Int J Hum Genet, 17(3): 97-102 (2017)
DOI:10.31901/24566330.2017/17.03.01

Novel Tag SNPs of Beta-Globin Gene Cluster in Chinese Han Population: Biological Marker for Genetic Backgrounds and Clinical Studies

Wittaya Jomoui

Department of Pathology, Maha Chakri Sirindhorn Medical Center, Faculty of Medicine, Srinakharinwirot University, Ongkharak, Nakhon Nayok, Thailand

KEYWORD Beta-thalassemia. Bioinformatics Tool. Genetics. Linkage Analysis. Polymorphism

ABSTRACT The beta-globin gene cluster is located on chromosome 11. SNPs database was published by Hap Map Project. Here, novel tag SNPs on beta -globin gene cluster in Chinese Han population was reported on using Haploview 4.2. The haplotype block and LD analysis showed three blocks that were constructed with 35 common SNPs. All common SNPs were generated with the Tag SNPs using Haploview's Tagger with LD threshold r2 e" 0.8, run with aggressive tagging; 2- and 3 marker haplotypes and default LOD cutoff 3.0. The results showed that seven SNPs capture all 35 common SNPs. Seven tag SNPs including rs4910736, rs5010979, rs4910548, rs11036634, rs7480526, rs10837628, rs11036635 could be used to predict the genetic background of beta-globin gene defect or clinical correlation with beta-thalassemia in a Chinese or Asian population. Further study should test these tag SNPs for genetic background and a clinical relation between the biological marker and beta-thalassemia disease.

INTRODUCTION

The beta-globin gene cluster is located on the short arm chromosome 11 in humans (11p15.5). It spans approximately 65 kb which consists of the locus control region (LCR) and six linked genes, including epsilon-globin gene, G gamma-globin gene , A gamma-globin gene, pseudo beta-globin gene, delta-globin gene and beta-globin gene (Haussecker and Proudfoot 2005; Moleirinho et al. 2013). Moreover, a recombination hot spot was found between the delta-globin gene and beta-globin gene, covering 9.1 kb in length (Currat et al. 2002). Beta-thalassemia is the result a decrease (beta^{+} -thalassemia) or absence of production of beta globin chains (beta^{0} -thalassemia). It is a common disorder in South China and Southeast Asia. More than 200 beta-thalassemia cases related to point mutation, while deletion types were rarely found (Thein et al. 1990; Rund and Rachmilewitz 2005). Beta-thalassemia has been found in malaria regions of the world. These mutations have shown population specific distributions (Kuesap et al. 2015). In Southern China, where haemoglobinopathies are the most endemically reported in China, the frequency of beta -thalassemia is about 6.43 percent. In Thailand, a frequency ranging between 3-9 percent has been reported for beta-thalassemia (Fucharoen and Winichagoon 1992; Xiong et al. 2010).

A haplotype is a type of haploid chromosome related to the gene of interest. Alternatively, it is the pattern of chromosome background usually constructed using multiple DNA polymorphisms within the chromosome (Delaneau and Marchini 2014). In the past, seventeen polymorphisms in beta-globin gene cluster were used to study population genetics. Clinical relation with beta-thalassemia and seven common restriction sites were constructed haplotypes.

Hap Map Project was launched in 2002. Many SNPs databases have been published for each chromosome (Frazer et al. 2007). SNPs database in beta globin gene cluster showed more than seventy polymorphisms. It seems that the new database for SNPs in beta-globin gene cluster makes it useful for population genetics or clinical relation studies for beta-thalassemia.

Address for correspondence:
Dr. Wittaya Jomoui
Department of Pathology, Faculty of Medicine,
Maha Chakri Sirindhorn Medical Center,
Srinakharinwirot University,
62 Moo 7 Rungsit-Nakhon Nayok Rd., Ongkharak,
Nakhon Nayok 26120, Thailand
Telephone: +6637 395085, *Extn:* 10338
E-mail: paper_zaa@hotmail.com

However, it is expensive and time-consuming to genotype every SNPs in beta-globin gene cluster in large-scale studies. Thus, it is highly critical to select an appropriate subset of SNPs in order to identify the remaining SNPs accurately. The selection of SNPs is designated as known-tag SNPs. In practice, choosing tag SNPs is based on linkage disequilibrium (LD) criteria. A number of algorithms are available to assess LD between SNPs and selected tag SNPs representative of groups of correlated SNPs (Sicotte et al. 2011; Ilhan et al. 2013). This paper would like to represent the potential biological marker as known genetics variants within the beta-globin gene cluster using the bioinformatics analysis to select the best tag SNPs. The model from The Chinese population is used for further studies about the genetics background of beta-thalassemia.

Objective

To determine novel tag SNPs on beta-globin gene cluster in the Chinese Han population using Haploview 4.2. for genetic background and clinical studies.

METHODOLOGY

Selection of SNPs within Beta-globin Gene Cluster

The SNPs within beta-globin gene cluster is located on the short arm chromosome 11 of humans (11p15.5). It spans approximately 70 kb, which consists of the locus control region (LCR). SNPs for 84 healthy Chinese Han were downloaded from the Hap Map (www.hapmap.org) Project; Version 3, Release R2, Chr 11: 5200..5270 kb using HAPLOVIEW v. 4.2 program with default setting; individuals with >50 percent missing genotypes were excluded and pairwise comparisons of markers more than 500 kb were ignored.

Construction of Haplotype Block and Linkage Disequilibrium Test

Common SNPs (MAF> 0.05 and Hardy Weinberg p-value cutoff 0.01) except SNPs within recombination hot spot between beta-globin and delta globin were constructed with haplotype block using Haploview, Version 4.2 (Broad Institute of MIT and Harvard, Cambridge, Mass), which is a software package that calculate linkage disequilibrium (LD) statistics and population patterns from the genotypes of SNPs. The default algorithm was taken from Gabriel et al. 2002. The D' value was generated with 95 percent confidence bounds, with each comparison called "strong LD", "inconclusive" or "strong recombination" (Gabriel et al. 2002).

Tag SNPs Analysis

Haploview's Tagger operates in either pairwise or aggressive mode. In either case, it begins by selecting a minimal group of markers that related to all alleles are correlated at an r^2 greater than the threshold with a marker in that set (Gabriel et al. 2002). All common SNPs were generated for the Tag SNPs using Haploview's Tagger with LD threshold $r^2 \geq 0.8$, run with aggressive tagging; 2- and 3 marker haplotypes and default LOD cutoff 3.0.

RESULTS

Common SNPs in Beta-globin Cluster Profile

The data SNPs of 84 healthy Chinese Han were downloaded from the HapMap (www.hapmap.org) Project; Version 3, Release R2, Chr 11: 5200…5270 kb using HAPLOVIEW v. 4.2 programs with the default setting. Overall, seventy-one SNPs were found in the analysis. However, only 35 SNPs showed MAF>0.05, which indicated that these SNPs were common in a population. P-values > 0.05 were found to be in Hardy-Weinberg Equilibrium as shown in Table 1.

Haplotype Block Analysis and LD Analysis

Three blocks were constructed that related to the common SNPs (as shown in Fig. 1 and Table 2); block 1 (3 kb) included 5 SNPs (rs12364872, rs10837628, rs11036351, rs1609812, rs7480526), block 2 (35 Kb) included 23 SNPs (rs7948416, rs7948668, rs3759074, rs10837643, rs4320977, rs4283007, rs4910736, rs2105819, rs10768687, rs2071348, rs10488676, rs6578592, rs11036455, rs10768707, rs11036474, rs11036476, rs2855123, rs2855122, rs5010979, rs5010981, rs4348933, rs3759071, rs3759067), block 3 (18 kb) included 7 SNPs (rs4910548, rs4910741, rs4910742, rs11036634, rs11036635, rs11036644, rs4601817). Globin gene in the cluster were separated in to three blocks, including; beta globin

Table 1: Distribution of common SNPs covered beta globin gene cluster in Chinese Han population

No. SNPs	*SNPs ID*	*Position*[*]	*HWpval*[**]	*MAF*[***]	*Alleles(A:B)*[****]
1	rs12364872	5200720	0.4183	0.179	A:G
2	rs10837628	5200980	1	0.327	A:G
4	rs11036351	5202576	0.0844	0.179	C:T
5	rs1609812	5203717	0.2075	0.494	A:G
7	rs7480526	5204309	0.4183	0.179	A:C
28	rs7948416	5213007	1	0.101	C:A
29	rs7948668	5213223	0.9435	0.262	A:G
31	rs3759074	5214354	1	0.167	A:G
33	rs10837643	5214614	1	0.28	T:A
34	rs4320977	5214738	0.8267	0.268	A:G
35	rs4283007	5215066	0.8267	0.268	T:A
36	rs4910736	5215565	0.8267	0.268	C:A
37	rs2105819	5216303	0.8267	0.268	C:G
40	rs10768687	5217815	0.6064	0.28	G:C
42	rs2071348	5220722	1	0.167	T:G
43	rs10488676	5225373	1	0.268	G:A
44	rs6578592	5225716	0.722	0.107	A:C
46	rs11036455	5226994	1	0.268	A:G
47	rs10768707	5230498	0.8136	0.101	C:T
48	rs11036474	5231754	1	0.167	T:C
49	rs11036476	5231919	1	0.274	C:T
52	rs2855123	5233654	1	0.274	A:T
53	rs2855122	5233812	1	0.274	T:C
54	rs5010979	5235729	0.8136	0.101	C:T
55	rs5010981	5235931	0.8136	0.101	A:C
56	rs4348933	5241554	1	0.268	A:G
57	rs3759071	5248108	1	0.274	G:A
58	rs3759067	5248768	0.722	0.107	T:C
59	rs4910548	5250977	1	0.071	C:T
66	rs4910741	5260908	1	0.06	C:A
67	rs4910742	5263085	1	0.071	A:G
68	rs11036634	5265304	0.2216	0.286	C:T
69	rs11036635	5265472	0.3426	0.22	G:A
70	rs11036644	5267639	0.5069	0.286	A:T
71	rs4601817	5268977	0.4166	0.292	A:G

[*]Position in chromosome 11 from database of NCBI
[**]P-value > 0.05 were found to be in Hardy-Weinberg Equilibrium
[***]MAF is minor allele frequency.
[****]Allele on positive strand, character A is ancestral allele and character B is derived allele

gene located on block 1, delta-globin gene, A gamma-globin gene, G gamma-globin gene and epsilon-globin gene located on block 2, and locus control region (LCR) located on block 3. A total of 35 common SNPs were contained in three blocks which is 100 percent of the selected common SNPs. The average of linkage disequilibrium, r^2 value and LOD value between SNPs in each block was calculated as shown in Table 2. The average LD in each block showed a high value (>90%).

Tag SNPs Analysis

Tag SNPs analysis (setting; LD threshold $r^2 \geq 0.8$, run with aggressive tagging; 2- and 3 marker haplotypes and default LOD cutoff 3.0.) showed seven SNPs that could capture all 35 common SNPs in this study. The seven tag SNPs included rs4910736 capture 14 SNPs, rs5010979 capture 6 SNPs, rs4910548 capture 3 SNPs, rs11036634 capture 3 SNPs, rs7480526 capture 3 SNPs, rs10837628 capture 1 SNP, rs11036635 capture 1 SNP, AA pattern of rs7480526, rs10837628 capture 1 SNP, and CCA pattern of rs11036634, rs4910736, rs11036635 capture 3 SNPs. Taken together, a total of seven tag SNPs selected in the present paper could capture most SNPs within the beta-globin gene cluster and might represent possible biological significance in the beta-globin gene variation (see Table 3).

Table 2: The average of linkage disequilibrium, r^2 value and LOD value between SNPs in the blocks of beta globin gene cluster in Chinese Han population

Block	*SNPs ID*	*Average*				
		r^2	*LOD*	*D'*	*CI low**	*CI high***
1	rs12364872, rs10837628, rs11036351, rs1609812, rs7480526	**0.41**	**15.56**	**0.97**	**0.79**	**0.96**
2	rs7948416, rs7948668, rs3759074, rs10837643, rs4320977, rs4283007, rs4910736, rs2105819, rs10768687, rs2071348, rs10488676, rs6578592, rs11036455, rs10768707, rs11036474, rs11036476, rs2855123, rs2855122, rs5010979, rs5010981, rs4348933, rs3759071, rs3759067	**0.57**	**20.89**	**0.96**	**0.77**	**0.99**
3	rs4910548, rs4910741, rs4910742, rs11036634, rs11036635, rs11036644, rs4601817	**0.38**	**11.70**	**0.91**	**0.61**	**0.96**

* The lower 95 percent confidence bound of the D' value
** The upper 95 percent confidence bound of the D' value

Table 3: Tag SNPs within beta globin gene cluster of Chinese Han population

Test	*No. SNPs*	*SNPs captured*
rs4910736	14	rs2105819, rs10837643, rs4320977, rs10768687, rs3759071, rs4910736, rs4348933, rs11036476, rs2855123, rs11036455, rs7948668, rs2855122, rs10488676, rs4283007
rs5010979	6	rs7948416, rs10768707, rs3759067, rs5010981, rs6578592, rs5010979
rs4910548	3	rs4910742, rs4910548, rs4910741
rs11036634	3	rs11036634, rs4601817, rs11036644
rs7480526	3	rs11036351, rs12364872, rs7480526
rs10837628	1	rs10837628
rs11036635	1	rs11036635
rs7480526, rs10837628: AA	1	rs1609812
rs11036634, rs4910736, rs11036635: CCA	3	rs2071348, rs11036474, rs3759074

DISCUSSION

Beta-thalassemia is a major public health problem in many countries, especially Southern China and Southeast Asia. Previous studies have reported about 0.5 – 6 percent for beta-thalassemia carriers in China; Guangxi province, Guangdong province, Hong Kong, Taiwan, and about 0.5 – 10 percent for beta-thalassemia carriers in Southeast Asian countries including Thailand, Myanmar, Malaysia, Singapore, Indonesia, Vietnam, Cambodia, and Laos (Fucharoen and Winichagoon 1992; Xiong et al. 2010). The cause of beta-thalassemia is the reduced or absent synthesis of the beta-globin chain, which is a component of the hemoglobin structure (tetramer) (Thein et al. 1990; Rund and Rachmilewitz 2005). Homozygous state of beta-thalassemia or compound heterozygous may represent either thalassemia major or thalassemia intermedia. Newborns with thalassemia major require regular blood transfusions for the first 2 years to survive. The survival of patients with well transfused and treated as well as appropriate chelation extends beyond the age of 30 years. The severity of thalassemia major to beta- thalassemia carrier ranges widely (Galanello and Origa 2010).

Previous methods for the detection of single nucleotide polymorphism (SNP) were first discovered in 1980 using restriction enzymes to recognise the presence or absence of polymorphic restrictions (Schork et al. 2000). Although more than 60 polymorphic restriction sites have been described in beta-globin gene cluster, 7-8 polymorphisms are commonly selected and used in haplotype study. The International HapMap Project was launched in 2002 with the aim of providing a public resource to accelerate medical genetic research (Frazer et al. 2007). This project is useful for haplotype or linkage analysis using bioinformatics tools. Recently, high-

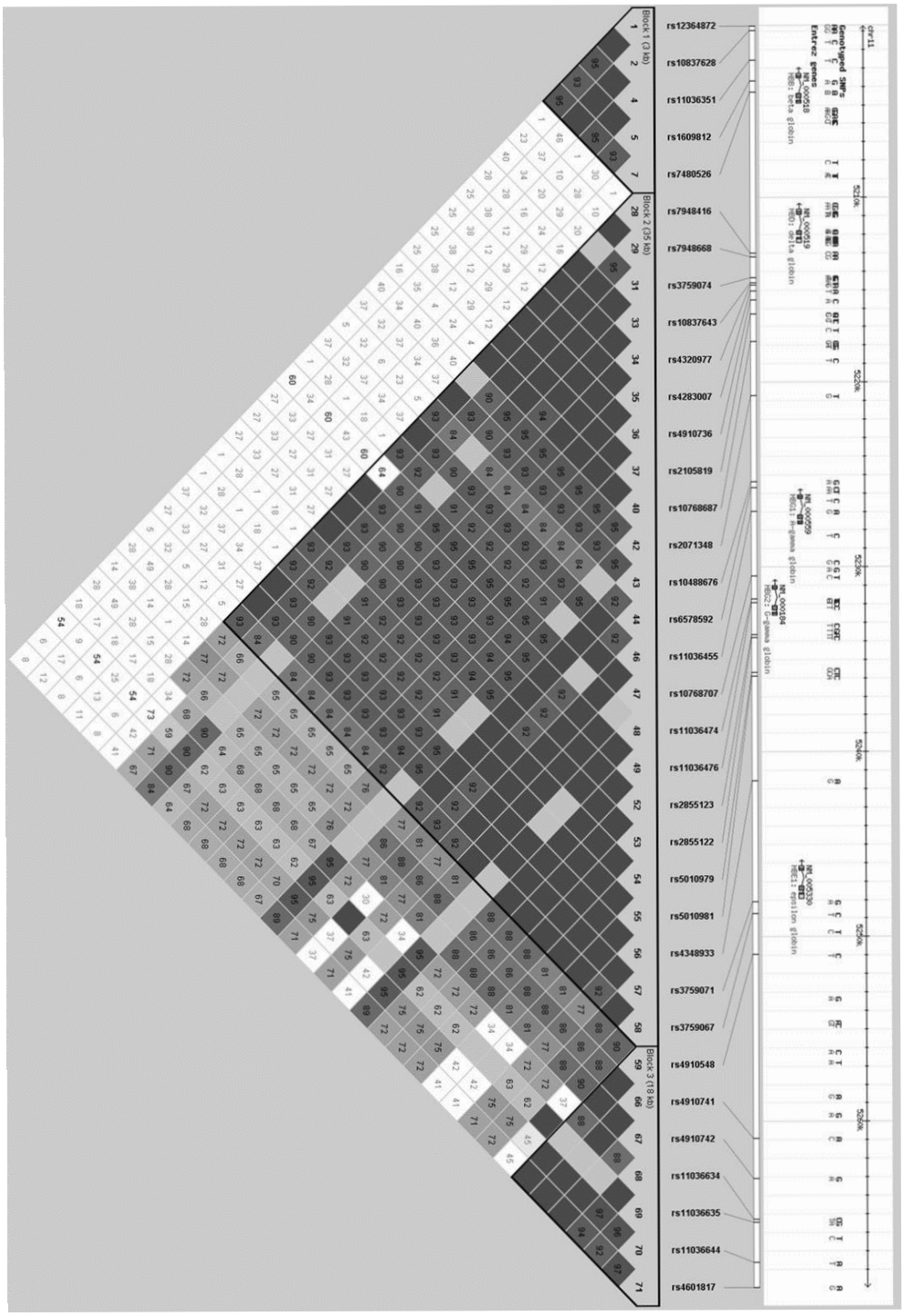

Fig.1. Beta globingene cluster map and LD plot of common SNPs in Chinese Han population. The D' values were indicated with numbers in squares. The LOD score (dark shade, D'=1 and LOD score e"3; lighter shades, D'< 1 and LOD score <3) indicated by color scheme

throughput SNP genotyping has been required for LD mapping. The most common SNP genotyping chemistries available are hybridisation, primer extension, and cleavage methods. LD is an important indicator for related population genetic studies and clinical association studies as well as the association of inherited disease with polymorphism. Furthermore, LD in each genomic region can also be used to describe the history of gene conversion, natural selection, and mutation (Haussecker and Proudfoot 2005).

Polymorphism within the beta-globin gene cluster should be revisited and updated for population genetics studies and clinical association studies. In recent decades, the evolutionary constraints of the beta-globin cluster have been reported. Comprehensive analysis, based on classic neutrality tests, empirical and haplotype-based studies, revealed that the delta-globin gene and its neighbour pseudogene have mainly evolved under purifying selection (Moleirinho et al. 2013). In this paper, evidence was revealed with the result of linkage within block 2.

This paper has shown haplotype block and LD analysis as shown in Tables 1 and 2 and Figure 1. The 35 SNPs of beta-globin gene cluster obtained from CHB a population showed MAF>0.05 which indicates that these SNPs are common in a population and associated as three blocks, as first reported. The average of linkage disequilibrium, r^2 value and LOD value between SNPs in the blocks with CI low and CI high were shown. As a result, bioinformatics analysis with Tag SNP was selected based on r^2 and LOD value between SNPs. Seven SNPs could capture all 35 common SNPs within beta-globin gene cluster in this paper. These SNPs can be used to predict the genetic background of beta-globin gene defects or clinical correlation with beta-thalassemia in a Chinese or Asian population. The lowest LD values between SNPs in block 1 and block 2, resulting from a recombination hot spot between delta and beta-globin gene were represented (see Fig. 1).

CONCLUSION

Use of the bioinformatics method to select tag SNPs in this paper may provide an effective way to select tag SNPs in an entire beta-globin gene cluster.

RECOMMENDATIONS

Further study is needed to test these tag SNPs for the genetic background of beta-thalassemia in Chinese and Asian populations. In addition, further guidance is needed concerning the clinical relationship between the biological marker and beta-thalassemia disease susceptibility.

REFERENCES

Currat M, Trabuchet G, Rees D, Perrin P, Harding RM et al. 2002. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. *Am J Hum Genet*, 70: 207-223.

Delaneau O, Marchini J 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun*, 5: 3934.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature,* 449: 851-861.

Fucharoen S, Winichagoon P 1992. Thalassemia in Southeast Asia: Problem and strategy for prevention and control. *Southeast Asian J Trop Med Public Health*, 23: 647–655.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J 2002. The structure of haplotype blocks in the human genome. *Science*, 296: 2225-2229.

Galanello R, Origa R 2010. Beta-thalassemia, *Orphanet J Rare Dis*, 5: 11.

Haussecker D, Proudfoot NJ 2005. Dicer-dependent turnover of intergenic transcripts from the human beta-globin gene cluster. *Mol Cell Biol*, 25: 9724-9733.

Ilhan I, Tezel G 2013. How to select tag SNPs in genetic association studies? The CLONTagger method with parameter optimization. *OMICS*, 17: 368-383.

Kuesap J, Chaijaroenkul W, Rungsihirunrat K, Pongjantharasatien K, Na-Bangchang K 2015. Coexistence of Malaria and Thalassemia in malaria endemic areas of Thailand. *Korean J Parasitol,* 53: 265-270.

Moleirinho A, Seixas S, Lopes AM, Bento C, Prata MJ et al. 2013. Evolutionary constraints in the beta-globin cluster: the signature of purifying selection at the delta-globin (HBD) locus and its role in developmental gene regulation. *Genome Biol Evol*, 5: 559-571.

Rund D, Rachmilewitz E 2005. Beta-thalassemia. *N Engl J Med*, 353: 1135-1146.

Schork NJ, Fallin D, Lanchbury JS 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet*, 58: 250-264.

Sicotte H, Rider DN, Poland GA, Dhiman N, Kocher JP 2011. SNPPicker: High quality tag SNP selection across multiple populations. *BMC Bioinformatics*, 12: 129.

Thein SL, Winichagoon P, Hesketh C, Best S, Fucharoen S et al. 1990. The molecular basis of beta-thalassemia in Thailand: Application to prenatal diagnosis. *Am J Hum Genet*, 47: 369-375.

Xiong F, Sun M, Zhang X, Cai R, Zhou Y et al. 2010. Molecular epidemiological survey of haemoglobinopathies in the Guangxi Zhuang Autonomous Region of southern China. *Clin Genet*, 78: 139-148.

Paper received for publication on August 2017
Paper accepted for publication on September 2017