The targeted exome sequencing strategy (NeoExome) for Chinese newborns with the pilot study of 3423 neonates

Lin Zou[1], Ziyang Cao[1], Xiaoyan He[2], Dongjuan Wang[2], Maosheng Gu[3], Feng Suo[3], Rong Qiang[4], Ruixue Zhang[4], ChengRong Song[4], Xiaohua Wang[5], Bo Zhu[5], Donghua Cao[6], Haihua Yu[6], Yiping Qu[7], Guosong Shen[8], Jian Wu[9], Xiaobin Wang[9], Zhengyu Jin[9], Pengpeng Wang[9], Jinxia Wang[1], Hongyang Zhang[1], Zijun Yan[1], and Guangjun Yu[1]

[1]Children's Hospital of Shanghai
[2]Chongqing Medical University
[3]Xuzhou Maternity and Child Health Care Hospital
[4]Northwest Women and Children's Hospital
[5]The Affiliated Hospital of Inner Mongolia Medical University
[6]Dalian Medical University
[7]Zhejiang University School of Medicine Children's Hospital Hubin Campus
[8]Huzhou Maternity and Child Care Hospital
[9]Research and Development Department MyGenostics Inc Beijing China

May 2, 2022

Abstract

Newborn screening (NBS) is an effective way for 3-step prevention of birth defects. The suitable technology and rational NBS screening diseases are critical for each country and area. High-throughput sequencing has shown high application potential in NBS. However, lack of sequencing strategy for monogenic inherited diseases NBS in China. In this study, we systematically evaluated the application efficiency of different sequencing approaches for NBS, and a gene-disease association list (NeoExome panel) for the Chinese population with 601 genes was designed based on the top rare disease list and databases. In the 1000 Genomes Project, 7.6% (23/301) were NGS positive. Among the 3249 neonates recruited, NGS positive rate was 12.0%. In the 200 conventional NBS (+) subgroup, 118 were NGS positive, with 76.3% (90/118) neonates harboring consistent results of conventional NBS and NGS; in the conventional NBS (-) subgroup, the NGS positive rate was 8.9% (271/3049). Our study designed a personal NBS targeted-sequencing NeoExome panel of monogenic inherited diseases for Chinese, which has shown acceptable performance.

1 Introduction

Birth defects (also known as congenital anomalies) are the major causes of neonatal deaths, worldwide. According to the report on prevention and treatment of birth defects-Ministry of Public Health of China in 2012, the incidence of birth defects in China is about 5.6%, with approximately 900000 new cases of birth defects every year[1]. To reduce the occurrence of birth defects, the World Health Organization (WHO) develops prevention strategies into three levels based on the causes and epidemiology of birth defects: preconception care to increase the likelihood of a healthy infant delivery, pregnancy care to reduce the birth rate of defective infant, newborn infant and child care to decrease disability, mortality, and serious consequences of birth defects[2]. 3 periods-screening (Preconception screening, Peri-conception screening, Neonatal screening) prevent different diseases of birth defects. Newborn screening (NBS) is the last step against birth defects,

which aims to identify seriously harmful diseases in the neonatal period. It is helpful for early diagnosis, early intervention (as reducing exposure to risk factors), and early management of the diseases, thus preventing mortality and morbidity of children[3,4].

NBS begins in the 1960s when Professor Guthrie firstly applied the method of bacterial inhibition to screen phenylketonuria (PKU) in the dried blood spots[5]. Since then, NBS has gradually expanded under the screening guiding principles with the progression of clearly-investigated diseases, the feasibility of screening methods, and the increment of public conscientious [6,7]. In 2006, the American college of medical genetics (ACMG) expert group published and continually updated the Recommended Uniform Screening Panel (RUSP) for NBS consensus with 35 primary and 26 secondary conditions[8]. While in China, NBS started in 1981, mainly focused on the screening of PKU, congenital hypothyroidism (CH) firstly, followed by deafness, glucose-6-phosphate dehydrogenase (G6PD) deficiency, congenital adrenal hyperplasia (CAH), and so on[9]. With the expansion of screening diseases, NBS methods have developed from the initial low throughput biochemical approach to high throughput molecular biology technology. Tandem mass spectrometry (MS) is the main conventional NBS method and has been widely used in clinical practice for the last 20 years, with the feature of rapid, sensitive, and high throughput[10]. Although MS accelerated the development of NBS, it is now well established from a variety of studies that MS has a high false-positive/negative ratio[4,11,12]. Besides, conventional NBS covered a limited range of diseases. New methods are urgently needed for complementing the shortcomings of MS and other conventional NBS.

Recently, high-throughput sequencing technology has been widely used in tumor-targeted genes test[13], pathogen detection[14], as well as 3 periods-screening of birth defects[15]. Laboratories have successfully carried out studies on sequencing technology as an NBS method[16-18] and indicated sequencing could be used as a second-tier confirmation for conventional NBS of some diseases[19,20]. Besides, the ability of sequencing for the gene diagnosis of Neonatal Intensive Care Units (NICU) patients has also been uncovered by multiple studies[21,22]. However, there are several concerns about the use of sequencing in NBS. 1) Interpretation of the gene variant site. Sequencing may identify lots of variants with uncertain significance (VUS), while the relationship between many genes and diseases is not clear. 2) Report of the results. There is no gene-disease association list about conditions that are treatable and preventable so far. 3) Personalized sequencing panel. Sequencing protocols and panels specifically designed for NBS are lacking. 4) Clinical trials of the whole population. Most of the previous studies focused on the use of sequencing in ill infants, or for the screening for a specific disease, studies about multiple monogenic-disease screening in healthy newborns are not fully understood. These big challenges for the application of sequencing in NBS are presented and aroused great interest in this field.

Based on the above-mentioned challenges, researchers devoted to establishing proper criteria of gene-disease association list and reporting strategies for newborn gene sequencing since the year 2012 [23,24]. However, this gene-disease association list and reporting strategy cannot be fully applied, as the high incidence of monogenic diseases in the Chinese population is quite different from that of other countries, due to different races and geographical areas. Here, we systemically evaluated the NGS methods used in NBS, designed the severe, actionable, and early-onset inherited gene-disease list for the Chinese population, established the featured NGS panel (NeoExome), performed a multi-center study to verify its feasibility with 3423 neonates, and compared the NeoExome panel with other NGS panels of NBS. (Supplementary Fig. 1)

2 Methods

2.1 Establishment of the NGS panel

Generation of gene-disease association : Monogenic inherited diseases were collected according to the criteria of covering diseases that are severe, actionable, and early onset (refer to the result section), and genes with relatively clear relationship with diseases were determined based on Online Mendelian Inheritance in Man (OMIM) database (https://www.omim.org/).

Age of onset : According to the diseases records in OMIM database, the known minimum onset age of diseases is manually sorted out, which can be divided into 6 categories: <1 year of age (Infants), 1-3 years

of age (Toddlers), 3-6 years of age (Preschoolers), 6-12 years of age (Middle Childhood), 12-18 years of age (Young Teens and Teenagers), > 18 years of age (Adulthood).

2.2 Ethical compliance

This study was approved by the Ethics Committee of Children's Hospital of Chongqing Medical University (2019-R-171-1). All guardians of neonates enrolled have signed the informed consent.

2.3 Subjects

The study has been registered in ClinicalTrials.gov (Number: NCT03984266). *Inclusion criteria* (meet the Both): All newborns (including hospitalized infants and infants with abnormal results of conventional NBS); the guardian signs the informed consent and agrees to participate. *Exclusion criteria* (Meet one or more of the following criteria): Other similar clinical studies are underway; received transfusion of allogeneic blood products in recent 2 weeks.*Rejection criteria* : Specimen cannot be tested due to improperly collected or stored; samples with unstandardized data; Samples without follow-up data; Guardian's request to withdraw.

From Oct 2019 to Sep 2021, a total of 3423 neonatal subjects were enrolled from 5 hospitals: Children's Hospital of Chongqing Medical University, Inner Mongolia Maternal and Child Health Hospital, Northwest Women and Children's Medical Center, Dalian Maternal and Child Health Hospital, and Xuzhou Maternal and Child Health Care Hospital.

2.4 Sample preparation

The heel blood of the neonatal subjects was collected into a special filter paper, dry naturally at room temperature to form dried blood spots (DBS). After conventional NBS, the DBS was stored at 4 . Genomic DNA was extracted from DBS using QIAamp DNA Mini Kit (Qiagen, German) according to the operation manual. Briefly, a punched-out circle from a DBS was placed into a 1.5 ml microtube with Buffer ATL to dissolve DNA and incubated at 85°C for 10 min. Protein K was added to digest proteins. After 70°C-incubation and adding ethanol, the buffer was transferred into a Mini spin column. Finally, the DNA solution was eluted with Elution buffer after a series of centrifugation. The quality and quantity of the DNA were assessed by Qubit® 3.0 Fluorometer (Thermo Fisher Scientific Inc., USA) according to the instruction.

2.5 Library construction and sequencing

Library construction was performed according to the standard procedures of Illumina (Illumina, Inc., USA), which included terminal repair, adaptor connection, and PCR enrichment briefly. The Neonatal Gene Capture Kit- NeoEXOME (MyGenostics GenCap(r) Enrichment technologies, China) was designed based on the targeted genes in our project, and used for hybridization capture of relevant target regions. The final library was constructed after hybridization capture, PCR, and then determined by Qubit(r) 3.0 Fluorometer and Agilent 2100 Bioanalyzer system (Agilent Technologies, USA). After QC (quality control), the enriched library was sequenced by using HiSeq X Ten System (Illumina, Inc., USA) for 150-bp double-terminal sequencing[22].

After sequencing, the reads were compared to the UCSC HG19 reference genome, and variants were identified using a bioinformatics process set by MyGenostics (based on GATK (The Genome Analysis Toolkit)). Variation was annotated using ANNOVAR databases. Mutations with high population frequency were filtered reference to dbSNP 1381000 Genome Project, ESP6500SI, and ExAC (the Exome Aggregation Consortium) browser.

2.6 Probe design description:

The design of liquid-phase hybridization capture probe was based on the GenCap technology of MyGenostics[25]. In addition to designing probes that cover the exon region of related genes (covering the exon region, exon flanks ˜50bp, and non-coding disease-causing regions reported by HGMD database), encryption-designing probes were carried out for 25 CNV high-risk genes (increasing the coverage of 300bp intron region

by designing probes on the exon flanks of related genes). On the other hand, designing probes also included full-length coverage of*HBA1 /HBA2 /HBB /SMN1 /SMN2* genes (coding region and intron region of coverage gene). The probes covered hot spot variation of mitochondrial genes, *MT-RNR1* , *MT-TL1* ,*MT-ND3,* and *MT-ATP6* . Some regional genes of Prader-William/Angleman, DiGeorge, and Williams in microdeletion syndrome were covered, and the probe encryption was beneficial for the analysis of copy number variation in related regions. Meanwhile, according to the GC content, region size, Tm value and other parameters of the target region, the length, density, and position of the probe are adjusted based on the Probe design software BaitDesigner of MyGenostics to improve the capture efficiency of the probe.

2.7 Gene variants interpretation

Based on the ACMG guidelines, the pathogenicity of the variation was classified into five subtypes, including pathogenic, likely pathogenic, unknown, likely benign, and benign. Then the disease risk of the examined neonates was graded to three-level risks (high risk, moderate risk, and low risk) and carrier, according to the pathogenicity of the variation and the genetic inheritance pattern (Supplementary Figure 2). Each grade is defined as follows:

High risk : A pathogenic or likely pathogenic variant of the genes that are autosomal dominant or Y-linked; Two pathogenic or likely pathogenic variants of the genes that are autosomal recessive; Female newborns carry two pathogenic or likely pathogenic variants of X-linked dominant genes; Male newborns carry a pathogenic or likely pathogenic variant of X-linked recessive genes.

Moderate risk : A pathogenic or likely pathogenic variant, and a potent variant of unknown significance of the genes that are autosomal recessive; Female newborns carry a pathogenic or likely pathogenic variant, and a potent unknown significance variant of X-linked recessive genes.

Low risk : A pathogenic, likely pathogenic, or unknown significance variant, and a nonvirulent variant with unknown meaning of the genes that are autosomal recessive; A unknown significance variant of the genes that are autosomal dominant or Y-linked; Female newborns carry a pathogenic, likely pathogenic, or unknown significance variant and a nonvirulent variant with unknown meaning of X-linked recessive genes.

Carrier : A pathogenic or potentially pathogenic variation of autosomal recessive genes; Female newborns carry a pathogenic or likely pathogenic variant of X-linked recessive genes.

2.8 Venny analysis

The Venny 2.1 software was used to compare different NBS NGS panels with our panel through an interactive tool[26](*https://bioinfogp.cnb.csic.es/tools/venny/*). Multivariate Venn's Diagrams was drawn by R version 4.0.5 software (*https://www.r-project.org/*).

3 Results

3.1 Selection of sequencing methods for neonatal screening

To select the appropriate detection method, we firstly compared different sequencing methods for NBS used in the previously published studies (Table 1). In the beginning, scientists conducted neonatal screening trials by using whole exome sequencing (WES) or whole genome sequencing (WGS), including WGS for STATseq project[16], WES for BabySeq project[17], NBSeq project[27], and North Carolina Newborn Exome Sequencing for Universal Screening (NC NEXUS) project[18]. These studies implied sequencing may screen out inborn errors before the disease onset and WGS/WES is operationally feasible in neonatal screening[28-32]. However, WGS/WES analysis produces lots of VUS, shows the higher cost and longer turn-around time (TAT). Consequently, scientists explored the application of targeted NGS panels that only analyzed a subset of gene loci in NBS. Notably, studies suggested that targeted sequencing has more advantages over WGS/WES as the operation of data analysis/follow-up is more feasible and easier, with lower cost, shorter TAT, and easier interpretation[33,34]. Additionally, there is little difference in positive rate between targeted NGS and WGS/WES[35]. Therefore, we choose targeted sequencing in our study.

3.2 Generation and feature of NeoEXOME panel

OMIM database was used to standardize and screen out the diseases and related genes. To design the NBS NGS panel covering monogenic diseases that are severe, actionable, and early onset for the Chinese population, we combined the following categorization of diseases in our study: 1) 248 related genes (249 diseases) that were included in China National Catalog of Rare Diseases. This catalog includes diseases with low incidence, high risk, and strong treatability based on the clinical data; 2) 360 genes (312 diseases) that were included in Preventable and Treatable Rare Diseases in China; 3) 79 genes (71 diseases) that were now in the conventional MS/MS NBS project in China; 4) 71 genes (71 diseases) that were included in the conventional NBS in China; 5) 149 genes (148 diseases) that were recommended by ACMG; 6) 55 genes (57 diseases) that were recommended by the experts of the Chinese NBS Laboratory Committee. Moreover, a list of diseases with high prevalence in the Chinese population based on the database about positive monogenetic diseases in 40000 ill newborns by MyGenostics was also included. This list included 141 diseases and 94 related genes. Then Venny analysis was conducted to identify and generate the disease-genes panel (named NeoEXOME) in our study (Fig 1A). Our NeoEXOME panel is comprised of 601 genes and 542 kinds of diseases (Supplementary Table1).

The panel covered diseases with multi-systems, including skeletal, respiratory, urinary, immune system, nervous system, cardiovascular system, blood, endocrine system, metabolism, and mitochondria-related diseases (Fig 1B). Among them, diseases of the metabolic system account for the most proportion in our NeoEXOME panel. Besides, the onset age of these gene-related diseases was also collected and shown in Fig 1C. 418 out of 601 gene-diseases were onset in the neonatal period (<1 year), accounting for the largest proportion. 577 out of 601 (96%) gene-diseases were onset in childhood (<18 years).

3.3 Performance of NeoEXOME

Results of " 1000 Genomes"

Firstly, we validated our panel using the data of unrelated 301 Chinese in the "1000 Genomes Project". According to the interpretation criteria, 72 were negative, 223 were carrier, and 23 were positive (7.6%). Variants of *FLG* and*GJB2* account for the most frequent mutant genes, with the frequency 47.8% (11/23) and 21.7% (5/23), respectively (Supplementary Table 2).

A pilot study of the performance of NeoEXOME for NBS in China

To evaluate the performance of NeoEXOME in clinical practice. A pilot, multi-center clinical trial was conducted. From Oct 2019 to Sep 2021, 3423 neonates were enrolled from 5 institutions (Fig 2). Of all these subjects, DBS was collected within 3 days of birth and next-generation sequencing was performed based on our NeoEXOME panel.

Among the 3249 neonates eligible for the analysis, 934 infants were NGS negative, and 389 infants were NGS positive (244 High risk, 4 Moderate risk, 141 Low risk), and 1926 were carriers. NGS positive rate was 12.0%. 343 infants showed one genetic mutation, who were genetically susceptible to mono gene-related diseases; 46 infants had more than one gene mutations, who were genetically susceptible to multi-diseases. Infants that were NGS reported as high-risk account for the highest proportion (Fig 3A-B). While autosomal recessive inheritance was the most common inheritance pattern (Fig 3A-B). Endocrine and metabolism system disorders accounted for the highest proportion in healthy neonates (Fig 3C). Most of the diseases detected were predicted to develop within the first year of life (Fig 3D). Variants of *DUOX2*account for the most frequent mutant genes (18.3%, 131/716), followed by *UGT1A1* , *PAH* , *GJB2* , *FLG* (Fig 3E). The top five related disorders were thyroid dyshormonogenesis, hyperbilirubinemia, phenylketonurics, deafness, hyperbilirubinemia, and ichthyosis vulgaris, with the frequency of variant 5.7%, 5.4%, 5.0 %, 4.1%, respectively. 1926 infants were tested as carrier of the diseases in our panel. There were 3462 variants, including 1224 pathogenic variants and 2238 likely pathogenic variants (Fig 4A). Variants of*GJB2* account for the most frequent mutant genes of all the centers, followed by *UGT1A1* , *DUOX2, FLG,* and*SLC25A13* (Fig 4B).

To further validate the use of NGS in NBS, conventional NBS results were collected to compare the consistency of both approaches. 200 out of 3249 neonates were conventionally NBS positive (+). In the NBS (+) subgroup, 118 were NGS positive (64 High risk, 2 Moderate risk, 52 Low risk), 14 were NGS negative, and 68 were NGS carrier (Table 2). Of the 118 NBS (+)/NGS (+) cases, 90 (76.3%) had consistent results with conventional NBS and NGS (Supplementary Table 3). In the 3049 NBS (-) subgroup, 271 (8.9%) were NGS positive (180 High risk, 2 Moderate risk, 89 Low risk), 920 were NGS negative, and 1858 were NGS carrier (Table 2). 168 of the 271 NGS (+) /NBS (-) neonates were followed up till Dec 2021, and 9 of them (including genes of *DUXO2* , *PAH* , *MUT* , *WAS* , and *SLC22A5*) were clinically diagnosed.

3.4 Comparison of NeoEXOME with other NBS panels

To find out the difference between our NeoEXOME panel and other NGS panels and WGS gene-disease lists worldwide, we compared our panel with other NBS panels in terms of gene-disease system and disease onset age. The gene-diseases information of "BabySeq"[23], "NC NEXUS"[24], and "NESTS"[36] were collected. There were 414 common genes in BabySeq and our panel (Fig 5A), most of which are endocrine and metabolism system disease-genes. Nervous, ENT (ear, nose, and throat) and syndromes system disease-genes were higher in the BabySeq panel. Also, the BabySeq panel included gene-disease pairs of digestive system and tumor, which was not covered in our panel. There were 268 common genes in NC NEXUS and our panel (Fig 5B). ENT, respiratory, and syndromes system disease-genes were higher in the NC NEXUS panel. Similar to the BabySeq panel, NC NEXUS also covered gene-disease pairs of digestive system and tumor. In addition, we compared our panel with the "NESTS" panel for Chinese newborns recently reported by Dr. Li. There were 191 common genes in the two panels (Fig 5C). Deafness-related genes dominated the gene list of the NESTS panel. Genes related to the respiratory system were higher in the NESTS panel. 2 genes *MUTYH* and *APC* that were related to multiple colorectal adenomas, and adenomatous polyposis coli were also included in the NESTS panel. However, the NESTS panel did not include genes of *DUOX2* , *UGT1A1* , *FLG,* and *ATP7B* , which showed a high positive rate in our NeoEXOME panel. In general, the gene-disease association list designed by different projects focused on varied disease catalog and have their own characteristics.

4 Discussion

NGS plays an important role in the whole 3-step prevention to control birth defects. For the primary prevention of birth defects, Carrier screening using the NGS approach should be carried out for couples who were willing to give birth, and genetic counseling and birth guidance should be given to couples who were positive for genetic screening. Genetic screening for Carrier focuses on diseases that were leading to severe genetic defects in children[37]. For the second-step prevention of birth defects, Peri-conception screening using the NGS approach could detect chromosomal abnormalities of fetal free DNA in maternal peripheral blood, mainly involved in diseases caused by chromosomal abnormalities, such as Trisomy 21, 18, 13 syndromes[38]. For the third-step prevention of birth defects, NBS using the NGS method was suggested to conduct in neonates to find out the diseases seriously harmful, mainly focused on serious and actionable diseases[3]. In our study, we aimed to determine the suitable technology, rational diseases and gene-disease association list for newborn sequencing screening. Firstly, we systemically evaluated the NGS methods to select the targeted sequencing for NBS. We then designed a NeoExome panel covering diseases that are severe, actionable, and early-onset for the Chinese population, and verified it in multiple NBS centers. We also compared our NeoExome panel with other NBS NGS panels to illustrate the respective characteristics. Our study designed a targeted-sequencing NeoExome panel for Chinese in NBS with applicable performance.

The gene-disease list is critical in NGS panel design. The study of BabySeq project curated a catalog of gene-disease pairs based on ACMG, the ClinGen clinical validity classification framework criteria, penetrance, and age of onset. They finally screened out 954 genes that are met the criteria reported in their project[23]. The NC NEXUS project, designed an age-based framework to assess the gene-disease list. They assessed 822 gene-disease pairs and divided them into 4 different categories [24]. In China, the NGS screening for NBS is also conducted recently. Dr. Yu has designed a panel of 573 genes for the screening of severe inherited disorders[36]. While Dr. Li explored the application of their 465-genes panel in clinical practise[34]. Dr. Zhao designed an NBS genetic sequencing panel including 134 genes of 74 inborn disorders using multiplex PCR[39],

and Dr. Xu investigated 164 pathogenic genes with 94 common genetic diseases[40,41]. However, all studies did not elaborate in detail on why their study included these genes. In our study, we comprehensively integrated the diseases catalog of Rare Diseases in China, routine NBS diseases in China, ACMG and mainland experts' recommendation, and database of genetic diseases in 40000 ill-children, to generate the severe, actionable, and early onset monogenic inherited diseases list of NeoExome panel with 601 genes for NBS.

The positive rate of our NGS panel in "1000 Genomes Project" is 7.6%, higher than other previous studies[42]. We found that*FLG* and *GJB2* are the most frequently mutated genes in our analysis. *FLG* is a pathogenic gene for ichthyosis vulgaris, which encodes filaggrin and plays a key role in epidermal terminal differentiation and skin barrier formation. The proportion of *FLG*gene variation in ichthyosis vulgaris was reported to be 55.6%, and the ichthyosis vulgaris subjects with *FLG* gene variant suffered more severe diseases[43]. However, *FLG* also has higher mutations in the normal population, Palmer and colleagues carried out *FLG* analysis with 1008 people of European origin and found functional deletion mutations of the*FLG* gene were approximately 9%[44]. Besides,*GJB2* gene variation is generally considered to be a common cause of non-syndromic deafness[45], while *GJB2* c.109G>A mutation was considered as a pathogenic variant with incomplete penetrance and high carrier rate in Asian[46]. Our analysis identified 5 *GJB2*variants, 4 of them was c.109G>A mutation (Supplementary Table2). Our study further demonstrated the importance of genetic counseling for such genetic variations.

It was reported that the healthcare cost of NBS using WGS is high and increased the financial burden[47,48].Subsequently, researchers implied target sequencing might highly decrease the cost as well as TAT [33,35]. In our study, the targeted sequencing of NeoExome approach is designed, carried out, and the TAT is 14 days, which is much shorter than that of WGS/WES. Moreover, the positive rate of our NeoExome panel for NBS (-) neonates was 8.9% (271/3049) in our pilot study, which is consistent with the results of BabySeq project[31]. Despite the advantages, the use of targeted sequencing in NBS increased the risk of missed detection as exome sequencing only covers approximately 1%-2% of the entire genome. Recently, one study conducted transcriptome sequencing (RNA-seq), WES, and WGS in 115 undiagnosed patients with diverse phenotypes, and the results found RNA-seq could help diagnose 17% patients that had negative results of WES/WGS[49]. This suggests that it is necessary to simultaneously conduct WGS, RNA-seq, and other methods for symptomatic children with negative results of WES/targeted NGS approach. In addition, the targeted panel should be updated with the increased evidence-based studies of some new gene-disease associations.

Moreover, it is also necessary to pay attention to whether the genes associated with these adult-onset diseases need to be detected at neonatal stages. In the previous study of the BabySeq project, a*BRCA2* pathogenic variant, which is related to a 45% risk of breast cancer and 11% risk of ovarian cancer in women, was found in one infant[31]. Because of moral distress, the researchers proposed to return these adult-onset genetic variants[50]. But it raised a strong ethical debate, Lainie FR and colleagues then published an article and argued that researchers should avoid identifying adult-onset genetic variants, as this may cause psychological impact to the child and his/her parents, and deprive the child of the right to an open future[51]. Our current panel included 12 adult-onset genes and 12 genes with unknown onset age, which we are planning to excluded in the next round. Probably, "Age-based genetic screening strategies" should be used for reference and applied in China[3].

Our study has several limitations. Firstly, lacking whole follow-up data of some recruited subjects in our validation cohort is the main disadvantage. In our study, low concordance between NGS and conventional NBS (Table 2) was indicated, which was also reported in the BabySeq project [52]. One of the reasons is that the conventional NBS data collected in our study are incomplete due to the great difference of detection indexes in different regions. On the other hand, sequencing could not fully represent phenotypes, there were always some subjects with gene-positive, but clinical phenotype-negative[34]. This may be caused by two reasons: one is that some diseases are later onset due to the differences in the time of diseases onset, while our study period is limited and these disease-related symptoms have not shown. The other is due to the low association of genes and disease. Once the penetrance of a gene variant is low, the individual with this

variant has a low chance to be symptomatic[53]. This is also one limitation in our study, we did not include penetrance in our report interpretation category. Our report category needs to be modified and the follow-up needs to be strengthened in our next study.

Although studies have shown that NGS may play a role as second-line screening in NBS and cannot replace MS, we believe that sequencing screening can be performed simultaneously with conventional NBS if the cost is reasonable. For the diseases that have no biochemical markers, such as spinal muscular atrophy (SMA), duchenne muscular dystrophy (DMD), and cardiovascular disease, sequencing could predict disease at early stage. Therefore, we proposed this NeoExome panel may have the following applications in the clinical practice: 1) As a first-tier NBS for genetic diseases with no biochemical detections; 2) As an adjunct diagnostic tool for monogenic inherited diseases with obvious abnormal phenotype; 3) As a second-tier NBS. A combination of NeoExome with conventional NBS can enhance the clinical utility of NBS. For example, sequencing screening can be further carried out in children with ambiguous-positive results in conventional NBS to improve diagnostic accuracy. In addition, the application of sequencing in healthy infants with negative conventional NBS results can help to find out the gene variants in the early stage of their lifetime to prevent mortality. 4) As an exclusion (quasi-first-line) screening for infants with low phenotypic specificity (jaundice, etc.), to assist the excluding the common genetic diseases. More specific clinical trials are needed in the future to validate the use of NeoExome in clinical practice.

In all, we designed a personal targeted-sequencing panel for Chinese NBS, evaluated this approach in a multi-center pilot study. Our study indicated our panel needs to be further optimized for the whole population screening. Besides, more specific clinical trials need to be carried out to verify the applicability of the panel in the next step.

Acknowledgments: We thank all the participants and their guardians in the research.

Conflict of Interest: The authors declare no conflicts of interest.

Funding : This work is partially supported by the Chongqing Science and Technology Bureau (cstc2019jscx-msxm0189), Shanghai Hospital Development Center (SHDC2020CR1047B) and the Natural Science Foundation of China (82070167, 81870126).

Authorship : Conceptualization, LZ, GJY, and JW.; Investigation, ZYC, XYH, DJW, MSG, FS, RQ, RXZ, CRS, XHW BZ, DHC, HHY, YPQ, GSS, JW, XBW, ZYJ, and PPW; Validation, ZYC, JXW, HYZ, ZJY, and LZ; Writing – Original Draft Preparation, ZYC and LZ; Writing – Review & Editing, ZYC and LZ; Visualization, ZYC and PPW; Supervision, LZ, and GJY; Funding acquisition, LZ, and GJY; All authors have read and agreed to the published version of the manuscript.

Data sharing Statement: All data are available in the main text or the supplementary materials.

References:

1. Ministry of Health of the People's Republic of China. The report on prevention and treatment of birth defects- Ministry of Public Health of China. 2012,*http://www.gov.cn/gzdt/2012-09/12/content_2223371.htm.*

2. World Health Organization, Sixty-third World Health Assembly, Birth defects. 2010.

3. Zhou WH, Zhao ZY. [Genomic newborn screening: opportunities and challenges]. *Zhonghua er ke za zhi = Chinese journal of pediatrics.* 2021;59(7):541-544.

4. Fabie NAV, Pappas KB, Feldman GL. The Current State of Newborn Screening in the United States. *Pediatric clinics of North America.* 2019;66(2):369-386.

5. Guthrie R, Susi A. A SIMPLE PHENYLALANINE METHOD FOR DETECTING PHENYLKETONURIA IN LARGE POPULATIONS OF NEWBORN INFANTS.*Pediatrics.* 1963;32:338-343.

6. Wilson JM, Jungner YG. [Principles and practice of mass screening for disease]. *Boletin de la Oficina Sanitaria Panamericana Pan American Sanitary Bureau.* 1968;65(4):281-393.

7. Rajabi F. Updates in Newborn Screening. 2018;47(5):e187-e190.

8. Newborn screening: toward a uniform screening panel and system–executive summary. *Pediatrics.* 2006;117(5 Pt 2):S296-307.

9. Zhan JY, Qin YF, Zhao ZY. Neonatal screening for congenital hypothyroidism and phenylketonuria in China. *World journal of pediatrics : WJP.* 2009;5(2):136-139.

10. Ombrone D, Giocaliere E, Forni G, Malvagia S, la Marca G. Expanded newborn screening by mass spectrometry: New tests, future perspectives.*Mass spectrometry reviews.* 2016;35(1):71-84.

11. Mendell JR, Shilling C, Leslie ND, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Annals of neurology.* 2012;71(3):304-313.

12. Slaughter JL, Meinzen-Derr J, Rose SR, et al. The effects of gestational age and birth weight on false-positive newborn-screening rates. *Pediatrics.* 2010;126(5):910-916.

13. Rizvi H, Sanchez-Vega F, La K, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2018;36(7):633-641.

14. Han D, Li Z, Li R, Tan P, Zhang R, Li J. mNGS in clinical microbiology laboratories: on the road to maturity. *Critical reviews in microbiology.* 2019;45(5-6):668-685.

15. Swedish Council on Health Technology A. SBU Systematic Review Summaries. In: *Prenatal Diagnosis Through Next Generation Sequencing (NGS).* Stockholm: Swedish Council on Health Technology Assessment (SBU)Copyright (c) 2016 by the Swedish Council on Health Technology Assessment.; 2016.

16. Willig LK, Petrikin JE, Smith LD, et al. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *The Lancet Respiratory medicine.* 2015;3(5):377-387.

17. Holm IA, Agrawal PB, Ceyhan-Birsoy O, et al. The BabySeq project: implementing genomic sequencing in newborns. *BMC pediatrics.*2018;18(1):225.

18. Roman TS, Crowley SB, Roche MI, et al. Genomic Sequencing for Newborn Screening: Results of the NC NEXUS Project. *American journal of human genetics.* 2020;107(4):596-611.

19. Lin Y, Zhang W, Huang C, et al. Increased detection of primary carnitine deficiency through second-tier newborn genetic screening.*Orphanet journal of rare diseases.* 2021;16(1):149.

20. Parad RB, Kaler SG, Mauceli E, Sokolsky T, Yi L, Bhattacharjee A. Targeted next generation sequencing for newborn screening of Menkes disease. *Molecular genetics and metabolism reports.*2020;24:100625.

21. Meng L, Pammi M, Saronwala A, et al. Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. *JAMA pediatrics.*2017;171(12):e173438.

22. Zhu T, Gong X, Bei F, et al. Application of Next-Generation Sequencing for Genetic Diagnosis in Neonatal Intensive Care Units: Results of a Multicenter Study in China. *Frontiers in genetics.*2020;11:565078.

23. Ceyhan-Birsoy O, Machini K, Lebo MS, et al. A curated gene list for reporting results of newborn genomic sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2017;19(7):809-818.

24. Milko LV, O'Daniel JM, DeCristo DM, et al. An Age-Based Framework for Evaluating Genome-Scale Sequencing Results in Newborn Screening.*The Journal of pediatrics.* 2019;209:68-76.

25. Yu X, Lin Y, Wu H. Targeted Next-Generation Sequencing Identifies Separate Causes of Hearing Loss in One Deaf Family and Variable Clinical Manifestations for the p.R161C Mutation in SOX10. *Neural plasticity.* 2020;2020:8860837.

26. Oliveros JC-VAitfclwVsdhbccetvih.

27. Adhikari AN, Gallagher RC, Wang Y, et al. The role of exome sequencing in newborn screening for inborn errors of metabolism.*Nature medicine.* 2020;26(9):1392-1397.

28. Murry JB, Machini K, Ceyhan-Birsoy O, et al. Reconciling newborn screening and a novel splice variant in BTD associated with partial biotinidase deficiency: a BabySeq Project case report. *Cold Spring Harbor molecular case studies.* 2018;4(4).

29. Bodian DL, Klein E, Iyer RK, et al. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2016;18(3):221-230.

30. 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. *Genetics in medicine : official journal of the American College of Medical Genetics.*2016;18(11):1090-1096.

31. Ceyhan-Birsoy O, Murry JB, Machini K, et al. Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project. *American journal of human genetics.*2019;104(1):76-93.

32. Petrikin JE, Cakici JA, Clark MM, et al. The NSIGHT1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *NPJ genomic medicine.* 2018;3:6.

33. van Campen JC, Sollars ESA, Thomas RC, et al. Next Generation Sequencing in Newborn Screening in the United Kingdom National Health Service. *International journal of neonatal screening.*2019;5(4):40.

34. Hao C, Guo R, Hu X, et al. Newborn screening with targeted sequencing: A multi-center investigation and a pilot clinical study in China. *Journal of genetics and genomics = Yi chuan xue bao.* 2021.

35. de Castro MJ, Gonzalez-Vioque E, Barbosa-Gouveia S, et al. Rapid Phenotype-Driven Gene Sequencing with the NeoSeq Panel: A Diagnostic Tool for Critically Ill Newborns with Suspected Genetic Disease.*Journal of clinical medicine.* 2020;9(8).

36. Luo X, Sun Y, Xu F, et al. A pilot study of expanded newborn screening for 573 genes related to severe inherited disorders in China: results from 1,127 newborns. *Annals of translational medicine.*2020;8(17):1058.

37. Gregg AR, Aarabi M, Klugman S, et al. Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genetics in medicine : official journal of the American College of Medical Genetics.* 2021;23(10):1793-1806.

38. Breveglieri G, D'Aversa E, Finotti A, Borgatti M. Non-invasive Prenatal Testing Using Fetal DNA. *Molecular diagnosis & therapy.*2019;23(2):291-299.

39. Huang X, Wu D, Zhu L, et al. Application of a next-generation sequencing (NGS) panel in newborn screening efficiently identifies inborn disorders of neonates. *Orphanet journal of rare diseases.*2022;17(1):66.

40. Wang X, Wang YY, Hong DY, et al. Combined genetic screening and traditional biochemical screening to optimize newborn screening systems.*Clinica chimica acta; international journal of clinical chemistry.* 2022;528:44-51.

41. Tong F, Wang J, Xiao R, et al. Application of next generation sequencing in the screening of monogenic diseases in China, 2021: a consensus among Chinese newborn screening experts. *World journal of pediatrics*

: WJP. 2022;18(4):235-242.

42. Zhang J, Walsh MF, Wu G, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. *The New England journal of medicine.*2015;373(24):2336-2346.

43. Qian Q, Cheng R, Li M, et al. Prevalence of atopic dermatitis, ichthyosis and filaggrin mutations in adolescents in a middle school in Shanghai. 2015;48:629-632.

44. Palmer CN, Irvine AD, Terron-Kwiatkowski A, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nature genetics.* 2006;38(4):441-446.

45. Koohiyan M, Koohian F, Azadegan-Dehkordi F. GJB2-related hearing loss in central Iran: Review of the spectrum and frequency of gene mutations. *Annals of human genetics.* 2020;84(2):107-113.

46. Shen J, Oza AM, Del Castillo I, et al. Consensus interpretation of the p.Met34Thr and p.Val37Ile variants in GJB2 by the ClinGen Hearing Loss Expert Panel. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2019;21(11):2442-2452.

47. Genomic Screening in Newborns Holds Promise, Challenges: Results of the BabySeq Project demonstrate the potential of newborn genomic screening, but challenges remain. *American journal of medical genetics Part A.* 2019;179(4):519-520.

48. Berg JS, Agrawal PB, Bailey DB, Jr., et al. Newborn Sequencing in Genomic Medicine and Public Health. *Pediatrics.* 2017;139(2).

49. Murdock DR, Dai H, Burrage LC, et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *The Journal of clinical investigation.* 2021;131(1).

50. Holm IA, McGuire A, Pereira S, Rehm H, Green RC, Beggs AH. Returning a Genomic Result for an Adult-Onset Condition to the Parents of a Newborn: Insights From the BabySeq Project. *Pediatrics.*2019;143(Suppl 1):S37-s43.

51. Ross LF, Clayton EW. Ethical Issues in Newborn Sequencing Research: The Case Study of BabySeq. *Pediatrics.* 2019;144(6).

52. Wojcik MH, Zhang T, Ceyhan-Birsoy O, et al. Discordant results between conventional newborn screening and genomic sequencing in the BabySeq Project. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2021;23(7):1372-1375.

53. Seaby EG, Ennis S. Challenges in the diagnosis and discovery of rare genetic disorders using contemporary sequencing technologies.*Briefings in functional genomics.* 2020;19(4):243-258.

Figure Legends

Figure 1. Characteristics of 601 gene-diseases associations. A) 7 categorizations of disease were collected and included to generate the gene panel. The disease system (B) and onset age (C) for those gene-disease associations were demonstrated. <1, <1 year of age (Infants); 1-3, 1-3 years of age (Toddlers); 3-6, 3-6 years of age (Preschoolers); 6-12, 6-12 years of age (Middle Childhood); 12-18, 12-18 years of age (Young Teens and Teenagers); >18, >18 years of age (Adulthood); UN, unknown.

Figure 2. The overall flowchart of NeoEXOME study.3423 neonates meet the inclusion criteria were enrolled from 5 institutions, 174 were excluded. CQ, Children's Hospital of Chongqing Medical University; XZ, Xuzhou Maternal and child Health Care Hospital; NM, Inner Mongolia Maternal and Child Health Hospital; XB, Northwest Women and Children's Medical Center; DL, Dalian Maternal and Child Health Hospital.

Figure 3. NeoEXOME detection results of positive cases. A) Gene variants interpretation (risk grade and inheritance patterns) of neonates with mono-gene change. B) Gene variants interpretation (risk

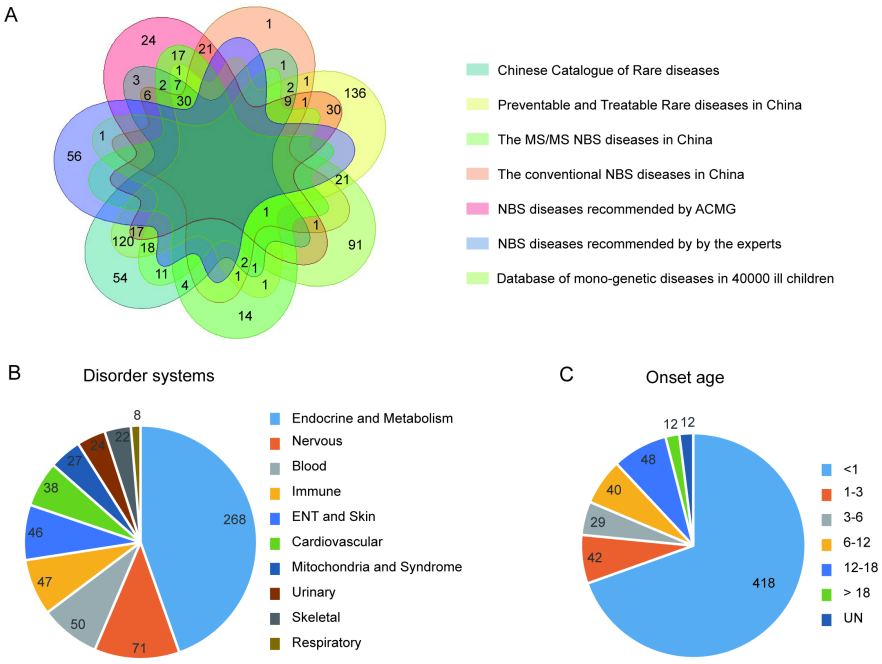
grade and genetic patterns) of neonates with multi-genes change. C) Disease system distribution of neonates that were NGS positive. D) Gene-disease associations' onset age of neonates that were NGS positive. E) Top 20 genes distribution of neonates with NGS positive cases. LR, Low risk; MR, Moderate risk; HR, High risk; AD, Autosomal dominant; AR, Autosomal recessive; XLD, X-linked dominant; Mito, Mitochondrial; Syns, Syndrome; ENT, ear, nose, and throat; Endo, Endocrine; Meta, metabolism.

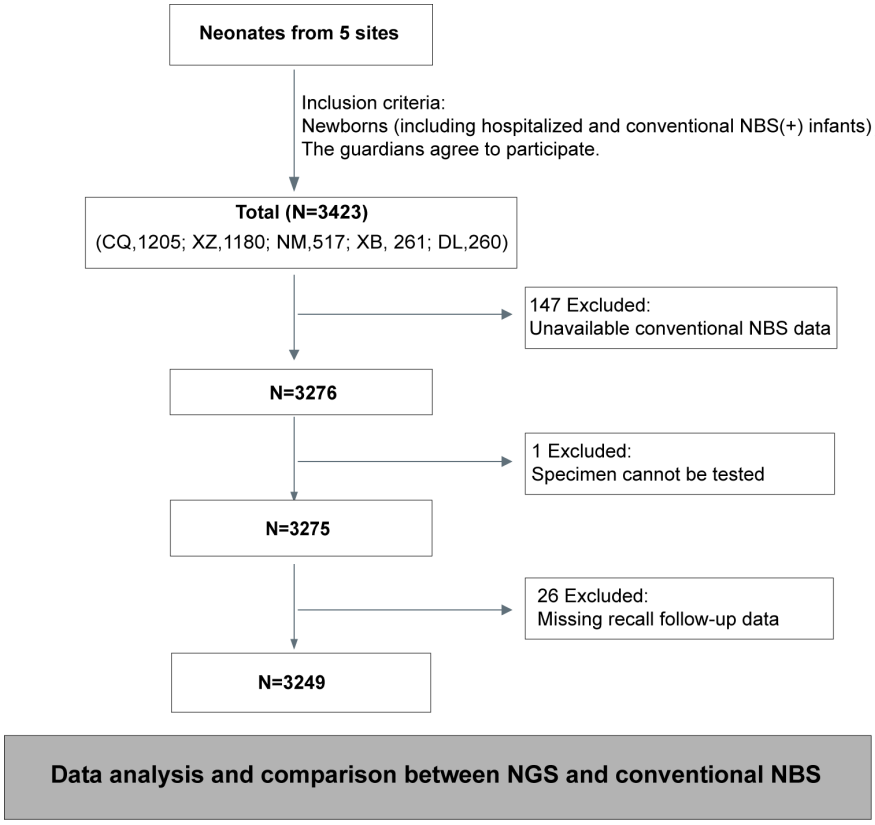
Figure 4. NGS detection results of carrier. A) Gene variants characteristics of neonates that were carrier. B) Top 20 genes distribution of neonates that were carrier.

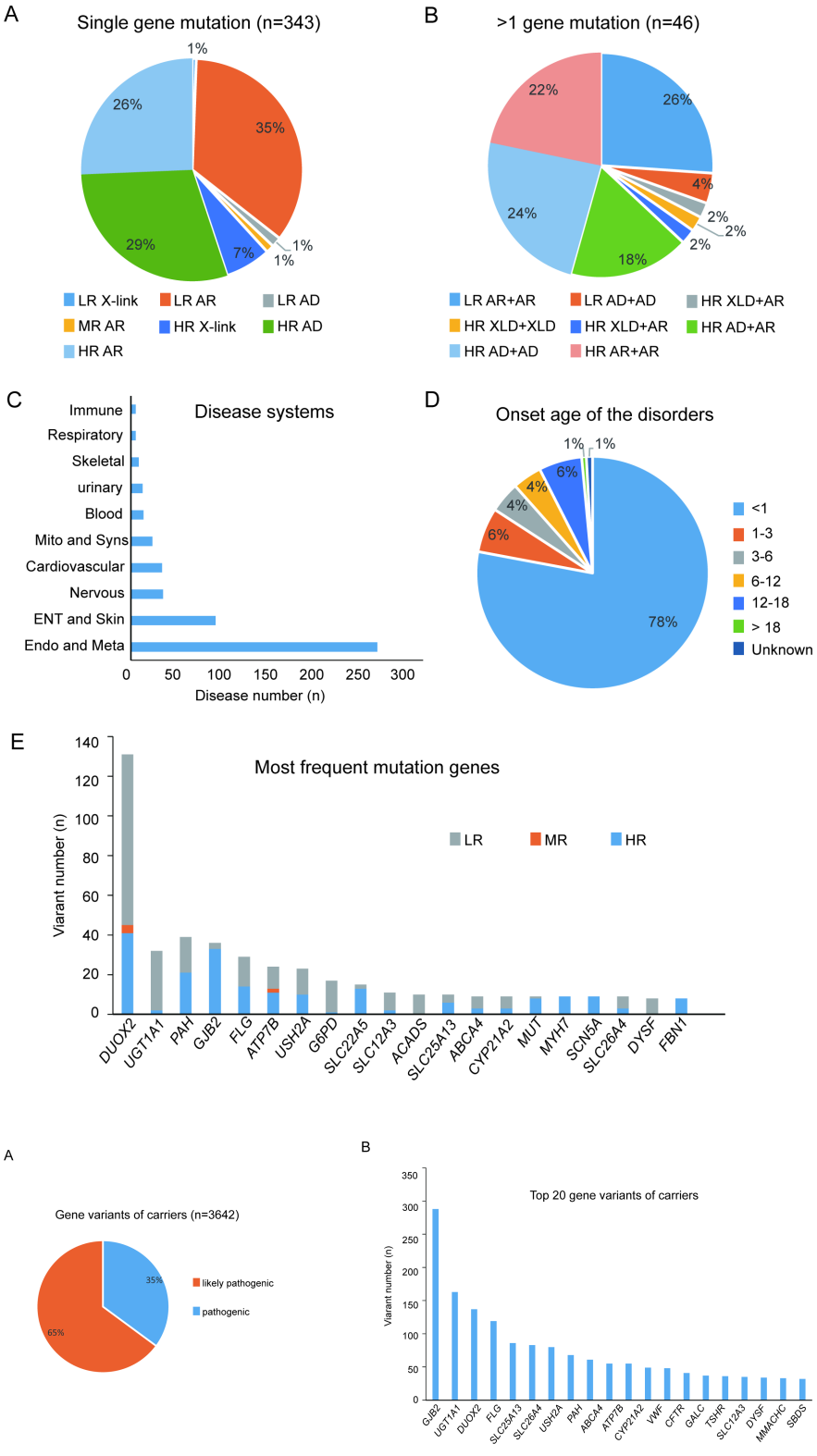
Figure 5. Comparison of NeoEXOME with other NBS panel. A) NeoEXOME and Babyseq. B) NeoEXOME and NC NEXUS. C) NeoEXOME and NESTS.

Supplementary Figure 1. The work flow of our study.

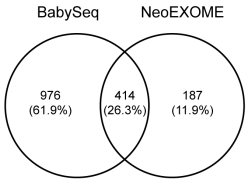
Supplementary Figure 2. The gene variants interpretation of NeoEXOME.





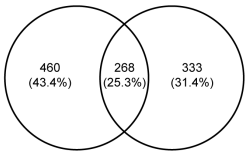


A



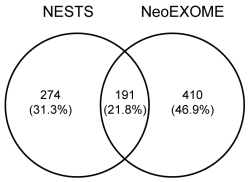
	BabySeq	NeoEXOME	Common genes	Only in BabySeq	Only in NeoEXOME
Gene-Number	1390	601	414	976	187
Disease systems, n (%)					
Endocrine and Metabolism	359 (25.83)	268(44.59)	204	155	64
Nervous	291 (20.94)	71(11.81)	42	249	29
Blood	62 (4.46)	50(8.32)	33	20	17
Immune	38 (2.73)	47(7.82)	24	14	23
ENT and Skin	167 (12.01)	46(7.65)	33	134	13
Cardiovascular	111 (7.99)	38(6.32)	27	84	11
Mitochondria and Syndrome	164 (11.80)	27(4.49)	13	151	14
Urinary	47 (3.38)	24(3.99)	19	28	5
Skeletal	65 (4.68)	22(3.66)	12	53	10
Respiratory	40 (2.88)	8(1.33)	7	33	1
Digestive	23 (1.65)	0	0	23	0
Tumor and others	23 (1.65)	0	0	23	0
Onset age, n (%)					
<18 years	1309(94.17)	577(96.00)	407	902	170
>18 years	81(5.83)	24(4.00)	7	74	17
BabySeq Category					
A	818		310		
B	55		29		
C	517		75		

B



	NC NEXUS	NeoEXOME	Common genes	Only in NC NEXUS	Only in NeoEXOME
Gene-Number	728	601	268	460	333
Disease systems, n (%)					
Endocrine and Metabolism	197(27.06)	268(44.59)	159	38	109
Nervous	34(4.67)	71(11.81)	17	17	54
Blood	79(10.85)	50(8.32)	19	60	31
Immune	33(4.53)	47(7.82)	16	17	31
ENT and Skin	121(16.62)	46(7.65)	15	106	31
Cardiovascular	73(10.03)	38(6.32)	20	53	18
Mitochondria and Syndrome	103(14.15)	27(4.49)	9	94	18
Urinary	11(1.51)	24(3.99)	3	8	21
Skeletal	12(1.65)	22(3.66)	4	8	18
Respiratory	54(7.42)	8(1.33)	6	48	2
Digestive	9(1.24)	0	0	9	0
Tumor and others	2(0.27)	0	0	2	0
Onset age, n (%)					
Birth	384(52.75)	418(69.55)	191	193	227
Childhood	114(15.66)	111(18.47)	53	61	58
Adolescent	20(2.75)	48(7.99)	15	5	33
Adulthood	81(11.13)	12(2.00)	9	72	3
Variable	107(14.70)	0	0	107	0
UN	22(3.02)	12(2.00)	0	22	12
NC NEXUS Category					
1	420		186		
2	214		73		
3	18		2		
4	17		4		
NR	59		5		

C



	NESTS	NeoEXOME	Common genes	Only in NESTS	Only in NeoEXOME
Gene-Number	465	601	191	274	410
Disease systems, n (%)					
Endocrine and Metabolism	120(25.81)	268(44.59)	94	26	174
Nervous	20(4.30)	71(11.81)	10	10	61
Blood	22(4.73)	50(8.32)	11	11	39
Immune	25(5.38)	47(7.82)	11	14	36
ENT and Skin	138(29.68)	46(7.65)	19	119	27
Cardiovascular	18(3.87)	38(6.32)	13	5	25
Mitochondria and Syndrome	42(9.03)	27(4.49)	10	32	17
Urinary	33(7.10)	24(3.99)	11	22	13
Skeletal	16(3.44)	22(3.66)	10	6	12
Respiratory	29(6.24)	8(1.33)	2	27	6
Digestive	2(0.43)	0	0	2	0
Onset age, n (%)					
<1	335(72.04)	420(69.88)	153	182	267
1-3	39(8.39)	43(7.15)	11	28	32
3-12	73(15.70)	74(12.31)	24	49	50
12-18	18(3.87)	49(8.15)	3	15	34
≥18	0	15(2.50)	0	0	15

Hosted file

table1 WGS WES.docx available at https://authorea.com/users/479913/articles/567572-the-targeted-exome-sequencing-strategy-neoexome-for-chinese-newborns-with-the-pilot-study-of-3423-neonates

Hosted file

table2 NBS results comparsion all newborn.docx available at https://authorea.com/users/479913/articles/567572-the-targeted-exome-sequencing-strategy-neoexome-for-chinese-newborns-with-the-pilot-study-of-3423-neonates