Rapid Paediatric Sequencing (RaPS): comprehensive real-life workflow for rapid diagnosis of critically ill children

Lamia Mestek-Boukhibar¹, Emma Clement², Wendy Jones², Suzanne Drury³, Louise Ocaka¹, Andrey Gagunashvili¹, Polona Le Quesne Stabej¹, Chiara Bacchelli¹, Nital Jani¹, Shamima Rahman⁴, Lucy Jenkins⁵, Jane Hurst², Maria Bitner-Glindzicz⁴, Mark Peters⁶, Philip L Beales¹ & Hywel J Williams¹

Supplementary Information

¹GOSgene, Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom
²Clinical Genetics and Genomic Medicine, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom
³Congenica Ltd, Bioinnovation Data Centre, Wellcome Genome Campus, Cambridge, United Kingdom
⁴Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom
⁵NE Thames Regional Genetics Laboratory, Great Ormond Street Hospital, London, United Kingdom
⁶UCL Great Ormond Street Institute of Child Health and Great Ormond Street NHS Foundation Trust, London, United Kingdom
# Table of Contents

**Supplementary text S1**  Whole genome sequencing ........................................ 3
**Supplementary text S2**  Variant analysis .......................................................... 4
**Supplementary text S3**  Sequencing analysis ..................................................... 6

References .............................................................................................................. 7

**Figure S1**  Specialties involved in care of probands recruited to RaPS ............... 8
**Figure S2**  Age of probands recruited to RaPS at the time of whole genome sequencing 9
**Figure S3**  RaPS referrals .................................................................................. 10
**Figure S4**  Variant breakdown overview .............................................................. 11
**Figure S5**  Identification and validation of CC2D2A inversion ............................. 12

**Table S1**  Patients’ phenotypes at the time of referral ......................................... 13
**Table S2**  Number of variants per trio ................................................................... 16
**Table S3**  Sample mean coverage and percentage of whole genome covered at the indicated read depth in both parents and proband .................................................. 18
**Table S4**  Gene variants shortlisted in each patient ............................................. 20
Supplementary Text S1.

Whole genome sequencing

Genomic DNA (gDNA) preparation: Whole genome sequencing was performed on a trio (father, mother and proband). gDNA was extracted from blood sample using Chemagic-STAR (Hamilton, USA) in a diagnostic accredited lab (NE Thames Regional Genetics Lab). High quality gDNA was used for whole genome library preparation. One µl of gDNA was run on 1% agarose gel to confirm absence of degradation. gDNA concentration was measured using Qubit dsDNA Broad Range Assay Kit (Thermo Fisher Scientific, Waltham, MA). DNA was diluted to 1.1 µg in total volume of 55 µL in HT1 buffer (Illumina, USA) and transferred to Covaris 50 µL individual tubes (Woburn, MA). gDNA was sheared to 350 bp using E220 Focused-ultrasonicator (Woburn, MA) for 60 seconds with the following parameters: target peak BP 400, peak incident power 140, duty factor 10% and 22 cycles of burst. Successful shearing was assessed on 1% agarose gel prior to starting library preparation.

Library preparation: Whole genome gDNA libraries were prepared using TruSeq DNA PCR-Free Library Prep (Illumina, USA) following manufacturer advice starting with 1 µg of sheared gDNA (in 50 µL). Libraries were single indexed using Illumina’s indexed adapters (Set A FC-121-3001 or Set B FC-121-3002, Illumina, USA). Library concentration was measured using by quantitative Polymerase Chain Reaction (qPCR) (KAPA Biosystems, Roche, Basel, Switzerland) following manufacturer’s advice. Briefly, 2 µL of library was diluted 10 000× and 20 000× in dilution buffer (100 µL Tween 20, 2 mL 1 M Tris and 198 mL dH₂O) and incubated overnight. qPCR was performed in triplicates in a total volume of 16 µL each and run on Applied Biosystems 7300 qPCR machine (Thermo Fisher Scientific, Waltham, MA).

Library normalization and sequencing: Libraries were normalised to 2 nM with Tris·HCl (10 mM pH 8.5) supplemented with 0.1% Tween 20. Libraries were denatured with 0.2 N NaOH and stabilised with 200 mM Tris·HCl. gDNA libraries from parents were pooled and 2.7 pM was sequenced on Illumina NextSeq 550 System with 2×150 bp for 29 hours. gDNA library from proband was sequenced on double flow cell on Illumina HiSeq 2500 System (Rapid Run Mode) with 9 pM loading concentration and 2×150 bp for 30 hours. For the last two trios (RaPS_23 and PaPS_24), proband samples were sequenced on NextSeq 550 System.
Supplementary Text S2.

Variant analysis

Variant analysis and filtering was performed using Ingenuity Variant Analysis (IVA) software (QIAGEN; [https://variants.ingenuity.com/](https://variants.ingenuity.com/)).

Summary of variant filtering and shortlisting: VCF files for each trio were uploaded to IVA and sample relationships were annotated using a ped file (father, mother and proband) where unaffected parents were used as controls and proband as a case.

The number of variants as shown in Supplementary Figure S4 is the total number of variants per trio to which pre-filters were then applied: first, common variants were filtered out by excluding those that are present at minor allele frequency (MAF) >0.5% in the public databases as unlikely to be causative; second, a filter was applied to only keep the variants that were predicted to effect protein function (missense variants, stop gains or losses, frameshifts, small insertions and deletions (indels) as well as variants affecting splicing sites present at ±7 nucleotides from exon-intron junctions).

Variant short listing was performed in Phases (I, II and III) as described in the main text Methods section. In Phase I analysis, variants that fell within genes that are part of Phase I gene list were kept the rest were excluded. These variants were then sorted based on mode of inheritance by applying a genetic filter in IVA to keep only variants that are present in the proband. The same principle was applied to Phase II and Phase III analysis. By applying this cascade of filters, large amount of variants were excluded to only focus on those that are relevant to patient and clinical manifestation.

Integrated Genome Viewer (IGV) was used in parallel to supplement the variant shortlisting. For each shortlisted variant (an average of 7–10 per trio) the genomic location was viewed in IGV to determine the quality of the read.

1. Pre-filtering step:
   a. Common variants with MAF >0.5% in 1000G (1000 Genomes Project Consortium et al., 2015), ExAC (Lek et al., 2016) and Exome Variant Server ([evs.gs.washington.edu/EVS/](http://evs.gs.washington.edu/EVS/)) databases were filtered out. For homozygous and hemizygous variants in proband, MAF is increased to ≤10% and variants with no homozygotes/hemizygotes in ExAC were investigated.
   b. Effect of variant on protein function was set to include predicted pathogenic, likely pathogenic and uncertain significance (benign or likely benign variants are investigated if further evidence of pathogenicity was available).
   c. Variants associated with loss of function were kept: frameshift causing, in-frame insertions/deletions, missense and splice site (±7 nucleotides).
   d. 5’ UTR and 3’ UTR variants were also investigated for genes known to be disease-causing or with compelling evidence for candidate genes.

2. Phase I analysis: This comprised setting a gene panel as a filter to investigate genes associated with the patients reported phenotypes as the first line of investigation. The gene panel was constructed by converting clinical phenotypes to HPO terms retrieving associated genes from different sources: The Genomics England PanelApp ([https://panelapp.](https://panelapp.))
genomicsengland.co.uk/), Phenotips (https://phenotips.org/), OMIM Gene Map (https://omim.org/search/advanced/geneMap), established panels from NE Thames Regional Genetics Lab, Great Ormond street Institute of Child Health-UCL experts and literature search in PubMed.

3. Phase II analysis workflow: This consisted of variants in disease-associated genes from OMIM (https://omim.org/search/advanced/geneMap) and DDG2P (Firth et al., 2009) databases.

4. Phase III analysis workflow: This is performed when Phase I and Phase II variants are exhausted and or when a variant is interesting from a research point of view and consists of variants that pass the following set of criteria: a. Effect on protein function is likely damaging b. Supportive research based evidence from the literature with link to patient phenotypes

Mode of inheritance: Here the “mode of inheritance” (MOI) term is used to refer to the inheritance of the variant in the proband and not a disease. MOI are prioritized based on parents consanguinity and family history to cover: autosomal recessive (homozygous in proband and heterozygous in parents), compound heterozygous, de novo, X-linked dominant (XLD) and X-linked recessive (XLR) to include hemizygous in male patients. The rapid mode of sequencing often does not yield read depth that permits detection of mosaicism, however, if suspected then Sanger sequencing is performed. Specifically, in the scenario where we had a recessive gene in which only one potentially pathogenic SNV variant was identified from the VCF file, the entire gene was inspected using IGV to search for a potential second variant that was not called using our standard analysis pipeline (such as possible deletions, translocations, or inversions). Detecting structural variants on IGV requires a trained professional, and a full IGV manual is freely accessible: http://software.broadinstitute.org/software/igv/UserGuide. Briefly, we searched for an obvious drop in coverage as an indication of deletion, and we inspected the orientation of the paired-reads which are colour coded (whole genome sequencing was perfumed as paired-end sequencing as indicated in our methods section), and so reads face each other. Paired-reads that have same orientation appear in different colour and are an indication of inversion which can be verified by Sanger sequencing.
Supplementary Text S3.

Sequencing analysis

Read mapping and variant calling: Basecalling of raw sequencing reads was performed on BaseSpace Sequence hub (http://basespace.illumina.com). Fastq files for each individual were downloaded from BaseSpace, and reads from different lanes were merged together. Mapping and variant calling were performed using a Genalice appliance running Genalice Map 2.5.5 including Mapping, Variant Calling and the Population Calling modules for trio analysis (Genalice Core BV, Netherlands). Human genome build 37 (GRCh37, hg19) and Genalice default configuration files were used for WGS mapping, and trio variant detection. Aligned reads were stored in the GAR format (Genalice Aligned Reads), using less than 5 GB per sample. Variants were stored in a GVM (Genalice Variant Map) per trio, using less than 200 MB per sample. A standard multi sample VCF with Mendelian inheritance annotation using Context Based Call Enhancement was extracted from each GVM.

Processing speed comparison: To compare the processing speed between using a Genalice appliance and open source software we calculated the time taken to analyse a randomly chosen trio from our study. The Genalice analysis is as described above and for the comparison we used BWA-MEM (Li and Durbin, 2010) for read mapping and GATK (DePristo et al., 2011; McKenna et al., 2010) for variant calling with the analysis run on our in-house UCL high performance computing cluster.

Variant calling accuracy comparison: To explore whether the quality of the variants called by Genalice and open source software were comparable we utilised the reference DNA sample NA12878 (http://jimb.stanford.edu/giab/). We prepared this sample for sequencing using the exact same protocol as that used for all our RaPS trios and sequenced it under the same conditions as we did for proband samples. The fastq files were then processed under the conditions detailed above and the resultant VCF file was processed to calculate the SNP-precision, SNP-recall and SNP-Fscore (Zook et al., 2014). High confidence regions values obtained are indicated below:

<table>
<thead>
<tr>
<th>True-positive</th>
<th>False-positive</th>
<th>False-negative</th>
<th>SNP-precision</th>
<th>SNP-recall</th>
<th>SNP-Fscore</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,048,709</td>
<td>49,430</td>
<td>103,717</td>
<td>0.984</td>
<td>0.9671</td>
<td>0.9755</td>
</tr>
</tbody>
</table>
References


Figure S1. Specialties involved in care of probands recruited to RaPS
Figure S2. Age of probands recruited to RaPS at the time of whole genome sequencing
Figure S3. RaPS referrals. The figure depicts the number of individuals referred to RaPS and whether they have passed the inclusion criteria or not. The accompanying table lists the reasons for exclusion for individuals who were not included in RaPS.
Figure S4. Variant breakdown overview. Schematic breakdown of number of variants in 24 trios in different steps of the RaPS workflow. In Phase I, the number of genes filtered is different in each trio. Phase II is composed of genes from GGD2P (1.643 genes) and OMIM Morbid Genes (7.737 genes) databases and is standard across all trios. The RaPS variant analysis workflow is phased to prioritise likely causative gene variants. Note that compound heterozygous and homozygous variants are counted once. Refer to Supplementary text S2 for detailed description of variant filtering and shortlisting.
Figure S5. Identification and validation of CC2D2A inversion. A Schematic diagram depicting the reference configuration of part of CC2D2A gene (only exons 8–10 are shown for simplicity) and an inversion spanning exons 8 and 9. B Mother and proband are heterozygous carriers of the inversion and both primer pairs iF/F and iF/R generate a PCR product of 2051 bp and 2444 bp, respectively. Father is not a carrier and only primer pair iF/R generates a PCR product of 2444 bp. Sequencing of the iF/F PCR product in both mother and proband identified the exact location of break points (Chr4:15510661–15514794; hg19). C IGV screenshot showing the discordant reads (coloured sea green and blue) in both mother and proband.
<table>
<thead>
<tr>
<th>Patient</th>
<th>Age at time of whole genome sequencing (sex)</th>
<th>Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>1 year 7 months 15 days (F)</td>
<td>Absent B cells, multiple septic episodes, disseminated infection with multiorgan involvement (lungs, brain, liver, gut), granulomatous lower back collection, tubular leak with occasional hypernatremia, hypertension, bcytopoenia (low platelets and Hb), generalised oedema, capillary leak, fragile skin</td>
</tr>
<tr>
<td>Patient 2</td>
<td>12 years 5 months 1 days (F)</td>
<td>Easy bruising, failure to thrive, splenic rupture, blue sclerae, joint hypermobility</td>
</tr>
<tr>
<td>Patient 3</td>
<td>0 years 5 months 18 days (F)</td>
<td>Talipes, fasciculation, abnormal dopamine turnover, abnormality of the brain, small cerebellum, simplified gyral pattern, contractures, low set ears, alveolar cleft gums, stiffness, small nasal cleft, polyhydramnios</td>
</tr>
<tr>
<td>Patient 4</td>
<td>0 years 7 months 28 days (M)</td>
<td>Craniofacial abnormalities, abnormality of the limbs, talipes, micrognathia, camptodactyly, short halluces</td>
</tr>
<tr>
<td>Patient 5</td>
<td>0 years 1 month 30 days (M)</td>
<td>Severe combined immunodeficiency (thymic), cerebellar dysplasia, delayed cortical development, visual impairment, cleft lip/palate, craniosynostosis, hypoparathyroidism</td>
</tr>
<tr>
<td>Patient 6</td>
<td>0 years 2 months 18 days (F)</td>
<td>High platelet count, hypoplastic aortic arch, double outlet right ventricle, bicuspid pulmonary valve, ventricular septal defects, severe combined immunodeficiency (thymic), cerebellar dysplasia, delayed cortical development, visual impairment, cleft lip/palate, craniosynostosis, hypoparathyroidism, radial aplasia, absent spleen, micrognathia, coarctation of the aorta, tracheobronchomalacia, gastro-oesophageal reflux disease, doubly committed ventricular septal defect (VSD) with additional small apical VSDs, left anterior descending artery from right coronary artery, periauricular pits, long fingers and toes, triphalangeal thumbs, hypermetropic astigmatism, intermittent exotropia</td>
</tr>
<tr>
<td>Patient 7</td>
<td>0 years 6 months 9 days (F)</td>
<td>Global developmental delay, seizures, cerebral haemorrhage, visual impairment, bilateral hydronephrosis, small patent foramen ovale, anaemia, recurrent upper tract infections, fungal upper tract infection, ectopic ureters, abnormal lymphocyte count (low) but normal proportion of T-cells, B-cells and NK-cells</td>
</tr>
<tr>
<td>Patient 8</td>
<td>2 years 7 months 16 days (M)</td>
<td>Multiple serositis, severe hypotonia (trunk and lower limbs), mild calcineurin inhibitor related microangiopathy, fragmentocytes, thrombocytopenia, graft versus host disease of the gut, chronic inflammatory, gut epithelial damage, microangiopathy, multiple sclerosis, severe combined immunodeficiency, immune cell abnormalities, neutropenia, viral infection, fluid retention, capillary leak, blood pressure, hypoalbuminaemia</td>
</tr>
<tr>
<td>Patient 9</td>
<td>1 year 9 months 14 days (F)</td>
<td>Pulmonary hypertension, alveolar hypoplasia, upper respiratory tract infection, anaemia, renal failure, thrombocytopenia, status epilepticus, encephalitis, metabolic acidosis respiratory failure, impaired cardiac function, atypical hemolytic-uremic syndrome</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Patient</th>
<th>Age at time of whole genome sequencing (sex)</th>
<th>Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 10</td>
<td>5 years 1 month 13 days (M)</td>
<td>Recurrent upper respiratory tract infections, tonsillitis, hypertriglyceridemia, hematochezia, splenomegaly, thrombocytopenia, neutropenia, anaemia due to reduced life span of red cells, colitis, increased serum ferritin, fever, lymphadenopathy, retinal vein occlusion, Epstein-Barr virus-induced hemophagocytic lymphohistiocytosis</td>
</tr>
<tr>
<td>Patient 11</td>
<td>0 years 1 month 2 days (M)</td>
<td>Acute oligonuric kidney injury, ossification defect of the skull</td>
</tr>
<tr>
<td>Patient 12</td>
<td>0 years 0 months 16 days (M)</td>
<td>Nonketotic hyperglycinaemia, continuing severe encephalopathy</td>
</tr>
<tr>
<td>Patient 13</td>
<td>0 years 10 months 16 days (F)</td>
<td>Primary immune deficiency, atypical haemolytic uremic syndrome</td>
</tr>
<tr>
<td>Patient 14</td>
<td>2 years 4 months 11 days (F)</td>
<td>Intrauterine growth retardation, failure to thrive, skeletal dysplasia, lung dysplasia, mild hypotonia resulting in delayed swallow and nocturnal hypoventilation, small ventricula septal defect, small subaortic shelf, suspected primary immunodeficiency, Epstein-Barr virus-driven high grade lymphoma, iliac crest serration, lacy pelvis, coarse facial features, bulbar palsy, platyspondyly, coronal notching at the lumbar region, thoraco lumbar kyphosis</td>
</tr>
<tr>
<td>Patient 15</td>
<td>0 years 3 months 7 days (M)</td>
<td>Profound bilateral sensorineural hearing loss, hypotonia, failure to thrive despite good feeding, metabolic acidosis, lactic acidosis</td>
</tr>
<tr>
<td>Patient 16</td>
<td>0 years 2 months 1 day (M)</td>
<td>Cardiac abnormalities, small patent ductus arteriosus, neurological problems, right duplex kidney, intracerebral cyst, soft dysmorphic features, adducted thumbs, slanting eyes, right ventricular hypertrophy, hyperinsulinemic hypoglycemia, multiple thrombi including large inferior vena cava clot and cerebral infarction, cystic encephalomalacia, deranged liver function, ontractures</td>
</tr>
<tr>
<td>Patient 17</td>
<td>0 years 2 months 24 days (M)</td>
<td>Antenatal hydrops fetalis, pancytopenia, hepatosplenomegaly, thrombocytopenia, very high ferritin, skin rash, XIAP deficiency, neonatal hypothyroidism, hypocholaemic metabolic acidosis, histiocytes on bone marrow biopsy, increased circulating low-density lipoprotein levels, high triglycerides, low HDL cholesterol levels, abnormal liver function tests, congenital dislocation of hip, inguinal hernia, rhizomelic short limbs, blue sclerae, short stature</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Patient</th>
<th>Age at time of whole genome sequencing (sex)</th>
<th>Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 18</td>
<td>0 years 0 months 27 days (M)</td>
<td>Bilateral cleft lip/palate, horseshoe kidney, ambiguous genitalia, hypospadias, undescended testes, hypoplastic aortic arch, patent ductus arteriosus, atrial septal defect, abnormal myocardium, coloboma (optic disk and iris), dilation of intrahepatic and extrahepatic bile ducts, right pleural effusion, microcephaly, hypertelorism, broad nasal bridge, exorbitism, arched eyebrows, low set ears, hypertrichosis, short neck with redundant skin at the nape, rhizomelic shortening of the upper limbs, clinodactyly, syndactyly, thickened skin over fingers, prominent anterior plantar pad, overlapping toes, short penis, scrotal oedema, failure to thrive, immune deficiency, coarctation of aorta, cystic, renal dysplasia, hearing impairment, vertebral abnormalities (butterfly vertebrae at T2), choledochal cyst in liver, deranged thyroid functional tests (lowT3)</td>
</tr>
<tr>
<td>Patient 19</td>
<td>0 years 2 months 5 days (F)</td>
<td>Heart defects, atrial septal defect, patent ductus arteriosus, mild aortic stenosis, dysplastic multicystic kidneys, hypertrichosis, natal tooth, small jaw, inverted nipples, restriction of elbow extension, conjugated bilirubinaemia</td>
</tr>
<tr>
<td>Patient 20</td>
<td>0 years 8 months 26 days (M)</td>
<td>Congenital heart defects (ventricular septal defect and double outlet right ventricle), hypoplastic corpus callosum, small pons, small cerebellum, severe tracheobronchomalacia, diaphragmatic eventration, heart block, inguinal hernia, hyperinsulinism, abnormal fat distribution</td>
</tr>
<tr>
<td>Patient 21</td>
<td>0 years 1 months 25 days (M)</td>
<td>Ambiguous genitalia, bifid scrotum hypocalcaemia, hypotension, hypoglycaemia, patent ductus arteriosus, hyperbilirubinemia, microcephaly, thin ribs, short limbs, brachydactyly, rocker, bottom feet, recurrent infection, short toes, small phallicus, depressed nasal bridge</td>
</tr>
<tr>
<td>Patient 22</td>
<td>0 years 3 months 3 days (M)</td>
<td>Congenital cataract, small pupils, dilated cardiomyopathy, lactic acidosis</td>
</tr>
<tr>
<td>Patient 23</td>
<td>0 years 1 month 6 days (M)</td>
<td>Intrauterine growth retardation, neurogenic arthrogryposis, hypotonia, scoliosis, anterior horn cell pathology</td>
</tr>
<tr>
<td>Patient 24</td>
<td>0 years 0 months 17 days (F)</td>
<td>Prenatal hydrocephalus, dystroglycanopathy, post axial polydactyly of hands, abnormal cerebellar vermis, respiratory distress</td>
</tr>
</tbody>
</table>

Human Phenotype Ontology terms were derived for each patient using the phenotype information shown here and used to construct Phase I gene panels.
Table S2. Number of variants per trio (frameshifts, in-frame indels, start/stop codon changes, missense variants)

<table>
<thead>
<tr>
<th>RaPS trio ID</th>
<th>Whole genome</th>
<th>Coding regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of variants per trio</td>
<td>Number of variants passed pre-filter per trio</td>
</tr>
<tr>
<td>RaPS_01</td>
<td>6,047,454</td>
<td>42,573</td>
</tr>
<tr>
<td>RaPS_02</td>
<td>5,489,792</td>
<td>40,354</td>
</tr>
<tr>
<td>RaPS_03</td>
<td>6,209,676</td>
<td>45,000</td>
</tr>
<tr>
<td>RaPS_04</td>
<td>6,962,558</td>
<td>48,977</td>
</tr>
<tr>
<td>RaPS_05</td>
<td>5,954,527</td>
<td>41,700</td>
</tr>
<tr>
<td>RaPS_06</td>
<td>6,582,174</td>
<td>40,678</td>
</tr>
<tr>
<td>RaPS_07</td>
<td>5,810,316</td>
<td>41,512</td>
</tr>
<tr>
<td>RaPS_08</td>
<td>6,137,998</td>
<td>43,474</td>
</tr>
<tr>
<td>RaPS_09</td>
<td>6,071,727</td>
<td>43,206</td>
</tr>
<tr>
<td>RaPS_10</td>
<td>5,841,339</td>
<td>41,891</td>
</tr>
<tr>
<td>RaPS_11</td>
<td>5,826,869</td>
<td>42,147</td>
</tr>
<tr>
<td>RaPS_12</td>
<td>5,475,165</td>
<td>40,466</td>
</tr>
<tr>
<td>RaPS_13</td>
<td>6,000,484</td>
<td>41,574</td>
</tr>
<tr>
<td>RaPS_14</td>
<td>6,066,300</td>
<td>42,794</td>
</tr>
<tr>
<td>RaPS_15</td>
<td>6,051,267</td>
<td>42,536</td>
</tr>
<tr>
<td>RaPS_16</td>
<td>5,991,679</td>
<td>42,633</td>
</tr>
<tr>
<td>RaPS_17</td>
<td>5,033,594</td>
<td>34,458</td>
</tr>
<tr>
<td>RaPS_18</td>
<td>5,061,461</td>
<td>33,915</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>RaPS trio ID</th>
<th>Whole genome</th>
<th>Coding regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of variants per trio</td>
<td>Number of variants passed pre-filter per trio</td>
</tr>
<tr>
<td>RaPS_19</td>
<td>5,581,553</td>
<td>37,792</td>
</tr>
<tr>
<td>RaPS_20</td>
<td>6,008,553</td>
<td>43,131</td>
</tr>
<tr>
<td>RaPS_21</td>
<td>5,334,898</td>
<td>39,662</td>
</tr>
<tr>
<td>RaPS_22</td>
<td>4,658,755</td>
<td>31,974</td>
</tr>
<tr>
<td>RaPS_23</td>
<td>5,859,596</td>
<td>48,115</td>
</tr>
<tr>
<td>RaPS_24</td>
<td>5,870,780</td>
<td>42,891</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>139,928,515</td>
<td>993,453</td>
</tr>
<tr>
<td><strong>Average/trio</strong></td>
<td>5,830,355</td>
<td>41,394</td>
</tr>
</tbody>
</table>

Table shows the number of variants per trio analysis at different stages of the variant selection workflow. Data was filtered at MAF ≤10%, effect of variant: missense, start/stop codon change, frameshifts, in-frame indels and splice site loss ± 7 nucleotides. MAF was later adjusted according to zygosity, it was decreased to ≤0.5% in de novo and heterozygous variants, and kept at ≤10% in homozygous, compound heterozygous and X-linked recessive variants in which case only variants with no reported homozygous/hemizygous were considered. Shortlisting of variants was based on effect on protein function, zygosity and association with patient phenotypes. Phase I filter consisted of a phenotype-specific gene list and was generated through Genomics England PanelApp, Phenotips, OMIM, established panels, NE Thames Regional Genetics Laboratory, GOSICH experts and literature search. The number of variants that passed Phase I gene filter are per trio. The parental genotypes were used to infer mode of inheritance in the patients and informed variants shortlisting. Shortlisted variants are thus per patient.

<sup>a</sup>Homozygous and compound heterozygous variants are counted once.


<sup>c</sup>Candidate gene identified in Phase I.

<sup>d</sup>5’ UTR expansion in EIF4A3 Phase I candidate gene was not picked up by our WGS analysis.
Table S3. Sample mean coverage and percentage of whole genome covered at the indicated read depth in both parents and proband

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Mean coverage</th>
<th>% bases above 10×</th>
<th>% bases above 15×</th>
<th>% bases above 20×</th>
<th>% bases above 30×</th>
</tr>
</thead>
<tbody>
<tr>
<td>RaPS_1 Father</td>
<td>11.3</td>
<td>65.9</td>
<td>22.3</td>
<td>3.2</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_1 Mother</td>
<td>9.7</td>
<td>50.9</td>
<td>10.5</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_1 Proband</td>
<td>16.4</td>
<td>90.4</td>
<td>62.9</td>
<td>27.1</td>
<td>1.4</td>
</tr>
<tr>
<td>RaPS_2 Father</td>
<td>29.6</td>
<td>96.3</td>
<td>92.6</td>
<td>86.4</td>
<td>55.1</td>
</tr>
<tr>
<td>RaPS_2 Mother</td>
<td>9.3</td>
<td>46.4</td>
<td>8.5</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_2 Proband</td>
<td>26.8</td>
<td>95.4</td>
<td>91.7</td>
<td>82.8</td>
<td>38.6</td>
</tr>
<tr>
<td>RaPS_3 Father</td>
<td>13.8</td>
<td>80.4</td>
<td>44.9</td>
<td>12.6</td>
<td>0.3</td>
</tr>
<tr>
<td>RaPS_3 Mother</td>
<td>14.3</td>
<td>84.4</td>
<td>48.9</td>
<td>14.1</td>
<td>0.3</td>
</tr>
<tr>
<td>RaPS_3 Proband</td>
<td>32.5</td>
<td>96.6</td>
<td>95</td>
<td>91.7</td>
<td>67.6</td>
</tr>
<tr>
<td>RaPS_4 Father</td>
<td>6.6</td>
<td>16.8</td>
<td>1.1</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>RaPS_4 Mother</td>
<td>8.1</td>
<td>32.7</td>
<td>3.7</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>RaPS_4 Proband</td>
<td>20.1</td>
<td>93.9</td>
<td>82.3</td>
<td>54.9</td>
<td>6</td>
</tr>
<tr>
<td>RaPS_5 Father</td>
<td>19.7</td>
<td>91</td>
<td>76.8</td>
<td>52</td>
<td>7.8</td>
</tr>
<tr>
<td>RaPS_5 Mother</td>
<td>23.2</td>
<td>94.2</td>
<td>86.8</td>
<td>69.8</td>
<td>20.3</td>
</tr>
<tr>
<td>RaPS_5 Proband</td>
<td>10.3</td>
<td>56.9</td>
<td>13.9</td>
<td>1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_6 Father</td>
<td>11.1</td>
<td>64.1</td>
<td>21.5</td>
<td>3.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_6 Mother</td>
<td>12.7</td>
<td>76.7</td>
<td>33.9</td>
<td>6.8</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_6 Proband</td>
<td>47.3</td>
<td>97</td>
<td>96.3</td>
<td>95.1</td>
<td>90.1</td>
</tr>
<tr>
<td>RaPS_7 Father</td>
<td>11.1</td>
<td>64.1</td>
<td>21.4</td>
<td>3.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_7 Mother</td>
<td>11.8</td>
<td>70.4</td>
<td>26.3</td>
<td>4.3</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_7 Proband</td>
<td>33.5</td>
<td>97.2</td>
<td>96.2</td>
<td>93.6</td>
<td>69.1</td>
</tr>
<tr>
<td>RaPS_8 Father</td>
<td>16.2</td>
<td>85.3</td>
<td>61.5</td>
<td>29.6</td>
<td>1.6</td>
</tr>
<tr>
<td>RaPS_8 Mother</td>
<td>19.6</td>
<td>91.9</td>
<td>78.9</td>
<td>51.8</td>
<td>6.4</td>
</tr>
<tr>
<td>RaPS_8 Proband</td>
<td>16.7</td>
<td>90</td>
<td>65.7</td>
<td>28.7</td>
<td>1</td>
</tr>
<tr>
<td>RaPS_9 Father</td>
<td>19.8</td>
<td>93.5</td>
<td>80.1</td>
<td>51.6</td>
<td>7.4</td>
</tr>
<tr>
<td>RaPS_9 Mother</td>
<td>13.1</td>
<td>76.1</td>
<td>37.9</td>
<td>9.9</td>
<td>0.2</td>
</tr>
<tr>
<td>RaPS_9 Proband</td>
<td>28.7</td>
<td>96.7</td>
<td>93.1</td>
<td>82.4</td>
<td>43.6</td>
</tr>
<tr>
<td>RaPS_10 Father</td>
<td>10.4</td>
<td>57.6</td>
<td>17.8</td>
<td>2.4</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_10 Mother</td>
<td>9.5</td>
<td>49.9</td>
<td>11.9</td>
<td>1.3</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_10 Proband</td>
<td>22.2</td>
<td>95.2</td>
<td>87</td>
<td>66.2</td>
<td>12.6</td>
</tr>
<tr>
<td>RaPS_11 Father</td>
<td>15.9</td>
<td>86.3</td>
<td>60.9</td>
<td>26.5</td>
<td>1.1</td>
</tr>
<tr>
<td>RaPS_11 Mother</td>
<td>10.3</td>
<td>56.6</td>
<td>14.9</td>
<td>1.9</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_11 Proband</td>
<td>9.6</td>
<td>48.7</td>
<td>10.4</td>
<td>1.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_12 Father</td>
<td>15.0</td>
<td>83.5</td>
<td>54.8</td>
<td>20.8</td>
<td>0.6</td>
</tr>
<tr>
<td>RaPS_12 Mother</td>
<td>11.8</td>
<td>69</td>
<td>27.9</td>
<td>5.4</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_12 Proband</td>
<td>33.6</td>
<td>96.5</td>
<td>93.7</td>
<td>88.8</td>
<td>69.5</td>
</tr>
<tr>
<td>RaPS_13 Father</td>
<td>10.8</td>
<td>60.4</td>
<td>20.5</td>
<td>3.3</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_13 Mother</td>
<td>21.4</td>
<td>93.7</td>
<td>83.8</td>
<td>61.5</td>
<td>12.3</td>
</tr>
<tr>
<td>RaPS_13 Proband</td>
<td>32.3</td>
<td>97.3</td>
<td>96.3</td>
<td>92.7</td>
<td>62.4</td>
</tr>
<tr>
<td>RaPS_14 Father</td>
<td>12.3</td>
<td>72.5</td>
<td>31.6</td>
<td>6.3</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_14 Mother</td>
<td>14.1</td>
<td>83</td>
<td>46.7</td>
<td>13.3</td>
<td>0.3</td>
</tr>
<tr>
<td>RaPS_14 Proband</td>
<td>22.0</td>
<td>96.2</td>
<td>89.1</td>
<td>66</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Continued on next page
Table S3. (continued from previous page)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Mean coverage</th>
<th>% bases above 10×</th>
<th>% bases above 15×</th>
<th>% bases above 20×</th>
<th>% bases above 30×</th>
</tr>
</thead>
<tbody>
<tr>
<td>RaPS_15 Father</td>
<td>13.3</td>
<td>78.2</td>
<td>40.4</td>
<td>10.2</td>
<td>0.2</td>
</tr>
<tr>
<td>RaPS_15 Mother</td>
<td>12.4</td>
<td>74.6</td>
<td>32.1</td>
<td>6.3</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_15 Proband</td>
<td>18.1</td>
<td>91.8</td>
<td>73.3</td>
<td>40</td>
<td>3.1</td>
</tr>
<tr>
<td>RaPS_16 Father</td>
<td>13.7</td>
<td>79.8</td>
<td>44.4</td>
<td>12.6</td>
<td>0.3</td>
</tr>
<tr>
<td>RaPS_16 Mother</td>
<td>14.9</td>
<td>85.6</td>
<td>53.9</td>
<td>18.4</td>
<td>0.5</td>
</tr>
<tr>
<td>RaPS_16 Proband</td>
<td>10.4</td>
<td>57</td>
<td>15.2</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_17 Father</td>
<td>6.1</td>
<td>14.1</td>
<td>0.9</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>RaPS_17 Mother</td>
<td>5.3</td>
<td>8.5</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RaPS_17 Proband</td>
<td>11.8</td>
<td>67.7</td>
<td>26.2</td>
<td>5.3</td>
<td>0.2</td>
</tr>
<tr>
<td>RaPS_18 Father</td>
<td>11.8</td>
<td>58.5</td>
<td>32.1</td>
<td>13.5</td>
<td>1</td>
</tr>
<tr>
<td>RaPS_18 Mother</td>
<td>9.0</td>
<td>42.6</td>
<td>15.5</td>
<td>3.7</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_18 Proband</td>
<td>21.3</td>
<td>94.7</td>
<td>85.1</td>
<td>61.6</td>
<td>9.5</td>
</tr>
<tr>
<td>RaPS_19 Father</td>
<td>16.9</td>
<td>83.1</td>
<td>62.3</td>
<td>36.3</td>
<td>4.7</td>
</tr>
<tr>
<td>RaPS_19 Mother</td>
<td>15.4</td>
<td>77.2</td>
<td>53</td>
<td>28.1</td>
<td>3.6</td>
</tr>
<tr>
<td>RaPS_19 Proband</td>
<td>16.5</td>
<td>91.8</td>
<td>65.2</td>
<td>26.3</td>
<td>0.8</td>
</tr>
<tr>
<td>RaPS_20 Father</td>
<td>15.5</td>
<td>84.8</td>
<td>58.1</td>
<td>23.8</td>
<td>0.8</td>
</tr>
<tr>
<td>RaPS_20 Mother</td>
<td>13.0</td>
<td>76.8</td>
<td>38.1</td>
<td>9.6</td>
<td>0.2</td>
</tr>
<tr>
<td>RaPS_20 Proband</td>
<td>15.5</td>
<td>87.3</td>
<td>57.2</td>
<td>20.5</td>
<td>0.6</td>
</tr>
<tr>
<td>RaPS_21 Father</td>
<td>18.7</td>
<td>90.9</td>
<td>72.6</td>
<td>43.5</td>
<td>5.4</td>
</tr>
<tr>
<td>RaPS_21 Mother</td>
<td>8.4</td>
<td>34.7</td>
<td>5.8</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_21 Proband</td>
<td>38.5</td>
<td>98.1</td>
<td>97.1</td>
<td>94.6</td>
<td>81.9</td>
</tr>
<tr>
<td>RaPS_22 Father</td>
<td>6.4</td>
<td>17.9</td>
<td>3</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_22 Mother</td>
<td>5.7</td>
<td>13</td>
<td>1.6</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_22 Proband</td>
<td>17.1</td>
<td>93</td>
<td>68.5</td>
<td>30.2</td>
<td>1.5</td>
</tr>
<tr>
<td>RaPS_23 Father</td>
<td>11.4</td>
<td>65.7</td>
<td>23.1</td>
<td>3.9</td>
<td>0.1</td>
</tr>
<tr>
<td>RaPS_23 Mother</td>
<td>14.0</td>
<td>82.2</td>
<td>44.5</td>
<td>13.4</td>
<td>0.3</td>
</tr>
<tr>
<td>RaPS_23 Proband</td>
<td>29.1</td>
<td>97.3</td>
<td>93.6</td>
<td>84.7</td>
<td>48.1</td>
</tr>
<tr>
<td>RaPS_24 Father</td>
<td>16.9</td>
<td>89.3</td>
<td>64.5</td>
<td>32.4</td>
<td>2.06</td>
</tr>
<tr>
<td>RaPS_24 Mother</td>
<td>14.2</td>
<td>82.2</td>
<td>46.2</td>
<td>15.7</td>
<td>0.5</td>
</tr>
<tr>
<td>RaPS_24 Proband</td>
<td>32.2</td>
<td>98.1</td>
<td>97.2</td>
<td>93.9</td>
<td>63.3</td>
</tr>
</tbody>
</table>

Read depth metrics were generated using GATK’s DepthOfCoverage tool with "--includeDeletions --countType COUNT_READS --minMappingQuality 20 --minBaseQuality 20 -interval_merging OVERLAPPING_ONLY". Mitochondrial DNA, unlocalized and unplaced contigs were excluded from the analysis.
### Table S4. Gene variants shortlisted in each patient

<table>
<thead>
<tr>
<th>Trio</th>
<th>Gene</th>
<th>cDNA variant</th>
<th>Protein variant</th>
<th>Phase</th>
<th>Genotype</th>
<th>Variant class</th>
<th>ClinVar accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>RaPS_01</td>
<td>POLE1</td>
<td>c.5912A&gt;G</td>
<td>p.N1971S</td>
<td>I</td>
<td>CH</td>
<td>4</td>
<td>SCV000778566</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c.5867A&gt;T</td>
<td>p.E1956V</td>
<td></td>
<td></td>
<td>4</td>
<td>SCV000778567</td>
</tr>
<tr>
<td>RaPS_02</td>
<td>COL3A1</td>
<td>c.2194G&gt;A</td>
<td>p.G732R</td>
<td>I</td>
<td>de novo</td>
<td>4</td>
<td>SCV000778568</td>
</tr>
<tr>
<td>RaPS_05</td>
<td>CHD7</td>
<td>c.3106C&gt;T</td>
<td>p.R1036*</td>
<td>I</td>
<td>de novo</td>
<td>5</td>
<td>SCV000778569</td>
</tr>
<tr>
<td>RaPS_08</td>
<td>IL2RG</td>
<td>c.665G&gt;C</td>
<td>p.R222R</td>
<td>I</td>
<td>XLR</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>RaPS_12</td>
<td>GLDC</td>
<td>c.2489C&gt;T</td>
<td>p.T830M</td>
<td>I</td>
<td>homozygous</td>
<td>4</td>
<td>SCV000778572</td>
</tr>
<tr>
<td>RaPS_16</td>
<td>NSD1</td>
<td>c.1066G&gt;T</td>
<td>p.G625*, p.G356*</td>
<td>I</td>
<td>de novo</td>
<td>5</td>
<td>SCV000778575</td>
</tr>
<tr>
<td>RaPS_18</td>
<td>BCHE</td>
<td>c.2T&gt;C</td>
<td>p.M1T</td>
<td>II</td>
<td>homozygous</td>
<td>4</td>
<td>SCV000778578</td>
</tr>
<tr>
<td>RaPS_21</td>
<td>TBCE</td>
<td>c.155_166delGCCACGAAGGGA</td>
<td>p.S52_G55del</td>
<td>I</td>
<td>homozygous</td>
<td>4</td>
<td>SCV000778576</td>
</tr>
<tr>
<td>RaPS_24</td>
<td>CC2D2A</td>
<td>c.585_586dupTA</td>
<td>p.T196fs*63</td>
<td>I</td>
<td>CH</td>
<td>5</td>
<td>SCV000778577</td>
</tr>
</tbody>
</table>

The table lists details of those gene variants identified in this study as likely causative or clinically relevant to feedback. The Phase column refers to the variant analysis Phase through which the variant was identified. The Variant class column refers to the variant classification based on the ACMG guidelines. Classes 4 and 5 refer to likely pathogenic and pathogenic variants, respectively. We have deposited full details of each of our variants to the ClinVar database and provide a link as shown in the ClinVar accession column. CH, compound heterozygous; XLR, X-linked recessive.