npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-025-00483-7

The Utah NeoSeq Project: a collaborative multidisciplinary program to facilitate genomic diagnostics in the neonatal intensive care unit

Sabrina Malone Jenkins[1,2] ✉, Rachel N. Palmquist[2,3], Barry Moore[4], Steven E. Boyden[4], Thomas J. Nicholas[4], Pinar Bayrak-Toydemir[5,6], Rong Mao[5,6], J. Andrew R. Farrell[4], Carson H. Holt[4], Shawn G. Rynearson[4], Chelsea M. Solorzano[1], Alistair Ward[4], D. Hunter Best[5,6], Najla Al-Sweel[5,6], Dawn L. Bentley[1], Luca Brunelli[1], Clement Y. Chow[7], Devin W. Close[8], Michael J. Cormier[4], Malia J. Deshotel[8], Jacob Durtschi[5], Erik J. Eide[4], Luaiva Floyd[1], Eric K. Fredrickson[8], Makenzie L. Fulmer[5,6,9], Edgar J. Hernandez[4], Ashley L. Kapron[10], Mary Anne Karren[4], Robert G. Lewis[5,6], Christine E. Miller[5], L. Charles Murtaugh[4], Kelsey E. Nicholson[8], Katherine Noble[8], Brendan D. O'Fallon[5], John M. O'Shea[5,6], David C. Pattison[8], Brent S. Pedersen[4], Brandy J. Petersen[1], Bennet D. Peterson[11], Lucilla Pizzo[5,6], Hayley M. Reynolds[9], Paul Rindler[8], Carrie B. Torr[1,12,13], Ting Wen[5,6], H. Joseph Yost[14], Jian Zhao[5,6], Mark Yandell[4], Gabor T. Marth[4], Aaron R. Quinlan[4,11], John C. Carey[15], Brian J. Shayota[15], Martin Tristani-Firouzi[16,17] & Joshua L. Bonkowsky[2,3,17]

Rapid genomic diagnostics in the Neonatal Intensive Care Unit represents a paradigm shift in medicine with increasing evidence of the utility of early diagnosis, impacting management. The goal of the Utah NeoSeq Project was to implement and evaluate a multidisciplinary and longitudinal rapid sequencing program while transitioning to CLIA-certified sequencing. Enrollment of 65 infants resulted in 26 (40%) with a diagnostic variant(s) and 7 (11%) harboring a strong candidate. This includes re-analyses resulting in four additional diagnoses. Parental surveys indicated that 7% (4/59) of parents had a decisional conflict after consent, and 3% (2/59) experienced decisional regret after the results. Fifty-two provider surveys were conducted. Seventy-nine percent (41/52) of results and 86% (19/22) of diagnostic results were "very useful" or "useful" and associated with management changes. The NeoSeq Project demonstrates that a multidisciplinary collaborative approach to diagnosis is feasible. We have developed a generalizable, collaborative protocol that addresses the need for expedited genetic evaluation with emerging technologies.

Genetic disorders lead to significant morbidity and mortality in critically ill infants. The rapid clinical progression of many genetic disorders necessitates a path to quickly identifying the underlying genetic variants to facilitate early, personalized treatment plans in the neonatal intensive care unit (NICU). Previous studies show that next-generation sequencing, including whole genome sequencing (WGS), leads to earlier diagnosis and improved outcomes due to rapid, more precise intervention in neonates[1–7]. However, barriers related to frontline clinical implementation of WGS in the care of critically ill newborns have restricted widespread use of WGS and left significant regulatory and logistical challenges to be overcome[8,9].

The NeoSeq Project was framed around the recognition that a multidisciplinary approach, integrating clinicians and scientists, could be used for clinical implementation and provide maximal benefit to patients[10]. We used a collaborative model between clinical staff at a regional hospital, molecular sequencing laboratory, and genome research center to address the needs of acutely ill patients at the University of Utah NICU while developing and testing genomic data interpretation tools with the potential for broad

A full list of affiliations appears at the end of the paper. ✉e-mail: Sabrina.malonejenkins@hsc.utah.edu

application[11–13]. Further, post-WGS, the NeoSeq Project integrated an undiagnosed pathway that included a protocol for routine short-read DNA re-analysis and utilization of additional investigative technologies, including long-read DNA sequencing, RNA sequencing, and functional analysis in animal models. Here, we summarize the application of our processes in an initial cohort of sixty-five critically ill neonates, aiming to expedite patient identification, improve the diagnostic yield of standard WGS analysis methods, and integrate workflows for ongoing analysis of unsolved cases. Based on key lessons learned, our goal is to expand these methods across regional NICUs to improve the care of patients in rural and underserved areas throughout the U.S. Mountain West.

Results

Population, recruitment, and consent

The NeoSeq study was conducted from February 2020 to May 2023. Screening, patient identification, and enrollment counts are included in Fig. 1. Of patients identified and assigned a kindred ID, some patients were not offered full consent discussion under multiple circumstances, including if the family declined full consent discussion or the clinical team determined that rWGS was not appropriate based on identification of a likely genetic diagnosis through other testing, a non-genetic etiology, complex psychosocial factors that would interfere with full participation in the study or preclude a diagnosis, demise (in utero or postnatal) prior to enrollment, palliative delivery, or delivery at an outside hospital. Sixty-five of sixty-six families (98.5%) who underwent full consent discussion chose to enroll in the study. Parents opted in to receive medically actionable incidental findings for 98% (64/65) of the probands. The overall opt-in rate was 94% (183/194), with 94% (61/65) for mothers, 90% (57/63) for fathers, and 100% (1/1) for siblings. Our enrollment period included the COVID-19 pandemic, during which clinical research at the University of Utah was temporarily placed on hold, so there were patients who met inclusion criteria that were unable to be included in the study.

Table 1 displays the demographic data of our participants, which reflect the current trends in birth rates across racial and ethnic groups in northern Utah. Over half of the participants were identified prenatally, reflecting our efforts for early identification of newborns at risk for genetic disorders. Eighty-six percent of the cohort was >32 weeks gestation at the time of birth. Eighteen (28%) of the probands died at some point after enrollment, underscoring the critical illness of infants undergoing genomic testing in the NICU.

Reporting and confirmation

The NeoSeq Project identified twenty-six patients (40%) with a diagnostic pathogenic, or likely pathogenic genetic variant(s) and seven cases (11%) with a strong candidate determined to be likely diagnostic (Fig. 2). Of the 26 cases who received a diagnosis (40%) and the seven cases (11%) with a strong candidate, seven cases (21%) had a structural variant, including five deletions, one inversion flanked by deletions, and one unbalanced translocation (Table 2). Among the eighteen probands who died, ten patients (56%) were diagnosed, and three patients (17%) had a strong candidate, suggesting that the deceased patients had a higher likelihood of possessing an identifiable genetic disorder.

The average turn-around time from enrollment to generation of a results letter was 11.8 days, and the average turn-around time for diagnostic

Fig. 1 | Screening, enrollment and overall study flow.

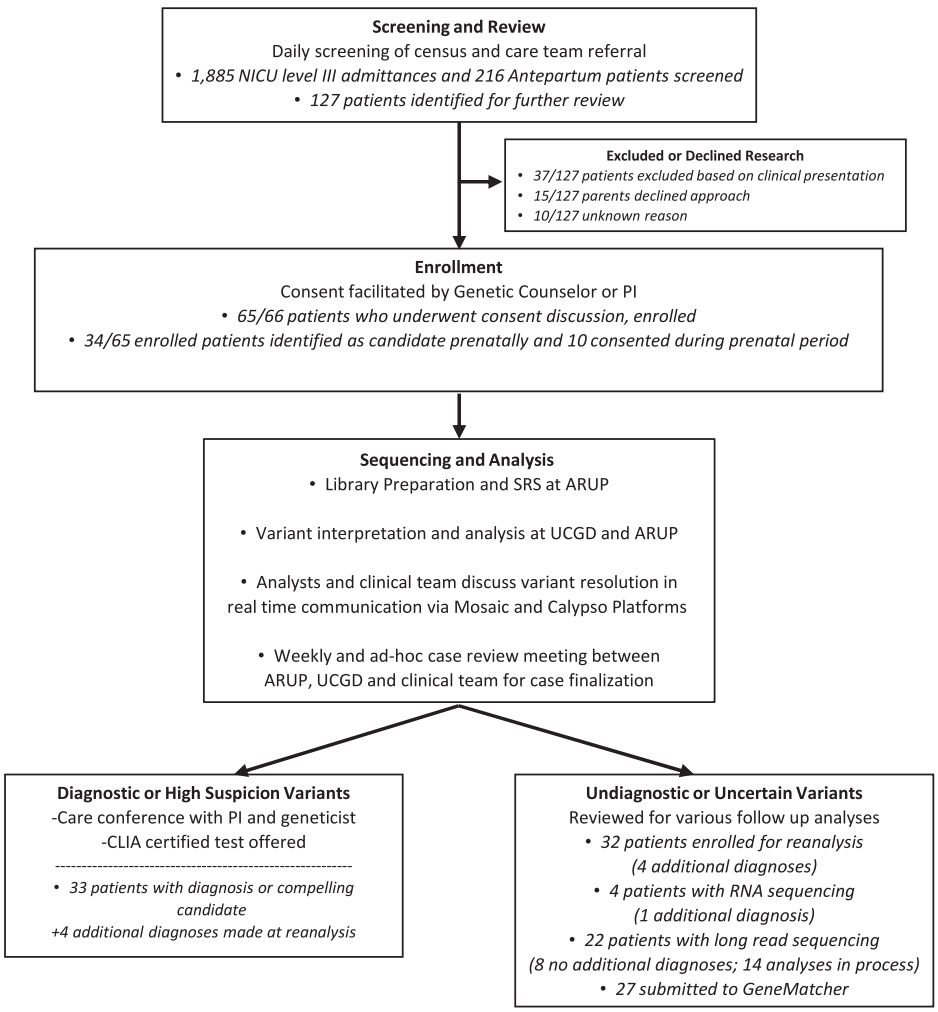


Table 1 | NeoSeq cohort characteristics

Gender, *N* (%)	
Male	43 (66)
Female	22 (34)
Self-reported race/ethnicity, *N* (%)	
American Indian or Alaska Native	0
Asian	1 (1.5)
Black or African American	1 (1.5)
Caucasian	39 (60)
Latino(a)	17 (26)
Multiple	6 (9)
Asian-Caucasian mixed	2
Latino-Caucasian mixed	2
Native American-Caucasian mixed	1
Native American-Latino mixed	1
Native Hawaiian or Pacific Island	1 (1.5)
Group characteristics*, N *(%)	
Gestational age	
<32 weeks	9 (14)
32–37 weeks	27 (41)
>37 weeks	29 (45)
Birth weight	
<2500 grams	28 (43)
>2500 grams	37 (57)
Identified prenatal period	34 (52)
Test ordering	
Test ordered as	
Trio	60 (92)
Duo (parent)	2 (3)
Duo (twin sibling)	2 (3)
Proband	0
Quad (trio and sibling)	1 (2)
Time to testing	
DOL test initiated	8 days (0–71)
Identification, discharge, mortality	
Discharged home prior to ROR	37 (57)
Deceased prior to ROR	6/18 (33)
Deceased during hospitalization	6/65 (9)
Discharged to Level IV NICU	23/65 (35)
Length of Hospital Stay	20.8 days (0–114 days)

DOL day of life; *ROR* return of results

findings was 6.4 days. Figure 2 shows the distribution of the diagnostic candidates and the turn-around time at each stage of testing. Figure 2b displays two cases with a total testing time of >30 days. The first case was delayed due to an expired IRB, and the second case was secondary to a delay in sequencing due to library prep failure, server issues delaying the transfer of files, and an extended variant review of a deceased patient.

Periodic re-analysis, long-read sequencing, RNA sequencing

All cases undiagnosed after initial analysis were scheduled for yearly re-analysis. Four cases were diagnosed at the 1-year re-analysis time-point due to a combination of recent literature review (three cases) and new variant discovery (one case). RNA sequencing was completed on four non-diagnostic cases with identified variants of uncertain significance (VUS) predicted to affect splicing. RNA sequencing confirmed the diagnosis of one case by providing significant evidence of altered splicing and nonsense-mediated decay caused by a variant of the diagnostic gene. Trio long-read DNA sequencing was completed for eight families, with no new diagnoses. The functional analysis service completed additional testing on one case that provided further evidence of the likely diagnosis; immunoblotting of patient skin biopsy samples showed greatly reduced *DERL1* protein expression compared to control samples, confirming that the compound heterozygous candidate variants resulted in a loss-of-function.

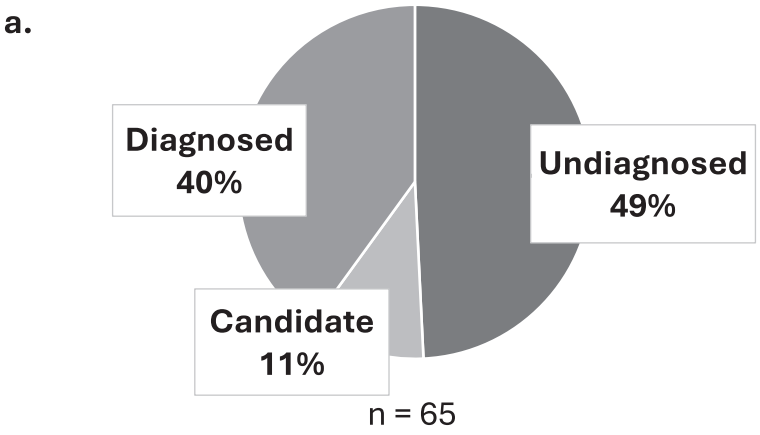
Measuring outcomes and utility of testing

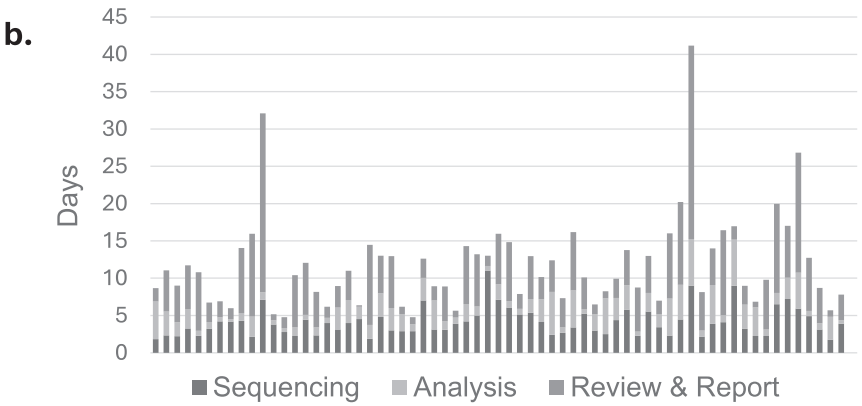
One parent declined to participate in parental surveys, leaving fifty-nine participants who completed both pre- and post-result surveys. Of the 59 participants, all but one reported that sequencing was explained to them in a way that they understood. Four participants (7%, 4/59) reported decisional conflict after consent, and two (3%, 2/59) went on to have decisional regret after the return of results. Of these two participants, one had a diagnostic result, and one had a non-diagnostic result. Despite having decisional conflict and regret, both participants agreed that the test was offered for their child's benefit prior to receiving the results. When asked to identify the most helpful part of the consent process, 70% (41/59) identified "speaking with someone with genetic expertise", 20% (12/59) answered "discussing with the primary team", 7% (4/59) identified "reading the consent", and 3% (2/59) answered "reading the summary sheet." Provider surveys collected information on the perceived clinical utility of WGS and identified changes in management (Table 3). Seventy-nine percent (41/52) of providers surveyed reported that rWGS was "useful" or "very useful," with 39% (16/41) of those reporting a change in management. Open-ended comments provided additional insight into clinician views and select quotes are summarized in Table 3.

Discussion

Rapid WGS is an emerging technology in the NICU. There is increasing recognition of the benefits of early genetic diagnosis for critically ill neonates by providers, families, hospitals, and insurers. However, major barriers in genomic diagnostics remain including limitations in technical and analytical identification of disease-causing variants as well as regulatory and implementation challenges for frontline clinical use. Additionally, accessibility to genomic testing has been limited by variable insurance coverage, by the cost of the test, and by the cost of test development and validation. We developed an integrative clinical research model that partnered with the expertise and interests of our university's hospital system, affiliated reference laboratory, and genomic research center. This model allowed for access to rapid WGS technologies for patients and families at no cost to them, and the data and experience gained from The Utah NeoSeq Project have been pivotal in supporting the approval of an rWGS test for clinical use in the University of Utah NICU starting in June 2023.

The Grant Scott Bonham Fetal Center at Primary Children's Hospital and the University of Utah Level III NICU provide a novel setting for exploring the utility of rWGS. The University of Utah NICU serves as an inborn unit for patients seen at the Fetal Center and is anticipated to need pediatric sub-specialty care at the connecting Level IV Primary Children's Hospital NICU, in addition to being a comprehensive Level III NICU managing premature and full-term infants. Most studies utilizing rWGS have been conducted in Level IV units[14,15], which are enriched for patients with complex medical presentations, and as expected, the proportion of admittances identified for enrollment in our study was lower than in Level IV studies. Of enrolled patients, 52% (34) were identified prenatally. Earlier recognition and diagnosis of genetic conditions can impact clinical care. The advancement of prenatal sequencing and the potential of expedited answers is an evolving area with positive potential. The risks and benefits of sequencing at different time points should continue to be explored. Although our consent rate after the consent discussion was very high





	Range (days)	Average (days)
Sequencing	1.8 - 11	4.1
Analysis	0.4 - 5	2.3
Review & Report	0.2 - 24	5.4
Total	4.8 - 32.1	11.8
Time to finding	3 - 11.7	**6.3**

Fig. 2 | NeoSeq diagnostic rate and time intervals. **a** NeoSeq diagnostic rate: cases were divided into diagnosed, strong genetic candidate, and undiagnosed. **b** NeoSeq time intervals: range and average days to complete sequencing, analysis, review, and time to preliminary finding.

(98.5%), some families declined to hear the consent after the initial approach.

This study served as a test cohort for the development of software tools and analysis workflows. The bioinformatics and development team also refined data transfer automation and data storage and communication platforms for visualizing data and documenting analyst and clinician reviews. Since the cost of clinical validation of a genome sequencing test is high, it is often difficult to generate sufficient data required to meet regulatory standards. This study contributed requisite data necessary for validation and subsequent use of the clinical rWGS test in ARUP Laboratories' CLIA-certified clinical laboratory, which uses the same sequencing methodology and sample preparation as NeoSeq.

Comparison of diagnostic rates between sites can be difficult given the variability of cohort selection, with ranges reported from 20-50%[1,14,16–20]. The NeoSeq Project resulted in 40% of cases with a diagnosis and an additional 11% of cases with a compelling candidate gene or likely diagnosis. Clinical review and follow-up testing were instrumental in determining the diagnostic status for several variants, highlighting the importance of an integrative model that partners clinicians and analysts and allows for custom analyses to be performed on a research basis. For example, a diagnostic complex inversion[11] was identified in one case that would have likely been missed by other WGS studies. In another case, we used RNA sequencing to confirm the pathogenicity of a variant in *NEB* causing nemaline myopathy, information that was vital to the parents for future pregnancy planning. We speculate there will be additional diagnoses in the future with re-analysis and increased RNA and long-read sequencing and analysis.

We found that our parental survey results were similar to published results[21], with low parental decisional conflict and regret. We also found that parents identified the most important part of the consent process as discussing the study with an individual with expertise in genetics; in our study, this was a genetic counselor or neonatologist with extensive genetic knowledge. This result highlights the importance of the consent process surrounding rapid WGS in minimizing decisional conflict and regret, an important consideration when expanding the use of this testing.

Provider surveys were conducted to assess clinical utility and change in management, with 52 providers completing the survey. Clinicians uniformly appreciated the benefits of integrating genetic testing into their clinical care of critically ill infants. Providers

Table 2 | Diagnostic and suspected diagnostic variants

Case	Duo trio	Gene variant(s), classification, inheritance	Diagnosis
Diagnosed cases			
1	Trio	TAF1 c.3544 C > T (p.Arg1182Cys), P, maternal (XR)	Intellectual developmental disorder, X-linked, syndromic 33 (MIM: 300966)
2	Trio	LGI4 c.504 G > C (p.Trp168Cys), LP, paternal (AR) c.322 G > A (p.Glu108Lys), LP, maternal (AR)	Arthrogryposis multiplex congenita 1, neurogenic, with myelin defect (MIM: 617468)
3	Trio	ABCC8 c.4307 G > A (p.Arg1436Gln), P, maternal (AR) c.695 G > A (p.Trp232Ter), P, paternal (AR)	Hyperinsulinemic hypoglycemia, familial, 1 (MIM: 256450)
4	Duo	COL1A c.1273 G > A (p.Gly425Ser), P, paternal or de novo (AD)	Osteogenesis imperfecta, type III (MIM: 259420)
5	Trio	BBS10 c.271dupT (p.Cys91Leufs*5), P, paternal (AR) c.909_912del (p.Ser303ArgfsTer3), P, maternal (AR)	Bardet-Biedl syndrome 10 (MIM: 615987)
6	Trio	ZFPM2 chr8:101080841-105439188 inversion, P, maternal (AD)	Diaphragmatic hernia 3 (MIM: 610187)
7	Trio	COL2A1 c.733_753del (p.Pro245_Pro251del), P, de novo (AD)	Spondyloepiphyseal dysplasia congenita (MIM: 183900)
8	Trio	CTSA c.184 C > T (p.Gln44Ter), P, paternal (AR) c.1372 T > G (p.Phe440Val), P, maternal (AR)	Galactosialidosis (MIM: 256540)
9	Trio	MDFIC c.386dupA (p.Met131Asnfs*3), P, homozygous (AR)	Lymphatic malformation 12 (MIM: 620014)
10	Trio	MYCBP2 c.11407 C > T (p.Arg3803Ter), P, de novo (AD)	MYCBP2-related developmental delay with corpus callosum defects (PMID: 36200388)
11	Trio	ATAD3A 1p36.33 interstitial deletion, chr1:1457189-1523860 (66671 bp), LP, maternal (AR) c.1696 C > T (p.Gln566Ter), LP, paternal (AR)	Pontocerebellar hypoplasia, hypotonia, and respiratory insufficiency syndrome, neonatal lethal (MIM: 618810)
12	Duo	PTPN11 c.417 G > C (p.Glu139Asp), P, maternal (AD)	Noonan syndrome 1 (MIM: 163950)
13	Trio	NSD1 c.4076 C > G (p.Ser1359Ter), P, de novo (AD)	Sotos syndrome (MIM: 117550)
14	Trio	KCNQ2 c.1076 C > T (p.Thr359Met), P, paternal (AD)	Seizures, benign familial neonatal, 1 (MIM: 121200)
15	Trio	TRPV4 c.806 G > A (p.Arg269His), P, de novo (AD)	Scapuloperoneal spinal muscular atrophy (MIM: 181405)
16	Trio	SNAP25 c.529 C > T (p.Gln177Ter), P, de novo (AD)	Myasthenic syndrome, congenital, 18 (MIM: 616330)
17	Trio	SOX9 c.1018 C > T (p.Gln340Ter), P, de novo (AD)	Campomelic dysplasia (MIM: 114290)
18	Trio	CHD7 and others 8q11.21-q12.3 interstitial deletion, chr8:51202481-61832156 (10629675 bp), P, de novo (AD)	CHARGE syndrome (MIM: 214800)
19	Trio	NEB c.[17965 T > A;18127 G > A] (p.[Tyr5989Ans;Ala6043Thr]), P, maternal (AR) c.23556 G > A (p.Ser7852 =), P, paternal (AR)	Nemaline myopathy 2 (MIM: 256030)
20	Trio	7q34-qter terminal deletion, chr7:140054895-159345973 (19291078 bp), P, de novo (AD)	7q terminal deletion syndrome
21	Trio	DNAH9 c.6431 G > C (p.Arg2144Pro), LP, paternal (AR) c.10600 C > T (p.Arg3534Ter), LP, maternal (AR)	Ciliary dyskinesia, primary, 40 (MIM: 618300)
22	Trio	LZTR1 c.742 G > A (p.Gly248Arg), P, paternal (AD)	Noonan syndrome 10 (MIM: 616564)
23	Trio	46,XY,del(8)t(7;8)(q21.3;p23.1) unbalanced translocation, P, de novo (AD)	Chromosome 8p deletion and 7q duplication syndrome
24	Trio	PKD1 c.11157-1 G > A, P, maternal (AD)	Polycystic kidney disease 1 with or without polycystic liver disease (MIM: 173900)
25	Trio	15q11.2-q13 interstitial deletion, chr15:23434321-28634351 (5200030 bp), P, de novo (AD)	Prader-Willi syndrome (MIM: 176270)
26	Trio	PAX3 c.1291 C > T (p.Gln431Ter), P, paternal (AD)	Waardenburg syndrome, type 1 (MIM: 193500)

Table 2 (continued) | Diagnostic and suspected diagnostic variants

Case	Duo trio	Gene variant(s), classification, inheritance	Diagnosis
Strong candidate cases			
1	Trio	PQBP1 c.487 C > T (p.Arg163Cys), VUS, maternal (XR)	Renpenning syndrome 1 (MIM: 309500)
2	Duo	DERL1 c.253 C > T (p.Arg85Ter), VUS, paternal or de novo (AR) c.154-2 A > G, VUS, maternal (AR)	
3	Trio	HUWE1 c.2513 C > G (p.Pro838Arg), VUS, maternal (XR)	Intellectual developmental disorder, X-linked, syndromic, Turner type (MIM: 309590)
4	Trio	ARSL c.1141 G > A (p.Gly381Ser), VUS, maternal (XR)	Chondrodysplasia punctata 1, X-linked recessive (MIM: 302950)
5	Trio	PIGA c.371 T > A (p.Ile124Lys), VUS, maternal (XR)	Multiple congenital anomalies-hypotonia-seizures syndrome 2 (MIM: 300868)
6	Trio	HAND2 c.406_407del (p.Ser136Glnfs*206), VUS, maternal mosaic (AD)	Tetralogy of Fallot (PMID: 25093829)
7	Trio	8p21.3-p12 interstitial deletion, chr8:22678458-32502968 (9824510 bp), LP, de novo (AD)	

P pathogenic, LP likely pathogenic, *VUS* variant of uncertain significance, AD autosomal dominant, AR autosomal recessive, *XR* X-linked recessive.
Genomic coordinates are for GRCh38.

Table 3 | Clinician survey results

Category	*N* (%)
Utility of result	
All results considered very useful or useful	41/52 (79)
Positive results considered very useful or useful	19/22 (86)
Changes in management in diagnostic cases	
Medical therapy initiated	0
Medical therapy stopped	1 (5)
Procedure/diagnostic test avoided	1 (5)
Early GT/trach	1 (5)
New specialist referral initiated	4 (18)
Additional medical screening	5 (23)
Palliative care initiated	3 (14)
Change in medical care unrelated to primary diagnosis	1 (5)
Other	2 (9)
It did not change management	8 (36)

Open-ended comments	
Positive results	"Parents seemed very happy to have a genetic diagnosis and appreciated knowing what to expect." "Happy to have helpful, positive results." "The baby's clinical condition was terrible, and we likely would have redirected care based on family wishes even without the genetics. However, having the genetics gave us more insight into the disease without multiple painful ancillary tests (muscle bx, EMG, etc)."
Negative results	"This is an interesting phenotype noted in 2 generations. The WGS approach very quickly excluded known genetic causes and provides for future re-analysis of sequencing results with the aim of gene discovery for isolated and inherited imperforate anus." "Still appreciate having this testing available"
Uncertain results	"While this result was unclear, on follow-up my impression is the family appreciated having a possible diagnosis. It didn't change planning. My understanding is he will be followed along with a group investigating this gene so may be helpful in the future?"

reported 79% (41/52) of all results to be "very useful" or "useful" and 86% (19/22) of positive results to be very useful or useful. Specific changes in management based on genetic testing results were identified in 40.7% (22/54) of cases, with "avoidance of procedure/diagnostic test" and "additional medical screening" being the most commonly identified changes. Providers expressed their own appreciation of having positive, negative, or uncertain results, as well as their perceptions that families appreciated such results.

The NeoSeq Project was limited by being a single-center study. Our enrollment period included the COVID-19 pandemic, during which research at the University of Utah was temporarily placed on hold, so there were patients who met inclusion criteria that were unable to be included in the study. Further, during the initial phase, not all patients were able to receive RNA and long-read sequencing, limiting our ability to measure their effect on diagnostic yield. Both of these modalities are now being offered retrospectively to many undiagnosed cases in this study as well as prospectively to patients with non-diagnostic results by clinical rWGS at the University of Utah Level III NICU, to further assess the utility of these technologies for neonates without a unifying diagnosis.

Genomic medicine is rapidly becoming part of pediatric and neonatal care. Given the critically ill condition of infants admitted to the NICU, establishing clear inclusion and exclusion criteria remains a complex challenge. It could be argued that all NICU patients are at risk of having an underlying genetic condition. Until comprehensive genomic sequencing is available to all NICU admissions, it remains difficult to determine who should have access to this technology. With the NeoSeq program, we have demonstrated the feasibility of developing and implementing a research-based rWGS diagnostic program at an academic hospital, as a means to test novel strategies while also maximizing benefits to our patients. Further, we found utility in providing early genetic diagnoses for clinical management, counseling, and family planning. Given the increasing recognition of the benefit of comprehensive genetic diagnosis in pediatric diseases, our protocol could provide a roadmap for generalized application to undiagnosed patients at other centers, especially where clinical sequencing is not an option due to institutional policy and coverage.

Methods

Study design

The purpose of the Utah NeoSeq Project is to provide diagnostic rapid whole genome sequencing (rWGS) to critically ill neonates and explore the utility and optimization of this technology. The study design of this project delivers a rapid return of results to clinical providers and families by operating under a research protocol, which allows for flexible and customizable research-based analysis workflows, comprehensive interrogation and reporting of all variant types, and prioritization of selected cases for experimental and analytical follow-up in the form of additional sequencing, clinical testing, or functional investigation of candidate variants. It was determined that a randomized control trial was not ethically appropriate given the growing evidence showing the improved outcomes and cost-efficiency of rWGS diagnosis in NICU populations. Therefore, all patients meeting the study criteria were offered enrollment. rWGS was completed in parallel with the clinical standard of care, which often includes sequencing of comprehensive phenotype-specific gene panels and/or chromosomal microarray analysis.

Ethical and legal considerations

We formed an expert panel of legal and ethical counsel, including members of the University of Utah Institutional Review Board (IRB), members of the University of Utah Center for Health Ethics, Arts and Humanities, pediatric legal counsel, genetic laboratory directors, and clinician researchers, with the goal of reviewing and evaluating applicable laws and weighing them with ethical considerations. Because rWGS was performed under a research protocol, we recognized that returning results prior to a Clinical Laboratory Improvement Amendments (CLIA)-certified confirmation could be interpreted as in conflict with CLIA regulations. Therefore, we created an IRB-approved protocol focused on informed consent in the setting of Health Insurance Portability and Accountability Act (HIPAA) right-to-access for families, to ensure that medically actionable variants with the potential for clinical intervention could be returned as research results[22]. Several factors were weighed in the development of this protocol, including the acute status of patients and the experience of our DNA extraction and research sequencing laboratory. Our consensus was that shared decision-making and transparency are fundamental to compassionate neonatal care. Parents were counseled regarding the limitations of genetic research results before and after testing, and the study team assisted in coordinating CLIA-certified confirmation testing for families with reported variants.

This research was conducted in full compliance with all relevant ethical regulations, including the Declaration of Helsinki. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Study investigators

Our study engaged three University of Utah-affiliated entities: ARUP Laboratories, a large non-profit University of Utah-affiliated reference laboratory, the Utah Center for Genetic Discovery (UCGD), an academic genome research center, and the University of Utah Department of Pediatrics. This collaboration included laboratory personnel and directors, genome scientists and analysts, molecular geneticists, and clinician researchers from Neonatology, Medical Genetics, Maternal Fetal Medicine, and other pediatric sub-specialties. Collaborations were established with the University of Utah Model Functional Analysis Service and the Utah Clinical and Translational Science Institute (CTSI), allowing for ongoing investigation of candidate variants.

Population, recruitment, and consent

Acutely ill infants suspected by their treating clinicians of having a genetic etiology were offered enrollment, along with their biological parents. Common indications included multiple congenital anomalies, unexplained seizures, arthrogryposis/skeletal dysplasia, non-immune hydrops fetalis, and acute clinical presentation and hospital course out of proportion to expectations in the NICU. Neonates with an established non-genetic cause for presentation, a known genetic diagnosis, or with high suspicion for aneuploidy were excluded from the study. We identified and enrolled patients both prenatally and postnatally. Prenatally, we identified patients and consented through the fetal care center and through the University of Utah hospital, which specializes in high-risk deliveries of patients who will require Level IV care at Primary Children's Hospital. Postnatally, we evaluated infants at our inborn Level III NICU. Clinical providers aided in the identification of potential patients and introduced the idea of genetic testing to the families.

Our consent[22] (Supplementary Note 1) was offered in both English and Spanish, with short forms available for 61 additional languages. We required an explicit opt-in or opt-out to receive incidental "medically actionable" findings for the patient and each parent. The consent outlined that participants' samples and information would be stored indefinitely for further research. If new results are identified through future research, participants are contacted and asked if they would like to receive these findings, with the understanding that their interest in research participation may have changed since they were initially consented.

Sample collection and sequencing

Proband umbilical or venous blood and parent venous blood were collected and couriered to ARUP Laboratories, typically within 1 hour. Genomic DNA was extracted from 0.5 to 1 mL whole blood using the Chemagic Magnetic Separation Module I (PerkinElmer, MA, USA) and quantified with a broad-range double-stranded assay kit on a Qubit 1.0 fluorometer (ThermoFisher Scientific, USA). For forty-three families, sequencing libraries were prepared with the Illumina DNA Prep workflow (Illumina, CA, USA). Dual-indexed, paired-end, whole genome libraries were prepared from 500 ng input DNA using on-bead transposons to normalize the DNA fragmentation and adapter ligation process. The normalized product underwent five PCR cycles to add unique 10-bp dual indices and sequencing adapters. After amplification, double-sided bead selection was used to select appropriately sized library fragments. The double-stranded libraries were quantified via electrophoresis on the 4200 TapeStation with high-sensitivity D5000 tapes (Agilent, USA), diluted to 3 nM, and pooled at equimolar ratios. The final library pool was diluted to 1.6 nM and spiked with 1% PhiX bacteriophage DNA as a sequencing control (Illumina, CA, USA)[11]. For twenty-two families, libraries were prepared for sequencing with the Illumina PCR-Free Prep workflow (Illumina, CA, USA) from 1 μg input DNA. Adapters, including 10-bp dual indices, were added via ligation. The final single-stranded libraries were pooled at fixed volumes and quantified via single-strand DNA Qubit (ThermoFisher, USA). These final pools were diluted to 2.6 nM and spiked with 1% PhiX (Illumina, CA, USA). For both library preparations, the final pool was denatured and loaded on a NovaSeq 6000 instrument for paired-end sequencing (2 × 150 bp) on an S1 flow cell

(Illumina, CA, USA). Sequence reads were demultiplexed and converted to FASTQ files with bcl2fastq (v2.20) and securely transferred from ARUP to UCGD.

Bioinformatics

SNV & indel variant calling. Automated calling of single nucleotide variants (SNVs) and short insertions and deletions (indels) was performed by the UCGD Core. The SNV/indel calling pipeline used the FastQForward parallelization platform. FastQForward wraps the BWA short-read aligner[23] and the Sentieon[24] variant calling toolkit, a computationally efficient re-implementation of GATK algorithms, for an overall workflow that follows GATK best practices[25]. Reads were aligned to a modified version of the GRCh38 reference genome that has false segmental duplications and other known erroneous sequences masked out by Ns, as per recommendations issued by the Genome Reference Consortium and Telomere-to-Telomere (T2T) Consortium. Variants were annotated using VEP, including five splice prediction plug-ins (SpliceAI, MaxEntScan, SpliceRegion, dbscSNV, and GeneSplicer), and the UTRannotator plug-in Automation of the NeoSeq pipeline used NICUWatch (https://github.com/srynobio/NICUWatch), a codebase to manage sample data transfer, multi-platform project creation, workflow implementation in Nextflow[26], and Amazon SNS messaging service to inform team members of data processing status and enable rapid downstream analysis.

Structural variation. Structural variants (SVs) were identified using Smoove[27] (https://github.com/brentp/smoove), Manta[28], and RUFUS (https://github.com/jandrewrfarrell/RUFUS). Smoove and Manta call structural events using split-mapped reads and discordant read pair insert sizes. SV calls were annotated with overlapping genes and duphold metrics[29] using Smoove, and with population allele frequencies using SVAFotate[30]. On select cases, data were re-aligned to the gapless T2T reference genome and reanalyzed with Smoove for further investigation of ambiguous SV calls. RUFUS performs de novo assembly of reads containing sequence kmers unique to the proband and absent from one or more parental controls, creating assembled contigs of mutant haplotypes and allowing identification of complex structural rearrangements as well as SNVs and indels. Mixed-type compound heterozygotes consisting of an SNV/indel and SV in trans were identified using RUFUS as well as manual comparison of SNV and SV candidate lists.

Other variant types. Short tandem repeat expansions at known disease-associated repeat loci were identified using GangSTR[31] and STRling[32] and filtered using dumpSTR and custom scripts. Aneuploidies and large copy number variants were identified using MoChA[33] and a RUFUS derivative that examines genome-wide kmer depth (unpublished). Uniparental disomy was identified by examining per-chromosome Mendelian inheritance error rates using MoChA. Extended regions of homozygosity indicative of monosomy, large deletions, consanguinity, or uniparental isodisomy were identified using the 'bcftools roh' subcommand[34].

Variant selection and prioritization. We used Slivar[35] to filter for high-impact, rare variants consistent with the possible modes of inheritance suggested by the pedigree structure. High-impact variants included nonsynonymous variants, splice region variants, intronic and synonymous variants predicted by any of five splice prediction programs to alter splicing, and 5' UTR start-gain variants. We also used the statistical approaches of VAAST[36,37] and Phevor[38] via Fabric Genomics and GEM[39] to prioritize variants according to predicted deleteriousness, conservation, population allele frequency, known associations of the gene to Human Phenotype Ontology (HPO) terms assigned to the patient, and other prior information about the specific variant (e.g. ClinVar classifications).

Quality control. Quality metrics were generated at various stages of the analysis. FASTQ files were evaluated for read quality using the fastp[40] tool. Alignments were evaluated for per-chromosome coverage using 'goleft indexcov'[41] and for general alignment metrics using alignstats (https://github.com/jfarek/alignstats). Variant calls were evaluated using 'bcftools stats'[42] for variant statistics and Peddy[43] to identify issues with assigned relatedness, sex, ancestry, and DNA contamination. Select QC results were aggregated and visually displayed for review using MultiQC.

Visual data review, communication, and project management. The software tool *Mosaic* (https://frameshift.io/) from Frameshift Genomics was used to support multiple aspects of project management, phenotype presentation, analysis, data quality and results review, and collaboration for team members. Extended candidate variant lists, prioritized gene lists, and detailed quality assessment data were uploaded and displayed within the tool, and its visualization functionality was used to compare data across NeoSeq cases. *Mosaic* also provided secure access to the underlying genomic data, allowing tools such as IGV and iobio[13,44–47] to be dynamically launched and saved.

Reporting and confirmation. The multidisciplinary team, consisting of clinicians, laboratory medical directors, and analysts, met weekly to review cases. Candidate variants were classified using the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines used for CLIA-certified clinical testing, and all reported variants were reviewed and signed out by an American Board of Medical Genetics and Genomics (ABMGG)-certified clinical molecular geneticist, a medical geneticist, and the study PI[48]. Variants were reported in the categories of diagnosis, selected VUS, and medically actionable incidental findings (for those patients who opted to receive such information during enrollment). VUS were reported when there was significant phenotypic overlap, a high suspicion of pathogenicity, or when pathogenicity could potentially be resolved through additional family testing or non-invasive patient workup. For diagnostic findings, the clinicians informed interpretation based on clinical presentation; sub-specialist reviewers were available from all pediatric divisions servicing the NICU in order to allow for precise clinical correlation. Medically actionable incidental findings included but were not limited to all ACMG-recommended secondary genes; it was determined that the 'medically actionable' status of variants should be evaluated on a case-by-case basis with consideration of current gene-disease associations and developments in management and therapies. Pathogenic and likely pathogenic variants were reported for incidental findings in line with ACMG guidelines[49].

Two documents were generated by the study team for each case: a research report and a results letter. The research report was a technical summary of all candidate variants that were identified and discussed, including variants that did meet the criteria for return to families and was kept as an internal reference for the analysis team to document evidence for and against pathogenicity and to consult upon yearly re-analysis. The results letter contained returnable results, was written in language intended for providers and patients, and was uploaded to the patient's electronic medical record under the research section and labeled as research results. The study team determined if additional follow-up was needed, including family genetic testing or biochemical laboratory testing, and these recommendations were included as a section in the result letter. Results were shared with the clinical care providers and returned to the parents by the study neonatologist and medical geneticist. Reported variants were clinically validated with CLIA-certified genetic testing, when available, and were confirmed using non-WGS research methods when CLIA-certified testing was not available.

Periodic reanalysis. Cases without a diagnostic result underwent yearly data re-analysis with a review of recent literature and medical records to refine the phenotype. Periodic re-analysis is considered imperative in the neonatal population, given that many phenotypes that inform analysis

may not appear until later in childhood. For candidate variants in genes of uncertain significance to human disease, we utilized GeneMatcher, a freely accessible web service that allows individuals to post a gene of interest and connects them internationally with other individuals who post the same gene, facilitating comparisons of variant type and clinical presentation if available.

Long-read sequencing. Diagnostically negative cases were candidates for long-read WGS. Long-read libraries were prepared from 5 µg genomic DNA using the standard PacBio HiFi protocol (PacBio, CA, USA). Genomic DNA was sheared on a MegaRuptor 3 (Diagenode) to 15–20 kb fragments. Prepared libraries were size-selected on a 0.75% BluePippin gel (Sage Sciences). Libraries were quantified using a high-sensitivity double-stranded DNA kit (ThermoFisher, USA) and size checked with a Genomic DNA 165 kb Kit on the Femto Pulse (Agilent, USA). HiFi Sequencing was performed on the Sequel IIe (PacBio, USA). Probands were sequenced with three 8 M SMRT cells, and parents were each sequenced with one SMRT cell. All SMRT cells had 30-hour movies collected. HiFi reads were aligned to the modified GRCh38 reference using pbmm2, and SNV/indels were identified using DeepVariant[50] and joint called using Glnexus. Short-aligned Illumina reads were then merged with aligned HiFi reads, and SNV/indels were recalled using DeepVariant's HYBRID_PACBIO_ILLUMINA model. Variants were annotated using the same protocol as with short reads. Structural variants were identified with pbsv using only HiFi reads.

RNA Sequencing. For selected diagnostically negative cases and cases with a candidate variant predicted to affect splicing, participant blood was drawn into a PAXgene Blood RNA Tube (IVD) (Qiagen: 762165). Cellular RNA from stabilized blood was isolated and purified using PAXgene Blood RNA Kit (Qiagen: 762164), including DNase I treatment to remove any residual DNA. Libraries were prepared using the Illumina TruSeq Stranded Total RNA Library Prep with Ribo-Zero Globin kit (20020613) with TruSeq RNA UD Indexes (20022371). Sequencing libraries were chemically denatured and applied to an Illumina NovaSeq flow cell at 1.3 nM using the NovaSeq XP workflow (20043131), and a 150 × 150 cycle paired-end sequence run was performed on a NovaSeq 6000 instrument using a S4 reagent Kit v1.5 (20028312).

RNA analysis. RNA sequence reads were aligned with STAR[51] using twopassMode=Basic to produce aligned BAM files and junction files. LeafCutterMD[52] was run on the indexed BAM files for all patients and parents currently sequenced (48 samples in total) to identify candidate outlier splice junctions for each patient. A custom python script was used to modify the STAR-generated junction files into a form compatible with IGV[53] and IOBIO[13,44–47]. Deeptools bamCoverage[54] was used to generate bigwig files representing read coverage in aligned BAM files. Custom Perl scripts were used to split LeafCutterMD results by patient and format them as BED files. Bedtools[55] was used to annotate the candidate splice junctions from LeafCutterMD with overlapping gene symbols. We collected candidate genes that had the potential to be affected by altered splicing from both identification of splice site/region variants from our WGS analyses and splice outlier candidates from the LeafCutterMD analysis. These RNASeq candidate genes were evaluated for overlap between the phenotypes described in the HPO for each candidate gene and the phenotypes derived from natural language processing-based extraction of HPO terms from the patient's medical records using ClinPhen[56]. Phenotype overlap was scored by a custom python script that implements a semantic similarity search similar to that used by Phenomizer[57] and originally described by Resnik et al.[58]. Candidate splicing events that impacted genes with phenotype overlap scored in the top 10th percentile for all genes annotated by HPO with disease associations were manually reviewed. Manual review of final candidates consisted of evaluating the read data supporting putative splicing events by viewing the RNASeq alignments supporting the event in IGV and the viewing the RNASeq junction files and bigwig coverage files in a custom IOBIO[13,44–47] application designed for reviewing splicing events. Genes and associated splicing events that were well supported by both phenotypic overlap and RNASeq read support were reviewed for clinical relevance by the full NeoSeq team. The steps in the analysis pipeline above were managed with a Snakemake[59] workflow.

Functional analysis evaluation. As needed, NeoSeq cases were reviewed by a functional analysis team consisting of collaborating investigators with expertise in functional genetics in cell lines, *Drosophila,* and zebrafish, in assessing the consequence of candidate variants in various model systems. Clinical and genetic information was integrated with a database and literature review to determine the most informative and feasible opportunities for experimental exploration of potential diagnostic variants.

Measuring outcomes and utility of testing. Parental satisfaction and understanding are crucial for enhancing the consent process for future parents and may increase the inclusion of more diverse participants in whole genome sequencing studies. We surveyed parents (Supplementary Note 2 and 3) who either participated in or declined participation in the study to evaluate their attitudes toward consent, research participation, satisfaction with genomic testing, and experiences of decisional conflict and regret. Following the consent process, or the decision to decline, the study coordinator approached parents to invite them to participate in a voluntary survey. This survey was conducted at the child's bedside and took ~3–5 mins to complete with the assistance of the study coordinator. Participants were able to skip any questions they did not feel comfortable answering. Surveys used multiple-choice, open-ended answers, and the Likert scale.

We also conducted a survey (Supplementary Note 4) among clinical providers to assess the utility of rWGS in the diagnosis, management, and counseling of their patients. The survey was distributed via REDCap within 5–7 days following the return of NeoSeq results. A cover letter accompanying the survey outlined its purpose, and consent was implied through the completion of the survey. The survey, estimated to take ~5 minutes, included questions regarding changes in clinical management, the perceived utility of rWGS testing, and its perceived impact on the family.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 2 August 2024; Accepted: 28 February 2025;
Published online: 23 March 2025

References

1. Petrikin, J. E. et al. The NSIGHT1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *NPJ Genom. Med.* **3**, 6 (2018).
2. Willig, L. K. et al. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir. Med.* **3**, 377–387 (2015).
3. van Diemen, C. C. et al. Rapid targeted genomics in critically ill newborns. *Pediatrics* **140**, e20162854 (2017).
4. Daoud, H. et al. Next-generation sequencing for diagnosis of rare diseases in the neonatal intensive care unit. *CMAJ Can. Med. Assoc.* **188**, E254–E260 (2016).
5. Saunders, C. J. et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl. Med.* **4**, 154ra135 (2012).
6. 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).
7. 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).
8. Green, E. D. et al. Strategic vision for improving human health at the forefront of genomics. *Nature* **586**, 683–692 (2020).
9. Stark, Z. et al. Integrating genomics into healthcare: a global responsibility. *Am. J. Hum. Genet.* **104**, 13–20 (2019).
10. 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).
11. Nicholas, T. J. et al. Comprehensive variant calling from whole-genome sequencing identifies a complex inversion that disrupts ZFPM2 in familial congenital diaphragmatic hernia. *Mol. Genet. Genom. Med.* **10**, e1888 (2022).
12. Peterson, B. et al. Automated prioritization of sick newborns for whole genome sequencing using clinical natural language processing and machine learning. *Genome Med.* **15**, 18 (2023).
13. Ward, A. et al. Clin.iobio: a collaborative diagnostic workflow to enable team-based precision genomics. *J. Pers. Med.* **12**, 73 (2022).
14. 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).
15. NICUSeq Study Group. et al. Effect of whole-genome sequencing on the clinical management of acutely ill infants with suspected genetic disease: a randomized clinical trial. *JAMA Pediatr.* **175**, 1218–1226 (2021).
16. Lunke, S. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* **29**, 1681–1691 (2023).
17. Krantz, I. D. et al. Effect of whole-genome sequencing on the clinical management of acutely ill infants with suspected genetic disease. *JAMA Pediatr.* **175**, 1–10 (2021).
18. Yang, L. et al. Genetic aetiology of early infant deaths in a neonatal intensive care unit. *J. Med. Genet.* **57**, 169–177 (2020).
19. Farnaes, L. et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *NPJ Genom. Med.* **3**, 10 (2018).
20. Meng, L. et al. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. *JAMA Pediatr.* **171**, e173438 (2017).
21. Cakici, J. A. et al. A prospective study of parental perceptions of rapid whole-genome and -exome sequencing among seriously ill infants. *Am. J. Hum. Genet.* **107**, 953–962 (2020).
22. Malone Jenkins, S. et al. Addressing ethical and laboratory challenges for initiation of a rapid whole genome sequencing program. *J. Clin. Transl. Sci*. **5**, e177 (2021).
23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
24. Weber, J. A., Aldana, R., Gallagher, B. D. & Edwards, J. S. Sentieon DNA pipeline for variant detection-Software-only solution, over 20× faster than GATK 3.3 with identical results. *PeerJ PrePrints* **4**, e1672v1672 (2016).
25. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics.***43**, 11.10.1–11.10.33 (2013).
26. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
27. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
28. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics***32**, 1220–1222 (2016).
29. Pedersen, B. S. & Quinlan, A. R. Duphold: scalable, depth-based annotation and curation of high-confidence structural variant calls. *GigaScience* **8**, giz040 (2019).
30. Nicholas, T. J., Cormier, M. J. & Quinlan, A. R. Annotation of structural variants with reported allele frequencies and related metrics from multiple datasets using SVAFotate. *BMC Bioinformatics* **23**, 490 (2022).
31. Mousavi, N., Shleizer-Burko, S., Yanicky, R. & Gymrek, M. Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* **47**, e90 (2019).
32. Dashnow, H. et al. STRling: a k-mer counting approach that detects short tandem repeat expansions at known and novel loci. *Genome Biol.* **23**, 257 (2022).
33. Loh, P.-R. et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).
34. Narasimhan, V. et al. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics.* **32**, 1749–1751 (2016).
35. Pedersen, B. S. et al. Effective variant filtering and expected candidate variant yield in studies of rare human disease. *Npj Genom. Med.* **6**, 1–8 (2021).
36. Hu, H. et al. VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet. Epidemiol.* **37**, 622–634 (2013).
37. Yandell, M. et al. A probabilistic disease-gene finder for personal genomes. *Genome Res.* **21**, 1529–1542 (2011).
38. Singleton, M. V. et al. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am. J. Hum. Genet.* **94**, 599–610 (2014).
39. De La Vega, F. M. et al. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Med.* **13**, 153 (2021).
40. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
41. Pedersen, B. S., Collins, R. L., Talkowski, M. E. & Quinlan, A. R. Indexcov: fast coverage quality control for whole-genome sequencing. *GigaScience* **6**, 1–6 (2017).
42. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
43. Pedersen, B. S. & Quinlan, A. R. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with peddy. *Am. J. Hum. Genet.* **100**, 406–413 (2017).
44. Ward, A. et al. Rapid clinical diagnostic variant investigation of genomic patient sequencing data with iobio web tools. *J. Clin. Transl. Sci.* **1**, 381–386 (2017).
45. Miller, C. A., Qiao, Y., DiSera, T., D'Astous, B. & Marth, G. T. bam.iobio: a web-based, real-time, sequence alignment file inspector. *Nat. Methods* **11**, 1189–1189 (2014).
46. Di Sera, T. et al. Gene.iobio: an interactive web tool for versatile, clinically-driven variant interrogation and prioritization. *Sci. Rep.* **11**, 20307 (2021).
47. Ekawade, A. et al. Genepanel.iobio - an easy to use web tool for generating disease- and phenotype-associated gene lists. *BMC Med. Genomics* **12**, 190 (2019).
48. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
49. Green, R. C. et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.***15**, 565–574 (2013).
50. Poplin, R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).

51. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Jenkinson, G. et al. LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics* **36**, 4609–4615 (2020).
53. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
54. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
55. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.***26**, 841–842 (2010).
56. Deisseroth, C. A. et al. ClinPhen extracts and prioritizes patient phenotypes directly from medical records to expedite genetic disease diagnosis. *Genet. Med. J. Am. Coll. Med. Genet.* **21**, 1585–1593 (2019).
57. Köhler, S. et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am. J. Hum. Genet.* **85**, 457–464 (2009).
58. Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. Available at https://doi.org/10.48550/arXiv.cmp-lg/9511007 (1995).
59. Mölder, F. et al. Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021).

Acknowledgements

This work was supported by the Utah Center for Genomic Medicine, ARUP Laboratories, the Mark and Kathie Miller Pediatric Genomics Fund, the Ben B. and Iris M. Margolis Foundation, and the R. Harold Burton Foundation. Sequence alignment, variant calling, and variant interpretation analyses were performed in part by the Utah Center for Genetic Discovery Core facility, part of the Health Sciences Center Cores at the University of Utah. This work utilized resources and support from the Center for High-Performance Computing at the University of Utah, partially funded by the NIH Shared Instrumentation Grant 1S10OD021644-01A1.

Author contributions

A.L.K., B.M., C.B.T., C.E.M., C.Y.C., J.A.R.F., J.L.B., L.B., M.TF., M.A.K., R.N.P., S.MJ., S.E.B. and T.J.N. conceived and designed the study. B.J.P., B.J.S., C.M.S., D.L.B., L.F., R.N.P. and S.MJ. recruited patients. B.J.P., C.M.S., D.L.B., L.F. and S.MJ. collected patient samples. B.J.P., C.B.T., C.M.S., D.L.B., J.C.C., K.N., R.N.P. and S.MJ. collected the data. D.H.B. and D.W.C. contributed to the technical design and oversight of all genomic sequencing and laboratory workflow design/implementation. D.C.P., E.K.F., E.J.E., J.D., K.N., K.E.N. and M.J.D. performed the sample processing, extracting, sequencing, and data transfer for experiments. A.R.Q., B.M., B.D.O., C.H.H., C.E.M., C.Y.C., G.T.M., J.A.R.F., M.Y., M.J.C., P.BT., R.N.P., R.M., S.MJ., S.G.R., S.E.B. and T.J.N. conceived, designed, implemented, oversaw, and/or contributed to the analysis and computational workflows. B.M., B.D.P., B.S.P., B.J.S., C.B.T., C.M.S., C.E.M., C.Y.C., D.H.B., D.L.B., D.W.C., E.J.H., H.J.Y., H.M.R., J.A.R.F., J.Z., J.C.C., J.M.O., K.E.N., L.C.M., L.P., M.L.F., M.A.K., M.J.C., N.AS., P.R., P.BT., R.N.P., R.G.L., R.M., S.MJ., S.E.B., T.J.N. and T.W. analyzed and interpreted the data. A.W., B.M., B.D.O., C.B.T., C.Y.C., D.L.B., E.J.H., J.A.R.F., M.L.F., M.J.C., N.AS., R.N.P., S.MJ., S.G.R., T.J.N. and T.W. contributed to data or analysis tools. B.M., B.J.S., C.M.S., M.TF., P.BT., R.N.P., R.M., S.MJ., S.E.B. and T.J.N. drafted the initial manuscript. A.L.K., A.R.Q., A.W., B.D.O., B.D.P., B.J.P., B.J.S., B.M., B.S.P., C.B.T., C.E.M., C.H.H., C.M.S., C.Y.C., D.C.P., D.H.B., D.L.B., D.W.C., E.J.E., E.J.H., E.K.F., G.T.M., H.J.Y., H.M.R., J.A.R.F., J.C.C., J.D., J.L.B., J.M.O., J.Z., K.E.N., K.N., L.B., L.C.M., L.F., L.P., M.A.K., M.J.C., M.J.D., M.L.F., M.TF., M.Y., N.AS., P.BT., P.R., R.G.L., R.M., R.N.P., S.E.B., S.G.R., S.MJ., T.J.N., T.W. read and approved the final manuscript.

Competing interests

A.W. is the CEO and co-founder of Frameshift Genomics. B.M. is a consultant for Fabric Genomics Inc. G.T.M. is the founder and Chief Scientific Officer of Frameshift Genomics, Inc. M.Y. is a co-founder, consultant, and science advisory board member of Fabric Genomics Inc. M.Y. is a co-founder of Backdrop Health Inc. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-025-00483-7.

Correspondence and requests for materials should be addressed to Sabrina Malone Jenkins.

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

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

[1]Division of Neonatology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA. [2]Center for Personalized Medicine, Primary Children's Hospital, Intermountain Healthcare, Salt Lake City, UT, USA. [3]Division of Pediatric Neurology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA. [4]Department of Human Genetics, Utah Center for Genetic Discovery, University of Utah, Salt Lake City, UT, USA. [5]Institute for Research and Innovation, ARUP Laboratories, Salt Lake City, UT, USA. [6]Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA. [7]Department of Human Genetics, University of Utah, Salt Lake City, UT, USA. [8]Institute for Clinical and Experimental Pathology, ARUP Laboratories, Salt Lake City, UT, USA. [9]University of Utah School of Medicine, Salt Lake City, UT, USA. [10]Center for Genomic Medicine, University of Utah, Salt Lake City, UT, USA. [11]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA. [12]Grant Scott Bonham Fetal Center at Primary Children's Hospital, Salt Lake City, UT, USA. [13]Center for Health Ethics, Arts and Humanities, University of Utah, Salt Lake City, UT, USA. [14]Molecular Medicine Program, Department of Pediatrics and Department of Neurobiology, University of Utah, Salt Lake City, UT, USA. [15]Division of Medical Genetics, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA. [16]Division of Cardiology, Department of Pediatrics, University of Utah School of Medicine, UT Salt Lake City, USA. [17]These authors contributed equally: Martin Tristani-Firouzi, Joshua L. Bonkowsky. ✉e-mail: Sabrina.malonejenkins@hsc.utah.edu