Study Protocol

Protocol of the China Neonatal Genomes Project: an observational study about genetic testing on 100,000 neonates

Feifan Xiao[1,2#], Kai Yan[1#], Huijun Wang[2], Bingbing Wu[2], Liyuan Hu[1], Lin Yang[3,4], Wenhao Zhou[1,2,4]

[1]Department of Neonatology, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China; [2]Center for Molecular Medicine, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China; [3]Clinical Genetic Center, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China; [4]Key Laboratory of Birth Defects, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China

[#]These authors contributed equally to this work.

Correspondence to: Lin Yang. Clinical Genetic Center, Children's Hospital of Fudan University, National Children's Medical Center, 399 Wanyuan Road, Minhang District, Shanghai 201102, China. Email: yanglin_fudan@163.com; Liyuan Hu. Department of Neonatology, Children's Hospital of Fudan University, National Children's Medical Center, 399 Wanyuan Road, Minhang District, Shanghai 201102, China. Email: nowadays921@126.com.

Background: Genetic diseases are caused by gene variants or chromosomal anomalies. The early diagnosis of genetic diseases is vital for patients. Next-generation sequencing has been widely used in the diagnosis of genetic diseases in recent years. Early genetic diagnosis by next-generation sequencing can guide clinical management and reduce the lifetime cost of care. However, no large study (>10,000 neonates) of the Chinese neonatal genome has been reported. Hence, the China Neonatal Genomes Project (CNGP) was created to address this deficiency.

Methods: The CNGP is an observational study performed in mainland China. In stage 1 (from August 2016 to December 2021), 30,000 neonates suspected of having genetic diseases were eligible for genetic sequencing. In stage 2 (from January 2022 to December 2025), 70,000 neonates will be screened using clinical exome sequencing. Clinical features, laboratory tests, and imaging examinations will be collected for the diagnosis of genetic disorders. The primary outcome is the diagnostic rate of gene variants. The secondary outcomes include clinical characteristics, clinical interventions, and patient outcomes.

Discussion: To our knowledge, CNGP is the largest project to explore the Chinese neonatal genome. The CNGP will build the Chinese neonatal genome database and establish a genetic testing workflow for neonatal genetic diseases. The results of the CNGP will provide useful genetic data for future studies.

Trial registration: ClinicalTrials.gov (NCT03931707).

Keywords: China; neonate; gene; sequencing

Received: 18 March 2021; Accepted: 12 June 2021; Published: 28 August 2021.

doi: 10.21037/pm-21-29

View this article at: https://dx.doi.org/10.21037/pm-21-29

Introduction

Genetic diseases are caused by gene variants or chromosomal anomalies. Common genetic diseases include congenital defects, chromosomal disorders, and metabolic disorders. Approximately 5% of newborns will be diagnosed with a genetic disease prior to until 25 years of age (1). Genetic diseases during the neonatal period can influence neonatal mortality (2). The neonatal period (the first 4 weeks of a child's life) is the most vulnerable time for a child's survival. According to data from the World Health Organization, the global average neonatal mortality rate was 17 per 1,000 live births in 2019 (3). Kingsmore *et al.* (4) found that approximately 21% of deceased infants were diagnosed with genetic diseases. It has been reported that genetic diseases impose a substantial economic burden on healthcare system (5,6). Moreover, genetic diseases cause significant psychological burdens for patients and their families.

Although most genetic diseases cannot be cured, some genetic disorders (such as hepatolenticular degeneration, favism, and phenylketonuria) can be controlled to avoid disease onset after early diagnosis. Thus, early diagnosis of genetic diseases is vital for patients.

In the early 1960s, biochemical tests became available to detect certain genetic diseases, such as phenylketonuria (7). Subsequently, other metabolic disorders were identified using tandem mass spectrometry (8). With technological advances, some methods, such as fluorescence in situ hybridization, chromosome microarray analysis, and Sanger sequencing, have been used for the diagnosis of genetic disorders. However, the disadvantages of these methods include their low throughput and high cost. With the development of sequencing and bioinformatics, next-generation sequencing (NGS) has enabled the simultaneous sequencing of thousands of genes. Moreover, the cost of NGS is relatively low. Therefore, NGS has been widely used in the diagnosis of genetic diseases since 2010 (9). Early genetic diagnosis by NGS can guide clinical management and reduce the lifetime cost of care (10-13). Moreover, NGS results can help in screening of relatives (14) and counseling families (2). Furthermore, NGS has been performed as a first-tier molecular test in infants with suspected monogenic disorders (15).

In China, the average annual births are estimated to be 17 million. The incidence of birth defects is estimated to be 5.6%, meaning that more than 900,000 newborns are born with congenital defects every year (16). Moreover, the under 5 mortality rate caused by birth defects was 0.16% in 2017 (17). Generally, congenital defects and metabolic disorders in newborns are caused mainly by genetic disorders (18). Thus, the need for early genetic diagnosis is clear. Considering this goal, the available target gene panel is more suitable for all newborns. Therefore, the spectrum of pathogenic/likely pathogenic genes is important for future genetic screening of newborns. However, no large study (>10,000 neonates) has previously reported the Chinese newborn genome. Hence, we conducted the China Neonatal Genomes Project (CNGP).

Our project aims to build a Chinese neonatal genome database, establish the genetic testing workflow of neonatal genetic diseases, promote the industrialization of neonatal genetic disease gene testing, and improve the training system for genetic counseling. We present the following article in accordance with the SPIRIT reporting checklist (available at https://pm.amegroups.com/article/view/10.21037/pm-21-29/rc).

Methods

Ethics and dissemination

The present study is performed in accordance with the Declaration of Helsinki (as revised in 2013), Good Clinical Practice, and related laws. This study was approved by the Ethics Committee of the Children's Hospital of Fudan University (CHFudanU_NNICU11). Informed consent was obtained from the parents or guardians of the neonates. The consent information includes the aims of this project, risks involved in participating, and participants' rights and responsibilities. A detailed description of the project will be provided to the parents or guardians of the neonates. The results of this project will be published in international peer-reviewed journals and all participant data will be non-identifiable in the publication of the study findings. The trial is registered at ClinicalTrials.gov (NCT03931707).

Study design and setting

The CNGP is an observational study that aims to enroll 100,000 neonates for genetic testing. This project involving two stages.

Stage 1: From August 2016 to December 2021, we enrolled 30,000 neonates to collect genetic data. At this stage, neonates suspected of having genetic diseases were eligible for genomic sequencing. Genetic testing by clinical exome sequencing (CES) or exome sequencing (ES) were directly ordered by physicians, and genome sequencing (GS) was approved by a laboratory-based physician's application.

Stage 2: From January 2022 to December 2025, 70,000 neonates will be enrolled and screened for genetic diseases using CES.

Sample collection will involve at least 2 mL venous blood samples from each neonate subject for genetic testing. In addition, clinical features, laboratory tests, and imaging examinations will be collected to diagnose genetic disorders. The basic clinical features of the previously enrolled neonates are shown in Supplementary file.

The Children's Hospital of Fudan University, the leading unit of the CNGP, is responsible for genetic testing and data analysis. Other participating units include children and maternal hospitals, pediatrics departments from general hospitals, and children hospitals from 31 provinces/

autonomous regions/municipalities in mainland China.

Study population

The participants were all members of the CNGP who were hospitalized in the neonatal department of each hospital. The inclusion criteria were as follows: (I) neonates (age ≤28 days), (II) Chinese parents, (III) at least 2 mL venous blood sample obtained, and (IV) informed consent from the parent or guardian. The exclusion criteria were as follows: (I) mothers with multiple pregnancies, (II) parents under 18 years of age who could not make consent decisions, and (III) patients or guardians who rejected genetic data for subsequent research analysis.

Laboratory processing

CES targets 2,742 specific genes, ES targets the exome (protein-coding genes), and GS targets the entire genome. The Agilent ClearSeq Inherited Disease panel kit (including 2,742 genes; Agilent, Santa Clara, CA, USA) and the Agilent SureSelect XT Human All Exon V5 kit were used for CES and ES, respectively. Sequencing was conducted using the Illumina HiSeq X10 (Illumina, San Diego, CA, USA). The average on-target sequencing depth was 200× for CES and 120× for ES. GS was performed using the Illumina NovaSeq 6000 sequencing platform. For CES and ES, sequencing reads were aligned to the reference human genome (hg19) and variant calling was performed using the Genome Analysis Toolkit Best Practices Pipeline (19). The ClinVar, Online Mendelian Inheritance in Man, and the Human Gene Mutation Database were searched for known pathogenic and likely pathogenic variations. Interpretation of sequence variants was conducted based on published standards and guidelines (20,21). For GS, sequencing reads were mapped to the human reference genome hg38. Copy number variations or structural variations were prioritized for interpretation, with reference to the literature and following genetic databases: Decipher, DGV, ClinGen, and the Human Genome Mutation Database. The detailed methods of sequencing and analysis of CES, ES, and GS can be found in our published studies (18,22-24). Variants in patients and parents were validated using Sanger sequencing. CNVs will be confirmed by multiplex ligation-dependent probe amplification, array-based comparative genomic hybridization, or quantitative real-time PCR.

Outcomes

The primary outcome is the diagnostic rate of gene variants. Taking the number of newborn babies as the denominator and the number of neonates with gene variants detected by gene sequencing, the Chinese neonatal gene variant rate will be obtained. The secondary outcomes include clinical characteristics (such as gender, family history, and symptoms), clinical interventions, and patient outcomes.

Data analysis plan

Neonates who are tested by NGS will be divided into two groups (those with genetic findings and those without). Clinical characteristics, clinical interventions, and outcomes will be collected for data analysis. Data will be analyzed using frequency counts and proportions. The chi-square test and Fisher's exact test will be used to analyze analysis of categorical variables. The Mann-Whitney U test will be performed to distinguish intergroup differences. Statistical significance will be set at $P<0.05$. All data analyses will be conducted using SPSS version 20 (IBM, Armonk, NY, USA). The raw clinical and genetic data will be stored on a local server.

Discussion

This project was initiated on August 8, 2016. As of February 1, 2021, 98 hospitals have participated. Currently, several studies based on CNGP have been published: (I) we performed optimized trio genome sequencing as a first-tier genetic test in 84 critically ill neonates (18); (II) we provided a precise and portable workflow for survival motor neuron gene copy number analysis based on exome sequencing (25); (III) we used 24-h rapid trio-exome sequencing for 10 critically ill neonates (6); and (IV) we used NGS to investigate the genetic causes in 588 neonates with multiple congenital anomalies (24).

This study has two main limitations. First, neonates will mainly be enrolled from the Children's Hospital of Fudan University; therefore, the generalizability of genetic results may be limited. Second, complex genetic diseases may be missed because of the current sequencing technology.

In conclusion, CNGP is the largest project to explore the Chinese newborn genome. The CNGP will build a Chinese neonatal genome database and establish a genetic testing workflow for neonatal genetic diseases. The results of the

CNGP will provide useful genetic data for future studies.

Acknowledgments

We would like to thank Editage English editing service for the help in polishing our paper.
Funding: None.

Footnote

Reporting Checklist: The authors have completed the SPIRIT reporting checklist. Available at https://pm.amegroups.com/article/view/10.21037/pm-21-29/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://pm.amegroups.com/article/view/10.21037/pm-21-29/coif). WZ serves as an Executive Editors-in-Chief of *Pediatric Medicine*. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The present study is performed in accordance with the Declaration of Helsinki (as revised in 2013), Good Clinical Practice, and related laws. This study was approved by the Ethics Commission of Children's Hospital of Fudan University (CHFudanU_NNICU11). Informed consent was obtained from the parents or guardians of the neonates.

References

1. 1. Verma IC, Puri RD, et al. Global burden of genetic disease and the role of genetic screening. Semin Fetal Neonatal Med 2015;20:354-63.
2. Carroll J, Wigby K, Murray S, et al. Genetic testing strategies in the newborn. J Perinatol 2020;40:1007-16.
3. World Health Organization. Available online: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/neonatal-mortality-rate-(per-1000-live-births)
4. Kingsmore SF, Henderson A, Owen MJ, et al. Measurement of genetic diseases as a cause of mortality in infants receiving whole genome sequencing. NPJ Genom Med 2020;5:49.
5. McCandless SE, Brunger JW, Cassidy SB, et al. The burden of genetic disease on inpatient care in a children's hospital. Am J Hum Genet 2004;74:121-7.
6. Wang H, Qian Y, Lu Y, et al. Clinical utility of 24-h rapid trio-exome sequencing for critically ill infants. NPJ Genom Med 2020;5:20.
7. Guthrie R, Susi A. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. Pediatrics 1963;32:338-43.
8. Wilcken B, Wiley V, Hammond J, et al. Screening newborns for inborn errors of metabolism by tandem mass spectrometry. N Engl J Med 2003;348:2304-12.
9. Hoppman-Chaney N, Peterson LM, Klee EW, et al. Evaluation of oligonucleotide sequence capture arrays and comparison of next-generation sequencing platforms for use in molecular diagnostics. Clin Chem 2010;56:1297-306.
10. Petrikin JE, Willig LK, Smith LD, et al. Rapid whole genome sequencing and precision neonatology. Semin Perinatol 2015;39:623-31.
11. Kingsmore SF, Cakici JA, Clark MM, 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 2019;105:719-33.
12. Farnaes L, Hildreth A, Sweeney NM, et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. NPJ Genom Med 2018;3:10.
13. Hays T, Wapner RJ, et al. Genetic testing for unexplained perinatal disorders. Curr Opin Pediatr 2021;33:195-202.
14. Malinowski J, Miller DT, Demmer L, et al. Systematic evidence-based review: outcomes from exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability. Genet Med 2020;22:986-1004.
15. Stark Z, Tan TY, Chong B, et al. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. Genet Med 2016;18:1090-6.
16. The National Health Commission of the People's Republic of China, 2012 (updated 2021/02/01). Available online: https://www.gov.cn/gzdt/att/att/site1/20120912/1c6f6506c

7f811bacf9301.pdf
17. The National Health Commission of the People's Republic of China, 2019 (updated 2020/02/01). Available online: https://www.nhc.gov.cn/fys/s7901/201905/bbd8e2134a7e47958c5c9ef032e1dfa2.shtml
18. Wang H, Lu Y, Dong X, et al. Optimized trio genome sequencing (OTGS) as a first-tier genetic test in critically ill infants: practice in China. Hum Genet 2020;139:473-82.
19. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43:11.10.1-33.
20. Kleinberger J, Maloney KA, Pollin TI, et al. An openly available online tool for implementing the ACMG/AMP standards and guidelines for the interpretation of sequence variants. Genet Med 2016;18:1165.
21. Richards S, Aziz N, Bale 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 2015;17:405-24.
22. Dong X, Liu B, Yang L, et al. Clinical exome sequencing as the first-tier test for diagnosing developmental disorders covering both CNV and SNV: a Chinese cohort. J Med Genet 2020;57:558-66.
23. Yang L, Kong Y, Dong X, et al. Clinical and genetic spectrum of a large cohort of children with epilepsy in China. Genet Med 2019;21:564-71.
24. Wang H, Xiao F, Dong X, et al. Diagnostic and clinical utility of next-generation sequencing in children born with multiple congenital anomalies in the China neonatal genomes project. Hum Mutat 2021;42:434-44.
25. Liu B, Lu Y, Wu B, et al. Survival Motor Neuron Gene Copy Number Analysis by Exome Sequencing: Assisting Spinal Muscular Atrophy Diagnosis and Carrier Screening. J Mol Diagn 2020;22:619-28.

doi: 10.21037/pm-21-29
Cite this article as: Xiao F, Yan K, Wang H, Wu B, Hu L, Yang L, Zhou W. Protocol of the China Neonatal Genomes Project: an observational study about genetic testing on 100,000 neonates. Pediatr Med 2021;4:28.

Supplementary

Basic clinical features table

1. Basic information
1.1 Sex: □ male □ female
1.2 Time of birth (mm/dd/yy):
1.3 Gestational age: _____weeks
1.4 Apgar score: _____
1.5 Resuscitation: □ Yes No □
1.6 The mode of delivery: □ Vaginal delivery □ Cesarean section
1.7 Birthweight: _____Kg
1.8 Weight on admission: _____Kg
1.9 Blood pressure: SBP_____/DBP_____mmHg
1.10 Heart rate: _____bpm
1.11 family history: □No □ Yes _____

2. The maternal basic information
2.1 Age of pregnancy: _____years
2.2 □ singleton □ twins □ multiple
2.3 G___P____
2.4 Abnormal placenta: □ Yes □ No
2.5 Abnormal amniotic fluid: □ Yes □ No
2.6 Abnormal umbilical cord □ Yes □ No
2.7 Condition:
□none □diabetes □hypertension □hyperthyroidism □hypothyroidism
□infection disease □others:_____
3 Clinical features
3.1 antenatal and prenatal period
□ normal □ intrauterine growth retardation, IUGR □ oligohydramnios
□ polyhydramnios □ nuchal Translucency thickening
□ echogenic intracardiac focus □ intestinal echo focus
□ others_____
3.2 facial features
□ normal □ cataract □ cleft lip and palate □ eye deformity
□ tooth abnormality □ abnormal external ear
□ others_____
3.3 cardiovascular system
□ normal □ cardiomyopathy (type_____) □ arrhythmia
□ stenosis of aorta □ atrial septal defect □ ventricular septal defect
□ others_____
3.4 skin
□ normal □ hair abnormalities □ nail abnormality □ blister
□ connective tissue abnormalities □ pigmentation □ hypopigmentation
□ others_____
3.5 nervous system
□ normal □ hypotonia □ hypertonia □ seizure □ epilepsy
□ stroke □ abnormal brain structure □ macrocephaly □ microcephaly
□ others_____

https://dx.doi.org/10.21037/pm-21-29

3.6 respiratory system
□ normal □ respiratory distress □ diaphragmatic hernia
□ tracheo-esophageal fistula □ chest deformity □ pulmonary mass
□ others_____
3.7 digestive system
□ normal □ abdominal defect □ pyloric stenosis □ gastroesophageal reflux
□ vomit □ diarrhea □ constipation □ imperforate anus
□ Hirschsprung's disease □ hepatomegaly □ others_____
3.8 skeletal system
□ normal □ contracture □ malformed foot □ digit malformation
□ syndactyly □ limb abnormality □ scoliosis □ others_____
3.9 endocrine system
□ normal □ diabetes □ hypothyroidism □ hyperthyroidism
□ pheochromocytoma □ others_____
3.10 metabolism and mass spectrometry
□ normal □ hypoglycemia □ ketone disease □ lactic acidosis
□ increased pyruvate □ increased alanine □ organic acidurias
□ organic acidemias □ hyperaminoaciduria □ hyperaminoacidemia
□ others_____
3.11 blood system
□ normal □ anemia □ neutropenia □ thrombopenia □ pancytopenias
□ bleeding disorder □ thrombosis □ immunodeficiency □ leukemia
□ others_____

https://dx.doi.org/10.21037/pm-21-29