Pharmacological Reports
2011, 63, 924–934
ISSN 1734-1140

Non-neutral nonsynonymous single nucleotide polymorphisms in human ABC transporters: the first comparison of six prediction methods

Da Cheng Hao[1], Yao Feng[1], Rongrong Xiao[1], Pei Gen Xiao[2]

[1]Laboratory of Biotechnology, College of Environment, Dalian Jiaotong University, Dalian 116028, China

[2]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Correspondence: Da Cheng Hao, e-mail: hao@djtu.edu.cn; Pei Gen Xiao, e-mail: xiaopg@public.bta.net.cn

Abstract:
Nonsynonymous single nucleotide polymorphisms (nsSNPs) in coding regions that can lead to amino acid changes may cause alteration of protein function and account for susceptibility to disease and altered drug/xenobiotic response. Abundant nsSNPs have been found in genes coding for human ATP-binding cassette (ABC) transporters, but there is little known about the relationship between the genotype and phenotype of nsSNPs in these membrane proteins. In addition, it is unknown which prediction method is better suited for the prediction of non-neutral nsSNPs of ABC transporters. We have identified 2,172 validated nsSNPs in 49 human ABC transporter genes from the Ensembl genome database and the NCBI SNP database. Using six different algorithms, 41 to 52% of nsSNPs in ABC transporter genes were predicted to have functional impacts on protein function. Predictions largely agreed with the available experimental annotations. Overall, 78.5% of non-neutral nsSNPs were predicted correctly as damaging by SNAP, which together with SIFT and PolyPhen, was superior to the prediction methods Pmut, PhD-SNP, and Panther. This study also identified many amino acids that were likely to be functionally critical but have not yet been studied experimentally. There was significant concordance between the predicted results of SIFT and PolyPhen. Evolutionarily non-neutral (destabilizing) amino acid substitutions are predicted to be the basis for the pathogenic alteration of ABC transporter activity that is associated with disease susceptibility and altered drug/xenobiotic response.

Key words:
phenotype, SNAP, PolyPhen, SIFT, Panther, Pmut, SNP, ABC transporter

Introduction

ATP-binding cassette (ABC) transporters are members of a protein superfamily that is one of the largest and most ancient families, with representatives in all extant phyla from prokaryotes to humans [19]. ABC transporters are transmembrane proteins that utilize the energy of adenosine triphosphate (ATP) hydrolysis to carry out certain biological processes, including translocation of substrates across membranes and non-transport-related processes such as translation of RNA and DNA repair [14, 42]. A wide variety of substrates are transported across extra- and intracellular membranes, including metabolic products, lipids and sterols, and drugs. Proteins are classified as ABC transporters based on the sequence and organization of their ABC domain(s).

There are 49 known ABC transporters in humans, which are classified into seven families by the Human Genome Organization (http://nutrigene.4t.com/human-abc.htm). Nonsynonymous SNPs (nsSNPs) of the human ABC transporter genes may cause absent or reduced transport activity, and polymorphisms of ABC transporters have been found to be closely related to altered drug clearance and/or drug response, cystic fibrosis, and various eye diseases and syndromes [21]. For example, nsSNPs of ABCB1 (P-glycoprotein) influence highly active antiretroviral therapy (HAART) efficacy and AIDS-free survival [17].

Human genetic variation may directly or indirectly influence responses to modern antiretroviral therapies for HIV. It is already known that some immunogenetic and other human genetic variations affect the natural history of HIV disease progression where individuals are untreated, but less information is available as to whether these differences are still relevant in the context of HAART [3]. Antiretroviral therapy adds additional opportunities for human genetic contributions to affect variable prognosis, in particular for those genes that influence pharmacokinetics and/or adverse events (e.g., ABC transporter genes). To date, the majority of studies investigating the influence of human genetic variation on HIV disease and treatment outcome have focused on a small number of SNPs, not including most of the ABC transporters.

The functional impact of most nsSNPs in human ABC transporter genes is still unknown. However, computational technologies aid the experimental exploration of nsSNPs and are indispensable in predicting the response to HAART in HIV-infected subjects. For example, Hao et al. [15] identified 923 nsSNPs from human phase II xenobiotic metabolizing enzyme genes and used SNAP, Panther, and PolyPhen to predict the impact of these nsSNPs on enzyme function. In addition, Wang et al. [41] predicted the phenotype of 1,632 nsSNPs of human ABC transporters using SIFT and PolyPhen. However, the numbers of nsSNPs in ABC transporter genes collected in public databases and publications are rapidly increasing, and new algorithms with better predictive power are becoming available. The present assessment of the presumably functionally essential residues of drug/xenobiotic transporters remains far less complete. In this study, we therefore investigated the potential effect of known human ABC transporter nsSNPs on protein function using six algorithms. The data set we compiled is the largest and most diverse one to date for the evaluation of prediction methods that require similar types of inputs.

Materials and Methods

Nonsynonymous SNP datasets

The data on human ABC transporter genes were collected from Ensembl (http://www.ensembl.org/Homo_sapiens/Search/) and Entrez Gene on the NCBI website (http://www.ncbi.nlm.nih.gov/sites/entrez). Expired and merged gene names were excluded from the study. The majority of the variants included in this analysis were identified during the screening of 12 human *ABCA*, 11 *ABCB*, 13 *ABCC*, 4 *ABCD*, 1 *ABCE*, 3 *ABCF*, and 5 *ABCG* genes from Ensembl (http://www.ensembl.org/Homo_sapiens/Gene/Variation_ Gene/). Ensembl integrates genetic variants from dbSNP, Uniprot, the personal genomes of Watson and Venter, and Illumina human genome sequencing results and links to information such as transcripts, population genetics, individual genotype, genomic context, phenotype data, and phylogenetic context. Information including gene symbol, gene name, mRNA accession number (ENST or NM), protein accession number (ENSP or NP), SNP ID, amino acid residue 1 (wild-type, wt), amino acid position, and amino acid residue 2 (missense) were collected. Supplementary variants were identified from Entrez Gene on NCBI and through PubMed literature searching and added to the dataset after cross-examination. The information on the effect of the nsSNPs on enzyme activity and the correlation between the nsSNPs and disease/adverse drug reaction/toxicant intake were extracted from *in vivo* and *in vitro* experiments (e.g., recombinant protein analysis) according to the literature.

Prediction of the phenotypes of nsSNPs in human ABC transporter genes

The effect of the variant amino acid substitution on protein function was predicted using PolyPhen (http://genetics.bwh.harvard.edu/pph/), Panther (http://www.pantherdb.org/tools/csnpScoreForm.jsp), SNAP (http://cubic.bioc.columbia.edu/services/SNAP/submit.html), SIFT (http://sift.jcvi.org/#), Pmut (http://mmb2.pcb.ub.es:8080/PMut/), and PhD-SNP (http://gpcr2.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi). The table with the above information is available upon request.

PolyPhen uses empirically derived rules based on previous research in protein structure, interaction, and

evolution that automatically predict whether a replacement is likely to be deleterious for the protein on the basis of three-dimensional structure and multiple alignments of homologous sequences [30]. In this study, PolyPhen input is a protein amino acid sequence together with sequence position and two amino acid variants characterizing the polymorphism. Generally, PolyPhen scores of 0–1.49 are classified as benign, 1.50–1.99 as possibly damaging, and ≥ 2 as probably damaging. In concordance analysis, scores of 0–0.99 are classified as benign, 1–1.24 as borderline, and 1.25–1.49 as potentially damaging. However, some predictions were presented as benign, possibly damaging, or probably damaging, but without scores. In addition, some prediction scores were below zero. These prediction results were excluded in concordance analysis between the functional consequences for nsSNPs predicted by two different algorithms.

The Panther program analyzes variants and provides subPSEC (position-specific evolutionary conservation) scores that range from –10 (most severe) to 0 (least severe) [36, 37]. It then uses a hidden Markov model (HMM) based on position specific independent counts to convert this data into a probability score that the particular amino acid substitution is damaging. To convert the Panther results, a score of 0 to –3 is considered to be tolerated and a score of –3 to –10 is not tolerated. The value of –3 was chosen because it represents a probability of 0.5 of being damaging.

SNAP combines many sequence analysis tools in a battery of neural networks to predict the functional effects of nsSNPs [2]. For example, SNAP uses functional effects from SIFT [24] and conservation information from position-specific independent counts (PSIC) [33]. For each mutant, SNAP returns three values: the binary prediction (neutral/non-neutral), the RI (Reliability Index, range 0–9) and the expected accuracy that estimates accuracy on a large dataset at the given RI (i.e., accuracy of test set predictions calculated for each neutral and non-neutral RI). The latter two values correlate; when both are provided, the server chooses the one yielding better predictions.

SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring nonsynonymous polymorphisms and laboratory-induced missense mutations. Generally, SIFT scores of 0–0.05 are classified as "AFFECT PROTEIN FUNCTION" and 0.05–1 as "TOLERATED". Pmut uses different kinds of sequence information to label mutations and neural networks to process this information [12]. It provides a very simple output: a yes/no answer and a reliability index. PhD-SNP, a support vector machine (SVM)-based classifier, is optimized to predict if a given single point protein mutation can be classified as disease-related or as a neutral polymorphism [6].

Validation of the prediction results

We first searched PubMed by keyword (i.e., the respective gene name), then downloaded the search results that had been recorded before November 2010 from the National Center for Biotechnology Information (NCBI). From this search, we obtained approximately 3,000 papers that contained the gene names. Next, we curated the data manually and retrieved phenotype data related to the nsSNPs. The criteria for inclusion were that there was evidence of altered transporting activity or disease association shown in the references. Different researchers double-checked all articles collected. nsSNPs with experimental evidences of altered transport activity or disease association were regarded as deleterious. The phenotypic data were from both *in vivo* and *in vitro* studies, in which the analysis of site-directed mutagenesis or enzymatic/transport changes often provided direct evidence indicating the functional impact of nsSNPs. Prediction accuracy was analyzed according to the positive findings from these experiments. As a test for the ability of six algorithms to identify substitutions impacting enzymatic activity, scores were obtained and compared for the collected nsSNPs of human ABC transporter genes related to the loss of transport/enzyme activity and disease based on experimental and clinical studies.

Statistical analysis

The Pearson's χ^2 test and Fisher's exact test were used to compare the percentages. Given that the six algorithms employ different approaches and also different data sets as foundations for their analyses, it is important to find the concordance of the six prediction tools on functional consequences of each nsSNP prediction. Concordance analysis of each nsSNP predicted by six methods were assessed using the linear correlation coefficient R. Prediction scores of each nsSNP were plotted on scatter graphs and analyzed using linear trend lines; p values below 0.05 were considered statistically significant.

Results

Validated nsSNPs of human ABC transporter genes

Two thousand one hundred and seventy two amino acid substitution variants were identified in the systemic screening of 49 human ABC transporter genes for the analysis of the potential impact of all nsSNPs in human ABC transporter genes. Among seven subfamilies, *ABCC* had the most nsSNPs (901), followed by *ABCB* (603), *ABCA* (538), *ABCG* (61), *ABCF* (38), and *ABCD* (29). *ABCE* had only two nsSNPs. We cross-examined the databases and removed invalid SNPs. No nsSNPs were identified in the screening of *ABCD2*; therefore, it was not included in this study. Among the other 48 genes, *ABCC6* (*MRP6*) had the most nsSNPs (269; table available upon request), followed by *ABCA4* (222), *ABCB3* (*TAP2*; 170), ABCB2 (*TAP1*; 143), and *ABCC7* (*CFTR*; 142).

Prediction of functional effects of nsSNPs of human ABC transporter genes

As shown in Figure 1, there was significant difference in the distribution of prediction results by the six algorithms (x^2 = 2193.2, df = 10, $p < 0.0001$). More deleterious predictions were made by Panther (52.3%) and SIFT (52.1%) than by SNAP (41.1%) and Pmut (37.2%), with PolyPhen (45.4%) and PhD-SNP (42.7%) in between. Notably, there were no Panther predictions for 508 (23.4%) nsSNPs. Similarly, there were no Panther predictions for 101 (23.5%) *UGT* (*UDP-glucuronosyltransferase*) nsSNPs [15]. The common reason for this result was that they failed to align to an HMM. For comparison, Thomas et al. [36] found that 100/115 missense SNPs (not predicted, 13.0%) aligned to a position in an HMM from Panther and could be given sub-PSEC scores. For the neurodegenerative disease-associated genes *SOD1* and *PARK2*, 18.3% (17/93) and 45.4% (15/33) of nsSNPs failed to align to the HMM employed by Panther, respectively, and could not be analyzed [38].

Potential effect of nsSNPs of ABC transporters on amino acid changes

Four hundred ninety four, 439, and 402 ABCC nsSNPs were predicted as non-neutral by SIFT, PolyPhen, and SNAP, respectively (Fig. 2). As many as

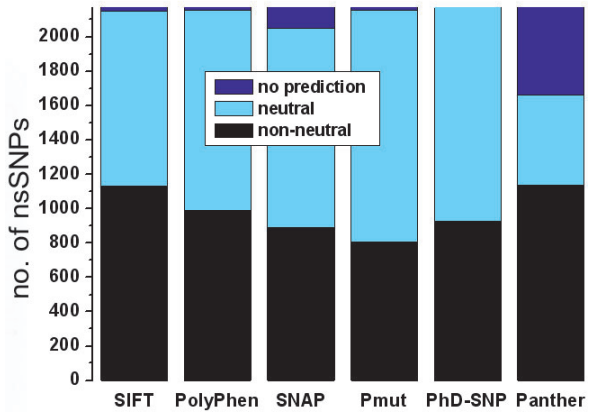

Fig. 1. Prediction results for nsSNPs of human ABC transporter genes. NS, not scored (not predicted); Neu, neutral; and Non, non-neutral

286 (31.7%) ABCC nsSNPs were predicted as deleterious by SNAP, PolyPhen and SIFT, followed by ABCA (170) and ABCB (96). In comparison, the overlap between predictions of Pmut, PhD-SNP, and Panther was low. Although 55.5% (1205/2172) of nsSNPs were predicted as non-neutral by Pmut or PhD-SNP, only 15.8% (344/2172) of predictions were shared across all three methods, which was significantly lower than for SNAP, PolyPhen, and SIFT (27.1% or 588; $p < 0.0001$). As many as 147 *ABCC6* nsSNPs and 120 *ABCA4* nsSNPs were predicted as deleterious by SNAP, PolyPhen, and SIFT, respectively, followed by *ABCC7* (47), *ABCC8* (*SUR1*; 27),

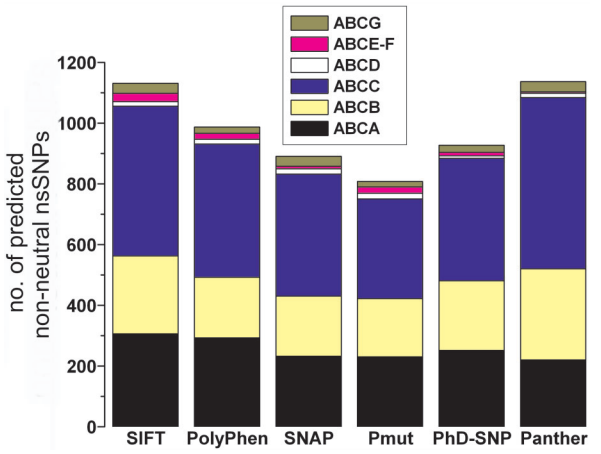

Fig. 2. Comparison of SIFT, Panther, PolyPhen, PhD-SNP, Pmut and SNAP predictions. The numbers of predictions made by the six methods are shown. Probably and possibly damaging mutations were used for PolyPhen

and *ABCB3* (22). In contrast, only 76 *ABCA4* nsSNPs and 57 *ABCC6* nsSNPs were predicted as deleterious by Pmut, PhD-SNP, and Panther, followed by *ABCC7* (30), *ABCC8* (24), and *ABCB2* (18). Personalized medicine is about making the treatment as individualized as the disease. It involves identifying genetic, genomic, and clinical information that allows accurate predictions to be made about a person's susceptibility of developing disease, the course of disease, and its response to treatment. Individuals with deleterious nsSNPs should be provided with the individualized therapeutic regimen [16]. It seemed that only a small subset of deleterious mutations could be reliably identified but that this subset provided the raw material for personalized medicine. Because there was only a small amount of overlap among the six methods, multiple methods should be used when trying to identify deleterious mutations in humans.

Table 1 shows common amino acid changes of non-neutral nsSNPs in human ABC transporter genes predicted by PolyPhen and SIFT. Arg was the most common amino acid of contig reference (wild-type), followed by Leu and Gly. Cys was the most common amino acid of missense, followed by Pro and Arg. Arg→Cys was the most frequent substitution caused by nsSNPs in ABC transporter genes, followed by Leu→Pro and Arg→Trp. Interestingly, Arg was also the most common amino acid of cytochrome P450 (CYP) contig reference [40], and Arg→Cys was the most frequent substitution caused by nsSNPs in *CYP* genes.

According to the PolyPhen algorithm, potential effects of some selected amino acid substitutions might result from nsSNPs in human ABC transporter genes. These predictions included the disruption of an annotated functional site (e.g., S753R in ABCC7), disruption of a ligand binding site (e.g., K461E and T463I in ABCB11 or Q698P in ABCC6), disruption of an annotated bond formation site (e.g., C75G, C1488R, and C1488F in ABCA4), and improper substitution in the transmembrane region (e.g., C764Y in ABCA4, T1980K in ABCA12, and G982R in ABCB11). Forty-six of these predictions were confirmed by *in vitro*, *in vivo*, and/or epidemiological studies (Tab. 2).

Tab. 1. Common amino acid changes of deleterious nsSNPs in human ABC transporter genes predicted by the SIFT and PolyPhen algorithms

Contig reference	Number	Missense	Number	Common amino acid change	Number
Arg	185	Cys	94	Arg→Cys	58
Leu	104	Pro	85	Leu→Pro	53
Gly	91	Arg	67	Arg→Trp	36
Ser	49	Ser	55	Arg→Gln	29
Val	36	Gly	44	Gly→Arg	26
Thr	33	Trp	44	Arg→His	20
Ala	32	Gln	42	Leu→Arg	18
Asp	29	Leu	40	Arg→Gly	14
Ile	26	Phe	39	Gly→Val	13
Pro	26	His	39	Pro→Leu	13
Glu	21	Thr	32	Asp→Asn	13
Lys	20	Glu	31	Glu→Lys	12
Asn	20	Lys	31	Ala→Pro	11
Phe	20	Met	29	Arg→Leu	11
Tyr	19	Val	27	Tyr→Cys	11
Cys	19	Asp	25	Val→Met	10
Met	15	Asn	21	Thr→Ile	10
Gln	14	Ile	18	Val→Gly	10
Trp	11	Tyr	9	Gly→Asp	10
His	11	Ala	9	Leu→Gln	9
TOTAL	781	TOTAL	781		

Concordance analysis of predicted results by PolyPhen and SIFT

Figure 3 and Table 3 show the concordance analysis between the functional consequences for 2,129 nsSNPs predicted by PolyPhen and SIFT. Raw scores of PolyPhen and SIFT were used for the correlation analysis. The scatter graph was plotted using prediction scores of PolyPhen and SIFT, and it showed negative correlation (Fig. 3). The correlation coefficient $r = -0.502$ (ANOVA, $p < 0.0001$) illustrated the significant concordance between the prediction scores from these two algorithms.

Validation of the prediction results

The confirmed phenotypes of nsSNPs manifest as alterations in transport/enzyme activity, susceptibility to

Tab. 2. Potential effect of amino acid substitutions for nsSNPs in human ABC transporter genes predicted by the PolyPhen algorithm and the analogous predictions by SIFT and SNAP

Gene	Allelic variant	Substitution Effect (According to PolyPhen)	SIFT prediction	SNAP prediction	Phenotype	Reference
ABCA4	C75G	Disruption of annotated bond formation site	TOLERATED	Non-neutral	STGD	[42]
	C764Y	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
	V767D	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
	G851D	Improper substitution in the TM region	AFFECT	Neutral	STGD	[42]
	E1087K	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
	P1380L	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
	C1488R	Disruption of annotated bond formation site	AFFECT	Non-neutral	STGD	[42]
	C1488 F	Disruption of annotated bond formation site	AFFECT	Non-neutral	STGD	[42]
	L1729P	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
	S1736P	Improper substitution in the TM region	TOLERATED	Non-neutral	STGD	[42]
	V1884E	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
	G1886E	Improper substitution in the TM region	AFFECT	Non-neutral	STGD	[42]
ABCA12	T1980K	Improper substitution in the TM region	TOLERATED	Non-neutral	Non-Bullous congenital ichthyosiform erythroderma	[30]
ABCB11	K461E	Disruption of ligand binding site	AFFECT	Non-neutral	Immature protein	[5]
	T463I	Disruption of ligand binding site	AFFECT	Non-neutral	Mild exon skipping	[5]
	G982R	Improper substitution in the TM region	AFFECT	Non-neutral	Immature protein	[5]
ABCC6	W38S	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	S317R	Improper substitution in the TM region	TOLERATED	Non-neutral	PXE	[28]
	L355R	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	T364R	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	C440G	Improper substitution in the TM region	TOLERATED	Non-neutral	PXE	[28]
	A455P	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	L463H	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	S535P	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	F551S	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	Q698P	Disruption of ligand binding site	AFFECT	Non-neutral	PXE	[28]
	T944I	Improper substitution in the TM region	TOLERATED	Neutral	PXE	[28]
	L1063R	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
	G1203D	Improper substitution in the TM region	AFFECT	Non-neutral	PXE	[28]
ABCC7	G85E	Improper substitution in the TM region	AFFECT	Non-neutral	produced only core-glycosylated protein	[13]
	G85V	Improper substitution in the TM region	AFFECT	Non-neutral	produced only core-glycosylated protein	[13]
	Y89C	Improper substitution in the TM region	AFFECT	Neutral	abnormal gating	[13]
	E92K	Improper substitution in the TM region	AFFECT	Non-neutral	produced only core-glycosylated protein	[13]
	K95S	Improper substitution in the TM region	AFFECT	Non-neutral	charge neutralizing	[44]
	Q98R	Improper substitution in the TM region	AFFECT	Non-neutral	cystic fibrosis	[17]
	I125T	Improper substitution in the TM region	AFFECT	Neutral	chronic pulmonary disease	[25]
	H199R	Improper substitution in the TM region	AFFECT	Non-neutral	cystic fibrosis	[10]
	R334W	Improper substitution in the TM region	AFFECT	Non-neutral	cystic fibrosis	[38]
	L346P	Improper substitution in the TM region	AFFECT	Non-neutral	significant loss of segment hydropathy	[9]
	R347H	Improper substitution in the TM region	AFFECT	Non-neutral	cystic fibrosis	[17]
	S753R	Disruption of annotated functional site	AFFECT	Non-neutral	CBAVD	[27]
	S1141K	Improper substitution in the TM region	TOLERATED	Non-neutral	increased susceptibility to channel block by cytoplasmic anions	[44]
ABCD1	R104C	Improper substitution in the TM region	AFFECT	Non-neutral	X-ALD	[34]
	P143S	Improper substitution in the TM region	AFFECT	Non-neutral	X-ALD	[26]
	S342P	Improper substitution in the TM region	TOLERATED	Neutral	X-ALD	[34]
ABCG8	G575R	Improper substitution in the TM region	AFFECT	Non-neutral	Sitosterolaemia	[31]

TM – transmembrane; STGD – Stargardt disease 1; CBAVD – congenital bilateral absence of the vas deferens; X-ALD – X-linked adrenoleukodystrophy

Tab. 3. Concordance analysis between the functional consequences of each nsSNP predicted by SIFT and PolyPhen

		SIFT prediction					Total
		Tolerant 1.000~0.201	Borderline 0.200~0.101	Potentially intolerant 0.100~0.050	Intolerant 0.049~0.000	Not scored	
PolyPhen prediction							
Benign	N/A	7	0	3	40	9	59
Benign	0.00~0.99	455	98	66	149	3	771
Borderline	1.00~1.24	58	15	19	49	0	141
Potentially damaging	1.25~1.49	57	25	27	76	1	186
Possibly damaging	1.50~1.99	59	36	56	250	0	401
Probably damaging	2.00	22	21	22	506	0	571
Total		658	195	193	1,070	13	2,129

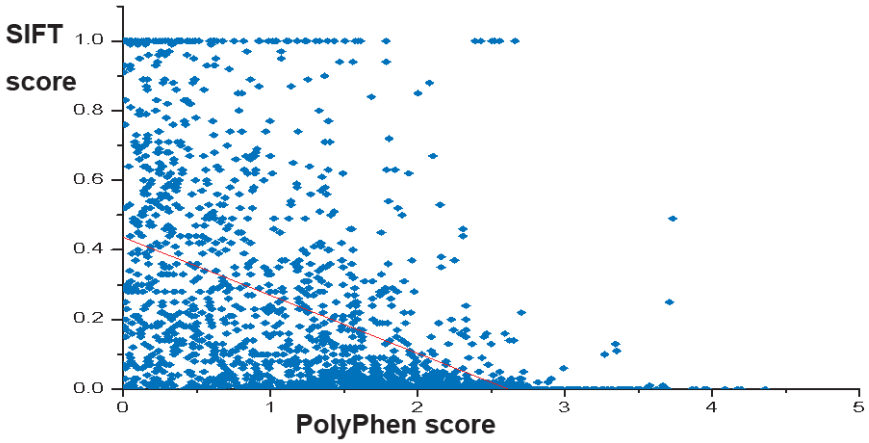

Fig. 3. Scatter graph showing the negative correlation between prediction scores for nsSNPs in human ABC transporter genes using SIFT and PolyPhen

diseases (e.g., cystic fibrosis, Dubin-Johnson syndrome, intrahepatic cholestasis of pregnancy, permanent neonatal diabetes mellitus, and Stargardt disease 1), drug/xenobiotic toxicity (e.g., drugs that are substrates of ABC transporters in the placenta) and resistance. To date, 548 nsSNPs (ABCC6: 114, ABCA4: 113, ABCC7: 62, others: 259) of human ABC transporters have been identified in relation to decreased activity, loss of transport activity, susceptibility to diseases or altered drug/xenobiotic response based on experimental and clinical studies. Using positive findings from these experiments, if the variants were predicted to be deleterious in this study, they were considered correct predictions (true positive, TP). An incorrect prediction was assumed when these nsSNPs were predicted as tolerant (false negative, FN). In addition, 17 neutral nsSNPs (ABCB1: 9, ABCC3: 8) were collected to determine the prediction specificity of the six methods.

The confirmed variants were collected from results derived from site-directed mutagenesis studies of the enzyme using biochemical characterization or clinical data from family-based and association studies (table available upon request). The distribution of prediction results made by six algorithms was significantly different ($x^2 = 1201.8$, $p < 0.0001$; Fig. 4A). SNAP made the most TP predictions (78.5%), followed by SIFT (77.4%) and PolyPhen (72.6%). The prediction accuracy of the theoretical RPLS model for the testing set was 80.4% [21], but this method was only tested in

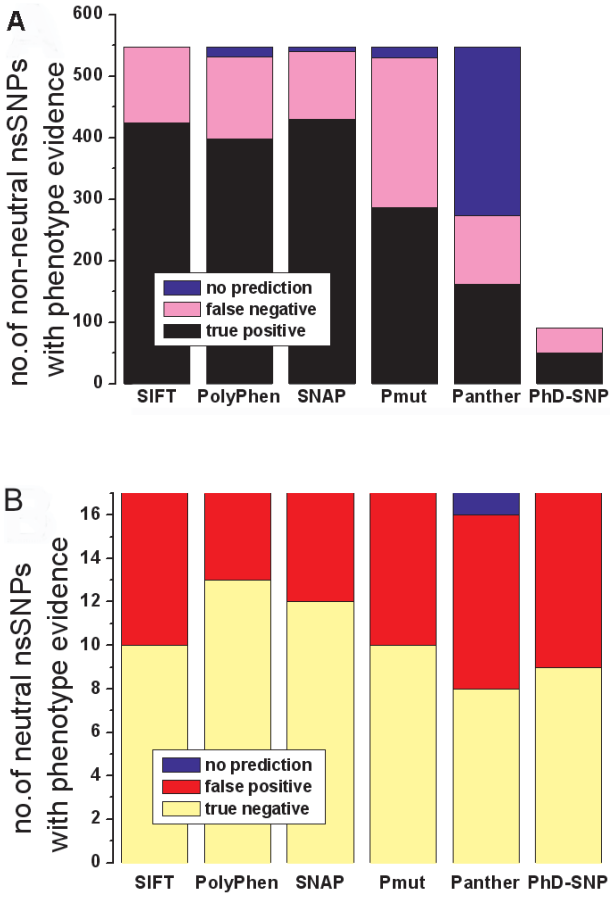

Fig. 4. Prediction performance of PolyPhen, SIFT, Panther, PhD-SNP, Pmut and SNAP on confirmed non-neutral nsSNPs of human ABC transporters. **A** – true positive prediction and false negative prediction; **B** – true negative prediction and false positive prediction

nsSNPs of ABCB transporters. Panther made the least TP predictions (29.6%), followed by Pmut (52.2%) and PhD-SNP (54.9%). The latter two were determined to be the most insensitive prediction methods for nsSNPs of ABC transporters, with 44.7% and 45.1% false negative rates, respectively. Furthermore, the prediction accuracy of Panther varied greatly across different protein families. For example, Panther correctly predicted the functional impact of greater than 94% of the naturally occurring ABCA1 variants, superior to PolyPhen [4]. In our previous study [15], however, Panther made 98.4% and 60.6% TP predictions for nsSNPs of UGT and other phase II metabolizing enzymes, respectively. Finally, the cross-validated performance of Pmut was an 84% overall success rate [12] and a 67% improvement over random. However, Pmut might not be suitable for nsSNP prediction of human ABC transporters.

Approximately 76.5%, 70.6%, and 58.8% of characterized non-deleterious nsSNPs were correctly predicted as neutral and "TN predictions" by PolyPhen, SNAP and SIFT, respectively (Fig. 4B). In addition, there was no significant difference in the results of TN and FP predictions made by the six algorithms (Fisher's exact test, $p = 0.60$). The numbers of TN and FP predictions made by the six methods were not reported previously; thus, this study represents the first report on the prediction specificity of the six methods of interest. Taken together, SNAP, PolyPhen, and SIFT are superior to Pmut, Panther, and PhD-SNP in the prediction of non-neutral nsSNPs of ABC transporter genes.

Discussion

Two thousand one hundred seventy two validated nsSNPs were obtained from 49 validated ABC transporter genes from Ensembl and NCBI dbSNP. Each *ABCA*, *ABCB*, *ABCC*, *ABCD*, *ABCF*, and *ABCG* gene had an average of 45, 55, 69, 7, 13, and 12 nsSNPs, respectively. However, only 25.2% (548 of 2172) of the nsSNPs in the data set of validated nsSNPs in ABC transporter genes were found to contribute to the alteration of transport/enzyme activity or correlate with disease according to published *in vivo* and *in vitro* studies. This result was significantly lower than the 33% found in phase I metabolizing enzyme *CYP* genes ($p < 0.0001$) [40]. These confirmed phenotypes of nsSNPs related to alteration in transport/enzyme activity were then connected to susceptibility to various diseases and poor transport of drugs/xenobiotics.

Although many sophisticated *in silico* approaches have been used to predict the impact of nsSNPs on protein structure and activity, the foundations for these algorithms are protein sequence alignment, physicochemical differences, mapping to known protein 3-D structures, or combinations thereof. Different *in silico* algorithms focus on different aspects of this information, among which the PolyPhen, SIFT, and SNAP algorithms are the main representatives in this field. Significant agreement was observed between the functional consequences of nsSNP predicted by the SIFT and PolyPhen algorithms (Fig. 3). In *CYP*s, 38.94% and 42.73% of the amino acid substitutions were predicted by SIFT and PolyPhen, respectively, to have functional effects on enzymatic activity [40].

However, we found significant differences in the distribution of PolyPhen prediction results for ABC transporters, *CYP*s, *UGT*s, and other phase II enzymes (x^2 = 72.6, $p < 0.0001$) [15, 40]. For example, more deleterious predictions (54.4%) were made for *UGT*s than for ABC transporters (45.4%), *CYP*s (42.7%) and other phase II enzymes (34.5%). In SNAP prediction, less non-neutral predictions (41.1%) were made for ABC transporters and other phase II enzymes (44%) than for *UGT*s (57.2%; $p < 0.0001$). Furthermore, more non-neutral predictions (52.1%) were made for ABC transporters in SIFT prediction than for *CYP*s (38.9%; $p < 0.0001$). These results were consistent with results from Xi et al. [44], who reported that 20 to 50% of the large number of amino acid substitutions observed in DNA repair genes affected function. Additionally, Doss and Sethumadhavan [11] reported that PolyPhen classified 40 of 125 amino acid substitutions (32%) in hereditary nonpolyposis colorectal cancer (HNPCC) genes as "probably or possibly damaging".

A number of *in vivo* and *in vitro* experiments have provided evidence for the functional effects of nsSNPs on transport/enzyme activity and metabolic dysfunction or their correlations with diseases. Prediction accuracy can then be analyzed based on this evidence. As such, 63 to 75% of amino acid substitutions were previously predicted correctly by the SIFT and PolyPhen algorithms [7, 23, 34], which was comparable to our results for ABC transporters. Specifically, PolyPhen correctly designated 60% of disease-associated mutations (DAMs) to be probably damaging [20]. However, 21% of DAMs were identified to be benign, which provided a lower limit on the false negative rate of inference and was comparable to our false negative rates (24.5%). Similar accuracies were also reported in other studies [8, 22, 25].

In this study, SNAP proved to be at least as good as PolyPhen and SIFT in predicting non-neutral nsSNPs of ABC transporters, and its prediction specificity was comparable to PolyPhen and SIFT. SNAP was therefore used to assess the functional necessity of all possible nonnative point mutants in the entire hMC4R protein, and the predictions largely agreed with the available experimental annotations [1]. In this study, although the data for evaluation were obtained from benchmarking studies, there might have been a bias because of the small sample set of only 548 substitutions, so one should be cautious to extrapolate these data to another gene family or genome.

Conclusion

In conclusion, the present study identified 41 to 52% of nsSNPs of human ABC transporter genes to be non-neutral using *in silico* methods. A prediction accuracy analysis found that 78.5% (SNAP) of non-neutral nsSNPs were predicted correctly as damaging. In predicting the non-neutral nsSNPs, the number of true positive predictions and the prediction accuracy of SNAP, PolyPhen, and SIFT were significantly higher than Pmut, PhD-SNP, and Panther. In contrast, no one single algorithm was superior to the others in prediction specificity. Of nsSNPs predicted as non-neutral, the prediction results of PolyPhen, SIFT, and SNAP cross-validated and complemented each other and particularly identified nsSNPs with phenotypes confirmed by disease association studies and biochemical analyses. These amino acid substitutions were thus postulated to be the pathogenetic basis for increased susceptibility to certain diseases, adverse drug reactions, and altered drug/xenobiotic response. Developing new algorithms for integrating heterogeneous data types is now essential to take advantage of the available information, which can then be used to infer the biological impact of SNPs. Detailed analyses of nsSNPs aid in the identification of key residues in ABC transporters. The prediction of nsSNPs in human ABC transporter genes is then helpful for further genotype-phenotype studies on individual differences in drug/xenobiotic transport and clinically unfavorable responses.

Acknowledgments:
The authors wish to thank the National Science and Technology major program 2008ZX10005-004 (Study of Chinese traditional medicine effecting HAART post immuno-reconstitution), the Liaoning Education Department (grant 2009A120) and the China Postdoctoral Science Foundation (grants 20080440019 and 200902069) for financial support.

References:

1. Bromberg Y, Overton J, Vaisse C, Leibel RL, Rost B: In silico mutagenesis: a case study of the melanocortin 4 receptor. FASEB J, 2009, 23, 3059–3069.
2. Bromberg Y, Rost B: SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res, 2007, 35, 3823–3835.

3. Brumme ZL, Harrigan PR: The impact of human genetic variation on HIV disease in the era of HAART. AIDS Rev, 2006, 8, 78–87.
4. Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR: Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the *ABCA1* gene. PLoS Genet, 2005, 1, e83.
5. Byrne JA, Strautnieks SS, Ihrke G, Pagani F, Knisely AS, Linton KJ, Mieli-Vergani G et al: Missense mutations and single nucleotide polymorphisms in *ABCB11* impair bile salt export pump processing and function or disrupt pre-messenger RNA splicing. Hepatology, 2009, 49, 553–567.
6. Capriotti E, Calabrese R, Casadio R: Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics, 2006, 22, 2729–2734.
7. Chasman D, Adams RM: Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol, 2001, 307, 683–706.
8. Cheng TM, Lu YE, Vendruscolo M, Lio' P, Blundell TL: Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single nucleotide polymorphisms. PLoS Comput Biol, 2008, 4, e1000135.
9. Choi MY, Partridge AW, Daniels C, Du K, Lukacs GL, Deber CM: Destabilization of the transmembrane domain induces misfolding in a phenotypic mutant of cystic fibrosis transmembrane conductance regulator. J Biol Chem, 2005, 280, 4968–4974.
10. D'Apice MR, Gambardella S, Bengala M, Russo S, Nardone AM, Lucidi V, Sangiuolo F, Novelli G: Molecular analysis using DHPLC of cystic fibrosis: increase of the mutation detection rate among the affected population in Central Italy. BMC Med Genet, 2004, 5, 8.
11. Doss CG, Sethumadhavan R: Investigation on the role of nsSNPs in HNPCC genes – a bioinformatics approach. J Biomed Sci, 2009, 16, 42.
12. Ferrer-Costa C, Orozco M, de la Cruz X: Sequence-based prediction of pathological mutations. Proteins, 2004, 57, 811–819.
13. Gené GG, Llobet A, Larriba S, de Semir D, Martínez I, Escalada A, Solsona C et al.: N-terminal CFTR missense variants severely affect the behavior of the CFTR chloride channel. Hum Mutat, 2008, 29, 738–749.
14. Goffeau A, de Hertogh B, Baret PV: ABC Transporters. In: Encyclopedia of Biological Chemistry. Eds. Lennarz WJ, Lane MD, Elsevier, Amsterdam, 2004, 1, 1–5.
15. Hao DC, Xiao PG, Chen SL: Phenotype prediction of nonsynonymous single nucleotide polymorphisms in human phase II drug/xenobiotic metabolizing enzymes: perspectives on molecular evolution. Sci China Life Sci, 2010, 53, 1252–1262.
16. He XJ, Zhao LM, Qiu F, Sun YX, Li-Ling J: Influence of *ABCB1* gene polymorphisms on the pharmacokinetics of azithromycin among healthy Chinese Han ethnic subjects. Pharmacol Rep, 2009, 61, 843–850.
17. Hendrickson SL, Jacobson LP, Nelson GW, Phair JP, Lautenberger J, Johnson RC, Kingsley L et al.: Host genetic influences on highly active antiretroviral therapy efficacy and AIDS-free survival. J Acquir Immune Defic Syndr, 2008, 48, 263–271.
18. Izumikawa K, Tomiyama Y, Ishimoto H, Sakamoto N, Imamura Y, Seki M, Sawai T et al.: Unique mutations of the cystic fibrosis transmembrane conductance regulator gene of three cases of cystic fibrosis in Nagasaki, Japan. Intern Med, 2009, 1327–1331.
19. Jones PM, George AM: The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci, 2004, 61, 682–699.
20. Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipski AJ: Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. Genome Res, 2009, 19, 1562–1569.
21. Li Y, Wang Y, Li Y, Yang L: Prediction of the deleterious nsSNPs in *ABCB* transporters. FEBS Lett, 2006, 580, 6800–6806.
22. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ et al.: Proportionally more deleterious genetic variation in European than in African populations. Nature, 2008, 451, 994–997.
23. Ng PC, Henikoff S: Accounting for human polymorphisms predicted to affect protein function. Genome Res, 2002, 12, 436–446.
24. Ng PC, Henikoff S: SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res, 2003, 31, 3812–3814.
25. Ng PC, Henikoff S: Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet, 2006, 7, 61–80.
26. Ngiam NS, Chong SS, Shek LP, Goh DL, Ong KC, Chng SY, Yeo GH, Goh DY: Cystic fibrosis transmembrane conductance regulator (*CFTR*) gene mutations in Asians with chronic pulmonary disease: a pilot study. J Cyst Fibros, 2006, 5, 159–164.
27. Perusi C, Gomez-Lira M, Mottes M, Pignatti PF, Bertini E, Cappa M, Vigliani MC et al.: Two novel missense mutations causing adrenoleukodystrophy in Italian patients. Mol Cell Probes, 1999, 13, 179–182.
28. Pieri Pde C, Missaglia MT, Roque Jde A, Moreira-Filho CA, Hallak J: Novel *CFTR* missense mutations in Brazilian patients with congenital absence of vas deferens: counseling issues. Clinics (Sao Paulo), 2007, 62, 385–390.
29. Plomp AS, Florijn RJ, Ten Brink J, Castle B, Kingston H, Martín-Santiago A, Gorgels TG et al.: *ABCC6* mutations in pseudoxanthoma elasticum: an update including eight novel ones. Mol Vis, 2008, 14, 118–124.
30. Ramensky V, Bork P, Sunyaev S: Human non-synonymous SNPs: server and survey. Nucleic Acids Res, 2002, 30, 3894–3900.
31. Sakai K, Akiyama M, Yanagi T, McMillan JR, Suzuki T, Tsukamoto K, Sugiyama H et al: *ABCA12* is a major causative gene for non-bullous congenital ichthyosiform erythroderma. J Invest Dermatol, 2009, 129, 2306–2309.
32. Solcà C, Stanga Z, Pandit B, Diem P, Greeve J, Patel SB: Sitosterolaemia in Switzerland: molecular genetics links

the US Amish-Mennonites to their European roots. Clin Genet, 2005, 68, 174–178.
33. Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN: PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. Protein Eng, 1999, 12, 387–394.
34. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P: Prediction of deleterious human alleles. Hum Mol Genet, 2001, 10, 591–597.
35. Takahashi N, Morita M, Maeda T, Harayama Y, Shimozawa N, Suzuki Y, Furuya H et al.: Adrenoleukodystrophy: subcellular localization and degradation of adrenoleukodystrophy protein (ALDP/ABCD1) with naturally occurring missense mutations. J Neurochem, 2007, 101, 1632–1643.
36. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K et al.: PANTHER: a library of protein families and subfamilies indexed by function. Genome Res, 2003, 13, 2129–2141.
37. Thomas PD, Kejariwal A: Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci USA, 2004, 101, 15398–15403.
38. Valdmanis PN, Verlaan DJ, Rouleau GA: The proportion of mutations predicted to have a deleterious effect differs between gain and loss of function genes in neurodegenerative disease. Hum Mutat, 2009, 30, E481–489.
39. Valle EP, Burgos RI, Valle JR, Egas Béjar D, Ruiz-Cabezas JC: Analysis of *CFTR* gene mutations and cystic fibrosis incidence in the Ecuadorian population. Invest Clin, 2007, 48, 91–98.
40. Wang LL, Li Y, Zhou SF: A bioinformatics approach for the phenotype prediction of nonsynonymous single nucleotide polymorphisms in human cytochromes P450. Drug Metab Dispos, 2009, 37, 977–991.
41. Wang LL, Liu YH, Meng LL, Li CG, Zhou SF: Phenotype prediction of non-synonymous single-nucleotide polymorphisms in human ATP-binding cassette transporter genes. Basic Clin Pharmacol Toxicol, 2011, 108, 94–114.
42. Wang Y, Hao DC, Stein WD, Yang L: A kinetic study of rhodamine123 pumping by P-glycoprotein. Biochim Biophys Acta, 2006, 1758, 1671–1676.
43. Webster AR, Héon E, Lotery AJ, Vandenburgh K, Casavant TL, Oh KT, Beck G et al.: An analysis of allelic variation in the ABCA4 gene. Invest Ophthalmol Vis Sci, 2001, 42, 1179–1189.
44. Xi T, Jones IM, Mohrenweiser HW: Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics, 2004, 83, 970–979.
45. Zhou JJ, Li MS, Qi J, Linsdell P. Regulation of conductance by the number of fixed positive charges in the intracellular vestibule of the CFTR chloride channel pore. J Gen Physiol, 2010, 135, 229–245.

Received: November 8, 2010; **in the revised form;** January 25, 2011; **accepted:** February 7, 2011.