Received: 2 November 2020 | Revised: 8 January 2021 | Accepted: 9 January 2021

DOI: 10.1111/cge.13923

SHORT REPORT

CLINICAL GENETICS WILEY

Early genetic screening uncovered a high prevalence of thalassemia among 18 309 neonates in Guizhou, China

Mei Tan[1] | Yue Bai[2] | Xiangmei Zhang[1] | Jian Sun[1] | Chengshuang Huang[1] | Runmei Tian[1] | Yuhang Yang[1] | Xi Luo[1] | Qiong Su[1] | Liusong Wu[1] | Libo Zheng[3] | Jing Xia[4] | Hongmei Murong[5] | Ping Zhu[6] | Fan Yang[2] | Xiaosong Zhong[2] | Jindong Chen[7] | Yan Chen[1]

[1]Department of Pediatrics, Affiliated Hospital of Zunyi Medical University, Zunyi, China

[2]The Clinical Center of Gene and Cell Engineering, Capital Medical University Affiliated Shiji Tan Hospital, Beijing, China

[3]Department of Obstetrics and Gynecology, Xingyi City Hospital, Xingyi, Guizhou, China

[4]Department of Pediatrics, Xingyi City Hospital, Xingyi, Guizhou, China

[5]Department of Obstetrics, Qianxinan Prefecture Hospital, Xingyi, Guizhou, China

[6]Department of Hematology, Peking University First Hospital, Beijing, China

[7]Department of Research, Exploring Health LLC, Guangzhou, China

Correspondence
Yan Chen, Department of Pediatrics, Affiliated Hospital of Zunyi Medical University, 201 Dalian Rd., Zunyi, China.
Email: cyz600@163.com

Funding information
Department of Science and Technology in Guizhou, Grant/Award Number: [2019]2806

Abstract

Thalassemia is a common monogenic disease in southwestern China, especially in Guizhou province. In this study, 18 309 neonates were examined for thalassemia. The thalassemia carrier rate was 12.90%, which is associated with geographical regions, with carrier frequencies significantly differing between regions ($p < 0.0001$). The carrier rates for α-thalassemia and β-thalassemia were 8.91% and 3.36%, respectively. There are 22 genotypes identified among 1632 α-thalassemia cases, and 18 genotypes detected among 615 β-thalassemia cases. The birthrates of individuals with intermediate thalassemia and β-thalassemia major were 0.153% and 0.055%, respectively. Methodologically, NGS-Gap-PCR is superior to traditional detection methods, with 65 more cases detected by NGS-Gap-PCR. Since thalassemia-rich genotypes were highly prevalent in this region, early detection of thalassemia carriers would be meaningful for genetic counseling and prevention/treatment of thalassemia. NGS-Gap-PCR provides a powerful tool for neonate genetic testing and clinical diagnosis of thalassemia, especially in high-prevalence regions.

KEYWORDS
gap-PCR, genetic counseling, hemoglobin, neonates, next-generation sequencing, thalassemia epidemiology

1 | INTRODUCTION

Thalassemia is manifested primarily as chronic progressive hemolytic anemia.[1] The degree of anemia varies, depending on the type and amount of hemoglobin that is synthesized. Thalassemia minor may present as mild anemia or a complete lack of clinical symptoms, while thalassemia major often leads to severe anemia that is a serious threat to the lives and health of children.[2] Each year, approximately 5500 severe α-thalassemia patients die during the perinatal period, and 50 500 β-thalassemia major patients are born throughout the entire world, which at least 30 000 require lifetime standard blood transfusions to survive.[3] Early identification of thalassemia is beneficial for timely prevention and control of thalassemia major.[4] Thalassemia gene screening and related genetic counseling are the most effective ways to reduce the incidence of thalassemia major.[5]

Guizhou province, China consists of nine states with a population of 36 229 500 under 176 167 km^2, bordering the southern provinces of Yunnan, Guangxi, Hunan, Sichuan, and Chongqing City. Like its neighbor provinces such as Yunnan and Guangxi, Guizhou province has a high prevalence of thalassemia. Qianxinan State, one of its states, located in southwestern Guizhou Province, and adjacent to Yunnan and Guangxi, is geographically linked to the high incidence of thalassemia in this region. Previous studies on thalassemia have been limited due to small sample size, and most of the study subjects were

* Mei Tan, Yue Bai, Jindong Chen and Yan Chen contributed equally to this work.

children plus adults.[6,7] The thalassemia carrier rate in the Guizhou population has not yet been subjected to a large-scale and more comprehensive epidemiological survey. Therefore, we used next-generation sequencing (NGS) combined with Gap-PCR to detect the prevalence of the thalassemia gene mutations in neonates born from June 2016 to June 2017 in Qianxinan State to assess the thalassemia carrier rate, incidence of thalassemia major and provide a theoretical basis for the screening, prevention, and treatment of thalassemia.

2 | MATERIALS AND METHODS

2.1 | Participants and study design

The study was approved by Beijing Genomics Institute (BGI) institutional review board on bioethics and biosafety (BGI-IRB), and all individuals' parents provided written informed consent. Qianxinan State in Guizhou Province is composed of eight cities/districts including Anlong, Ceheng, Puan, Qinglong, Wangmo, Xingren, Xingyi and Zhenfeng. From June 2016 to June 2017, 18 309 umbilical cord blood samples were collected from all the neonates born in all eight cities/districts of Qianxinan State Four blood spots of umbilical cord blood were collected from each neonate at birth.

2.2 | Testing methods

DNA extraction was performed using the Magen nucleic acid extraction kit (GenMagbio DOF-9696PRO, Guangzhou, China). Screening of thalassemia gene mutations was carried out by Beijing Genomics Institute-Shenzhen (BGI-Shenzhen) through its NGS plus Gap-PCR routine protocol that has been described in previous study.[8] Briefly, full length of *HBA1*, *HBA2*, *HBZ*, *HBQ1*, *HBB*, *HBD*, *HBG1*, *HBG2*, and *HBE1* and the key regulatory regions (*5′HVR*, *3′HVR*, *HS40*, *HS-1*, *HS-2*, *HS-3*, *HS-4*, and *HS-5*) of the human α-and β-globin gene clusters, which covers nearly all reported pathogenic *SNVs/Indels* coding and non-coding regions in the two gene clusters. Sequencing libraries were constructed according the Illumina HiSeq sequencing library preparation protocol. These libraries were further paired-end-sequenced for 100 base pairs (PE100) with an Illumina HiSeq 2000 machine. The bioinformatic analysis of identifying hemoglobin gene variations was described in detail previously.[8] Variations from 63 cases were further validated with Sanger sequencing. Gap-PCR was used to identify α-globin deletions and β-globin deletions including the three most common α-thalassemia-associated deletions ($-^{SEA}$, $-\alpha^{3.7}$, and $-\alpha^{4.2}$), two rare deletions ($-^{FIL}$ and $-^{THAI}$), and two common β-globin gene deletions (SEA-HPFH and $G\gamma^{+}[^{As}\gamma\delta\beta]$).

2.3 | Statistical analysis

A chi-square (χ^2) test was used to evaluate the differences in thalassemia carrier rates (α-thalassemia, β-thalassemia and combined α-/β-thalassemia) between different regions and sexes. Two separate analyses were conducted to address the different types of thalassemia compared to a healthy group: α and combined α-/β-thalassemia and β and combined α-/β-thalassemia. The crude thalassemia carrier rate for each type was calculated for each region and sex; crude odds ratios (ORs) and 95% confidence intervals (CIs) for comparing each region to Xingren (reference, the largest population surveyed) and females to males were calculated using a univariate logistic regression model. A multivariate logistic regression model was used for the adjusted analysis. The variables in the multivariate models included sex and region. The adjusted ORs (adj.ORs) and 95% CIs were determined. A 95% CI that did not include 1.00 was considered statistically significant at an $\alpha = .05$ level of significance. The analyses were performed using SAS 9.4 (Proc Logistic, SAS Institute, Cary, NC, USA).

3 | RESULTS

3.1 | Thalassemia carrier rates

Among the 18 309 subjects, 10 057 were males and 8252 were females (Table 1). Of them, 2362 (12.9%) were identified to be carriers of the thalassemia genes, including 1632 α-thalassemia, 615 β-thalassemia, 115 combined α-/β-thalassemia (Table 1). In addition, we observed 93 abnormal hemoglobinopathies. The characterization and annotation of the globin gene mutations/deletions were performed as described in previous studies.[9,10] Geographically, Ceheng presented the highest carrier rate (27.98%, 394/1408), and Anlong showed the lowest rate (5.56%, 3/54). These frequencies showed statistically significant differences between the regions ($p < 0.0001$). In addition, variants from 63 cases were further confirmed through Sanger sequencing, and the results were consistent with the NGS data. The variant types and corresponding numbers were listed in the Supplementary Table S1.

For the frequencies of the α-thalassemia, β-thalassemia, and combined α-/β-thalassemia, no statistically significant associations were observed between the gender and the detection rates ($p > 0.05$). The crude ORs and adjusted ORs (adj.ORs) for each region compared to the reference (Xingren, which had the largest survey population and was used as a reference baseline) are presented in (Table 2). For α only and combined α-/β-thalassemia, the odds were 3.2 times higher in Cehen and 2.6 times higher in Wangmo (adj.OR [95% CI]: 4.15 (3.52–4.90); adj.OR [95% CI]: 3.56 (3.01–4.22)). A similar pattern was observed for β only and combined α-/β-thalassemia in these two regions. With respect to sex, we did not observe any statistically significant associations.

3.2 | Frequency of the α-globin genotypes and α-thalassemia

There were 1632 (8.91%) α-thalassemia mutation carriers identified, including 1604 cases of α-thalassemia minor and 28 cases of

TABLE 1 Carrier frequencies of α-thalassemia, β-thalassemia, and combined α-/β-thalassemia

	Health (%)	α-thalassemia (%)	β-thalassemia (%)	Combined α-/β- thalassemia (%)	Total	Frequency (%)
Total *N* (%)	15 947 (87.10)	1632 (8.91)	615 (3.36)	115 (0.63)	18 309	12.9
Region						
Anlong	51 (−94.45)	1 (1.85)	2 (3.70)	0 (0.00)	54	5.56
Ceheng	1014 (72.02)	271 (19.25)	84 (5.96)	39 (2.77)	1408	27.98
Puan	1060 (92.09)	67 (5.82)	19 (1.65)	5 (−0.44)	1151	7.91
Qinglong	1822 (−86.85)	192 (9.15)	76 (3.62)	8 (0.38)	2098	13.16
Wangmo	1055 (74.14)	251 (17.64)	91 (6.39)	26 (1.83)	1423	25.86
Xingren	4967 (90.8)	350 (6.4)	137 (−2.51)	16 (0.29)	5470	9.20
Xingyi	3846 (91.35)	250 (5.94)	106 (2.52)	8 (0.19)	4210	8.65
Zhenfeng	2132 (85.45)	250 (10.02)	100 (4.01)	13 (0.52)	2495	14.55
Sex						
Female	7213 (87.41)	723 (8.76)	266 (3.22)	50 (0.61)	8252	12.59
Male	8734 (86.84)	909 (9.04)	349 (3.47)	65 (0.65)	10 057	13.16

Abbreviation: Total *N*, total subject number.

TABLE 2 Comparison of crude and adjusted odds ratios between regions and sexes by thalassemia type

	α-thalassemia and combined α-/β-thalassemia		β-thalassemia and combined α-/β-thalassemia	
Regions	**Crude odds ratio (95% CI)**	**Adjusted odds ratio (95% CI)**	**Crude odds ratio (95% CI)**	**Adjusted odds ratio (95% CI)**
Anlong	0.27 (0.04–1.93)	0.27 (0.04–1.94)	1.27 (0.31–5.28)	1.28 (031–5.30)
Ceheng	4.15 (3.52–4.90)	4.15 (3.52–4.90)	3.94 (3.08–5.04)	3.95 (3.08–5.05)
Puan	0.92 (0.71–1.20)	0.92 (0.71–1.20)	0.74 (0.48–1.14)	0.73 (0.48–1.13)
Qinglong	1.49 (1.24–1.78)	1.49 (1.24–1.78)	1.50 (1.14–1.96)	1.50 (1.14–1.96)
Wangmo	3.56 (3.01–4.22)	3.56 (3.01–4.22)	3.60 (2.81–4.62)	3.60 (2.80–4.62)
Xingyi	0.91 (0.77–1.07)	0.91 (0.77–1.07)	0.96 (0.75–1.23)	0.96 (0.75–1.23)
Zhenfeng	1.67 (1.41–1.98)	1.67 (1.42–1.98)	1.72 (1.34–2.21)	1.72 (1.34–2.20)
Xingren	Ref			
Sex				
Female	0.96 (0.87–1.06)	0.96 (0.87–1.07)	0.92 (0.80–1.07)	0.92 (0.79–1.07)
Male	Ref			

intermediate thalassemia, caused by 22 distinct genotypes including deletion (1315 cases), nondeletion mutations (292 cases), and deletion combined with nondeletion mutations (25 cases). More than 42.71% (697/1632) of the carriers harbored an αα/-$\alpha^{3.7}$ gene deletion, and αα/-^{SEA} was less common, at 26.47% (432/1632). Hb CS heterozygosis was the most common genotype (12.75%, 208/1632). There was no association between genotype and sex ($p > 0.05$). The 28 cases (1.72%, 28/1632) of intermediate thalassemia were identified as having a genotype-predictive phenotype (Table 3).

3.3 | Frequency of the β-globin genotypes and β-thalassemia

The β-thalassemia mutation frequency was 3.36% (615/18309). There were 18 genotypes found, all of which were point mutations. Among these mutations, codon 17 (A- > T) β^0 heterozygosis was the most common alteration (44.39%, 273/615), followed by codon 41/42 (-TTCT) β^0 heterozygosis (27.48%, 169/615) and IVS-II-654 C > T β^+ heterozygosis (9.27%, 57/615). In addition, 10 cases (0.055%, 10/18309) of β-thalassemia major and five related genotypes were detected. The codon 17 (A > T) β^0 homozygosis accounted for 60% (6/10) (Table 3).

3.4 | Composite α-/β-thalassemia

We identified a total of 115 cases of combined α-/β-thalassemia, with a carrier rate of 0.63% (115/18309) across 65 males and 50 females. There were 29 types of genotypes detected. Among these genotypes, the most frequently occurring gene mutation was αα/-$\alpha^{3.7}$ combined with codon 17 (A- > T) β^0 heterozygosis, accounting for 19.13% (22/115) (Table 3).

TABLE 3 The genotype distribution pattern of all types of hemoglobinopathies

Genotype		All		Female			Male			χ^2	*p* value
		Case	Case col %	Case	Case col %	Case row %	Case	Case col %	Case row %		
α-thalassemia	All	1632	100	723	100	44.3	909	100	55.7	19.4883	0.615
	Hb Agrinio Heterozygosis combined with αα/−$\alpha^{3.7}$	1	0.06	1	0.14	100	0	0	0		
	Hb Constant Spring (Hb CS) combined with αα/−^{SEA}	10	0.61	4	0.55	40	6	0.66	60		
	Hb Constant Spring (Hb CS) combined with αα/−$\alpha^{3.7}$	5	0.31	2	0.28	40	3	0.33	60		
	Hb Constant Spring (Hb CS) combined with αα/−$\alpha^{4.2}$	4	0.25	2	0.28	50	2	0.22	50		
	Hb Westmead combined with αα/−^{SEA}	1	0.06	0	0	0	1	0.11	100		
	Hb Westmead combined with αα/−$\alpha^{3.7}$	3	0.18	1	0.14	33.33	2	0.22	66.67		
	-$\alpha^{3.7}$/−^{SEA}	10	0.61	1	0.14	100	0	0	0		
	-$\alpha^{3.7}$/−$\alpha^{3.7}$	14	0.86	3	0.41	33.33	6	0.66	66.67		
	-$\alpha^{3.7}$/−$\alpha^{4.2}$	9	0.55	7	0.97	50	7	0.77	50		
	-$\alpha^{4.2}$/−-^{SEA}	5	0.31	3	0.41	33.33	6	0.66	66.67		
	-α4.2/−α4.2	1	0.06	1	0.14	20	4	0.44	80		
	αα/−-^{SEA}	432	26.47	0	0	0	1	0.11	100		
	αα/−-THAI	5	0.31	179	24.76	41.44	253	27.83	58.56		
	αα/−$\alpha^{3.7}$	697	42.71	5	0.69	100	0	0	0		
	αα/−$\alpha^{4.2}$	143	8.76	324	44.81	46.48	373	41.03	53.52		
	Hb Constant Spring (Hb CS) Heterozygosis	208	12.75	60	8.3	41.96	83	9.13	58.04		
	Hb Constant Spring (Hb CS) Homozygosis	2	0.12	90	12.45	43.27	118	12.98	56.73		
	Hb Phnom Penh Heterozygosis	1	0.06	1	0.14	50	1	0.11	50		
	Hb Quong Sze Heterozygosis	9	0.55	0	0	0	1	0.11	100		
	Hb Westmead Heterozygosis	70	4.29	4	0.55	44.44	5	0.55	55.56		
	Initiation codon (−T) Heterozygosis	1	0.06	35	4.84	50	35	3.85	50		
	Poly A (A- > G) AATAAA- > AATGAA Heterozygosis	1	0.06	0	0	0	1	0.11	100		
β-thalassemia	All	615	100	266	100	43.25	349	100	56.75	21.719	0.1958
	−28 (A > G) (HBB:c.-78A > G) β^+ Heterozygosis	9	1.46	1	0.38	11.11	8	2.29	88.89		
	−29 (A > G) (HBB:c.-79A > G) β^+ Heterozygosis	5	0.81	2	0.75	40	3	0.86	60		
	−50 (G > A) (HBB: c.-100G > A) β^+Heterozygosis	33	5.37	17	6.39	51.52	16	4.58	48.48		
	Codon 17 (A > T) (HBB:c.52A > T) β^0 Heterozygosis combined with (HBB:c.316_197C > T)IVS-II-654 (C > T) β^+Heterozygosis	1	0.16	0	0	0	1	0.29	100		
	Codon 17 (A > T) (HBB:c.52A > T)β^0 Heterozygosis combined with Codons 41/42 (-TTCT) (HBB:c.124_127delTTCT) β^0Heterozygosis	1	0.16	0	0	0	1	0.29	100		

(Continues)

TABLE 3 (Continued)

Genotype		All		Female			Male			χ^2	*p* value
		Case	Case col %	Case	Case col %	Case row %	Case	Case col %	Case row %		
	5′UTR +43 to +40 (-AAAC) (HBB:c.-11_-8delAAAC)β^+ Heterozygosis	3	0.49	3	1.13	100	0	0	0		
	Codon 17 (A- > T) (HBB:c.52A > T)β^0 Heterozygosis/Codons 71/72 (+A) HBB:c.216_217insAβ^0 Heterozygosis	1	0.16	0	0	0	1	0.29	100		
	Codon 17 (A > T) (HBB:c.52A > T) β^0 Heterozygosis	273	44.39	114	42.86	41.76	159	45.56	58.24		
	Codon 17 (A > T) (HBB:c.52A > T)β^0 Homozygosis	6	0.98	3	1.13	50	3	0.86	50		
	Codon 37 (TGG > TAG) (HBB:c.113G > A)Heterozygosis	3	0.49	2	0.75	66.67	1	0.29	33.33		
	Codon 43 (G > T) (HBB:c.130G > T)β^0 Heterozygosis	6	0.98	5	1.88	83.33	1	0.29	16.67		
	Codon 5 (-CT) (HBB:c.17_18delCT)β^0 Heterozygosis	1	0.16	0	0	0	1	0.29	100		
	Codons 27/28 (+C) (HBB:c.84_85insC)β^0 Heterozygosis	2	0.33	0	0	0	2	0.57	100		
	Codons 41/42 (-TTCT) (HBB:c.124_127delTTCT) β^0 Heterozygosis	169	27.48	74	27.82	43.79	95	27.22	56.21		
	Codons 71/72 (+A) (HBB:c.216_217insA)β^0 Heterozygosis	7	1.14	4	1.5	57.14	3	0.86	42.86		
	Hb E (HBB:c.79G > A)Heterozygosis	33	5.37	14	5.26	42.42	19	5.44	57.58		
	IVS-II-654 (C > T) (HBB:c.316_197C > T)β^+ Heterozygosis	57	9.27	23	8.65	40.35	34	9.74	59.65		
	IVS-II-705 (T > G) (HBB: c.316-146 T > G)β^+Heterozygosis	5	0.81	4	1.5	80	1	0.29	20		
α combined with β	All	115	100	50	100	43.48	65	100	56.52	28.1591	0.456
	Hb Constant Spring (Hb CS) Heterozygosis combined with −50 (G > A) β^+ Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	Hb Constant Spring (Hb CS) Heterozygosis combined with Codon 17 (A > T) β^0Heterozygosis	15	13.04	6	12	40	9	13.85	60		
	Hb Constant Spring (Hb CS) Heterozygosis combined with Codons 41/42 (-TTCT) β 0Heterozygosis	6	5.22	5	10	83.33	1	1.54	16.67		
	Hb Constant Spring (Hb CS) Heterozygosis combined with Hb E Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	Hb Constant Spring (Hb CS) Heterozygosis combined with IVS-II-654 (C > T) β^+ Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	Hb Westmead Heterozygosis combined with 5′UTR +43 to +40 (-ααAC) β^+ Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	Hb Westmead Heterozygosis combined with Codon 17 (A > T) β^0 Heterozygosis	3	2.61	1	2	33.33	2	3.08	66.67		
	Hb Westmead Heterozygosis combined with Codons 41/42 (-TTCT) β^0 Heterozygosis	2	1.74	1	2	50	1	1.54	50		
	$-\alpha^{3.7}/--^{SEA}$ combined with −50 (G > A) β^+ Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	$-\alpha^{3.7}/-\alpha^{3.7}$ combined with Codon 17 (A > T) β^0 Heterozygosis	1	0.87	1	2	100	0	0	0		

TABLE 3 (Continued)

Genotype		All		Female			Male			χ^2	*p* value
		Case	Case col %	Case	Case col %	Case row %	Case	Case col %	Case row %		
	-$\alpha^{4.2}$/−$\alpha^{4.2}$ combined with Codons 41/42 (-TTCT) β^0 Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	αα/−-^{SEA} combined with −28 (A > G) β^+ Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	αα/−-^{SEA} combined with −50 (G > A) β + Heterozygosis	4	3.48	2	4	50	2	3.08	50		
	αα/−-^{SEA} combined with Codon 17 (A > T) β^0 Heterozygosis	9	7.83	2	4	22.22	7	10.77	77.78		
	αα/−-^{SEA} combined with Codons 41/42 (-TTCT) β^0 Heterozygosis	8	6.96	2	4	25	6	9.23	75		
	αα/−-^{SEA} combined with Hb E Heterozygosis	3	2.61	1	2	33.33	2	3.08	66.67		
	αα/−$\alpha^{3.7}$ combined with −28 (A > G) β^+ Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	αα/−$\alpha^{3.7}$ combined with −50 (G > A) β^+ Heterozygosis	2	1.74	1	2	50	1	1.54	50		
	αα/−$\alpha^{3.7}$ combined with Codon 17 (A > T) β^0 Heterozygosis 2 and Codons 41/42 (-TTCT) β^0 Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	αα/−$\alpha^{3.7}$ combined with Codon 17 (A > T) β^0 Heterozygosis	22	19.13	12	24	54.55	10	15.38	45.45		
	αα/−$\alpha^{3.7}$ combined with Codons 27/28 (+C) β^0 Heterozygosis	1	0.87	1	2	100	0	0	0		
	αα/−$\alpha^{3.7}$ combined with Codons 41/42 (-TTCT) β^0 Heterozygosis	13	11.3	7	14	53.85	6	9.23	46.15		
	αα/−$\alpha^{3.7}$ combined with Hb E Heterozygosis	1	0.87	0	0	0	1	1.54	100		
	αα/−$\alpha^{3.7}$ combined with IVS-I-1 (G > T) β^0 Heterozygosis	1	0.87	1	2	100	0	0	0		
	αα/−$\alpha^{3.7}$ combined with IVS-II-654 (C > T) β^+ Heterozygosis	3	2.61	2	4	66.67	1	1.54	33.33		
	αα/−$\alpha^{4.2}$ combined with −50 (G > A) β^+ Heterozygosis	3	2.61	2	4	66.67	1	1.54	33.33		
	αα/−$\alpha^{4.2}$ combined with Codon 17 (A > T) β^0 Heterozygosis	5	4.35	0	0	0	5	7.69	100		
	αα/−$\alpha^{4.2}$ combined with Codons 41/42 (-TTCT) β^0 Heterozygosis	2	1.74	2	4	100	0	0	0		
	αα/−$\alpha^{4.2}$ combined with Hb E Heterozygosis	2	1.74	1	2	50	1	1.54	50		
Abnomal hemoglobinopathy	All	93	100	50	100	53.76	43	100	46.24		
	Hb Alesha Heterozygosis Hb Hekinan II Heterozygosis	1	1.08	0	0	0	1	2.33	100		
	Hb Fuchu-I Heterozygosis	1	1.08	0	0	0	1	2.33	100		
	Hb G-Honolulu Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb G-Taipei Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Hekinan II	1	1.08	0	0	0	1	2.33	100		
	Hb Hekinan II Heterozygosis	30	32.26	14	28	46.67	16	37.21	53.33		
	Hb J-Bangkok Heterozygosis	7	7.53	3	6	42.86	4	9.3	57.14		
	Hb New York Heterozygosis	11	11.83	7	14	63.64	4	9.3	36.36		

(Continues)

TABLE 3 (Continued)

Genotype		All		Female			Male			χ^2	*p* value
		Case	Case col %	Case	Case col %	Case row %	Case	Case col %	Case row %		
	Hb Port Phillip Heterozygosis	1	1.08	0	0	0	1	2.33	100		
	Hb Q-Thailand Heterozygosis	16	17.2	6	12	37.5	10	23.26	62.5		
	Hb Shizuoka Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Zurich-Langstrasse Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Broomhill Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb G-Coushatta Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Groene Hart Heterozygosis	2	2.15	0	0	0	2	4.65	100		
	Hb Hekinan II Heterozygosis	4	4.3	3	6	75	1	2.33	25		
	Hb Hekinan II Homozygosis	1	1.08	1	2	100	0	0	0		
	Hb Hope Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Izmir Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb J-Bangkok Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb J-Lome Heterozygosis	1	1.08	0	0	0	1	2.33	100		
	Hb J-Wenchang-Wuming Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb New York Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Q-Thailand Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Queens Park Heterozygosis	1	1.08	0	0	0	1	2.33	100		
	Hb Rio Claro Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Shenyang Heterozygosis	1	1.08	1	2	100	0	0	0		
	Hb Zurich-Langstrasse Heterozygosis	2	2.15	2	4	100	0	0	0		

Abbreviation: col, column.

3.5 | Other types of hemoglobinopathy

We identified 93 (0.51%, 93/18309) cases of other types of hemoglobinopathy with 21 types of gene mutation. Hb Hekinan II heterozygosis, Hb Q-Thailand heterozygosis and Hb New York heterozygosis ranked among the top three, accounting for 32.26% (30/93), 17.2% (16/93), and 11.83% (11/93), respectively. In addition, 23 cases of hemoglobinopathy were also suffers of α-thalassemia; seven cases were suffers of β-thalassemia; and four cases were associated with combined–/β-thalassemia (Table 3).

3.6 | NGS combined with gap-PCR is superior to traditional screening technology

Of the2,362 thalassemia-carriers, 65 (2.75%, 29 males, 36 females) were undetectable by traditional gene screening technology, including 8 cases of α-thalassemia, 45 cases of β-thalassemia, and 12 cases of combined α-/β-thalassemia. -50 (G > A) β^+ heterozygosis was the most common genotype, accounting for 50.77% (33/65).

4 | DISCUSSION

Thalassemia is considered an epidemic in Guizhou province of southwestern China. Early identification and diagnosis of thalassemia are essential for predicting the clinical manifestations based on genotype–phenotype correlation for genetic counseling. Herein, by using NGS combined with Gap-PCR, we for the first time performed a large-scale screening of thalassemia, focusing only on the neonates (18 309 cases) born in Qianxinan State between June 2016 and June 2017. Since no adults were included, the carrier rate obtained in this study accurately reflected the incidence of the thalassemia carriers in this subpopulation.

The incidence of thalassemia carriers in Qianxinan State is 12.90%, which is significantly higher than the national average carrier rate of thalassemia of 3.62%,[11] and indicating thalassemia is a serious public health problem in this region. The region with the highest prevalence was Ceheng (27.98%), followed by Wangmo (25.86%). The reason for this trend may be that Ceheng and Wangmo are inhabited by many ethnic minorities with interethnic marriage and living habits that can increase the risk of thalassemia.

The carrier rate (8.91%) for α-thalassemia in this study was significantly higher than the previously reported 3.28% in Guizhou.[7] These differences may be related to the sample sizes and the detection method of thalassemia gene. In addition, we detected 20 types of α-thalassemia genotypes. While previous reports indicated that αα/$-^{SEA}$ was most common in southern China,[7] we identified that αα/$—^{3.7}$ was the most common mutation (42.71%) in this subpopulation, followed by αα/$—^{SEA}$ (26.47%) and Hb CS heterozygosis (12.75%).

The prevalence of β-thalassemia (3.36%) in this study is much higher than the reported average 0.67% in China. Of the18 detected β-thalassemia genotypes, the codon 17 (A- > T) β^0 heterozygote (44.39%), codon 41/42 (-TTCT) β^0 heterozygote (27.48%), and IVS-II-654 (C > T) β^+ heterozygote (9.27%) were the three most common mutations, which is consistent with the reported data in Guizhou Province.[6,12] However, these results differed from findings from Guangdong Province and Hainan Province, where the most common gene mutation was codon 41/42.[13,14] These results suggest that prevalence of β-thalassemia and the genotype distribution are geographically associated.

By using NGS-Gap-PCR, we detected 65 thalassemia-carriers and related mutations that are undetectable by traditional method, including the high-prevalent -50 G > A mutation and a series of rare mutations such as the poly A (A- > G) AATAAA- > AATGAA, Hb Phnom Penh, and initiation codon (-T) mutations. This finding further suggests that NGS combined with Gap-PCR can effectively identify new mutations and reduce the rate of misdiagnosis.[15]

Approximately 0.51% of the neonates were carriers of abnormal hemoglobinopathy in Qianxinan State, caused by 21 genotypes. The Hb Hekinan II heterozygote was the most common mutation (32.26%), followed by the HbQ-Thailand heterozygote (17.2%). Hb Hekinan and HbQ-Thailand were not previously reported in Guizhou. In addition, we detected four rare genotypes for the first time in China: the Hb Zurich-Langstrasse hybrid, Hb Fuchu-I heterozygous, Hb Shizuoka hybrid and Hb Broomhill heterozygous types.

In addition, we identified 28 cases of intermediate α-thalassemia and 10 cases of β-thalassemia major. During the follow-up observations, three cases of β-thalassemia major presented clinical symptoms, and one of them had started to receive regular blood transfusions. For β-thalassemia major, if regular blood transfusions and iron chelation therapy are administered in a timely and correct manner, patient life expectancy may approach normal. Therefore, a precise diagnosis of β-thalassemia major during the neonatal period and early intervention at the onset of symptoms are of particular importance.

In summary, we first used NGS combined with Gap-PCR to screen 18 309 neonates for thalassemia carriers in Qianxinan State, Guizhou Province. This study established an epidemiological thalassemia database in Qianxinan State, Guizhou Province, and enriched the genetic spectrum of thalassemia. Since thalassemia-rich genotypes were highly prevalent in this region, early detection of thalassemia carriers would be meaningful for genetic counseling and prevention/treatment of thalassemia in this region. In addition, NGS-Gap-PCR provides a powerful tool for the neonate genetic testing and clinical diagnosis of thalassemia.

ACKNOWLEDGEMENTS

This study was financially supported by the Department of Science and Technology in Guizhou (No. [2019]2806).

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13923.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated.

ORCID

Jindong Chen https://orcid.org/0000-0003-4253-6850

REFERENCES

1. Martin M, Haines D. Clinical management of patients with thalassemia syndromes. *Clin J Oncol Nurs*. 2016;20:310-317.
2. Ip HW, So CC. Diagnosis and prevention of thalassemia. *Crit Rev Clin Lab Sci*. 2013;50:125-141.
3. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ*. 2008;86:480-487.
4. Jindatanmanusan P, Riolueang S, Glomglao W, et al. Diagnostic applications of newborn screening for alpha-thalassaemias, haemoglobins E and H disorders using isoelectric focusing on dry blood spots. *Ann Clin Biochem*. 2014;51:237-247.
5. Smith EC, Orkin SH. Hemoglobin genetics: recent contributions of GWAS and gene editing. *Hum Mol Genet*. 2016;25:R99-R105.
6. Huang SW, Liu XM, Li GF, Su L, Wu X, Wang RL. Spectrum of beta-thalassemia mutations in Guizhou Province, Pr China, including first observation of codon 121 (GAA>TAA) in Chinese population. *Clin Biochem*. 2013;46:1865-1868.
7. Huang SW, Xu Y, Liu XM, et al. The prevalence and spectrum of alpha-thalassemia in Guizhou province of south China. *Hemoglobin*. 2015;39:260-263.
8. Shang X, Peng Z, Ye Y, et al. Rapid targeted next-generation sequencing platform for molecular screening and clinical genotyping in subjects with hemoglobinopathies. *EBioMedicine*. 2017;23:150-159.
9. Borgio JF. Molecular nature of alpha-globin genes in the Saudi population. *Saudi Med J*. 2015;36:1271-1276.
10. Borgio JF. Impact of annotation error in alpha-globin genes on molecular diagnosis. *PloS One*. 2017;12:e0185270.
11. Zeng YT, Huang SZ. Disorders of haemoglobin in China. *J Med Genet*. 1987;24:578-583.
12. Yu F, Zhong C, Zhou Q, et al. Genetic analysis of beta -thalassemia mutations in the minority populations of Guizhou province. *Chin J Med Genet*. 2010;27:700-703.
13. Xiong F, Sun M, Zhang X, et al. Molecular epidemiological survey of haemoglobinopathies in the Guangxi Zhuang autonomous region of southern China. *Clin Genet*. 2010;78:139-148.
14. Yao H, Chen X, Lin L, et al. The spectrum of alpha- and beta-thalassemia mutations of the Li people in Hainan Province of China. *Blood Cells Mol Dis*. 2014;53:16-20.
15. He J, Song W, Yang J, et al. Next-generation sequencing improves thalassemia carrier screening among premarital adults in a high prevalence population: the Dai nationality, China. *Genet Med*. 2017;19:1022-1031.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Tan M, Bai Y, Zhang X, et al. Early genetic screening uncovered a high prevalence of thalassemia among 18 309 neonates in Guizhou, China. *Clinical Genetics*. 2021;99:704–712. https://doi.org/10.1111/cge.13923