HGG Advances

ARTICLE

Refinements and considerations for trio whole-genome sequence analysis when investigating Mendelian diseases presenting in early childhood

Courtney E. French,[1,8] Helen Dolling,[1,2,3,9] Karyn Mégy,[1,3] Alba Sanchis-Juan,[1,3] Ajay Kumar,[1] Isabelle Delon,[2] Matthew Wakeling,[4] Lucy Mallin,[5] Shruti Agrawal,[2] Topun Austin,[1,2] Florence Walston,[6] Soo-Mi Park,[2] Alasdair Parker,[2] Chinthika Piyasena,[7] Kimberley Bradbury,[7] Next Generation Children's Project Consortium, Sian Ellard,[4,5] David H. Rowitch,[1,2,3] and F. Lucy Raymond[1,2,3,*]

Summary

To facilitate early deployment of whole-genome sequencing (WGS) for severely ill children, a standardized pipeline for WGS analysis with timely turnaround and primary care pediatric uptake is needed. We developed a bioinformatics pipeline for comprehensive gene-agnostic trio WGS analysis of children suspected of having an undiagnosed monogenic disease that included detection and interpretation of primary genetic mechanisms of disease, including SNVs/indels, CNVs/SVs, uniparental disomy (UPD), imprinted genes, short tandem repeat expansions, mobile element insertions, *SMN1/2* copy number calling, and mitochondrial genome variants. We assessed primary care practitioner experience and competence in a large cohort of 521 families (comprising 90% WGS trios). Children were identified by primary practitioners for recruitment, and we used the UK index of multiple deprivation to confirm lack of patient socio-economic status ascertainment bias. Of the 521 children sequenced, 176 (34%) received molecular diagnoses, with rates as high as 45% for neurology clinics. Twenty-three of the diagnosed cases (13%) required bespoke methods beyond routine SNV/CNV analysis. In our multidisciplinary clinician user experience assessment, both pediatricians and clinical geneticists expressed strong support for rapid WGS early in the care pathway, but requested further training in determining patient selection, consenting, and variant interpretation. Rapid trio WGS provides an efficacious single-pass screening test for children when deployed by primary practitioners in clinical settings that carry high *a priori* risk for rare pediatric disease presentations.

Introduction

Rapid genomic analysis of children in intensive care units (ICUs) has been heralded as the most significant change in clinical practice in the past 10 years.[2] More than 20 studies using a range of case selection and genomic technologies, including gene panels, whole-exome and whole-genome analysis, have reported diagnostic yields of 20%–50%.[1,3–6] Although analysis of single-nucleotide variants (SNVs) is well reported, accurate analysis of other types of diagnostic variants varies among sequencing strategies, bioinformatics pipelines, and publications. Whole-genome sequencing (WGS) has the advantage that multiple mechanisms of disease that cause Mendelian disease can be analyzed without additional experimentation. However, there are still some barriers and challenges to implementation in routine care.[1,4,7] For example, in which clinical settings should rapid WGS be deployed by primary practitioners at the first medical encounter or admission? Iterative clinician feedback is a crucial tool to improve this implementation.

We previously described results of WGS trio analysis of 195 patients recruited primarily from neonatal intensive care units (NICUs) and pediatric intensive care units (PICUs) with a 2 week turnaround time.[1] Here, we expanded our cohort and developed a more comprehensive bioinformatics pipeline to determine the genomic architecture of 521 young, seriously ill children, applied in several pediatric service settings (NICU, PICU, and neurology and genetics clinics) without socio-economic ascertainment bias in the National Health Service (NHS) in England. The data illustrate the multiple different disease-causing mechanisms that need to be considered in a comprehensive first-pass WGS analysis. Furthermore, the data show the importance and utility of primary clinician education and uptake in the initiation of rapid WGS diagnostic testing in both the intensive care and pediatric clinic settings.

Subjects and methods

Participants were recruited under NRES Cambridge South Research Ethics 13/EE/0325. Of the 1,554 individuals recruited, 189 consented to initial diagnostic analysis, 1,352 (87.7%) consented to longitudinal research within the National Institute for Health and Care Research (NIHR) Bioresource, and 13 withdrew. The study protocol

[1]School of Clinical Medicine, University of Cambridge, Cambridge CB2 1TN, UK; [2]Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK; [3]NIHR Bioresource, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK; [4]Institute of Biomedical and Clinical Science, University of Exeter Medical School, University of Exeter, Exeter EX4 4PY, UK; [5]Royal Devon and Exeter NHS Foundation Trust, Exeter EX2 5DW, UK; [6]Norfolk and Norwich University Hospital NHS Foundation Trust, Norwich NR4 7UY, UK; [7]Guy's and St Thomas' NHS Foundation Trust, London SE1 9RT, UK; [8]Boston Children's Hospital, Boston, MA 02115, USA; [9]Centre for Family Research, Department of Psychology, University of Cambridge, Cambridge CB2 3RQ, UK
*Correspondence: flr24@cam.ac.uk
https://doi.org/10.1016/j.xhgg.2022.100113.

Table 1. Special cases of pediatric rare disease requiring refined filtering

Exceptions to filtering rules	Refinement of bioinformatics pipeline	NGC cases affected
Non-Mendelian inheritance pattern: imprinted genes	Reports heterozygous inherited variants for a list of known disease-associated imprinted genes.	Case 484 (heterozygous stop gain variant in *MAGEL2* inherited from the unaffected father).
Non-Mendelian inheritance pattern: *PCDH19*	Reports heterozygous variants inherited from the father for this X-linked gene (female patients only).	Case 456 (heterozygous stop gain variant in *PCDH19* inherited from the hemizygous unaffected father).
Non-Mendelian inheritance pattern: incomplete penetrance/variable expressivity	Reports heterozygous inherited variants for a list of genes associated with the patient's HPO terms or provided by the clinician. Also appropriate for cases with an affected parent.	Sixteen cases had inherited dominant variants, including case 466 (*SLC2A1*, *G LUT1* deficiency/epilepsy, incomplete penetrance) and case 68 (*PTPN11*, Noonan syndrome, mildly affected parent).
Common in population datasets: incomplete penetrance/variable expressivity	Higher population thresholds for recessive variants in genes known to have incomplete penetrance or variable expressivity and variants that are common (e.g., *CYP21A2*).	Case 274 had compound heterozygous variants in *CYP21A2* that were found multiple times as homozygous in gnomAD, and p.Val282Leu has AF > 1%.
Common in population datasets: somatic mutations	Higher population thresholds for *de novo* variants in genes known to be enriched for somatic mutations in blood (*ASXL1*, *DNMT3A*).	Case 420 (*de novo* frameshift variant in *ASXL1* that has an allele count of 121 in gnomAD).
Common in population datasets: founder variants	Higher population thresholds for recessive variants in genes known to have common pathogenic founder mutations (e.g., *CFTR*:p.Phe508del, the most common pathogenic allele for cystic fibrosis in Caucasians, which make up the bulk of most large population datasets).	–
Common in population datasets: functional polymorphisms	Higher population thresholds for recessive variants in genes known to have common pathogenic functional polymorphisms (e.g., hypomorphs such as NM_005105.5:c.-21G>A in *RBM8A* that are only disease causing for TAR syndrome when in *trans* with a loss-of-function variant).	–

was as previously described[1] except that (1) WGS was aligned to the GRCh38 human reference genome build, at 50–60×; (2) an expanded data analysis pipeline was used (see below); and (3) research reports were sent to referring clinicians. Variants of uncertain significance (VUS) were not reported except in specific cases as detailed.

SNVs/indels (insertion/deletions), copy number variants (CNVs)/structural variants (SVs), and mitochondrial genome (MT) variants were annotated and filtered as previously described,[1,8–10] with a few improvements (Table 1; Supplemental methods). The SNV/indel annotation and filtering pipeline was validated using 487 trio exomes with reportable findings provided by the Exeter Genomics Laboratory. In addition, mobile element insertions (MEIs) were called by MELT,[11] annotated and filtered similarly to SNVs. Copy numbers for *SMN1* and *SMN2* were called by SMNCopyNumberCaller,[12] and tandem repeat expansions were called for specific genes with ExpansionHunter[13] and ExpansionHunter DeNovo.[14] Uniparental hetero- and isodisomy (UPD) was determined for specific disease-associated regions. Details of additional analyses are described in supplemental methods.

The impact and clinician user experience outlining the perceived advantages and challenges of recruiting patients and implementation of routine rapid trio WGS in pediatric care were assessed using a survey at the end of the project.

Results

Comprehensive gene-agnostic trio WGS analysis pipeline for pediatric genetic diseases

WGS analysis was performed on 521 severely ill neonatal or pediatric patients recruited over the course of 3.5 years (2016–2020) from NICU (194 [37%]), PICU (118 [23%]), pediatric neurology (122 [23%]), and clinical genetics (87 [17%]) clinics. Recruitment rates of eligible patients in pediatric neurology and clinical genetics were high (82%–88%), while recruitment rates in NICU and PICU improved over time, with higher rates in the last 1.5 years (80% and 73%, respectively) compared with the first 2 years (47% and 49%, respectively) (Table S2). Families were consented to the study by a trained research coordinator after referral to the study by clinicians. Consent discussions used a standard script that was adapted to individual parents' needs and questions. Results from the first 195 patients were previously published.[1] The overall median age was 8.5 months (Table S1; Figure S1), and no socio-economic bias in eligibility and consent to the study was observed using UK index of multiple deprivation (IMD) scores (Figure S2), reflecting equity of access provided by the UK National Health Service. A gene-agnostic trio analysis was performed in 90% of cases, as both parents contributed DNA samples (467 families).

We developed a comprehensive bioinformatics pipeline designed to identify the cause of severe early-onset rare diseases. Analysis included rare SNVs and small insertion/deletions with a predicted impact on gene function; CNVs and SVs of all sizes including deletions, duplications, insertions, inversions, translocations, and MEIs; variants in the mitochondrial genome; known pathogenic tandem repeat expansions; imprinted genes and UPD; and deletions

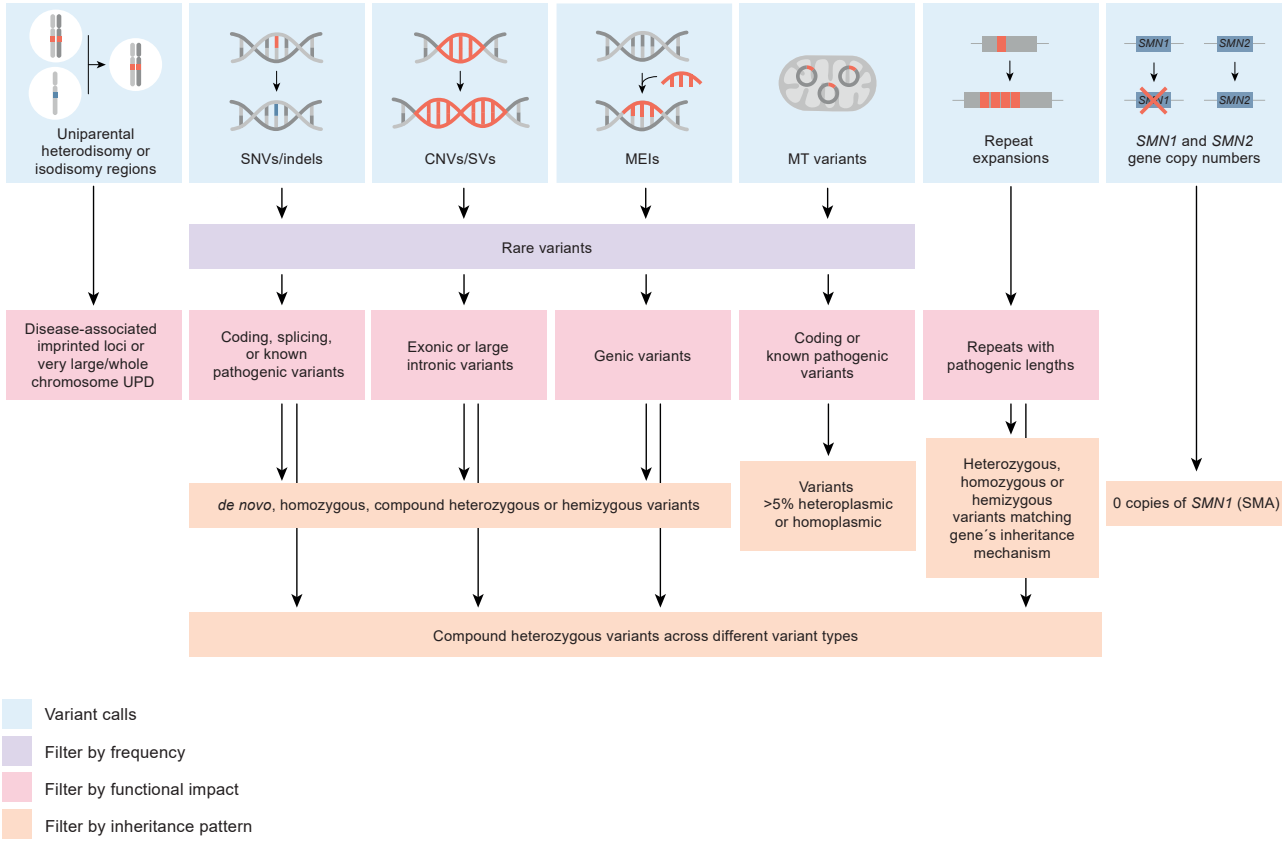

Figure 1. Comprehensive trio WGS bioinformatics pipeline
Schema of the bioinformatics pipeline. The color corresponds to the main steps (variant calls and various filtering types). CNV, copy number variant; MEI, mobile elements insertion; MT, mitochondria; SNV, single-nucleotide variant; SV, structural variant; UPD, uniparental disomy.

affecting *SMN1*, causative of spinal muscular atrophy (SMA; MIM: 253300) (Figure 1).[7,12,13,15–18]

Several exceptions and refinements were incorporated to accommodate genes with different disease-causing mechanisms (Table 1). Potential causal variants were discussed in multidisciplinary team meetings (MDTs), and variants were interpreted following the American College of Medical Genetics (ACMG) guidelines[19,20] (Table S3). Pathogenic, likely pathogenic, and a few variants of uncertain significance were reported, or a negative report was issued. All positive results were independently confirmed using an appropriate alternative method. Reports were understandable to both non-specialist physicians and patients (Figure S3), with technical details as appendices.[21] Variants in genes of certain significance were submitted to ClinVar with ACMG classification. Variants in candidate genes of uncertain significance were submitted to GeneMatcher.

Results from the application of the comprehensive analysis pipeline

One hundred seventy-six of the 521 families (34%) received confirmed molecular diagnoses (Tables S3 and S4). Of these, 56% (n = 99) had *de novo* variants, 34% (n = 60) had recessive inherited variants, and 10% had inherited variants with reduced penetrance or affected family members (n = 14), different disease mechanisms such as *PCDH19* (MIM: 300460) (n = 1), imprinted genes (n = 1), or mitochondrial genome variants (n = 1) (Figure 2). Trio analysis and frequency filtering for rarity permitted efficient interpretation of *de novo* and compound heterozygous variants (76% of cases with causal variants). Ethnicity was determined from genotyping, and the majority of probands were of European ancestry (79%, compared with 86% in the UK population[22]). Increased diagnostic rates for families of South Asian ancestry (47%) compared with European ancestry (31%) were observed, driven by an increased number of reported homozygous causative variants (Figure 2).

A total of 137 cases had disease-causing SNVs/indels in the nuclear genome, 54 (31%) with predicted loss-of-function variants (42 coding, 8 canonical splice sites, 4 non-canonical splice sites) and 83 (47%) with missense variants or in-frame indels (Figure 3A). Four cases (2%) were explained by variants in the mitochondrial genome: one was homoplasmic in both patient and mother (case 909) and the others were *de novo* and heteroplasmic in blood from the patient (case 445, 16% heteroplasmy; case 268, 61%; and case 223, 69%). Finally, disease-causing SVs or CNVs were present in 35 cases (20%). Of these, 12 were large CNVs (>500 kb), and 15 cases had smaller intragenic deletions

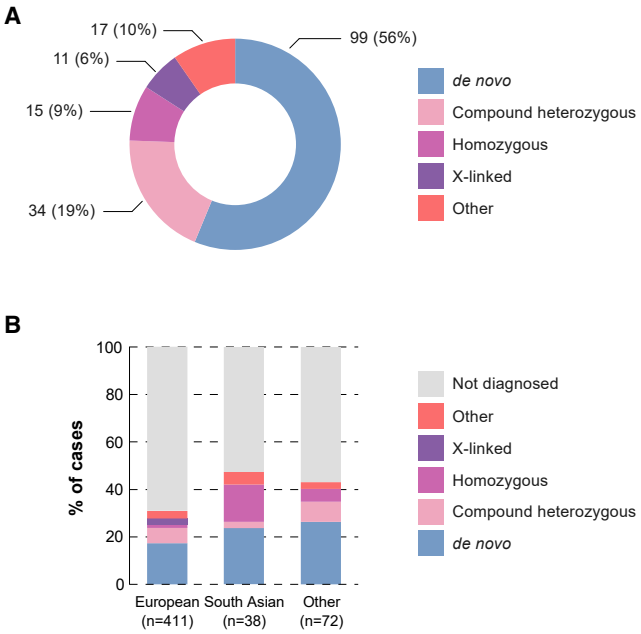

Figure 2. Inheritance patterns of variants in diagnosed cases
Distribution of inheritance patterns of variants in (A) all diagnosed cases and (B) split by genomic ethnicity.

and duplications, ranging in size from 77 to 126 kb, 50% of which were in *trans* with disease-causing SNVs/indels. The remaining 8 cases required specialized analysis and molecular genetic confirmation, including two with non-coding structural variants affecting the promoter region where biochemical or immunological analysis also confirmed pathogenicity (cases 440 and 490), three with *SMN1* deletions causing SMA (cases 142, 299, and 481; Figure 3B), two with disease-causing uniparental disomy (cases 437 and 441; Figure 3C), and one with a pathogenic repeat expansion in *DMPK* causing myotonic dystrophy (MIM: 160900) (case 388; Figure 3D).

The diagnostic rate varied by referral specialty location, with the highest yields in the pediatric neurology department (46%) and clinical genetics clinic (40%) and lower yields in the PICU (31%) and NICU (25%) (Table 2). Patients with neurodevelopmental delay, hypotonia, seizures, or suspected mitochondrial disorder had higher diagnostic rates (45%, 46%, 35%, and 46%, respectively; Table 2). Of the 16 cases with molecular diagnoses and suspected mitochondrial disease, 7 (44%) had pathogenic variants in genes not previously associated with mitochondrial biology. Children born very or extremely preterm (23–32 weeks) and moderate to later preterm (32–37 weeks) had lower genetic diagnostic rates compared with term infants in both the PICU and NICU (10% versus 25% versus 32%) and non-ICU settings (0% versus 33% versus 45%) (Table S5). A higher diagnostic rate (51% versus 32%) was observed in children with the most severe disease who did not survive beyond 1 year of age.

Gene-agnostic trio analysis allows both the identification of novel phenotypes for known disease-causing genes and gene discovery. Multiple cases expanded the phenotypic spectrum of known genes, including *TTN*,[23] *TELO2*,[24] *UNC45A*,[25] *PPCS*,[26] *NSD1*,[27] and *FBN1*[28] (MIM: 188840, 611140, 611219, 609853, 606681, and 134797, respectively). Variants in these genes were not suspected prior to testing. The variants in *TTN* were associated with an antenatal diagnosis of arthrogryposis, polyhydramnios, and neonatal respiratory difficulties in keeping with a rare titinopathy, and the child with a *de novo* PPCS variant presented with severe hypotonia, necrotizing myopathy, and intermittent rhabdomyolysis who only later developed a severe cardiomyopathy post-diagnosis. *TELO2* and *UNC45A* are only recently reported genes, and such young patients had not been identified before. For *NSD1* and *FBN1*, these syndromes had not been considered in the clinical differential diagnosis, but in retrospect the phenotype of the cases fitted well as the children developed. Nearly 15% of the cases with causal variants (n = 24) had variants in genes first reported since the start of the study in 2016, and for 9, the variants were reported after iterative re-analysis, indicating the importance of regular review and "updating." Three families contributed to the first identification of novel disease-causing genes or phenotypes: *VPS4A*,[29] *NDUFA6*,[30] and *TARS2*[31] (MIM: 609982, 602138, and 600124, respectively). Nine negative cases (not included in the overall diagnostic rate) have variants in candidate disease genes submitted to GeneMatcher and are undergoing further investigations.

Clinical impact and professional user experience

To assess the impact and user experience of the Next Generation Children's (NGC) project, referring pediatricians and clinical geneticists completed a survey. The response rate was 34% from the NICU, the PICU, pediatric immunology, neurology, and clinical genetics (32 of 94) (Figure S4). The numbers of respondents from pediatrics (n = 18) and clinical genetics (n = 14) were comparable, and although some differences in responses between the groups can be observed, they are not statistically significant by Fisher's exact test (Table S6; Figure S5).

More than 90% of clinicians found the study "very useful" or "somewhat useful" with regard to contributing to research (n = 32 [100%]); increased confidence in decision making, especially when a molecular diagnosis was made (n = 31 [97%]); improved communication with parents (n = 31 [97%]); gaining knowledge about a genetic etiology (n = 31 [97%]); the ability to offer an additional test to families (n = 30 [94%]); helping explain a poor outcome (n = 30 [94%]); and improved clinical management of patients (n = 29 [91%]) (Figure 4A; Table S7).

There were also documented challenges to the study. Clinicians described an increased workload, but the majority (n = 23 [72%]) regarded this as "insignificant." More challenging was the requirement to decide on patients' suitability for testing (n = 13 [41%] "somewhat significant" or "very significant") and the need to perform extra research into reported variants (n = 13 [44%] "somewhat

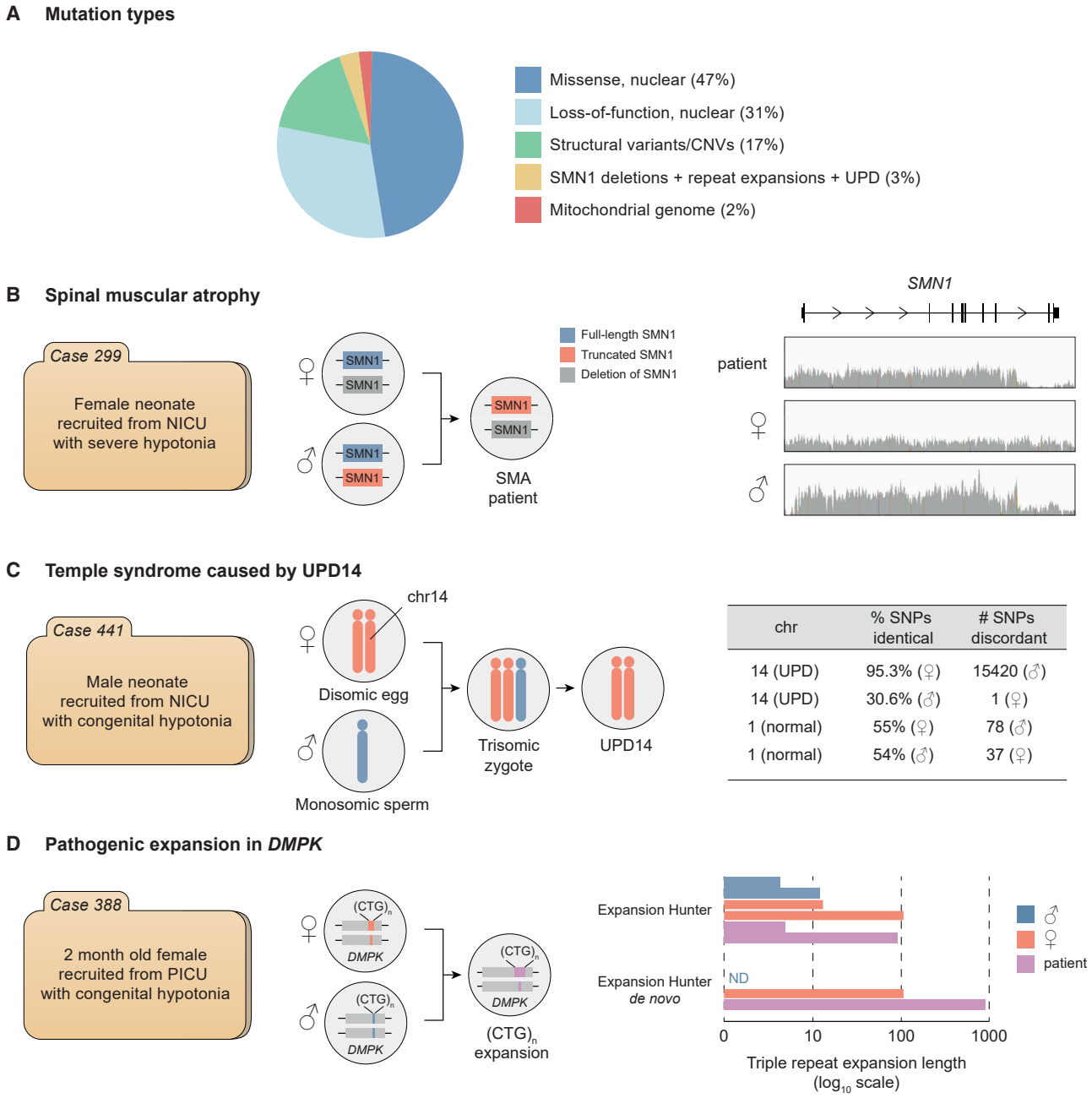

Figure 3. Different mechanisms of disease found in the cohort

(A) Proportion of the different inheritance and variant types identified in the cohort. Examples of cases with (B) spinal muscular atrophy from *SMN1* deletion and truncation (exon 7 and 8 deletion), (C) Temple syndrome from maternal uniparental heterodisomy 14, and (D) congenital muscular dystrophy in the child due to a large expansion of a trinucleotide repeat expansion in the 3′ UTR of *DMPK* inherited from the mother. For the latter, ExpansionHunter detected heterozygous anchored repeat lengths of 5/15 in the father, 13/107 in the mother, and 5/92 in the child. ExpansionHunter DeNovo detected 9 times as many paired in-repeat reads in the child compared with the mother, indicating early-onset myotonic dystrophy (DM1) in the child. Triplet PCR in mother and child confirmed expansions in the pathogenic range.

significant or "very significant"). The biggest challenges occurred around having to communicate with the parents, with almost half of clinicians reporting the need for additional communication about genetic testing (n = 15 [47%] "somewhat significant" or "very significant") and uncertainty about explaining genetic results to parents (n = 16 [50%] "somewhat significant" or "very significant"), including discussing negative results (Figures 4B; Table S7).

Finally, we asked clinicians to comment on the future use of WGS by scoring questions and entering free text. The majority (n = 29 [91%]) strongly (score 5 out of 5) recommended the use of trio WGS in their specialty, but pediatricians were much less confident about their ability in determining

Table 2. Diagnostic rates by medical specialty and phenotype

	All		With seizures[a]		With delay[a]		With hypotonia[a]		Suspected mitochondrial[a]	
Specialty	Total	Diagnosed	Total	Diagnosed	Total	Diagnosed	Total	Diagnosed	Total	Diagnosed
NICU	194	48 (25%)	44	8 (18%)	4	1 (25%)	25	12 (48%)	5	1 (20%)
PICU	118	37 (31%)	43	16 (37%)	46	16 (35%)	20	8 (40%)	5	2 (40%)
Pediatric neurology	122	56 (46%)	63	29 (46%)	65	34 (52%)	28	13 (46%)	20	12 (60%)
Clinical genetics	87	35 (40%)	19	6 (32%)	57	26 (46%)	21	10 (48%)	5	1 (20%)
ALL	521	176 (34%)	169	59 (35%)	172	77 (45%)	94	43 (46%)	35	16 (46%)

NICU, neonatal intensive care unit; PICU, pediatric intensive care unit.
[a]The phenotype categories are not mutually exclusive (patients may belong to more than one).

patient suitability (41% scoring 5 out of 5) and consenting for testing (56%), although some geneticists were also lacking in confidence in selecting suitable children and the consenting process (Figure S6). Half of the respondents wanted additional training in determining eligibility, consenting, analysis, and variant interpretation.

Discussion

We performed a large prospective gene-agnostic WGS trio analysis of 521 severely ill children and their parents (trios) within the National Health Service in England. The NHS provides universal health care for almost 100,000 children admitted to ICUs each year and >0.5 million new hospital outpatient pediatric appointments per year. Our demographic analysis of recruitment showed that WGS can be offered and accepted equitably across the socio-economic spectrum and yields beneficial insights into the burden of genetic disease across all social groups. As rapid genome-wide sequencing becomes integrated into routine pediatric care, a key question is how to expand the analysis pipeline to its full potential, including deployment by primary practitioners. Our study reports recruitment from several clinical specialist units (NICU, PICU, pediatric neurology, and genetics clinic) and demonstrates the occurrence of a high *a priori* risk for genetic disease in these patients. Although referrals came from the specialist consultants (e.g., a neonatologist in the NICU), identifying appropriate patients was frequently based on a multidisciplinary discussion with attending neurologists and geneticists who also supported the consenting process.

We developed an in-house comprehensive bioinformatics pipeline to determine the genomic architecture of this young, seriously ill cohort, which required specific considerations.

Multiple mechanisms of disease

The phenotypic spectrum of presentation of disease in very young children is often limited and non-specific (e.g., hypotonia). The primary analysis therefore included the common exon 7 and 8 deletion in *SMN1* (spinal muscular atrophy), abnormal expansions in *DMPK* (congenital myotonic dystrophy), UPD (e.g., Prader-Willi syndrome [MIM: 176270]), genome-wide SNVs, CNVs, and mitochondrial genome defects, which yielded an overall diagnostic rate of 46%. WGS analysis removes the need for multiple separate specialized laboratory tests and facilitates the rapid turnaround by simultaneous testing to inform therapy in seriously ill children. The use of WGS analysis permits structural variant calling, including mobile element insertions and complex SVs, that cannot be deciphered with most other sequencing strategies. For example, we identified cases in which a promoter inversion (case 490), a promoter duplication (case 440), and a complex structural rearrangement (case 375) caused disease but were not apparent by exome sequence analysis. Overall, 22% of diagnoses would have been missed without such broad systematic CNV, SV, and mitochondrial genome analyses.[5,32]

Gene-agnostic approach and frequency filters

A gene-agnostic approach ensures that all genes are considered equally initially, regardless of phenotype, but requires a trio design with low population frequency filters to limit the number of rare variants to be considered. A trio design enabled rapid assessment of rare variants with respect to *de novo* inheritance and assessment of rare variants in *trans* for recessive disease. Importantly, human phenotype ontology (HPO) terms are not used to select genes to be considered in the initial analysis. This enables new and previously unrecognized phenotypes to be assigned to known genes at a significant rate that would not otherwise be on a panel.[1] Subsequently, it is also necessary to consider higher population frequency filters during the analysis to ensure that a comprehensive analysis is performed, as some pathogenic variants are more frequent in the general population. To make the analysis rapid and thus timely, only biologically relevant genes associated with the phenotype of the child are considered. This reduces the number of less rare variants that need to be considered on the basis of the secondary HPO driven gene prioritization. This facilitates variant identification of (1) founder mutations in certain populations (e.g., *CFTR* [MIM: 602421]); (2) reduced penetrance alleles in certain genes (e.g., *SLC2A1* [MIM: 138140], *SERPINA1* [MIM: 107400]) for which one parent may carry the gene abnormality that would be filtered out in a strict *de novo*

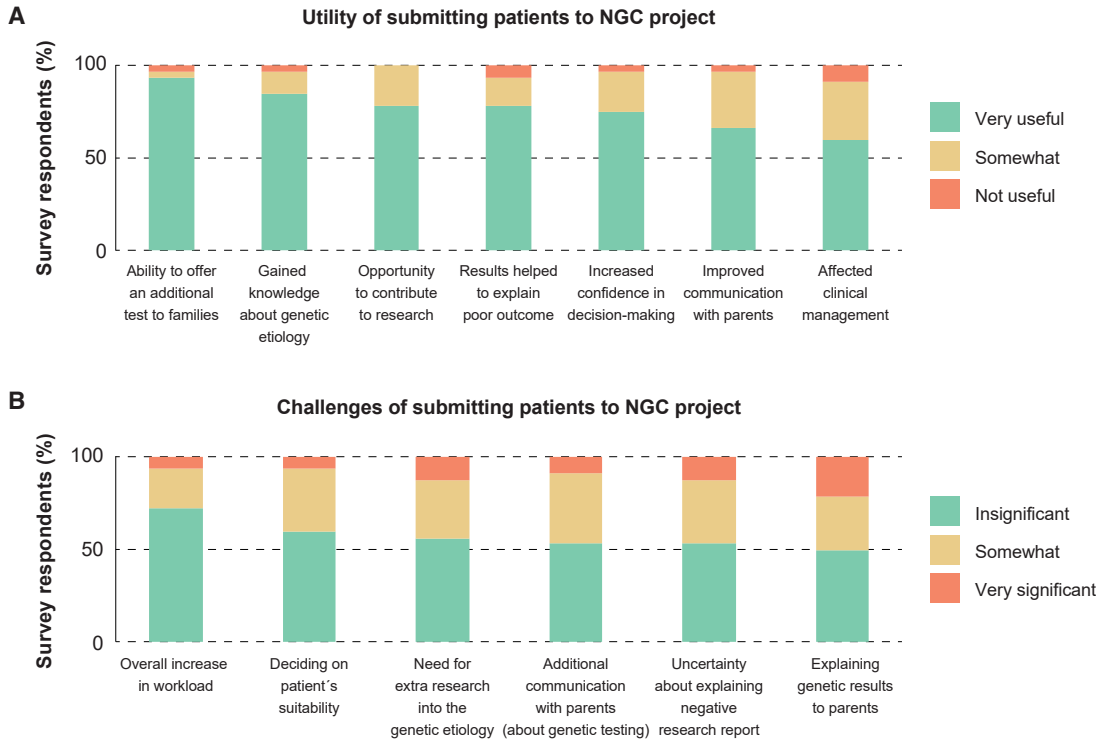

Figure 4. NGC impact survey results: utility and challenges
Results of the survey assessing the impact of the NGC project. Thirty-two pediatricians and clinical geneticists were asked to rate (A) advantages of the study as not, somewhat, or very useful and (B) challenges as not, somewhat, or very significant.

analysis (variants in imprinted genes, such as *MAGEL2* [MIM: 605283], can be missed when filtering using Genome Aggregation Database [gnomAD] variant frequencies or similar adult control datasets, as the parent of origin is not documented in controls and will appear too frequent for rare diseases); (3) diseases associated with rare germline mutations in *ASXL1* (MIM: 612990) and *DNMT3* (MIM: 602769)[33] which can be missed if frequencies alone in gnomAD are used, as variants in these genes have high somatic frequencies in blood from adult control populations[34]; and (4) rare recessive Mendelian diseases (e.g., thrombocytopenia-absent radius syndrome caused by mutations in *RBM8A* [MIM: 605313]), in which one rare variant is in association with a common SNP in *trans*, causing disease.[35]

We report an overall diagnostic rate of 34%, emphasizing the high burden of monogenic disease in pediatric and sub-specialty care. The yield was higher in older children with chronic neurological symptoms, as distinguishing transient from chronic pathology is challenging in the neonatal period, especially among preterm infants, in whom fewer phenotypic features are present. The diagnostic rate reported is likely to be an underestimate, as mainly pathogenic and likely pathogenic variants in genes of certain significance were returned to the families.[19] An unexpectedly high percentage of diagnoses had autosomal-recessive inheritance patterns (28%), including several cases with compound heterozygous variants (n = 34 [19% of diagnosed cases]). This is true for both NICU or PICU and non-ICU settings and is driven partially by the identification of CNVs in *trans* with SNVs in 8 cases (5% of diagnosed cases) and the inclusion of the recessive SMA cases (n = 3 [2%]). We did not observe higher metabolic diagnoses in the sick ICU patients, although we did identify several mitochondrial diseases in patients within the severely sick cohort. These were due to recessive nuclear encoded mitochondrial gene variants. Many of these children died. Prior to the availability of rapid trio genome analysis, they may have died without diagnoses.

Reporting of variants of uncertain significance, particularly in critically ill children, was deemed of limited immediate benefit. Of cases with nuclear SNVs/indels, only 23 VUSs were returned (18%). These were reported when in *trans* with a pathogenic or likely pathogenic variant (n = 5), additional studies were planned to test functional impact (n = 7), the variant had already been reported (n = 2), and communicating clinical uncertainty of variant interpretation in older children was deemed to be of benefit (n = 9). In addition, 6 VUSs were reported to be benign after the recommended functional biochemical testing, cytogenetics, or histopathology had been performed, for example, very long chain fatty acid levels (*ABCD1* [MIM: 300100]), bleeding times and clotting factor tests (*F8* [MIM: 301071]), centromeric instability (*DNMT3B* [MIM: 242860]), and immunohistochemistry of muscle (*PYGM* [MIM: 608455]).

Frequent iterative re-analysis of the data improved the diagnostic rate as new genes were published. Submission

of variants to ClinVar facilitated future variant interpretation, and submission of variants in genes of uncertain significance to GeneMatcher ensured that further novel disease genes can be identified. Continuing follow-up of families within the NIHR Bioresource enables psychological, health economic, and mechanistic research together with recruitment to stratified medicine and clinical trials.

Furthermore, pipelines need improvement, and each iteration must be validated using a known truth set. Areas for future improvements include highly repetitive regions that map poorly but harbor important causal variants that need specific variant calling and quality control; improved variant interpretation using advanced prediction tools for the impact on mRNA splicing or protein function; increasing availability of ethnicity-specific population controls; and additional knowledge of the non-coding space.[36–38] However, some relevant mechanisms of disease, such as methylation defects and tissue-specific mosaicism, cannot be called by routine blood WGS analysis.

Critical to the success of WGS in a clinical diagnostic setting is primary practitioner uptake, positive user experience, and acceptance. Although initial anecdotal comments (e.g., "It's good not to have too many VUSs," "We don't want to go back [to not having the test as an option]") were encouraging, formal feedback was essential for evaluation. There were no significant differences between pediatricians' and clinical geneticists' views regarding issues of implementation, although with <20 responses to the survey per group, there was limited power to detect statistically significant differences. The small differences observed were related to the degree of concern. To offer rapid extensive testing to families and to participate in research were perceived by both groups as very useful and outweighed the additional workload needed for consenting. However, clinicians did not consent patients to this study, and thus the workload associated with consenting may be underestimated. Non-genetic specialist pediatricians noted the value of genomic results that explained poor outcomes, reassuring professionals and parents that optimum care had been provided. Our results highlight the need for more education for all professional groups (including genetic specialists) around selecting those suitable for testing, explaining genetic investigations, and variant interpretation.

In summary, on the basis of the genomic architecture of a large cohort of children identified using a comprehensive WGS pipeline, we recommend rapid trio analysis early in the diagnostic pathway, ordered by the primary practitioner on admission to a hospital or specialist clinic. We describe the need for both stringent and lenient frequency filtering on the basis of the biology of each early-onset disease gene. The use of rapid trio WGS was welcomed by clinicians as a significant improvement to patient care. More training in using and communicating genomics to families was requested by all professional groups, and further evaluation of the workload and skills needed to offer this testing is required in a much larger sample of professionals. Finally, future research is needed to evaluate not only the benefits to systematic deployment of WGS in the NHS to mitigate the "diagnostic odyssey" and the facilitation of precision medicine but also to understand the long-term psychosocial impact of a genetic diagnosis, the optimal timing of delivery of a result driven by the needs of the family rather than the speed of the technology, and the support needs of families across the range of socio-economic spectrum once a diagnosis is made.

Consortia

The members of the Next Generation Children's Project Consortium are Evan Reid, Geoff Woods, Hillary Wong, James Whitworth, Kathryn Beardsall, Matina Prapa, Nazima Pathan, Richard Sandford, Rita Horvath, Sarah Morley, Serena Nik-Zainal, Patrick Yu Wai Man, Helen Butcher, Helen Dolling, David Rowitch, James Rowe, Lucy Raymond, Marc Tischkowitz, Patrick Chinnery, Alba Sanchis-Juan, Christopher Penkett, Karyn Mégy, Olga Shamardina, Willem Ouwehand, Andrea Katrinecz, Ajay Kumar, Courtney E. French, Salih Tuna, Elizabeth Radford, Audrienne Sammut, Topun Austin, John Bradley, Ajay Thankamony, Alasdair Parker, Alexander Ross, Amanda Ogilvy-Stuart, Ana Lisa Taylor Tavares, Angela D'Amore, Anna Maw, Anne Kelly, Bina Mukthyar, Caterina Garone, Cristine Sortica da Costa, David Hopkins, Deepa Krishnakumar, Doris Wari-Pepple, Gautam Ambegaonkar, Gusztav Belteki, Helen Baxendale, Helen Firth, Isabelle Delon, Jennifer Hague, Joan Paterson, Jonathon Holland, June-Anne Gold, Katrina Andrews, Manali Chitre, Marion Bohatschek, Mike Gattens, Phil Buttery, Pooja Harijan, RagaMallika Pinnamaneni, Riaz Kayani, Ricardo Branco, Richard Brown, Roddy O'Donnell, Rosalie Campbell, Ruth Armstrong, Sajeev Job, Sam O'Hare, Sarah Bowdin, Saraswathy Sabanathan, Shazia Hoodbhoy, Shruti Agrawal, Simon Holden, Soo-Mi Park, Sue Broster, Thomas Rossor, Tim Ham, Wilf Kelsall, Timothy Ham, Yogen Singh, Katharine Riley, Miss Brinda Muthusamy, Miss Louise Allen, Raul Silva, Annie Fitzsimmons, Joanne Harley, Kelly Spike, Nathalie Kingston, Sian Ellard, Matthew Wakeling, Lucy Malin, Florence Walston, Paul Clarke, Vivek Kalra, Camilla Kingdon, A. David Edwards, Chinthika Piyasena, Dragana Josifova, Francesca Forzano, Hugh Lemonde, Kimberley Bradbury, Melita Irving, Michael Carter, Muriel Holder, Rahul Singh, Cristina Dias, Emma Clement, Jane Hurst, Melissa Lees, Emma Wakeling, Virginia Clowes, Tessa Homfray, Angeliki Menounou, Anand Iyer, Rachel Kneen, Rajiv Chaudhary, Stefan Spinty, Robert Hegarty, Richard Thompson, Sophie Hambleton, and Abhijit Dixit. Affiliations are listed in Table S8.

Data and code availability

The sequencing data generated during this study are available in EGA: initial publication (French et al.,[1] *Intensive Care Medicine*, 2019) (GRCh37) EGAD00001004357; this publication (GRCh37) EGAD00001007780; this publication (GRCh38) EGAD00001007868.

Pathogenic variants and others of interest were submitted to ClinVar. The code for annotation and filtering SNVs and for structural variants is publicly available at GitHub (https://github.com/ajaarma/snv and https://github.com/ajaarma/sv). All other code is available upon request.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100113.

Acknowledgments

We would like to thank Antonio Garcia for figure preparation. This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014), the NIHR Rare Disease Bioresource, The Rosetrees Trust, and the Isaac Newton Trust. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of interests

The authors declare no competing interests.

Received: January 25, 2022
Accepted: April 19, 2022

Web resources

OMIM, http://www.omim.org/
Index of deprivation, https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
MELT, https://melt.igs.umaryland.edu/
Vcfanno, https://github.com/brentp/vcfanno
ExpansionHunter, https://github.com/Illumina/ExpansionHunter
MToolBox, https://github.com/mitoNGS/MToolBox
SMNCopyNumberCaller, https://github.com/Illumina/SMNCopyNumberCaller
ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
GeneMatcher, https://genematcher.org/

References

1. French, C.E., Delon, I., Dolling, H., Sanchis-Juan, A., Shamardina, O., Megy, K., Abbs, S., Austin, T., Bowdin, S., Branco, R.G., et al.; Next Generation Children Project (2019). Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. Intensive Care Med. *45*, 627–636. https://doi.org/10.1007/s00134-019-05552-x.
2. Stark, Z., and Ellard, S. (2021). Rapid genomic testing for critically ill children: time to become standard of care? Eur. J. Hum. Genet. *30*, 142–149. https://doi.org/10.1038/s41431-021-00990-y.
3. Gubbels, C.S., VanNoy, G.E., Madden, J.A., Copenheaver, D., Yang, S., Wojcik, M.H., Gold, N.B., Genetti, C.A., Stoler, J., Parad, R.B., et al. (2020). Prospective, phenotype-driven selection of critically ill neonates for rapid exome sequencing is associated with high diagnostic yield. Genet. Med. *22*, 736–744. https://doi.org/10.1038/s41436-019-0708-6.
4. Sanford, E.F., Clark, M.M., Farnaes, L., Williams, M.R., Perry, J.C., Ingulli, E.G., Sweeney, N.M., Doshi, A., Gold, J.J., Briggs, B., et al. (2019). Rapid whole genome sequencing has clinical utility in children in the PICU. Pediatr. Crit. Care Med. *20*, 1007–1020. https://doi.org/10.1097/PCC.0000000000002056.
5. Australian Genomics Health Alliance Acute Care, F., Lunke, S., Eggers, S., Wilson, M., Patel, C., Barnett, C.P., Pinner, J., Sandaradura, S.A., Buckley, M.F., Krzesinski, E.I., et al. (2020). Feasibility of ultra-rapid exome sequencing in critically ill infants and children with suspected monogenic conditions in the Australian public health care system. JAMA *323*, 2503–2511. https://doi.org/10.1001/jama.2020.7671.
6. Raymond, F.L. (2020). Clinical genomics in critically ill infants and children. JAMA *323*, 2480–2482. https://doi.org/10.1001/jama.2020.8112.
7. Carss, K.J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., Megy, K., Grozeva, D., Dewhurst, E., Malka, S., et al. (2017). Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am. J. Hum. Genet. *100*, 75–90. https://doi.org/10.1016/j.ajhg.2016.12.003.
8. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.
9. Calabrese, C., Simone, D., Diroma, M.A., Santorsola, M., Gutta, C., Gasparre, G., Picardi, E., Pesole, G., and Attimonelli, M. (2014). MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. Bioinformatics *30*, 3115–3117. https://doi.org/10.1093/bioinformatics/btu483.
10. Pedersen, B.S., Layer, R.M., and Quinlan, A.R. (2016). Vcfanno: fast, flexible annotation of genetic variants. Genome Biol. *17*, 118. https://doi.org/10.1186/s13059-016-0973-5.
11. Gardner, E.J., Lam, V.K., Harris, D.N., Chuang, N.T., Scott, E.C., Pittard, W.S., Mills, R.E., Genomes Project, C., and Devine, S.E. (2017). The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res. *27*, 1916–1929. https://doi.org/10.1101/gr.218032.116.
12. Chen, X., Sanchis-Juan, A., French, C.E., Connell, A.J., Delon, I., Kingsbury, Z., Chawla, A., Halpern, A.L., Taft, R.J., NIHR BioResource, and Bentley, D.R., et al. (2020). Spinal muscular atrophy diagnosis and carrier screening from genome sequencing data. Genet. Med. *22*, 945–953. https://doi.org/10.1038/s41436-020-0754-0.
13. Dolzhenko, E., Deshpande, V., Schlesinger, F., Krusche, P., Petrovski, R., Chen, S., Emig-Agius, D., Gross, A., Narzisi, G., Bowman, B., et al. (2019). ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics *35*, 4754–4756. https://doi.org/10.1093/bioinformatics/btz431.
14. Dolzhenko, E., Bennett, M.F., Richmond, P.A., Trost, B., Chen, S., van Vugt, J., Nguyen, C., Narzisi, G., Gainullin, V.G., Gross, A.M., et al. (2020). ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol. *21*, 102. https://doi.org/10.1186/s13059-020-02017-z.
15. Sanchis-Juan, A., Stephens, J., French, C.E., Gleadall, N., Megy, K., Penkett, C., Shamardina, O., Stirrups, K., Delon, I.,

Dewhurst, E., et al. (2018). Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. Genome Med. *10*, 95. https://doi.org/10.1186/s13073-018-0606-6.

16. Turro, E., Astle, W.J., Megy, K., Graf, S., Greene, D., Shamardina, O., Allen, H.L., Sanchis-Juan, A., Frontini, M., Thys, C., et al. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. Nature *583*, 96–102. https://doi.org/10.1038/s41586-020-2434-2.
17. Wei, W., Tuna, S., Keogh, M.J., Smith, K.R., Aitman, T.J., Beales, P.L., Bennett, D.L., Gale, D.P., Bitner-Glindzicz, M.A.K., Black, G.C., et al. (2019). Germline selection shapes human mitochondrial DNA diversity. Science *364*, eaau6520. https://doi.org/10.1126/science.aau6520.
18. Dolzhenko, E., van Vugt, J., Shaw, R.J., Bekritsky, M.A., van Blitterswijk, M., Narzisi, G., Ajay, S.S., Rajan, V., Lajoie, B.R., Johnson, N.H., et al.; The US–Venezuela Collaborative Research Group (2017). Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. *27*, 1895–1903. https://doi.org/10.1101/gr.225672.117.
19. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. Genet. Med. *17*, 405–424. https://doi.org/10.1038/gim.2015.30.
20. ACGS. https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf.
21. Farmer, G.D., Gray, H., Chandratillake, G., Raymond, F.L., and Freeman, A.L.J. (2020). Recommendations for designing genetic test reports to be understood by patients and non-specialists. Eur. J. Hum. Genet. *28*, 885–895. https://doi.org/10.1038/s41431-020-0579-y.
22. UK_gov. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest.
23. Oates, E.C., Jones, K.J., Donkervoort, S., Charlton, A., Brammah, S., Smith, J.E., 3rd, Ware, J.S., Yau, K.S., Swanson, L.C., Whiffin, N., et al. (2018). Congenital Titinopathy: comprehensive characterization and pathogenic insights. Ann. Neurol. *83*, 1105–1124. https://doi.org/10.1002/ana.25241.
24. You, J., Sobreira, N.L., Gable, D.L., Jurgens, J., Grange, D.K., Belnap, N., Siniard, A., Szelinger, S., Schrauwen, I., Richholt, R.F., et al. (2016). A syndromic intellectual disability disorder caused by variants in TELO2, a gene encoding a component of the TTT complex. Am. J. Hum. Genet. *98*, 909–918. https://doi.org/10.1016/j.ajhg.2016.03.014.
25. Esteve, C., Francescatto, L., Tan, P.L., Bourchany, A., De Leusse, C., Marinier, E., Blanchard, A., Bourgeois, P., Brochier-Armanet, C., Bruel, A.L., et al. (2018). Loss-of-function mutations in UNC45A cause a syndrome associating cholestasis, diarrhea, impaired hearing, and bone fragility. Am. J. Hum. Genet. *102*, 364–374. https://doi.org/10.1016/j.ajhg.2018.01.009.
26. Iuso, A., Wiersma, M., Schuller, H.J., Pode-Shakked, B., Marek-Yagel, D., Grigat, M., Schwarzmayr, T., Berutti, R., Alhaddad, B., Kanon, B., et al. (2018). Mutations in PPCS, encoding phosphopantothenoylcysteine synthetase, cause autosomal-recessive dilated cardiomyopathy. Am. J. Hum. Genet. *102*, 1018–1030. https://doi.org/10.1016/j.ajhg.2018.03.022.
27. Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y., et al. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. Nat. Genet. *30*, 365–366. https://doi.org/10.1038/ng863.
28. Stheneur, C., Collod-Beroud, G., Faivre, L., Buyck, J.F., Gouya, L., Le Parc, J.M., Moura, B., Muti, C., Grandchamp, B., Sultan, G., et al. (2009). Identification of the minimal combination of clinical features in probands for efficient mutation detection in the FBN1 gene. Eur. J. Hum. Genet. *17*, 1121–1128. https://doi.org/10.1038/ejhg.2009.36.
29. Rodger, C., Flex, E., Allison, R.J., Sanchis-Juan, A., Hasenahuer, M.A., Cecchetti, S., French, C.E., Edgar, J.R., Carpentieri, G., Ciolfi, A., et al. (2020). De novo VPS4A mutations cause multisystem disease with abnormal neurodevelopment. Am. J. Hum. Genet. *107*, 1129–1148. https://doi.org/10.1016/j.ajhg.2020.10.012.
30. Alston, C.L., Heidler, J., Dibley, M.G., Kremer, L.S., Taylor, L.S., Fratter, C., French, C.E., Glasgow, R.I.C., Feichtinger, R.G., Delon, I., et al. (2018). Bi-allelic mutations in NDUFA6 establish its role in early-onset isolated mitochondrial complex I deficiency. Am. J. Hum. Genet. *103*, 592–601. https://doi.org/10.1016/j.ajhg.2018.08.013.
31. Zheng, W.Q., Pedersen, S.V., Thompson, K., Bellacchio, E., French, C.E., Munro, B., Pearson, T.S., Vogt, J., Diodato, D., Diemer, T., et al. (2021). Elucidating the molecular mechanisms associated with TARS2-related mitochondrial disease. Hum. Mol. Genet. *31*, 523–534. https://doi.org/10.1093/hmg/ddab257.
32. Clark, M.M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., Watkins, K., Ellsworth, K., Camp, B., Kint, C.I., Yacoubian, C., et al. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. Sci. Transl. Med. *11*, eaat6177. https://doi.org/10.1126/scitranslmed.aat6177.
33. Carlston, C.M., O'Donnell-Luria, A.H., Underhill, H.R., Cummings, B.B., Weisburd, B., Minikel, E.V., Birnbaum, D.P., Exome Aggregation, C., Tvrdik, T., MacArthur, D.G., and Mao, R. (2017). Pathogenic ASXL1 somatic variants in reference databases complicate germline variant interpretation for Bohring-Opitz Syndrome. Hum. Mutat. *38*, 517–523. https://doi.org/10.1002/humu.23203.
34. Karczewski, K.J., Francioli, L.C., MacArthur, D.G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
35. Albers, C.A., Paul, D.S., Schulze, H., Freson, K., Stephens, J.C., Smethurst, P.A., Jolley, J.D., Cvejic, A., Kostadima, M., Bertone, P., et al. (2012). Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat. Genet. *44*, 435–439. S1–S2. https://doi.org/10.1038/ng.1083.
36. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature *538*, 201–206. https://doi.org/10.1038/nature18964.
37. Sanchis-Juan, A., Hasenahuer, M.A., Baker, J.A., McTague, A., Barwick, K., Kurian, M.A., Duarte, S.T., BioResource, N., Carss, K.J., Thornton, J., and Raymond, F.L. (2020). Structural analysis of pathogenic missense mutations in GABRA2 and

identification of a novel de novo variant in the desensitization gate. Mol. Genet. Genomic Med. *8*, e1106. https://doi.org/10.1002/mgg3.1106.

38. Whiffin, N., Karczewski, K.J., Zhang, X., Chothani, S., Smith, M.J., Evans, D.G., Roberts, A.M., Quaife, N.M., Schafer, S., Rackham, O., et al.; Genome Aggregation Database Consortium (2020). Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. Nat. Commun. *11*, 2523. https://doi.org/10.1038/s41467-019-10717-9.