Short Communication

Screening for Deleterious Nonsynonymous Single-Nucleotide Polymorphisms in Genes Involved in Steroid Hormone Metabolism and Response

Melissa M. Johnson,[1] John Houck,[1] and Chu Chen[1,2]

[1]Program in Epidemiology, Fred Hutchinson Cancer Research Center and [2]Department of Epidemiology, University of Washington, Seattle, Washington

Abstract

To facilitate selection of single-nucleotide polymorphisms (SNP) for molecular epidemiologic studies investigating the hormonal carcinogenesis hypothesis, we used two sequence homology–based tools [Sort Intolerant from Tolerant (SIFT) and Polymorphism Phenotype (PolyPhen)] to predict the potential impact a nonsynonymous SNP (nsSNP), which results in an amino acid substitution, may have on the activity of proteins encoded by genes involved in the steroid hormone metabolism and response pathway. We screened 137 variants. Of these, 28% were predicted by SIFT and PolyPhen as having a potentially damaging effect on protein function. Investigation into the association of these variant alleles with hormone-related cancers may prove to be fruitful. (Cancer Epidemiol Biomarkers Prev 2005;14(5):1326–9)

Introduction

Epidemiologic association studies focus a great amount of effort into identifying single-nucleotide polymorphisms (SNP) in genes that may have an association with disease risk. Often, the SNPs that have an association with disease are those that are known as nonsynonymous SNPs (nsSNP), which result in an amino acid substitution. These types of polymorphisms are believed to be more likely to cause a change in structure and as such compromise the function of a protein. The structure of a protein can change in various ways due to the biochemical differences of the amino acid variant (acidic, basic, or hydrophobic) and by the location of the variant in the protein sequence (by affecting tertiary or quaternary structure or the active site where substrate binds). Many molecular epidemiologic studies focus on studying SNPs found in coding regions in hopes of finding significant association between SNPs and disease susceptibility, but often find little or no association (1-3). Our ability to better select a nsSNP for an association study can be enhanced by first examining the potential impact an amino acid variant may have on the function of the encoded protein with the use of two innovative sequence homology–based programs, Sort Intolerant from Tolerant (SIFT; http://blocks.fhcrc.org/sift/SIFT.html) and Polymorphism Phenotype (PolyPhen; http://www.bork.embl-heidelberg.de/PolyPhen/).

SIFT uses sequence homology among related genes and domains across species to predict the impact of all 20 possible amino acids at a given position, allowing users to determine which nsSNPs would be of most interest to study by sorting variants by this prediction score (1, 2, 4-9). The SIFT algorithm has been shown to predict a phenotype for a nsSNP more accurately than previously used substitution scoring matrices, such as BLOSUM62, as these matrices do not incorporate information specific to the protein of interest (4, 10). Another advantage to using SIFT is the potential to analyze a larger number of nsSNPs than methods that are dependent on the availability of protein structure alone (4, 11).

The PolyPhen algorithm, like SIFT, also takes an evolutionary approach in distinguishing deleterious nsSNPs from functionally neutral ones. PolyPhen differs from SIFT in that it predicts how damaging a particular variant may be by using a set of empirical rules based on sequence, phylogenetic, and structural information characterizing a particular variant. In addition to using sequence alignments, PolyPhen utilizes protein structure databases, such as PDB (Protein Data Bank) or PQS (Protein Quarternary Structure), DSSP (Dictionary of Secondary Structure in Proteins), and three-dimensional structure databases to determine if a variant may have an effect on the protein's secondary structure, interchain contacts, functional sites, and binding sites (3, 7-9).

These two algorithms have great potential in their use to screen for potentially damaging nsSNPs in genes that may be associated with disease risk, such as hormone-related cancers like breast, prostate, and endometrial cancer. Interindividual variation in the biosynthesis and metabolism of sex hormones has been shown in population studies and may be related to risk of these types of cancers. For instance, Dunning et al. (12) found that circulating levels of several steroid hormones, such as estrogen, androgen, and their precursors, are directly related to the risk of breast cancer. There is evidence that variation in circulating levels of steroid hormones may be associated with polymorphisms of genes involved in the steroid hormone metabolism and response pathways (13-15). In this study, we screened 137 nsSNPs identified in genes in these pathways, using the SIFT and PolyPhen algorithms, to predict which of these nsSNPs would most likely have a damaging effect on the function of the encoded proteins.

Materials and Methods

nsSNP Screening Process. Two public SNP databases, dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and

Received 11/8/04; revised 12/28/04; accepted 1/6/05.

Grant support: National Cancer Institute, NIH, grant R01CA84141.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Chu Chen, Program in Epidemiology, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Mailstop M5-C800, 1100 Fairview Avenue North, Seattle, WA 98109-1024. Phone: 206-667-6644; Fax: 206-667-2537. E-mail: cchen@fhcrc.org

Table 1. SIFT and PolyPhen results for 137 amino acid variants of genes involved in steroid hormone metabolism and response

Gene (protein ID)	SNP identifier	AA variant*	SIFT score †	SIFT prediction ‡	PolyPhen prediction	Gene (protein ID)	SNP identifier	AA variant	SIFT score	SIFT prediction	PolyPhen prediction
AR	rs1800053	A646D	0.19	Tolerant	Benign	*CYP19A1*	rs2236722	W39R	(0.11)	(Tolerant)	POS
(NP_000035)	rs9332968	I665T	0.06	Tolerant	Benign	(NP_000094)	cyp19a1-44	T201M	0.06	Tolerant	Benign
	rs9332969	R841H	0.07	Tolerant	POS		rs700519	R264C	0.03	INTOL	Benign
	rs9332970	I843T	0.03	INTOL	POS		rs2304462	R264H	0.21	Tolerant	Benign
	rs9332971	R856H	0.01	INTOL	POS						
						ESR1	rs9340773	G77S	1.00	Tolerant	Benign
COMT	comt-11	G32S	(0.00)	(INTOL)	Benign	(NP_000116)	esr1-22	M264I	1.00	Tolerant	Benign
(NP_000745)	rs6270	C34S	(0.04)	(INTOL)	Benign						
	rs6267	A72S	0.31	Tolerant	Benign	*GSTM1*	rs1065411	K173N	0.33	Tolerant	Benign
	rs13306281	V92M	0.00	INTOL	Benign	(NP_000552)	rs449856	S210T	0.90	Tolerant	Benign
	rs5031015	A102T	0.56	Tolerant	Benign						
	rs4986871	A146V	0.03	INTOL	Benign	*GSTT1*	rs2266635	A21T	0.00	INTOL	Benign
	rs4680	V158M	0.02	INTOL	Benign	(NP_000844)	rs11550605	T104P	0.10	Tolerant	POS
	rs13306279	P199L	0.01	INTOL	PRB		rs2266633	D141N	0.75	Tolerant	Benign
							rs2266637	V169I	0.08	Tolerant	Benign
CYP1A1	rs4646422	G45D	0.06	Tolerant	Benign		rs2234953	E173K	0.06	Tolerant	PRB
(NP_000490)	cyp1a1-58	M66V	0.01	INTOL	POS						
	cyp1a1-96	I78T	0.00	INTOL	PRB	*HSD3B1*	rs17023677	T54I	0.53	Tolerant	Benign
	rs2229150	R93W	0.00	INTOL	PRB	(NP_000853)	rs4986952	R71I	0.01	INTOL	PRB
	cyp1a1-98	A132P	0.09	Tolerant	Benign		rs6201	I79V	1.00	Tolerant	Benign
	cyp1a1-57	I171M	0.01	INTOL	POS		rs6684974	G90S	0.02	INTOL	POS
	cyp1a1-56	T173R	0.68	Tolerant	Benign		rs6205	F286L	1.00	Tolerant	POS
	cyp1a1-21	P238S	0.01	INTOL	POS		rs1047303	T367N	0.30	Tolerant	Benign
	cyp1a1-23	R279W	0.00	INTOL	PRB						
	rs4987133	I286T	0.00	INTOL	PRB	*HSD3B2*	rs4986954	D74N	0.02	INTOL	Benign
	cyp1a1-26	M331I	0.12	Tolerant	Benign	(NP_000189)	rs6211	E94Q	1.00	Tolerant	Benign
	rs2856833	F381L	0.03	INTOL	Benign						
	cyp1a1-28	R431G	0.00	INTOL	PRB	*HSD17B1*	rs605059	S313G	(0.85)	(Tolerant)	Benign
	rs1799814	T461N	0.49	Tolerant	Benign	(NP_000404)					
	rs1048943	I462V	0.13	Tolerant	Benign						
	rs2278970	A463G	0.52	Tolerant	Benign	*LHB*	rs1800447	W28R	0.40	Tolerant	POS
						(NP_000885)	lhb-02	I35T	1.00	Tolerant	Benign
CYP1A2	cyp1a2-20	D104N	0.00	INTOL	POS		rs5030773	Q74R	0.00	INTOL	PRB
(NP_000752)	cyp1a2-31	S298R	0.06	Tolerant	POS		rs5030774	G122S	0.64	Tolerant	Benign
	cyp1a2-32	G299S	0.11	Tolerant	Benign						
	cyp1a2-44	R457W	0.00	INTOL	PRB	*NQO1*	rs11555215	R119P	0.00	INTOL	PRB
						(NP_000894)	rs4986998	R139W	0.01	INTOL	Benign
CYP1B1	rs10012	R48G	(0.27)	(Tolerant)	POS		rs1800566	P187S	0.11	Tolerant	PRB
(NP_000095)	rs9282670	Q68R	0.34	Tolerant	Benign		nqo1-04	I193T	0.33	Tolerant	Benign
	rs9282671	Y81N	0.00	INTOL	PRB						
	rs1056827	A119S	0.02	INTOL	Benign	*PGR*	rs11571143	A50T	(0.01)	(INTOL)	Benign
	rs9341248	S206N	0.03	INTOL	Benign	(NP_000917)	rs3740754	G57R	(0.44)	(Tolerant)	Benign
	rs9341250	R266L	0.26	Tolerant	Benign		rs11571144	A120V	(0.00)	(INTOL)	Benign
	rs4398252	M372V	0.01	INTOL	PRB		rs11571145	P186L	(0.00)	(INTOL)	PRB
	rs1056836	V432L	0.22	Tolerant	Benign		rs11571146	M301R	(0.00)	(INTOL)	PRB
	rs4986887	D441H	0.16	Tolerant	Benign		rs3740753	T344S	(0.68)	(Tolerant)	Benign
	rs4986888	A443G	0.05	INTOL	Benign		rs11571147	C347S	(1.00)	(Tolerant)	Benign
	rs1056837	D449E	0.05	INTOL	Benign		rs11571148	P352L	(0.00)	(INTOL)	PRB
	rs1800440	N453S	0.03	INTOL	Benign		rs11571149	S378R	(0.00)	(INTOL)	Benign
							rs11571150	A444S	(0.33)	(Tolerant)	Benign
CYP3A4	rs12721634	L15P	(0.00)	(INTOL)	PRB		rs11571151	V529L	(0.02)	(INTOL)	Benign
(NP_059488)	cyp3a4-05	G56D	0.00	INTOL	Benign		rs11571152	Q536P	(0.03)	(INTOL)	PRB
	rs3091339	K96E	0.00	INTOL	Benign		rs2020874	R625I	(0.00)	(INTOL)	POS
	cyp3a4-06	I118V	0.57	Tolerant	Benign		rs11571222	L651V	(0.11)	(Tolerant)	Benign
	rs4986907	R162Q	0.64	Tolerant	Benign		rs1042838	L660V	(1.00)	(Tolerant)	Benign
	cyp3a4-09	V170I	0.46	Tolerant	Benign		rs2020880	S865L	(0.06)	(Tolerant)	Benign
	rs4986908	D174H	0.06	Tolerant	Benign						
	rs4986908	D174N	0.02	INTOL	Benign	*SULT1A1*	rs1042008	H149Y	1.00	Tolerant	Benign
	rs12721627	T185S	0.01	INTOL	Benign	(NP_001046)	rs1042011	E151Q	0.21	Tolerant	Benign
	rs4987161	F189S	0.00	INTOL	PRB		rs1042014	E151D	1.00	Tolerant	Benign
	cyp3a4-13	P218R	0.01	INTOL	Benign		rs9282861	R213H	0.00	INTOL	POS
	cyp3a4-14	S222P	0.13	Tolerant	POS		rs1801030	V223M	0.00	INTOL	Benign
	rs3208363	S252A	1.00	Tolerant	Benign						
	rs10250778	T349N	0.02	INTOL	Benign	*SULT1A2*	rs1136703	I7T	0.47	Tolerant	Benign
	cyp3a4-17	T363M	0.00	INTOL	PRB	(NP_001045)	rs10797300	P19L	0.34	Tolerant	Benign
	rs12721629	L373F	0.12	Tolerant	Benign		rs4987024	Y62F	0.33	Tolerant	Benign
	rs4986909	P416L	0.00	INTOL	PRB		rs1126451	V86A	0.68	Tolerant	Benign
	rs1041988	I431T	0.31	Tolerant	Benign		rs11569763	H226D	0.01	INTOL	PRB
	rs4986910	M445T	0.00	INTOL	PRB		rs1059491	N235T	0.00	INTOL	Benign
	rs4986913	P467S	0.42	Tolerant	POS		rs11569766	N239S	0.00	INTOL	POS
							rs27742	E282K	1.00	Tolerant	Benign
CYP3A5	rs13220949	R439K	0.00	INTOL	PRB						
(NP_000768)	rs13233803	R495T	0.00	INTOL	PRB	*UGT2B4* (NP_066962)	rs13119049	D458E	0.00	INTOL	POS

(Continued on the following page)

Table 1. SIFT and PolyPhen results for 137 amino acid variants of genes involved in steroid hormone metabolism and response (Cont'd)

Gene (protein ID)	SNP identifier	AA variant*	SIFT score †	SIFT prediction ‡	PolyPhen prediction	Gene (protein ID)	SNP identifier	AA variant	SIFT score	SIFT prediction	PolyPhen prediction
CYP11A1	rs11544450	G15C	(0.01)	(INTOL)	Benign						
(NP_000772)	rs1130841	Y16C	(1.00)	(Tolerant)	PRB	*UGT2B7*	rs12233719	A71S	0.67	Tolerant	Benign
	rs1130843	F274L	0.00	INTOL	PRB	(NP_001065)	rs7439366	H268Y	0.00	INTOL	PRB
	rs1049968	M301I	0.82	Tolerant	Benign						
	rs6161	E314K	0.31	Tolerant	Benign	*UGT2B15*	rs1902023	D85Y	0.06	Tolerant	Benign
						(NP_001067)	rs4148269	T523K	(0.04)	(INTOL)	POS
CYP17A1	rs762563	C22W	(0.19)	(Tolerant)	PRB						
(NP_000093)											

Abbreviations: AA, amino acid; INTOL, intolerant; PRB, probably damaging; POS, possibly damaging.
*Variants listed were obtained from dbSNP (rs_ID number) and SIFT500cancer (gene name_ID number).
†TI scores ≤0.05 are predicted to be intolerant, whereas TI scores >0.05 are tolerant variants. TI scores in parenthesis have a MSCS ≥3.25 and should be interpreted with caution.
‡Predictions in parenthesis were made in low confidence (MSCS ≥ 3.25) and should be interpreted with caution.

SNP500cancer (http://snp500cancer.nci.nih.gov/home.cfm), were used to identify nsSNPs for the 25 genes of interest. From dbSNP, we identified 113 nsSNPs. An additional 24 nsSNPs were identified from the SNP500cancer database.

Obtaining nsSNP Tolerance Scores with SIFT and PolyPhen. For each gene of interest, we analyzed the protein sequence identified from dbSNP using the SIFT-version 2 database. In general, for the peptide sequence, SIFT performs multiple alignments of a number of sequences until a median conservation for the peptide sequence is reached at the default of 3.0 and predicts whether substitution with any of the other amino acids is tolerated or deleterious for every position in the submitted sequence. The SIFT prediction is given as a tolerance index (TI) score ranging from 0.0 to 1.0, which is the normalized probability that the amino acid change is tolerated. A nsSNP with a TI score of ≤0.05 is considered to be deleterious. The confidence measure of this result is given as the median sequence conservation score (MSCS) for the given codon. Deleterious predictions with a MSCS ≥3.25 were made with low confidence because the sequences used to determine the score were not diverse enough.

We adjusted the PolyPhen algorithm settings to include all sequences homologous to the peptide sequence of interest for the calculation of the difference in structural parameters instead of using the single most homologous sequence as a default. All other settings of PolyPhen were set as the default. Predictions of how a particular nsSNP may affect protein structure by PolyPhen are assigned as ''probably damaging,'' a score made with high confidence that the nsSNP should affect protein structure and/or function; ''possibly damaging,'' where it may affect protein function and/or structure; and ''benign,'' as most likely having no phenotypic effect (3). It arrives at these predictions by determining the position-specific independent count difference of the two allelic variants in the polymorphic position and also measures the degree of damaging effect a variant may have on structural parameters of a protein. The position-specific independent count is a logarithmic ratio of the likelihood of a given amino acid occurring at a particular position to the likelihood of the same amino acid occurring at any position in the sequence, also known as background frequency. Thus, PolyPhen, as mentioned before, uses a set of empirical rules based on sequence, phylogenetic, and structural information to determine if protein structure may be compromised.

Results

The screen for nsSNPs of the two online databases, dbSNP and SNP500cancer, yielded 137 nsSNPs in 25 genes associated with sex steroid biosynthesis, catabolism, and response. Of the 25 genes selected, only *UGT1A1* lacked any submitted nsSNPs in the two databases. The results of the SIFT and PolyPhen analyses are listed in Table 1. SIFT scores ≤0.05 are considered to be deleterious (TI ≤ 0.05) and SIFT scores with low confidence (MSCS ≥ 3.25) are in parentheses. Of the 137 nsSNPs screened with SIFT and PolyPhen, the 111 nsSNPs that were scored with confidence by SIFT (MSCS < 3.25) were used in determining concordance between SIFT and PolyPhen.

Table 2 gives the concordance of SIFT and PolyPhen predictions for only the 111 variants scored with confidence by SIFT. Of these 111 nsSNPs, 31 (28%) were predicted to have a negative effect on protein function (SIFT: TI ≤ 0.05 with confidence; PolyPhen: probably damaging/possibly damaging) and 50 (45%) were predicted to be tolerated variants (TI > 0.05 with confidence; benign). The remaining 30 nsSNPs (27%) of these variants had conflicting results between the SIFT and PolyPhen algorithms.

Discussion

We identified a total of 66 of 137 nsSNPs that were scored as intolerant by SIFT; 14 of these were scored with low confidence. A total of 30 nsSNPs were scored as ''probably damaging'' and 21 as ''possibly damaging'' by PolyPhen. Out of the 66 nsSNPs classified as intolerant by SIFT, 17 variants are distributed among a number of genes: one in *LHB*, one in *CYP11A1*, five in *CYP1A1*, one in *CYP1A2*, one in *CYP1B1*, four in *CYP3A4*, two in *CYP3A5*, one in *NQO1*, and one in *UGT2B7*, and were scored as ''most intolerant'' by SIFT (TI = 0.0 with confidence) and probably damaging by PolyPhen. Thirty-one variants were scored as having some damaging effect on protein function (TI ≤ 0.05; probably damaging/possibly damaging). It would be interesting to study these particular variants in an epidemiologic study to assess possible associations with the risk of hormone-related cancers. Of the 111 nsSNPs scored with confidence by SIFT, 30 nsSNPs (27%) were found to have conflicting results. A possible explanation as to why SIFT and PolyPhen may differ in their prediction scores could be that PolyPhen, in addition to using sequence information, incorporates structural information of the protein (obtained from PQS and DSSP).

Several studies have used SIFT and/or PolyPhen to assess function of nsSNPs in candidate genes for association studies, but none have reported such assessment of the steroid hormone metabolism pathway (2, 9). Based on an evaluation of published association of nsSNPs and cancer risk, Zhu et al. (1) concluded that the risk of a number of cancers were significantly inversely correlated with SIFT

Table 2. Concordance of SIFT and PolyPhen predictions for the 111 variants scored with confidence by SIFT of genes involved in steroid hormone metabolism and response

PolyPhen prediction	SIFT prediction and score			
	Most intolerant (TI score, 0.00)	Intolerant (TI score, 0.01-0.05)	Borderline tolerant (TI score, 0.06-0.10)	Tolerant (TI score, 0.11-1.00)
Probably damaging	17 (15.3)*	4 (3.6)	1 (0.9)	1 (0.9)
Possibly damaging	4 (3.6)	6 (5.4)	3 (2.7)	4 (3.6)
Benign	6 (5.4)	15 (13.5)	7 (6.3)	43 (38.7)

*Number (%) of variants. These numbers are representative of only the 111 nsSNPs that were scored with confidence by SIFT (MSCS < 3.25) of the 137 nsSNPs screened by SIFT and PolyPhen.

predictions. Their results also suggest that variants that occur in conserved sequences are more likely to be associated with cancer susceptibility. Two studies thus far have combined both the SIFT and PolyPhen algorithms to screen for deleterious nsSNPs. In one paper, Xi et al. (7) examined the DNA repair pathways and found that the predictions made by the two programs are highly correlated, with ~62% concordance in the predictions of deleterious or potentially damaging nsSNPs. We observed a similar concordance of 73% between SIFT and PolyPhen predictions, but in genes involved in steroid hormone metabolism and response. In a second, more recent, study, Livingston et al. (8) used SIFT and PolyPhen to identify 57 potentially deleterious nsSNPs involved in DNA repair, cell cycle regulation, apoptosis, and drug metabolism.

Whereas useful, SIFT and PolyPhen are nonetheless somewhat limited. This is because both require available sequence data. However, as more whole genome sequences become publicly available, prediction of functionality of more nsSNPs will become possible and the results more reliable (5, 9). When there are too few sequences available on a particular gene for comparison, for example, the *PGR* gene, or when the sequences seem to be too homologous to one another, a nonfunctional variant may be predicted to be ''intolerant'' (6, 9). When the MSCS is >3.25, Savas et al. (2) interpreted the SIFT predictions as ''possibly affecting'' or ''possibly tolerated.'' However, like Ng and Henikoff (4, 5), we chose to disregard nsSNPs that were scored with low confidence by SIFT until more sequence data become publicly available to allow more reliable predictions to be made. Another limitation of these algorithms is that the impacts of a combination of variants are not assessed and the dependence of functional impact of a variant on genotype of other genes or on exposure risk is not addressed (9). One final restriction of SIFT and PolyPhen is that the algorithms are unable to predict the impact of SNPs that occur outside of the coding region, such as promoter and enhancer regions, and splice sites that may affect protein levels or protein function.

To those conducting large-scale population-based epidemiologic studies, the idea of prioritizing nsSNPs in the investigation of association of SNPs with disease risk is of great interest. The use of SIFT and PolyPhen to select potentially intolerant nsSNPs for epidemiology studies can be an efficient way to explore the role of genetic variation in disease risk and to contain cost. Furthermore, predicted impact of these nsSNPs can be tested with the use of animal models and/or cell line systems to determine if functionality of the protein has indeed been altered.

References

1. Zhu Y, Spitz MR, Amos CI, Lin J, Schabath MB, Wu X. An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology. Cancer Res 2004;64:2251–7.
2. Savas S, Kim DY, Ahmad MF, Shariff M, Ozcelik H. Identifying functional genetic variants in DNA repair pathway using protein conservation analysis. Cancer Epidemiol Biomarkers Prev 2004;13:801–7.
3. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894–900.
4. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001;11:863–74.
5. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002;12:436–46.
6. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4.
7. 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;83:970–9.
8. Livingston RJ, von Niederhausern A, Jegga AG, et al. Pattern of sequence variation across 213 environmental response genes. Genome Res 2004;14:1821–31. Epub 2004 Sep 13.
9. Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. Nat Rev Genet 2004;5:589–97.
10. Henikoff S, Henikoff JG. Automated assembly of protein blocks for database searching. Nucleic Acids Res 1991;19:6565–72.
11. Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet 2000;16:198–200.
12. Dunning AM, Dowsett M, Healey CS, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. J Natl Cancer Inst 2004;96:936–45.
13. Kristensen VN, Borresen-Dale AL. Molecular epidemiology of breast cancer: genetic variation in steroid hormone metabolism. Mutat Res 2000;462:323–33.
14. Henderson BE, Feigelson HS. Hormonal carcinogenesis. Carcinogenesis 2000;21:427–33.
15. Mitrunen K, Hirvonen A. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. Mutat Res 2003;544:9–41.