  102. Shroyer NF, Lewis RA, Yatsenko AN, Wensel TG, Lupski JR. Cosegregation and functional analysis of mutant *ABCR* (*ABCA4*) alleles in families that manifest both Stargardt disease and age-related macular degeneration. *Hum Mol Genet*. 2001;10(23):2671–8.
  103. Allikmets R. Further evidence for an association of *ABCR* alleles with age-related macular degeneration. The International ABCR Screening Consortium. *Am J Hum Genet*. 2000;67(2):487–91.
  104. Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, et al. Mutation of the Stargardt disease gene (*ABCR*) in age-related macular degeneration. *Science*. 1997;277(5333):1805–7.
  105. Bernstein PS, Leppert M, Singh N, Dean M, Lewis RA, Lupski JR, et al. Genotype-phenotype analysis of *ABCR* variants in macular degeneration probands and siblings. *Invest Ophthalmol Vis Sci*. 2002;43(2):466–73.
  106. Cohn JA, Friedman KJ, Noone PG, Knowles MR, Silverman LM, Jowell PS. Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. *N Engl J Med*. 1998;339(10):653–8.
  107. Goldstein JL, Brown MS. Familial hypercholesterolemia: identification of a defect in the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity associated with overproduction of cholesterol. *Proc Natl Acad Sci U S A*. 1973;70(10):2804–8.
  108. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344–7.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

