

# Prevalence of Thalassemia-carrier of Couples at Childbearing Age and Risk Prediction of Thalassemia

**Liu-song Wu**

The Affiliated Hospital of Zunyi Medical University

**Li-Jun Zhang**

The Qian-Nan-Zhou People's Hospital of Guizhou

**Hong-Fang Luo**

The 2nd affiliated Hospital of Guizhou Medical University

**Ge Huang**

The Qian-Dong-Nan People's Hospital of Guizhou

**Xi Luo**

The Affiliated Hospital of Zunyi Medical University

**Mei Tan**

The Affiliated Hospital of Zunyi Medical University

**Jindong Chen** (✉ [jindong\\_chen@hotmail.com](mailto:jindong_chen@hotmail.com))

Exploring Health, LLC.

**Yan Chen**

The Affiliated Hospital of Zunyi Medical University

---

## Research Article

**Keywords:** Next-generation sequencing, thalassemia, childbearing-age couples, molecular epidemiology,  $\alpha$ -globin,  $\beta$ -globin

**Posted Date:** March 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-273797/v1>

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

# Abstract

**Background:** Thalassemia is highly prevalent hematologic disease in Guizhou, China. This study aims to determine the epidemiological characteristics of thalassemia for couples at childbearing age in this subpopulation.

**Results:** There were 4481 couples at childbearing age recruited for thalassemia-carrier screening through both traditional hematological tests and next-generation sequencing (NGS). Of them, 1314 (14.66%) thalassemia-carriers were identified, including 857 (9.76%)  $\alpha$ -thalassemia, 391 (4.36%)  $\beta$ -thalassemia, and 48 (0.54%) composite  $\alpha$  and  $\beta$ -thalassemia. Of them, 38 couples were high-risk thalassemia carriers. In addition, 12  $\alpha$ -globin gene alterations and 16  $\beta$ -globin mutations were detected including four novel thalassemia mutations. SEA is the most common  $\alpha$ -thalassemia genotype (26.86%), CD41-42 is the most prevalent  $\beta$ -thalassemia genotype (36.57%); the  $\alpha\alpha/\alpha^{-3.7} + CD41-42$  is the most frequent composite  $\alpha$  and  $\beta$ -thalassemia genotype (18.75%). Ethnically, the Zhuang has the highest rate of thalassemia-gene carriers among the ethnic groups. Geographically, Qiannan presented the highest rate of thalassemia-gene carrier.

**Conclusion:** This result enriched the genetic map of thalassemia and provided thalassemia genetic counseling and fertility-guidance for thalassemia-carriers in Guizhou, China. The NGS is so far the most accurate method for population thalassemia screening.

## Background

Thalassemia is the most prevalent monogenic hematologic disorder that affects millions of people and leads to thousands of deaths around the world every year. Approximately 5% of the population worldwide is thalassemia carriers [1, 2]. Currently, a cure or treatment for children with thalassemia is not always available or affordable. For severe thalassemia patients, hematopoietic stem cell transplantation is the only cure, but is a costly treatment that imposes a heavy burden on families and the society. Therefore, prevention of thalassemia through conducting genetic screening and counseling to childbearing-age couples in the regions with high-incidence of thalassemia is an effective measure.

Thalassemia is highly epidemic in southern provinces of China, and it has been a serious public health problem in those regions. Thus, prevention of thalassemia has become a strategic need to reduce birth defects in southern China [3]. Guizhou, as one of the high-incidence province of thalassemia, was populated by multiple ethnic groups, and genetic screening and counseling is usually not available due to economic difficulties. Thus, thalassemia becomes more and more prevalent in this subpopulation.

Next-generation sequencing has been shown to allow rapid, multiplex and high-throughput detection of genetic variants [4]. NGS related technologies—applied to the whole genome, the exome, or targeted gene panels—have been effectively used in research settings, as well as in clinical testing and diagnosis of genetic disorders [5, 6]. In this study, we adopted NGS to determine the mutation types, prevalence, and distribution of thalassemia in 4481 couples at childbearing age from various ethnic groups inhabited in Guizhou, and then to assess the risk of having a child with thalassemia for each couple.

## Methods

### Participants

A cohort of 4481 couples (8962 subjects), at age between 19-45 years, were recruited by simple random sampling in Guizhou Province, China. These participants came from multiple ethnicities inhabiting in 9 regions across Guizhou

province (Figure 1). The subjects were divided into four age groups: 19–25 years old, 1792; 26–30 years old, 3136; 31–35 years old, 3402; 36–45 years old, 632. All participants have signed the informed consent, and all experimental protocols in this study were approved by the Medical Ethics Committee at Affiliated Hospital of Zunyi Medical University in accordance with the Declaration of Helsinki. Approximately 10 ml peripheral blood was drawn with EDTA anticoagulation from each subject, and stored at 4°C until further use.

### **Traditional screening of thalassemia carriers using hematological phenotype analysis**

All samples were first screened using traditional hematological methods, included routine blood examinations and hemoglobin electrophoresis. Routine blood examinations were performed using Automated Hematology Analyzer XE5000 (Sysmex, Japan). Hematological positivity for thalassemia is defined when the average red blood cell volume (MCV) < 80fl or average red blood cell hemoglobin content (MCHC) < 27pg. Hemoglobin electrophoresis was carried out using Automatic Hydrasys Capillary electrophoresis system (Sebia, French). Subjects are considered as thalassemia carriers when their hemoglobin HbF >1.2% or HbA2 > 3.5% or HbA2 < 2.5%. According to the International Cutout Association recommended cutoff value (MCV < 80fL, MCH < 27pg) for thalassemia diagnosis, subjects are identified as possible sufferers of  $\beta$ -thalassemia or composite  $\alpha$ - and  $\beta$  thalassemia [16].

### **Genomic DNA extraction and Genotypic Analysis Using Traditional Methods**

Blood genomic DNA was extracted with the MagPure Buffy Coat DNA Midi KF Kit (Magen, China) and the GenMag Nucleic Acid Isolation kit (GenMagBio, Beijing,China). Subjects positive for thalassemia by traditional hematological tests were further subjected to genotyping with Gap-PCR and multiplex ligation-dependent probe amplification (MLPA) for CNVs, reverse dot blot (RDB), high-resolution melting analysis (HRMA) for SNVs, and the Sanger sequencing. All techniques were performed following the routine protocols.

### **NGS Genotyping**

To compare and determine whether NGS genotyping is superior to traditional hematological tests plus Gap-PCR genotyping, we analyzed all the samples with NGS. NGS combined with Gap-PCR was used to screen thalassemia gene mutations, and the specific experimental methods were the same as those used in our previous studies [3, 8].

### **Data Analysis and Interpretation**

The original sequencing data processing and gene mutation description were implemented with references. All statistical analyses were performed using SPSS 21.0 software. The counting data was illustrated as number (n) and percentage (%). The  $\chi^2$  test was used for determination of sample difference, and  $P$ -value $\geq$ 0.05 was considered to be statistically significant.

Other related analyses have been described in our previous report [17] (Fig. 1).

## **Results**

### **Thalassemia carriers identified by hematological examinations and traditional DNA sequencing**

Among 8,962 participants, 1961 subjects were initially identified as thalassemia carriers by hematological examinations, and the detection rate was as much as 21.88%. However, only 988 subjects were confirmed to be thalassemia by traditional DNA sequencing, including 658  $\alpha$ -thalassemia carriers, 294  $\beta$ -thalassemia carriers, and 36 composite  $\alpha$ - and  $\beta$ -thalassemia carriers. The general detection rate is 11.02% (988/8962).

## Thalassemia carriers identified by NGS

In contrast, when all 8,962 individuals were directly analyzed by NGS, 1314 subjects were diagnosed as thalassemia carriers, including 875  $\alpha$ -thalassemia carriers, 391  $\beta$ -thalassemia carriers, and 48 composite  $\alpha$ - and  $\beta$ -thalassemia carriers. The overall detection rate of thalassemia is 14.66% (1314/8962). Of them,  $\alpha$ -thalassemia carriers and  $\beta$ -thalassemia carriers accounted for 9.76% (875/8962) and 4.36% (391/8962), respectively. In addition, the rate of composite  $\alpha$  and  $\beta$ -thalassemia was 0.54% (48/8962), which was determined for the first time in Guizhou (Table 1, Fig. 2).

## Alpha- and beta- gene alterations associated with thalassemia

To date, five  $\alpha$ -deletion alterations ( $-\alpha^{3.7}$ ,  $-\alpha^{4.2}$ ,  $-\text{SEA}$ ,  $-\text{FIL}$ ,  $-\text{THAI}$ ) and two common large fragment deletions of  $\beta$  thalassemia [SEA-HPFH, G $\gamma$  + (A $\gamma\delta\beta$ )] have been reported. In this study, a total of 17 different  $\alpha$  variations with 22 distinct genotypes were identified (Table 1). Of them,  $-\text{SEA}/\alpha\alpha$  was the most abundant  $\alpha$ -thalassemia genotype (26.86%), and  $-\alpha^{3.7}/\alpha\alpha$ ,  $-\alpha^{4.2}/\alpha\alpha$ , and  $\alpha\alpha^{\text{CS}}/\alpha\alpha$  genotypes are also common and represented 19.66%, 7.77%, and 7.31%, respectively. In addition, rare variations such as  $\alpha\alpha^{-\text{THAI}}$ , Poly A (A $\rightarrow$ G), Hb Phnom Penh, and Initiation codon (-T) were detected for the first time in Mainland, China.

In addition, 16 different  $\beta$  variations with 16 distinct genotypes were detected in this study (Table 1). Of them,  $\beta\text{CD41-42}/\beta\text{N}$  and  $\text{IVS-II-654 (C>T)}/\beta\text{N}$  are the two most common genotypes. This finding is consistent with the previous report [7]. The frequencies of the  $\beta\text{CD41-42}/\beta\text{N}$  and  $\text{IVS-II-654 (C>T)}/\beta\text{N}$  were 36.57% and 26.09%, respectively, in the study. The remaining common genotypes include  $\beta\text{CD17}/\beta\text{N}$ ,  $\beta\text{-50}/\beta\text{N}$ , and  $\beta\text{-28}/\beta\text{N}$  with corresponding frequencies of 10.74%, 9.72%, and 8.70%. In addition, a rare  $\beta$ -thalassemia mutation  $\beta\text{147(HC3)Stop}\rightarrow\text{Gln;HBB: c.442TC}$  was only found in Guizhou, China and has been named as Hb Zunyi [8].

Among the 48 carriers of composite  $\alpha$  and  $\beta$ -thalassemia (22 males, 26 females), 77.08% of genotypes are common deletions of  $\alpha$ -globin gene ( $\alpha\alpha^{-\text{SEA}}$ ,  $\alpha\alpha^{-\alpha^{3.7}}$ ,  $\alpha^{\text{WS}}\alpha/\alpha\alpha$ ,  $\alpha\alpha^{-\alpha^{4.2}}$ ) combined with a  $\beta$ -globin gene point mutation. Of the genotypes, composite  $\alpha\alpha^{-\alpha^{3.7}}$  and Codons 41/42 (-TTCT)/ $\beta\text{N}$  was the most frequent one. In addition, 10 abnormal hemoglobin variants including an  $\alpha$ -triplication ( $-\alpha\alpha\alpha^{\text{anti4.2}}/\alpha\alpha$ ) combined with the  $\beta$ -globin gene mutation, CD17, were identified in the 48 subjects. Among the 857 subjects of  $\alpha$ -thalassemia, 76 subjects were  $\alpha$ -triplication carriers (8.87%), including 56  $\alpha\alpha/\alpha\alpha\alpha^{\text{anti4.2}}$  and 20  $\alpha\alpha/\alpha\alpha\alpha^{\text{anti3.7}}$ .

## Geographic Distribution of thalassemia gene carriers in Guizhou

Among the 9 regions in Guizhou province, Qiannan was the region with highest carrier rate of  $\alpha$ -thalassemia (13.34%, 186/1934), while Guiyang presented the highest carrier rate of  $\beta$ -thalassemia (6.43%, 36/560). Overall, Qiannan presented the highest carrier rate of thalassemia (20.23%, 282/1934), followed by Qiandongnan (17.64%, 339/1922), Qianxinan (15.46%, 107/692), Anshun (14.82%, 126/850), and Tongren (14.49%, 80/552). The difference of the thalassemia gene carrier rates between those regions was statistically significant ( $P<0.05$ ) (Table 2, Fig. 3).

## Ethnic Distribution of thalassemia gene carriers

Among the ethnic groups including Han, Yi, Miao, Buyi, Shui, Tujia, Yao, Gelao, and Zhuang, the ethnic group with the highest carrier rate of  $\alpha$ -thalassaemia was Buyi (13.27%, 186/1402), while Gelao was the group with the highest rate of  $\beta$ -thalassaemia (9.09%, 12/132). The highest overall carrier rate of thalassaemia was Zhuang (21.05%, 16/76), followed by Buyi (19.90%, 279/1402), Miao (18.20%, 294/1615), Gelao (17.74%, 185/1043), and Tujia (17.33%,

26/150). The difference of the thalassemia carrier rates between the ethnic groups was also significantly different ( $P < 0.05$ ). (Table 3, Fig. 3).

### Identification of high-risk couples with thalassemia by NGS

In this study, we identified 0.85% (38/4881) of the couples with high risk for thalassemia using the NGS, and only 0.36% (16/4481) of the couples was found to be high-risk carriers for thalassemia by routine techniques (Fig. 3, Table 4). Thus, compared with the traditional screening/detection methods, there are 22 (23.2%) more high-risk couples for thalassemia that were identified by NGS. Among the 38 couples, 10 couples carried  $-^{SEA}/\alpha\alpha$  genotype and are high-risk for Hb Bart's edema, 11 couples were at high risk for the H disease, and the other 16 couples are carriers of heterozygous mutations. Interestingly, about half of these couples were from the same ethnic group and live in an isolated area of Guizhou Province, suggesting founder variations may exist in this particular subpopulation.

### Higher detection rate of thalassemia using NGS than traditional techniques

When traditional techniques including hematological tests and Sanger DNA sequencing were adopted, only 988 subjects were confirmed to be thalassemia carriers, the **detection rate** was 11.02% (988/8962). In contrast, 1314 subjects were identified to be thalassemia carriers by high-throughput NGS, and the detection rate was 14.46% (1314/8962). There are 326 cases or 3.64% (326/8962) of thalassemia carriers missed by routine detection techniques. Of the 326 cases, 74 were undetectable by conventional thalassemia detection techniques (Table 5), and another 252 were missed due to the lower sensitivity of the hematological examination or the defects of conventional thalassemia gene detection technology. In addition, 38 couples (0.85%, 38/4481) with high risk for thalassemia were identified by NGS while only 16 couples (0.36%, 16/4481) were detected by traditional methods.

## Discussion

Thalassemia is a serious lethal and disabling monogenic hereditary hematological disease that is common in particular areas around the world. Implementation of genetic testing and counseling to the couples at childbearing age in those regions will play a central role for preventing or reducing the birth of thalassemia sufferers [9, 10]. Guizhou province, located in the plateau mountainous area in southwestern China, is a multi-ethnic province with a high incidence of thalassemia. Thus, in this study, we determined the prevalence of young thalassemia carriers in Guizhou by recruiting 4881 couples at childbearing age for thalassemia screening using both traditional methods and NGS.

Our results indicated that the overall frequency of thalassemia gene carriers was as high as 14.66%. Geographically, of the 9 regions in Guizhou Province, Qiannan displayed the highest carrier rate of thalassemia, followed by Qianxinan, Qiandongnan, and Anshun. The higher carrier rate of thalassemia genes may be related to the poor economic condition, geographical environment, and ethnicity in those regions. In addition, ethnic groups including Buyi, Yi, Miao, and Zhuang inhabit those regions with consanguineous marriage, which lead to higher prevalence of thalassemia due to founder genetic alterations. In contrast, the majority Han showed the lowest carrier rate of thalassemia genes (8.31%) in those regions.

To date, there are 17 common types of  $\alpha$ -globin gene alterations identified in the Chinese population, and the static  $\alpha$ -thalassemia ( $-\alpha^{3.7}/$ ,  $-\alpha^{4.2}/$ ,  $\alpha^{CS}\alpha/$ ) represents 2.3% [11]. In this study, we have detected 12 of the 17 types of the  $\alpha$ -globin gene mutations. Of them,  $-^{SEA}/\alpha\alpha$ ,  $-\alpha^{3.7}/\alpha\alpha$ , and  $-\alpha^{4.2}/\alpha\alpha$  are the most common types. The genotype pattern of the  $\alpha$ -thalassemia in Guizhou is similar to that in other Southern provinces such as Fujian, Guangdong, and Guangxi [12, 13]. In addition, since the frequency of the SEA-deficiency type ( $-^{SEA}$ ) is as high as 26.86%, there is a higher possibility that both couples carry the  $-^{SEA}$  [14]. Actually, among the 38 couples of high-risk for thalassemia identified in this study, 10

couples carried the  $-^{SEA}/\alpha\alpha$  mutation that could have given birth to a child with Hb Bart's edema, and unfortunately, three couples gave birth to three children with thalassemia before the results were delivered to them. One of the children died, and the other two are under regular blood transfusion therapy. The other 28 couples of high-risk for thalassemia, including 7 couples for H disease, and 8 couples with heterozygous mutations, have regular followed ups and participate in our fertility-guidance. Thus, to prevent the birth of severe thalassemia children, it is very necessary to conduct genetic counseling and fertility guidance combined with prenatal diagnosis and follow-up for high-risk couples.

In addition, if one of the couple carries the composite  $\alpha$ - and  $\beta$ -thalassemia and the other is not the  $-^{SEA}/\alpha\alpha$  carrier, then no intermediate or severe thalassemia children will be born. In this study, we identified 48 cases with composite  $\alpha$ - and  $\beta$ -thalassemia, but only 4 cases of their spouses carried the  $-^{SEA}/\alpha\alpha$  gene. These 4 affected couples are currently under fertility-guidance.

We identified 16  $\beta$ -globin gene mutations. The common mutations include IVS-II-654 (C→T), CD41-42 (-TCTT), CD17 (A→T), and -28 (A→G). The most common genotype was  $\beta$ CD41-42/ $\beta$ N heterozygote. In addition, we also identified quite a few rare types of thalassemia gene mutations. Of them, Poly A (A→G) AATAAA→AATGAA, Hb Phnom Penh, and Initiation codon (-T) have not been previously reported in China [7]. Whether these mutations are Guizhou-population unique or not merits further investigation. Actually, we have previously reported such a rare  $\beta$ -thalassemia mutation  $\beta$ 147(HC3)Stop→Gln;HBB: c.442T>C, named Hb Zunyi [8].

Thalassemia is difficult to diagnose at the early stages of life. In this study, we demonstrated that NGS is superior to traditional hematological tests. A total of 1314 thalassemia carriers were detected with NGS, while only 988 carriers of thalassemia were confirmed when traditional methods were used for screening. There were 326 (3.64%) thalassemia carriers missed with traditional methods. Of the 326 subjects, 252 cases were missed due to the negative hematological test results, 74 cases were undetectable through Gap-PCR/PCR-RDB-Sanger sequencing. Traditional hematological tests usually presented false negative/positive results due to confusion with some small cell hypopigmentemia such as hemoglobin disease, iron deficiency anemia, etc. [15], while Gap-PCR or PCR-RDB + Sanger DNA sequencing apparently miss unknown site mutations and many large Ins/Del alterations [2]. Thus, the high-throughput NGS is currently the most accurate method for the detection of thalassaemia genes.

In summary, we identified 1314 (14.66%) thalassemia-carriers from 4481 couples (8962 subjects) at childbearing age in Guizhou Province, and determined their geographical and ethnic distribution/pattern and genetic map of thalassemia-gene clusters in this subpopulation. Of the carriers, 38 couples are identified as high-risk thalassemia-carriers, and currently under follow-up and fertility-guidance. In addition, we demonstrated that NGS is more accurate than the traditional methods for thalassemia-carrier screening. These results enriched the genetic map of thalassemia in China and provide a theoretical basis for formulating prevention and management of thalassemia sufferers.

## Abbreviations

NGS: next-generation sequencing; Hb: hemoglobin; Gap-PCR: gap-polymerase chain reaction; PCR-RDB: Polymerase chain reaction-reverse dot blot.

## Declarations

## Acknowledgements

We thank all the participants for their cooperation and contribution. We are grateful to The BGI-Shenzhen for performing and analyzing the NGS sequencing.

### **Ethics approval and consent to participate**

This study was approved by the Research Ethics Committee at Zunyi Medical University. Written informed consent was obtained from all participants. For participants under 16 years old, written informed consent was obtained from a parent or guardian.

### **Consent for publication**

Written informed consent for publication of identifying images or other personal or clinical details was obtained from all of the participants included in the study. For participants under 18 years old, written informed consent was obtained from a parent or legal guardian.

### **Availability of data and materials**

All data and materials are included in this article.

### **Competing interests**

The authors have no declaration of conflicts of interest. All the experiments undertaken in this study comply with the current laws of China, where the research was performed.

### **Funding**

This study was financially supported by the Department of Science and Technology in Guizhou (No. [2019]2806). The funder had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### **Authors' Contributions**

YC and JC designed the experiments and drafted the manuscript; YC supervised the experiments; LW, LZ, HL, and GH organized and collected samples from participants; LW, XL, and MT completed the DNA extraction and most of the experiments, LW conducted the data analysis; YC and JC reviewed the data analysis.

## **References**

1. Shang X, Xu X: **Update in the genetics of thalassemia: What clinicians need to know.** *Best practice & research Clinical obstetrics & gynaecology* 2017, **39**:3-15.
2. Viprakasit V, Ekwattanakit S: **Clinical Classification, Screening and Diagnosis for Thalassemia.** *Hematology/oncology clinics of North America* 2018, **32**(2):193-211.
3. Huang SW, Xu Y, Liu XM, Zhou M, Li GF, An BQ, Su L, Wu X, Lin J: **The Prevalence and Spectrum of alpha-Thalassemia in Guizhou Province of South China.** *Hemoglobin* 2015, **39**(4):260-263.
4. Tan M, Lu S, Wu LS, Jin DW, Peng ZY, Chen Y: **[Application of Next Generation Sequencing to Screen the Neonatal Thalassemia Genes].** *Zhongguo shi yan xue ye xue za zhi* 2015, **23**(5):1404-1409.
5. Korf BR, Rehm HL: **New approaches to molecular diagnosis.** *Jama* 2013, **309**(14):1511-1521.

6. Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, Yeung A, Peters H, Mordaunt D, Cowie S *et al*: **A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders.** *Genetics in medicine : official journal of the American College of Medical Genetics* 2016, **18**(11):1090-1096.
7. Zhang H, Li C, Li J, Hou S, Chen D, Yan H, Chen S, Liu S, Yin Z, Yang X *et al*: **Next-generation sequencing improves molecular epidemiological characterization of thalassemia in Chenzhou Region, P.R. China.** *Journal of clinical laboratory analysis* 2019, **33**(4):e22845.
8. Su Q, Chen S, Wu L, Tian R, Yang X, Huang X, Chen Y, Peng Z, Chen J: **Severe Thalassemia Caused by Hb Zunyi [beta147(HC3)Stop→Gln; HBB: c.442T>C] on the beta-Globin Gene.** *Hemoglobin* 2019, **43**(1):7-11.
9. Piel FB, Weatherall DJ: **The alpha-thalassemsias.** *The New England journal of medicine* 2014, **371**(20):1908-1916.
10. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Dewi M, Temperley WH, Williams TN, Weatherall DJ, Hay SI: **Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates.** *Lancet* 2013, **381**(9861):142-151.
11. Lai K, Huang G, Su L, He Y: **The prevalence of thalassemia in mainland China: evidence from epidemiological surveys.** *Scientific reports* 2017, **7**(1):920.
12. Li J, Xie XM, Liao C, Li DZ: **Co-inheritance of alpha-thalassaemia and beta-thalassaemia in a prenatal screening population in mainland China.** *Journal of medical screening* 2014, **21**(4):167-171.
13. Xu C, Liao B, Qi Y, Huangfu Z, Chen J, Chen Y: **Analysis of Gene Mutation Types of alpha- and beta-Thalassemia in Fuzhou, Fujian Province in China.** *Hemoglobin* 2018, **42**(3):143-147.
14. Zhang Q, Fan X, Xu M, Zhang Y, Xu H, Wen X, Zhou W: **Hb H Disease Caused by Multiple Mutations in the Polyadenylation Signal Site and -(SEA)/alphaalpha.** *Hemoglobin* 2017, **41**(3):189-192.
15. He J, Song W, Yang J, Lu S, Yuan Y, Guo J, Zhang J, Ye K, Yang F, Long F *et al*: **Next-generation sequencing improves thalassemia carrier screening among premarital adults in a high prevalence population: the Dai nationality, China.** *Genetics in medicine : official journal of the American College of Medical Genetics* 2017, **19**(9):1022-1031.
16. von Elm E, Altman DG, Egger M, Pocock SJ, Gotszche PC, Vandenbroucke JP, Initiative S: **The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.** *Lancet* 2007, **370**(9596):1453-1457.
17. Shang X, Peng Z, Ye Y, Asan, Zhang X, Chen Y, Zhu B, Cai W, Chen S, Cai R *et al*: **Rapid Targeted Next-Generation Sequencing Platform for Molecular Screening and Clinical Genotyping in Subjects with Hemoglobinopathies.** *EBioMedicine* 2017, **23**:150-159.

## Tables

**Table 1** The  $\alpha$ -thalassemia,  $\beta$ -thalassemia, and composite  $\alpha$  and  $\beta$ -thalassemia genotypes identified in Guizhou, China

Genotype	NO.	F(%)	CR(%)	AN	P(%)
$\alpha$ -thalassemia chromosomes				875	9.76
$-\alpha^{3.7}/-\alpha^{3.7}$	58	0.65	6.63		
$-\alpha^{3.7}/-\alpha^{4.2}$	46	0.51	5.26		
$-\alpha^{3.7}/\alpha\alpha^{CS}$	38	0.42	4.34		
$-\alpha^{3.7}/\alpha\alpha$	172	1.92	19.66		
$-\alpha^{4.2}/\alpha\alpha$	68	0.76	7.77		
$\alpha\alpha^{CS}/\alpha\alpha^{CS}$	18	0.20	2.06		
$\alpha\alpha^{CS}/\alpha\alpha$	64	0.71	7.31		
$\alpha\alpha^{QS}/\alpha\alpha$	19	0.21	2.17		
$_{-SEA}/-\alpha^{3.7}$	36	0.40	4.11		
$_{-SEA}/\alpha\alpha$	235	2.62	26.86		
$\alpha\alpha^{WS}/\alpha\alpha$	16	0.18	1.83		
$\alpha\alpha^{WS}/\alpha\alpha^{WS}$	16	0.18	1.83		
$\alpha\alpha\alpha^{anti4.2}/\alpha\alpha$	56	0.62	6.40		
$\alpha\alpha\alpha^{anti3.7}/\alpha\alpha$	20	0.22	2.29		
c.96-9T>C	3	0.03	0.34		
c.95+9C>T	1	0.01	0.11		
CD13(GCC>TCC)	1	0.01	0.11		
IVS-I-116	1	0.01	0.11		
$\alpha\alpha/-$ THAI	4	0.04	0.46		
Poly A (A->G)	1	0.01	0.11		
Hb Phnom Penh	1	0.01	0.11		
Initiation codon (-T)	1	0.01	0.11		
$\beta$ -thalassemia chromosomes				391	4.36
$\beta$ -28/ $\beta$ N	31	0.35	7.93		
$\beta$ -29/ $\beta$ N	4	0.04	1.02		
$\beta$ CD41-42/ $\beta$ N	143	1.60	36.57		
$\beta$ 43/ $\beta$ N	9	0.10	2.30		
$\beta$ IVS- $\square$ -654/ $\beta$ N	102	1.14	26.09		
$\beta$ CD26/ $\beta$ N	4	0.04	1.02		

$\beta$ CD27/28/ $\beta$ N	3	0.03	0.77	
$\beta$ E/ $\beta$ N	1	0.01	0.26	
$\beta$ CD17/ $\beta$ N	42	0.47	10.74	
204k dup heterozygotes	1	0.01	0.26	
$\beta$ CD37/ $\beta$ N	2	0.02	0.51	
$\beta$ -50/ $\beta$ N	38	0.42	9.72	
$\beta$ 5'UTR +43 to +40 (-AAAC)	3	0.03	0.77	
$\beta$ CD5	1	0.01	0.26	
$\beta$ 147(HC3)Stop→Gln;HBB: c.442T>C	3	0.03	0.77	
IVS-II-705 (T>G)	4	0.04	1.02	
$\alpha$ and $\beta$ -thalassemia chromosomes				48 0.54
$\alpha\alpha$ / $-\alpha$ <sup>3.7</sup> ,CD41-42	9	0.10	18.75	
Hb CS,CD41-42	6	0.07	12.50	
$\alpha\alpha$ / $-\text{SEA}$ ,CD41-42	4	0.04	8.33	
Hb QS,CD41-42	2	0.02	4.17	
CS,CD41-42	2	0.02	4.17	
$\alpha\alpha$ / $-\text{SEA}$ ,CD17	4	0.04	8.33	
Hb WS,CD17	3	0.03	6.25	
$\alpha\alpha$ / $-\alpha$ <sup>3.7</sup> ,CD17	4	0.04	8.33	
$\alpha\alpha$ / $-\alpha$ <sup>4.2</sup> ,CD17	1	0.01	2.08	
$\alpha\alpha$ <sup>anti4.2</sup> / $\alpha\alpha$ ,CD17	1	0.01	2.08	
$\alpha\alpha$ / $-\text{SEA}$ $\beta$ -50/ $\beta$ N	4	0.04	8.33	
$\alpha\alpha$ / $-\alpha$ <sup>4.2</sup> , $\beta$ -50 / $\beta$ N	3	0.03	6.25	
$\alpha\alpha$ / $-\alpha$ <sup>3.7</sup> , $\beta$ -50 / $\beta$ N	2	0.02	4.17	
$-\alpha$ <sup>3.7</sup> / $-\text{SEA}$ , $\beta$ -50 / $\beta$ N	1	0.01	2.08	
Hb WS, IVS-II-705 (T>G)	1	0.01	2.08	
Hb CS, $\beta$ -50 / $\beta$ N	1	0.01	2.08	
<b>Total</b>				<b>1314 14.66</b>

Note: No.: case number; F: genotype frequency; CR: constituent ratio; AN: allele number; P: percentage;

**Table 2 Geographical distribution of the thalassemia gene carriers in various regions of Guizhou**

Region	$\alpha$ -thal	CR(%)	F(%)	$\beta$ -thal	CR(%)	F(%)	$\alpha$ - & $\beta$ -thal chr	CR(%)	F(%)	Total	CR(%)	P(%)
Zunyi	84	5.50	0.94	62	4.06	0.69	4	0.26	0.04	150	9.83	1.67
Qiandongnan	231	12.02	2.58	95	4.94	1.06	13	0.68	0.15	339	17.64	3.78
Qiannan	186	13.34	2.08	86	6.17	0.96	10	0.72	0.11	282	20.23	3.15
Tongren	62	11.23	0.69	18	3.26	0.20		0.00	0.00	80	14.49	0.89
Liupanshui	37	5.91	0.41	14	2.24	0.16	2	0.32	0.02	53	8.47	0.59
Qianxinan	68	9.83	0.76	39	5.64	0.44		0.00	0.00	107	15.46	1.19
Guiyan	26	4.64	0.29	36	6.43	0.40	9	1.61	0.10	71	12.68	0.79
Bijie	85	10.12	0.95	15	1.79	0.17	6	0.71	0.07	106	12.62	1.18
Anshun	96	11.29	1.07	26	3.06	0.29	4	0.47	0.04	126	14.82	1.41
<b>Total</b>	<b>875</b>		<b>9.76</b>	<b>391</b>		<b>4.36</b>	<b>48</b>		<b>0.54</b>	<b>1314</b>		<b>14.66</b>

Note:  $\alpha$ -thal, cases of  $\alpha$ -thalassemia; F: genotype frequency;  $\beta$ -thal, cases of  $\beta$ -thalassemia;  $\alpha$ - &  $\beta$ -thal chr,  $\alpha$  and  $\beta$ -thalassemia chromosome; CR: constituent ratio; AN: allele number; P: percentage;

**Table 3 Ethnic distribution of thalassemia gene carriers in Guizhou**

Nationality	$\alpha$ -thal	CR(%)	F(%)	$\beta$ -thal	CR(%)	F(%)	$\alpha$ - & $\beta$ -thalchr	CR(%)	F(%)	Total	CR(%)	P (%)
Han	305	7.12	3.40	146	3.41	1.63	8	0.19	0.09	459	10.72	5.12
Dong	120	11.51	1.34	63	6.04	0.70	2	0.19	0.02	185	17.74	2.06
Miao	206	12.76	2.30	70	4.33	0.78	18	1.11	0.20	294	18.20	3.28
Buyi	186	13.27	2.08	78	5.56	0.87	15	1.07	0.17	279	19.90	3.11
Shui	20	10.05	0.22	5	2.51	0.06		0.00	0.00	25	12.56	0.28
Tujia	16	10.67	0.18	8	5.33	0.09	2	1.33	0.02	26	17.33	0.29
Gelao	9	6.82	0.10	12	9.09	0.13		0.00	0.00	21	15.91	0.23
Yao	2	4.26	0.02	4	8.51	0.04		0.00	0.00	6	12.77	0.07
Zhuang	10	13.16	0.11	4	5.26	0.04	2	2.63	0.02	16	21.05	0.18
Others	1	5.88	0.01	1	5.88	0.01	1	5.88	0.01	3	17.65	0.03
<b>Total</b>	<b>875</b>		<b>9.76</b>	<b>391</b>		<b>4.36</b>	<b>48</b>		<b>0.54</b>	<b>1314</b>		<b>14.66</b>

Note:  $\alpha$ -thal, cases of  $\alpha$ -thalassemia carrier; F: genotype frequency;  $\beta$ -thal, cases of  $\beta$ -thalassemia carrier;  $\alpha$ - &  $\beta$ -thal chr,  $\alpha$  and  $\beta$ -thalassemia chromosomes; CR: Constituent ratio; P: percentage;

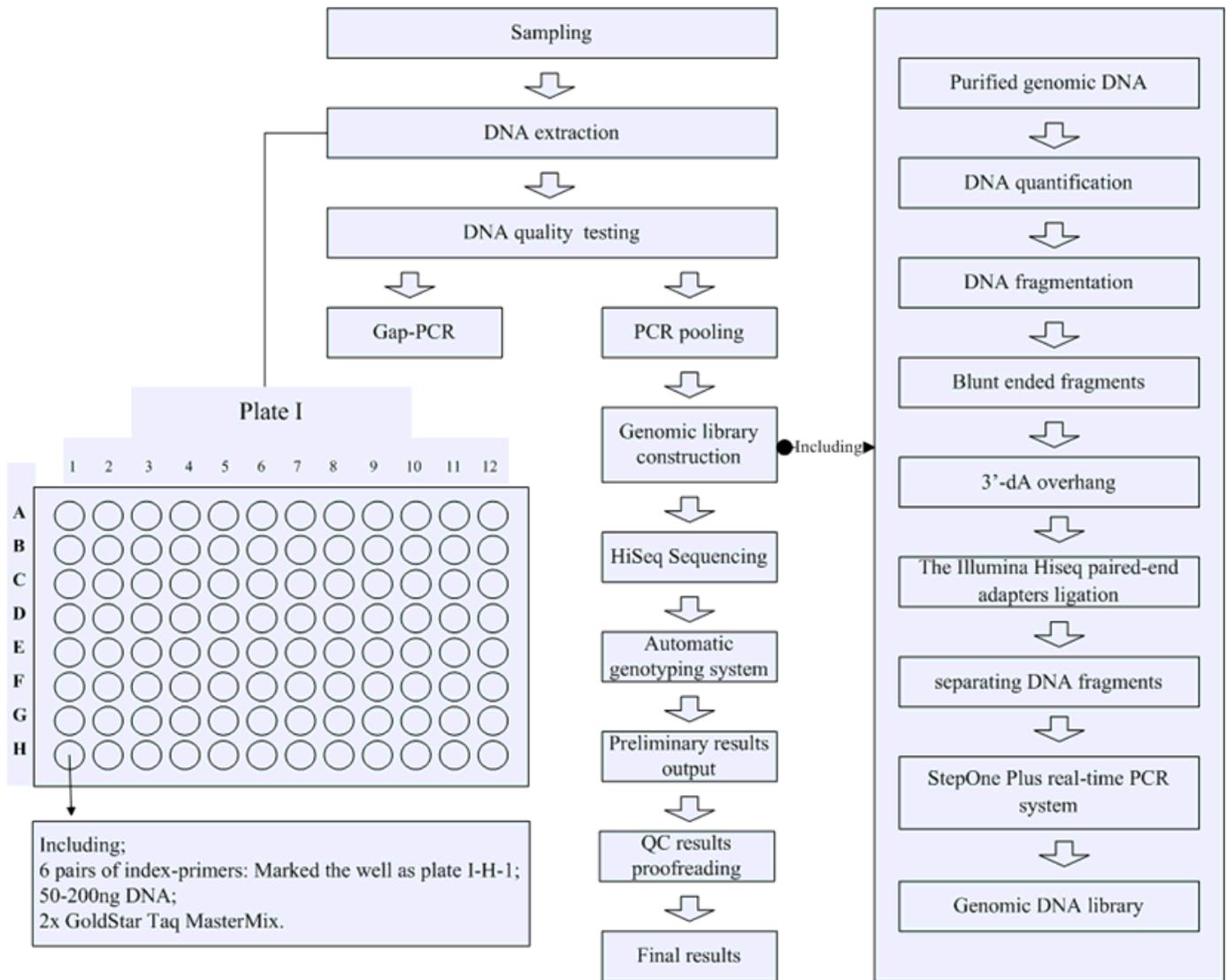
**Table 4. The globin genotypes of the 38 high-risk couples for thalassemia**

Male	Female	Number
$-\text{SEA}/\alpha\alpha$	$-\text{SEA}/\alpha\alpha$	9
$-\text{SEA}/\alpha\alpha$	$-\text{SEA}/\alpha\alpha \beta^{\text{CD17}}/\beta^{\text{N}}$	1
$-\text{SEA}/\alpha\alpha$	$-\alpha^{3.7}/\alpha\alpha$	5
$-\text{SEA}/\alpha\alpha$	$-\alpha^{4.2}/\alpha\alpha$	2
$\alpha\alpha^{\text{CS}}/\alpha\alpha$	$-\text{SEA}/\alpha\alpha$	2
$\alpha\alpha^{\text{CS}}/\alpha\alpha$	$\alpha\alpha^{\text{CS}}/\alpha\alpha$	2
$\beta^{\text{CD28}}/\beta^{\text{N}}$	$-\alpha^{3.7}/\alpha\alpha \beta^{\text{CD41-42}}/\beta^{\text{N}}$	1
$\beta^{\text{CD17}}/\beta^{\text{N}}$	$\beta^{\text{CD17}}/\beta^{\text{N}}$	3
$\beta^{\text{CD43}}/\beta^{\text{N}}$	$\beta^{\text{CD43}}/\beta^{\text{N}}$	2
$\beta^{\text{CD17}}/\beta^{\text{N}}$	$\beta^{\text{CD41-42}}/\beta^{\text{N}}$	4
$\beta^{\text{CD41-42}}/\beta^{\text{N}}$	$\beta^{\text{CD41-42}}/\beta^{\text{N}}$	6
$\beta^{\text{CD17}}/\beta^{\text{N}}$	$-\alpha^{3.7}/\alpha\alpha \beta^{\text{CD41-42}}/\beta^{\text{N}}$	1
<b>Total</b>		<b>38</b>

Table 5. 74 genotypes missed by traditional thalassemia screening techniques

Genotype	Number
$\beta$ -50/ $\beta$ N	38
IVS-II-705 (T->G)	4
$\alpha\alpha$ / <sup>-THAI</sup>	4
$\alpha\alpha$ / <sup>-SEA</sup> $\beta$ -50/ $\beta$ N	4
$\beta$ 5'UTR +43 to +40 (-AAAC)	3
$\alpha\alpha$ / $\alpha$ <sup>4.2</sup> , $\beta$ -50 / $\beta$ N	3
$\beta$ CD37/ $\beta$ N	2
$\alpha\alpha$ / $\alpha$ <sup>3.7</sup> , $\beta$ -50 / $\beta$ N	2
Poly A (A->G)	1
Hb Phnom Penh	1
Initiation codon (-T)	1
$\beta$ CD5	1
$\alpha$ <sup>3.7</sup> / <sup>-SEA</sup> , $\beta$ -50 / $\beta$ N	1
Hb WS, IVS-II-705 (T->G)	1
Hb CS, $\beta$ -50 / $\beta$ N	1
IVS-I-116	1
c.96-9T>C	3
c.95+9C>T	1
CD13(GCC>TCC)	1
$\beta$ 204k dup	1
Total	74

## Figures



**Fig. 1**

**Figure 1**

The diagram of NGS sequencing

## Distribution of participants

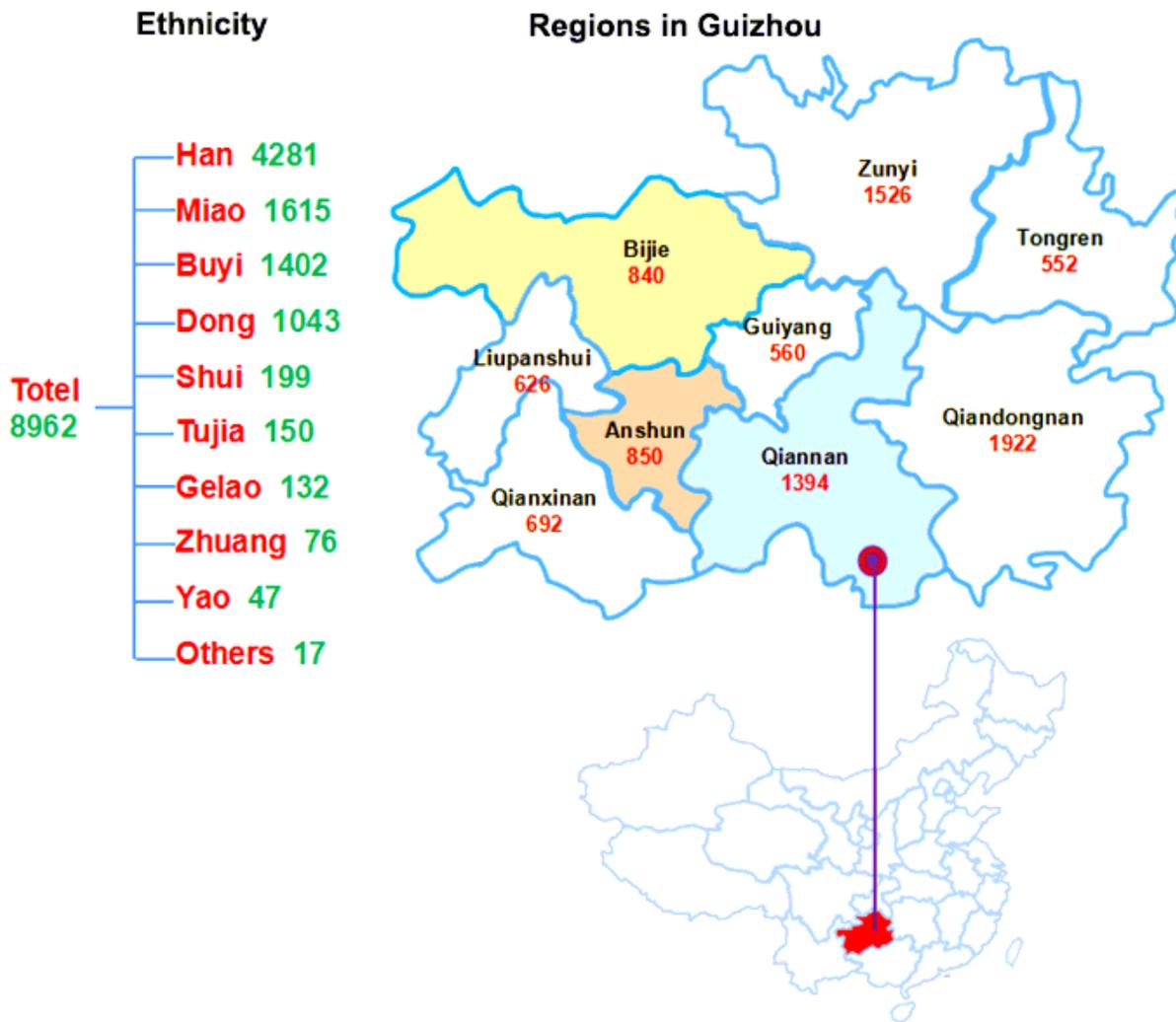
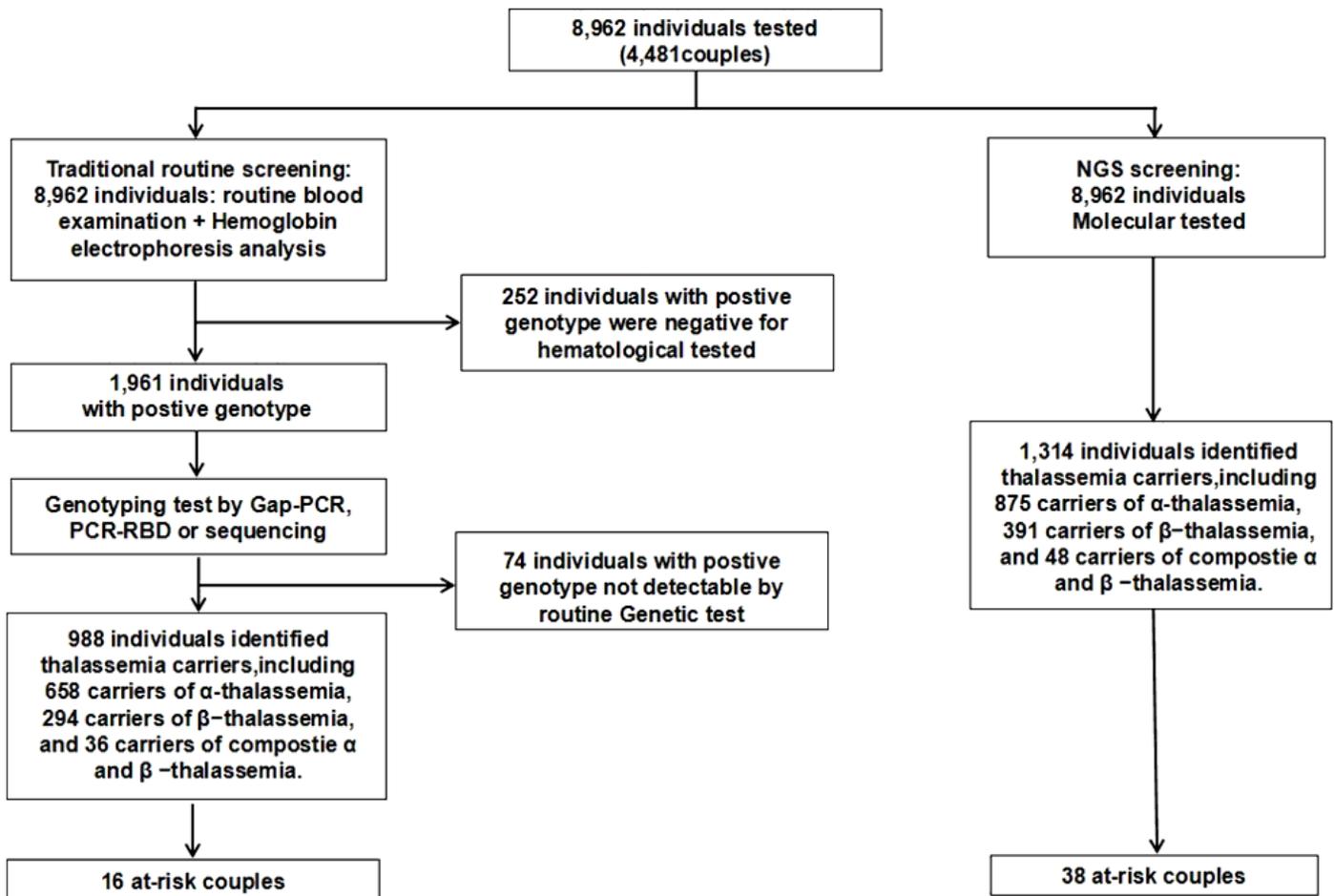


Fig 2

### Figure 2

The geographical and ethnic distribution of the 8,962 participants from Guizhou Province, China. Among the 8,962 subjects, 1922 were from Qiandongnan, 1526 from Zunyi, 1394 from Qiangnan, 850 from Anshun city, 840 from Bijie city, 692 from Qianxinan, 626 from Liupanshui, 560 from Guiyang, 552 from Tongren. The ethnic groups with numbers of participants are listed at the left side. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Fig. 3**

**Figure 3**

Comparison of couple numbers at high risk for hemoglobinopathies detected by NGS and the traditional methods. Of the 8962 participants, 1314 carriers for thalassemia were identified by NGS, while only 988 subjects were found to be thalassemia carriers by traditional methods. There are 326 thalassemia carriers missed by traditional techniques. Of the 326 cases, 252 were missed by the routine blood tests, and 74 were undetectable by GAP-PCR and Sanger sequencing. There are 38 couples defined as high-risk carriers for  $\alpha$ - and  $\beta$ -thalassemia in this study.