Genetics inMedicine | ORIGINAL RESEARCH ARTICLE

Next-generation sequencing improves thalassemia carrier screening among premarital adults in a high prevalence population: the Dai nationality, China

Jing He, MS[1,2], Wenhui Song, PhD[3], Jinlong Yang, BS[3,5], Sen Lu, PhD[4], Yuan Yuan, MS[4], Junfu Guo, MS[4], Jie Zhang, PhD[2], Kai Ye, BS[4], Fan Yang, BS[4], Fangfang Long, MS[3], Zhiyu Peng, PhD[4], Haijing Yu, PhD[5], Le Cheng, PhD[3,6] and Baosheng Zhu, MS[1,2]

Purpose: Thalassemia is one of the most common monogenic diseases in southwestern China, especially among the Dai ethnic group. Here, we explore the feasibility of a next-generation sequencing (NGS) screening method specifically for the Dai people.

Methods: Blood samples were obtained from Dai people for premarital screening. Double-blind, parallel hemoglobinopathy screening was conducted using both traditional hematological methods (red cell indexes and hemoglobin electrophoresis, then DNA sequencing) and an NGS approach.

Results: Among 951 tested individuals, we found a thalassemia carrier rate of 49.5% (471/951) using the NGS screen, in contrast to 22.0% (209/951) found using traditional methods. Almost 74.8% (217/290) of α-thalassemia carriers and 30.5% (25/82) of composite α- and β-thalassemia carriers were missed by traditional screens. The proportion of such α- and β-thalassemia carriers among the Dai people is 8.6% (82/951). For β-thalassemia carriers, the high ratio (66/99) of CD26 mutations may suggest a correlation between CD26 and the environmental adaption of the Dai people.

Conclusions: Methodological comparisons demonstrate the superiority of NGS for both sensitivity and specificity, provide a comprehensive assessment of thalassemia screening strategies, and indicate that NGS is a competitive screening method, especially among populations with a high prevalence of disease.

Genet Med advance online publication 26 January 2017

Key Words: carrier screen; China; high-prevalence population; next-generation sequencing; thalassemia

INTRODUCTION

Thalassemia is caused by a mutated or missing globin gene resulting in a decrease or structural abnormalities in one or more globin peptides. Such imbalances in globin chain production disrupt the proper assembly of hemoglobin tetramers, causing red blood cell embrittlement and hemolytic anemia disease.[1] Thalassemia is distributed mainly in coastal areas of the Mediterranean, Africa, the Middle East, India, and southeastern Asia.[2,3] Other studies have shown an unusually high prevalence of thalassemia among people in the southern provinces of China such as Guangxi, Guangdong, and Hainan.[4–6] Our recent investigations added the Yunnan province to the high-prevalence regions of thalassemia.[7] Most individuals carry four α-globin genes and two β-globin genes. Two β-globins and two α-globins serve as a scaffold that holds four heme molecules with iron to form hemoglobin. The α-thalassemia carriers resulting from a deletion or dysfunction of one allele are also called "silent carriers" because the hematological profile is generally entirely normal. When deletion or dysfunction occurs in two alleles, either in *trans* or in *cis*, a mild asymptomatic microcytic anemia (α^0-thalassemia) is present. Inactivation of three α-globin alleles, "hemoglobin H disease," has variable presentations from mild to severe. A lack of four α genes (α^0-thalassemia homozygote) is fatal: children are born with lethal hydrops fetalis unless they receive a transfusion *in utero*.[8] β-thalassemia carriers, also referred to as minor, who result from a deletion or dysfunction of one allele, have mild clinical symptoms. The lack of two β-globin alleles, β-thalassemia major, usually presents as severe anemia requiring lifelong transfusions; however, at times, this has a variable, milder presentation.[9] Therefore, effective premarital screening in regions of the world where α^0-thalassemia carriers or β-thalassemia minor are prevalent is important.

Yunnan province is located in southwestern China and links several southeastern Asian neighboring populations. Some ethnic minorities of Yunnan, like the Dai, represent a cross-border nationality. The Dai people have a high population density and frequent ethnic intermarriage. Genetic studies of Yunnan Dai

The first three authors contributed equally to this work.

[1]Yunnan Provincial Key Laboratory For Birth Defects and Genetic Diseases, The First People's Hospital of Yunnan Province, Kunming, China; [2]Medical School of Kunming University in Science and Technology, Kunming, China; [3]BGI-Yunnan, BGI-Shenzhen, Kunming, China; [4]BGI-Shenzhen, Shenzhen, China; [5]College of Life Sciences, Yunnan University, Kunming, China; [6]College of Clinical Medicine, College of Basic Medical Sciences, Dali University, Dali, China. Correspondence: Le Cheng (chengle@genomics.cn) or Baosheng Zhu (bszhu@aliyun.com)

Submitted 15 December 2015; accepted 28 November 2016; advance online publication 26 January 2017. doi:10.1038/gim.2016.218

people including investigation of thalassemia are limited. Next-generation sequencing (NGS) allows the generation of vast amounts of genomic data in order to reveal the genetic constitution of people and to evaluate potential health risks. NGS has been widely used for noninvasive prenatal diagnosis and novel mutation detection in thalassemia.[10,11] In this study, we used NGS for a large-scale population screening program in order to assess the thalassemia carrier frequency among the Dai people in Yunnan and to explore its potential use in preventing severe thalassemia and reducing pediatric mortality.

The Dai people are one of several ethnic groups residing primarily in Xishuangbanna Dai Autonomous Prefecture and Dehong Dai and Jingpo Autonomous Prefecture in Yunnan Province, southwestern China. They are closely related to the Lao and Thai people, who form a majority in Laos, Thailand and other countries and regions in south and southeastern Asia. The Thai people—the largest ethnic group in Thailand—are also called Thai in Cambodia and Vietnam, Dai in China, Shan in Myanmar, Lao in Laos, and Assam in India, and are thought to share a recent common origin.[12] People are classified as Dai in China if at least one of their parents belongs to the Dai ethnic group and they typically speak one of the southwestern Thai languages .

MATERIALS AND METHODS

Samples and demographic data

The study was approved by BGI-IRB (BGI's institutional review board on bioethics and biosafety) and all individuals provided informed written consent. A total of 1,451 people who were premarital or newlywed ethnic minorities from Dehong (Ruili, Mangshi, Lianghe, Yinjiang, Longchuan) and Xishuangbanna (Menghai, Menghun, Mengzhe) prefectures, Yunnan Province, China, were involved in the screening. Among these samples, 951 premarital or newlywed individuals were of Dai ethnicity. Only individuals aged 18 to 45 years old were included. The age distribution was as follows: 18–25 years old, 583; 26–30 years old, 266; 31–35 years old, 79; 36–41 years old, 20; unrecorded age, 3 (**Supplementary Table S1** online details the demographic data).

Traditional carrier screening using hematological phenotype analysis

All samples were screened using traditional hematological methods. This included routine blood examinations and hemoglobin electrophoresis for each sample. Hematology phenotypes were identified if a positive result was obtained for at least one of following: (i) RBC indexes of low cellular pigment, including mean corpuscular volume (MCV) ≤80 fl and/or mean corpuscular hemoglobin (MCH) ≤27 pg,[13] and (ii) HbA_2 ≤2.5% (abnormal hemoglobin concentration for a suspected α-thalassemia carrier) or HbA_2 ≥3.5% (abnormal hemoglobin concentration for a suspected β-thalassemia carrier) associated with fetal hemoglobin (HbF) ≥2.0% in some cases.

Hematological tests were performed using Automated Hematology Analyzer XS 500i (Sysmex, Kobe, Japan) for routine blood examinations and V8 Capillary electrophoresis system (Helena Biosciences Europe, Tyne and Wear, UK) for the hemoglobin analysis. Sequential hematological screening was defined as positive if MCV or MCH was positive first and then their HbA_2 was positive. Parallel hematological screening was defined as the sum of the MCV and MCH positive results and HbA_2 positive results.

NGS screen using targeted capture

Preparation for DNA samples. Genomic DNA was extracted from 200-μl blood samples using the Kingfisher Flex (Thermo Scientific, Rockford, IL) and isolated using the GenMag Nucleic Acid Isolation kit (Magnetic bead method) (GenMagBio, Beijing,China). DNA extracts were arrayed in 96-well plates and the concentration was quantified by Nanodrop-8000 (Thermo Scientific). We restricted our analysis to samples with a DNA concentration >20 ng/ml and an A260/A280 ratio between 1.8 and 2.0.

PCR amplification, pooling, library construction, and next-generation sequencing

We designed six pairs of primers for polymerase chain reaction (PCR) amplification corresponding to four gene mutations (HBA1, HBA2, HBB-1, and HBB-2) and two deletion mutations (HBA-Q and HBB-Q). The amplicons, in principle, should detect most known disease-causing point mutations and copy-number variations (CNVs) in the HbVar Database. The primers are related to the following patents: WO/2014/023076, WO/2014/023167, and CN102952877. PCR reactions were performed in 96-well plates, with each sample corresponding to one library. Ninety-six kinds of index sequences were designed, corresponding to each well of the plate. The six primers marked by the index sequence were known as the index primers. All samples were barcoded using these index primers. PCR reactions (25 μl) were performed with the index primers, 50–200 ng DNA, and 2× GoldStar Taq MasterMix (CoWin Bioscience, Beijing, China). Amplicons were sequenced using the ABI 9700 (PerkinElmer Applied Biosystems, Foster City, CA) and L69G (LongGene Scientific Instruments, Hangzhou, China) platforms. Point mutation thermal cycling conditions were 95 °C for 10 min, 95 °C for 30 s, annealing temperature for 30 s, 72 °C for 50 s, 35 cycles, 72 °C for 5 min, and 15 °C until the amplicons were pooled. CNV thermal cycling conditions were as follows: 95 °C for 10 min, 95 °C for 30 s, annealing temperature for 1 min, 24 cycles, and 15 °C until the amplicons were pooled. The four-point mutation PCR amplicons were pooled into one centrifuge tube with equal volume, and the two CNV amplicons were pooled into a second tube. We required ≥5 μg (pooled point mutation) and ≥1 μg (pooled CNV) amplicons.

We adopted the Illumina Hiseq sequencing library preparation protocol for library construction, including purified genomic DNA (Qiagen DNA Purification kit), DNA quantification (NanoDrop 8000 UV-Vis Spectrophotometer; Thermo Fisher Scientific), DNA fragmentation (excluding CNV amplicons), blunt-ended fragmentation (Enzymatics kits), 3′-dA overhang, Illumina Hiseq paired-end adapters ligation

(Illumina HiSeq), and DNA fragment separation (CNVs not included), followed by size selection using agarose gel electrophoresis and the StepOne Plus real-time PCR system. Sequencing was performed using the paired-end tag (PE100) protocol with an Illumina HiSeq2000 machine. We generated a total of 1.5 Gbp per genomic library (**Supplementary Figure S1** online).

All samples were tested and sequenced in batches by a second laboratory. Routine blood examinations were performed in five hospitals in Dehong and three hospitals in Xishuangbanna. The hemoglobin electrophoresis was also examined by two laboratories in Dehong and Xishuangbanna. All NGS sequencing was completed in four batches.

Data analysis and allele assignment

Based on the resequencing strategy, a bioinformatics process focused on detecting Hb gene point mutations and deletions was developed (**Supplementary Figure S2** online). We excluded low-quality sequences from further analysis. Filtered sequence reads were partitioned by samples based on the respective adapter information (index primer). We processed single-nucleotide polymorphism and InDel versus CNVs using different strategies. The mutation-associated strategy was as follows: raw reads were aligned on the target region reference using the BWA program[14] with default parameters and the consensus sequence was generated by the SAMtools[15] software package. Coverage, depth, and length were recorded for each consensus using ReSeqTools.[16] Single-nucleotide polymorphism and InDel results were filtered based on sequencing quality and read depth. Mutation categories were assigned based on the results of alignment between filtered consensus sequence and the HbVar Database. Based on the normalization of the target gene data with endogenous references, we estimated the relative ratio between the samples and normal controls using the read-depth statistics of the HBA1-Q, HBA2-Q, HBB-Q, and internal control genes. The variance and standard deviation of each cluster were obtained using the clustering method. The shortest distances were selected as the optimal value from the distances between each value and the mean. This was used to generate the absolute CNV for each sample.

We validated our NGS approach for detection of thalassemia carriers using three approaches: (i) we compared NGS results of 51 random samples with their Sanger sequencing results; (ii) we tested nine NGS-positive samples with the Sanger sequencing results; (iii) 23 samples that had at least one of the hematological indexes and a negative NGS result were selected to make a comparison between their NGS and Sanger sequencing results.

For the purpose of this study, codon 26 mutation is included in the term β-thalassemia.

RESULTS

NGS methodological validation

We initially performed a series of validation experiments by comparing results generated by NGS and Sanger sequencing independently as follows:

1. Sanger sequencing results from 51 random samples were found to match the NGS results completely
2. Sanger sequencing confirmed nine samples detected as positive by NGS
3. Twenty-three cases were positive according to routine blood testing and hemoglobin electrophoresis screening but negative according to NGS and Sanger sequencing

Thalassemia carriers found by NGS

In total, 471 thalassemia mutation carriers were identified from 951 samples (**Figure 1**). We determined that the Dehong population had a higher carrier rate for composite α-thalassemia and β-thalassemia carriers when compared with those from the Xishuangbanna population (**Supplementary Table S1** online).

Among composite α-thalassemia and β-thalassemia carriers, more than 62.2% (51/82) of mutations consisted of a specific deletion ($-\alpha^{3.7}/\alpha\alpha$) in addition to an HBB gene point mutation. In addition, composite carriers consisting of the deletion or gene mutation ($\alpha^{CS}\alpha$ or $--^{SEA}/\alpha\alpha$) and an HBB gene mutation were also common. The composite $-\alpha^{3.7}/\alpha\alpha$ and codon $26/\beta^{A}$ carriers are the most common and occur mainly in the Dehong population. Several rare hemoglobin gene mutations, such as c.95+1G>A, c.1delA, and Hb Queens Park, were detected in this study although they have not previously been reported in Mainland China (**Supplementary Table S2** online).

We identified 21 distinct types of α-thalassemia mutations. More than 44.5% (129/290) of carriers harbor a gene deletion, namely $-\alpha^{3.7}/\alpha\alpha$. The $--^{SEA}/\alpha\alpha$ is less common. Although α-thalassemia carrier frequencies were similar in the Dehong and Xishuangbanna populations, the rank order of the two major mutant alleles differed. For example, 16.3% (111/680) and 5.4% (37/680) of the carriers were $-\alpha^{3.7}/\alpha\alpha$ and $--^{SEA}/\alpha\alpha$, respectively, in the Dehong population, whereas 14.0% (38/271) and 6.6% (18/271) of the carriers were $--^{SEA}/\alpha\alpha$ and $-\alpha^{3.7}/\alpha\alpha$, respectively, in the Xishuangbanna population (**Supplementary Table S3** online).

We identified 10 β-globin gene mutations in the screened cohort. More than 63.6% (63/99) of samples had the codon $26/\beta^{A}$ gene mutation. The codon $17/\beta^{A}$ and codons $41–42/\beta^{A}$ genotypes are the second and third most frequent. The top three most abundant β-thalassemia gene mutations were codon $26/\beta^{A}$, codon $17/\beta^{A}$, and codons $41–42/\beta^{A}$: 8.4% (57/680), 0.6% (4/680), and 0.6% (4/680) in the Dehong population and 2.2% (6/271), 3.3% (9/271), and 2.6% (7/271) in the Xishuangbanna population, respectively (**Supplementary Table S4** online).

α-globin and β-globin mutations and ranks of population carrier frequencies

There were 12 types of α-globin gene mutations detected in this study. Among them, the two most frequent gene mutations were $-\alpha^{3.7}$ and $--^{SEA}$, representing 80.0% (335/419) of all mutations. The mutations of $\alpha^{CS}\alpha$, $\alpha^{WS}\alpha$, and $-\alpha^{4.2}$ occur frequently and occurred in a total of 15.0% (63/419) of all α-globin mutations (**Table 1**). The most common α-globin gene mutation from the

ORIGINAL RESEARCH ARTICLE

Table 1 Carrier rates of α-globin and β- globin gene mutations and constituent ratios in two populations

Mutations	Cases (*n* = 951)	Constituent ratio, %	Dehong (*n* = 680)	Local carrier rate	Xishuangbanna (*n* = 271)	Local carrier rate	Sum carrier rate (%)
$-\alpha^{3.7}$	236	56.3	201	29.6%	35	12.9%	24.8
$--^{SEA}$	99	23.6	53	7.8%	46	17.0%	10.4
$\alpha^{CS}\alpha$	26	6.2	21	3.1%	5	1.8%	2.7
$\alpha^{WS}\alpha$	24	5.7	11	1.6%	13	4.8%	2.5
$-\alpha^{4.2}$	13	3.1	6	0.9%	7	2.6%	1.4
c.95+1G>A	7	1.7	7	1.0%	-	-	0.7
Hb Owari	5	1.2	2	0.3%	3	1.1%	0.5
Hb Hekinan	4	1.0	-	-	4	1.5%	0.4
c.1delA	2	0.5	2	0.3%	-	-	0.2
$\alpha^{QS}\alpha$	1	0.2	-	-	1	0.4%	0.1
Hb Queens Park	1	0.2	-	-	1	0.4%	0.1
αααα	1	0.2	-	-	1	0.4%	0.1
Total	419	100	303	44.7%	116	42.8%	44.1
Codon 26	125	65.8	115	16.9%	10	3.7%	13.1
Codons 41–42	23	12.1	12	1.8%	11	4.1%	2.4
Codon 17	19	10.0	5	0.7%	14	5.2%	2.0
-50 G>A	6	3.2	1	0.1%	5	1.8%	0.6
-28 A>G	4	2.1	1	0.1%	3	1.1%	0.4
Hb Dhonburi	5	2.6	5	0.7%	-	-	0.5
c.316-238C>T	2	1.1	2	0.3%	-	-	0.2
Codons 71–72	2	1.1	1	0.1%	1	0.4%	0.2
Hb Hope	2	1.1	1	0.1%	1	0.4%	0.2
Codon 41 (-C)	1	0.5	1	0.1%	-	-	0.1
IVS-I-1 (G>T)	1	0.5	-	-	1	0.4%	0.1
Total	190	100	144	21.2%	46	17.0%	20.0

Dehong population differed from the Xishuangbanna population, despite the fact that these Dai populations are thought to share a common recent origin. The carrier rate of the $-\alpha^{3.7}$ deletion was estimated to be as high as 23.0% (219/951) in this study and thus differed from all previously published reports.[7,17,18]

Eleven types of β-globin gene mutations were detected in this study. The three most common are codon 26, codons 41–42, and codon 17, which represent 87.9% (167/190). Once again, the rank order differed significantly between the Dehong and Xishuangbanna populations (**Table 1**). Codon 26/β^A predominates in the Dehong in contrast to the Xishuangbanna. Mutations -50 G>A, -28 A>G, and Hb Dhonburi are also common. Three cases of Hb Dhonburi, which were reported in populations in Italy, Iran, and Thailand, were first detected in mainland China.[19–22] One case of Hb Hope matched the previous records.[7,17,18] Another two cases of c.316-238C>T were verified; years ago, such cases were reported only in India.[23]

Comparison between traditional hematological and NGS screening methods

Detection rate differences. Although 452 cases of low cellular pigment were screened by RBC indexes from 951 samples (a positive rate of 47.5%), only 77 suspected α-thalassemia carriers remained after RBC indexes and hemoglobin electrophoresis results were combined. The detection rate using the traditional screen method was only 16.4% (61/372); 83.6% (311/372) of α-thalassemia carriers were missed using traditional approaches. Similarly, β-thalassemia gene mutation detection rates were 72.9% (132/181) based on RBC indexes combined with hemoglobin electrophoresis (**Figure 2**). By contrast, NGS predicts much higher carrier frequencies. We predicted an α-thalassemia carrier frequency of 39.1% (372/951) and a β-thalassemia carrier frequency of 19.0% (181/951). We estimated the false-negative rates of α-thalassemia detection to be 23.4% by RBC indexes (87/372, including 1 α^0-thalassemia carriers) and 79.8% (297/372, including 72 α^0-thalassemia carriers) by hemoglobin electrophoresis. The false-negative rates of β-thalassemia detection by RBC indexes were 17.1% (31/181) and 10.5% (19/181) by hemoglobin electrophoresis. The predominant false-negative genotypes by RBC indexes and hemoglobin electrophoresis are reported in the **Supplementary Data** online.

There were 99 carriers with $--^{SEA}$ (24 composited with other mutations); 47 carriers had false-negative results in HbA_2 indexes (2.5–3.5), implying that the hemoglobin electrophoresis was not sensitive enough for $--^{SEA}$. No one had false-negative results in RBC indexes.

Both RBC indexes and hemoglobin electrophoresis had a high missed diagnosis ratio for thalassemia detection

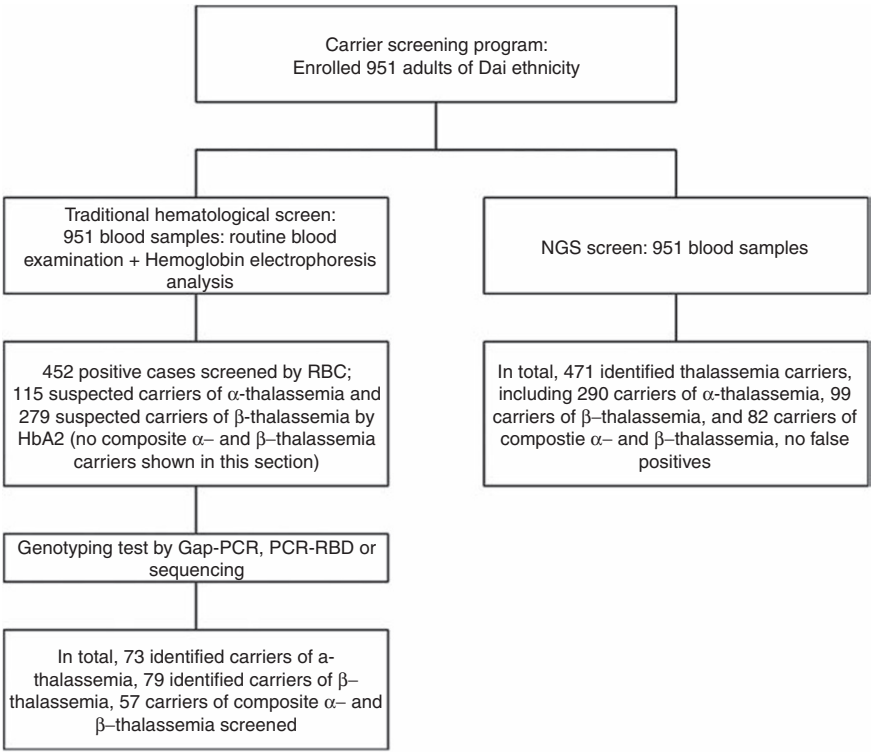

Figure 1 Diagram of work flow and outcomes.

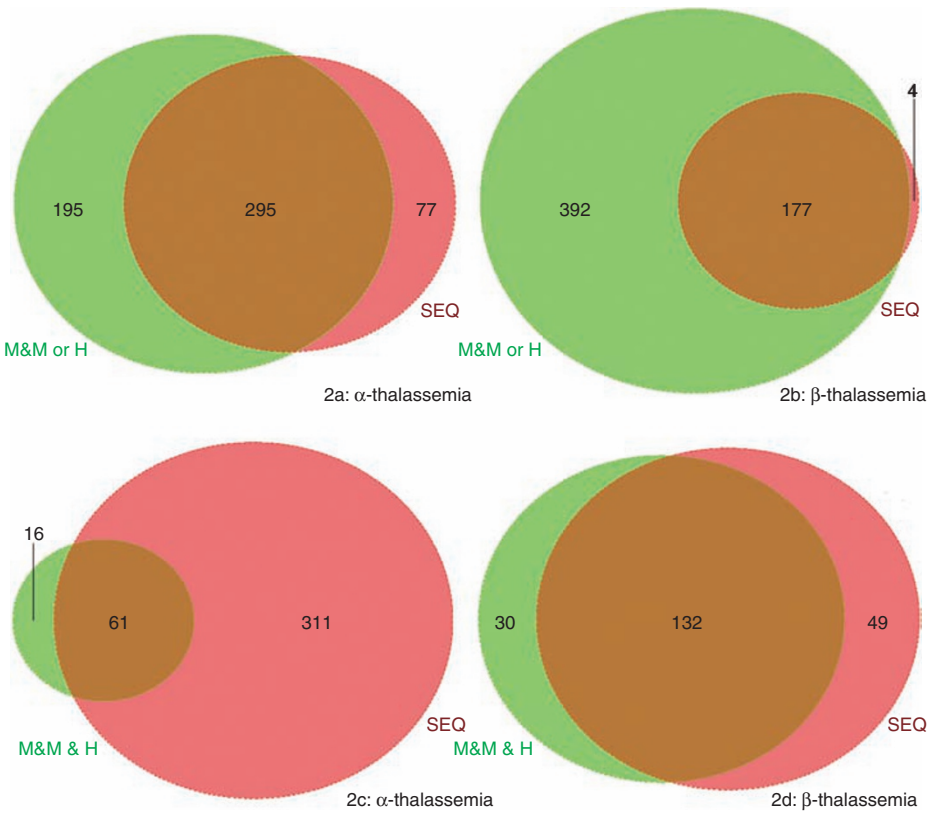

Figure 2 Methodological comparison of thalassemia mutations. (**a** and **c**) α-thalassemia and (**b** and **d**) β-thalassemia. M&M or H=MCV&MCH-positive or HbA_2-positive; M&M&H=MCV&MCH-positive and HbA_2-positive; SEQ=NGS-positive.

Table 2 Sensitivity and specificity of hematological indexes for thalassemia carriers screened by NGS

Screen parameters	Carriers	Mean sensitivity, % (95% CI)	Mean specificity, % (95% CI)	True positive	False positive	True negative	False negative
MCV+MCH only (1 sample without RBC recorded)	α-thalassemia	76.55 (71.90–80.77)	70.98 (67.10–74.65)	284	168	411	87
	$\alpha^M\alpha/\alpha\alpha$	64.46 (50.08–70.49)	58.19 (54.46–61.86)	156	296	412	86
	$\alpha^M\alpha^M/\alpha\alpha$ or $\alpha^M\alpha/\alpha^M\alpha$	99.11 (95.13–99.98)	59.31 (55.89–62.66)	111	341	497	1
	$\alpha^M\alpha^M/\alpha^M\alpha$	100 (80.49–100.00)	53.38 (50.12–56.62)	17	435	498	0
	β-thalassemia (including HbE)	82.78 (76.45–87.99)	60.65 (57.10–64.12)	149	303	467	31
	Codon 26/β^A	77.88 (69.10–85.14)	56.51 (53.08–59.90)	88	364	473	25
	Codon 26/codon 26	100.00 (54.07–100.00)	52.75 (49.51–55.98)	6	446	498	0
	Other β-thalassemia	90.16 (79.81–96.30)	55.34 (52.01–58.64)	55	397	492	6
HbA_2 only (3 samples without HbA_2 recorded)	α- thalassemia	19.51 (15.59–23.93)	92.57 (90.13–94.57)	72	43	536	297
	$\alpha^M\alpha/\alpha\alpha$	10.33 (6.80–14.87)	87.25 (84.56–89.62)	25	90	616	217
	$\alpha^M\alpha^M/\alpha\alpha$ or $\alpha^M\alpha/\alpha^M\alpha$	34.23 (25.49–43.84)	90.80 (88.64–92.67)	38	77	760	73
	$\alpha^M\alpha^M/\alpha^M\alpha$	56.25 (29.88–80.25)	88.63 (86.41–90.59)	9	106	826	7
	β-thalassemia (including HbE)	89.39 (83.92–93.49)	84.53 (81.77–87.01)	160	119	650	19
	Codon 26/β^A	93.75 (87.55–97.45)	79.19 (76.27–81.89)	105	174	662	7
	Codon 26/codon 26	100.00 (54.07–100.00)	71.02 (68.01–73.90)	6	273	669	0
	Other β-thalassemia	80.33 (68.16–89.40)	74.07 (71.05–76.93)	49	230	657	12
Sequential hematological screen MCV+MCH and Hb A_2	α-thalassemia	16.53 (12.89–20.72)	97.24 (95.55–98.41)	61	16	563	308
	$\alpha^M\alpha/\alpha\alpha$	5.79 (3.20–9.52)	91.08 (88.73–93.07)	14	63	643	228
	$\alpha^M\alpha^M/\alpha\alpha$ or $\alpha^M\alpha/\alpha^M\alpha$	34.23 (25.49–43.84)	95.34 (93.69–96.67)	38	39	798	73
	$\alpha^M\alpha^M/\alpha^M\alpha$	56.25 (29.88–80.25)	92.70 (90.84–94.29)	9	68	864	7
	β-thalassemia (including HbE)	73.74 (66.66–80.03)	96.10 (94.48–97.35)	132	30	739	47
	Codon 26/β^A	72.32 (63.07–80.36)	90.31 (88.10–92.23)	81	81	755	31
	Codon 26/codon 26	100.00 (54.07–100.00)	83.44 (80.91–85.76)	6	156	786	0
	Other β-thalassemia	73.77 (60.93–84.20)	86.81 (84.40–88.97)	45	117	770	16
Parallel hematological screen MCV+MCH or HbA_2	α-thalassemia	79.40 (74.91–83.42)	66.32 (62.31–70.17)	293	195	384	76
	$\alpha^M\alpha/\alpha\alpha$	69.10 (62.77–74.78)	54.53 (50.78–58.25)	167	321	385	75
	$\alpha^M\alpha^M/\alpha\alpha$ or $\alpha^M\alpha/\alpha^M\alpha$	99.10 (95.08–99.98)	54.84 (51.40–58.25)	110	378	459	1
	$\alpha^M\alpha^M/\alpha^M\alpha$	100.00 (79.41–100.00)	49.36 (46.10–52.62)	16	472	460	0
	β-thalassemia (including HbE)	98.32 (95.18–99.65)	49.15 (45.57–52.75)	176	391	378	3
	Codon 26/β^A	99.11 (95.13–99.98)	45.45 (42.04–48.90)	111	456	380	1
	Codon 26/codon 26	100.00 (54.07–100.00)	40.45 (37.29–43.66)	6	561	381	0
	Other β-thalassemia	96.72 (88.65–99.60)	42.73 (39.45–46.06)	59	508	379	2

$\alpha^M\alpha/\alpha\alpha$ means a deletion or gene mutation; $\alpha^M\alpha^M/\alpha\alpha$ or $\alpha^M\alpha/\alpha^M\alpha$ means two deletions or gene mutations; $\alpha^M\alpha^M/\alpha^M\alpha$ means three deletions or gene mutations. There is one sample without RBC and HbA_2 recorded ($\alpha^{CS}\alpha/\alpha\alpha$+ codons 41–42/$\beta^A$) and two samples with only RBC recorded (--^{SEA}/c.1delA+ codon 26/β^A and --^{SEA}/$\alpha\alpha$). We kept these three in our genotype statistics.

(**Supplementary Figure S3** online). Moreover, the MCV+MCH and HbA_2 sequential combined detection strategy resulted in low sensitivity and a high missed diagnosis ratio for combined carriers of α- and β- thalassemia. The sensitivity improved with the MCV+MCH and HbA_2 parallel combined detection screen when compared with only routine blood detection for β-thalassemia. This observation was not true for α-thalassemia carriers. The MCV+MCH and HbA_2 parallel combined detection screen only moderately improved detection sensitivity, with a concomitant significant loss in specificity (**Table 2**, **Supplementary Tables S5–S9** online, **Figure 2**).

DISCUSSION

Carrying rate, mutation types, and rare mutations by NGS method

In the present study, we found a much higher thalassemia carrier rate of 49.5% among the Dai people screened by an NGS method than did previously reported datasets using a hematological method of screening. This is also the first study to reveal a precise carrier rate of thalassemia in an adult cohort of the Dai people in China. Our results seemed incredibly high; however, they matched some regional studies of thalassemia. A carrier rate of 43.17% was reported for the Dai people in

Table 3 Most common allele frequencies or mutated genotypes of the countries and regions where Dai people occupy a high proportion of the total population

	α-thalassemia: α-globin variant allele frequency or mutated genotypes	β-thalassemia: β-globin variant allele frequency or mutated genotypes	Complex thalassemia: Genotypes of α- and β-thalassemia mutation compound carriers	Ref.
Dai people in Yunnan	$-\alpha^{3.7}/\alpha\alpha$	Codons 41–42	Codon 26/β^A +$-\alpha^{3.7}/\alpha\alpha$	Table 1, **Supplementary Tables S3 and S4** online
	$--^{SEA}/\alpha\alpha$	Codon 17	Codons 41–42/β^A+$-\alpha^{3.7}/\alpha\alpha$	
	$\alpha^{CS}\alpha/\alpha\alpha$	-50 G>A	Codon 17/β^A+$-\alpha^{3.7}/\alpha\alpha$	
	$\alpha^{WS}\alpha/\alpha\alpha$	-28 A>G	Codon 26/β^A +$\alpha^{CS}\alpha/\alpha\alpha$	
	$-\alpha^{3.7}/-\alpha^{3.7}$	Hb Dhonburi	Codon 26/β^A + $--^{SEA}/\alpha\alpha$	
Cambodia	$-\alpha^{3.7}$	Codons 41–42	Codon 26/β^A +$-\alpha^{3.7}/\alpha\alpha$	30
	$\alpha^{CS}\alpha$		Codon 26/β^A +$-\alpha^{3.7}/\alpha^{CS}\alpha$	
	$--^{SEA}$		Codon 26/β^A +$-\alpha^{3.7}/-\alpha^{3.7}$	
	$-\alpha^{4.2}$, $\alpha\alpha\alpha^{anti3.7}$,$\alpha\alpha\alpha^{anti4.2}$		Codon 26/codon 26 +$-\alpha^{3.7}/\alpha\alpha$	
			Codon 26/codon 26 +$-\alpha^{3.7}/\alpha^{CS}\alpha$	
Hong Kong	$--^{SEA}$	Codons 41–42	-	31
	$-\alpha^{3.7}$	IVSII-654		
	$-\alpha^{4.2}$	-28 A>G		
		Codon 17		
		Codon 43		
India	-	IVSI-5	-	32,33
		619-bp del		
		IVSI-1		
		Codons 8/9		
		Codons 41–42		
Laos	$--^{SEA}/\alpha\alpha$	Codon17	Codon 26/β^A +$-\alpha^{3.7}/\alpha\alpha$	34,35
	$-\alpha^{3.7}/\alpha\alpha$	Codons 41–42	Codon 26/β^A +$--^{SEA}/\alpha\alpha$	
	$\alpha^{CS}\alpha/\alpha\alpha$	-28 A>G	Codon 26/β^A +$\alpha^{CS}\alpha/\alpha\alpha$	
	$-\alpha^{4.2}/\alpha\alpha$	Hb Hope	Codon 26/β^A +$-\alpha^{4.2}/\alpha\alpha$	
	$-\alpha^{3.7}/-\alpha^{3.7}$	Codons 71-72, HbAbn, HbKorle-Bu	Codon 26/codon 26 +$--^{SEA}/\alpha\alpha$	
Myanmar	$-\alpha^{3.7}/\alpha\alpha$	IVS I-1	-	33,36
	$-\alpha^{3.7}/-\alpha^{3.7}$	IVSI-5		
	$--^{SEA}/\alpha\alpha$	Codons 41–42		
	$-\alpha^{4.2}/\alpha\alpha$, $-\alpha^{3.7}/-\alpha^{4.2}$	Codon 17		
	$--^{SEA}/-\alpha^{3.7}$	-28 A>G		
Thailand	$-\alpha^{3.7}/\alpha\alpha$	Codons 41–42	α^0-thalassemia: Codon 26/β^A +$\alpha^{CS}\alpha/\alpha\alpha$	37–39
	$\alpha^{CS}\alpha/\alpha\alpha$	Codon 17	Codon 26/β^A +$-\alpha^{3.7}/\alpha\alpha$	
		IVSII-654	Codon 26/β^A +$\alpha^{PS}\alpha/\alpha\alpha$	
		-28 A>G	Codon 26/codon 26 +$\alpha^{CS}\alpha/\alpha\alpha$	
		IVSI-5	β^{thal}/codon 26+$\alpha^{CS}\alpha/\alpha\alpha$	
Vietnam	$-\alpha^{3.7}/\alpha\alpha$	Codons 41–42	-	40
	$-\alpha^{3.7}/-\alpha^{3.7}$	Codon 17		
	$--^{SEA}/\alpha\alpha$	Codon 95		
	$\alpha^{CS}\alpha/\alpha\alpha$	Codons 71-72		
	$-\alpha^{3.7}/\alpha^{CS}\alpha$	IVSII-654		

Dehong (unpublished data). Another report of Dai children in Yunnan also revealed high carrier frequencies in Dehong and Xishuangbanna (unpublished data).

The α-thalassemia mutation was estimated to be up to 30.5% (290/951) primarily due to the α^{CS} and α^{WS} mutations, which are rarely reported.[24] A series of rare mutations, such as c.95 + 1G>A, c.1delA, were also first detected in this study. The $-\alpha^{3.7}$ and $--^{SEA}$ are the first and second among the most abundant α-thalassemia mutations in the Dai people, which matched previous reports.[25] Compared with other countries and regions with a high proportion of Thai people, the α-thalassemia mutations of the Dai people in Yunnan, Myanmar, Thailand, and

Vietnam share a common set of frequent genotypes. The most common α-thalassemia genotype is $-\alpha^{3.7}/\alpha\alpha$, followed by a set of other less common α-thalassemia genotypes, including $--^{SEA}/\alpha\alpha$, $\alpha^{CS}\alpha/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$. In addition, we report that the Dai people in Yunnan have a high proportion of $\alpha^{WS}\alpha/\alpha\alpha$ carriers—an observation that was not observed in the records of any other country or region.

Among β-globin mutants, codon 26 showed the highest carrier rate, matching previous investigations[26]; it forms abnormal hemoglobin E (HbE). Codons 41–42 show a higher carrier frequency than codon 17. Predominant mutations of β-thalassemia are codon 26, codons 41–42, codon 17, and -50 G>A in Yunnan. Our previous study suggested that the HbE mutation may be relevant for human adaptation in the Yunnan province.[10] Based on our present findings and previously published data, all of the countries and regions in southeastern Asia show a high proportion of codon $26/\beta^A$ and codon 26/codon 26. Codons 41–42 were the most common in five countries, including China (including Hong Kong and Yunnan), Cambodia, Thailand, and Vietnam. The Dai people in Yunnan showed the highest proportion of -50 G>A mutations; this was not reported in other countries.

With respect to composite genotypes, codon $26/\beta^A+-\alpha^{3.7}/\alpha\alpha$ and codon $26/\beta^A+\alpha^{CS}\alpha/\alpha\alpha$ are more prevalent than others. As expected, the Dai people in this study showed genotype distributions similar to those of Thailand and Vietnam populations. The similarities in composite genotypes and mutational spectrums of the hemoglobin gene suggest that the Dai and Thai nationalities may share a common ancestry and closer genetic kinship with the Kinh in Vietnam (**Table 3**). The Thai, Kinh, and Dai people share frequent commercial exchanges and intermarriage, reinforcing the genetic relationship among these three groups. This is in contrast to the Dai and native populations in Cambodia, Hong Kong, India, Laos, and Myanmar where genetic exchange is rarer.[27,28]

Methodology comparison

We demonstrate the superiority of NGS as a screen and confirmation method in high-prevalence populations. One possible explanation for the low sensitivity of traditional hematological screening may be due to the relatively higher detection rate of the "silent" α-thalassemia. The genotypes (including $-\alpha^{3.7}/\alpha\alpha$, $-\alpha^{4.2}/\alpha\alpha$, $\alpha^{CS}\alpha/\alpha\alpha$, $\alpha^{WS}\alpha/\alpha\alpha$) usually present normal values in MCV, MCH, and HbA_2 indexes; 33.3% (43/129) of $-\alpha^{3.7}/\alpha\alpha$ carriers (unincorporated with β-thalassemia) were missed using the routine hematological screen method. Similarly, 63.6% (7/11) of the $-\alpha^{4.2}/\alpha\alpha$ genotype, 57.1% (8/14) of the $\alpha^{WS}\alpha/\alpha\alpha$ genotype, and 33.3% (5/15) of the $\alpha^{CS}\alpha/\alpha\alpha$ genotype were misdiagnosed using routine hematological screen methods. In the present study, 91.5% (118/129) of $-\alpha^{3.7}/\alpha\alpha$-type carriers were missed due to HbA_2 >2.5 using the hemoglobin electrophoresis (**Table 2**, **Supplementary Tables S3, S6, and S9** online); 15.9% (10/63) of carriers with the codon $26/\beta^A$ genotype ratio were missed using the routine blood method. Diagnosis of 1 out of 24 carriers of the common β^0 type (codons $41–42/\beta^A$ and codon $17/\beta^A$) were missed using the routine hematological screen method (**Supplementary Tables S4 and S5** online). Our statistical results support the observation that some silent thalassemia carriers are most likely not detected. The routine blood test was less likely to miss β-thalassemia carriers and had fewer limitations than the α-thalassemia.

The high proportion of hematology abnormalities suggests that the Dai population is more likely to exhibit iron deficiency. The significantly decreased red cell MCV and MCH, even among α^+ thalassemia carriers, that are usually "silent" may be compounded by iron deficiency in the Dai population.[29] Further research will help define this; however, our findings provide a genetic basis for this difference.

We also found significant differences in both α-thalassemia and β-thalassemia carrier rates between groups from Dehong and Xishuangbanna, respectively. The overall thalassemia carrier rate was estimated to be 49.9 % (339/680) in the Dai people from Dehong. Their α-thalassemia carrier rate, including composite α-thalassemia and β-thalassemia genotypes, was 39.3% (267/680), with $-\alpha^{3.7}/\alpha\alpha$ and $--^{SEA}/\alpha\alpha$ being among the most common. Their β-thalassemia carrier rate, including composite α-thalassemia and β-thalassemia carriers, was 20.3% (138/680), predominantly codon $26/\beta^A$. The overall thalassemia carrier rate was 48.7% (132/271) in the Dai people from Xishuangbanna. Their α-thalassemia carrier rate, including composite α-thalassemia and β-thalassemia carriers, was 38.7% (105/271), with $--^{SEA}/\alpha\alpha$ being the most common. Their β-thalassemia carrier rate, includin g composite α-thalassemia and β-thalassemia carriers, was 15.9% (43/271), predominantly codon $17/\beta^A$.

The genetic heterogeneity of the Dai people in Xishuangbanna is greater than that for those in Dehong. This may relate to Xishuangbanna's unique geographical position as a transportation hub to southeastern Asia and mainland China, resulting in more genetic exchange and greater diversity. Although the two prefectures have similar α-thalassemia and β-thalassemia carrier frequencies, the Dai people in Xishuangbanna are at greater risk for giving birth to children with severe thalassemia because of a more complex and varied set of α-globin and β- globin gene mutations. Codons 41–42 and codon 17 mutations, for example, are more severe (based on the degree of hemophthisis) than codon 26. Similarly, the deletion allele $--^{SEA}/\alpha\alpha$ results in a serious clinical manifestation of anemia. We propose that requiring thalassemia carrier screening, particularly in premarital or newlywed Dai people in Xishuangbanna, may be crucial for preventing the birth of children with severe thalassemia.

Thirteen hemoglobin-H (HbH) carriers (not including composite α- and β-thalassemia carriers) were found in our survey. HbH patients present obvious microcytic hypochromic anemia, hepatosplenomegaly, mild jaundice, and other symptoms. HbH patients often showed clinical manifestations that varied greatly. The genotype of HbH patients matched the α-globin gene mutation types mainly in two prefectures. HbH carriers with $--^{SEA}/-\alpha^{3.7}$ were frequently from Dehong, whereas those with $--^{SEA}/\ \alpha^{WS}\alpha$ mainly originated from Xishuangbanna.

Because our samples were taken from an adult cohort of reproductive age, the HbH patients detected in this research did not manifest serious clinical symptoms.

HbE is enriched in southeastern Asia, primarily in Laos, Thailand, Cambodia, and China. HbE carriers generally show no clinical manifestations and are difficult to diagnose, especially among Yunnan minority populations; we hypothesize that local intermarriage has increased the risk of disease in this area. HbE and β-thalassemia carriers have a complex and diverse set of phenotypes ranging from asymptomatic to requiring frequent clinical blood transfusions. No HbE/β^0 carriers were detected in our study.

In summary, we report a high frequency of missed thalassemia carriers based on conventional hematological methods. Using an NGS approach, we analyzed more than 300 α-hemoglobin and β-hemoglobin mutations using a single test with a cost-effective price for each sample. Our approach significantly reduces false-negative results and misdiagnoses, and also reduces the need for repeated blood sampling and further referral tests. Our strategy may facilitate carrier screen programs in areas with a high prevalence of thalassemia. However, considering the complexity of α- and β-globin gene mutations in this population, there is always the possibility of misinterpretation of results, incorrectly assigning increased risk when there is none, and vice versa. Genotypic diagnoses must be interpreted by health-care workers and counselors properly trained in globin gene genetics and all its clinical manifestations.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

This project was supported by funding from the Special Fund for BGI-Yunnan's High-Throughput Sequencing Platform (Yunnan Province 2013DA008, P.R. China), the Research of High-Throughput Sequencing Technology in the Application of Prevention and Control of the Mediterranean Anemia High-Risk Groups (Yunnan Province 2014FC003, P.R. China), the Construction of China National Genebank (Yunnan Genebank) (Yunnan Province 2015DA008, P.R. China), the Cultivation of Backup Young and Middle-Aged Academic Technology Leaders in Yunnan Province (Yunnan Province 2014HB053, P.R. China), the Fund of the National Natural Science Foundation (81260415, P.R. China), and the National Key Research and Development Program: Precision Medical Research (2016YFC0900503, P.R. China). We thank Evan Eichler, School of Medicine, University of Washington, Seattle, WA, for helpful comments regarding the manuscript.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Muncie HL Jr, Campbell J. Alpha and beta thalassemia. *Am Fam Physician* 2009;80:339–344.
2. Weatherall DJ. The thalassaemias. *BMJ* 1997;314:1675–1678.
3. Cao A, Kan YW. The prevention of thalassemia. *Cold Spring Harb Perspect Med* 2013;3:a011775.
4. 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.
5. Xu XM, Zhou YQ, Luo GX, et al. The prevalence and spectrum of alpha and beta thalassaemia in Guangdong Province: implications for the future health burden and population screening. *J Clin Pathol* 2004;57:517–522.
6. Yao H, Chen X, Lin L, et al. The spectrum of α- and β-thalassemia mutations of the Li people in Hainan Province of China. *Blood Cells Mol Dis* 2014;53:16–20.
7. Zhang J, Zhu BS, He J, et al. The spectrum of α- and β-thalassemia mutations in Yunnan Province of Southwestern China. *Hemoglobin* 2012;36: 464–473.
8. Origa R, Moi P, Galanello R, Cao A. Alpha-thalassemia. In: Pagon RA, Adam MP, Ardinger HH, et al. (eds). *GeneReviews*. University of Washington: Seattle, WA, 1 November 2005; updated 21 November 2013.
9. Origa R. Beta-thalassemia. In: Pagon RA, Adam MP, Ardinger HH, et al. (eds). *GeneReviews*. University of Washington: Seattle, WA, 28 September 2000; updated 14 May 2015.
10. Papasavva T, van Ijcken WF, Kockx CE, et al. Next generation sequencing of SNPs for non-invasive prenatal diagnosis: challenges and feasibility as illustrated by an application to β-thalassaemia. *Eur J Hum Genet* 2013;21:1403–1410.
11. Xiong L, Barrett AN, Hua R, et al. Non-invasive prenatal diagnostic testing for β-thalassaemia using cell-free fetal DNA and next generation sequencing. *Prenat Diagn* 2015;35:258–265.
12. Sun H, Zhou C, Huang X, et al. Autosomal STRs provide genetic evidence for the hypothesis that Tai people originate from Southern China. *PloS One* 2013;8:e60822.
13. Ma ES, Chan AY, Ha SY, Lau YL, Chan LC. Thalassemia screening based on red cell indices in the Chinese. *Haematologica* 2001;86:1310–1311.
14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
15. Li H, Handsaker B, Wysoker A, et al.; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
16. He W, Zhao S, Liu X, et al. ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis. *Genet Mol Res* 2013;12:6275–6283.
17. Zhang J, He J, Zeng XH, et al. Genetic heterogeneity of the β-globin gene in various geographic populations of Yunnan in southwestern China. *PLoS One* 2015;10:e0122956.
18. Zhu BS, He J, Zhang J, et al. [A study on gene mutation spectrums of α- and β-thalassemias in populations of Yunnan Province and the prenatal gene diagnosis]. *Zhonghua Fu Chan Ke Za Zhi* 2012;47:85–89.
19. Grosso M, Rescigno G, Zevino C, Matarazzo M, Poggi V, Izzo P. A rare case of compound heterozygosity for delta(+)27 and Hb Neapolis (Dhonburi) associated to an atypical beta-thalassemia phenotype. *Haematologica* 2001;86:985–986.
20. Moghimi B, Yavarian M, Oberkanins C, et al. Hb Dhonburi (Neapolis) [beta126(H4)Val–>Gly] identified in a family from northern Iran. *Hemoglobin* 2004;28:353–356.
21. Pagano L, Viola A, Fioretti G, Ammirabile M, Ricchi P, Prossomariti L. Neapolis (CD 126 beta+ GGT->GGG): a result of a screening in Campania, a region in Southern Italy. *Haematologica* 2007;92:990–991.
22. Yamsri S, Singha K, Prajantasen T, et al. A large cohort of β(+)-thalassemia in Thailand: molecular, hematological and diagnostic considerations. *Blood Cells Mol Dis* 2015;54:164–169.
23. Edison ES, Shaji RV, Devi SG, et al. Analysis of beta globin mutations in the Indian population: presence of rare and novel mutations and region-wise heterogeneity. *Clin Genet* 2008;73:331–337.
24. Xu JJ, Qiu XX, Du J, Li M, Huang PL, Li J. Clinical characteristics of heterozygote and double heterozygote of three rare mutations in alpha-thalassemia. *J Int Reprod Health Fam Planning* 2014;3:172–174.
25. Li J, Li R, Zhou JY, Xie XM, Liao C, Li DZ. Prenatal control of nondeletional α-thalassemia: first experience in mainland China. *Prenat Diagn* 2013;33: 869–872.
26. He J, Zeng X, Zhang Y, et al. Prevalence of hemoglobin E in Yunnan Province of Southwest China. *Hematology* 2016;21:54–59.
27. He P. The migration of early Tai groups and the formation of modern Dai, Lao, Thai and Shan Peoples. *Study Ethnics Guangxi* 2005;2:134–142.
28. Peng MS, Quang HH, Dang KP, et al. Tracing the Austronesian footprint in Mainland Southeast Asia: a perspective from mitochondrial DNA. *Mol Biol Evol* 2010;27:2417–2430.

29. Yang F-b, Li-sha H, Tuan-biao Z, Li-qin Y, Jing-tao L. Investigation of anemia epidemiology to under 10 years-old children who are minorities in three borders in Yunnan province. *Soft Sci Health* 2011;25:649–652.
30. Carnley BP, Prior JF, Gilbert A, et al. The prevalence and molecular basis of hemoglobinopathies in Cambodia. *Hemoglobin* 2006;30:463–470.
31. Lau YL, Chan LC, Chan YY, et al. Prevalence and genotypes of alpha- and beta-thalassemia carriers in Hong Kong – implications for population screening. *N Engl J Med* 1997;336:1298–1301.
32. Colah R, Nadkarni A, Gorakshakar A, et al. Impact of beta globin gene mutations on the clinical phenotype of beta thalassemia in India. *Blood Cells Mol Dis* 2004;33:153–157.
33. Brown JM, Thein SL, Weatherall DJ, Mar KM. The spectrum of beta thalassaemia in Burma. *Br J Haematol* 1992;81:574–578.
34. Sicard D, Lieurzou Y, Lapoumeroulie C, Labie D. High genetic polymorphism of hemoglobin disorders in Laos: complex phenotypes due to associated thalassemic syndromes. *Hum Genet* 1979;50:327–336.
35. Savongsy O, Fucharoen S, Fucharoen G, Sanchaisuriya K, Sae-Ung N. Thalassemia and hemoglobinopathies in pregnant Lao women: carrier screening, prevalence and molecular basis. *Ann Hematol* 2008;87:647–654.
36. Than AM, Harano T, Harano K, Myint AA, Ogino T, Okadaa S. High incidence of 3-thalassemia, hemoglobin E, and glucose-6-phosphate dehydrogenase deficiency in populations of malaria-endemic southern Shan State, Myanmar. *Int J Hematol* 2005;82:119–123.
37. Sanchaisuriya K, Fucharoen G, Sae-ung N, Jetsrisuparb A, Fucharoen S. Molecular and hematologic features of hemoglobin E heterozygotes with different forms of alpha-thalassemia in Thailand. *Ann Hematol* 2003;82:612–616.
38. Fucharoen G, Trithipsombat J, Sirithawee S, et al. Molecular and hematological profiles of hemoglobin EE disease with different forms of alpha-thalassemia. *Ann Hematol* 2006;85:450–454.
39. Yamsri S, Sanchaisuriya K, Fucharoen G, Sae-Ung N, Fucharoen S. Genotype and phenotype characterizations in a large cohort of β-thalassemia heterozygote with different forms of α-thalassemia in northeast Thailand. *Blood Cells Mol Dis* 2011;47:120–124.
40. O'Riordan S, Hien TT, Miles K, et al. Large scale screening for haemoglobin disorders in southern Vietnam: implications for avoidance and management. *Br J Haematol* 2010;150:359–364.