High-risk phenotypes of genetic disease in a Neonatal Intensive Care Unit population

Tiantian Xiao[1], Qi Ni[2], Huiyao Chen[2], Huijun Wang[2], Lin Yang[2], Bingbing Wu[2], Yun Cao[1], Guoqiang Cheng[1], Laishuan Wang[1], Liyuan Hu[1], Hongfang Mei[1], Yulan Lu[2], Mengchun Gong[3], Xinran Dong[2], Wenhao Zhou[1,2]

[1]Department of Neonatology, Key Laboratory of Neonatal Diseases, Ministry of Health, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai 201102, China;
[2]Center for Molecular Medicine, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai 201102, China;
[3]Institute of Health Management, Southern Medical University, Guangzhou, Guangdong 510515, China.

To the Editor: Genetic diseases contribute to 35% of deaths during the first year of life and are a significant cause of intensive care.[1] A previous study based on the China Neonatal Genomes Project investigated the genetic causes of early infant deaths and found that >25% of deceased neonates with genetic diagnoses can be cured if diagnosed in time.[2] Therefore, it is crucial to target and diagnose neonates with genetic diseases as early as possible. According to our experience, the typical phenotypes, such as special facial features or multiple congenital anomalies (MCAs), indicate a high risk of genetic disease and lead physicians to perform genetic testing in neonates as early as possible. However, in practice, infants without typical phenotypes typically undergo a long and costly diagnostic process before genetic diagnoses are confirmed. Moreover, a recent survey by the American College of Medical Genetics and Genomics (ACMG) and other national professional organizations indicated that there are insufficient numbers of qualified geneticists to fulfil genetic service needs.[3] The ACMG published the general clinical features for genetic testing indications. For example, patients with phenotypes or family history data that strongly implicate a genetic cause may undergo genetic testing.[1] However, the study indicated that many genetic conditions arise *de novo* or are inherited with no family history.[1] A previous study attempted to apply the non-phenotype-driven panel approach in neonates admitted to the neonate intensive care unit (NICU).[4] However, at present, the diagnostic yield is only 3.45% (1/29).[4] In addition, the economic and ethical issues associated with genomic screening remain challenging. Therefore, the available indications for genetic testing may improve the management of genetic diseases.

The electronic medical record (EMR) contains a massive amount of data representing the diversity of the patients' clinical information. The EMR data are critical for research on genetic diseases, and several phenotyping pipelines have already been validated to extract clinical features from it.[5] However, to our knowledge, clinical findings are variable and documented based on the physician's experience and training. Moreover, the massive expressions of clinical findings hamper the analysis of character phenotypes in genetic diseases. Accordingly, we used Human Phenotype Ontology (HPO) which can normalize expressions of clinical findings and provide the most comprehensive resource for computational phenotyping. We developed an Auto-Neo-HPO pipeline and investigated the risk phenotypes suggestive of genetic diagnoses in NICU population via this pipeline.

The study was approved by the Ethics Committee of the Children's Hospital of Fudan University (No. CHFudanU_NNICU11). Informed consent for genetic testing was provided by biological parents or guardians. The study was conducted in accordance with the *Declaration of Helsinki* (as revised in 2013). The clinical trial registration number was NCT02551081.

We conducted an observational study in a large tertiary NICU in the Children's Hospital of Fudan University. The patients were enrolled from June 1, 2016 to June 30, 2020. The inclusion criteria were as follows: (1) postnatal age of <28 days; (2) infants aged above 35 weeks; (3) hospital stay for at least 24 h; and (4) informed consent for genetic testing provided by biological parents or guardians. The exclusion criteria were as follows: (1) neonates with MCAs

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000001959

Correspondence to: Xinran Dong, Center for Molecular Medicine, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai 201102, China
E-Mail: xrdong@fudan.edu.cn

Chinese Medical Journal 2022;135(5)
Received: 13-06-2021; **Online:** 13-01-2022 **Edited by:** Jing Ni

defined as two or more structural malformations at birth[6]; (2) missing or low-quality clinical information.

We targeted EMR data and clinical exome sequencing data for all participants. We assessed the demographic characteristics, gestational age, birthweight, clinical phenotypes at discharge, and outcomes at discharge. The risk phenotype candidates were the HPO terms that were over-represented ($P < 0.05$, odds ratio [OR] > 1) in infants with genetic diagnoses compared with those in infants without genetic diagnoses. We selected neonates with genetic diagnoses as the case group. The control group included neonates without genetic diagnoses. Genetic diagnoses, sequencing data analysis, and pathogenic variant judgment are described in Supplementary Material, http://links.lww.com/CM9/A905. The detailed definitions of clinical features of the infants are illustrated in Supplementary Material, http://links.lww.com/CM9/A905.

A tool named Auto-Neo-HPO (Registration No. V1.0, 2018SR215790) was applied to assist in HPO term extraction. Briefly, the core of Auto-Neo-HPO included the local semantic database in both Chinese and English versions (with information initially obtained from the HPO and ChinaHPO databases) and a natural language processing pipeline to transfer the non-database-matched phrases into HPO terms. Two experienced geneticists and two neonatologists revised the Auto-Neo-HPO output HPO terms and updated the semantic database if required. Next, several HPO terms were modified according to the special consideration of phenotype conditions in neonates. The performance of the Auto-Neo-HPO pipeline is described in Supplementary Material, http://links.lww.com/CM9/A905.

Descriptive statistics were used to establish the clinical parameters of the study group. For continuous variables, the median, maximum, and minimum values were reported; for categorical variables, percentages were reported. We only calculated the frequency of HPO terms counted >10 times in the study and compared the frequency of HPO terms between the group with genetic diagnoses and the group without genetic diagnoses. The *P* value was determined by Fisher's exact test and adjusted by multivariate logistic regression, and the significance threshold was set at 0.05. The estimated ORs and 95% profile likelihood confidence intervals (CIs) were reported.

The study population is illustrated in Supplementary Figure 1, http://links.lww.com/CM9/A905. Demographic information and clinical features are described in Supplementary Table 1, http://links.lww.com/CM9/A905. A total of 2600 neonates with a wide range of neonatal diseases and genetic etiologies were enrolled, including 248 neonates (9.5%) with genetic diagnoses. Among the 2600 neonates, 1554 (59.8%) were male and 1046 (40.2%) were female. In total, 168 (6.5%) neonates died in the hospital or received palliative care, including 33 (13.3%) neonates with genetic diagnoses and 135 (5.7%) without genetic diagnoses.

Of 93 HPO terms counted >10 times in the study, 25 HPO terms were documented in the group with genetic diagnoses. Abnormal heart morphology (human phenotype [HP]: 0001627; 49.2%, 1279/2600) was the most common phenotype in an NICU population, followed by jaundice (HP: 0000952; 47.3%, 1229/2600), and sepsis (HP: 0100806; 42.3%, 1101/2600) [Supplementary Table 2, http://links.lww.com/CM9/A905]. Compared to HPO terms in neonates without genetic diagnoses, muscular hypotonia (HP: 0001252; $P_{adjust} < 0.0001$, OR_{adjust}: 3.41, 95% CI 1.63–6.80), seizure (HP: 0001298; $P_{adjust} < 0.0001$, OR_{adjust}: 2.47, 95% CI 1.73–3.50), and cryptorchidism (HP: 0000028; $P_{adjust} = 0.0233$, OR_{adjust}: 3.36, 95% CI 1.11–9.27) were the high-risk phenotypes in neonates with genetic diagnoses, whereas jaundice ($P_{adjust} < 0.0001$, OR_{adjust}: 0.57, 95% CI 0.47–0.77) and meningitis (adjust $P_{adjust} = 0.0153$, OR_{adjust}: 0.17, 95% CI 0.03–0.56) were low-risk phenotypes in neonates with genetic diagnoses. The abnormality of metabolism/homeostasis (HP: 0001939) was not significantly different between neonates with genetic diagnoses and those without genetic diagnoses ($P = 0.6100$, OR: 0.90, 95% CI 0.62–1.27) [Figure 1 and Supplementary Table 3, http://links.lww.com/CM9/A905].

Genetic diagnosis is difficult. Many neonates with a high risk of genetic diagnoses may not undergo proper examination and effective treatments in a timely manner in the non-tertiary NICU because of limited genetic resources and physicians' knowledge. In some cases, neonates underwent ineffective treatments, which may have had adverse effects or exacerbated symptoms before they were transferred to the referred NICU. In practice, physicians usually perform genetic testing in neonates with MCAs or typical syndromes owing to the high clinical suspicion of genetic diseases. However, in neonates without typical clinical findings, it is difficult for physicians to decide whether to perform genetic testing or determine the type of genetic testing to be performed. To our knowledge, we fully describe the clinical phenotypes related to genetic diagnoses based on data science in an NICU population, and our study suggested the three risk phenotypes suspected of genetic diagnoses.

Our study suggested that 9.5% of neonates admitted to NICU at our institution had genetic diagnoses. We identified three high-risk phenotypes including muscular hypotonia (HP: 0001252), seizure (HP: 0001298), and cryptorchidism (HP: 0000028). Therefore, in addition to structural malformations and special facial features, these

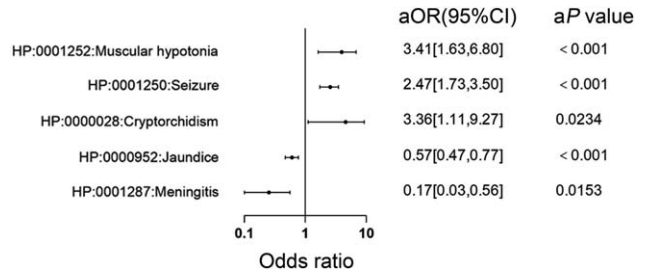

Figure 1: The risk phenotypes associated with genetic diagnoses. aOR: Adjusted odds ratio; a*P* value: Adjusted *P* value; CI: Confidence interval; HP: Human phenotype; OR: Odds ratio.

three phenotypes could be considered as the clinical indicators to perform further genetic testing. Among these risk phenotypes, jaundice (HP: 0000952) should be interpreted cautiously. Our study suggested that it was a low-risk phenotype, indicating that the near-term infants with jaundice may be at a low risk of genetic diagnoses. However, the age of the onset time of jaundice was not analyzed in our cohort. Infants with prolonged jaundice may be related to liver disease. This condition is highly suggestive of a genetic condition. On the other hand, neonates with metabolic phenotypes are suspected to have a genetic disease in practice. However, the abnormality of metabolism/homeostasis (HP: 0001939) was not a risk phenotype for genetic diagnoses in our study. Some metabolic phenotypes, such as hyperinsulinemia (HP: 0000842), glutaric aciduria (HP: 0003150), lactic acidosis (HP: 0003128), and hyperlipidemia (HP: 0003077) were counted <10 times, and we manually merged them into the parental HPO. However, those phenotypes were strongly associated with genetic diagnoses based on clinical experience. Therefore, further study should be investigated for these metabolic phenotypes in detail. Those infants presenting the metabolic disorders had to be investigated for the underlying etiologies based on the current clinical experience.

Regarding the clinical terms mapped to the HPO terms, we faced some issues. First, one clinical term could be translated to the different HPO terms via the Auto-Neo-HPO pipeline. For example, there could be jaundice or neonatal hyperbilirubinemia documented in the EMR, and the Auto-Neo-HPO pipeline could map them into different HPO terms, such as jaundice (HP: 0000952), hyperbilirubinemia (HP: 0002904), neonatal hyperbilirubinemia (HP: 0003265), yet, they were the same clinical feature. Therefore, we merged these same or similar HPO terms to one HPO term based on the suggestion of the experienced neonatologists. Second, when the infants were admitted to NICU, there could be ambiguous features. For example, some neonates could have congenital heart diseases (CHD) found during the fetal period, yet, the types of the CHD are usually not definitive. Therefore, we merged all HPO terms related to the different types of CHD but PDA in one HPO term (Abnormal heart morphology [HP: 0001627]) in our cohort. PDA is not considered to be a congenital malformation if it is identified in the early neonatal period, thereby, we analyzed the PDA independently. The merged HPO terms were described in Supplementary Table 4, http://links.lww.com/CM9/A905.

There are several limitations to our study that might have led to bias. First, we only extracted the phenotypes from a clinical context and HPO terms did not include a description of the severity of the clinical findings. Further, in-depth analyses of phenotypes, such as electroencephalography, magnetic resonance imaging, and the severity of the clinical features, may provide more accurate phenotypic information indicating the genetic disease. Second, we analyzed the clinical findings using HPO terms; however, not every clinical finding can be mapped to HPO terms. Instead, we defined similar HPO terms or the parental HPO terms based on the topological structure of HPO. Thus, it may be more accurate to use clinical terms and HPO terms together. Finally, we did not consider the age of the onset time of every phenotype, this may be important to physicians to make a decision.

In conclusion, with the Auto-New-HPO pipeline assistance, we effectively and fully investigated the phenotypes in term and near-term infants without MCAs or special facial features admitted in a large NICU population. We identified the common clinical features in an NICU population based on a data science. Furthermore, we identified that muscular hypotonia (HP: 0001252), seizure (HP: 0001298), and cryptorchidism (HP: 0000028) were high-risk phenotypes suggestive of genetic diagnoses. These risk phenotypes may be the indicators for further genetic testing.

Acknowledgments

We express our deep gratitude to the patients and the families for their willingness and cooperation in the study. The authors also wish to acknowledge various doctors in NICU and Molecular diagnostic center of the Children's Hospital of Fudan University, and the contributing members of the "China Neonatal Genomes Project (CNGP)."

Funding

The study was supported by a grant from the Shanghai Municipal Science and Technology Major Project (No. 2017SHZDZX01).

Conflicts of interest

None.

References

1. Bick D, Jones M, Taylor SL, Taft RJ, Belmont J. Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. J Med Genet 2019;56:783–791. doi: 10.1136/jmedgenet-2019-10.
2. Yang L, Liu X, Li Z, Zhang P, Wu B, Wang H, *et al*. Genetic aetiology of early infant deaths in a neonatal intensive care unit. J Med Genet 2020;57:169–177. doi: 10.1136/jmedgenet-2019-10.
3. Maiese DR, Keehn A, Lyon M, Flannery D, Watson M. Working Groups of the National Coordinating Center for Seven Regional Genetics Service Collaboratives. Current conditions in medical genetics practice. Genet Med 2019;21:1874–1877. doi: 10.1016/j.ajhg.2018.11.016.
4. Ceyhan-Birsoy O, Murry JB, Machini K, Lebo MS, Yu TW, Fayer S, *et al*. Interpretation of genomic sequencing results in healthy and Ill newborns: Results from the BabySeq project. Am J Hum Genet 2019;104:76–93. doi: 10.1016/j.ajhg.2018.11.016.
5. Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet 2011;12:417–428. doi: 10.1038/nrg2999.
6. Wang H, Xiao F, Dong X, Lu Y, Cheng G, Wang L, *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;42:434–444. doi: 10.1002/humu.24170.

How to cite this article: Xiao TT, Ni Q, Chen HY, Wang HJ, Yang L, Wu BB, Cao Y, Cheng GQ, Wang LS, Hu LY, Mei HF, Lu Y, Gong MC, Dong XR, Zhou WH. High-risk phenotypes of genetic disease in a Neonatal Intensive Care Unit population. Chin Med J 2022;135:625–627. doi: 10.1097/CM9.0000000000001959