Integrated multi-omics for rapid rare disease diagnosis on a national scale

Sebastian Lunke [1,2,3], Sophie E. Bouffler [3], Chirag V. Patel[4],
Sarah A. Sandaradura[5,6], Meredith Wilson [5,6], Jason Pinner[7,8],
Matthew F. Hunter[9,10], Christopher P. Barnett[11,12,13], Mathew Wallis[14,15],
Benjamin Kamien[16], Tiong Y. Tan [1,2], Mary-Louise Freckmann[17], Belinda Chong [1],
Dean Phelan[1], David Francis[1], Karin S. Kassahn [12,13], Thuong Ha [12,18,19],
Song Gao[12], Peer Arts [12,13,18,19], Matilda R. Jackson[3,12], Hamish S. Scott [3,12,13,18,19],
Stefanie Eggers[1], Simone Rowley[1], Kirsten Boggs[3,5,7], Ana Rakonjac[3,5,7],
Gemma R. Brett [1,2], Michelle G. de Silva [1,2,20], Amanda Springer[9,10],
Michelle Ward[16], Kirsty Stallard[11], Cas Simons[20], Thomas Conway [20],
Andreas Halman [1], Nicole J. Van Bergen[2,20], Tim Sikora [20], Liana N. Semcesen[2],
David A. Stroud[1,2,20], Alison G. Compton [1,2,20], David R. Thorburn [1,2,20],
Katrina M. Bell[1,20], Simon Sadedin[1,2,20], Kathryn N. North[2,3,20],
John Christodoulou [2,3,6,20] & Zornitza Stark [1,2,3] ✉

Critically ill infants and children with rare diseases need equitable access to rapid and accurate diagnosis to direct clinical management. Over 2 years, the Acute Care Genomics program provided whole-genome sequencing to 290 families whose critically ill infants and children were admitted to hospitals throughout Australia with suspected genetic conditions. The average time to result was 2.9 d and diagnostic yield was 47%. We performed additional bioinformatic analyses and transcriptome sequencing in all patients who remained undiagnosed. Long-read sequencing and functional assays, ranging from clinically accredited enzyme analysis to bespoke quantitative proteomics, were deployed in selected cases. This resulted in an additional 19 diagnoses and an overall diagnostic yield of 54%. Diagnostic variants ranged from structural chromosomal abnormalities through to an intronic retro-transposon, disrupting splicing. Critical care management changed in 120 diagnosed patients (77%). This included major impacts, such as informing precision treatments, surgical and transplant decisions and palliation, in 94 patients (60%). Our results provide preliminary evidence of the clinical utility of integrating multi-omic approaches into mainstream diagnostic practice to fully realize the potential of rare disease genomic testing in a timely manner.

Genomic testing is transforming rare disease diagnosis. The unparalleled acceleration of rare disease gene discovery, dramatic reductions in the cost of genomic sequencing and government investments to drive clinical application have significantly improved both diagnostic rates and the timeliness of genetic diagnosis[1]. Genomic testing is around five times more likely to achieve a diagnosis than previous 'gold standard' tests such as chromosomal microarray[2] and is increasingly delivered with rapid turnaround times to guide clinical management in real

time[3–6]. The diagnostic and clinical utility, as well as cost-effectiveness of rapid genomic testing in critically ill infants and children with suspected genetic conditions are now well established, with over 30 studies totaling 2,000 patients published worldwide[5,7]. Multiple healthcare systems have funded this type of testing as standard of care[8].

Targeted panels, exomes and genomes have all been used in rapid genomic testing programs, but as whole-genome sequencing (WGS) becomes increasingly deliverable at scale across healthcare systems[9,10], it is expected to supersede other modalities. WGS can comprehensively assess multiple variant types, including structural and copy-number variants (CNVs), short tandem repeats (STRs) and mitochondrial variants, in a single test and the shorter sample preparation times will further decrease time to result. Despite these advantages, adoption is hampered by higher costs, at times immature analysis tools and lack of robust evidence for a significant increase in diagnostic yield[11,12]. In addition, the improved analytical performance of WGS and ever earlier test initiation driven by rapid diagnosis programs exacerbate existing interpretive challenges. Improvements in bioinformatic analysis and integration of multi-omic approaches will further optimize diagnostic performance, as there is a growing appreciation for the need to closely integrate discovery research with clinical testing to maximize diagnostic benefits for current and future patients[13]. However, these approaches rarely form part of current diagnostic practice and are typically the domain of specialized research programs, limiting access.

In the present study, we expanded our rapid genomic diagnosis program to a national scale in a prospectively ascertained cohort of critically ill infants and children with rare disease. In addition, we assessed the diagnostic performance of WGS and the impact of systematically integrating additional analysis types, transcriptome analysis and functional assays.

Results

Participant demographics and indications for testing
A total of 333 infants and children were referred to the Acute Care Genomics program. The panel deemed 26 ineligible and a further 17 were withdrawn by the referring team after approval (Fig. 1). Of the 290 participants, 135 (47%) were karyotypically female; the median age was 29 d (range 0 d to 17 years); and 64% presented with symptoms at birth (Fig. 2c).

Trio analysis was performed in 273 families (94%), with the remaining 17 (6%) proceeding as duos due to parental unavailability. Patients admitted to neonatal intensive care units (NICUs) accounted for 47% of the cohort (135 of 290), 39% (112) were admitted to pediatric intensive care units (PICUs) and 15% (43) were other critically unwell hospital patients (for example, awaiting transplant). Genetic counselors and/or clinical geneticists provided pre-test counseling to 286 of 290 families (99%). The cohort was ethnically and clinically diverse, with 694 unique Human Phenotype Ontology (HPO) terms recorded at an average of five per patient (Fig. 2d,e)[14,15]. The commonest reasons for referral were neurological disorders (hypotonia and seizures) and complex multi-system disorders indicative of either syndromic or metabolic disease.

Time to clinical WGS report
The average laboratory turnaround time as measured from receipt of all samples to clinical report was 2.9 d (95% CI 2.85–2.99), with a fastest time to result of 45 h. All reports were issued within 5 calendar days of sample receipt (Fig. 3a).

Diagnostic yield of standard WGS analysis
Of the 290 individuals tested, 137 were diagnosed during initial clinical analysis. The majority of diagnoses (127 of 137; 93%) were in genes contained in the virtual panels assigned by clinicians and 10 (7%) were from panel-agnostic analyses. Of note, eight diagnoses (6%) were due to parentally inherited variants in genes associated with dominant

conditions (*SHH*, *FOXF1*, *SCN5A* X2, *FGFR3*, *RANBP2*, *BAG3* and *ENG*), where the parents were not recognized as affected before testing. Of these, five parents were subsequently recognized as clinically affected (*FGFR3*, *SCN5A* X2, *BAG3* and *ENG*), with important implications for further cascade testing within the family; one parent was found to be mosaic (*SHH*, 20% mosaicism). Parental transmission of *FOXF1* and *RANBP2* variants on the other hand, are thought to be explained by a parent-of-origin effect[16] and incomplete penetrance[17], respectively. Eight individuals received a partial diagnosis, whereas another five had a dual diagnosis (Supplementary Tables 1–3). Diagnostic yield was highest in those presenting with isolated renal, dermatological and hematological phenotypes and lowest in those with unexplained interstitial lung disease (Fig. 3d). Nineteen genes accounted for diagnoses in multiple individuals (Fig. 3d), with all other diagnoses being unique, highlighting the wide diversity of diagnoses encountered in this clinical setting.

Diagnostic yield of extended bioinformatic analysis
A triplet repeat expansion in *DMPK* was identified using the STRipy tool[18] in an infant with hypotonia (A0131129). STRipy detected an expanded allele predicted to contain over 200 repeats but was unable to accurately size the expansion. Clinically accredited testing confirmed the presence of a ~900 repeat allele in the proband, consistent with a diagnosis of congenital myotonic dystrophy and 48 repeats (premutation range) in the mother, who was not clinically affected. Paternal allele sizes were normal. Nanopore long-read sequencing in the proband indicated expansion size of 1,048 repeats.
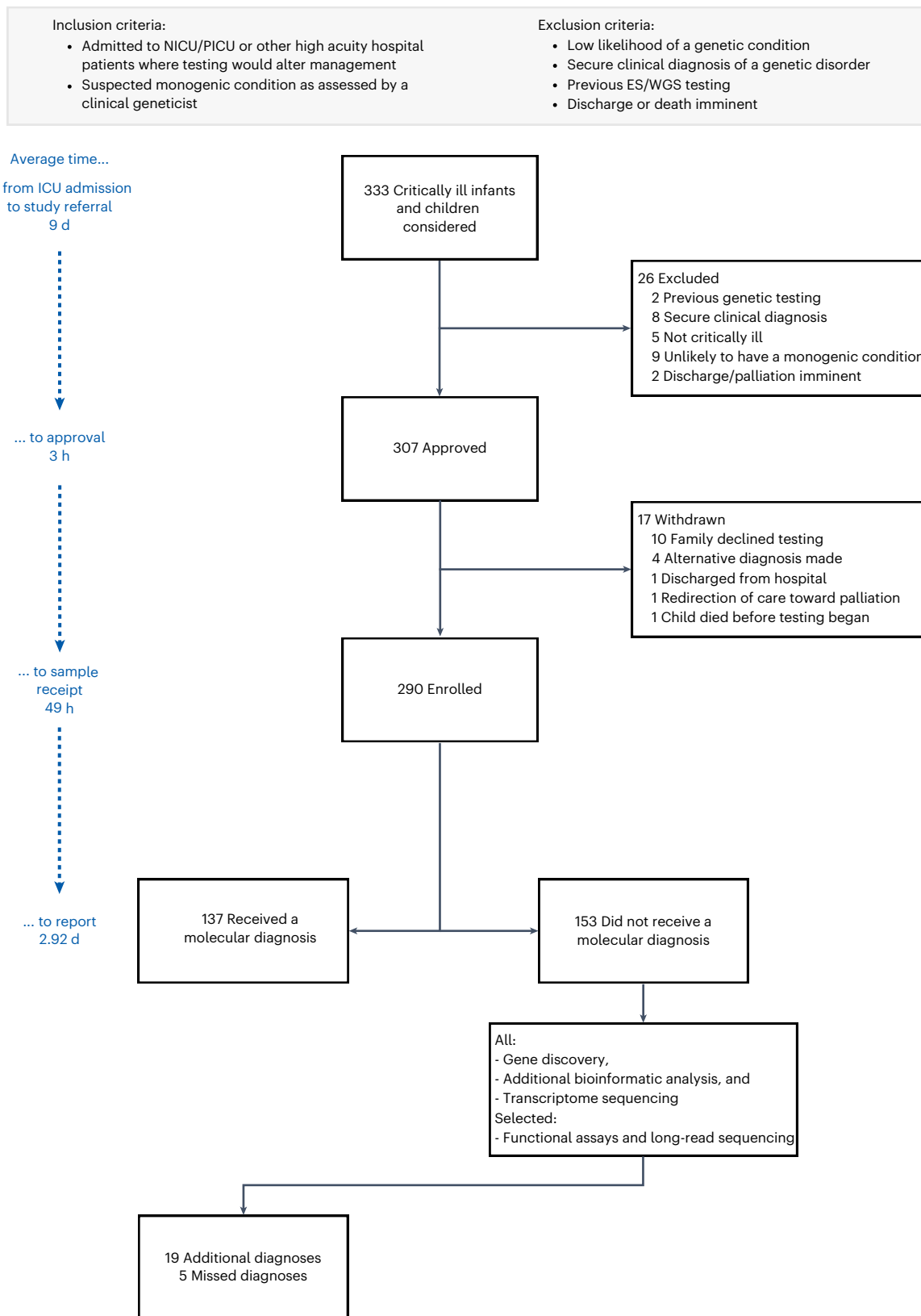
Analysis using Manta and Schism for complex structural variants indicated a large, up to 2.6-kb, intronic insertion variant, likely derived from a SINV-VNTR-Alus (SVA) retrotransposable element, in the last intron of *MECP2* in a male child with microcephaly, developmental delay and epileptic encephalopathy (A0131084). The presence and origin of the insertion were confirmed using Nanopore long-read sequencing, demonstrating that the full 2.5-kb SVA sequence is present in the last intron of *MECP2* (Fig. 4b). Testing of the proband and the parents using long-range PCR also confirmed this to be a de novo event. Manual analysis of RNA data determined that the intronic insertion disrupted normal splicing; however, an estimated 66% of wild-type *MECP2* splicing remained, consistent with reduced clinical severity. Of note, the overall expression of *MECP2* was not reduced, nor was our independent bioinformatic splicing analysis able to detect this event, highlighting the value of multiple approaches to identify pathogenic variants.

Transcriptome yield
RNA-sequencing data were successfully generated from 335 individuals (115 patients and their parents), resulting in three new diagnoses.

In A0131122, presenting with features of skeletal dysplasia, a known pathogenic heterozygous promoter variant in *RMRP* was identified (NR_003051.3(*RMRP*):n.-5_-4insAACTACTCTGTGAAGCTGA), consistent with a diagnosis of cartilage-hair hypoplasia. This was confirmed to reduce expression of the affected allele in the proband and paternal sample. A second single-nucleotide insertion variant, initially of uncertain significance, was identified in the maternal allele (NR_003051.3(*RMRP*), n.93_94insA). Analysis of maternal RNA data showed this insertion conferred strong (>99%) preferential expression from the wild-type allele, indicating instability of the variant transcript. In combination with the paternal pathogenic variant and a 9.2-fold downregulation of *RMRP* in the proband, this was considered sufficient evidence to ascribe pathogenicity to the maternal insertion, securing the diagnosis.

In A0831013, presenting with neonatal hypotonia, a new 403-bp hemizygous, maternally inherited deletion spanning parts of the gene promoter and 5′ UTR of *MTM1* was identified, including the transcription start site (TSS) (NM_000252.2(*MTM1*): c.-76_-11del), consistent with a diagnosis of myotubular myopathy. Analysis of RNA

Inclusion criteria:
- Admitted to NICU/PICU or other high acuity hospital patients where testing would alter management
- Suspected monogenic condition as assessed by a clinical geneticist

Exclusion criteria:
- Low likelihood of a genetic condition
- Secure clinical diagnosis of a genetic disorder
- Previous ES/WGS testing
- Discharge or death imminent

Average time...

... from ICU admission to study referral
9 d

333 Critically ill infants and children considered

26 Excluded
    2 Previous genetic testing
    8 Secure clinical diagnosis
    5 Not critically ill
    9 Unlikely to have a monogenic condition
    2 Discharge/palliation imminent

... to approval
3 h

307 Approved

17 Withdrawn
    10 Family declined testing
    4 Alternative diagnosis made
    1 Discharged from hospital
    1 Redirection of care toward palliation
    1 Child died before testing began

... to sample receipt
49 h

290 Enrolled

... to report
2.92 d

137 Received a molecular diagnosis

153 Did not receive a molecular diagnosis

All:
- Gene discovery,
- Additional bioinformatic analysis, and
- Transcriptome sequencing
Selected:
- Functional assays and long-read sequencing

19 Additional diagnoses
5 Missed diagnoses

Fig. 1 | Recruitment workflow for the Acute Care Genomics program. Critically ill infants and children with suspected genetic conditions were proposed to a national panel of experts, with those approved undergoing ultra-rapid WGS. Additional bioinformatic analyses and transcriptome sequencing were performed in all undiagnosed patients. Long-read sequencing and functional assays were deployed in selected cases. ICU, intensive care unit.

data confirmed reduction in *MTM1* expression in the male proband, effectively abolishing gene expression with no evidence of alternative TSS utilization.

The *GLB1* variant in A1031002 was a new homozygous exon 7 splice donor region variant (NM_000404.3(*GLB1*): c.733+6 T > C) predicted to cause a splice defect, which would result in a premature stop codon
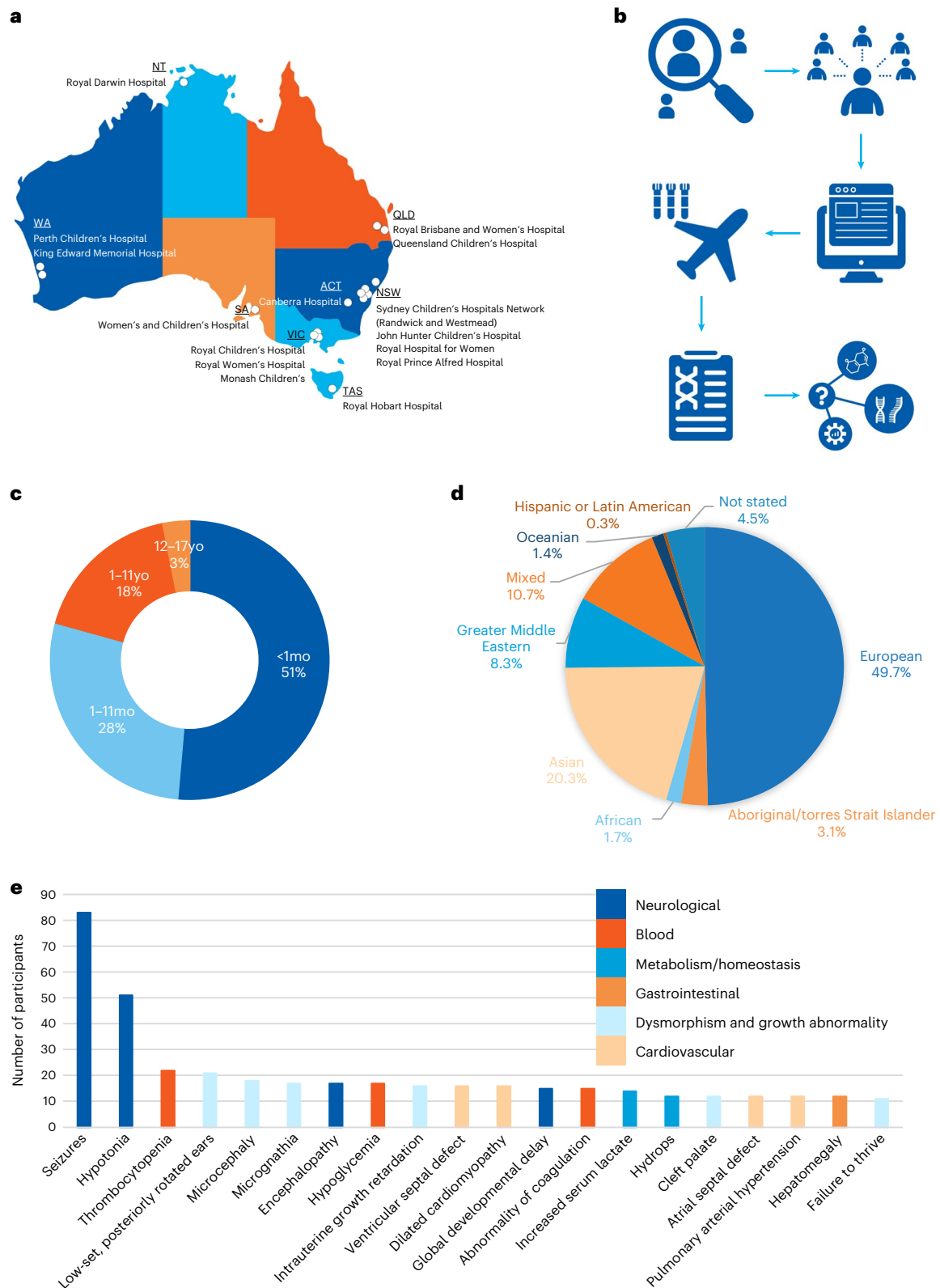
Fig. 2 | Patient recruitment and key characteristics. a, Recruitment sites. NT, Northern Territory; QLD, Queensland; WA, Western Australia; ACT, Australian Capital Territory; NSW, New South Wales; SA, South Australia; VIC, Victoria; TAS, Tasmania. **b**, Study workflow, including patient selection using guidelines and virtual expert panel; electronic resources to support test ordering and consent; sample shipping; diagnostic reporting; and extended analysis and multi-omic approaches in unsolved cases. **c**, Age of study participants. **d**, Ancestry/ethnicity of participants. **e**, Twenty most common HPO terms, coded by major groups.

and nonsense-mediated decay. Pathogenic variants in *GLB1* are associated with GM1 gangliosidosis; however, the patient's phenotype was milder than expected. Analysis of RNA data indicated exon skipping with residual wild-type transcription in the proband, consistent with the milder clinical presentation. Confirmation of a pathogenic splice variant determined access to a clinical trial.
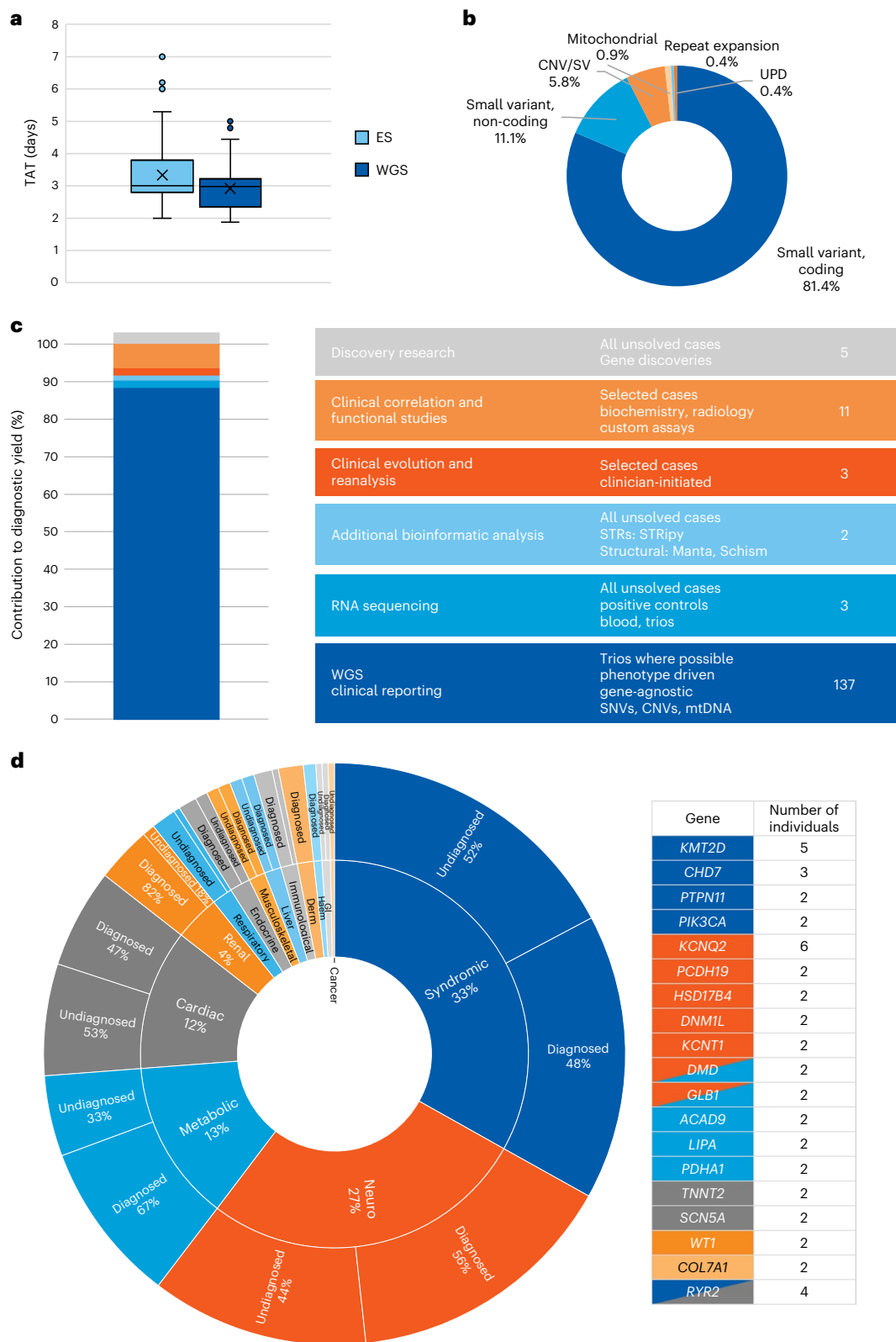
Fig. 3 | Summary of diagnostic outcomes. a, Time to clinical report for ultra-rapid WGS cohort ($n = 290$), compared to previous ultra-rapid ES cohort ($n = 108$). X represents the mean; central line represents the median; top and bottom edges of the boxes are the first and third quartiles; the whiskers show the minima to maxima no greater than $1.5\times$ the interquartile range with remaining outliers plotted individually. TAT, turnaround time. **b**, Variant types detected by WGS. **c**, Incremental gain in diagnostic yield from extended analysis and multi-omic approaches. CNV, copy number variant; SV, structural variant; UPD, uniparental disomy. **d**, Sunburst representing the spectrum of diagnoses. Arranged by number of patients, clockwise, the inner ring represents the principal clinical presentation and the second ring represents diagnostic yield in each group. Genes responsible for diagnoses in multiple individuals represented in the adjacent table, color-coded by principal clinical presentation. Full names of each disorder and Online Mendelian Inheritance in Man numbers are included in Supplementary Tables 1–3.
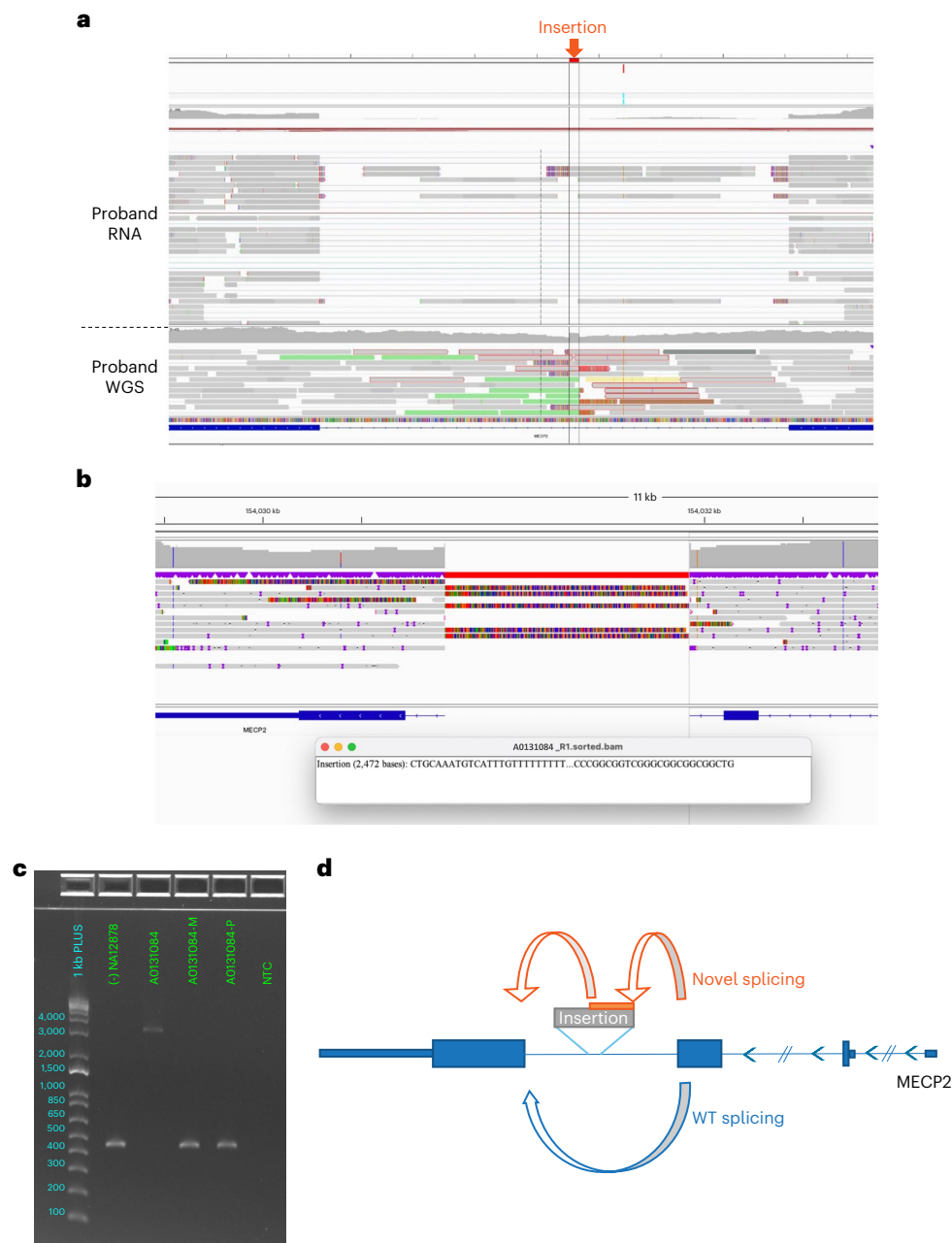
Fig. 4 | Identification and confirmation of intronic insertion variant, likely derived from an SVA retrotransposable element, in the last intron of *MECP2*. a, Integrated Genomics Viewer (IGV) proband RNA (top) and DNA (bottom) short-read sequencing data indicating the presence of a DNA insertion resulting in the inclusion of a pseudo-exon in the last intron of *MECP2*. **b**, Nanopore sequencing data demonstrating the insertion. **c**, PCR gel electrophoresis of relevant *MECP2* region in the proband (A0131084), parents (A0131084-M and A0131084-P) and a control sample (NA12878), consistent with an insertion of approximately 2.6 kb. This clinically accredited assay was performed once. **d**, Schematic of the observed splicing outcomes of *MECP2* in the proband with the presence of a transposon-derived pseudo-exon, residual canonical splicing, skipping of the penultimate exon and intronic read-through. WT, wild-type.

Confirmed splice variants in another seven samples were included as positive controls, with a splice defect confirmed in two of these, whereas five had insufficient expression in blood to be informative. This included the confirmation of an in-frame exon skipping event caused by the NM_005188.3(*CBL*): c.1096-2 A > T canonical splice variant, in a child with Moyamoya disease and developmental delay. Consistent with the known mechanism of disease[19,20], this splice variant resulted in the upregulation of *CBL*, which was detected in our expression outlier analysis, highlighting the need to carefully assess upregulated outliers.

Clinical and functional correlation

Thirty variants of uncertain significance (VUSs), deemed highly likely to be clinically relevant, were reported. One individual underwent specialized clinical re-review including international consultation to establish variant pathogenicity[21]. In ten individuals, additional functional studies were performed (Table 1). This consisted of a range of clinically accredited assays in seven individuals and bespoke research assays in three.

NUP214. We identified *NUP214* biallelic variants in A1131048; one variant (NM_005085.3(*NUP214*): c.112 C > T; p.(Arg38Cys)) was

Table 1 | Individuals with variants of uncertain significance (VUS) where additional functional validation resulted in variant reclassification

Study ID	Presenting clinical features	Gene variant(s), original classification	Additional studies	Final variant classification(s), time to report
Clinically accredited tests				
A3231005	Hydrops	*GLB1* c.130 G > T,p.(Asp44Tyr), VUS	Enzyme studies	LP 5 days
A0831018	Immunological dysfunction, bone marrow failure	*PNP* c.97 T > C, p.(Ser33Pro), VUS	Enzyme studies	LP 5 days
A3331002	Seizures	*HSD17B4* c.1132 G > A, p.(Gly378Arg), hmz, VUS	Very-long-chain fatty acids	LP 12 days
A2131008	Hydrops, dilated cardiomyopathy, thrombocytopenia	*TBX19* c.666-2 A > T, VUS	Endocrine testing	LP 17 days
A0831021	Lactic acidosis, central hypotonia	*GTPBP3* c.521 G > C, p.(Arg174Pro), hmz, VUS	Respiratory chain enzymes	LP 7 weeks
A0731002	Severe intrauterine growth restriction	*RNU4ATAC* n.18 G > A, VUS n.50 G > A, LP	Bone dysplasia specialist review of skeletal survey	LP, LP 2 months
A0431022	Neurological deterioration	*PDHA1* c.1045 G > A, p.(Ala349Thr), VUS	Enzyme studies	LP 4 months
Research assays				
A1431031	Severe lactic acidosis	*ACAD9* c.1636 G > A, p.(Val546Met), LP and c.1376_1381delins, p.(Lys459_Ser461delinsThrCys) VUS	Western blot[23]	P, LP 4 months
A1131048	Encephalopathic episode	*NUP214* c.112 C > T; p.(Arg38Cys) P and c.929 T > C; p.(Ile310Thr) VUS	Western blot, nuclear pore quantification, quantitative proteomics	P, LP 6 months
A0131063	Recurrent liver failure	*NBAS* c.2951 T > G, p.(Ile984Ser), VUS and c.406 A > G, p.(Arg136Gly), VUS	Western blot and p31 level[22]	P, LP 10 months

hmz, homozygous; P, pathogenic; LP, likely pathogenic.

previously reported and was classified as pathogenic[22]. The second variant (NM_005085.3(*NUP214*): c.929 T > C; p.(Ile310Thr)) was new, absent in gnomAD and predicted to be damaging by multiple in silico tools. Western blot of proteins from patient fibroblasts demonstrated a significant decrease ($P < 0.05$) in NUP214 levels compared to controls (Fig. 5a,b). Quantification of NUP214-containing pores in the nuclear region by super-resolution microscopy imaging showed a significant reduction ($P < 0.0001$) in the NUP214-containing nuclear pore density in the nuclear envelope compared to controls (Fig. 5c,d), confirming previous reports[22]. Quantitative proteomics confirmed decreased NUP214 protein levels (Fig. 5e and Supplementary Table 4) and revealed a decrease in two other nucleoporins; POM121C and NUP88, the latter a physical interactor of NUP214 within the human nuclear pore complex (Fig. 5f) and previously reported as reduced in patients with *NUP214* pathogenic variants[22]. Last, we observed a decrease in viability of patient fibroblasts after a 2-h heat shock, validating previous reports[22], without changes in the apoptotic response to heat stress. These multiple lines of additional evidence to support pathogenicity of the *NUP214* variants in A1131048 were used to reclassify the second variant as likely pathogenic.

Reanalysis

Three variants were reported as diagnostic following reanalysis prompted by evolving clinical features. A0731004 initially presented with multiple congenital anomalies, but subsequently developed seizures at 5 months of age, leading to the reporting of a de novo new missense variant in *GABRB3*. A0131108 initially presented in infancy with focal seizures, remained undiagnosed but developed nystagmus, resulting in the reporting of a pathogenic variant in *FRMD7*. Finally,

infant A0131089 presented with neonatal seizures and had a pathogenic de novo variant identified in *KCNQ2*, as well as a maternally inherited variant of uncertain significance in *GABRG2*. The *GABRG2* variant was upgraded as pathogenic 2 years later based on family history, evolving clinical presentation and newly published data.

Gene discovery

Ten gene candidates were submitted to GeneMatcher. Five matched to multiple other similarly affected individuals and are currently being pursued to establish gene–disease relationships.

Diagnoses not obtained by WGS

Five individuals had uninformative WGS and were subsequently diagnosed through orthogonal genetic testing. In three, the causative variants were mosaic, one each in the *KCNT1*, *PIK3CA* and *IKBKG* genes. In each of these cases, the relevant variants were filtered from the analysis as failing quality parameters (the minimum variant allele threshold of ≥15% of reads with the variant). The *PIK3CA* variant was detected in a fibroblast sample (mosaicism level 21%), whereas the *KCNT1* variant was detected on exome sequencing (mosaicism level 16%). The mosaic *IKBKG* variant was detected using Sanger sequencing (mosaicism level not determined). Review of the available trio WGS data identified that the variant was maternally inherited, with the apparent mosaicism in the proband reflecting somatic reversion.

A further case was diagnosed by microarray with uniparental isodisomy of chromosome 15 (UPD), causing Prader–Willi syndrome. Notably, RNA-sequencing data were also able to identify this diagnosis based on the lack of expression of *SNHG14* and *SNRPN*, which are only expressed from the paternally inherited allele, illustrating the scope
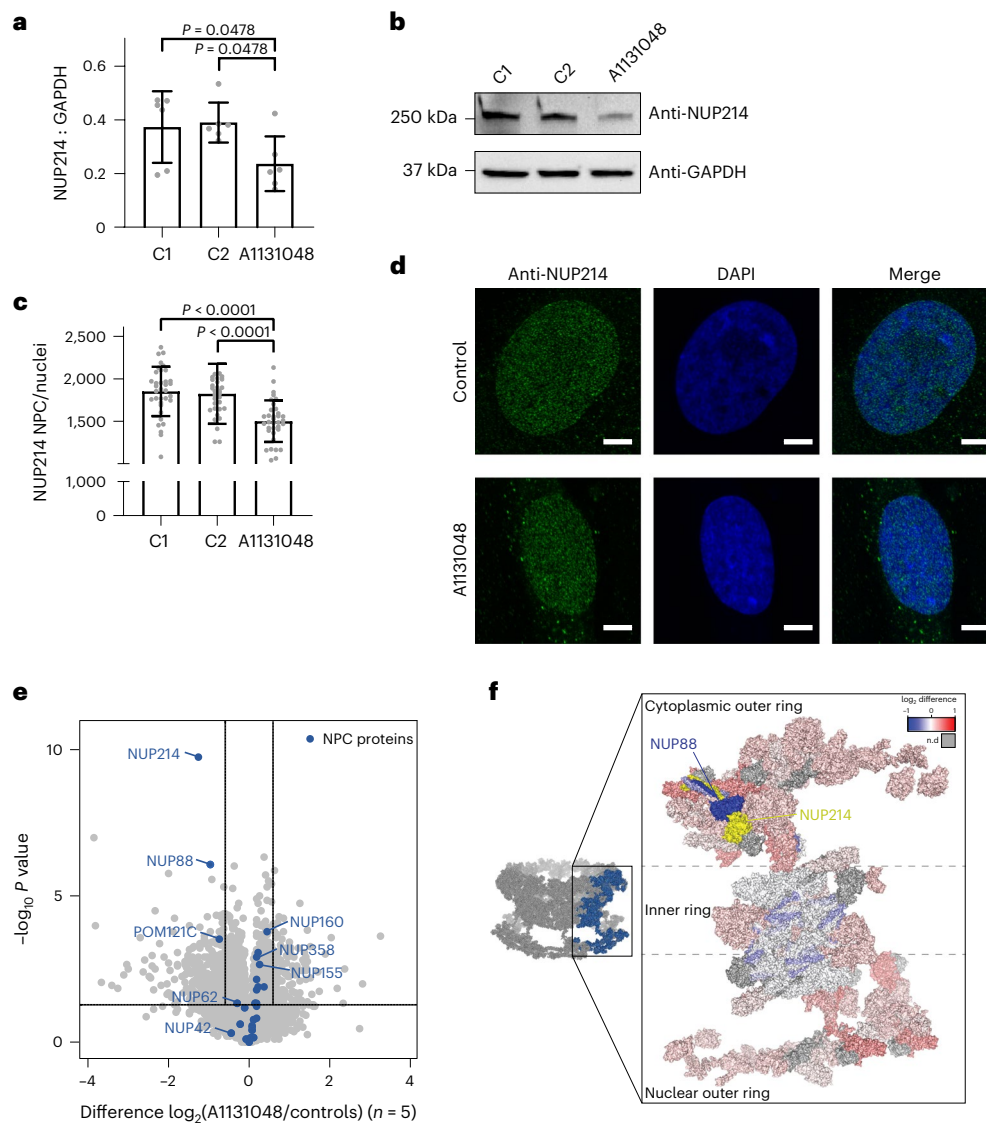
Fig. 5 | Decreased NUP214 steady-state levels and decreased NUP214-containing nuclear pore complex density in fibroblasts from A1131048.
a, Densitometry shows significantly reduced NUP214 levels in fibroblasts compared to controls. $n = 6$ biological samples per cell line. Data represent mean ± s.d. One-way analysis of variance (ANOVA) with Holm–Sidak's multiple comparisons test. **b**, Representative western blot of NUP214 in fibroblasts from A1131048 and two controls (C1 and C2). **c**, Quantification of NUP214-containing nuclear pore complex pore density (pores per nucleus) in fibroblasts from A1131048 compared to controls shows a significant decrease in fibroblasts from A1131048. $n = 3$ biological samples per cell line, from three independent experiments, represented is pooled data from from $n = 35$ for C1, $n = 31$ for C2 and $n = 35$ for A1131048. Data represent mean ± s.d. One-way ANOVA with Holm–Sidak's multiple comparisons test. **d**, Compressed Z-stack representative

images of NUP214 immunostaining (green spots) and nucleus (4,6-diamidino-2-phenylindole (DAPI); blue) in fibroblasts from A1131048 and controls. Scale bars, 5 μm. Images are representative from three experiments. **e**, Volcano plot showing protein abundances in the fibroblasts from A1131048 relative to healthy controls ($n = 5$). Nuclear pore complex (NPC) components are indicated in blue. The horizonal line represents $P = 0.05$ and the vertical lines represent fold changes of ±1.5. Data were derived from a two-sided Student's t-test. No adjustments were made for multiple comparisons. **f**, Topographical heat map showing fold changes of NPC proteins identified by proteomics mapped onto the structure of the NPC cytosolic face (Protein Data Bank, 7TBL). Yellow indicates NUP214 subunit. Other subunits, including NUP88, which is coiled around NUP214, are colored according to their fold change relative to controls, as indicated in the inset; gray indicates no data.

---

of RNA-sequencing data to identify dysregulation of imprinted genes. The final missed diagnosis had a polyalanine repeat expansion in the *PHOX2B* gene, a well-described variant causing congenital central hypoventilation syndrome, missed due to lack of read coverage in the responsible low-complexity gene region.

Clinical utility
Of the 156 individuals diagnosed through WGS, 120 (77%) had changes in critical care management as reported by treating clinicians (Supplementary Tables 1 and 2). Many of these related to improvements in the

overall process of care through rationalizing investigations, referrals and treatments. In 94 patients (60%), results had major implications for management, including informing precision treatments; surgical and transplant decisions; and palliation. Enzyme replacement therapy was commenced in three individuals and six infants were diagnosed with *KCNQ2*-related epileptic encephalopathy, informing choice of anti-epileptic agent. Ten diagnoses (6%) informed decisions regarding transplantation and 35 diagnoses (22%) formed part of decisions to redirect care toward palliation. As an example of a diagnosis with major implications for clinical management, A1131034 presented in

adolescence with unexplained renal failure and was diagnosed with Frasier syndrome. This prompted further imaging, which revealed dysplastic gonads and identification of bilateral gonadoblastomas on biopsy. The individual underwent successful surgery to resect these.

Discussion

In this prospectively ascertained national cohort, we demonstrate diagnostic benefits of using ultra-rapid WGS as a first-tier test, due to a combination of increased analytical performance and use of multi-omic approaches to improve variant interpretation. We also increased robustness of the national approach by implementing ultra-rapid data sharing, analysis and reporting in a second diagnostic laboratory, while expanding clinical recruitment to all Australian states and territories. We provide detailed data on diagnostic yields in different patient groups to guide future implementation.

Laboratory testing times for WGS were not significantly different from those in our previously described cohort using exome sequencing (ES)[3], with an average time from sample receipt to report of 2.9 d (95% CI 2.85–2.99) for WGS and 3.1 (95% CI 2.98–3.21) for ES. While sample preparation times were shortened with the removal of the exome enrichment step, average bioinformatic processing time increased due to increased data volumes and addition of CNV calling, highlighting this as an area for future improvement. The range of variant types identified using WGS was much broader, including CNVs ranging in size from 57 kb to a ring chromosome, deep intronic and regulatory variants, mitochondrial DNA variants, STRs and a transposon insertion. Six diagnoses were due to a combination of a small coding variant with a CNV, further highlighting the utility of WGS to combine analysis for multiple variant types in a single test, shortening time to diagnosis.

It is also clear that the analytical potential of WGS currently outstrips our ability to robustly detect, validate and classify many variant types under accredited conditions, with additional testing modalities critical in demonstrating downstream effects. This was most notably exemplified by the male proband with atypical features of Rett syndrome caused by a retrotransposon insertion causing splice defect in *MECP2*. This diagnosis was made using a combination of extensive, cutting-edge bioinformatic analysis, trio RNA-sequencing and orthogonal validation by established molecular techniques.

Transcriptome analysis was also instrumental in securing three additional diagnoses. In these cases, both aberrant splicing and reduction of expression were identified, with a mixture of variant types, including non-canonical splice site, promoter deletion and expression changes in a noncoding gene. Availability of parental RNA data was critical for resolution in at least one case. We elected to perform RNA-sequencing using the original whole blood samples submitted for WGS to reduce burden on families and clinicians. The additional diagnostic yield obtained from RNA-sequencing in this cohort (2.5%) is lower than reported in other studies (7–36%) (refs. [23–26]), which likely reflects the inclusion of all, rather than selected, unsolved patients and limitations of tissue-specific expression. Despite these challenges, our approach of using trio transcriptome analysis on RNA from blood samples has proven both feasible and useful and we will further explore implementation in parallel with ultra-rapid WGS.

The largest gain in diagnostic yield was from pursuing clinical and functional correlation of VUSs identified by WGS analysis. This included radiological, metabolic and immunological tests that are clinically available through to custom-designed studies, such as western blots, ultra-resolution immunofluorescent imaging and quantitative proteomics. Achieving timely and systematic functional validation remains challenging, with the range of assays currently clinically available focusing on specific metabolic and immunological disorders. As genome-wide tests become increasingly incorporated at the start of the diagnostic pathway, there will be a growing need for high-throughput, gene-agnostic functional assays to resolve VUSs. Such functional assays need to be designed so that the results can be incorporated

into diagnostic reporting using established criteria[27]. Many of these assays will need to transition from research to clinical laboratories to scale up and ensure appropriate test funding, validity, reproducibility and timeliness of results.

There is ongoing debate about the relative advantages of clinically focused and gene-agnostic analysis approaches. We used a combination of the two, with 7% of diagnoses achieved outside of the virtual gene panels nominated by clinicians. By contrast, analysis outside of the 'Mendeliome' virtual gene panel did not reveal any additional diagnoses but yielded ten gene candidates, which were submitted to GeneMatcher[28]. We anticipate that several of these will be confirmed as disease genes, with the same approach applied in our previously published ES cohort (*n* = 108; 2018 to 2019) having yielded six gene discoveries to date[29–33]. Identifying gene candidates has traditionally been the domain of dedicated research programs[34–36], with selected patients entering these at the end of the diagnostic trajectory. With genomic testing increasingly performed through clinical pathways, identifying gene candidates should form an integral part of diagnostic analysis. Several diagnostic laboratories have now reported their experience of systematically contributing to gene discovery[37–39] and while this requires additional resourcing, wider implementation will both reduce the time to gene discovery and expand the effort.

We are aware of five diagnoses that were missed by WGS, highlighting some important limitations and the need to consider alternative testing modalities when clinical suspicion of a specific disorder is high. Three were due to mosaicism and while we have amended our practice to review known pathogenic variants failing quality checks, detection of mosaicism will continue to be challenging due to the relatively lower coverage of WGS compared to exome or targeted sequencing. This limitation is only likely to be overcome once reduction in the cost of WGS data generation allows deeper sequencing. Deeper sequencing of variants in low-complexity regions would also have improved the chance of detecting the missed polyalanine repeat variant in *PHOX2B*. The missed UPD, as well as the identification of an STR variant and a complex structural variant using unaccredited bioinformatic analysis in this cohort, highlight the need to continuously improve and expand the range of analysis tools used in the diagnostic setting. Some of these challenges may be overcome by increased use of long-read sequencing technology. While we used Nanopore as a post hoc validation tool, long-read sequencing as first-line diagnostic testing would have substantially reduced the time to diagnosis in both the repeat expansion and the transposon insertion cases and may have identified other complex variants. The feasibility of using long-read sequencing in the acute setting has recently been demonstrated and holds the promise of dramatically shortening turnaround times and potentially decentralizing sequencing and interpretation[6,40,41]. From a broader implementation perspective, the service delivery model in our study, which included patient selection by clinical geneticists and centralized sequencing, may only be applicable to a subset of healthcare systems. While a substantial body of work has now accumulated about the utility of rapid genomic testing in the acute setting, long-term outcome data are lacking and this remains a major limitation of this and other studies.

In this prospectively ascertained national cohort undergoing ultra-rapid WGS, we demonstrate scalability of the model and highlight gains in diagnostic yield by rapidly integrating genomic, transcriptomic and proteomic data. There is a need to systematically integrate these approaches as part of diagnostic pathways to fully realize the potential of genomic sequencing to alter outcomes in rare disease patients and families in a timely manner.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-023-02401-9.

References

1. Stark, Z. et al. Integrating genomics into healthcare: a global responsibility. *Am. J. Hum. Genet.* **104**, 13–20 (2019).
2. Clark, M. M. et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom. Med.* **3**, 16 (2018).
3. Lunke, S. et al. 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 (2020).
4. Dimmock, D. et al. Project Baby Bear: rapid precision care incorporating rWGS in 5 California children's hospitals demonstrates improved clinical outcomes and reduced costs of care. *Am. J. Hum. Genet.* **108**, 1231–1238 (2021).
5. Stark, Z. & Ellard, S. Rapid genomic testing for critically ill children: time to become standard of care? *Eur. J. Hum. Genet.* **30**, 142–149 (2022).
6. Gorzynski, J. E. et al. Ultrarapid nanopore genome sequencing in a critical care setting. *N. Engl. J. Med.* **386**, 700–702 (2022).
7. Goranitis, I. et al. Is faster better? An economic evaluation of rapid and ultra-rapid genomic testing in critically ill infants and children. *Genet. Med.* https://doi.org/10.1016/j.gim.2022.01.013 (2022).
8. Kingsmore, S. F. 2022: a pivotal year for diagnosis and treatment of rare genetic diseases. *Cold Spring Harb. Mol. Case Stud.* **8**, a006204 (2022).
9. Smedley, D. et al. 100,000 Genomes pilot on rare-disease diagnosis in health care - preliminary report. *N. Engl. J. Med.* **385**, 1868–1880 (2021).
10. Stranneheim, H. et al. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med.* **13**, 40 (2021).
11. Kingsmore, S. F. et al. A randomized, controlled trial of the analytic and diagnostic performance of singleton and trio, rapid genome and exome sequencing in ill infants. *Am. J. Hum. Genet.* **105**, 719–733 (2019).
12. Brockman, D. G. et al. Randomized prospective evaluation of genome sequencing versus standard-of-care as a first molecular diagnostic test. *Genet Med.* **23**, 1689–1696 (2021).
13. Seaby, E. G., Rehm, H. L. & O'Donnell-Luria, A. Strategies to uplift novel Mendelian gene discovery for improved clinical outcomes. *Front. Genet.* **12**, 674295 (2021).
14. Morales, J. et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS catalog. *Genome Biol.* **19**, 21 (2018).
15. Robinson, P. N. et al. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* **83**, 610–615 (2008).
16. Alsina Casanova, M. et al. Maternal mutations of FOXF1 cause alveolar capillary dysplasia despite not being imprinted. *Hum. Mutat.* **38**, 615–620 (2017).
17. Neilson, D. E. et al. Infection-triggered familial or recurrent cases of acute necrotizing encephalopathy caused by mutations in a component of the nuclear pore, RANBP2. *Am. J. Hum. Genet.* **84**, 44–51 (2009).
18. Halman, A., Dolzhenko, E. & Oshlack, A. STRipy: a graphical application for enhanced genotyping of pathogenic short tandem repeats in sequencing data. *Hum. Mutat.* **43**, 859–868 (2022).
19. Martinelli, S. et al. Heterozygous germline mutations in the CBL tumor-suppressor gene cause a Noonan syndrome-like phenotype. *Am. J. Hum. Genet.* **87**, 250–257 (2010).
20. Niemeyer, C. M. et al. Germline CBL mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia. *Nat. Genet.* **42**, 794–800 (2010).
21. Smallwood, K. et al. POLR1A variants underlie phenotypic heterogeneity in craniofacial, neural, and cardiac anomalies. *Am. J. Hum. Genet.* **110**, 809–825 (2023).
22. Fichtman, B. et al. Pathogenic variants in NUP214 cause 'plugged' nuclear pore channels and acute febrile encephalopathy. *Am. J. Hum. Genet.* **105**, 48–64 (2019).
23. Gonorazky, H. D. et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019).
24. Murdock, D. R. et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J. Clin. Invest.* **131**, e14500 (2021).
25. Lee, H. et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med.* **22**, 490–499 (2020).
26. Maddirevula, S. et al. Analysis of transcript-deleterious variants in Mendelian disorders: implications for RNA-based diagnostics. *Genome Biol.* **21**, 145 (2020).
27. Brnich, S. E. et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* **12**, 3 (2019).
28. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).
29. Stolz, J. R. et al. Clustered mutations in the GRIK2 kainate receptor subunit gene underlie diverse neurodevelopmental disorders. *Am. J. Hum. Genet.* **108**, 1692–1709 (2021).
30. Cooper, M. S., Stark, Z., Lunke, S., Zhao, T. & Amor, D. J. IREB2-associated neurodegeneration. *Brain* **142**, e40 (2019).
31. Lee, R. G. et al. Deleterious variants in CRLS1 lead to cardiolipin deficiency and cause an autosomal recessive multi-system mitochondrial disease. *Hum. Mol. Genet.* https://doi.org/10.1093/hmg/ddac040 (2022).
32. Rehman, A. U. et al. Biallelic loss of function variants in PPP1R21 cause a neurodevelopmental syndrome with impaired endocytic function. *Hum. Mutat.* **40**, 267–280 (2019).
33. Amarasekera, S. S. C. et al. Multi-omics identifies large mitoribosomal subunit instability caused by pathogenic MRPL39 variants as a cause of pediatric onset mitochondrial disease. *Hum. Mol. Genet.* https://doi.org/10.1093/hmg/ddad069 (2023).
34. Cloney, T. et al. Lessons learnt from multifaceted diagnostic approaches to the first 150 families in Victoria's Undiagnosed Diseases Program. *J. Med. Genet.* **59**, 748–758 (2022).
35. Osmond, M. et al. Outcome of over 1500 matches through the Matchmaker Exchange for rare disease gene discovery: the 2-year experience of Care4Rare Canada. *Genet. Med.* **24**, 100–108 (2022).
36. Baxter, S. M. et al. Centers for Mendelian Genomics: a decade of facilitating gene discovery. *Genet. Med.* **24**, 784–797 (2022).
37. McWalter, K., Torti, E., Morrow, M., Juusola, J. & Retterer, K. Discovery of over 200 new and expanded genetic conditions using GeneMatcher. *Hum. Mutat.* **43**, 760–764 (2022).
38. Taylor, J. P. et al. A clinical laboratory's experience using GeneMatcher-building stronger gene–disease relationships. *Hum. Mutat.* **43**, 765–771 (2022).
39. Towne, M. C. et al. Diagnostic testing laboratories are valuable partners for disease gene discovery: 5-year experience with GeneMatcher. *Hum. Mutat.* **43**, 772–781 (2022).
40. Goenka, S. D. et al. Accelerated identification of disease-causing variants with ultra-rapid nanopore genome sequencing. *Nat. Biotechnol.* **40**, 1035–1041 (2022).

41. Lunke, S. & Stark, Z. Can rapid nanopore sequencing bring genomic testing to the bedside? *Clin. Chem.* https://doi.org/10.1093/clinchem/hvac111 (2022).

[1]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [2]Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Victoria, Australia. [3]Australian Genomics, Melbourne, Victoria, Australia. [4]Genetic Health Queensland, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia. [5]Sydney Children's Hospitals Network – Westmead, Sydney, New South Wales, Australia. [6]Children's Hospital Westmead Clinical School, University of Sydney, Sydney, New South Wales, Australia. [7]Sydney Children's Hospitals Network – Randwick, Sydney, New South Wales, Australia. [8]Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia. [9]Monash Genetics, Monash Health, Melbourne, Victoria, Australia. [10]Department of Paediatrics, Monash University, Melbourne, Victoria, Australia. [11]Paediatric and Reproductive Genetics Unit, Women's and Children's Hospital, North Adelaide, South Australia, Australia. [12]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia. [13]Adelaide Medical School, The University of Adelaide, Adelaide, South Australia, Australia. [14]Tasmanian Clinical Genetics Service, Tasmanian Health Service, Hobart, Tasmania, Australia. [15]School of Medicine and Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia. [16]Genetic Services of Western Australia, Perth, Western Australia, Australia. [17]Department of Clinical Genetics, The Canberra Hospital, Canberra, Australian Capital Territory, Australia. [18]Centre for Cancer Biology, An alliance between SA Pathology and the University of South Australia, Adelaide, South Australia. [19]UniSA Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, Australia. [20]Murdoch Children's Research Institute, Melbourne, Victoria, Australia. ✉e-mail: zornitza.stark@vcgs.org.au

Methods

Ethics

The Australian Genomics Acute Care study has Human Research Ethics Committee approval (HREC/16/MH/251). Parents provided informed consent for participation in the study, following genetic counseling.

Study design and participants

The Acute Care Genomics program is a national multi-site study delivering ultra-rapid genomic testing to critically ill pediatric patients with suspected genetic conditions. The study adopts a highly coordinated clinical and laboratory approach, informed by implementation science principles and theory[3,42]. Participants were recruited prospectively from a network of 17 hospitals, including all children's hospitals in Australia, between January 2020 and April 2022. Patients were eligible if they were admitted to a participating NICU or PICU with a suspected monogenic condition as assessed by a clinical geneticist. Other hospital patients were considered if a rapid result was likely to alter management. Patients were ineligible if a monogenic etiology was considered unlikely; if a secure clinical diagnosis (such as Apert syndrome) was in place; if previous exome or genome testing had been performed; or if death or discharge were imminent (Fig. 1).

Electronic referrals were peer reviewed by a panel of study investigators for approval. Electronic test ordering and consent were used to collect clinical and other metadata in standardized formats, including HPO terms[15], as well as to manage remote recruitment during COVID-19 restrictions. Patient sex was determined karyotypically. Parents self-reported ancestry and this was recorded by referring clinicians based on categories in the Human Ancestry Ontology[14]. Virtual gene panels to guide phenotype-driven analysis were assigned by requesting clinicians; all panels used in the study are publicly available from PanelApp Australia[43] (https://panelapp.agha.umccr.org). Data were managed using REDCap, a secure, web-based application designed to support data capture for research studies[44]. Chromosomal microarray (SNP-CMA) was performed before enrollment if the likelihood of a chromosomal condition was considered high (for example, patients with multiple congenital abnormalities or isolated hypotonia to exclude Prader–Willi syndrome). Laboratory reports were issued to referring clinicians and were accompanied by plain language reports for families[45]. Additional findings unrelated to the reason for testing were not deliberately sought. A separate sub-study will explore a two-step model for offering these to families 3–6 months post-result.

Genome sequencing, data analysis and interpretation

WGS data generation and clinical analysis was performed using diagnostically accredited methods by the Victorian Clinical Genetics Services (VCGS) in Melbourne, Australia or SA Pathology, South Australia, Australia. For samples processed by the VCGS, DNA was extracted manually from blood collected in EDTA vacutainers using the QIAamp DNA blood mini kit (QIAGEN). DNA quantity and quality were assessed using the Qubit dsDNA BR (broad-range) Assay kit (Thermo Fisher) and TapeStation genomic DNA kit (Agilent), respectively. Whole-genome DNA libraries were created using Nextera DNA Flex Library Prep kit/ Illumina DNA prep kit (Illumina) followed by 2 × 150-bp paired-end DNA sequencing on a NovaSeq 6000 instrument (Illumina), variably using S2 or S4 flow cells. Targeted mean sequencing depth was 30×, with a minimum of 90% of bases sequenced to at least 10× for nuclear DNA (nDNA) and a minimum of 800× mean coverage for mitochondrial DNA (mtDNA).

Data were bioinformatically processed using commercially available and in-house analysis pipelines. Alignment to the reference genome (GRCh38) and calling of nuclear/germline DNA variants was performed using the Dragen v.3.3.7 (Illumina) workflow. Alignment to the revised Cambridge Reference Sequence (rCRS) mitochondrial genome (NC_012920.1) and calling of mitochondrial DNA variants was performed using an in-house analysis pipeline based on the Broad Institute best practice workflow (https://gatk.broadinstitute.org/hc/en-us/articles/4403870837275-Mitochondrial-short-variant-discovery-SNVs-Indels).

Automated sex determination, relatedness and contamination checks were performed using a combination of Somalier[46] and in-house tools.

For nDNA, variant analysis and interpretation within the selected target regions (RefSeq genes ± 1 kb) was performed using the trio (where available) analysis approach in the Alissa Interpret software suite (Agilent). CNVs were screened for and interpreted using an internal CNV detection tool, CXGo[47], incorporating four CNV detection tools (Delly, Lumpy, CNVNator and Canvas)[48–51]. For mtDNA variant interpretation, a custom in-house analysis pipeline was used to visualize large deletions and annotate the VCF file with variant information before manual filtering. An mtDNA variant was regarded as homoplasmic or apparently homoplasmic when it was present in at least 97%, respectively, of sequence reads aligned to the genomic position.

Genes with established disease association (Mendeliome) are considered during routine analysis. Curation of nDNA and CNV variants was phenotype-driven with pre-curated or custom gene lists used for variant prioritization (PanelApp Australia https://panelapp.agha.umccr.org/)[43]. Further analysis of all genes to identify research candidate genes was performed for undiagnosed cases (see below).

Classification of nDNA, CNVs and mtDNA variants was based on the appropriate American College of Medical Genetics and Genomics guidelines. Reported high-confidence small variants were generally not confirmed by an orthogonal method. All reported CNVs were orthogonally validated unless otherwise stated.

For South Australian patients with referral after May 2021, clinical data analysis was performed by a separate laboratory, SA Pathology, following data generation at VCGS. For these cases, relevant fastq and/ or bam and vcf files were shared via BaseSpace (Illumina) followed by analysis and reporting using SA Pathology's clinically accredited systems for small variants (SNVs and indels). Briefly, variant annotation was performed using the in-house VariantGrid v.3 interpretation support software. Rare inherited variants (consistent with Mendelian inheritance modes) and de novo variants were reviewed for overlap with the clinical presentation. In addition, phenotype-driven and virtual gene panel analyses were performed, independent of inheritance modes with the proband as singleton, focusing on genes most strongly associated with the clinical phenotypes. Frequency in population databases (gnomAD) and in-house database (VariantGrid) was used to prioritize or exclude variants for review. Variants were further prioritized based on in silico pathogenicity predictions, sequence conservation scores, protein function and expression and known disease associations. Results were integrated with copy-number and mtDNA analyses performed by VCGS. Candidate variants were reviewed in IGV. All reportable variants met quality criteria and were thus not confirmed by an orthogonal method.

Extended bioinformatic analysis and long-read sequencing

Data from all undiagnosed cases were analyzed for STR expansions using STRipy[18] and screened for balanced structural variants such as inversions and structural variants within 50 kb upstream and downstream of relevant genes using Manta[52] and Schism (https://github.com/ssadedin/schism). Events were manually assessed to confirm inheritance along with plausible proximity to a clinically relevant gene. Findings were orthogonally validated using clinically accredited methods before reporting. A complex structural variant and a large repeat expansion were confirmed using long-read sequencing (Nanopore).

Gene candidates

In all undiagnosed cases, analysis was expanded to variants in all genes, focusing on de novo and recessive high impact variants. Biologically

plausible gene candidates, based on gene properties and/or literature research, were submitted to GeneMatcher[28].

Trio transcriptome sequencing, data analysis and interpretation

In all undiagnosed cases, RNA was extracted from EDTA blood using the Omega Bio-Tek EZNA DNA/RNA Isolation kit. The quantitative and qualitative analysis of the extracted RNA was performed using the Qubit RNA Assay kit (Thermo Fisher) TapeStation High Sensitivity RNA and standard RNA kits (Agilent). Libraries for RNA-sequencing were constructed using the Illumina TruSeq stranded total RNA gold kit (including probes for ribosomal and globin RNA transcription inhibition) with subsequent NovaSeq 6000 2 × 150-bp paired-end sequencing (Illumina), variably using S2 or S4 flow cells with an aim of ~80 million fragments (~160 M paired-end reads) per sample. To identify true outliers of gene expression, the Bioconductor package OUTRIDER[53] was utilized. One hundred and seven of the probands were analyzed together, with parental samples analyzed separately. For quality control, samples with size factor <0.1 were removed from the analysis. Overall, 9,651 genes were filtered out due to zero counts (not expressed), which accounted for 16.2% of the original annotated genes from Gencode v.37. Genes with an adjusted P value < 0.1 per sample were interrogated, with a focus on downregulated genes.

Nanopore sequencing

DNA for Nanopore sequencing was extracted from EDTA blood as described above (Genome sequencing). Data were generated on a Promethion using v.10.4.1 flow cells and chemistry, followed by alignment using Minimap2 before visualization in IGV.

Long-range PCR

Primers were designed to bind in the intronic regions flanking the retrotransposon insertion (MECP2-int4-gDNA-Fwd: 5′-GCCTCTCCAAAGTTCAGCAAC-3′; MECP2-int4-gDNA-Rev: 5′-TGCCCTGAGTGGGAAGTTCT-3′). PCR was performed using Prime-StarGXL DNA Polymerase (Takara Bio) according to manufacturer's instructions. Then, 50 ng genomic DNA from the proband (A0131084), both parents (A0131084-M and A0131084-P) and NA12878 (Coriell) was amplified, alongside a No Template Control reaction. Reactions were run on a 1% E-Gel EX (Thermo Fisher) and visualized using a GelDoc XR (Bio-Rad).

Functional assays

VUSs related to phenotype and close to 'likely pathogenic' classification were reported and a pathway was sought to access functional studies and/or additional clinical correlation (such as imaging studies). Clinically accredited functional assays were used where possible, with some variants referred to research groups. Western blots were performed as previously described[54,55].

NUP214 functional genomic analysis

Cell culture. Primary cultures of fibroblasts from individual A1131048 and two unrelated pediatric control fibroblasts (from a 2-month-old female and female <16 years old) were established from skin biopsies as previously described[56]. Fibroblasts were cultured in basal medium (high-glucose DMEM (Gibco) with 10% fetal bovine serum (Gibco), 100 U ml^{-1} penicillin and 100 µg ml^{-1} streptomycin) at 37 °C with 5% CO_2. All fibroblast control cell lines were established in-house and were established from pediatric individuals without any suspected genetic disorders. All cell lines were mycoplasma negative and the *NUP214* genotype was validated by PCR and Sanger sequencing using the following primers: *NUP214* chr9:131127590c.112 C > T; p.(Arg38Cys); Fwd: GAGACAGACCTTGGTCTCAGTAA; Rev: AGCATGCCACCATACTCCTC and *NUP214* chr9:131134995c.929 T > C; p.(Ile310Thr); Fwd: CGGTTGA TGGCCAATGTTTGT; Rev: CAAGGCATCTCAGCCTCCATT.

Western blot

For denaturing gels, proteins from fibroblasts were extracted, 30 µg of proteins separated by SDS–PAGE and transferred to PVDF (Merck, cat. no. IPVH00010) as previously described[57]. Primary antibodies were specific to human NUP214 (cat. no. ab70497) and human GAPDH (cat. no. G9545, Sigma Aldrich; 1:10,000 dilution) and appropriate anti-rabbit horseradish peroxidase-conjugated antibodies (GE Healthcare; 1:5,000 dilution), enhanced chemiluminescence reagents (Bio-Rad) and Bio-Rad ChemiDoc were used to capture band intensity. Antibodies were validated by the manufacturer against a titrated amount of control human fibroblast lysate and used according to manufacturer's protocols. Protein band intensities were quantified using Image Lab v.6.0 software.

Immunofluorescence microscopy

Patient and control fibroblasts were fixed in 4% paraformaldehyde for 20 min and washed in PBST (PBS + 0.5% Triton X-100). Cells were blocked and permeabilized with 1% BSA in PBST for 1 h at room temperature. Cells were probed with primary antibody anti-NUP214 antibody (rabbit, cat. no. ab70497) diluted with 1% BSA in PBST (1:500) for 1 h at room temperature and washed in PBS. Secondary incubation was with Alexa Fluor 488-conjugated goat anti-rabbit antibody (Thermo Fisher Scientific, cat. no. A-11008) in 1% BSA in PBST (1:1,000 dilution) for 1 h at room temperature and washed in PBS. Coverslips were mounted on a microscopy slide using ProLongTM Gold Antifade with DAPI (Life Technologies).

Quantitative mass spectrometry and data analysis

Fibroblast cellular pellets for the A1131048 patient and five control fibroblast cell lines were resuspended in 5% SDS and 50 mM triethyl-ammonium bicarbonate. Total protein concentration for each sample was quantified with Pierce BCA protein assay kit (Thermo Fisher) and samples were aliquoted for 25 µg of protein in triplicates for the patient and single aliquots for each control cell line. Samples were processed using S-trap micro-spin columns according to the manufacturer's instructions, with reduction and alkylation being performed with 40 mM chloroacetamide (Sigma) and 10 mM tri(2-carboxyethyl)phosphine hydrochloride (BondBreaker, Thermo Fisher). Protein digestion was performed at a 1:10 trypsin to protein ratio at 37 °C overnight and eluted peptides were dried down using a CentriVap Benchtop Vacuum Concentrator (Labconco). Samples were reconstituted in 45 µl 2% acetonitrile, 0.1% trifluoroacetic acid buffer and 2 µl each sample injected for liquid chromatography (LC)-tandem mass spectrometry (MS/MS). Data were acquired on an Orbitrap Eclipse mass spectrometer (Thermo Fisher) coupled with an Ultimate 3000 HPLC (Thermo Fisher) and NanoESI interface. The system was equipped with an Acclaim Pepmap nano-trap column (Dionex-C18, 75 µm × 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex-C18, 75 µm × 50 cm), running in a data-independent acquisition mode with a previously described method[58]. Raw files were processed using Spectronaut (v.16.2.220903.53000, Rubin) against a data-dependent-acquired peptide library generated from deeply fractionated control fibroblast samples containing 150,106 peptide precursors. Default Spectronaut BGS Factory search parameters were used with changes made to select 'Exclude single hit proteins' settings and de-select 'Major Group Top N' and 'Minor Group Top N' options, allowing all identified peptides to be considered for quantitation. Proteins were searched using the UniProt reviewed human canonical and isoform database (42,360 entries).

Data were imported into Perseus (v.1.6.15.0) (ref. 59) for processing where known contaminants were filtered for and all MS2 quantity levels were log$_2$ transformed. The A1131048 patient and control groups were filtered for proteins having at least two valid values and a two-sided t-test was conducted. A volcano plot was generated using the scatter-plot function with significance set at ±1.5 fold change (log$_2$ ± 0.585) and P value = 0.05 ($-$log$_{10}$ = 1.301), with NPC components

being manually annotated. Log$_2$-transformed changes of NPC components were exported and the values were used to color chains representing specific subunits according to the fold change of the relevant protein on the cryoelectron microscopy-derived complex structure (Protein Data Bank accession code 7TBL) using the PyMOL Molecular Graphics System, v.1.7.2.1 (Schrödinger)[60,61].

Microscopy and analysis

For NUP214-containing nuclear pore density images were captured with a Zeiss LSM 900 confocal microscope with Airyscan 2 super-resolution imaging. NUP214 immunostained cells were captured with a ×63 oil immersion lens and Z-stacks with a series of 30 slices. For NUP214 quantification, Z-stack images were flattened in ImageJ using Z-Projection and 'Maximum Projection'. Images were loaded in CellProfiler and a pipeline developed to mask the area of the nucleus to exclude cytoplasmic staining and to measure the number of 'spots' per nucleus. For nuclear morphology, cells were imaged with a Zeiss Axiovert microscope at ×20 magnification and nuclear morphology was scored as either normal or nuclei with dysmorphic morphology were classed as blebs, invagination, micro-nucleoli or herniated nuclei as previously described[62] by blind assessment.

Cellular stress and apoptosis

Apoptosis, viability and cytotoxicity were assessed after heat shock, with the ApoTox-Glo Triplex Assay kit (Promega, cat. no. G6320) according to previously described methods with minor modifications[22]. To expose cells to heat shock stress, 10,000 fibroblasts per well were first grown at 37 °C in 96-well ELISA microplates and allowed to attach overnight. The following day the growth medium was exchanged with pre-warmed medium and the cells were moved into a 43 °C incubator for 2 h of heat stress exposure. Cells were then returned to 37 °C to recover for various time points. To provide a positive control for the induction of apoptosis, 2.5 μM staurosporine was added to the growth medium for 6 h. To provide a positive control for cytotoxicity, 70 μM digitonin was added to the culture medium for 15 min. Fluorescence (viability (400Ex/505Em) and cytotoxicity (485Ex/520Em)) and luminescence were measured on a FLUOstar Optima microplate reader (BMG Labtech).

Orthogonal tests

Data were collected from referring clinicians about any diagnoses achieved using alternative genetic testing modalities.

Clinical utility of results

Data on changes in management following ultra-rapid WGS were collected from referring clinicians via a REDCap survey 3 months post-result using a structured data collection instrument (Supplementary Table 5). These are grouped into three main categories: targeted treatments; redirection of care toward palliation; targeted surveillance (investigations and subspecialist referrals aimed at known complications).

Statistics and reproducibility

No statistical method was used to predetermine sample size. Sample size was determined by the available funding. No data were excluded from the analyses.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

De-identified genomic and associated data from this study are available for ethically approved research. Data access requests are accepted via an online application form that will require approval from the Australian Genomics Data Access Committee. For access to the data, please email AG-datarequest@mcri.edu.au. Data access requests are reviewed by the committee once a month. Access to the data will require a Data Transfer Agreement; once signed, the data will be transferred to the requestor from the Australian Genomics' Genomic Data Repository. All variants reported in this study have been deposited in ClinVar (SUB13026601, SCV003921769 to SCV003922018). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD042001.

The web resources used are:

ClinVar: https://www.ncbi.nlm.nih.gov/clinvar
GATK: https://gatk.broadinstitute.org
gnomAD: https://gnomad.broadinstitute.org
HPO: https://hpo.jax.org/app
The Human Ancestry Ontology: https://www.ebi.ac.uk/ols/ontologies/hancestro
IGV: https://software.broadinstitute.org/software/igv
NCBI RefSeq: https://www.ncbi.nlm.nih.gov/refseq
Manta: https://github.com/Illumina/manta
OMIM: https://omim.org
PanelApp Australia: https://panelapp.agha.umccr.org/
Schism: https://github.com/ssadedin/schism
STRipy: https://stripy.org
UniProt: https://www.uniprot.org/
VariantGrid: https://variantgrid.com
VEP: https://ensembl.org/info/docs/tools/vep. Source data are provided with this paper.

Code availability

Schism identifies new variation in short-read whole-genome data without imposing a preconceived model, by comparing raw reads in test samples to those sampled from large scale aggregated control sets. The Schism code used during the study is available on GitHub at https://github.com/ssadedin/schism. The VariantGrid code (www.variantgrid.com) used during this study is available for research use under business source license 1.1 on GitHub at https://github.com/SACGF/variantgrid.

References

42. Best, S. et al. Learning from scaling up ultra-rapid genomic testing for critically ill children to a national level. *NPJ Genom. Med.* **6**, 5 (2021).

43. Stark, Z. et al. Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am. J. Hum. Genet.* **108**, 1551–1557 (2021).

44. Harris, P. A. et al. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inf.* **42**, 377–381 (2009).

45. Brett, G. R. et al. Co-design, implementation, and evaluation of plain language genomic test reports. *NPJ Genom. Med.* **7**, 61 (2022).

46. Pedersen, B. S. et al. Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches. *Genome Med.* **12**, 62 (2020).

47. Sadedin, S. P., Ellis, J. A., Masters, S. L. & Oshlack, A. Ximmer: a system for improving accuracy and consistency of CNV calling from exome data. *Gigascience* **7**, giy112 (2018).

48. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

49. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).

50. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).

51. Roller, E., Ivakhno, S., Lee, S., Royce, T. & Tanner, S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* **32**, 2375–2377 (2016).

52. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

53. Brechtmann, F. et al. OUTRIDER: A statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am. J. Hum. Genet.* **103**, 907–917 (2018).

54. Akesson, L. S. et al. Distinct diagnostic trajectories in NBAS-associated acute liver failure highlights the need for timely functional studies. *JIMD Rep.* **63**, 240–249 (2022).

55. Frazier, A. E. et al. Fatal perinatal mitochondrial cardiac failure caused by recurrent de novo duplications in the ATAD3 locus. *Medicine* **2**, 49–73 (2021).

56. Fowler, K. J. Storage of skin biopsies at -70 degrees C for future fibroblast culture. *J. Clin. Pathol.* **37**, 1191–1193 (1984).

57. Van Bergen, N. J. et al. Pathogenic variants in nucleoporin TPR (translocated promoter region, nuclear basket protein) cause severe intellectual disability in humans. *Hum. Mol. Genet.* **31**, 362–375 (2022).

58. Kumar, R. et al. Oligonucleotide correction of an intronic TIMMDC1 variant in cells of patients with severe neurodegenerative disorder. *NPJ Genom. Med.* **7**, 9 (2022).

59. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

60. Stroud, D. A. et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature* **538**, 123–126 (2016).

61. Lake, N. J. et al. Biallelic mutations in MRPS34 lead to instability of the small mitoribosomal subunit and Leigh syndrome. *Am. J. Hum. Genet.* **102**, 713 (2018).

62. Shamseldin, H. E. et al. NUP214 deficiency causes severe encephalopathy and microcephaly in humans. *Hum. Genet.* **138**, 221–229 (2019).

Acknowledgements

Author contributions

S.L., S.E.B. and Z.S. drafted the manuscript. C.V.P., S.A.S., M. Wilson, J.P., M.F.H., C.P.B., M. Wallis, B.K., T.Y.T., M.L.F., J.C. and Z.S. coordinated study recruitment and patient selection and approval. B.C., D.P., D.F., K.S.K., R.H., S.G., P.A., M.R.J., H.S.S., S.E., S.R., S.L. and Z.S. processed genomic data and performed data analyses. K.B., A.R., G.R.B., M.C.d.S., A.S., M. Ward and K.S. provided genetic counseling support including patient consent and result disclosure. C.S., T.C., K.M.B., S.S., S.L. and Z.S. processed and analyzed transcriptomic data. A.H., S.S. and K.M.B. performed additional bioinformatic analyses. N.J.V.B., T.S., L.N.S., D.A.S., A.G.C. and D.R.T. performed functional analyses. C.V.P., S.A.S., M. Wilson, J.P., M.F.H., C.P.B., M. Wallis, B.K., T.Y.T., M.L.F., J.C., S.L., K.N.N., S.S., M.G.d.S., G.R.B., S.E.B. and Z.S. were involved in the conception of the study. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-023-02401-9.

Correspondence and requests for materials should be addressed to Zornitza Stark.

Peer review information *Nature Medicine* thanks Kym Boycott and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Anna Maria Ranzoni, in collaboration with the *Nature Medicine* team.

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

	Corresponding author(s):	Zornitza Stark
	Last updated by author(s):	2022/12/24

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	REDCap database for clinical data collection V12.5.16 (reference provided)
Data analysis	Illumina next generation sequencing software (Dragen V3.3.7); Alissa Interpret software suite V5.2.9 (Agilent); custom CNV detection tool, CxGo V1.0.8 - V1.6.0 (reference provided), PanelApp software V3.1.1 (reference provided); STRipy for STR detection V2.0 (reference provided); Manta for structural variant detection V1.6.0 (reference provided); Schism for structural variant detection V1.0 (https://github.com/ssadedin/schism); Outrider for RNA analysis V1.12.0 (reference provided); R V4.2; Variant Grid V3 (variantgrid.com; https://github.com/SACGF/variantgrid); Image Lab Version 6.0; PyMOL Molecular Graphics System version 1.7.2.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

De-identified genomic and associated data from this study are available for ethically approved research. Data access requests are accepted via an online application form that will require approval from the Australian Genomics Data Access Committee. For access to the data, please email AG-datarequest@mcri.edu.au. Data access requests are reviewed by the committee once a month. Access to the data will require a Data Transfer Agreement (DTA). Once a DTA is signed, the data will be transferred to the requestor from AWS S3 storage. All variants reported in this study have been deposited in ClinVar (SUB13026601, SCV003921769 - SCV003922018). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD042001.
Web resources used:
ClinVar: https://www.ncbi.nlm.nih.gov/clinvar
GATK: https://gatk.broadinstitute.org
gnomAD: https://gnomad.broadinstitute.org
The Human Phenotype Ontology: https://hpo.jax.org/app
The Human Ancestry Ontology: https://www.ebi.ac.uk/ols/ontologies/hancestro
IGV: https://software.broadinstitute.org/software/igv
NCBI RefSeq: https://www.ncbi.nlm.nih.gov/refseq
Manta: https://github.com/Illumina/manta
OMIM: https://omim.org
PanelApp Australia: https://panelapp.agha.umccr.org/
Schism: https://github.com/ssadedin/schism
STRipy: https://stripy.org
UniProt: https://www.uniprot.org/
VariantGrid: https://variantgrid.com
VEP: https://ensembl.org/info/docs/tools/vep

Human research participants

Reporting on sex and gender	Sex and gender are reported in line with current laboratory standards. No sex or gender-based analyses of the data have been performed.
Population characteristics	Demographic information (particularly relating to age) is reported in the manuscript. Majority of participants recruited were under 1 month of age (Fig. 1C). No age-dependent analysis of the data has been performed. There was no inclusion or exclusion criteria based on this or other characteristics (e.g. ethnicity) and we expect the cohort to be representative of the Australian population.
Recruitment	Participants were recruited prospectively using patient eligibility criteria outlined in the manuscript. They were assessed as likely to have an underlying monogenic disorder by the referring clinical geneticist and a group of study investigators. The study provided formal onboarding sessions for new sites and specific guidelines for patient assessment and recruitment, as well as a national expert panel to ensure similar patients were approved from all recruiting sites. Despite efforts to have nationally consistent recruitment, there may have been some variability in individual clinician practice in terms of patient assessment which could impact diagnostic yield.
Ethics oversight	The Australian Genomics Acute Care study has Human Research Ethics Committee approval from Melbourne Health (HREC/16/MH/251).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Eligible patients were recruited prospectively into the study based on study eligibility criteria. Sample size was determined by the available funding.
Data exclusions	No data were excluded.
Replication	The study used clinically accredited genomic analysis, and as such replication was not required if the variant call reached QC requirements.
Randomization	All participants followed the same study pathway, so randomisation and blinding are not applicable to study design.
Blinding	All participants followed the same study pathway, so randomisation and blinding are not applicable to study design.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-NUP214 antibody (rabbit, Abcam cat. #ab70497), Alexa Fluor 488 conjugated goat anti-rabbit antibody (ThermoFisher Scientific, Can. #A-11008), GAPDH (Cat. #G9545, Sigma Aldrich)
Validation	All utilised antibodies are commercially available and validated by the respective manufacturers. Relevant positive and negative controls were included as described in the manuscript. Antibodies were validated by the manufacturer against a titrated amount of control human fibroblast lysate and used according to manufacturer's protocols.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Patient-derived primary fibroblast cell line was established for participant A1131048 who was karyotypically female.
Authentication	The NUP214 variants were Sanger-confirmed in the patient-derived cell line to authenticate it.
Mycoplasma contamination	Not performed.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	N/A, not a clinical trial
Study protocol	Included as Methods in the manuscript.
Data collection	Study sites and period of data collection are included as Methods in the manuscript.

Outcomes

The primary outcome is diagnostic yield.