Measurement of genetic diseases as a cause of mortality in infants receiving whole genome sequencing

Stephen F. Kingsmore, Audrey Henderson, Mallory J. Owen, Michelle M. Clark, Christian Hansen, David Dimmock, Christina D. Chambers, Laura L. Jeliffe-Pawlowski and Charlotte Hobbs

INTRODUCTION

Infant mortality has been tracked annually worldwide for many years (http://www.geoba.se/population.php?pc=world&type=0198&year=2017&st=rank&asde=d&page=1, https://data.worldbank.org/indicator/SP.DYN.IMRT.IN?locations=US). Knowledge of causes of infant mortality shapes public health policy and prioritizes diseases for investments in surveillance, intervention and medical research. As a result of such prioritization, programs such as Back to Sleep/Safe to Sleep® have significantly reduced rates of many causes of infant mortality over the past 50 years in the United States (https://safetosleep.nichd.nih.gov/, https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm). Knowledge of the underlying causes of infant mortality is based on death certificates, which are estimated to be inaccurate or incomplete in 33–53% of cases. In addition, the vast majority of death certificates are not informed by molecular diagnoses.

Congenital malformations, deformations and chromosomal abnormalities (International Classification of Diseases (ICD-10) codes Q00–Q99) have been reported to be the leading causes of infant death in the United States for the past 50 years. Many, but not all, of these are associated with genetic diseases. Several of the other nine leading causes of infant mortality are also associated with genetic diseases, including low birth weight and prematurity, sudden infant death syndrome, newborn sepsis, diseases of the circulatory system and neonatal hemorrhage. Molecular autopsies have transformed the forensic evaluation of sudden unexplained death in the young, leading to new guidelines and a decreased number of deaths. Despite this success, and while the molecular basis of 14,000 genetic diseases has been determined, it is not known which genetic diseases are leading contributors to infant mortality other than for sudden unexplained death and chromosomal aneuploidies.

Knowledge of the contribution of genetic diseases to infant mortality has grown indirectly. Many infant deaths occur in hospitals, particularly in neonatal, pediatric and cardiovascular intensive care units (ICUs). Rapid genomic or precision medicine has created a new opportunity to decrease infant mortality associated with genetic diseases in ICUs. Many genetic diseases have specific, effective, evidence-based treatments. However, when diagnosis of a genetic disease is delayed or absent, treatments must be chosen empirically, with suboptimal outcomes, including unnecessary infant mortality. In response, rapid and ultra-rapid clinical genomic sequencing methods have been developed for prompt diagnosis in infants and children in ICUs who may have a genetic disorder for which effective treatment exists. It is estimated that genetic diseases occur in about 16% of infants in regional ICUs. Over the past 5 years, rapid genomic sequencing has demonstrated consistent utility in this population. In 873 children in ICUs described in 13 studies, 36% received a genetic disease diagnosis by rapid genomic sequencing. The average clinical or therapeutic utility of those diagnoses was 71%. Changes in patient outcomes were reported in 27%. The latter included both avoidance of infant mortality by therapeutic interventions and palliative care decisions. These data suggest that clinical-grade genomic sequencing has the analytic and diagnostic performance to ascertain the contribution of genetic diseases to infant mortality. Furthermore, determination of the leading genetic diseases associated with infant mortality may provide information to direct clinical and public health interventions to reduce infant mortality and decrease suffering. The current manuscript reports...
results of a scoping study and literature review to assess whether genetic diseases in toto are a leading cause of infant mortality.

RESULTS

Cohort Studies
Among 312 probands who had received rapid genomic sequencing in two research datasets, we examined the characteristics of infants who died prior to their first birthday. Both studies enrolled inpatient infants at Rady Children’s Hospital in San Diego between 11/23/2015 and 9/7/2018, primarily from a level IV neonatal intensive care unit. Of these, 205 infants were enrolled in the NSIGHT2 randomized controlled trial and 107 infants were enrolled under a biorepository protocol. All infants had a higher mortality rate (0.45%, 304 of 66,928) than non-white infants (0.41%, 199 of 58,224) than those of other races (9.6%; 10 of 104, p < 0.05) (https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/documents/MchSt-InfMort.pdf).

We compared the demographic characteristics of infants who died with those who survived in the two cohorts, and with those of all infant deaths in San Diego County during the same period. Among participants in the two research cohorts, infant mortality did not differ significantly between males (9.5%; 16 of 169) and females (12.3%; 14 of 114; P = 0.45). In contrast, in San Diego County from 2015–2019, infant mortality was lower among females (0.35%, 282 of 81,612) than males (0.42%, 359 of 85,710, p < 0.02) (https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/documents/MchSt-InfMort.pdf).

Eight infants of unknown ethnicity, mortality did not differ significantly between white infants (10.3%; 19 of 185 infants) and those of other races (9.6%; 10 of 104, P = 0.86). In contrast, in San Diego County from 2015–2019, white infants had a lower mortality (0.34%, 199 of 58,224) than non-white infants (0.41%, 444 of 109,100; p < 0.05) (https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/documents/MchSt-InfMort.pdf).

In the two RCHSD studies, 81 (26%) of the 312 infants were of unknown race, mortality in the two studies did not differ significantly between white infants (10.3%; 19 of 185 infants) and those of other races (9.6%; 10 of 104, P = 0.86). Excluding 8 infants of unknown ethnicity, mortality did not differ significantly between Hispanic infants (11.7%; 16 of 137 infants) and non-Hispanic infants (8.4%; 14 of 167; P = 0.34) in the two studies. In San Diego County from 2015–2019, Hispanic infants had a higher mortality rate (0.45%, 304 of 66,928) than non-Hispanic infants (0.34%, 339 of 100,396; P < 0.0002) (https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/documents/MchSt-InfMort.pdf).

Among the 30 infants in the research studies who died, we evaluated whether the likelihood of death had been anticipated or palliative care discussed with parents by EHR review. Eleven (37%) of 30 deceased infants had received a palliative care consult, and 18 (60%) of 30 deceased infants had a modified “Code Blue” resuscitation status in the event of cardiorespiratory arrest. Of these, 16 (53%) had “allow natural death” as their Code Blue status at time of death and 2 (7%) had a partial resuscitation Code Blue status. Of the infants who had received a genetic diagnosis, 4 (40%) out of 10 had a palliative care consult, and 7 (70%) out of 10 had a modified “Code Blue” resuscitation status. Of the 20 infants who did not receive a genetic diagnosis, 7 (35%) had a palliative care consultation and 11 (55%) had a modified “Code Blue” status.

We examined phenotypic characteristics of the infants who died and survived in the two studies. Using methods that had previously been validated for identification of patients with orphan diseases, we extracted HPO terms corresponding to the clinical features of the 312 infants from unstructured text in the Rady Children’s Hospital EHR at time of enrollment (prior to death) using clinical natural language processing. Of 2,313 HPO terms extracted from Epic only four terms differed in frequency significantly between enrollees who died in infancy and those who survived (Table 1)—abnormal urinary system, respiratory failure, hypotension and cardiac arrest. All occurred in a higher proportion of records of infants who died.

In the two RCHSD studies, 81 (26%) of the 312 infants were diagnosed with a genetic disease by rapid genomic sequencing. Of the 30 infants who died, 10 (33%) were diagnosed with 11 genetic diseases (Table 3). The mortality rate, although higher among those with genetic diagnoses, did not differ significantly between infants with or without genetic diseases (12.3%; 10 of 81 versus 8.7%; 20 of 231; P = 0.85), respectively. Nor did the average age at time of death differ significantly between infants with or without genetic diseases (104 days versus 89 days, P = 0.72). All of the genetic diseases diagnosed in infants who died had previously been associated with infant or childhood mortality, with the exception of Chr 12q21.33q22 del, which has only been described in six children. In addition, the symptoms and signs observed in these infants were consistent with causal association of genetic diseases and infant death (Table 3). No genetic disease was recurrent. Of the genetic diseases identified in infants who died, only CHARGE syndrome was diagnosed in at least one of the 282 infants who survived.

### Systematic literature review
To evaluate whether the findings reported herein were generalizable, we searched the literature for genomic sequencing studies in acutely ill infants in which mortality was recorded. Six additional studies were identified that utilized current genetic disease databases and current variant detection and interpretation methods (Table 4). Three studies had similar experimental designs as herein—diagnostic genomic sequencing was performed in infant inpatients in United States hospitals who had acute diseases of unknown etiology and in whom genetic diseases were suspected. In two studies, genome sequencing was

| HPO term       | HPO ID    | Alive (%) | Deceased (%) | P-value |
|----------------|-----------|-----------|--------------|---------|
| Abnormal urinary system | HP:0000079 | 91 (32.3%) | 22 (73.3%) | <0.001 |
| Respiratory failure     | HP:0002878 | 84 (29.8%) | 21 (70%)    | <0.001 |
| Hypotension            | HP:0002615 | 53 (18.8%) | 16 (53.3%)  | <0.001 |
| Cardiac arrest          | HP:0001695 | 4 (1.4%)   | 6 (20%)     | <0.001 |

Terms were extracted from EHRs of 312 infants at time of enrollment for rapid clinical genomic sequencing for diagnosis of genetic diseases.
Table 2. ICD10 codes of 1,179 infant deaths in San Diego County between 2005 and 2011.

| ICD10 Code | All deaths, n (%) | Out-of-hospital deaths, n (%) | In-hospital deaths, n (%) |
|------------|------------------|------------------------------|--------------------------|
| Total      | 1,179            | 744 (63.1%)                  | 435 (36.9%)              |
| No ICD10 code provided | 523            | 523 (70.3%)                  | <5 (0%)                  |
| Gestation <32 weeks | 304 (46%) | 304 (40.9%)                  | n.d.                     |
| Gestation 32–36 weeks | 62 (9%) | 62 (8.3%)                    | n.d.                     |
| Gestation 37–38 weeks | 69 (11%) | 69 (9.3%)                    | n.d.                     |
| Gestation 39–44 weeks | 88 (13%) | 88 (11.8%)                   | n.d.                     |
| A - Certain infectious and parasitic diseases | 5 (1%) | 0 (0%)                       | 5 (1.1%)                 |
| C – Neoplasms | 5 (1%) | <5 (0.3%)                    | <5 (0.7%)                |
| D – Blood/immune system diseases | 9 (1%) | <5 (0.5%)                    | 5 (1.1%)                 |
| E – Endocrine, nutritional and metabolic diseases | 11 (2%) | <5 (0.5%)                    | 7 (1.6%)                 |
| G – Diseases of nervous system | 13 (2%) | 5 (0.7%)                     | 8 (1.8%)                 |
| I – Diseases of circulatory system | 16 (2%) | <5 (0.4%)*                   | 13 (3.0%)*               |
| JKM,N - Diseases of respiratory, digestive, genitourinary and musculoskeletal systems | 8 (1%) | <5 (0.3%)                    | 6 (1.4%)                 |
| P - Certain conditions originating in the perinatal period | 295 (45%) | 20 (2.7%)*                  | 275 (63%)*               |
| Q - Congenital malformations, deformations & chromosomal abnormalities | 192 (29%) | 44 (5.9%)*                   | 148 (34%)*               |
| R95 - Sudden infant death syndrome | 82 (13%) | 74 (9.9%)*                   | 8 (1.8%)*                |
| R99,Y34 - Ill-defined, unknown, undetermined | 6 (1%) | <5 (0.5%)                    | <5 (0.5%)                |
| WX – Accidents, Assault | 14 (2%) | 8 (1.1%)                     | 6 (1.4%)                 |

*p-value: < 0.001.

performed post mortem63,64. One study had both ante and post mortem genomic sequencing65. While all included inpatient and community deaths, the inclusion criteria differed in each study (Table 4). In toto, the studies examined 918 infant deaths. The overall contribution of genetic disease to infant mortality was 21% (weighted average, median 31%, range 6 to 86%). Locus heterogeneity (number of affected loci divided by number of infant deaths) was 70% (137 genetic diseases in 195 infant deaths). The most common genetic causes of death were trisomy 21 (9%), spinal muscular atrophy (5%), 22q11 deletion syndrome (4%), trisomy 18 (4%), CHARGE syndrome (3%), and trisomy 13 (2%). It should be noted that most studies excluded infants with known chromosomal anomalies and those who died prior to enrollment, resulting in under-reporting of these. In two of the studies, the studies examined 918 infant deaths. The overall contribution of genetic disease to infant mortality was 21% (weighted average, median 31%, range 6 to 86%). Locus heterogeneity (number of affected loci divided by number of infant deaths) was 70% (137 genetic diseases in 195 infant deaths). The most common genetic causes of death were trisomy 21 (9%), spinal muscular atrophy (5%), 22q11 deletion syndrome (4%), trisomy 18 (4%), CHARGE syndrome (3%), and trisomy 13 (2%). It should be noted that most studies excluded infants with known chromosomal anomalies and those who died prior to enrollment, resulting in under-reporting of these. In two of the studies, the studies examined 918 infant deaths.
Genetic diseases diagnosed by rapid genomic sequencing in ten of thirty infants who died.

| Patient ID | Genetic disease | Affected gene or locus | Age at death (days) | Clinical presentation |
|------------|-----------------|------------------------|---------------------|----------------------|
| 244        | 22q11.2 del syn  | Chr 14q31.2q32.2 del    | 25                  | Metabolic acidosis, RF, sepsis, hypoxic ischemic encephalopathy, coagulopathy, pulmonary hypertension, cardiomyopathy, heart failure, respiratory distress syndrome, tachypnea, hypoxemia, hypomagnesemia, respiratory failure, anemia, neutropenia, neutrophilic granulocyte, small for gestational age, respiratory failure, sepsis, pulmonary hypertension, chronic lung disease, osteopenia, recurrent infections, renal failure, coagulopathy, anemia, hypoglycemia, death, MA, MA, metabolic acidosis, death, MA. |
| 247        | CHARGE syn      | CHD7                   | 23                  | CHARGE syn, CHD7      |
| 302        | Barth syn       | Barth syn (not specified) | 99                  | Barth syndrome, heart failure, hypoxia, anemia, metabolic acidosis, death, MA. |
| 325        | Concentric dysplasia w. sep | TAZ         | 81                  | Concentric dysplasia w. septum, heart failure, respiratory failure, anemia, neutropenia, neutrophilic granulocyte, small for gestational age, respiratory failure, sepsis, pulmonary hypertension, chronic lung disease, osteopenia, recurrent infections, renal failure, coagulopathy, anemia, hypoglycemia, death, MA. |
| 6012       | Coffin-Siris syn | ANDR8                  | 250                 | Coffin-Siris syndrome, CHARGE syndrome, CHD7, Barth syndrome, heart failure, respiratory distress syndrome, tachypnea, hypoxemia, hypomagnesemia, respiratory failure, anemia, neutropenia, neutrophilic granulocyte, small for gestational age, respiratory failure, sepsis, pulmonary hypertension, chronic lung disease, osteopenia, recurrent infections, renal failure, coagulopathy, anemia, hypoglycemia, death, MA. |
| 6024       | Chr 13q21.3-22q2 del syn | 12q21.3-24q2 del       | 35                  | Cof, CHD7, Barth syn, heart failure, respiratory distress syndrome, tachypnea, hypoxemia, hypomagnesemia, respiratory failure, anemia, neutropenia, neutrophilic granulocyte, small for gestational age, respiratory failure, sepsis, pulmonary hypertension, chronic lung disease, osteopenia, recurrent infections, renal failure, coagulopathy, anemia, hypoglycemia, death, MA. |
| 6144       | Mitochondrial complex II deficiency | TTN1; GPD1 | 335                 | Mitochondrial complex II, respiratory failure, anemia, metabolic acidosis, death, MA. |
| 6147       | Mitochondrial complex I deficiency | SPFD2, Btg2a3 syn 2 | 335                 | Mitochondrial complex I, respiratory failure, anemia, metabolic acidosis, death, MA. |
| 6207       | Syn syn, chronic respiratory failure, del deletion, MA, metabolic acidosis, pulmonary, mitochondrial, sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay, SM, sensorineural deafness, death, MA. | NDUFEV1 | 596                  | Syn syn, chronic respiratory failure, del deletion, MA, metabolic acidosis, pulmonary, mitochondrial, sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay, SM, sensorineural deafness, death, MA. |
| 6596       | Syn syn, chronic respiratory failure, del deletion, MA, metabolic acidosis, pulmonary, mitochondrial, sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay, SM, sensorineural deafness, death, MA. | NDUFEV1 | 674                  | Syn syn, chronic respiratory failure, del deletion, MA, metabolic acidosis, pulmonary, mitochondrial, sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay, SM, sensorineural deafness, death, MA. |

A surprising finding was that published treatment guidelines existed for 70% of infant deaths associated with genetic diseases. Although some patients may not be suitable for treatment even if guidelines exist, it is possible that if substantiated in a larger, unbiased cohort this would imply that there exists a considerable opportunity to decrease infant mortality and morbidity associated with genetic diseases by early neonatal diagnosis. It should be noted that it is now possible to diagnose a genetic disease in 19 h by ultra-rapid whole genome sequencing, albeit such testing is not yet reimbursed by most healthcare payers. There is now a growing literature of infants whose lives have been saved by early diagnosis of genetic diseases and timely implementation of targeted treatments.

Another interesting finding herein was that palliative care counseling was provided to 40% of parents of infants who died of genetic diseases. Palliative care counseling was often associated with changes in parental decisions with regard to treatment intensity, such as modified "Code Blue" resuscitation status in the event of cardiorespiratory arrest. If substantiated in a larger, unbiased cohort this would imply that there exists a potential opportunity to decrease suffering in infants with genetic diseases that have fatal prognosis by early diagnosis, and to facilitate a healthy grieving process among parents of deceased infants.

In summary, this study demonstrates that it is possible to identify the contribution of genetic diseases to infant mortality by genomic sequencing. Larger studies using the methods described herein have potential for molecular redefinition of the leading causes of infant death. By analysis of the specific diseases that cause infant mortality it is possible to identify primary and secondary strategies that have the potential to reduce infant mortality and decrease suffering. There was notable heterogeneity in the studies included here, particularly regarding inclusion criteria, geography and initial selection of the cohort. Future studies are therefore warranted that examine large, sequential cohorts of individuals from a single geography over several years in order to more accurately quantify the contribution of genetic.
| Ref. | Subjects | Infant deaths abstracted | Time at molecular diagnosis | Age at enrollment | Other inclusion criteria | Percent receiving genomic sequencing | Suspected genetic disease prior to sequencing | Type of sequencing | Enrollment dates | Type of study | Location | Infant deaths associated with genetic disease |
|------|----------|--------------------------|-----------------------------|-------------------|-------------------------|-----------------------------------|------------------------------------------|-------------------|----------------|-------------|----------|--------------------------------------------|
| 68   | 3989     | 223                      | PM                          | <28 DOL           | ICU cohort; Death in hospital or within 1 week of discharge<sup>b</sup> | 100%                              | No                                       | WES               | Jan 2015 - Dec 2017 | Retrospective | Fudan University, China | 6% (13/223) |
| 29   | 272      | 58                       | AM                          | <100 DOL          | ICU cohort; Death by DOL 120<sup>b</sup> | 100%                              | Yes                                      | WES               | Dec 2011 - Jan 2017 | Retrospective | Houston, TX, USA | 52% (30/58) |
| 49   | 35       | 14                       | AM                          | <4 months         | ICU cohort; Death by DOL 120<sup>b</sup> | 100%                              | Yes                                      | WGS               | Nov 2011 - Oct 2014 | Retrospective | Kansas City, MO, USA | 86% (12/14) |
| 32   | 65       | 10                       | AM                          | <4 months         | ICU cohort; Infant death<sup>b</sup> | 85%                               | Yes                                      | WGS, WES, panel | Oct 2014 - Jun 2016 | Randomized controlled trial | Kansas City, MO, USA | 30% (3/10) |
| 69   | 16       | 10                       | PM                          | <1 year           | Infant death without molecular cause at postmortem<sup>e</sup> | 100%                              | No                                       | WGS               | Not specified | Postmortem | Brisbane, Queensland, Australia | 30% (3/10) |
| 70   | 573      | 573                      | Both                        | <1 year           | In- or out-patient evaluation; Infant death | 100%                              | No                                       | WES, panel        | Jan 2011 - Jun 2017 | Retrospective | Boston MA, USA | 22% (124/573) |
| Herein, 28 | 205     | 14                       | AM                          | <4 months         | ICU cohort<sup>b</sup> | 100%                              | Yes                                      | WGS, WES         | Jun 2017 - Oct 2018 | Randomized controlled trial | San Diego, CA, USA | 36% (5/14) |
| Herein | 107      | 16                       | AM                          | <1 year           | Inpatient cohorts<sup>b</sup> | 100%                              | Yes                                      | WGS               | Nov 2015 - Sep 2020 | Prospective  | San Diego, CA, USA | 31% (5/16) |

Weighted Average (Median).

<sup>a</sup>PM: Post-mortem, AM: ante-mortem.
<sup>b</sup>Excluded infants with known chromosomal anomaly.
Cohort studies

We examined infant mortality among 312 children who had received rapid genomic sequencing in two research datasets. The first was NSIGHT2, a prospective, blinded, randomized controlled trial in infants in the neonatal, pediatric and cardiovascular ICUs at Rady Children’s Hospital, San Diego (RCHSD). NSIGHT2 compared the effectiveness and outcomes of rapid whole genome sequencing, rapid whole exome sequencing and ultra-rapid whole genome sequencing (ClinicalTrials.gov NCT03211039, registered July 7, 2017). NSIGHT2 was approved by the institutional review board (IRB) at RCHSD, was designated non-significant risk by the Food and Drug Administration (FDA) in an Investigational Device Exemption (IDE) submission, and was performed in accordance with the Declaration of Helsinki. Informed consent was obtained from at least one parent or guardian for each infant. The inclusion criteria were age <4 months and time from admission or time from development of a feature suggestive of a genetic condition of <96 h. The clinical inclusion criteria were broad in order to include all acutely ill infants with diseases of unknown etiology, as specified by the IRB at RCHSD and the FDA (ClinicalTrials.gov NCT02917460, registered September 28, 2016). Informed consent was obtained from at least one parent or guardian. The clinical rapid whole genome sequencing, rapid whole exome sequencing, analysis and interpretation

Clinical rapid whole genome sequencing, rapid whole exome sequencing and ultra-rapid whole genome sequencing were performed in laboratories accredited by the College of American Pathologists and certified through the Clinical Laboratory Improvement Amendments. Experts selected a few clinical features representative of each child’s illness from the Electronic Health Record (EHR, Epic) and mapped them to simple genetic diseases with VAAST (Fabric Genomics). Trio EDTA-blood samples were obtained where possible. Genomic DNA was isolated with an EZ1 Advanced XL robot and the EZ1 DSP DNA Blood kit (Qiagen). DNA quality was assessed with the Quant-iT Picogreen dsDNA assay kit (ThermoFisher Scientific) using the Gemini EM Microplate Reader (Molecular Devices). Genomic DNA was fragmented by sonication (Covaris) and bar-coded, paired-end, PCR-free libraries were prepared for rapid whole genome sequencing with TruSeq DNA LT kits (Illumina) or Hyper kits (KAPA Biosystems). Sequencing libraries were analyzed with a Library Quantification Kit (KAPA Biosystems) and High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical), respectively. 2 ×101 nt rapid and ultra-rapid whole genome sequencing was performed to at least 40-fold coverage with Illumina HiSeq 2500 (rapid run mode), HiSeq 4000, or NovaSeq 6000 (S1 or 52 flow cell) instruments, as described.

Table 5. 23 examples of genetic diseases associated with infant mortality that have published, effective treatment guidelines.

| Locus /Loci | Mendelian inheritance in man or orphanet disease name | Treatment guidance |
|-------------|--------------------------------------------------------|---------------------|
| JAG1        | Alagille syndrome                                      | 66                  |
| FASLG       | Autoimmune lymphoproliferative syndrome                | 67                  |
| PKHD1       | Autosomal recessive polycystic kidney disease          | 68                  |
| TAZ         | Barth syndrome                                         | 69                  |
| SCNSA, GPD1L| Brugada syndrome                                       | 70, 71              |
| LMNA        | Cardiomyopathy, dilated, 1A                            | 72, 73              |
| SLC25A20    | Carnitine-acylcarnitine translocase deficiency         | 74, 75              |
| CYBB        | Chronic granulomatous disease, X-linked                | 76                  |
| RYR1        | Cong. neuromuscular disease with uniform type 1 fiber  | 77                  |
| CPT2        | CPT II deficiency, lethal neonatal                     | 78                  |
| SCN2A       | Epileptic encephalopathy, early infantile, 11          | 79–81               |
| ETFDH       | Glutaric acidemia IIC                                  | 82–84               |
| PRF1, STXBP2| Hemophagocytic lymphohistiocytosis, familial           | 85, 86              |
| IL10RA      | Inflammatory bowel disease 28, early onset, AR         | 87                  |
| TRMU        | Liver failure, transient infantile                     | 88                  |
| MUT         | Methylmalonic aciduria, mut(t) type                    | 89                  |
| OTC         | Ornithine transcarbamylase deficiency                  | 90                  |
| TCFG1       | Osteopetrosis, autosomal recessive 1                    | 91                  |
| GAA         | Pompe disease                                          | 92                  |
| PDHA1       | Pyruvate dehydrogenase E1-alpha deficiency             | 93, 94              |
| SMN1        | Spinal muscular atrophy                                | 95                  |
| CACNA1C     | Timothy Syndrome                                       | 96                  |
| TSC2        | Tuberous sclerosis                                      | 97                  |

disease to overall infant mortality. A large study could also provide an estimate of the potential impact of genomic medicine on reducing infant mortality. Finally, this study has demonstrated the potential for the SOMI platform to determine whether spatial, environmental, sociodemographic and health-related factors differ among those with and without a known genetic contribution to infant mortality. As with previous population studies of San Diego County, many of the results will have generalizable implications.

METHODS

Cohort studies

We examined infant mortality among 312 children who had received rapid genomic sequencing in two research datasets. The first was NSIGHT2, a prospective, blinded, randomized controlled trial in infants in the neonatal, pediatric and cardiovascular ICUs at Rady Children’s Hospital, San Diego (RCHSD). NSIGHT2 compared the effectiveness and outcomes of rapid whole genome sequencing, rapid whole exome sequencing and ultra-rapid whole genome sequencing (ClinicalTrials.gov NCT03211039, registered July 7, 2017). NSIGHT2 was approved by the institutional review board (IRB) at RCHSD, was designated non-significant risk by the Food and Drug Administration (FDA) in an Investigational Device Exemption (IDE) submission, and was performed in accordance with the Declaration of Helsinki. Informed consent was obtained from at least one parent or guardian for each infant. The inclusion criteria were age <4 months and time from admission or time from development of a feature suggestive of a genetic condition of <96 h. The clinical inclusion criteria were broad in order to include all acutely ill infants with diseases of unknown etiology, as well as those with presentations highly suspicious for a genetic cause. Infants in whom there was a very low likelihood that a genetic disease diagnosis would change management were excluded. Specific exclusion criteria included infection or sepsis with normal response to therapy, isolated prematurity, and previously confirmed genetic diagnosis that explained their clinical condition. Infants with a confirmed molecular diagnosis of certain common genetic conditions that are detected by prenatal screening were therefore excluded. Full inclusion and exclusion criteria have been previously published. Genome interpretation was performed as singleton probands. Infants undiagnosed as singletons were re-analyzed as familial trios.

The second dataset was from a biorepository cohort. Inpatient infants at RCHSD without etiologic diagnoses, in whom a genetic disorder was possible, were enrolled between 11/23/2015 and 9/7/2018. Prospective comparison of outcomes and healthcare utilization was approved by the IRB at RCHSD and the FDA (ClinicalTrials.gov NCT02917460, registered September 28, 2016). Informed consent was obtained from at least one parent or guardian. Inclusion criteria were symptomatic inpatients less than 1 year of age who were affected by an illness potentially attributable to a genetic disorder. There were no specific exclusion criteria. All patients enrolled underwent whole genome sequencing.

Clinical, rapid whole genome and exome sequencing, analysis and interpretation

Clinical rapid whole genome sequencing, rapid whole exome sequencing and ultra-rapid whole genome sequencing were performed in laboratories accredited by the College of American Pathologists and certified through the Clinical Laboratory Improvement Amendments. Experts selected a few clinical features representative of each child’s illness from the Electronic Health Record (EHR, Epic) and mapped them to simple genetic diseases with VAAST (Fabric Genomics). Trio EDTA-blood samples were obtained where possible. Genomic DNA was isolated with an EZ1 Advanced XL robot and the EZ1 DSP DNA Blood kit (Qiagen). DNA quality was assessed with the Quant-iT Picogreen dsDNA assay kit (ThermoFisher Scientific) using the Gemini EM Microplate Reader (Molecular Devices). Genomic DNA was fragmented by sonication (Covaris) and bar-coded, paired-end, PCR-free libraries were prepared for rapid whole genome sequencing with TruSeq DNA LT kits (Illumina) or Hyper kits (KAPA Biosystems). Sequencing libraries were analyzed with a Library Quantification Kit (KAPA Biosystems) and High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical), respectively. 2 ×101 nt rapid and ultra-rapid whole genome sequencing was performed to at least 40-fold coverage with Illumina HiSeq 2500 (rapid run mode), HiSeq 4000, or NovaSeq 6000 (S1 or 52 flow cell) instruments, as described.
Sample preparation and sequencing for rapid whole exome sequencing was performed by an external clinical laboratory (GeneDx). Exome enrichment was with the xGen Exome Research Panel v1.0 (Integrated DNA Technologies), and amplification used the Herculase II Fusion polymerase (Agilent)\textsuperscript{28}. FASTQ files for rapid whole exome sequencing were transferred to Rady Children’s Institute for Genomic Medicine (RCGIM) for analysis and interpretation.\textsuperscript{29}

Rapid whole genome, rapid whole exome and ultra-rapid whole genome sequence alignments were aligned to human genome assembly GRCh37 (hg19), and variants were identified with the DRAGEN Platform (v.3.5, Illumina, San Diego)\textsuperscript{29}. Structural variants were identified with Manta and CNVnator (using DNAxensus)\textsuperscript{29}. Structural variants were filtered to retain those affecting coding regions of known disease genes and with allele frequencies <2% in the RCIGM database. Nucleotide and structural variants were annotated, analyzed, and interpreted by clinical molecular geneticists using Opal Clinical (Fabric Genomics), according to standard guidelines\textsuperscript{24}. Opal annotated variants with respect to pathogenicity, generated a ranked-ordered differential diagnosis based on the disease gene algorithm VAAST, a gene burden test, and the algorithm PHEVOR (Phenotype Driven Variant Ontological Re-ranking), which combined the observed Human Phenotype Ontology (HPO) phenotype terms from patients, and re-ranked disease genes based on the phenotypic match and the gene score\textsuperscript{24}. Automatically generated, ranked results were manually interpreted through iterative Opal searches. Initially, variants were filtered to retain those with allele frequencies <1% in the Exome Variant Server, 1000 Genomes Samples, and Exome Aggregation Consortium database\textsuperscript{24}. Variants were further filtered for de novo, recessive and dominant inheritance patterns. The evidence supporting a diagnosis was then manually evaluated by comparison with the published literature. Analysis, interpretation and reporting required an average of six hours of expert effort. If genomic sequencing established a provisional diagnosis for which a specific treatment was available to prevent morbidity or mortality, this was immediately conveyed to the clinical team, as described\textsuperscript{24}. All causative variants were confirmed by Sanger sequencing or chromosomal microarray, as appropriate. Secondary findings were not reported, but medically actionable incidental findings were reported if families consented to receive this information.

Infant mortality data collection

Characteristics of infant mortality (up to one year of age) were assessed among the 312 infants by structured EHR queries and manual EHR review. Characteristics included sex, race, ethnicity, clinical presentation, age at time of enrollment for rapid genomic sequencing, molecular diagnosis, age at time of death, “Code Blue” resuscitation status in the event of cardiorespiratory arrest, and whether a palliative care consultation had been performed. HPO terms at time of enrollment were extracted from the Epic EHR using clinical natural language processing as previously described\textsuperscript{7}.

The San Diego study of outcomes in mothers and infants (SOMI)

SOMI is a community-level, spatially referenced research platform with data derived from a complete collection of birth and death records and hospital discharge summaries for all births in San Diego County\textsuperscript{47}. Integrated into the platform are spatially referenced data on a wide variety of environmental measures, including census, air and water quality, climate, economic, crime, the built environment, transportation and social system data. The SOMI dataset represents the entire population of ~1 million mothers and their children born between 2005 and 2024. These data can also be linked to banked maternal serum samples from up to 70% of mothers who have received prenatal care in San Diego County, and archived newborn blood spots for virtually all live-born infants in the county. The SOMI dataset can be queried with a unique data visualization platform that incorporates many sources and types of data into one interactive mapping tool. The SOMI platform enables exploration of multi-level data to identify relationships between infant mortality and many pre- and postnatal environmental, genetic and epigenetic factors. The IRBs at the University of California San Diego and San Francisco have approved a waiver of consent for queries of the de-identified SOMI dataset.

Systematic literature review

A literature review to compare published evidence with findings from the two cohort studies and from data about overall infant deaths occurring in San Diego was conducted according to the MOOSE and PRISMA guidelines.

Data sources and record identification

We searched PubMed from January 1, 2010, to February 22, 2020 with the terms “infant mortality” and “genetic disease”. There were no language restrictions. There were no study design restrictions.

Study screening and eligibility

Studies that evaluated the etiologic contribution of single locus genetic diseases to infant mortality using comprehensive genomic methods (genome or exome sequencing with or without chromosomal microarray) were eligible. We limited eligibility to studies of cohorts with a broad range of genetic diseases, rather than one or a few disease types or clinical presentations, and in which the majority of probands were less than 18 years old.

Inclusion criteria and data extraction

Data extraction was manual. Data was reviewed for completeness and accuracy by at least two expert investigators and discrepancies were reconciled by consensus. Data extraction was limited to infant deaths which occurred at an age less than 1 year. Inclusion criteria for data abstraction included the use of genome or exome sequencing with or without chromosomal microarray, and cases were considered positive if they had a confirmed diagnosis of a single locus genetic disease detected by sequencing. The primary outcome considered was infant mortality associated with a molecular diagnosis of a genetic disease. Molecular diagnoses were defined as pathogenic or likely pathogenic diplotype affecting single genes or loci with definitive, strong or moderate associations with infant mortality causation. Variants of uncertain significance and secondary findings were not extracted. Where more than one publication reported results from a cohort, we included the most recent value for diagnostic utility.

Testing was performed clinically in hospital laboratories and reference laboratories, and experimentally in research laboratories. Hospital and reference laboratory clinical tests were defined primarily by the site of testing and, as disclosed in the methods, and, secondarily, by the affiliations of the authors. Clinical testing was defined as testing under fixed protocols that were attested to comply with state or national regulatory guidelines for in vitro diagnostic testing. Experimental research tests were those that explored the utility of novel or bespoke methods of testing or analysis.

Statistical analysis

For continuous outcomes, such as age at death, P-values were determined with a two-tailed t-test for independent samples. Counts of outcomes were compared between groups with chi-square. A z-test was used to evaluate the association of proportions between populations, using the binomial approximation of the normal distribution. A significance cutoff of 0.05 was used for all comparisons.

Reporting summary

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

DATA AVAILABILITY

All data associated with this study are present in the paper or are available at the Longitudinal Pediatric Data Resource under a data use agreement and subject to the limitations of the informed consent documents for each subject (Accession Number dbSNP000003.v1.p, https://www.ncbi.nlm.nih.gov/research-tools/longitudinal-pediatric-data-resource).

Received: 17 March 2020; Accepted: 2 October 2020; Published online: 02 November 2020

REFERENCES

1. Murphy, S. L., Xu, J., Kochanek, K. D. & Arias, E. Mortality in the United States, 2017. NCHS Data Brief. 1–8 (2018).
2. Xu, J., Kochanek, K. D., Murphy, S. L. & Tejada-Vera, B. Deaths: final data for 2007. Natl. Vital Stat. Rep. 58, 1–19 (2010).
18. Farnaes, L. et al. Rapid whole-genome sequencing decreases infant morbidity.

24. Kingsmore, S. F. et al. A Randomized, Controlled Trial of the Analytic and

31. Sanford, E. et al. Concomitant diagnosis of immune deficiency and Pseudo-

monas sepsis in a 19 month old with echyma gangrenosum by host whole-

ome sequencing. Cold Spring. Harb. Mol. Case Stud. 4, a003244 (2018).

34. Schofield, D., Rynehart, L., Shrestha, R., White, S. M. & Stark, Z. Long-term
economic impacts of exome sequencing for suspected monogenic disorders:
diagnosis, management, and reproductive outcomes. Genet. Med. 21,
2586–2593 (2019).

35. Soden, S. E. et al. Effectiveness of exome and genome sequencing guided by
acuity of illness for diagnosis of neurodevelopmental disorders. Sci. Transl.
Med. 6, 265ra168 (2014).

36. Stark, Z. et al. Meeting the challenges of implementing rapid genomic testing in
acute pediatric care. Genet. Med. 20, 1554–1563 (2018).

37. Stark, Z. et al. Does genomic sequencing early in the diagnostic trajectory make
a difference? A follow-up study of clinical outcomes and cost-effectiveness.
Genet. Med. 21, 173–180 (2019).

38. Stark, Z. et al. Prospective comparison of the cost-effectiveness of clinical whole-
exome sequencing with that of usual care overwhelmingly supports early use
of tiered molecular test in infants with suspected monogenic disorders.
Genet. Med. 19, 867–874 (2017).

39. Stark, Z. et al. A prospective evaluation of whole-exome sequencing as a first-
tier molecular test in infants with suspected monogenic disorders. Genet. Med.
18, 1090–1096 (2016).

40. Tan, T. Y. et al. Diagnostic impact and cost-effectiveness of whole-exome
sequencing for ambulant children with suspected monogenic conditions.
JAMA Pediatr. 171, 855–862 (2017). Sep 1.

41. Tan, N. B. et al. Diagnostic and service impact of genomic testing technologies
in a neonatal intensive care unit. J. Pediaetr. Child Health. 55, 1309–1314 (2019).

42. van Diemen, C. C. et al. Rapid targeted genomics in critically ill newborns.
Pediatrics 140, eic20162854 (2017).

43. Wang, H. et al. Optimized trio genome sequencing (OGTS) as a first-tier generic
test in critically ill infants: practice in China. Hum. Genet. 139, 473–482 (2020).

44. 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).

45. Ceyhan-Birsoy, D. et al. Interpretation of genomic sequencing results in healthy
and ill newborns: results from the BabySeq Project. Am. J. Hum. Genet. 104,
76–93 (2019).

46. Araneta, M. R. G. et al. Health advantages and disparities in preterm birth among
immigrants despite disparate sociodemographic, behavioral, and maternal risk
factors in San Diego, California. Matern. Child Health J. 20, 173–180 (2016).

47. van der Werf, I. M. et al. Novel microdeletions on chromosome 14q32.2 suggest
a novel SOX9 nonsense mutation, q401x, in a case of campomelic dysplasia with XY sex reversal.
Genet. Med. 21, 1009–1016 (2016).

48. Ceyhan-Birsoy, O. et al. Interpretation of genomic sequencing results in healthy
and ill newborns: results from the BabySeq Project. Am. J. Hum. Genet. 104,
76–93 (2019).

49. Araneta, M. R. G. et al. Health advantages and disparities in preterm birth among
immigrants despite disparate sociodemographic, behavioral, and maternal risk
factors in San Diego, California. Matern. Child Health J. 20, 173–180 (2016).

50. van der Werf, I. M. et al. Novel microdeletions on chromosome 14q32.2 suggest
a novel SOX9 nonsense mutation, q401x, in a case of campomelic dysplasia with XY sex reversal.
Genet. Med. 21, 1009–1016 (2016).

51. Corsten-Janssen, N., van Ravenswaaij-Arts, C. M. A. & Kapusta, L. Congenital arch
vessel anomalies in CHARGE syndrome: preliminary results from a Canadian
cohort molecular test in infants with suspected monogenic disorders.
Genet. Med. 21, 1090–1096 (2016).

52. Hendrix, N. W., Clemens, M., Canavan, T. P., Surti, U. & Rajkovic, A. Prenatally
diagnosed 17q12 microdeletion syndrome with a novel association with con-
granular encephalopathy improved outcome. Pediatr. Neurol. 105, 69–70 (2018).

53. Briggs, B. et al. Novel Factor XII variant identified through whole-genome
sequencing in a child with intracranial hemorrhage. Cold Spring Harb. Mol. Case
Stud. 4, a003525 (2018).

54. Kang, S. L., Forsey, J., Dudley, D., Steward, C. G. & Tsai-Goodman, B. Clinical
features and assessment of factors for unexplained death in the young. Am. J. Forensic Med.
Pathol. 37, e173438 (2017).

55. Willig, L. K. et al. Whole-exome sequencing for identified 17q12 microdeletion syndrome: consideration of genotype-phenotype correlations.
Am. J. Med. Genet. A. 149A, 1012–1018 (2009).

56. van den Berghe, L., Ganguly, S., Vermeulen, S., de la Bana, M., van den Berghe, P.
and Vermolen, I. A fetal proteome-based approach to identify novel candidate genes
involved in the pathogenesis of preterm birth. PLoS One 7, e48372 (2012).

57. Suzumura, H., Sakurai, K., Kano, K. & Ichimura, T. Coflacs and assessment of factors for
unexplained death in the young. Am. J. Forensic Med. Pathol. 37, e173438 (2017).

58. 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).

59. Petrikin, J. E., Willig, L. K., Smith, L. D. & Kingsmore, S. F. Rapid whole genome
sequencing and precision neonatology. Semin. Perinatol. 39, 623–631 (2015).

60. Sanford, E. F. et al. Rapid whole genome sequencing has clinical utility in chil-
dren in the pediatric intensive care unit. In Press.
AUTHOR CONTRIBUTIONS
S.F.K. led the project, obtained funding, and wrote the manuscript. M.O. revised the manuscript. A.H., M.M.C. and C.H. performed data collection and analysis. C.D.C. provided oversight of part of the project and obtained funding. L.L.J.-P. extracted SOMI data. C.H. provided oversight of part of the project. All authors reviewed and edited the manuscript.

COMPETING INTERESTS
D.D. received funding from Biomarin (consultant for Pegvaliase trials), Audentes Therapeutics (Scientific Advisory Board), and Ichorion Therapeutics (consultant for mitochondrial disease drugs). The remaining authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information is available for this paper at https://doi.org/10.1038/s41525-020-00155-8.

Correspondence and requests for materials should be addressed to S.F.K.