

Newborn Screening in Unselected Children Using Genomic Sequencing

Min Jian

BGI-Shenzhen <https://orcid.org/0000-0003-4985-5520>

Xiaohong Wang

BGI-Shenzhen

Yuanyuan Sui

Maternal and child health and family planning service center of Huangdao District, Qingdao

Mingyan Fang

BGI-Shenzhen

Ya Gao

BGI-Shenzhen

Ruidong Guo

BGI-Shenzhen

Yingping Huang

BGI Qingdao

Chunhua Liu

Maternal and child health and family planning service center of Huangdao District

Chenchen Feng

BGI-Shenzhen

Yuanning Guan

BGI-Shenzhen

Yuxiao Gao

Maternal and child health and family planning center of Huangdao District

Zhiwei Wang

BGI qingdao

Shuli Li

Traditional Chinese medical hospital of Huangdao District

Bochen Cheng

BGI Shenzhen

Lina Sun

BGI Qingdao

Fenghua Cui

The Affiliated Hospital of Qingdao University

Jia Guo

BGI Shenzhen

Ying Zhan

Qingdao west coast new area central hospital, Qingdao

Guohong Zhang

BGI SHENZHEN

Ling Zheng

Traditional Chinese medical hospital of Huangdao District, Qingdao

Fengxia Su

BGI shenzhen

Wei Xue

The Affiliated Hospital of Qingdao University

Puyi Qian

China National GeneBank, BGI Shenzhen,

Shaobo Gao

Qingdao west coast new area central hospital, Qingdao

Jiayu Chen

China National GeneBank, BGI Shenzhen,

Qian Zhao

China National GeneBank, BGI Shenzhen,

Lingyao Guan

China National GeneBank, BGI Shenzhen,

Haorong Lu

Guangdong Provincial Key Laboratory of Genome Read and Write, BGI Shenzhen

Chao Nie

BGI-shenzhen

Karsten Kristiansen

University of Copenhagen

Lennart Hammarström

BGI Shenzhen

Xiaojing Jiang

Maternal and child health and family planning service center of Huangdao District, Qingdao,

Junnian Liu

BGI Qingdao

Fang Chen (✉ fangchen@genomics.cn)

BGI-Shenzhen <https://orcid.org/0000-0002-7095-0541>

Research article

Keywords: Newborn screening, Mendelian Diseases (MD), Primary Immunodeficiency (PID), pharmacogenetics (PGx), genomic sequencing(GS)

Posted Date: January 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-143405/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The aim of this study is to investigate potentially curable or treatable medical conditions in unselected newborns using genomic sequencing(GS).

Methods: 321 newborns from a cohort of pregnant women from Qingdao, China, underwent high-depth GS (average 47.42 fold), with the approval of the ethics committee. 61 Mendelian Diseases, 151 Primary Immunodeficiency Diseases(PID) and 5 DPWG recommended Essential pharmacogenetic(PGx) genes were analyzed.

Results: 121 Mendelian pathogenic or likely pathogenic variants associated with 31 inherited diseases were detected, among these hearing loss, congenital hypothyroidism, methylmalonic acidemia, methylmalonic acidemia with homocystinuria, phenylketonuria(PKU) and benign hyperphenylalaninemia accounted for half of the carrier variants. Three children with compound heterozygous variants at *GJB2* and *PAH* were confirmed by Sanger sequencing. Follow-up of the three families confirmed that one child was diagnosed with PKU and two children with *GJB2* variants were scheduled to undergo hearing loss testing every six months after genetic counseling due to the nature of incomplete penetrance of hearing loss. 11 heterozygous pathogenic/ likely pathogenic variants in eight PID genes were identified in 11 infants. All 321 newborns carried at least one variant at the five DPWG recommended PGx genes. Codeine and clopidogrel require more attention in giving prescription for 25% and 8% of newborns have a decreased function of CYP2D6 and CYP2C19 enzymes respectively.

Conclusions: Our study is the largest to date using GS to sequence unselected newborns. The results suggest that using GS may be a suitable method for screening newborns for variants in a large number of disease associated genes.

Background

Newborn screening for metabolic diseases was initiated in the 1960-ties and was based on the early identification of children with phenylketonurea (Føllings disease) as a cause of childhood mental retardation (1). The development of an assay based on examination of a dried blood spot (DBS) on a filter paper in 1963 (Guthrie card) (2) formed the basis of a technically simple screening assay, allowing the identification and subsequent dietary treatment of affected children.

Extended screening for a large number of additional metabolic diseases, aided by the introduction of mass spectrometry methods, has been gradually implemented, albeit to a different extent in various countries (<https://membership.isns-neoscreening.org/disorders/>). The ethical guidelines published by Wilson and Jungner in 1968 (3), remains the gold standard for which diseases to screen for and the current list of recommended disorders in the US contains 35 core and 26 secondary conditions (RUSP/HHS) (<https://www.hrsa.gov/advisory-committees/heritable-disorders/rusp/index.html>) (3).

DNA based screening using the DBS has recently been introduced, and screening for cystic fibrosis is included in screening programs in selected countries. Severe Combined Immunodeficiency (SCID) (T cell lymphopenia) using quantification of T cell receptor excision circles (TREC) (4) has also been implemented, starting in the USA in 2010 (5). Screening for kappa receptor excision circles (KREC) (B cell lymphopenia in X-linked agammaglobulinemia (XLA)) followed in 2011 (6), and a combined assay is used in selected countries. Currently, testing for other diseases such as Spinal Muscular Atrophy (SMA) is being implemented and additional conditions are being considered for national genetic screening of newborns.

The NSIGHT initiative (NIH) 2012–2018 (<https://www.genome.gov/Funded-Programs-Projects/Newborn-Sequencing-in-Genomic-Medicine-and-Public-Health-NSIGHT>), aimed at large scale Newborn Sequencing in Genomic Medicine and Public Health using exome sequencing (ES), or genomic sequencing (GS). The supported projects were related both to data collection, identification of specific disorders and ethical, legal and social implications of genomic sequencing of newborns.

Different methods for identification of mutated genes, including Targeted Region Sequencing (TRS), ES and GS, is currently standard procedure in newborn children with a suspected disease. However, sequencing of apparently healthy newborns remains controversial as it may violate the Wilson and Jungner criteria. Yet, ES and GS has already been performed on small cohorts of "healthy" newborn children (7, 8). The latter study identified actionable adult onset disease in 3.5% and actionable childhood onset disease in 9.4% of the tested children. Furthermore, relevant pharmacogenomic variants were identified in 5% of the newborns, supporting the concept of introducing ES/GS based screening of all newborns.

The current study is the largest study to date using GS on unselected newborns, looking at a wide panel of inherited diseases, primary immunodeficiency diseases and pharmacogenomically relevant variants.

Methods

Study Subjects

Pregnant women were enrolled from May 2018 to December 2018 in this study where 321 consecutive newborns (47% (151) males and 53% (170) females) from these families underwent genomic sequencing (GS). Pregnant women were recruited at the Maternal and Child Health and Family Planning Service Center of Huangdao District, Qingdao, China to join in a prospective cohort study, and then umbilical cord blood and umbilical cord of each pregnancy was collected at three collaborated hospitals.

The study was approved by the medical ethics committee (Maternal and child health and family planning service center of Huangdao district Ethics Reviews board, ethical permit ID: 20180001) and the Institutional Review Board of BGI (No. BGHRB18029 and No. BGIHRB17073). Each family signed an informed consent form before participating the study.

Each participant was provided with a report on the results of the genetic testing, a report interpretation and genetic counselling. Positive genetic test results will be followed up annually until the child is three years old.

Sample collection

Umbilical cord blood (5 ml) and umbilical cord (3 tubes, 1 cm per tube) were collected. When GS detection was carried out, umbilical blood DNA was preferred for detection. If the umbilical blood collection failed, umbilical cord DNA was extracted. The GS cohort contained DNA from 303 umbilical cord blood samples and 18 umbilical cords.

Processing of samples

Umbilical blood DNA was extracted with the HiPure Blood DNA Mini Kit (Magen, Guangzhou, China) whereas umbilical cord DNA was extracted with Salting-out Self-dispensing Kit. After DNA extraction, Qubit 3.0 fluorometer (Life Technologies, Paisley, UK) was used to measure the DNA concentration, and an 2% agarose gel electrophoresis was used to detect DNA fragment integrity.

Sequencing

Extracted DNA subsequently underwent library construction and sequenced using the sequencing platform DIPSEQ of MGI (MGI, Shenzhen, China) with 100-bp paired-end reads. Briefly, genomic DNA was normalized and processed for circularization (9). Genomic DNA was heat-denatured at 95 °C for 3 minutes to make a single strand DNA circle (ssDNA circle), which were then mixed reagents of MGIEasy™ DNA Library Prep Kit (MGI, Shenzhen, China) and incubated at 37 °C for 30 minutes to complete the circularization. The resulting ssDNA circle was then used to generate DNA nanoballs (DNBs) by rolling circle amplification (RCA) (10). After RCA and the formation of DNBs, the final product was measured by Qubit using the ssDNA HS Assay kit (Invitrogen), and loaded on a DNBSEQ-500 platform (MGI, Shenzhen, China) for sequencing (11) following the manufacturer's instructions.

Analysis pipeline

The alternative contigs in GRCh38 assembly was deleted to improve the alignment accuracy, and BWA mem was used to align the read to the human reference genome (GRCh38 / UCSC hg38). The Genome Analysis Software Kit (GATK 4.0) best practice pipeline was used to perform variants calling including SNVs (Single nucleotide variants) and short InDels (insertions/deletions). After variation calling, bcftools was used to extract variation from 61 Mendelian Diseases (MD), 151 Primary Immunodeficiency (PID) genes and 5 pharmacogenetic genes associated with severe adverse drug reactions. Subsequently all the samples were merged using bcftools merge. All the missing alleles were assumed to have no variant.

Polymorphic sites were then annotated by BGI in-house databases and public databases, such as the 1000 Genomes Project (KG, <http://www.1000genomes.org/>), the Exome Variant Server (ESP, <http://evs.gs.washington.edu/EVS/>), the Exome Aggregation Consortium (ExAC, <http://exac.broadinstitute.org>) and the Genome Aggregation Database (gnomAD, <https://gnomad.broadinstitute.org>). The data that support the findings of this study have been deposited into CNGB Sequence Archive(12) of CNGBdb with accession number CNP0001264(13).

Quality control (QC)

In order to ensure the high-quality of the data for each sample, stringent quality control criteria were applied, which required the GC content of the sequencing read fell within 40%-44%, the average Q30 above 80%, the duplicated rate below 10%, the average depth above 20x, the percentage $\geq 4x$ coverage of NoN-regions above 96%, the Ti/Tv ratio within 1.96–2.02 and the het/hom ratio within 1.3–1.7.

Gene lists

The 61 Mendelian Diseases related 109 RUSP annotated genes are shown in Supplementary Table 1. 151 genes suggested by the International Union of Immunologic Societies Expert Committee for Primary Immunodeficiency (14, 15), which are associated with the most severe conditions (including immunodeficiencies affecting cellular and humoral immunity, combined immunodeficiencies with associated or syndromic features and predominantly antibody deficiencies), were manually reviewed and used to explore the known PID genes and are shown in Supplementary Table 2, and the five genes suggested by the Dutch Pharmacogenetics Working Group to be associated with adverse drug reactions are given in Supplementary Table 3. Each gene and its inheritance manually was reviewed according to OMIM or the published literature.

Validation of variants

Pathogenic or likely pathogenic variants of inherited metabolic diseases were verified by Sanger sequencing.

Results

Quality control

The demographic data of 321 newborns are summarized in Table 1. The average sequencing depth of the 321 samples was 47.42 (28.84 ~ 82.90), and the average coverage was 99.48 (99.01 ~ 99.89) (Supplementary Fig. 1).

Table 1
Summary of demographic data collected from 321 Qingdao cohort

	Type	Number	Percentage
Pregnancy	Natural pregnancy	306	95.4%
	Assisted reproduction	11	3.4%
	Unknown	4	1.2%
Gestational weeks	Premature birth	7	2.2%
	Delivery pregnancy week	39 weeks plus 5 days	
	SD	9.47	
Gender of newborns	Boys	151	47%
	Girls	170	53%
Parental age at delivery	Father's age (ave.)	33	
	Mother's age (ave.)	32	
	SD of father's age	5	
	SD of mother's age	4	

Inherited diseases

We chose to analyse only the 109 genes underlying the 61 diseases listed in the recommended Uniform Screening Panel (RUSP). 121 pathogenic or likely pathogenic variants were detected in 321 children (30.53% of the samples), mainly in a heterozygous state (Table 2). Thirtyone inherited diseases were associated with the 121 pathogenic and likely pathogenic variants, while Hearing Loss was the most common disease. Twentyone newborns carried more than two genetic variants (Supplementary Table 4).

Table 2
Overview of the heritable diseases identified in 321 newborn children from Qingdao

No.	Core Condition	Inheritance	Gene	Variants	Class
1	Propionic Acidemia	AR	<i>PCCB</i>	c.1364A > G	P
				c.793G > A	LP
2	Methylmalonic Acidemia (methylmalonic-CoA mutase)	AR	<i>MUT</i>	c.2179C > T	P
3	Methylmalonic Acidemia (Cobalamin disorders)	AR	<i>MMAA</i>	c.658G > A	LP
				<i>MMACHC</i>	c.315C > G*
			c.445_446del*		P
			c.482G > A*		P
			c.609G > A*		P
			c.658_660del*		P
			<i>MMADHC</i>	c.80A > G*	P
	c.748C > T*	P			
4	3-Methylcrotonyl-CoA Carboxylase Deficiency	AR	<i>MCCC1</i>	c.639 + 2T > A	P
5	Holocarboxylase synthase Deficiency	AR	<i>HLCS</i>	c.782del	P
6	Glutaric Acidemia Type I	AR	<i>GCDH</i>	c.1213A > G	P
7	Carnitine Uptake Defect/Carnitine Transport Defect	AR	<i>SLC22A5</i>	c.1472C > G	P
				c.468G > A	P
8	Medium-chain Acyl-CoA Dehydrogenase Deficiency	AR	<i>ACADM</i>	c.548_551del	P
9	Trifunctional Protein Deficiency	AR	<i>HADHB</i>	c.1175C > T	LP
10	Citrullinemia, Type I	AR	<i>ASS1</i>	c.352G > A	LP
11	Homocystinuria	AR	<i>MMADHC</i>	c.748C > T*	P
12	Classic Phenylketonuria	AR	<i>PAH</i>	c.1301C > A*	LP
				c.611A > G*	P
				c.728G > A*	P
				c.740G > T*	P
				c.842 + 2T > A*	P
13	Primary Congenital Hypothyroidism	AR,AD	<i>TSHR</i>	c.1349G > A	P
		AR	<i>DUOX2</i>	c.1588A > T	P
				c.1883del	P
				c.1946C > A	LP
				c.3329G > A	P
		AR	<i>TPO</i>	c.2422del	P
14	Congenital adrenal hyperplasia	AR	<i>CYP21A2</i>	c.518T > A	P
				c.844G > T	P
				c.92C > T	P
15	S, beta-Thalassemia	AR,AD	<i>HBB</i>	c.126_129del*	P

Notes: Recommended Uniform Screening panel as of July 2018 (<https://www.hrsa.gov/advisory-committees/heritable-disorders/rusp/index.html>). The class the ACMG guidelines, referring to Clinvar annotation and literature reported. Variants with asterisk (*) were associated with more than one disease.

No.	Core Condition	Inheritance	Gene	Variants	Class
16	Cystic Fibrosis	AR	<i>CFTR</i>	c.2052_2053insA	P
17	Classic Galactosemia	AR	<i>GALT</i>	c.821-7A > G	P
		AR	<i>GALT</i>	c.844C > G	LP
18	Glycogen Storage Disease Type II (Pompe)	AR	<i>GAA</i>	c.2237G > C	P
				c.2662G > T	P
				c.2647-7G > A	LP
19	Hearing Loss	AD,AR,DD(Digenic dominant)	<i>GJB2</i>	c.109G > A	P
				c.235del	P
				c.299_300del	P
				c.605_606insAGAAGACTGTCTTCACAGTGTTTCATGATTGCAGTGTCTGGAATTTG	P
		AR	<i>SLC26A4</i>	c.1174A > T	P
		AR		c.1229C > T	P
		AR		c.1262A > C	LP
		AR		c.2027T > A	LP
		AR		c.2168A > G	P
		AR		c.919-2A > G	P
		AR		<i>USH2A</i>	c.2802T > G
	c.99_100insT	P			
	c.8559-2A > G	P			
20	Severe Combined Immunodeficiencies	AR	<i>ADA</i>	c.424C > T	P
				c.872C > T	P
		AR	<i>JAK3</i>	c.1744C > T	LP
				c.307C > T	LP
21	Spinal Muscular Atrophy due to homozygous deletion of exon 7 in SMN1	AR	<i>SMN1</i>	-	P
No.	Secondary Condition	Inheritance	Gene	Variants	Class
22	Methylmalonic acidemia with homocystinuria	AR	<i>MMACHC</i>	c.315C > G*	P
				c.445_446del*	P
				c.482G > A*	P
				c.609G > A*	P
				c.658_660del*	P
		c.80A > G*	P		
AR	<i>MMADHC</i>	c.748C > T*	P		
23	Short-chain acyl-CoA dehydrogenase deficiency	AR	<i>ACADS</i>	c.1031A > G	P
24	Glutaric acidemia type II	AR	<i>ETFDH</i>	c.1211T > C	LP
25	Citrullinemia, type II	AR	<i>SLC25A13</i>	c.1180 + 1G > A	P
				c.852_855del	P
26	Hypermethioninemia	AR	<i>GNMT</i>	c.149T > C	LP

Notes: Recommended Uniform Screening panel as of July 2018 (<https://www.hrsa.gov/advisory-committees/heritable-disorders/rusp/index.html>). The class the ACMG guidelines, referring to Clinvar annotation and literature reported. Variants with asterisk (*) were associated with more than one disease.

No.	Core Condition	Inheritance	Gene	Variants	Class
27	Benign hyperphenylalaninemia	AR	PAH	c.1301C > A*	LP
				c.611A > G*	P
				c.728G > A*	P
				c.740G > T*	P
				c.842 + 2T > A*	P
28	Biopterin defect in cofactor biosynthesis	AR	PTS	c.166G > A	P
				c.259C > T	P
				c.84-291A > G	P
29	Biopterin defect in cofactor regeneration	AR	PTS	c.166G > A	P
				c.259C > T	P
				c.84-291A > G	P
30	Various other hemoglobinopathies	AR	HBB	c.126_129del*	P
31	Galactose epimerase deficiency	AR	GALE	c.505C > T	P

Notes: Recommended Uniform Screening panel as of July 2018 (<https://www.hrsa.gov/advisory-committees/heritable-disorders/rusp/index.html>). The class the ACMG guidelines, referring to Clinvar annotation and literature reported. Variants with asterisk () were associated with more than one disease.*

Three children with compound heterozygous variants were detected, predicted computationally to be pathogenic or likely pathogenic in the 321 subjects. Two children had variants at the *GJB2* gene (NM_004004.5, c.109G > A, p.V37I; NM_004004.5, c.235del, p.L79Cfs3 and NM_004004.5, c.109G > A, p.V37I; NM_004004.5, c.299_300del, p.H100Rfs14), and one carried two variants at the *PAH* gene (NM_000277.1, c.611A > G, p.Y204C and NM_000277.1, c.842 + 2T > A). Sanger sequencing confirmed that one variant was inherited from her/his mother. However, as infant father's sample was not available, we could not determine if the small deletion and insertion was inherited from the father or whether it was a *de novo* variant (Supplementary Fig. 1). Follow-up of the three families confirmed that one child with compound heterozygosity in *PAH* has been diagnosed with PKU, while the other two children with *GJB2* variation have not shown characteristics of hearing loss yet. Previous studies report that homozygous or compound heterozygous variants of c.109G > A are associated with light to mild deafness, and show incomplete penetrance, which can lead to late-onset deafness (16, 17). After genetic counseling, the two children with *GJB2* variants were therefore scheduled to undergo hearing testing every six months.

Primary immunodeficiency diseases

The IUIS summary information of PID genes (15) was used to identify potential variants in 151 immunodeficiency associated genes. Altogether 11 heterozygous pathogenic/ likely pathogenic variants in eight genes were identified in 11 of the 321 newborn children (Table 3). However, all of these variants were detected in heterozygous state, no child was found to carry homozygous or compound variants in the immunodeficiencies genes that are recorded as being recessive gene in IUIS summary.

Table 3
Overview of the IUIS recommended PIDs identified in 321 newborn children from Qingdao

No.	Condition	Inheritance	Gene	Variants	Classification*	No. of Variants	Het/Hom
1	Adenosine deaminase (ADA) deficiency	AR	ADA	c.424C > T	P	1	Het
				c.872C > T	P	1	Het
2	Ataxia-telangiectasia	AR	ATM	c.67C > T	P	1	Het
3	MOPD1 deficiency (Roifman syndrome)	AR	CLASP1	c.196-562G > A	P	1	Het
4	Immunoskeletal dysplasia with neurodevelopmental abnormalities (EXTL3 deficiency)	AR	EXTL3	c.1970A > G	P	1	Het
5	EDA-ID due to NEMO/IKBKG deficiency (ectodermal dysplasia, immune deficiency)	XLR	IKBKG	c.518G > A	LP	1	Het (Female carrier)
6	JAK3 deficiency	AR	JAK3	c.1744C > T	LP	1	Het
				c.307C > T	LP	1	Het
7	DNA ligase IV deficiency	AR	LIG4	c.1271_1275del	P	1	Het
8	TACI deficiency (Immunodeficiency, common variable)	AD/AR	TNFRSF13B	c.542C > A	LP	2	Het
<i>P: Pathogenic</i>							
<i>LP: Likely pathogenic</i>							

Of note, we found a girl carries a heterozygous variant in *IKBKG*, which is associated with X-linked NEMO deficiency (ectodermal dysplasia, immune deficiency). Two children carried a likely pathogenic variant (NM_012452.2, c.542C > A, p.A181E) in *TNFRSF13B*, which has been reported to be associated with primary immunodeficiency, in particular common variable immunodeficiency although with a low penetrance. The A181E variant has been identified in affected patients in heterozygous, homozygous and compound heterozygous states, and the genetic pattern of the disease is marked as both autosomal dominant and autosomal recessive in the OMIM database. It should be noted however, that the variant has also been seen in normal individuals, and it has been found in 603/24,990 (2.4%) Finnish alleles and 1,540 of 282,092 (0.55%) alleles of different ethnic backgrounds according to the gnomAD database, whereas the prevalence of CVID is around 1/10–50,000 in North America and Europe (18).

Pharmacogenomics

Gene-drug selection was based on published PGx criteria, of which the Clinical Pharmacogenetics Implementation Consortium (CPIC) (19) and the Dutch Pharmacogenetics Working Group (DPWG) (20) guidelines are the most widely recognized. The CPIC and DPWG guidelines both provide clinical recommendations to patients with a known genotype. Minor difference exists between them (20), as the methodologies and clinical practice vary among countries. However, up to now, there is no recommendation on prioritizing which PGx tests should be recommended to an individual beforehand. Recently, the DPWG has developed the Clinical Implication Score (CIS) (20) with the goal to set up a guideline for which drugs testing of specific genetic variants is warranted. The CIS is translated into a three-category recommendation for testing: Essential, Beneficial and Potentially Beneficial (Supplementary Table 4).

In this study, we only focused on the gene-drug pairs according to the DPWG Essential category. We observed that every newborn in the Qingdao cohort carried at least one clinically relevant variant (Fig. 1) of the Essential PGx genes. Among the gene-drug pairs, *CYP2D6* had the highest variant carrying rate (Table 4), where 266 out of 321 infants carrying at least one relevant variant. In total, 150 infants carried one copy of *10 (rs106585), while 81 infants carried two copies of *10, suggesting that at least 25% infants have a decreased function of *CYP2D6* in Codeine metabolism (21, 22). Gene *CYP2C19* showed the second highest variation carrying rate, 209 out of 321 infants carrying at least one clinically relevant variant. Newborns carrying homozygous variants at the *CYP2C19* gene with subtypes *2/*2 and *3/*3 are 25 and 1 respectively, which would lead to lack of enzyme activity and a low metabolization of clopidogrel via the *CYP2C19* pathway (23). In addition, 133 and 122 infants carried variants at *UGT1A1* and *NUDT15* respectively. Homozygous variants at *UGT1A1* (24, 25) and *NUDT15* (26, 27) result in a reduced metabolism of Irinotecan, Azathioprine, Mercaptopurine and Tioguanine. No clinical related variation was detected at *DPYD* (Table 4).

Table 4
Overview of PGx gene-variant statistics in 321 and an allele frequency comparison by population

Gene-drug Pairs		Allele variation	dbSNP RS ID	Number of Carriers in the QD cohort			Total carriers related to the gene	Number of QD cohort	MAF*				
Drug	Gene			Het	Hom	Total			Qingdao	EAS	SAS	AFR	EUR
Azathioprine, Mercaptopurine, Tioguanine	NUDT15	*3	rs116855232	74	5	79	122	321	13.08%	9.52%	6.95%	0.08%	0.20
		*4	rs147390019	1	0	1			0.16%	0.10%	0.00%	0.00%	0.00
		*5	rs186364861	4	0	4			0.62%	1.39%	0.10%	0.00%	0.00
		*6	rs554405994	36	2	38			6.23%	4.76%	0.20%	0.15%	0.30
Irinotecan	UGT1A1	*6	rs4148323	106	15	121	133	321	21.18%	13.79%	1.74%	0.08%	0.70
		*27	rs35350960	12	0	12			1.87%	1.39%	0.00%	0.00%	0.00
Clopidogrel	CYP2C19	*2	rs4244285	141	25	166	209	321	29.75%	31.25%	35.79%	17.02%	14.5
		*3	rs4986893	33	1	34			5.45%	5.56%	1.23%	0.23%	0.00
		*4A/B	rs28399504	1	0	1			0.16%	0.10%	0.00%	0.00%	0.10
		*5	rs56337013	0	0	0			0.00%	0.00%	0.00%	0.00%	0.00
		*6	rs72552267	1	0	1			0.16%	0.00%	0.00%	0.00%	0.00
		*8	rs41291556	0	0	0			0.00%	0.00%	0.10%	0.08%	0.30
		*9	rs17884712	0	0	0			0.00%	0.00%	0.00%	0.98%	0.00
		*10	rs6413438	0	0	0			0.00%	0.00%	0.00%	0.15%	0.00
		*17	rs12248560	7	0	7			1.09%	1.49%	13.60%	23.52%	22.3
Codeine	CYP2D6	*3	rs35742686	0	0	0	266	321	0.00%	0.00%	0.20%	0.23%	1.89
		*4	rs3892097	5	0	5			0.78%	0.20%	10.94%	6.05%	18.5
		*6	rs5030655	0	0	0			0.00%	0.00%	0.10%	0.08%	1.99
		*8	rs5030865	10	1	11			1.87%	0.00%	0.00%	0.00%	0.00
		*9	rs5030656	0	0	0			0.00%	0.00%	0.00%	0.08%	2.58
		*10	rs1065852	150	81	231			48.60%	57.14%	16.46%	11.27%	20.1
		*14A/B	rs5030865	0	0	0			0.00%	0.99%	0.00%	0.00%	0.00
		*17	rs28371706	0	0	0			0.00%	0.00%	0.00%	21.79%	0.20
		*41	rs28371725	18	1	19			3.12%	3.77%	12.17%	1.82%	9.34
Fluorouracil, Capecitabine, Tegafur + DPD-inhibitor	DPYD	*2A	rs3918290	0	0	0	0	321	0.00%	0.00%	0.82%	0.08%	0.50
		*13	rs55886062	0	0	0			0.00%	0.00%	0.00%	0.00%	0.10
		X	rs67376798	0	0	0			0.00%	0.00%	0.10%	0.08%	0.70
		X	rs56038477	0	0	0			0.00%	0.00%	1.94%	0.08%	2.39

Notes: The gene-drugs pairs refer to the DPWG "Essential" category. MAF data refers to 1000 Genome phase 3 dataset. NA indicates no available data from 1 dataset and can thus not be detected by the current pipeline.

We further investigated the differences in allele frequency between the Qingdao cohort dataset and five subpopulations of the 1000 Genome dataset, including East Asians (EAS), South Asians (SAS), Africans (AFR), Europeans (EUR), and Americans (AMR). In most cases, the allele frequency of the Qingdao cohort is consistent with the EAS dataset, while the other four subpopulations differ significantly (Table 4).

Discussion

Around 15 million infants are born in China every year. Previous studies have underlined the importance of newborn screening in early detection and prevention of death or disability. Phenylketonuria (PKU)(28), Congenital Hypothyroidism (CH)(28) and Hearing Loss(29) are included in the national-wide newborn screening program in China, however, many newborn patients and/or carriers remain genetically undiagnosed. This is partly because the gene-by-gene or small panel strategies leave little room for the expansion of diseases association. In this study, we sought to investigate the power of GS in unselected newborns to identify potentially curable or treatable medical conditions in 321 children from Qingdao.

By expanding to 61 RUSP recommended diseases, we detected 121 pathogenic and likely pathogenic variants in 98 (30.53% of samples) newborns from 321 children, while 45.45% (55 out of 121) of variants were recurrent. Hearing loss, CH, methylmalonic acidemia, MMA, PKU and benign hyperphenylalaninemia accounted for half of the pathogenic and likely pathogenic variants in 321 children altogether, while other 25 inherited diseases accounted for the rest 47.60% of variants. Our finding that two and one newborns carried compound heterozygous variants in *GJB2* and *PAH* respectively, were confirmed to be associated with Hearing Loss and PKU, has demonstrated the clinical benefits of GS for newborns and their families. Notably, every assessed newborn was detected at least one important PGx variant at DPWG recommended essential PGx genes. This result was in line with one European 44,000 biobank participants study, where “99.8% of participants had a genotype associated with increased risks to at least one medication”(30). Moreover, a retrospective analysis of 1000 Genome dataset on 120 pharmacogenomics genes across 26 global populations have reported a median of three clinical variants per individual, and East Asian topped super-populations with the highest percentage of loss-of-function variants (60.9%)(31). Therefore, the results suggest that using GS may be a suitable method for screening newborns for variants in a large number of disease-associated genes.

The Wilson-Jungner guidelines (3), has provided the golden standard for what conditions which should be screened for in newborns since 1968. These guidelines have served us well and state that only disorders amenable to treatment should be investigated. However, although not curable, Cystic Fibrosis has been included in the screening programs in several countries. Similarly, screening for Spinal Muscular Atrophy has recently been added to the RUSP list (5), in spite of the prohibitively expensive therapy needed to mitigate the disease which precludes its implementation in many low and middle income countries. However, novel forms of therapy for hitherto incurable diseases, including SMA, is changing the indications for therapy and consequently, the newborn screening process. Thus, gene therapy promises to change the list of diseases which could/should be analyzed in the future. Furthermore, severe combined immunodeficiency (SCID) has recently been added to the RUSP list and screening has been initiated in many countries worldwide, in spite of the fact that “incurable” diseases are identified during the process (including trisomy 21, Nijmegen Breakage Syndrome and Ataxia-Telangiectasia). The latter also raises the question what actually constitutes a curable/treatable medical condition. In the case of Ataxia-telangiectasia, correct identification allows prophylaxis against the accompanying infections (due to immunodeficiency) but will not affect the neurological problems which will subsequently develop. Yet, parents are clearly in favor of being informed about the diagnosis (32), and the mitigating therapy available, even though curable treatment is currently not possible.

Previous studies using ES/GS on newborns have used an indiscriminate inclusion of OMIM defined diseases (7, 8). In our study, however, we have chosen a limited set of disorders including the currently recommended inherited disorders, selected primary immunodeficiency disorders and genes associated with drug reactions (using a standardized set of genes for the latter based on the available guidelines listed in the Clinical Pharmacogenetics Implementation Consortium and the Dutch Pharmacogenetics Working Group), thus limiting the search for genetic variants with known clinical importance. Although the number of children who were affected by variants in the two first categories was limited, every assessed child carried at one genetic variant potentially associated with adverse drug reactions, suggesting a major potential for improvement of personalized drug safety for children.

One important aspect when screening for disorders in a given population is to use a matched control database as variants can be highly specific for a given ethnic group (33, 34). Most databases published to date are based on individuals of European descent and many populations are poorly represented. The Genome Asia 100K project (35) aims to address this gap by sequencing a large number of individuals from different Asian populations and can be used as a reference and is essential as an unexpectedly high allele frequency of a given variant may be highly restricted to a particular ethnic population.

Conclusion

In this study, we applied GS to sequence unselected newborns to investigate potentially curable or treatable medical conditions in 321 children from Qingdao. Selective identification of genetic data, where therapeutic options are available, does not violate the Wilson-Jungner criteria and also provides a basis for future research on variants of unknown significance in an expanding number of genes and should therefore be considered in future screening programs for all newborns.

Abbreviations

GS: Genomic sequencing

ES: Exome sequencing

PID: Primary immunodeficiency diseases

PGx: Pharmacogenetic

CH: Congenital hypothyroidism

PKU: Phenylketonuria

MMA: Methylmalonic acidemia with homocystinuria

SMA: Spinal Muscular Atrophy

RUSP: the recommended Uniform Screening Panel

IUIS: the International Union of Immunological Societies

CPIC: the Clinical Pharmacogenetics Implementation Consortium

Declarations

Ethics approval and consent to participate

The study was approved by the medical ethics committee (Maternal and child health and family planning service center of Huangdao district Ethics Reviews board, ethical permit ID: 20180001) and the Institutional Review Board of BGI (No. BGI-IRB18029 and No. BGI-IRB17073). Each family signed an informed consent form before participating the study.

Consent for publication

Not applicable.

Availability of data and material

The data that support the findings of this study have been deposited into CNGB Sequence Archive(12) of CNGBdb with accession number CNP0001264(13).

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by Guangdong Provincial Key Laboratory of Genome Read and Write (No. 2017B030301011) and the Shenzhen Municipal Government of China (JCY20170817145047361) .

Author's contributions

YS and YG designed the study. MJ, MY and XW wrote the first version of the manuscript. XW, YH, CL, YG, YG, FC, PQ, SG, JC, QZ and LG recruited patients, collected samples, and performed clinical follow-up. ZW, SL, LS and HL performed sample preparation and sequencing experiments. RG, BC and MY conducted bioinformatics analysis. MJ, CF, JG, GZ, YZ, LZ, FS and WX interpreted sequencing results and clinical implications. CN, KK, LH, XJ, JL and FC overviewed the study and finalized the manuscript.

Acknowledgements

We appreciate the participation of the volunteers and their families, without their support, this work would not have been possible. This work was also supported by the China National GeneBank(CNGB).

References

1. Folling A, Sydnese S. A method for the estimation of phenylpyruvic acid in urine with some examples of its use in dietary treatment of phenylpyruvic oligophrenia. *Scand J Clin Lab Invest*. 1958;10(4):355-8.
2. Scriver CC. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants, by Robert Guthrie and Ada Susi, *Pediatrics*, 1963;32:318-343. *Pediatrics*. 1998;102(1 Pt 2):236-7.
3. Wilson JMG, Jungner G, World Health O. Principles and practice of screening for disease / J. M. G. Wilson, G. Jungner. Geneva: World Health Organization; 1968.
4. Chan K, Puck JM. Development of population-based newborn screening for severe combined immunodeficiency. *The Journal of allergy and clinical immunology*. 2005;115(2):391-8.
5. Amatuni GS, Currier RJ, Church JA, Bishop T, Grimbacher E, Nguyen AA, et al. Newborn Screening for Severe Combined Immunodeficiency and T-cell Lymphopenia in California, 2010-2017. *Pediatrics*. 2019;143(2).
6. Nakagawa N, Imai K, Kanegane H, Sato H, Yamada M, Kondoh K, et al. Quantification of κ -deleting recombination excision circles in Guthrie cards for the identification of early B-cell maturation defects. *J Allergy Clin Immunol*. 2011;128(1):223-5.e2.
7. 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(1):76-93.
8. Pavey AR, Bodian DL, Vilboux T, Khromykh A, Hauser NS, Huddlestone K, et al. Utilization of genomic sequencing for population screening of immunodeficiencies in the newborn. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2017;19(12):1367-75.
9. Xu Y, Lin Z, Tang C, Tang Y, Cai Y, Zhong H, et al. A new massively parallel nanoball sequencing platform for whole exome research. *BMC bioinformatics*. 2019;20(1):153.
10. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*. 2010;327(5961):78-81.
11. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, et al. A reference human genome dataset of the BGISEQ-500 sequencer. *GigaScience*. 2017;6(5):1-9.
12. Guo X, Chen F, Gao F, Li L, Liu K, You L, et al. CNSA: a data repository for archiving omics data. *Database (Oxford)*. 2020;2020.

13. Chen Fengzhen YL, Yang Fan, Wang Lina, Guo Xueqin, Gao Fei, Hua Cong, Tan Cong, Fang Lin, Shan Riqiang, Zeng Wenjun, Wang Bo, Wang Ren, Xu Xun, Wei Xiaofeng. CNGbDb: China National GeneBank DataBase. *Hereditas(Beijing)*. 2020;42(8):799-809.
14. Picard C, Bobby Gaspar H, Al-Herz W, Bousfiha A, Casanova JL, Chatila T, et al. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. *Journal of clinical immunology*. 2018;38(1):96-128.
15. Bousfiha A, Jeddane L, Picard C, Ailal F, Bobby Gaspar H, Al-Herz W, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *Journal of clinical immunology*. 2018;38(1):129-43.
16. Li L, Lu J, Tao Z, Huang Q, Chai Y, Li X, et al. The p.V37I exclusive genotype of GJB2: a genetic risk-indicator of postnatal permanent childhood hearing impairment. *PLoS One*. 2012;7(5):e36621.
17. Dai ZY, Sun BC, Huang SS, Yuan YY, Zhu YH, Su Y, et al. Correlation analysis of phenotype and genotype of GJB2 in patients with non-syndromic hearing loss in China. *Gene*. 2015;570(2):272-6.
18. Saikia B, Gupta S. Common Variable Immunodeficiency. *Indian J Pediatr*. 2016;83(4):338-44.
19. Relling MV, Klein TE, Gammal RS, Whirl-Carrillo M, Hoffman JM, Caudle KE. The Clinical Pharmacogenetics Implementation Consortium: 10 Years Later. *Clin Pharmacol Ther*. 2020;107(1):171-5.
20. Swen JJ, Nijenhuis M, van Rhenen M, de Boer-Veger NJ, Buunk AM, Houwink EJJ, et al. Pharmacogenetic Information in Clinical Guidelines: The European Perspective. *Clin Pharmacol Ther*. 2018;103(5):795-801.
21. Crews KR, Gaedigk A, Dunnenberger HM, Klein TE, Shen DD, Callaghan JT, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. *Clin Pharmacol Ther*. 2012;91(2):321-6.
22. Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoeconomic and clinical aspects. *Pharmacol Ther*. 2007;116(3):496-526.
23. Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther*. 2013;94(3):317-23.
24. Zhang X, Yin JF, Zhang J, Kong SJ, Zhang HY, Chen XM. UGT1A1*6 polymorphisms are correlated with irinotecan-induced neutropenia: a systematic review and meta-analysis. *Cancer Chemother Pharmacol*. 2017;80(1):135-49.
25. Akaba K, Kimura T, Sasaki A, Tanabe S, Ikegami T, Hashimoto M, et al. Neonatal hyperbilirubinemia and mutation of the bilirubin uridine diphosphate-glucuronosyltransferase gene: a common missense mutation among Japanese, Koreans and Chinese. *Biochem Mol Biol Int*. 1998;46(1):21-6.
26. Yang JJ, Whirl-Carrillo M, Scott SA, Turner AJ, Schwab M, Tanaka Y, et al. Pharmacogene Variation Consortium Gene Introduction: NUDT15. *Clin Pharmacol Ther*. 2019;105(5):1091-4.
27. Relling MV, Schwab M, Whirl-Carrillo M, Suarez-Kurtz G, Pui CH, Stein CM, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for Thiopurine Dosing Based on TPMT and NUDT15 Genotypes: 2018 Update. *Clin Pharmacol Ther*. 2019;105(5):1095-105.
28. Mei L, Song P, Xu L. Newborn screening and related policy against Phenylketonuria in China. *Intractable Rare Dis Res*. 2013;2(3):72-6.
29. Wang Q, Xiang J, Sun J, Yang Y, Guan J, Wang D, et al. Nationwide population genetic screening improves outcomes of newborn screening for hearing loss in China. *Genetics in Medicine*. 2019;21(10):2231-8.
30. Reisberg S, Krebs K, Lepamets M, Kals M, Mägi R, Metsalu K, et al. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. *Genetics in Medicine*. 2019;21(6):1345-54.
31. Wright GEB, Carleton B, Hayden MR, Ross CJD. The global spectrum of protein-coding pharmacogenomic diversity. *The Pharmacogenomics Journal*. 2018;18(1):187-95.
32. Blom M, Schoenaker MHD, Hulst M, de Vries MC, Weemaes CMR, Willemsen M, et al. Dilemma of Reporting Incidental Findings in Newborn Screening Programs for SCID: Parents' Perspective on Ataxia Telangiectasia. *Front Immunol*. 2019;10:2438.
33. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
34. Whiffin N, Ware JS, O'Donnell-Luria A. Improving the understanding of genetic variants in rare disease with large-scale reference populations. *JAMA: Journal of the American Medical Association*. 2019;322:1305-6.
35. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019;576(7785):106-11.

Figures

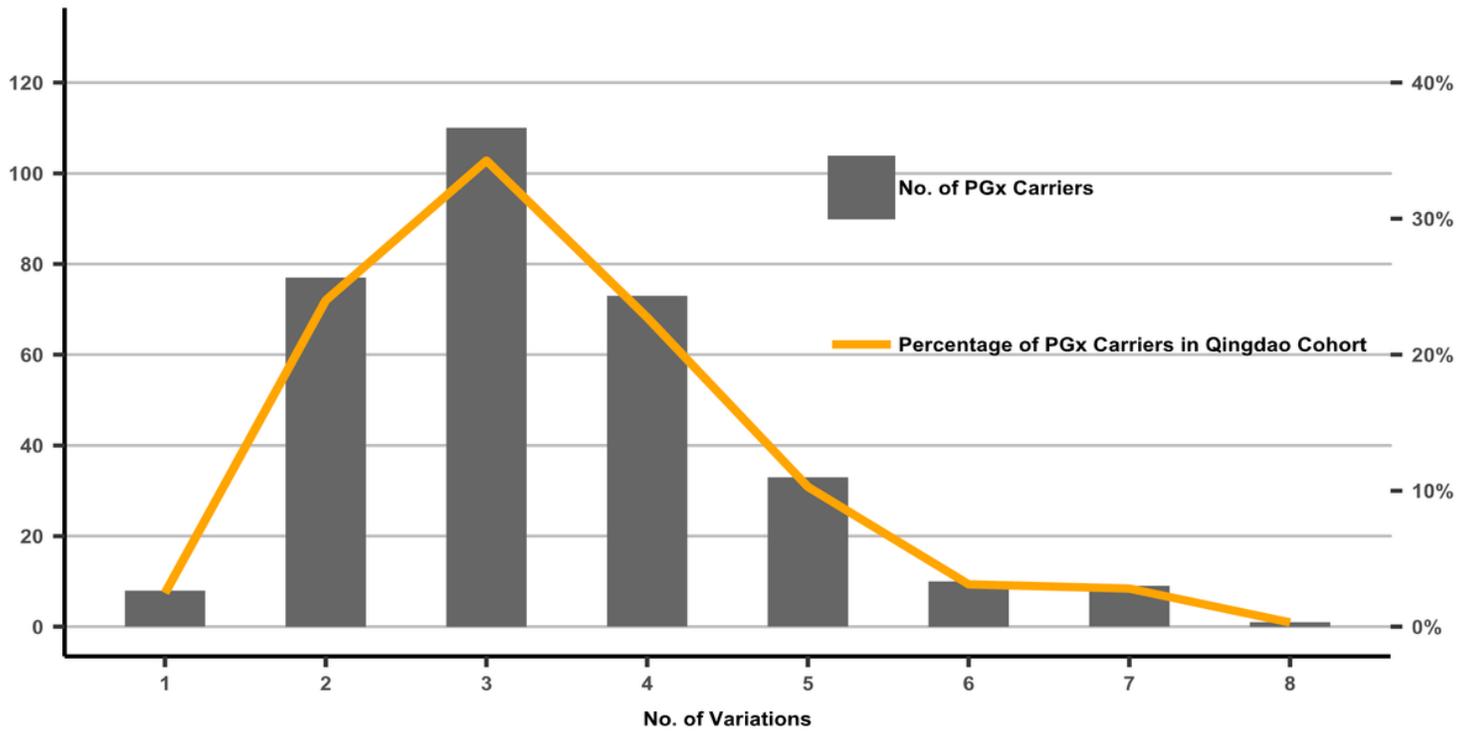


Figure 1

Distribution of Essential PGx variations in 321 Qingdao cohort (n=321). A clinical management strategy can be adopted for every carrier regarding DPWG guidelines. Among 321 newborns, every newborn in the Qingdao cohort carried at least one clinically relevant variant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementarymaterialsNBS20201107.docx](#)