

The Identification of Functional Non-Synonymous SNP in Human ATP-Binding Cassette (ABC), Subfamily Member 7 Gene: Application of Bioinformatics Tools in Biomedicine

Pornlada Nuchnoi^{1,2*}, Duangdao Nantakomol³, Vasunun Chumchua⁴, Chotiros Plabplueng^{1,2}, Chartchalerm Isarankura-Na-Ayudhya⁵ and Virapong Prachayasittikul^{5*}

¹Department of Clinical Microscopy, Faculty of Medical Technology, Mahidol University, Bangkok, 10700, Thailand

²Center for Innovation Development and Technology Transfer, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand

³Department of Clinical Microscopy, Faculty of Allied Health Science, Chulalongkorn University, Bangkok, 10330, Thailand

⁴National Institute for Child and Family Development, Mahidol University, Nakhonpathom, 73170, Thailand

⁵Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok, 10700, Thailand

Abstract

The prediction of functional single nucleotide polymorphism (SNP) is promising in modern genetics analysis. Computational biology technology has facilitated an increase in the successful rate of genetic association study and reduced the cost of genotyping. In the present study, we applied various bioinformatics tools for the selection of high potentially functional nsSNP and determined the linkage disequilibrium (LD) structure of ATP-binding cassette transporter member 7 (*ABCA7*) genes in HapMap populations. Two functional polymorphisms (rs3752233 and rs3752246) were identified on the basis of less protein stability, a low likelihood of mutability, a changing of protein structure and function. Interestingly, a completed LD between rs3752233 (R463H) and rs4147918 (Q1686R) was detected in Utah residents with ancestry from northern and western Europe (CEU) populations. In addition, the difference of the LD pattern between the populations observed highlighted the essential role of the construction of an LD map for designing and interpreting genetic association study. Studies herein convey the empirical guidelines for conduction of *ABCA7* genetic association study via bioinformatics and computational application.

Keywords: Bioinformatics; Computational biology; *ABCA7*; nsSNP; Linkage disequilibrium

Introduction

The public availability of the human genome sequence and the advance of genotyping technology have facilitated the study of genetic predispositions to various human complex diseases. The most abundant form of human genetic variation is single nucleotide polymorphism (SNP) which is commonly used in genetic association study [1]. The large number of SNPs accumulated in the database at the National Center for Biological Technology (NCBI) provides a great opportunity for mapping loci responsible for phenotypic variation e.g. severity of or susceptibility to complex diseases and variation in pharmacological responses [2,3]. Functional SNPs reported to be associated with human diseases can be categorized into 3 types: 1) regulatory SNP which is located in the regulatory region of genes e.g. the promoter region and 5'- or 3'-untranslated region has been shown to be involved in transcriptional regulation by affecting the putative binding site of transcription factor [4,5] 2) coding SNP which is classified into synonymous SNP (sSNP: no amino acid change) and non-synonymous SNP (nsSNP: causing amino acid change) 3) intronic SNP located in the intron region has been shown to activate cryptic splice sites leading to alternative splicing. It has been revealed that approximately 500,000 SNPs are localized in coding region [6]. The nsSNP is the most interesting owing to the direct effect on protein structure and function as shown in various diseases e.g. type 2 diabetes [7], essential hypertension [8], and malarial infection [9]. Unsurprisingly, the most effective way of assessing the role of candidate gene resulting in related diseases is by focusing on the region of the gene containing nsSNPs. With the hypothesis that nsSNP may directly affect protein structure stability and the efficiency of protein interactions resulting in pathological conditions [6], however, many nsSNPs show no evidence of biological involvement with various conditions [10,11,12]. Therefore, a closer investigation of the polymorphism in question by in silico analysis is needed prior

to the conducting of phenotype/genotype association study. The advantages of the in silico analysis of candidate genes for prediction functional nsSNP are the increasing of the successful rate of genetic association study and the reduction of the cost of genotyping. Moreover, the biological explanation for disease associated gene could be provided using computational biology.

ATP-binding cassette (ABC) transporter is a superfamily of highly conserved membrane protein which transports a variety of macromolecules such as amino acid, peptide, sugar and lipid across cell membranes. ABC member 7 (*ABCA7*) is a novel transporter protein mediating the efflux of phospholipids to apolipoprotein 1. The 24-kb genomic DNA of *ABCA7* consists of 46 exons on chromosome 19p13.3 region. A 220-kD *ABCA7* protein was reported to be expressed in a variety of cells such as macrophage [13], retinoblastoma cell line [14]. The function of *ABCA7* is proposed to be involved in the engulfment of apoptotic cells [13] and lipid transporters in the human brain [15]. However, little is known about *ABCA7* function and human disease.

***Corresponding authors:** Dr. Pornlada Nuchnoi, Department of Clinical Microscopy, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand 10700, Tel:+66-2-4414370 ext 2724; Fax:+66-2-441-4380; E-mail: mtpnn@mahidol.ac.th

Dr. Virapong Prachayasittikul, Professor, Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok, 10700, Thailand, Tel:+66-2-4414370; Fax :+66-2-441-4380; E-mail: mtvpr@mahidol.ac.th

Received December 07, 2010; **Accepted** December 23, 2010; **Published** January 07, 2011

Citation: Nuchnoi P, Nantakomol D, Chumchua V, Plabplueng C, Isarankura-Na-Ayudhya C, et al. (2011) The Identification of Functional Non-Synonymous SNP in Human ATP-Binding Cassette (ABC), Subfamily Member 7 Gene: Application of Bioinformatics Tools in Biomedicine. J Bioanal Biomed 3: 026-031. doi:10.4172/1948-593X.1000039

Copyright: © 2011 Nuchnoi P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

As mentioned above the role of *ABCA7* highlights a significant contribution in the discovery of the deleterious nsSNP responsible for human diseases.

The aims of this study are to distinguish *ABCA7* functional nsSNP from *ABCA7* neutral non-functional SNP in order to prioritize the proper marker prior to conducting genetic association study, functional analysis and to describe a linkage disequilibrium (LD) map for SNP selection. We therefore applied bioinformatics tools for analyzing the functional impact of *ABCA7* nsSNP deposited in a public database. We performed computational algorithm tools namely Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotype (PolyPhen), I-Mutant 2.0, Functional Analysis and Selection Tool for Single Nucleotide Polymorphisms (FASTSNP) to identify putative functional nsSNP that have a high possibility of affecting protein structure, function, and subsequently the cellular function. We identified the putative functional *ABCA7* nsSNPs as a functional marker and provided useful information for designing and interpreting the results of genetic association study by analyzing the linkage disequilibrium structure across 4 multi-ethnic populations in HapMap project.

Methods

SNP mining

The SNP data set of *ABCA7* gene was retrieved from dbSNP of NCBI build 129 (<http://www.ncbi.nlm.nih.gov/SNP>) and the International HapMap project (<http://www.hapmap.org>, phase 2 public release up to February 2009) for our computational analysis. It should be noted that the fraction of the DNA sequence deposited in the database could have resulted from sequencing errors or paralogous sequence variant. This makes artifacts in SNP depository database. To obtain true SNP for genetic association study, it is necessary to exclude SNP entries that show a high possibility of being artifacts or monomorphic in the HapMap population. The criteria for selecting proper polymorphism for our analysis were the following 1) Using the Basic Local Alignment Search Tool (BLAST) available at NCBI (<http://blast.ncbi.nlm.nih.gov/>) against the human genome, a flanking sequence of SNP that can be mapped to multiple regions of the genome were then filtered out as unreliable SNP 2) a SNP with supporting information as validated SNP by NCBI or SNP with reported allele/genotype frequency from HapMap as well as multiple submission SNP with independent submission was considered to be truly polymorphic. True polymorphism with minor allele frequency >5 % was regarded as common polymorphism. The alleles of polymorphism throughout the study were designated as ancestral/derived allele.

Analysis of the functional consequence of nsSNP by sequence homology based tool

In order to detect a deleterious nsSNP, we used the SIFT software (15). SIFT, Sorting Intolerant from Tolerant, is a sequence homology based tool. A deleterious nsSNP was detected based on the assumption that an amino acid position that has a significant biological function will be conserved throughout the evolutionary time. Amino acid substitution (AAS) at well conserved amino acid position was predicted as deleterious. We submitted a batch of *ABCA7* nsSNPs in the form of SNP ID. The SIFT predicted whether submitted *ABCA7* nsSNP affected protein function based on sequence homology and amino acid properties. The cut-off value was a tolerance index of ≥ 0.05 . The higher the tolerance index, the lesser the functional impact of a particular amino acid substitution. Hence, the calculating score represented the likelihood of mutability at the site of AAS.

Analysis of protein stability change upon single point amino acid substitution based on a support vector machine

We predicted nsSNP causing protein stability change using I-Mutant 2.0 software [16]. I-Mutant 2.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability change upon single amino acid substitution. The software was trained on a data set derived from ProTherm [17] which is presently the most comprehensive database of thermodynamic experimental data of free energy changes in protein and protein mutation. The protein stability change was predicted from the *ABCA7* protein sequence (NP_061985). The software computed the predicted free energy change value or sign (DDG) which is calculated from the unfolding Gibbs free energy value of the mutated protein minus unfolding Gibbs free energy value of the native protein (kcal/mol). A positive DDG value indicates that the mutated protein possess high stability and vice versa. A high reliability index (RI) is also important for interpreting the output data.

Simulation of functional change in nsSNP by a sequence homology based tool

We predicted a damaging nsSNP at the structural level using PolyPhen software [18]. PolyPhen is an automatic tool for the prediction of the possible impact of amino acid substitution on the structure and function of a human protein. We submitted the *ABCA7* protein sequence (NP_061985) together with position of AAS and two amino acid variants. The software analyzes the impact of nsSNP by mapping the AAS on the protein 3D structure to explore whether the AAS is likely to destroy the hydrophobic core of the protein, the electrostatic interaction and other important features of protein. PolyPhen uses empirically derived rules for predicting such a SNP. A nsSNP with high confidence is supposed to affect the protein structure and the function is assigned as "Probably damaging". nsSNP is supposed to affect protein structure and the function is assigned as "Possibly damaging". A benign nsSNP is likely lacking in any phenotypic effect. PolyPhen also calculates position-specific independent (PSIC) scores for each of the two variants and then computes the PSIC score difference between them. The higher the PSIC score difference, the higher is the possible functional impact of a particular AAS.

Analysis of functional nsSNPs and estimation of risk score by FASTSNP

The Functional Analysis and Selection Tool for Single Nucleotide Polymorphism (FASTSNP) is a web wrapper agent which connects many softwares and databases for processing analysis [19]. We used FASTSNP for the prediction of the functional effect of nsSNPs and an estimation of their risk score. The FASTSNP uses a decision tree for prioritizing the functional effect and estimating risk score. The eight nsSNPs were submitted for FASTSNP analysis and output files were displayed as a decision tree.

Structural feature analysis

TMHMM version 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) [20] and TopPred II (<http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred>) [21] were applied to predict transmembrane (TM) segment of *ABCA7* topology and to analyze other features of these membrane proteins based on the Hidden Markov Model and hydrophobicity analysis and a charge-biased analysis, respectively. The correlation measurement between the two different methods was conducted by methods agreement consensus on their prediction and loop orientation of all the prediction coincided.

Linkage disequilibrium (LD) analysis

To evaluate the LD structure of the *ABCA7* gene, we calculated pairwise LD coefficient (D') and r^2 . The Haploview software was applied to calculate LD statistics and visualize the LD structure and [22].

Results

SNP dataset

The nsSNPs of the *ABCA7* gene investigated in this study were retrieved from dbSNP. [23] Two hundred and thirty one SNPs of the *ABCA7* gene were deposited in dbSNP until build 129. Of the coding substitution, 5 (14.71 %) were frame ship substitutions, 11 (32.35 %) were non-synonymous SNPs and 18 (52.94 %) were synonymous SNPs. We classified eight of eleven reported nsSNPs in dbSNP as validated nsSNPs in our study. A true validated nsSNPs was designated when SNP show a localization at the *ABCA7* gene and data of genotype/allele frequency by HapMap project information or multiple-independent submissions could be obtained. The genotype/allele frequency data of *ABCA7* nsSNPs from the HapMap project was preferentially accepted as a high reliable measurement because of information on the multi-ethnic population across the Asia, Europe and the Africa continent. Seven nsSNPs deposited in dbSNP were reported by HapMap and one nsSNP was reported by Perlegen and Japan Biological Informatics Consortium (JBIC). The derived allele frequency of eight nsSNPs from 4 ethnic populations including CHB: Han Chinese in Beijing, China, JPT: Japanese in Tokyo, Japan, CEU: CEPH (Utah residents with ancestry from northern and western Europe, YRI: Yoruba in Ibadan, Nigeria are shown in Table 1. A total of 8 validated nsSNPs were therefore selected for investigating the effect of *ABCA7* nsSNPs on protein structure and function.

Identification of deleterious nsSNPs by sequence-based amino acid substitution prediction method SIFT software, a sequence homology based tool was used for detecting deleterious nsSNP of the *ABCA7* gene. SIFT predicts whether amino acid substitution affects protein function based on sequence homology and the physical properties of amino acid. Amino acid substitution at evolutionary conserved sites tends to impair protein function. The eight *ABCA7* nsSNPs were submitted to SIFT for analyzing the tolerance index. The higher tolerance index, the lesser the functional impact the AAS site is likely to have. Among the 8 nsSNPs submitted, only rs3752246 had been detected to be deleterious having a tolerance index score ≤ 0.05 (Table 2).

No.	SNP ID	Allele ^a	Derived Allele Frequency (HAPMAP)			
			CHB	JPT	CEU	YRI
1	rs3764645	A/G	0.51	0.33	-	0.08
2	rs3752232	G/A	0.94	0.92	0.94	0.72
3	rs3764647	G/A	0.94	0.92	0.94	0.73
4	rs3752233	G/A	0.12	0.19	0.06	0.00
5	rs3752239	A/C	1.00	1.00	-	-
6	rs3745842	C/T	0.33 ^b	0.37 ^c	0.42 ^d	0.45 ^e
7	rs3752246	C/G	0.26	0.43	0.18	0.00
8	rs4147918	A/G	0.12	0.16	0.06	0.00

^aderive/ancestral
^bPerlegen Chinese
^cJBIC
^dPerlegen European
^ePerlegen African

Table 1: Derived Allele Frequency of *ABCA7* nsSNPs analyzed in this study.

SNP	allele	Amino acid	Location	DDG	Tolerance Index	PSIC
rs3764645	A/G	E188G	Exon 7	0.12	0.79	1.620
rs3752232	G/A	T319A	Exon 10	-1.19	0.48	0.368
rs3764647	G/A	H395R	Exon 11	-0.60	0.41	0.455
rs3752233	G/A	R463H	Exon 12	-1.58	0.85	1.511
rs3752239	A/C	N718T	Exon 16	0.41	1.00	2.348
rs3745842	C/T	R1349Q	Exon 30	-0.55	0.58	0.403
rs3752246	C/G	G1527A	Exon 33	-0.59	0.02	1.187
rs4147918	A/G	Q1686R	Exon 37	-1.13	0.58	0.376

Table 2: List of *ABCA7* nsSNP predicted to be functional significant.

Identification of *ABCA7* nsSNPs causing protein stability change

In order to identify *ABCA7* functional nsSNPs, we applied I-Mutant 2.0 as a prediction tool for the automatic prediction of protein stability change upon single amino acid substitution. The *ABCA7* protein sequence (NP_061985) was submitted to I-Mutant 2.0 software as an input file. As a result, we obtained 6 nsSNPs with negative DDG values (Table 2). Out of 6 nsSNPs, 3 nsSNPs namely rs3752232, rs3752233 and rs4147918 showed a DDG value > -1.0 . The remaining 3 nsSNPs namely rs3764647, rs3745842 and rs3752246 showed a DDG value < -1.0 as shown in Table 2. A negative DDG value means less stability of the mutated protein. A more negative DDG value indicates a lesser protein stability. Of the 6 nsSNP which showed a negative DDG value, rs3752232, rs3764647, rs3752233, rs3745842, rs4147918 changed their amino acid from polar to non-polar amino acid, aromatic polar amino acid to polar amino acid, polar amino acid to aromatic polar amino acid, positively charged amino acid to uncharged polar amino acid, uncharged polar amino acid to positively charged amino acid respectively. The remaining rs3752246 had no alteration of the amino acid properties. Therefore the former 5 nsSNPs were considered to be functional nsSNP affected on protein stability change by I-Mutant 2.0 software.

Identification of functional *ABCA7* nsSNPs influenced on protein structural level

To identify the *ABCA7* nsSNP affected protein structure, the *ABCA7* nsSNPs were analyzed for predicting a possible impact of AAS on the structure and function of the protein using PolyPhen software. The *ABCA7* protein sequence (NP_061985) with each nsSNP position and their 2 amino acid variants was submitted as input for analyzing the protein structural change due to AAS. A PSIC score difference of 1.1 or above was considered to be significant. Of eight nsSNPs, 4 nsSNPs namely rs3764645, rs3752233, rs3752239 and rs3752246 were considered to be damaging. The PSIC score difference of the 4 functional nsSNPs showed values between 1.187 and 2.348 (Table 2). The rs3752233 which was observed to be the cause of protein less stability by I-Mutant 2.0 software were also predicted to be damaging by PolyPhen software. In addition the rs3752239 was highly confidently predicted as probably damaging nsSNP. The others namely rs3764645 and rs3752246 were supposed to affect protein structure and function.

Investigation of functional effect and estimated risk of ABCA7 nsSNPs

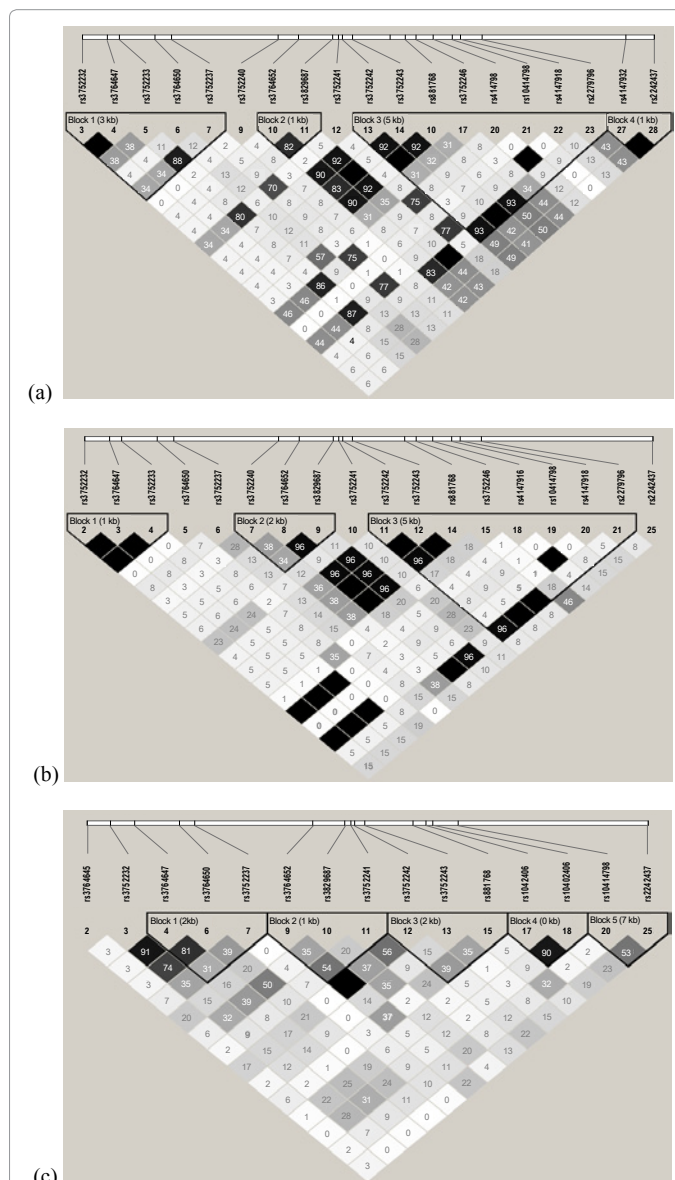
In order to efficiently identify nsSNP with a high possibility of having a functional effect, a Functional Analysis and Selection Tool for Single Nucleotide Polymorphism (FASTSNP) was applied for the detection of nsSNP influence on cellular and molecular biological function e.g. transcriptional and splicing regulation. In addition the estimation of risk score was also calculated by FASTSNP. The functional effect and estimated risk of eight ABCA7 nsSNPs are shown in Table 3. Four ABCA7 nsSNPs exhibited medium-high risk score (risk score = 3-4). The functional nsSNPs were rs3764645, rs3752233, rs3752239 and rs3752246. The remaining four nsSNPs showed low-medium risk (risk score=2-3). The four functional nsSNPs (rs3764645, rs3752233, rs3752239 and rs3752246) detected by FASTSNP were also predicted to be damaging by PolyPhen software. Interestingly the rs3752233 was independently predicted to be the functional nsSNP by I-Mutant 2.0, PolyPhen and FASTSNP. Furthermore, rs3752246 was also predicted to be functionally polymorphic by I-Mutant 2.0, SIFT, FASTSNP as well as PolyPhen software. For the rs3752239 predicted with high confidence to be damaged by PolyPhen (the highest PSIC score = 2.348) was also identified as high risk SNP by FASTSNP. The rs3764645 also had a high PSIC score (1.620) and was also predicted to be a putative functional SNP by FASTSNP. We therefore indicated the four nsSNPs namely rs3752246, rs3752233, rs3764645 and rs3752239 as the potential functional polymorphisms in our study.

Linkage disequilibrium of ABCA7 nsSNPs

An LD map of ABCA7 gene is essential for understanding the physical and biological association between each SNP. This contributes the effective designing and interpreting of the genetic association study. In order to investigate the LD structure of the ABCA7 gene, we constructed the ABCA7 LD map based on ABCA7 SNPs with a minor allele frequency ≥ 0.05 . We showed that the derived allele frequency and the LD pattern of the ABCA7 gene were similar between CHB+JPT and CEU, while the LD pattern of YRI was greatly different from CHB+JPT and CEU (Figure 1). We observed that rs3752233 which was predicted as the potential functional polymorphisms was in the completed LD ($D'=1$ and $r^2=1$) with rs3752232 and rs3764647 in the CEU population while rs3752233 exhibited a low LD level ($D'=1$ and $r^2=0.38$) with rs3752232 and rs3764647 in the Asian population (CHB+JPT). For the remaining predicted functional nsSNP, rs3752246 showed a varying degree of random association with other SNPs in the ABCA7 genomic region of CHB+JPT as well as the CEU. In addition, we also found the completed LD between the SNPs in the LD block 1 including 3 nsSNPs namely rs3752232, rs3764647 and rs3752233 with the rs4147918 which was located 13.4 kb away from the LD block 1 of CEU population.

SNP ID	allele	Amino acid	Possible Effect	Risk Score
rs3764645	A/G	E188G	Non-conservative change	3-4
rs3752232	G/A	T319A	conservative change	2-3
rs3764647	G/A	H395R	conservative change, splicing regulation	2-3
rs3752233	G/A	R463H	Non-conservative change, splicing regulation	3-4
rs3752239	A/C	N718T	Non-conservative change, splicing regulation	3-4
rs3745842	C/T	R1349Q	conservative change, splicing regulation	2-3
rs3752246	C/G	G1527A	Non-conservative change	3-4
rs4147918	A/G	Q1686R	conservative change, splicing regulation	2-3

Table 3: Functional Effect and Estimated Risk (FASTSNP).



(a) pairwise r^2 of CHB+JPT (b) pairwise r^2 of CEU (c) pairwise r^2 of YRI

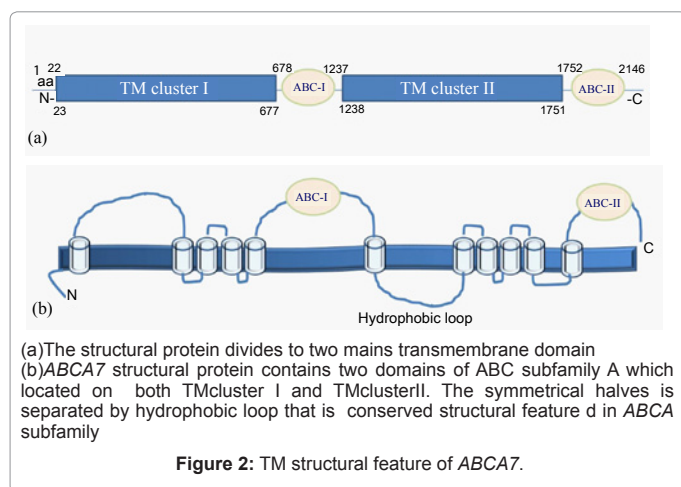
Figure 1: Linkage disequilibrium structure of ABCA7 gene in HapMap population. White, shade of grey and black square indicate no LD ($r^2 = 0$), intermediate LD ($0 < r^2 < 1$) and strong LD ($r^2 = 1$), respectively. The haplotype block indicated by bold black line are defined by the default definition of confidence intervals (Gabriel et al., 2002) in Haploview software.

Structural feature of ABCA7

The ABCA7 structural protein contains two symmetrical halves (Figure 2a) that are interrupted by a stretch of highly hydrophobic residue. The first half of structure is distributed by six TMhelix. The second half contains five TMhelix. Both of ATP binding cassettes are located on the outside loop region (Figure 2b). The ABC region consists of ABC transporter signature motif and Walker A and Walker B motif

Discussion

The increase in the number of putative novel genes and known polymorphisms deposited in the database as well as high-throughput genotyping technology have notably highlighted the



need for prioritization of the potential functional polymorphism used in genetic association study. Consequently, the selection of high potential functional polymorphism facilitates the success of identifying the genetic predisposition to complex diseases. In this study, we applied computational algorithm tools with a difference of biological principles such as the analysis of evolutionary conserve site, protein stability and structural change for the prediction of deleterious polymorphism. Of eleven *ABCA7* nsSNPs deposited in dbSNP of NCBI, we selected 8 validated nsSNPs based on BLAST analysis, information of allele/genotype frequency from HAPMAP database or multiple hits in dbSNP with independent submissions. We proposed 4 *ABCA7* nsSNPs namely rs3764645 (E188G), rs3752233 (R463H), rs3752239 (N718T) and rs3752246 (G1527A) as deleterious functional polymorphisms by I-Mutant 2.0, SIFT, PolyPhen and FASTSNP. Interestingly, rs3752246 (G1527A) is located in the N-myristoylation site. There is a study reported the mutation at N-myristoylation site of endothelial cell nitric oxide synthase (ECNOS) resulting in the changing of ECNOS localization from membrane to cytosolic compartment [24]. The cellular and molecular biology studies need to investigate the influence of rs3752246 on *ABCA7* localization in the human cells. Several mutations in extracellular domain of *ABCA1*, which is highly homologous to *ABCA7*, have been shown to be associated with Tangier and high density lipoprotein (HDL) deficiency. This suggested that the extracellular domain of *ABCA1* might be essential for the *ABCA1* function. A part of *ABCA7* extracellular domain is the epitope (SS-N) of the autoantigen in Sjögren syndrome [25]. The *ABCA7* protein has been recently detected on the plasma cells obtained from salivary gland of Sjögren patients [26]. The rs3752233 (R463H) causing changed the chemical properties from non-aromatic polar amino acid to aromatic polar amino acid is localized in the extracellular domain. The part of *ABCA7* extracellular domain which extrudes outside the cell and acts as autoantigen in Sjögren syndrome was thought to be excised and presented in salivary gland. This part of *ABCA7* protein could activate immune cell to produce autoantibody. However, amino acid changed from arginine to histidine at position 463 of *ABCA7* extracellular domain caused by the rs3752233 might change the antigenicity level of the autoantigen. This is supported by biochemical evidences that arginine exhibits stronger antigenicity potential than histidine [27]. The alteration of antigenicity of *ABCA7* protein might partly responsible for immunological mechanism change in salivary gland and related disease susceptibility. Furthermore the expression of *ABCA7* has also been detected in mouse spermatozoa. The *ABCA7* function as transporter of cholesterol efflux from mouse spermatozoa to lipid

acceptor. This process mediates the maturation and capacitation of spermatozoa. It was suggested that mutation causing the impairment of *ABCA7* structure and expression could influence on protein-protein interaction and function which resulting in reduction of male fertility [28]. The predicted functional nsSNPs in our study have been shown to effect on protein structural level. Thus the biological evidence presented here suggested the possible functional role of *ABCA7* nsSNPs and phenotypic variation related to biomedical field. The four predicted functional nsSNPs could be the potential candidate SNPs for the genetic association study.

The *ABCA7* is a membrane protein composed of 2146 amino acids therefore the protocol for creating the valid protein structure needs to be developed and molecular modeling is required in a further study. In the present study we applied structural bioinformatics tool to analyze the topology of membrane protein structure. The predicted functional SNPs were found to be located in the transmembrane region. The functional consequences of predicted SNPs need to be investigated further.

As a growing number of nsSNP have been deposited in the database, however, it is not practical to genotype all the reported nsSNPs in a large scale population study. The study of *ABCA7* LD structure is particularly useful for selecting a proper and cost-effective marker for the conduction of *ABCA7* genetic association study and functional analysis of *ABCA7* gene. Furthermore, the completed LD between rs4147918 and SNPs in LD block 1 composing of rs3752232, rs3764647 and rs3752233 emphasized the usefulness of the construction of LD map in a study population prior interpreting the result of genetic association study. Analysis of the LD map in HapMap population for inferring LD information to related population can help in designing and selecting tag SNP in genetic association study [4,5].

Our results from this study suggest the application of computational algorithm tools as well as the publicly available database such as dbSNP and HapMap for efficiently selecting a functional SNP for the conduct of genetic association study. Finally, we hope that this study will provide an empirical guideline for researchers to prioritize the known nsSNPs on the basis of molecular consequence. Although high throughput genotyping technology is currently available in post-genomic era, a discrimination of functional nsSNP from neutral nsSNP is an effective way of mapping genes associated with disease risk and also reducing the cost of genotyping. In addition, the LD structure is varied across genomic regions of the human genome. Understanding the structure of LD is helpful in selecting and interpreting the genetic association study. The combination of bioinformatics tools for application in biomedical research has a great impact on the ability to uncover the cause of genetic variation in complex diseases and drug responses.

Acknowledgement

We thank the Faculty of Medical Technology, Mahidol University for providing the facilities to carry out this work. We thank two anonymous reviewers for valuable comments and suggestions regarding this manuscript. This study was supported by the Young Scholar-Mahidol University Grant to P.N. The study was partially supported by the grant from the Faculty of Allied Health Sciences, Chulalongkorn University. This project is also supported by the Office of the Higher Education Commission and Mahidol University under the National Research Universities Initiative.

References

1. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. *Nat Genet* 27: 234-236.

Citation: Nuchnoi P, Nantakomol D, Chumchua V, Plabplueng C, Isarankura-Na-Ayudhya C, et al. (2011) The Identification of Functional Non-Synonymous SNP in Human ATP-Binding Cassette (ABC), Subfamily Member 7 Gene: Application of Bioinformatics Tools in Biomedicine. *J Bioanal Biomed* 3: 026-031. doi:10.4172/1948-593X.1000039

2. Ennis S, Maniatis N, Collins A (2001) Allelic association and disease mapping. *Brief Bioinform* 2: 375-387.
3. Downey PM, Petro R, Simon JS, Devlin D, Lozza G, et al. (2009) Identification of single nucleotide polymorphisms of the human metabotropic glutamate receptor 1 gene and pharmacological characterization of a P993S variant. *Biochem Pharmacol* 77: 1246-1253.
4. Nuchnoi P, J Ohashi, Kimura R, Hananantachai H, Naka I, et al. (2008a) Significant association between TIM1 promoter polymorphisms and protection against cerebral malaria in Thailand. *Ann Hum Genet* 72: 327-336.
5. Nuchnoi P, Ohashi J, Naka I, Nacapunchai D, Tokunaga K, et al. (2008b) Linkage disequilibrium structure of the 5q31-33 region in a Thai population. *J Hum Genet* 53: 850-856.
6. Wang Z, Moutl J (2001) SNP's protein structure and disease. *Hum Mutat* 17:263-270.
7. Nordman S, B Ding, Ostenson CG, Kärvestedt L, Brismar K, et al. (2005) Leu7Pro polymorphism in the neuropeptide Y (NPY) gene is associated with impaired glucose tolerance and type 2 diabetes in Swedish men. *Exp Clin Endocrinol Diabetes* 113: 282-287.
8. Zhao W, Wang L, Lu X, Yang W, Huang J, et al. (2007) A coding polymorphism of the kallikrein 1 gene is associated with essential hypertension: a tagging SNP-based association study in a Chinese Han population. *J Hypertens* 25: 1821-1827.
9. Das A, Das TK, Sahu U, Das BP, Kar SK, et al. (2009) CD36 T188G gene polymorphism and severe falciparum malaria in India. *Trans R Soc Trop Med Hyg* 103: 687-690.
10. Wang HJ, Wermter AK, Nguyen TT, Scherag A, Reichwald K, et al. (2007) No association of sequence variants in the neuropeptide Y2 receptor (NPY2R) gene with early onset obesity in Germans. *Horm Metab Res* 39: 840-844.
11. Knowles JW, Assimes TL, Boerwinkle E, Fortmann SP, Go A, et al. (2008) Failure to replicate an association of SNPs in the oxidized LDL receptor gene (OLR1) with CAD. *BMC Med Genet* 2: 9-23.
12. Naka I, Ohashi J, Nuchnoi P, Hananantachai H, Looareesuwan S, et al. (2008) Lack of association of the HbE variant with protection from cerebral malaria in Thailand. *Biochem Genet* 46: 708-711.
13. Jehle AW, Gardai SJ, Linsel-Nitschke P, Morimoto K, Janssen WJ, et al. (2006) ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. *J Cell Biol* 174: 547-556.
14. Hendig D, Langmann T, Zarbock R, Schmitz G, Kleesiek K et al. (2009) Characterization of the ATP-binding cassette transporter gene expression profile in Y79: a retinoblastoma cell line. *Mol Cell Biochem* 328: 85-92.
15. Kim WS, Weickert CS, Garner B (2008) Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. *J Neurochem* 104: 1145-1166.
16. Capriotti E, P Fariselli, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 33: 306-310.
17. Bava KA, Gromiha MM (2004) ProTherm version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res* 32: D120-121.
18. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30: 3894-3900.
19. Yuan HY, Chiou JJ, Liu CH, Liu CK, Lin YJ, et al. (2006) FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* 34: W635-641.
20. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567-580.
21. Claros MG, von Heijne G (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* 10: 685-686.
22. Barrett JC, Fry B (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265.
23. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29: 308-311.
24. Sessa WC, Barber CM, Lynch KR (1993) Mutation of N-myristoylation site converts endothelial cell nitric oxide synthase from a membrane to a cytosolic protein. *Circ Res* 72: 921-924.
25. Tanaka AR, Ikeda Y, Kidera A, Nakagawa S, Nagase T, et al. (2001) Human ABCA1 contains a large amino-terminal extracellular domain homologous to an epitope of Sjogren's Syndrome. *Biochem Biophys Res Commun* 283: 1019-1025.
26. Toda Y, Aoki R, Ikeda Y, Azuma Y, Kioka N, et al. (2005) Detection of ABCA7 positive cells in salivary glands from patients with Sjögren syndrome *Patho Int* 55:639-643.
27. Masseyeff R, Albert W (1993). *Method of Immunological Analysis*. New York, Wiley.
28. Morales C, Marat A, Ni X, Yu Y, Oko R, et al. (2008) ATP-Binding Cassette Transporter ABCA1, ABCA7 and ABCG1 in Mouse Spermatozoa. *Biochem Biophys Res Commun* 376: 472-477.