

RESEARCH ARTICLE

Open Access

# Gene spectrum analysis of thalassemia for people residing in northern China



Zhuo Yang<sup>1†</sup>, Wenzhe Zhou<sup>3†</sup>, Quexuan Cui<sup>3†</sup>, Ling Qiu<sup>1</sup> and Bing Han<sup>2\*</sup>

## Abstract

**Background:** Southern China provinces have high incidence of thalassemia, however, sporadic cases can be found in northern China as well.

**Methods:** People resided in north China who were suspected to have thalassemia were detected mutations by gap-polymerase chain reaction (Gap-PCR) and reverse dot blot (RDB) analyses. Those with positive findings from 2012 to 2017 were further analyzed for basic clinical data and ancestral information either by medical records or by telephone follow-up or both.

**Results:** Most people enrolled in our study had no or mild symptoms. For those with positive gene findings, people originated from the north had higher percentage of  $\beta$ -thalassemia gene mutations compared with those originated from the south (72.8% vs. 62.4%,  $\chi^2 = 9.92$ ,  $P = 0.001$ ). Analysis of the individual gene distribution of people from south and north areas did not show significant difference either in  $\alpha$ -thalassemia ( $P = 0.221$ ) or  $\beta$ -thalassemia ( $P = 0.979$ ). No significant difference was found in the frequency of  $\alpha$  mutation in people living in different altitudes. However, for  $\beta$ -thalassemia, the frequency of the 6 most common mutations was significantly different in people living in different provinces with altitude below 500 m, 500–1000 m, and above 1000 m ( $\chi^2$  test,  $P < 0.05$ ).

**Conclusion:** Most of people in north China with thalassemia mutation gene were thalassemia carriers. People originated from the north had higher frequency of  $\beta$  mutation than those originated from the south, but the north people had similar individual gene mutation profile compared with south people both for  $\alpha$  and  $\beta$  mutations. People lived in different altitudes had different spectrum of  $\beta$  mutations.

**Keywords:** Thalassemia, Gene mutation, Spectrum, Altitude, China

## Background

Thalassemia is an autosomal inherited defect caused by the reduced or absent synthesis of the alpha or beta globin chains of the hemoglobin (Hb) tetramer, which leads to hereditary anemia and made it one of the most pervasive monogenic diseases worldwide [1]. The thalassemia heterozygotes have shown resistance to malaria caused by *Plasmodium falciparum* [2].

The two main types of thalassemia,  $\alpha$  and  $\beta$  thalassemia, each can be subdivided into another two forms, the  $\alpha 0$  and  $\alpha +$  thalassemia, and  $\beta 0$  and  $\beta +$  thalassemia.

The heterozygous state of  $\alpha +$  thalassemia or  $\alpha 0$  thalassemia are often clinically asymptomatic while the compound heterozygous states for  $\alpha +$  thalassemia and  $\alpha 0$  thalassemia usually results in hemoglobin H. Hemoglobin Bart's, which represents the homozygous state for  $\alpha 0$  thalassemia, leads to death in the uterus or just after birth. The heterozygous inheritance of a  $\beta$  thalassemia, which called  $\beta$  thalassemia minor, usually demonstrates asymptomatic microcellular anemia while others are silent carriers. Both  $\beta$  thalassemia major and intermediate can result from the homozygous or compound heterozygous inheritance of  $\beta$  mutations. Patients with  $\beta$  thalassemia major usually present with severe anemia in infancy and become transfusion dependent for life, whereas patients with  $\beta$  thalassemia intermediate

\* Correspondence: hanbing\_li@sina.com

<sup>†</sup>Yang Zhuo, Zhou Wenzhe and Cui Quexuan contributed equally to this work.

<sup>2</sup>Department of hematology, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing, China

Full list of author information is available at the end of the article



may develop mild to moderate anemia and variable blood transfusion requirements [3].

The high frequency of inherited hemoglobin variants is present in the area extending from sub-Saharan Africa to the Indian subcontinent and East and Southeast Asia, especially the Mediterranean region [4]. In China, southern provinces, such as Guangxi, Guangdong, and Fujian, are known to be high incidence areas of the disease [5]. Thalassemia is an endemic disease which is mainly found in South China. The incidence in northern China is extremely low, and the sample size of the previous studies in the north was all less than 100.

Very few reports focused on thalassemia mutation detection and their clinical relevance in north China previously due to the rarity [6]. Because of continued migration, these diseases are now becoming increasingly common in international metropolitan cities such as Beijing. Furthermore, there indeed exist sporadic clinical cases with pure northern ancestry within three generations.

This study, through the analysis of data from 1059 people who lived in the north part of China with thalassemia mutation genes, including detailed family histories and gene spectrum, aims to trace the precise geographic origins of the various alleles identified in north China, give a geographical distribution of mutations of people resident in north China.

## Methods

### Probands

Among a total of 2136 people who came to Peking Union Medical College Hospital (PUMCH) from 2012 to 2017 to screen for thalassemia mutations, 1059 (299 males and 760 females) people had at least one positive finding (any mutations including heterozygosity or homozygosity) either for  $\alpha$ - or  $\beta$  mutation or both. PUMCH is a fully equipped hematology clinic and the only center for the detection of thalassemia mutation gene in the north China, so our data were considered representative for the northern China. The majority of people with positive findings originated from the 15 provinces in southern China; while the rest had ancestral home sporadically distributed in the 15 north provinces. The confirmed diagnosis was based on 'Criteria for diagnosis and treatment of hematologic diseases' and 'Guidelines for the prevention and control of thalassemia'. [1]

### Analysis of mutations

Samples were collected by standard methods: 3 tubes of fresh venous blood (2 ml each), anticoagulated by EDTA-K2, were used for gene detection, hemoglobin electrophoresis, and blood cell analysis, respectively. Genomic DNA was isolated from the blood samples by

the DNA rapid extraction kit (QIamp DNA mini blood kit, QIAGEN, Hilden, Germany).

### Gap-polymerase chain reaction (GAP-PCR)

The most common 3 large  $\alpha$ -globin deletions (NG\_000006.1:g.34164\_37967del3804, - 4.2, NG\_000006.1:g.26264\_45564del19301) which had been reported in Chinese people previously were detected by the gap-PCR method [7, 16]. PCR reaction system: total volume was 25  $\mu$ l, containing 21  $\mu$ l of reaction liquid, 2  $\mu$ l of DNA template and 2  $\mu$ l of pure H<sub>2</sub>O.

The reaction condition was 96 °C for 5 min, 98 °C for 45 s, 65 °C for 90 s, 72 °C for 3 min for 10 cycles; 98 °C for 30 s, 65 °C for 45 s, 72 °C for 3 min for 25 cycles; 72 °C for 10 min. Samples were preserved at 4 °C.

### Reverse dot blot (RDB) analysis

RDB analysis was employed for the detection of seventeen Chinese most common mutations in  $\beta$ -globin gene, including HBB:c.126\_129delCTTT, HBB:c.316-197C > T, HBB:c.-78A > G, HBB:c.216\_217insA, HBB:c.52A > T, HBB:c.79G > T, HBB:c.94delC, HBB:c.84\_85insC, IVS-I-1 (G > T), HBB:c.92 + 1G > T, HBB:c.130G > T, HBB:c.-82C > A, HBB:c.-79A > G, HBB:c.-80 T > C, HBB:c.45\_46insG, HBB:c.-50A > C, and HBB:c.92 + 5G > C, according to the instructions [8].

The reaction system and condition for the detection of the 3 common  $\alpha$ -globin point mutations (HBA2:c.427 T > C, HBA2:c.369C > G, and HBA2:c.377 T > C, non deletion mutation) was similar to that of the  $\beta$ -globin genes. PCR reaction system was as follow: total volume was 25  $\mu$ l, containing 23  $\mu$ l of PCR reaction liquid, and 2  $\mu$ l of DNA template. The reaction condition was 96 °C for 5 min, 98 °C for 45 s, 65 °C for 90 s, 72 °C for 3 min by 10 cycles; 98 °C for 30 s, 65 °C for 45 s, 72 °C for 3 min for 25 cycles; 72 °C for 10 min. Samples were preserved at 4 °C. The genotype was determined by hybrid membrane spot color characteristics of PCR products.

### Ancestral home statistics

People with no searchable identity cards or follow-up information were excluded from the study. Using hospital information system, people were first classified by identity card numbers, then telephone follow-up survey was explored to confirm their ancestral information, including a household registration and immigration history of their ancestors within three generations of the family, to recognize whether these northern inhabitants had southern origin.

### Data analysis

The frequency of 6  $\alpha$ - and 14  $\beta$ -thalassemia mutations which had been shown to be the most common mutations

in China was summarized for all the 1059 patients, and people carrying the mutations were collected with their medical data and basic laboratory parameters, and further distinguished by their ancestor origin to see the difference between north and south. People from areas with different average altitude: below 500 m, 500 to 1000 m and above 1000 m were further studied for the difference of the gene distribution.

SPSS 24.0 software (IBM, NY, USA) was used for statistical analysis. The ratio of  $\alpha$ - and  $\beta$ -thalassemia alleles was calculated, and data were analyzed by Fisher's exact test and Pearson's chi-squared test.  $P < 0.05$  was considered statistically significant.

**Results**

**Demographic data**

All subjects investigated here were north dwellers who had been resided in the northern China for more than 3 years. There were 1059 patients involved in this study, including 359  $\alpha$ -thalassemia carriers, 683  $\beta$ -thalassemia carriers, and 17  $\alpha$ - and  $\beta$ -thalassemia carriers at the same time. The mean age of these patients was  $30.2 \pm 13.9$  years ranging from 0 to 82 years. Their age and sex distribution exhibited relatively nonhomogenous that most patients were aged between 21 and 40 years (79.9%), and more than two-third of carriers were females (71.8%), probably due to the greater likelihood of hypochromic anemia in this particular population. Although thalassemia is rare in north part of China, there were 42  $\alpha$ -thalassemia (9.5%) and 140  $\beta$ -thalassemia (15.6%) carriers with pure northern descent within three generations in our study, spreading all over the north provinces (Table 1).

The  $\alpha$  mutations accounted for 27.2% of the all north originated (northerners) thalassemia genes and  $\beta$  mutations accounted for the rest 72.8%, while the percentage of  $\alpha$  and  $\beta$  mutations in south originate people (southerners) were 37.6 and 62.4%, respectively. Northerners seemed to have higher percentage of  $\beta$  mutations and the south descendants had higher frequency of  $\alpha$  mutation ( $\chi^2 = 6.76, P < 0.05$ ).

Not surprisingly, the majority of people with thalassemia mutated genes in our study can be defined as 'carriers' which had no symptoms, and the rest were 'patients' with mild clinical manifestation. More than 95% people in our study had mild symptoms, including mild anemia (Hb: 9-11 g/dl) and normal or moderately elevated serum ferritin level ( $< 500 \mu\text{g/L}$ ). However, there were still some people with severe anemia, elevated total bilirubin, enlarged spleen size, accompanied with iron overload and other complications (less than 5%). Totally, 18  $\beta$ -thalassemia cases (3 with pure north origin) were diagnosed as  $\beta$ -thalassemia intermediate (Hb: $< 9 \text{ g/dl}$ ), with 4 of them transfusion-dependent (TDT). No  $\beta$ -thalassemia major, hemoglobin H and hemoglobin Bart's were found in our research.

**Mutation analysis**

The techniques used for mutation detection in our study could cover more than 99.9% of the reported  $\alpha$  and  $\beta$  globin genes in the Chinese population. Nearly 10% of the carriers with  $\alpha$  mutations and 20% of the carriers with  $\beta$  mutations came from the north part of China including Beijing, Tianjin, Hebei, Shanxi, Shandong, Henan, Inner Mongolia, Shaanxi, Gansu, Xinjiang, Liaoning, Jilin, and Heilongjiang provinces, while the other carriers came

**Table 1** Demographic data of the patients enrolled

Demographic Data	N(%)	Individuals only with $\alpha$ -thalassemia	Individuals only with $\beta$ -thalassemia	Individuals combined $\alpha$ -/ $\beta$ -thalassemia carriers
Total	1059	359	683	17
Sex distribution				
males	299 (28.2)	116 (32.3)	177 (25.9)	6 (35.3)
females	760 (71.8)	243 (67.7)	506 (74.1)	11 (64.7)
Age distribution				
0-10	89 (8.4)	28 (7.8)	61 (8.9)	-
11-20	40 (3.8)	12 (3.3)	28 (4.1)	9 (52.9)
21-30	443 (41.8)	154 (42.9)	280 (41.0)	7 (41.2)
31-40	372 (35.1)	133 (37.0)	232 (34.0)	1 (5.9)
41-50	56 (5.3)	16 (4.5)	39 (5.7)	-
>50	59 (5.6)	16 (4.5)	43 (6.3)	-
Geographic distribution				
Northern region of China	183 (17.3)	42 (11.7)	140 (20.5)	1 (5.9)
Southern region of China	876 (82.7)	317 (88.3)	543 (79.5)	16 (94.1)

Table Legend: N = number of patients; % = number of events/total number of patients in column\*100

from the south parts of the country including Hubei and Hunan, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian and Jiangxi, Chongqing, Sichuan, Guizhou, Yunnan, Guangdong, Guangxi and Hainan provinces (Table 2).

The -NG\_000006.1:g.26264\_45564del19301 and HBB:c.316-197C > T mutations were the most common  $\alpha$  and  $\beta$  mutations in our study, with a frequency of 74.8 and 36.8%, respectively. The other common  $\alpha$  mutations were NG\_000006.1:g.34164\_37967del3804 (16.1%) and - $\alpha$ 4.2 deletion (5.4%), and the three mutations accounted for 96.3% of all  $\alpha$ -thalassemia genes analyzed. Other common  $\beta$  mutations were HBB:c.316-197C > T (36.8%), HBB:c.126\_129delCTTT (28.3%), HBB:c.52A > T (25.3%), HBB:c.-78A > G (3.0%) and HBB:c.216\_217insA (2.8%). All these mutations accounted for 96.0% of all the  $\beta$ -thalassemia genes analyzed in this study.

Detail analysis of the spectrum of  $\alpha$ -thalassemia mutations in different regions showed that there were no significant difference in the percentage of different  $\alpha$ -thalassemia mutations between people from north and south ( $P = 0.275$ ) (Fig. 1).

Similarly, no difference was found in the percentage of different  $\beta$  thalassemia mutations between north and south regions, either ( $P = 0.661$ , Fig. 1).

**Spectrum in different altitudes**

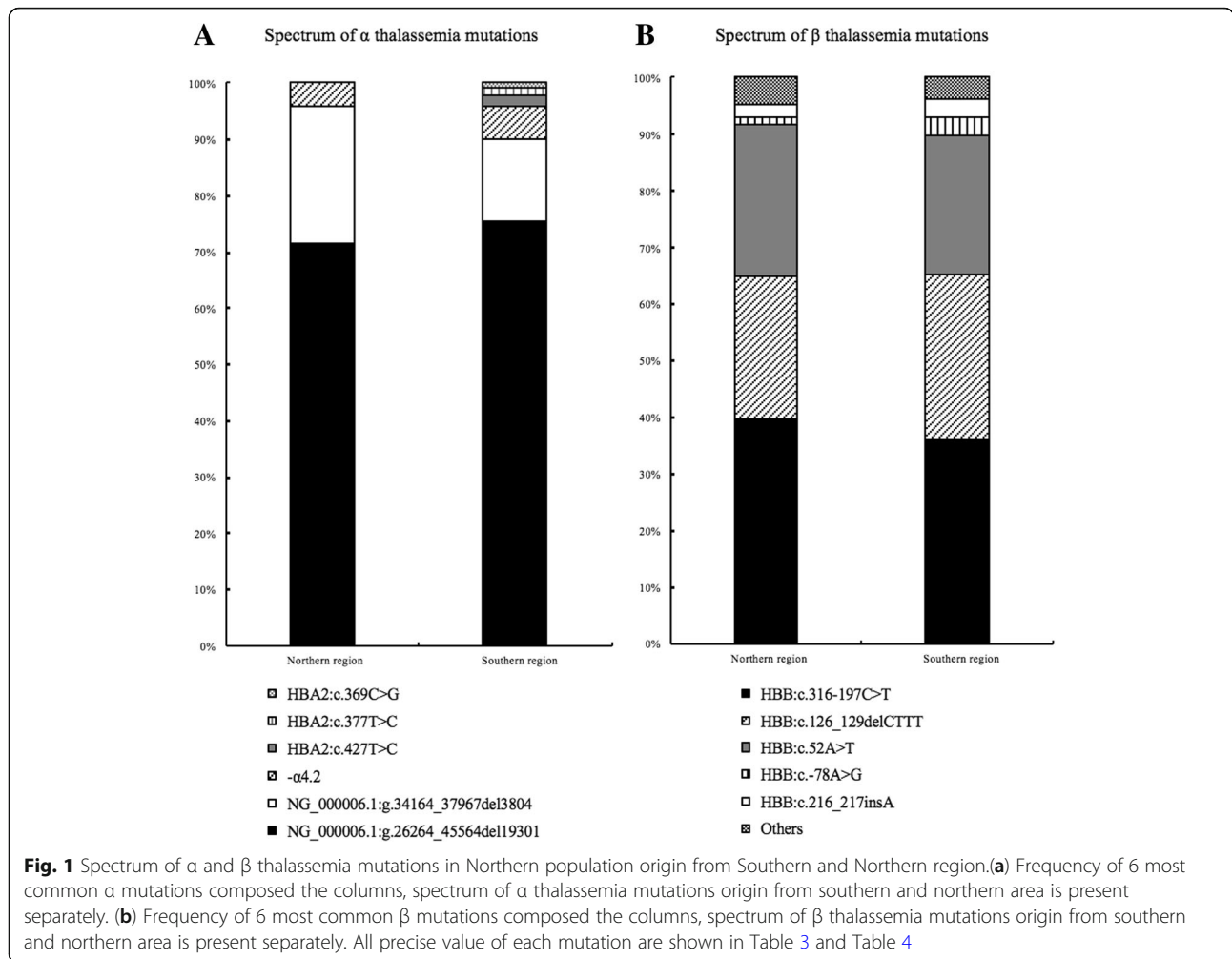
As shown in Tables 3 and 4, the percentage of different  $\alpha$ -thalassemia gene were evenly distributed in people lived in different level of altitude; however, detail analysis of the percentage of different  $\beta$ -thalassemia genes showed a remarkable difference in geographical areas with different elevations.

HBB:c.316-197C > T, HBB:c.126\_129delCTTT and HBB:c.52A > T were the top three frequent  $\beta$ -thalassemia mutations and their percentage was significantly different in various provinces. Higher percentage of certain gene seemed to concentrate on the same altitude level although different in provinces. For instance, the mutation HBB:c.316-197C > T was mainly distributed in Campagna and eastern coastal areas, such as Shanghai (100%), Shandong (72.7%), and Fujian (51.7%). The mutation HBB:c.126\_129delCTTT was

**Table 2** Frequency and distribution of  $\alpha$  and  $\beta$  thalassemia mutations

Mutation	Type	Northern region		Southern region		Overall distribution in China	
		N	%	N	%	N	%
$\alpha$ -thalassemia							
Total		49	100.0%	356	100.0%	405	100.0%
NG_000006.1:g.26264_45564del19301	$\alpha 0$	35	71.4%	268	75.2%	303	74.8%
NG_000006.1:g.34164_37967del3804	$\alpha +$	12	24.5%	53	14.9%	65	16.1%
- $\alpha 4.2$	$\alpha +$	2	4.1%	20	5.6%	22	5.4%
HBA2:c.427T > C	$\alpha +$	-	-	7	2.0%	7	1.7%
HBA2:c.377T > C	$\alpha +$	-	-	5	1.4%	5	1.2%
HBA2:c.369C > G	$\alpha +$	-	-	3	0.9%	3	0.7%
$\chi^2$ test		$\chi^2 = 2.582, p = 0.275$					
$\beta$ -thalassemia							
Total		131	100.0%	541	100.0%	672	100.0%
HBB:c.316-197C > T	$\beta 0$	52	39.7%	195	36.0%	247	36.8%
HBB:c.126_129 delCTTT	$\beta 0$	33	25.2%	157	29.0%	190	28.3%
HBB:c.52A > T	$\beta 0$	35	26.7%	134	24.8%	169	25.3%
HBB:c.-78A > G	$\beta 0$	2	1.5%	18	3.3%	20	3.0%
HBB:c.216_217insA	$\beta +$	3	2.3%	16	3.0%	19	2.8%
HBB:c.84_85insC	$\beta +$	1	0.8%	6	1.1%	7	1.0%
HBB:c.130G > T	$\beta 0$	1	0.8%	6	1.1%	7	1.0%
HBB:c.79G > T	$\beta E$	2	1.5%	4	0.7%	6	0.9%
HBB:c.-79A > G	$\beta 0$	1	0.8%	3	0.6%	4	0.6%
HBB:c.45_46insG	$\beta +$	-	-	1	0.2%	1	0.1%
HBB:c.92 + 1G > T	$\beta 0$	-	-	1	0.2%	1	0.1%
HBB:c.92 + 5G > C	$\beta +$	1	0.8%	-	-	1	0.1%
$\chi^2$ test		$\chi^2 = 2.409, p = 0.661$					

Table Legend: N = number of patients, % number of events/total number of patients\*100; p-value for the  $\chi^2$  test are performed. Significant differences ( $p < 0.05$ ) are highlighted in grey



mainly found in provinces that pass through the mountain range, such as Heilongjiang (Changbai Mountains), Shaanxi (Qin Mountains), Chongqing (Wu mountain), Guizhou (Yungui Plateau), Guangxi (South of the Five Ridges), and Hainan (Wuzhi Mountain), and all these provinces had a frequency of higher than 40%. On the other hand, mutation HBB:c.52A > T was more common in higher altitude areas, especially in Xinjiang (100.0%), Gansu (75.0%), and Sichuan (47.3%).

We found that the gene distribution in provinces with similar average altitudes was similar. We then classified the provinces according to their average altitude, 15 provinces were below 500 m, 8 were about 500–1000 m, and 5 were above 1000 m. The frequency of HBB:c.316-197C > T was gradually declined from the altitude below 500 m, to above 1000 m, and to about 500–1000 m. While the HBB:c.126\_129delCTTT was more common in the provinces with elevation between 500 and 1000 m. And the frequency of HBB:c.52A > T was gradually increased from plain area to plateau area. The number of genes in these divided regions were summarized and listed in Table 5.

### Discussion

Although the majority of the people in our study were descendants of southerners, there were indeed some people who were ‘original’ northerners and this is the report including the largest cases from the northerners. The purpose of this study was to compare the discrepancy of distribution of thalassemia gene between low incidence areas (north) and ‘hot spot’ (south) areas. To this aim, people with the 6 most common  $\alpha$  mutations and 17  $\beta$  mutations detected by Gap-PCR or RDB analysis and with traceable clinical data were studied.

It is noteworthy that the south originated people had higher frequency of  $\alpha$  mutation (37.6%) than the north ones (27.2%,  $P < 0.05$ ) in our study. This finding is consistent with the former literatures [9–16], showing that  $\beta$ -thalassemia is more common in China, either in the South or the North [18–20]. Our study gathered the largest thalassemia mutation cohort in the north, and demonstrated that North origin people seemed to have even higher percentage of  $\beta$ -thalassemia. Reason for the bias of different thalassemia distribution is not clear. Clinical

**Table 3** Frequency and distribution of  $\alpha$  and  $\beta$  thalassaemia mutations in Northern Chinese provinces

Mutation	Type	Beijing	Tianjin	Hebei	Shanxi	Inner Mongolia	Shandong	Henan	Shaanxi	Gansu	Xinjiang	Liaoning	Jilin	Heilong jiang
<b><math>\alpha</math>-thalassaemia</b>														
N (total)		15	1	10	-	6	5	6	1	-	-	2	1	2
NG_000006.1:g.26264_45564del19301	$\alpha 0$	73.3%	100.0%	60.0%	-	66.7%	80.0%	83.3%	100.0%	-	-	50.0%	-	100.0%
NG_000006.1:g.34164_37967del3804	$\alpha +$	20.0%	-	40.0%	-	33.3%	20.0%	16.7%	-	-	-	-	100.0%	-
$-\alpha^{4,2}$	$\alpha +$	6.7%	-	-	-	-	-	-	-	-	-	50.0%	-	-
<b><math>\beta</math>-thalassaemia</b>														
N (total)		36	6	26	10	3	11	12	9	4	1	6	4	3
HBB:c.316-197 C > T	$\beta 0$	36.1%	33.3%	38.5%	70.0%	33.3%	72.7%	50.0%	22.2%	-	-	16.7%	50.0%	-
HBB:c.126_129 delCTTT	$\beta 0$	36.1%	33.3%	23.1%	10.0%	-	9.1%	16.7%	44.4%	25.0%	-	16.7%	-	66.7%
HBB:c.52A > T	$\beta 0$	22.2%	16.7%	30.8%	20.0%	33.33%	9.1%	25.0%	22.2%	75.0%	100.0%	50.0%	50.0%	-
HBB:c.-78A > G	$\beta 0$	-	16.7%	-	-	-	9.1%	-	-	-	-	-	-	-
HBB:c.216_217 insA	$\beta +$	2.8%	-	-	-	33.3%	-	8.3%	-	-	-	-	-	-
HBB:c.84_85insC	$\beta +$	-	-	3.8%	-	-	-	-	-	-	-	-	-	-
HBB:c.130G > T	$\beta 0$	-	-	-	-	-	-	-	-	-	-	-	-	33.3%
HBB:c.79G > T	$\beta E$	2.8%	-	-	-	-	-	-	11.1%	-	-	-	-	-
HBB:c.-79A > G	$\beta 0$	-	-	-	-	-	-	-	-	-	-	16.7%	-	-
HBB:c.92 + 5G > C	$\beta +$	-	-	3.8%	-	-	-	-	-	-	-	-	-	-

Table Legend: N Number of patients; % number of events/total number of patients\*100

**Table 4** Frequency and distribution of  $\alpha$  and  $\beta$  thalassaemia mutations in Southern Chinese provinces during 2012–2017

Mutation	Type	Shang hai	Zhe jiang	Jiangsu	Anhui	Fujian	Jiangxi	Hubei	Hunan	Chong qing	Sichuan	Guizhou	Yunnan	Guang dong	Guangxi	Hainan
<b><math>\alpha</math>-thalassaemia</b>																
N (total)		3	14	6	12	40	45	26	32	4	43	6	4	33	83	5
NG_000006.1:g.26264_45564del19301	$\alpha 0$	66.7%	57.1%	83.3%	75.0%	70.0%	84.4%	76.9%	81.3%	75.0%	76.7%	66.7%	75.0%	84.9%	69.9%	60.0%
NG_000006.1:g.34164_37967del3804	$\alpha +$	33.3%	28.6%	-	16.7%	22.5%	11.1%	11.5%	12.5%	-	11.6%	16.7%	25.0%	9.1%	15.7%	40.0%
- $\alpha 4.2$	$\alpha +$	-	-	-	8.3%	-	2.2%	7.7%	3.1%	25.0%	11.6%	-	-	3.0%	9.6%	-
HBA2:c.427 T > C	$\alpha +$	-	-	-	-	-	2.2%	-	3.1%	-	-	16.7%	-	3.0%	3.6%	-
HBA2:c.377 T > C	$\alpha +$	-	-	16.67%	-	5.0%	-	3.9%	-	-	-	-	-	-	1.2%	-
HBA2:c.369 C > G	$\alpha +$	-	14.3%	-	-	2.5%	-	-	-	-	-	-	-	-	-	-
<b><math>\beta</math>-thalassaemia</b>																
N (total)		3	23	12	18	29	45	71	79	15	112	18	6	35	66	9
HBB:c.316-197 C > T	$\beta 0$	100.0%	39.1%	16.7%	44.4%	51.7%	57.8%	56.3%	41.8%	40.0%	28.6%	22.2%	33.3%	25.7%	6.1%	22.2%
HBB:c.126_129 delCTTT	$\beta 0$	-	21.7%	25.0%	16.7%	34.5%	26.7%	25.4%	22.8%	46.7%	20.5%	50.0%	16.7%	42.9%	42.4%	55.6%
HBB:c.52A > T	$\beta 0$	-	26.1%	33.3%	33.3%	10.3%	4.4%	11.3%	16.5%	13.3%	47.3%	27.8%	33.3%	17.1%	36.4%	-
HBB:c.-78A > G	$\beta 0$	-	8.7%	-	-	-	6.7%	2.8%	1.3%	-	0.9%	-	-	8.6%	9.1%	-
HBB:c.216_217 insA	$\beta +$	-	-	25.0%	-	-	-	1.4%	7.6%	-	0.9%	-	-	-	6.1%	11.1%
HBB:c.84_85insC	$\beta +$	-	4.4%	-	5.6%	3.5%	2.2%	1.4%	-	-	0.9%	-	-	-	-	-
HBB:c.130G > T	$\beta 0$	-	-	-	-	-	2.2%	1.4%	3.8%	-	0.9%	-	-	-	-	-
HBB:c.79G > T	$\beta E$	-	-	-	-	-	-	-	2.5%	-	-	-	16.7%	-	-	11.1%
HBB:c.-79A > G	$\beta 0$	-	-	-	-	-	-	-	1.3%	-	-	-	-	5.7%	-	-
HBB:c.45_46insG	$\beta +$	-	-	-	-	-	-	-	1.3%	-	-	-	-	-	-	-
HBB:c.92 + 5G > C	$\beta 0$	-	-	-	-	-	-	-	1.3%	-	-	-	-	-	-	-

Table Legend: N Number of patients, % Number of events/total number of patients\*100

**Table 5** Frequency and distribution of  $\alpha$  and  $\beta$  thalassemia mutations according to different elevation

Mutation	Type	Plain area		Mountain range		Continental plateau		Overall distribution in China	
		N	%	N	%	N	%	N	%
<b><math>\alpha</math>-thalassemia</b>									
Total		236	100.0%	116	100.0%	53	100.0%	405	100.0%
NG_000006.1:g.26264_45564del19301	$\alpha 0$	181	75.1%	82	74.0%	40	75.5%	303	74.8%
NG_000006.1:g.34164_37967del3804	$\alpha +$	38	17.8%	19	13.4%	8	15.1%	65	16.1%
- $\alpha 4.2$	$\alpha +$	7	3.1%	10	7.9%	5	9.4%	22	5.4%
HBA2:c.427 T > C	$\alpha +$	3	0.9%	4	3.9%	–	–	7	1.7%
HBA2:c.377 T > C	$\alpha +$	4	1.8%	1	0.8%	–	–	5	1.2%
HBA2:c.369 C > G	$\alpha +$	3	1.3%	–	–	–	–	3	0.7%
$\chi^2$ test		$\chi^2 = 5.53,$		$P = 0.063$		$\chi^2 = 0.11,$		$P = 0.947$	
				$\chi^2 = 4.20,$		$P = 0.122$			
<b><math>\beta</math>-thalassemia</b>									
Total		380	100.0%	166	100.0%	126	100.0%	672	100.0%
HBB:c.316-197C > T	$\beta 0$	174	45.8%	38	22.9%	35	27.8%	247	36.8%
HBB:c.126_129 delCTTT	$\beta 0$	96	25.3%	69	41.6%	25	19.8%	190	28.3%
HBB:c.52A > T	$\beta 0$	66	17.4%	43	25.9%	60	47.6%	169	25.2%
HBB:c.-78A > G	$\beta 0$	13	3.4%	6	3.6%	1	0.8%	20	3.0%
HBB:c.216_217insA	$\beta +$	11	2.9%	6	3.6%	2	1.6%	19	2.8%
HBB:c.84_85insC	$\beta +$	6	1.6%	–	–	1	0.8%	7	1.0%
HBB:c.130G > T	$\beta 0$	5	1.3%	1	0.6%	1	0.8%	7	1.0%
HBB:c.79G > T	$\beta E$	2	0.5%	3	1.8%	1	0.8%	6	0.9%
HBB:c.-79A > G	$\beta 0$	4	1.1%	–	–	–	–	4	0.6%
HBB:c.45_46insG	$\beta +$	1	0.3%	–	–	–	–	1	0.1%
HBB:c.92 + 1G > T	$\beta 0$	1	0.3%	–	–	–	–	1	0.1%
HBB:c.92 + 5G > C	$\beta +$	1	0.3%	–	–	–	–	1	0.1%
$\chi^2$ test		$\chi^2 = 29.74,$		$P < 0.001$		$\chi^2 = 24.21,$		$P < 0.001$	
				$\chi^2 = 45.59,$		$P < 0.001$			

Table Legend: N Number of patients, % number of events/total number of patients\*100, p-value for the  $\chi^2$  test, plain area: provinces with elevation below 500 m; mountain range: mountain region with an altitude about 500–1000 m; Continental plateau: provinces with altitude above 1000 m

analysis showed that most of the people had no symptoms or with mild anemia. No symptomatic  $\alpha$ -thalassemia or  $\beta$ -thalassemia major was found in our study. Since most of people carrying thalassemia mutations in north China have no or mild symptoms, they were suspected to have thalassemia only because they had microcytic anemia or sometimes, low mean corpuscular volume (MCV), abnormal erythrocyte morphology, or electrophoresis or for the differential diagnosis. Silent  $\alpha$ -thalassemia may not have low mean corpuscular volume (MCV) and can be ignored during regular medical check. Of course, the low incidence of thalassemia mutation in northerners may also contribute to the deviation. Therefore, data in our research cannot represent the true prevalence of  $\alpha$ - or  $\beta$ -thalassemia [16], Even though,

it is a cumulative data of five years from the largest thalassemia gene detection center, it may reflect the true situation of “detectable” thalassemia in the North China to some extent. Considering the harmlessness of those silence carriers, it is important to understand the general distribution of thalassemia and their clinical relevance, as shown in our study.

As for the percentage of individual thalassemia gene, there was no significant difference between people from north and south China, both for  $\alpha$ - and  $\beta$ -thalassemia, showing the ethnic coherence of North and South China.

Although mutations of  $\alpha$ -thalassemia were evenly distributed in different provinces, with NG\_000006.1:g.26264\_45564del19301 the highest one, mutations of



$\beta$ -thalassemia showed significant difference in the different geographical regions. After trying to classify the regions in different ways, we finally found that the altitude of provinces accurately reflected the distribution of  $\beta$ -thalassemia phenotype. The frequency of HBB:c.316-197C > T was relatively high in the provinces below 500 m, HBB:c.126\_129delCTTT centralized in the provinces with elevation between 500 and 1000 m, and HBB:c.52A > T mutation gradually increased from plain area to plateau area.

This is the first report of this interesting phenomenon, and we did not find any literature to verify it so far. Although there are numerous reports about gene distribution in different provinces in south China, but the studying objects came from the limit provinces, which mostly located in the southern part of China with a relatively low altitude.

All genetic mutations of thalassemia may have been influenced by natural selection [17]. It has been shown that malaria is a strongly selective factor for many genotypes (e.g., G6PD deficiency, thalassemia, ABO, Rh, MN, Duffy, secretory types (Ss), human leukocyte antigens (HLA), etc.) [21] Malaria may have variations from different altitudes probably due to change of distribution of mosquitoes across an altitudinal gradient [20]. However, few studies have focused on the malaria transition in different altitudes in China. Therefore, the change of gene pattern of  $\beta$ -thalassemia in different altitudes may give a clue for the malaria transition in the history of China [22, 23].

Race difference may also play a significant role in gene difference, and we couldn't exclude racial diversity at different altitudes. But in this study, we had checked that all the people detected in our study were Han Chinese, so the difference from ethnic diversity was small.

The reasons for thalassemia gene diversity are very complicated. One report has shown that even the frequency of PC allele for acid phosphatase in fourteen Sardinian villages positively correlates with the altitude and negatively with past malarial morbidity. Thus, thalassemia trait exerts a protective action only in subjects carrying PA allele for acid phosphatase [24]. In addition to malaria, other environmental factors related to altitude may also play a role in shaping the present pattern of distribution of thalassemia which need to be further investigated. There are also some reports pointed out that some thalassemia mutations in northern China may be a result of genetic founder effects, which need further research for verification and falsification [25, 26].

For the first time, we reported the major thalassemia gene profile of people who resided in the north part of China, where thalassemia was considered only sporadic onset. Although this study needs to be verified in larger cohort, we described the possibility of relationship between altitude and pattern of  $\beta$ -thalassemia. Future

whole genome sequencing studies which may better define the genetic polymorphisms, together with detail analysis of the change related to altitude which may cause the shift of thalassemia gene are worth doing. These findings will enrich our understanding of etiology and mechanism of thalassemia in China.

## Conclusions

This study demonstrated a geographical distribution of mutations of people resident in north China, which showed that most of people in north China with thalassemia mutation gene were thalassemia carriers and indicated that in both north and south China, the spectrum of  $\alpha$  and  $\beta$  mutations may have no significant difference. Furthermore, this research presented that people originate from regions with different level of altitudes may have different spectrum of  $\beta$  mutations, which broadened our understanding of etiology and mechanism of thalassemia in China.

## Acknowledgements

The authors would like to appreciate all patients who consented to disclose their medical records and answered our review calls. Their cooperation and helpful comments throw light upon our data collection and statistical analysis.

## Availability of data and materials

The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

## Funding

This study was supported by grants from the Chinese Academy of Medical Sciences innovation fund for medical sciences (Grant No. 2016-I2M-3-004), Beijing Natural Science of Foundation (7192168) and MANDARIN study from Novartis (Protocol Number: C1CL670AAU05), they both supported the collection of the careful history and follow-up of the subjects.

## Authors' contributions

HB conceived the idea of the study; YZ, ZWZ, CQX, and QL collected the data; Z WZ and CQX analysed the data; ZWZ interpreted the results, depicted the figures and wrote the paper; all authors discussed the results and revised the manuscript.

## Ethics approval and consent to participate

Our study was approved by the medical ethics committee of the Peking Union Medical College, registration number: S-K438. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## Consent for publication

Based on the Declaration of Helsinki, verbal informed consent was obtained from all patients included in the trial, parental consent was obtained for participants below 18 years old. The ethics committee approved this procedure. This article does not contain any studies with animals performed by any of the authors.

## Competing interests

The authors declare that they have no conflicts of interest.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Department of clinical laboratory, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing, China. <sup>2</sup>Department of hematology, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing, China. <sup>3</sup>Peking Union Medical College, Chinese Academy of Medical Science, Beijing, China.

Received: 25 October 2018 Accepted: 30 April 2019

Published online: 22 May 2019

**References**

- Muncie HL, James C. Alpha and beta thalassemia.[J]. *Rass Clin Sci*. 2009; 80(4):333–6.
- Hedrick PW. Population genetics of malaria resistance in humans. *Heredity*. 2011;107(4):283–304.
- Taher AT, Musallam KM, Cappellini MD, et al. Thalassaemia. *Lancet*. 2017;17: 31822–6.
- Weatherall DJ, Williams TN, Allen SJ, et al. The population genetics and dynamics of the thalassemsias. *Hematol Oncol Clin North Am*. 2010;24(6): 1021–31.
- Cao A, Galanello R B-t. *Genetics in Medicine Official Journal of the American College of Medical Genetics*. 2010;12(2):61–76.
- Yang Z, Han B, Gan Y, et al. Genetic analysis of intermediate severe and compound thalassemia in Peking union medical college hospital from 2012 to 2015 [J]. *Chin J lab med*. 2016;39(7):491–5.
- Lin M, Zhong TY, Chen YG, et al. Molecular epidemiological characterization and health burden of thalassemia in Jiangxi Province, P. R. China. *PLoS One*. 2014;9(7):e101505.
- Giardine B, Borg J, Viennas E, Pavlidis C, Moradkhani K, Joly P, Bartsakoulia M, Riemer C, Miller W, Tzimas G, Wajcman H, Hardison RC, Patrinos GP. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. *Nucleic Acids Res*. 2014 Jan;42:D1063–9.
- Wang S, Zhang R, Xiang G, et al. Mutation screening for thalassaemia in the Jino ethnic minority population of Yunnan Province, Southwest China. *BMJ Open*. 2015;5(12):e010047.
- Todd D. Prevalence and genotypes of  $\alpha$ - and  $\beta$ -thalassemia carriers in Hong Kong — implications for population screening. *N Engl J Med*. 1997;336(18): 1298–301.
- Yao XY, Yu J, Chen SP, et al. Prevalence and genetic analysis of  $\alpha$ -thalassemia and  $\beta$ -thalassemia in Chongqing area of China. *Gene*. 2013; 532(1):120–4.
- Li B, Zhang XZ, Yin AH, et al. High prevalence of thalassemia in migrant populations in Guangdong Province, China. *BMC Public Health*. 2014;14(1):1–8.
- Tang W, Zhang C, Lu F, et al. Spectrum of  $\alpha$ -thalassemia and  $\beta$ -thalassemia mutations in the Guilin region of southern China. *Clin Biochem*. 2015;48(16–17):1068.
- Huang SW, Xu Y, Liu XM, et al. The prevalence and Spectrum of  $\alpha$ -thalassemia in Guizhou Province of South China. *Hemoglobin*. 2015;39(4):260–3.
- Xia Y, Li-Ye Y, Hui-Tian Y, et al. Molecular epidemiological investigation of thalassemia in the Chengdu region, Sichuan Province, Southwest China. *Hemoglobin*. 2015;39(6):393–7.
- Lai K, Huang G, Li S, et al. The prevalence of thalassemia in mainland China: evidence from epidemiological surveys. *Sci Rep*. 2017;7(1):920.
- Voskarides K. Combination of 247 genome-wide association studies reveals high Cancer risk as a result of evolutionary adaptation[J]. *Mol Biol Evol*. 2017.
- Li CK. New trend in the epidemiology of thalassaemia. *Best Practice & Research Clinical Obstetrics & Gynaecology*. 2017;39:16–26.
- Gorello P, Arcioni F, Palmieri A, et al. The molecular Spectrum of  $\beta$ - and  $\alpha$ -thalassemia mutations in non-endemic Umbria, Central Italy. *Hemoglobin*. 2016;40(6):371–6.
- Mahdieh N, Rabbani B. Beta thalassemia in 31,734 cases with HBB, gene mutations: pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East. *Blood Rev*. 2016;30(6):493–508.
- Tzoneva M, Bulanov AG, Mavrudieva M, et al. Frequency of glucose-6-phosphate dehydrogenase deficiency in relation to altitude: a malaria hypothesis.[J]. *Bull World Health Organ*. 1980;58(4):659.
- Asigau S, Hartman DA, Higashiguchi JM, et al. The distribution of mosquitoes across an altitudinal gradient in the Galapagos Islands. *Journal of Vector Ecology Journal of the Society for Vector Ecology*. 2017;42(2):243.
- Palmarino R, Agostino R, Gloria F, et al. Red cell acid phosphatase: another polymorphism correlated with malaria. *Am J Phys Anthropol*. 1975;43(2): 177–85.
- Zeil G1, Lisa A, Astolfi P. Fertility and malaria in Sardinia. *Ann Hum Biol* 1990; 17(4):315–330.
- Voskarides K, Patsias C, Pierides A, Deltas C. COL4A3 founder mutations in Greek-Cypriot families with thin basement membrane nephropathy and focal segmental glomerulosclerosis dating from around 18th century. *Genet Test*. 2008 Jun;12(2):273–8.
- Marchi N, Pissard S, Cliquennois M, Vasseur C, Le Metayer N, Mureau C, Jouet JP, Georget AF, Genin E, Rose C. Confirmation of a founder effect in a northern European population of a new  $\beta$ -globin variant: HBB:c.23\_26dup (codons 8/9 (+AGAA)). *Eur J Hum Genet*. 2015 Sep;23(9):1158–64.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

