

Jun Ohashi · Izumi Naka · Jintana Patarapotikul
Hathairad Hananantachai · Gary Brittenham
Sornchai Looareesuwan · Andrew G. Clark
Katsushi Tokunaga

Strong linkage disequilibrium of a HbE variant with the (AT)₉(T)₅ repeat in the BP1 binding site upstream of the β -globin gene in the Thai population

Received: 7 September 2004 / Accepted: 1 October 2004 / Published online: 18 November 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract A binding site for the repressor protein BP1, which contains a tandem (AT)_x(T)_y repeat, is located approximately 530 bp 5' to the human β -globin gene (HBB). There is accumulating evidence that BP1 binds to the (AT)₉(T)₅ allele more strongly than to other alleles, thereby reducing the expression of HBB. In this study, we investigated polymorphisms in the (AT)_x(T)_y repeat in 57 individuals living in Thailand, including three homozygotes for the hemoglobin E variant (HbE; $\beta^{26}\text{Glu} \rightarrow \text{Lys}$), 22 heterozygotes, and 32 normal homozygotes. We found that (AT)₉(T)₅ and (AT)₇(T)₇ alleles were predominant in the studied population and that the HbE variant is in strong linkage disequilibrium with the (AT)₉(T)₅ allele, which can explain why the β^E chain is inefficiently synthesized compared to the normal

β^A chain. Moreover, the mildness of the HbE disease compared to other hemoglobinopathies in Thai may be due, in part, to the presence of the (AT)₉(T)₅ repeat on the HbE chromosome. In addition, a novel (AC)_n polymorphism adjacent to the (AT)_x(T)_y repeat (i.e., (AC)₃(AT)₇(T)₅) was found through the variation screening in this study.

Keywords HbE · Thai · (AT)_x(T)_y repeat · Linkage disequilibrium · BP1

MIM and accession numbers and URLs for data presented herein are as follows: Online Mendelian Inheritance of Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for HBB [MIM 141900]), GenBank, <http://www.ncbi.nlm.nih.gov/> (accession number [NG_000007.2] reference sequence information).

J. Ohashi (✉) · I. Naka · K. Tokunaga
Department of Human Genetics,
Graduate School of Medicine,
The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-0033, Japan
E-mail: juno-ky@umin.ac.jp
Tel.: +81-3-58413693
Fax: +81-3-58028619

J. Ohashi · A. G. Clark
Department of Molecular Biology and Genetics,
Cornell University, Ithaca, NY, USA

J. Patarapotikul · H. Hananantachai · S. Looareesuwan
Department of Microimmunology and Clinical
Tropical Medicine, Faculty of Tropical Medicine,
Mahidol University, Bangkok, Thailand

G. Brittenham
Department of Pediatrics and Medicine,
College of Physicians and Surgeons,
Columbia University, New York, NY, USA

Introduction

Hemoglobin E (HbE; $\beta^{26}\text{Glu} \rightarrow \text{Lys}$) is the most common variant of the β -globin gene (HBB; MIM 141900) in Southeast Asia. The HbE variant causes a structural defect in hemoglobin, while the HbE disease (homozygotes for HbE) shows milder clinical severity than hemoglobin disorders caused by HbS (HbS; $\beta^{6}\text{Glu} \rightarrow \text{Val}$) and HbC (HbC; $\beta^{6}\text{Glu} \rightarrow \text{Lys}$). In addition, the HbE trait (heterozygotes for HbE) is clinically benign with no anemia. The severity is considered to be proportional to the degree of the structural defect and the expression level of the variant chromosome. The β^E chain is inefficiently synthesized compared to the normal β^A chain so that heterozygotes for HbE express less variant β chain than the HbS and HbC heterozygotes (Traeger et al. 1980). This lower expression of the β^E chain seems to be one reason why the HbE disease and trait are mostly benign.

The HbE variant has a point mutation at codon 26 that changes a glutamic acid in the normal β chain to a lysine. In addition to creating a different hemoglobin structure, this exon mutation also activates a cryptic 5' splicing site at codon 25 of the HbE pre-mRNA. The mRNA produced by the aberrant splicing is nonfunctional, and the amount of correctly spliced HbE mRNA

is decreased (Orkin et al. 1982), resulting in the low concentration of β^E chain compared to β^A chain in HbE heterozygotes.

The expression of HBB is also known to be regulated by many critical cis-acting sequences (Grosveld et al. 1993). There is a binding site for the repressor protein BP1 located upstream of HBB at position -530 (Berg et al. 1989; Chase et al. 2002). The BP1 binding site contains a tandem $(AT)_x(T)_y$ repeat, alterations in which change the binding affinity for BP1, thereby influencing the transcription of HBB. To date, several alleles with different numbers of repeats have been found in various populations. The $(AT)_9(T)_5$ allele has been reported to be associated with silent β -thalassemia (Semenza et al. 1984; Murru et al. 1990). In addition, compared to other repeats on the HbS chromosome, the $(AT)_9(T)_5$ repeat binds BP1 tightly, and, in Indians with the HbS trait, the HbS is in linkage disequilibrium with $(AT)_9(T)_5$ (Elion et al. 1992). This may explain why the expression level of HbS is lower and the disease is milder in Indians than in Africans with the sickle cell trait (Elion et al. 1992). More recently, transient expression experiments have revealed that the $(AT)_9(T)_5$ repeat affects the promoter activity in an erythroid environment (Kalotychnou et al. 2002). These results strongly suggest that the expression level of the β chain coded by the chromosome with the $(AT)_9(T)_5$ repeat is reduced compared to other chains coded by the chromosome without the $(AT)_9(T)_5$ repeat.

The HbE variant was reported to be on the chromosome with $(AT)_7(T)_7$ or $(AT)_9(T)_5$ in the Chinese population (Zhou et al. 1995). However, it is uncertain whether or not the HbE in the Thai population is also in linkage disequilibrium with these $(AT)_x(T)_y$ alleles because HbE in Southeast Asia has been suggested to have multiple origins (Antonarakis et al. 1982). In Thailand, the overall frequency of HbE is estimated to be 13%, and in some areas, the frequency of HbE equals that of HbA (Wasi et al. 1980). The HbE variant is thought to

have arisen recently by a single mutation and rapidly increased its population frequency in Thailand (Ohashi et al. 2004). Thus, it is interesting to examine which $(AT)_x(T)_y$ allele is in linkage disequilibrium with such a major HbE variant. In this study, we investigated variations in the $(AT)_x(T)_y$ repeat in three HbE homozygous, 22 HbE heterozygous, and 32 normal (HbA) homozygous individuals living in Thailand. For the genotyping of this complicated repeat, we developed an efficient polymerase chain reaction–single-strand conformation polymorphism (PCR-SSCP) method.

Materials and methods

Subjects and polymorphisms

For this study, we recruited Thai patients with *Plasmodium falciparum* malaria living in Suan Pung, Thailand, near the border with Myanmar. The diagnosis of mild malaria is described elsewhere (Ohashi et al. 2002). The HbE variant genotype of malaria was described in our previous study (Ohashi et al. 2004). In the present study, we analyzed five polymorphic sites, including the $(AT)_x(T)_y$ repeat at the putative BP1 binding site, in 57 subjects. This study was approved by the institute review board of the Faculty of Tropical Medicine, Mahidol University, and informed consent was obtained from all participants.

PCR amplification and sequencing

PCR was performed for the putative BP1 binding site using the sense primer BP1-F (5'-gcattgcatgagcaaatgaaga-3') and the antisense primer BP1-R (5'-cagcagaatggatgaaaactc-3'). The positions of detected polymorphisms and primers on NG_000007.2 are shown in Fig. 1. PCR was carried out in a GeneAmp PCR system 9700 thermal cycler (Perkin-Elmer, Applied Biosystems) using an initial denaturation for 10 min at 96°C, followed by 40 cycles of denaturation for 30 s at 96°C, annealing for 30 s at 58°C, and extension for 30 s

Fig. 1 Five polymorphisms detected in this study. The position of each polymorphism was described based on the GenBank sequence entry NG_000007.2

NG_000007.2

69601

ttttatctcttgtttcccaaacctaataagtaactaatgcacagagcacattgatttgattttatttcta

BP1-F

tttttagacataatttattatgcatgcatgagcaaatgaaga¹aaaacaacaacaaatgaatgcatatatat

T/C

(AC)_n

(AT)_x

(T)_y

gtatatgtatgtgtgtat²atata³acacatatatatatat⁴ttttttt⁵cttttcttaccagaagggttttaa

69758

T/C

BP1-R

tccaaataaggagaagatagcttagaact⁶gaggtag⁷gagttttcatccattctgtcctg⁸taagtattttg

69840

catattctggagacgcaggaagagatccatctacatatcccaagctgaattatggtagacaaagctctt

69950

Table 1 Haplotypes including the BPI binding site detected in this study

Haplotype	TGTA	69758T/C	ATAT	(AC) _n	(AT) _x	(T) _y	CTTTT	69840T/C	BPI binding site
A1	TGTA	C	ATAT	ACAC	ATATATATATATATATAT	TTTTT	CTTTT	T	(AT) ₉ (T) ₅
A2 ^a	TGTA	T	ATAT	ACAC	ATATATATATATAT	TTTTTTTT	CTTTT	T	(AT) ₇ (T) ₇
A3	TGTA	T	ATAT	ACACAC	ATATATATATATAT	TTTTT	CTTTT	C	(AT) ₇ (T) ₅
A4	TGTA	T	ATAT	ACAC	ATATATATATATATAT	TTTTT	CTTTT	C	(AT) ₈ (T) ₅
A5	TGTA	C	ATAT	ACAC	ATATATATATATATAT	TTTTTTTT	CTTTT	T	(AT) ₈ (T) ₇
A6	TGTA	C	ATAT	ACAC	ATATATATATATAT	TTTTTTTT	CTTTT	T	(AT) ₇ (T) ₇
A7	TGTA	C	ATAT	ACAC	ATATATATATATATATAT	TTTTT	CTTTT	T	(AT) ₁₀ (T) ₅

^aHaplotype A2 corresponds to the reference sequence NG_000007

at 72°C. PCR products were sequenced for all samples on both strands using an ABI Prism 3100 automated sequencer (Perkin-Elmer, Applied Biosystems).

PCR-SSCP analysis

PCR-SSCP analysis for the amplified fragment including the (AT)_x(T)_y repeat was performed for all samples. The amplified fragment included five polymorphic sites (Table 1 and Fig. 1). One microliter of solution containing the PCR product was mixed with 7 µl of denaturing solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF). The mixture was denatured at 96°C for 5 min and immediately cooled on ice. One microliter of the mixture was applied to a 10% polyacrylamide gel (acrylamide:bis-acrylamide = 49:1) containing 10% glycerol. Electrophoresis was carried out at 20 mA/gel for 70 min at 10°C in a minigel electrophoresis unit (ATTO, Tokyo, Japan) and in 0.5X TBE (45 mM Tris-borate, pH 8.0, 1 mM EDTA). The separated single-stranded DNA fragments in the gel were visualized by silver staining (Daiichi Pure Chemicals, Tokyo, Japan). PCR-SSCP of the fragments showed seven different banding patterns (Fig. 2). To obtain the reference samples for the electrophoresis, each silver-stained band showing a unique position was cut out and subjected to direct PCR-based sequencing. Accordingly, we observed seven haplotypes corresponding to the unique seven banding patterns (Table 1). The results from direct PCR-based sequencing and PCR-SSCP were consistent.

Statistical analysis

Because our SSCP analysis could distinguish the seven haplotypes, we treated the SSCP haplotypes as alleles at a single polymorphic site. A phase or pair of haplotypes for 57 subjects (i.e., 28 HbE and 86 HbA chromosomes) was inferred using the Phase-Standard analysis version 2.0.2 program (Stephens et al. 2001; Stephens and Donnelly 2003) with the default settings. In this study, we reported the number of haplotypes based on the most likely haplotype pair for each subject.

Results

In this study, we investigated variations in the (AT)_x(T)_y repeat in three HbE homozygous, 22 HbE heterozygous, and 32 normal (HbA) homozygous individuals living in Thailand. Five polymorphisms, 69758T/C, (AC)_n, (AT)_x, (T)_y, and 69840T/C were genotyped successfully using both direct PCR-based sequencing and PCR-SSCP, allowing determination of the phase or pair of haplotypes in each individual. In the studied population, we observed seven haplotypes defined by these five polymorphisms (Table 1 and Fig. 1). Four alleles were

Table 2 Estimated number of haplotype frequencies among 57 subjects (114 chromosomes)

Haplotype	$((AT)_x(T)_y)$ allele	Number of haplotypes ^a	(Sample frequency)
HbE—A1	$((AT)_9(T)_5)$	27	(0.237)
HbE—A2	$((AT)_7(T)_7)$	1	(0.009)
HbA—A1	$((AT)_9(T)_5)$	19	(0.167)
HbA—A2	$((AT)_7(T)_7)$	18	(0.158)
HbA—A3	$((AT)_7(T)_5)$	3	(0.026)
HbA—A4	$((AT)_8(T)_5)$	21	(0.184)
HbA—A5	$((AT)_8(T)_7)$	5	(0.044)
HbA—A6	$((AT)_7(T)_7)$	19	(0.167)
HbA—A7	$((AT)_{10}(T)_5)$	1	(0.009)

^aNumbers of haplotypes were calculated based on the most likely pair of haplotypes for each individual obtained by using the Phase-Standard analysis version 2.0.2 program (Stephens et al. 2001; Stephens and Donnelly 2003)

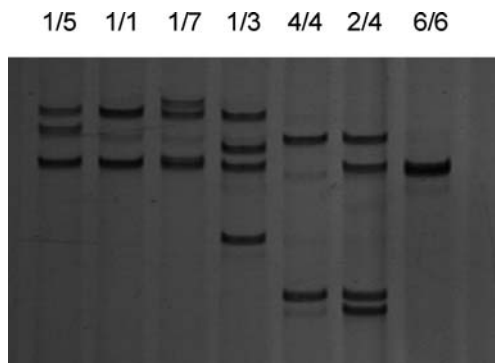


Fig. 2 Polymerase chain reaction–single-strand conformation polymorphism (PCR-SSCP) analyzes for five polymorphic sites at the BP1 binding site. The sample genotype (pair of haplotypes) is shown at the top of the lane (the numbers from 1 to 7 represent the haplotypes A1 to A7)

detected at $(AT)_x$ whereas only two alleles were observed at other sites. To our knowledge, this is the first study reporting an $(AC)_n$ polymorphism adjacent to the $(AT)_x(T)_y$ repeat.

The frequencies of haplotypes consisting of both HbA/HbE polymorphism and SSCP haplotype were estimated based on the group of 57 subjects (Table 2). Haplotypes with low estimated frequencies may not exist in the studied subjects (e.g., HbE—A2) because the results are based on statistical inference from a small sample. SSCP haplotypes A1, A2, A4, and A6 were common, and the $(AT)_9(T)_5$ and $(AT)_7(T)_7$ alleles were predominant, in the Thai population. The HbE variant was in strong linkage disequilibrium with haplotype A1 (i.e., $(AT)_9(T)_5$ allele) whereas HbA was on various chromosomes with different SSCP haplotypes.

Discussion

We investigated the $(AT)_x(T)_y$ repeat on HbE and HbA chromosomes in the Thai population. The $(AT)_9(T)_5$ allele is thought to associate with decreased β chain synthesis (Semenza et al. 1984; Murru et al. 1990; Elion

et al. 1992) and affect the HBB expression in an erythroid environment (Kalotychou et al. 2002). Our results show a strong linkage disequilibrium of HbE with $(AT)_9(T)_5$, implying that this repeat may lead to the low concentration of β^E chain in individuals with HbE. However, some results contradict this prospect. For example, the expression level of β chain in normal individuals is not significantly affected by a polymorphism in the $(AT)_x(T)_y$ repeat (Galanello et al. 1993), and phenotypic diversity is not associated with $(AT)_9(T)_5$ (Wong et al. 1989). Because it is known that the low concentration of β^E chain can be attained by the aberrant splicing of HbE mRNA (Orkin et al. 1982), further functional studies are needed to determine whether the $(AT)_9(T)_5$ repeat influences HbE expression.

Hemoglobin E was found to be in strong linkage disequilibrium with $(AT)_9(T)_5$ in this study. The same haplotype has been reported in Chinese (Zhou et al. 1995). Thus, HbE variants in the studied Thai population and Chinese may have the same origin (i.e., the HbE mutation may have occurred on a chromosome containing the $(AT)_9(T)_5$ repeat and rapidly spread into Southeast Asia) although HbE variants with different origins are suggested to exist in Thailand (Fucharoen et al. 2002). Assuming that the present relative allele frequencies of the $(AT)_x(T)_y$ repeats on HbA chromosomes correspond to the frequencies of these repeats at the time of HbE mutation, $(AT)_9(T)_5$ would have been one of the dominant alleles. Although the functional effect of this repeat on the HbE chromosome still remains unclear, HbE may not have been such a common variant in Southeast Asia if the HbE mutation occurred on a chromosome with another repeat not affecting β -chain production.

We found the $(AC)_n$ polymorphism adjacent to the $(AT)_x(T)_y$ repeat in the Thai population. The frequency of $(AC)_3(AT)_7(T)_5$ was low in Thai, while the frequency was 11.3% in Japanese (Our unpublished data). Thus, the $(AC)_3(AT)_7(T)_5$ allele would have arisen in common ancestors of Asian populations. Because the $(AC)_n$ polymorphism seems to affect the binding affinity of BP1 (Elion et al. 1992), it may reveal the functional significance of the $(AC)_n(AT)_x(T)_y$ repeat in the regulation of HBB expression by the BP1 repressor. Because the allele frequency of $(AC)_3(AT)_7(T)_5$ is approximately 11.3% in Japanese (Our unpublished data), the genotyping of not $(AT)_x(T)_y$ but $(AC)_n(AT)_x(T)_y$ is required, at least in Asians, in the future. The present PCR-SSCP analysis allowed us to distinguish seven haplotypes including the $(AC)_n(AT)_x(T)_y$ repeat. Furthermore, some additional Japanese haplotypes different from these seven could be also distinguished (data not shown). Thus, although it is necessary to initially perform direct PCR-based sequencing for the reference bands cut out of the SSCP gels, the PCR-SSCP method is very useful for rapidly and simply determining the phase or pair of haplotypes in a subject.

The BP1 binding site has been predicted to be part of a putative recombination hotspot located upstream of

the HBB coding sequence (Chakravarti et al. 1984; Smith et al. 1998; Schneider et al. 2002; Wall et al. 2003). To examine whether the BP1 site is included in the hotspot, a larger sample group or population must be studied. The PCR-SSCP method presented here should be helpful for such a study because it can be used to determine the phase, or the combination of haplotypes, of each sample (Wall et al. 2003).

Acknowledgements We are grateful to the patients who participated in this study. This study was partly supported by the Core University System Exchange Programme under the Japan Society for the Promotion of Science, coordinated by the University of Tokyo and Mahidol University, The National Research Council of Thailand and Mahidol University Grant, a Grant-in-Aid for Young Scientists (B) and Scientific Research on Priority Areas (JC) "Medical Science") from the Ministry of Education, Culture, Sports, Science and Technology, Japan, The Genetic Diversity Project supported by the New Energy and Industrial Technology Development Organization (NEDO), and NIH D43TW00620/CFDA#93.989.

References

- Antonarakis SE, Orkin SH, Kazazian HH Jr, Goff SC, Boehm CD, Waber PG, Sexton JP, Ostrer H, Fairbanks VF, Chakravarti A (1982) Evidence for multiple origins of the beta E-globin gene in Southeast Asia. *Proc Natl Acad Sci U S A* 79:6608–6611
- Berg PE, Williams DM, Qian RL, Cohen RB, Cao SX, Mittelman M, Schechter AN (1989) A common protein binds to two silencers 5' to the human beta-globin gene. *Nucleic Acids Res* 17:8833–8852
- Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human beta-globin gene cluster. *Am J Hum Genet* 36:1239–1258
- Chase MB, Fu S, Haga SB, Davenport G, Stevenson H, Do K, Morgan D, Mah AL, Berg PE (2002) BP1, a homeodomain-containing isoform of DLX4, represses the beta-globin gene. *Mol Cell Biol* 22:2505–2514
- Elion J, Berg PE, Lapoumeroulie C, Trabuchet G, Mittelman M, Krishnamoorthy R, Schechter AN, Labie D (1992) DNA sequence variation in a negative control region 5' to the beta-globin gene correlates with the phenotypic expression of the beta s mutation. *Blood* 79:787–792
- Fucharoen G, Fucharoen S, Sanchaisuriya K, Sae-Ung N, Suyasanonond U, Sriwilai P, Chinorak P (2002) Frequency distribution and haplotypic heterogeneity of beta(E)-globin gene among eight minority groups of Northeast Thailand. *Hum Hered* 53:18–22
- Galanello R, Meloni A, Gasperini D, Saba L, Cao A, Rosatelli MC, Perseu L (1993) The repeated sequence (AT)x(T)y upstream to the beta-globin gene is a simple polymorphism. *Blood* 81:1974–1975
- Grosveld F, Dillon N, Higgs D (1993) The regulation of human globin gene expression. *Baillieres Clin Haematol* 6:31–55
- Kalotychoy V, Kollia P, Voskaridou E, Patargias T, Anagnou NP, Loukopoulos D (2002) Functional role of the four different types of (AT)(x)T(y) motifs 5' to the beta-globin gene and their distribution in the Greek population. *Blood Cells Mol Dis* 28:39–46
- Murru S, Loudianos G, Cao A, Vaccargiu S, Pirastu M, Sciaratta GV, Agosti S, Parodi MI (1990) A beta-thalassemia carrier with normal sequence within the beta-globin gene. *Blood* 76:2164–2165
- Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Looareesuwan S, Tokunaga K (2002) Significant association of longer forms of CCTTT microsatellite repeat in the inducible nitric oxide synthase promoter with severe malaria in Thailand. *J Infect Dis* 186:578–81
- Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, Looareesuwan S, Clark AG, Tokunaga K (2004) Extended linkage disequilibrium surrounding the hemoglobin e variant due to malarial selection. *Am J Hum Genet* 74:1198–1208
- Orkin SH, Kazazian HH Jr, Antonarakis SE, Ostrer H, Goff SC, Sexton JP (1982) Abnormal RNA processing due to the exon mutation of beta E-globin gene. *Nature* 300:768–769
- Schneider JA, Peto TE, Boone RA, Boyce AJ, Clegg JB (2002) Direct measurement of the male recombination fraction in the human beta-globin hot spot. *Hum Mol Genet* 11:207–215
- Semenza GL, Delgrosso K, Poncz M, Malladi P, Schwartz E, Surrey S (1984) The silent carrier allele: beta thalassemia without a mutation in the beta-globin gene or its immediate flanking regions. *Cell* 39:123–128
- Smith RA, Ho PJ, Clegg JB, Kidd JR, Thein SL (1998) Recombination breakpoints in the human beta-globin gene cluster. *Blood* 92:4415–4421
- Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Traeger J, Wood WG, Clegg JB, Weatherall DJ (1980) Defective synthesis of HbE is due to reduced levels of beta E mRNA. *Nature* 288:497–499
- Wall JD, Frisse LA, Hudson RR, Di Rienzo A (2003) Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. *Am J Hum Genet* 73:1330–1340
- Wasi P, Pootrakul S, Pootrakul P, Pravatmuang P, Winichagoon P, Fucharoen S (1980) Thalassemia in Thailand. *Ann N Y Acad Sci* 344:352–363
- Wong SC, Stoming TA, Efremov GD, Huisman TH (1989) High frequencies of a rearrangement (+ATA; -T) at -530 to the beta-globin gene in different populations indicate the absence of a correlation with a silent beta-thalassemia determinant. *Hemoglobin* 13:1–5
- Zhou G, Chen MJ, Ren ZR, Huang SZ, Zeng YT (1995) Patterns of the (AT)xTy motif at the -530 region 5' to the beta-globin gene in the Chinese population. *Hemoglobin* 19:311–316