Software Prediction of the Effects of Single Nucleotide Polymorphisms

Angela Collie
CSE 527
December 13, 2004

A. Introduction

Single Nucleotide Polymorphisms (SNPs)

Although 99.9% of the base pairs in the human genome are the same between individuals, the remaining variation is crucial to the understanding of genetics and disease.[1] Much of this variation is in single nucleotide polymorphisms (SNPs), which are single base pair differences in the DNA sequence that are, by definition, present in at least 1% of a population group.[2] This means that the minor allele, the allele containing the rare SNP rather than the more common base pair, is present in greater than 1% of a population such as European Americans. It has been estimated that there are approximately 6 million "common" SNPs with a minor allele frequency greater than 10%.[1,3] These common SNPs are estimated to be present about every 600 base pairs.[1]

Several methods can be used to identify SNPs including traditional sequencing, microarray genotyping, Molecular Beacon genotyping, 5' Nuclease Assay with Taqman probes, allele-specific PCR, and primer extension-based assays.[4,5] The type of identification method has been shown by one group to affect the reproducibility of SNP data.[3]

Several SNP databases are available. One of the largest is the publicly available Single Nucleotide Polymorphism database (dbSNP), established by the National Center for Biotechnology Information (NCBI). Each unique SNP is assigned a reference SNP ID or rs ID. The most recent build, build 123, of October 28, 2004 contains 10,079,771 rs IDs with 5,007,794 as validated.[6]

SNPs can be present in any portion of a gene, including regulatory motifs, promoters, 5'untranslated region, 3'untranslated region, splice sites, exons and introns. The SNPs affecting the protein sequence, and potentially the protein structure, are located in the coding region of the relevant gene. SNPs in the coding region of the gene are defined as synonymous, those resulting in no amino acid change in the protein, or nonsynonymous, those resulting in an amino acid change in the protein.

In the most recent build of dbSNP, there are 51,220 nonsynonymous SNP alleles in 15,704 genes.[7] As the number of SNPs is large, it is beneficial for biological researchers to prioritize or reduce the number of candidate SNPs to examine in a gene. Software prediction of the effect of a nonsynonymous SNP (nsSNP) on a gene allows this prioritization.

SNP Software Prediction

Several groups have attempted to predict the effect of nsSNPs on the resulting protein.[8-15] For example, the database of topographic mapping of Single Nucleotide Polymorphism (topoSNP) examines nsSNPs and displays them onto a relevant protein structure with geometric location data for the altered amino acid and an entropy calculation based on a Hidden Markov Model.[14] This database contained no data on the genes or proteins of interest and was not used in our study. Two additional programs, PolyPhen[12] and SIFT[9-11] will analyze any SNP and were utilized in this study.

PolyPhen

Polymorphism Phenotyping (PolyPhen) also attempts to predict the effect of an nsSNP on a protein. PolyPhen is a web-based software tool for the prediction of the effect of nsSNPs on the resulting protein. It is available at http://www.bork.embl-heidelberg.de/PolyPhen/. PolyPhen uses sequence homology and protein structure to makes its predictions based on a set of rules, as shown in Figure 1. The input to PolyPhen is an amino acid sequence or corresponding ID, the position of the amino acid varied, and the amino acid variants. One amino acid variant will correspond to the amino acid in the reference sequence and the other will correspond to the amino acid resulting from the nsSNP. The PolyPhen algorithm takes the input data and analyzes it in three steps, utilizing multiple other programs and several databases.[12]

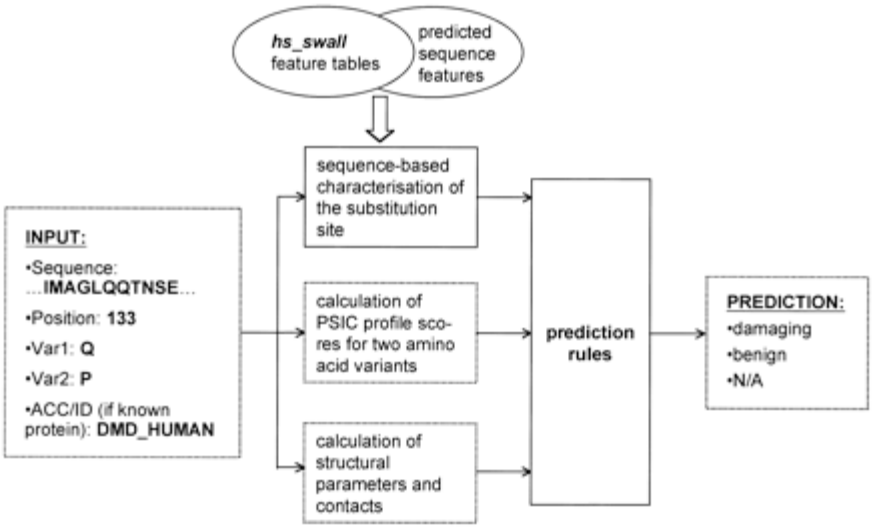

Figure 1. PolyPhen Data Flow[12]

PolyPhen begins with sequence-based characterization of the substitution site. First, it uses the corresponding protein description from the SWALL database to identify if the amino acid is part of a protein feature, such as a bond, a binding site, an active site, or a transmembrane region. At this point, the program stores some data to be analyzed in 3D structure comparisons. Next, several prediction programs TMHMM, Coils2, and SignalP are used to identify transmembrane regions, coils, and signal peptides, respectively. If the substitution is in the transmembrane region, the program uses the PHAT transmembrane-specific score to estimate the effect of the substitution.

The second step is the profile analysis of homologous sequences. PolyPhen uses BLAST against NRDB to locate proteins with 30-94% sequence identity. The aligned sequences are put into another software program called Position-Specific Independent Counts (PSIC).[16] This returns a matrix of scores. The score is based on the log likelihood ratio of the probability of the amino acid variant occurring at the specified position to the likelihood of the amino acid at any position, shown in Equation 1.

$$W(a,i) = \ln\left[\frac{p(a,i)}{q_a}\right]$$

Equation 1[16]

Where W (a,i) is the profile score and p(a,i) is the probability of amino acid a being seen at position I and q_a is the probability of amino acid in the background sequence.

This is a simplification, and there are further normalizations and estimations made due to the limited number of sequences available to be analyzed and the potential interdependence of sequences. The program calculates the PSIC scores for the two amino acids entered. It utilizes the difference between the scores in its analysis. The number of aligned sequences is also returned.

Finally, the program calculates structural parameters and contacts. PolyPhen BLASTs the sequence against the user-chosen PDB or PQS databases to find proteins of sequence identity of at least 50%. Second, several structural parameters are then calculated using the DSSP database and the HBplus program. Thirdly, PolyPhen checks contacts of the amino acid with ligands, interactions between parts of the protein, and critical residues. Critical residues are those found in the sequence-based characterization.

The program then uses the data described in the three steps above and the set of empirical rules shown in Table 1. It returns an output with a prediction of "probably damaging," "possibly damaging," "benign" or "unknown" for the given nsSNP-related amino acid change.

Table 1. PolyPhen Prediction Rules[12]

Rules (connected with logical AND)			**Prediction**
PSIC score difference	**Substitution site properties**	**Substitution type properties**	
Arbitrary	Annotated as a functional[a] or bond formation[b] site	Arbitrary	Probably damaging
Not considered	In a region annotated or predicted as transmembrane	PHAT matrix difference resulting from substitution is negative	Possibly damaging
Less than 0.5	Arbitrary	Arbitrary	Benign
Greater than 1.0	Atoms are closer than 3.0 Å to atoms of a ligand or residue annotated as BINDING, ACT_SITE, LIPID, METAL	Arbitrary	Probably damaging
Between 0.5 and 1.5	Normed accessibility ACC 15%	Absolute change of accessible surface propensity is 0.75 or absolute change of side chain volume is 60	Possibly damaging
Between 0.5 and 1.5	Normed accessibility ACC 5%	Absolute change of accessible surface propensity is 1.0 or absolute change of side chain volume is 80	Probably damaging
Between 1.5 and 2.0	Arbitrary	Arbitrary	Possibly damaging

Greater than 2.0	Arbitrary	Arbitrary	Probably damaging

The output report also shows multiple other factors that were calculated or on which the prediction was made. It is of note that not all calculated or observed parameters are used in the prediction.

Sorting Intolerant From Tolerant (SIFT)

Sorting Intolerant From Tolerant (SIFT) is another web-based software prediction tool for nsSNPs that was developed at the Fred Hutchinson Cancer Research Center. It is available at http://blocks.fhcrc.org/sift/SIFT.html. This program utilizes sequence homology to determine the effect of an nsSNP on a protein sequence. Several versions of this program have been described and differ in the protocol used to select homologous sequences.[9-11] The most recent version SIFT BLink Beta was utilized in this study.

In the SIFT BLink Beta version, the input to the program was the GI# for each protein of interest and the amino acid substitutions. The amino acid substitutions were in the format X#Y, where X was the one letter code for the amino acid in the reference sequence, # was the position of the amino acid, and Y was the nsSNP variant amino acid. Several SNPs could be submitted simultaneously. Homologous sequences in this version are chosen from BLink (BLAST Link), NCBI's precomputed BLAST searches for homologous proteins.[17] Blink returns an aligned set of homologous proteins.

The aligned proteins are analyzed using $R_c = \log_2 20 - \Sigma_{20aa}\, p_{ca} \log p_{ca}$ where p_{ca} is the probability that amino acid *a* is at position *c*.[9] This provides median sequence information.

The alignment is then converted into a position-specific scoring matrix (PSSM), which is an l x 20 matrix, where l is the length of the query sequence. Each matrix component is the probably of amino acid *a* at position *c*, or p_{ca}. p_{ca} is estimated based on Equation 2:

$$p_{ca} = \frac{N_c}{(N_c + B_c)} * g_{ca} + \frac{B_c}{(N_c + B_c)} * f_{ca}$$

Equation 2.[9]

where N_c is the total number of sequences, g_{ca} is the sequence-weighted frequency that a appears at c in the alignment. f_{ca} is a term that takes care of pseudocounts and is estimated based on a Dirichlet mixture. B_c is the number of pseudocounts. B_c is 0 at a position with no variability but is equal to exp($SUM_a(r_a * g_{ca})$. r_a is the rank of an amino acid as calculated from the BLOSUM2 matrix column for the reference amino acid. The BLOSUM62 matrix is a 20x20 substitution matrix of positive and negative integers that can be used for protein alignment and will select for sequences with 62% identical homology.[18,19]

The probabilities for amino acid substitutions at a given position *c* are then divided by the highest probability for an amino acid at position *c* to give normalized probabilities. SIFT then predicts that amino acids in a position with normalized probabilities of less than 0.05 are deleterious. Amino acids with probabilities greater than 0.05 are predicted to be tolerated.

The output of SIFT is a table of probabilities for each amino acid at each position as well as predictions on not tolerated or tolerated amino acids for each position. If specific

positions were input, SIFT returns a prediction, the number of sequences that were used in the alignment as well as the median sequence information. SIFT also warns that median sequence information above 3.25 represents sequences that are very similar and may lead to false results.

A recent study examined genes of interests with two of the described programs, PolyPhen and SIFT.[20] In the current study, a similar effort has been made to examine genes involved with inflammation. The inflammatory pathway chosen is that of interleukin 1, beta (IL-1B), which is involved in many proinflammatory processes, including the response to biomaterials (Anderson, JM), making it of interest to our group.[21,22]

Five genes related to the IL-1B pathway, were chosen. Interestingly, IL-1B did not contain any nsSNPs. Interleukin 1 receptor, type 1 (IL1R1) is a cell surface receptor for IL-1B. Two of the proteins downstream of IL1R1 are IL-1 receptor-associated kinase 1 (IRAK1) and the recently described IL-1 receptor-associated kinase 4 (IRAK4).[23] In addition, stimulation of IL1R1 leads to activation of the NF-KB transcription pathway, resulting in transcription of two cytokine ligands that have proinflammatory activities similar to IL-1B. These two cytokine ligands are tumor necrosis factor (TNF) and interleukin 6 (IL-6) and are also of interest in our work.

B. Materials and Methods

Selection of SNPs

Five genes related to the IL-1B inflammatory pathway were selected for the SNP study. They were IL1R1, IRAK1, IRAK4, TNF and IL-6. Each gene was identified in LocusLink and then linked to dbSNP. All nsSNPs, validated or not, were included in the analysis. There were 18 nsSNPs in dbSNP for the five genes. The SNPs are shown in Table 2 with the data available from dbSNP.

Table 2. Nonsynonymous SNPs in Five Genes of the Il-1B Signaling Pathway

Gene	Locus ID	NCBI GI #	dbSNP rs#	Minor Allele Frequency	Validated	Date of First dbSNP Submission	dbSNP Allele (Reference/ SNP)	Amino acid (Reference residue/SNP residue)	Codon position	Amino Acid Position
IL1R1	3554	4504659	rs2228139	0.09	Yes	6/19/2001	C/G	Ala [A]/Gly [G]	2	124
IRAK1	3654	4504717	rs1059703	0.713	Yes	9/13/2000	C/T	Ser [S]/ Leu [L]	2	532
			rs12860727	N.D.	No	3/19/2004	C/G	Arg [R]/Gly [G]	1	315
			rs10127175	0.042	Yes	11/15/2003	T/A	Cys [C]/Ser [S]	1	203
			rs1059702	0.13	Yes	9/13/2000	T/C	Phe [F]/Ser [S]	2	196
			rs11465830	0.138	Yes	11/14/2003	G/A	Arg [R]/His [H]	2	194
			rs11465829	0.062	Yes	11/14/2003	C/T	Thr[T]/Ile [I]	2	113
IRAK4	51135	7705841	rs4251469	0.049	No	12/11/2002	T/G	Ser [S]/Arg [R]	3	98
			rs4251583	0.044	No	12/11/2002	A/G	His [H]/Arg [R]	2	390
			rs4251545	0.303	Yes	12/11/2002	G/A	Ala [A]/Thr [T]	1	428
TNF	7124	25952111	rs3179060	N.D.	No	4/25/2002	C/A	His[H]/Asn [N]	1	52
			rs4645843	0.011	No	1/15/2003	C/T	Pro [P]/Leu [L]	2	84
			rs1800620	N.D.	No	11/7/2000	G/A	Ala [A]/Thr [T]	1	94
			rs11574936	0.049	No	1/28/2004	T/A	Ile [I]/Asn [N]	2	194
IL-6	3569	10834984	rs2069830	0.08	Yes	5/18/2001	C/T	Pro [P]/Ser [S]	1	32
			rs11544633	N.D.	No	11/18/2003	T/C	Leu [L]/Pro [P]	2	119
			rs2069860	0.042	Yes	5/18/2001	A/T	Asp [D]/Val [V]	2	162
			rs13306435	N.D.	No	3/22/2004	T/A	Asp [D]/Glu [E]	3	162

Validation for each gene could be determined through several methods: multiple independent submissions, validation by the HAPMAP project, frequency or number of observation data. The date of first dbSNP submission is the date when the polymorphism was submitted to dbSNP. Data would have been accessible with the next dbSNP build. Color coding for the amino acids corresponds to the color code used in SIFT. Nonpolar residues are colored black, uncharged polar residues are green, acidic residues are blue, and basic residues are red.

PolyPhen

PolyPhen, available at http://www.bork.embl-heidelberg.de/PolyPhen/, was run with the default parameters. For each of the 18 SNPs, the input was the protein sequence in fasta format, the amino acid position, and the amino acids of the reference sequence and the SNP. Each of the amino acid substitutions was predicted as probably damaging," possibly damaging," "benign", or "unknown."

SIFT

SIFT Blink Beta, available at http://blocks.fhcrc.org/sift/SIFT_BLink_submit.html, was used in this analysis with the default parameters of best BLAST hits to each organism and removal of sequences with 90% identity to the query sequence. The input to the program was the GI# for each protein of interest and the amino acid substitutions. The amino acid substitutions were in the format X#Y, where X was the one letter code for the amino acid in the reference sequence, # was the position of the amino acid, and Y was the one letter code for the nsSNP variant amino acid. Amino acid substitutions with scores of less than 0.05 are predicted as "affect protein function." All other amino acids are predicted as "tolerated."

C. Results

The nsSNPs for each of the five genes were analyzed using PolyPhen and SIFT. The results are shown in Tables 3 and 4.

Table 3. PolyPhen Results

Gene	Reference amino acid, amino acid position, SNP amino acid	PolyPhen Prediction	PSIC Score Difference	Number of Sequence Observations Used in Alignment	# of Protein Structures Used	Basis For Prediction	Notes
IL1R1	A124G	Benign	0.101	16	3	alignment	
IRAK1	S532L	Possibly damaging	1.350	6	0	alignment	
	R315G	Possibly damaging	1.575	8	0	alignment	
	C203S	Probably damaging	2.250	8	0	alignment	
	F196S	Benign	1.250	8	0	alignment	
	R194H	Benign	0.639	8	0	alignment	
	T113I	Benign	1.350	8	0	alignment	
IRAK4	S98R	Benign	0.900	2	0	alignment	*Based on only 2 sequence alignments
	H390R	Benign	0.225	2	0	alignment	*Based on only 2 sequence alignments
	A428T	Benign	0.900	2	0	alignment	*Based on only 2 sequence alignments
TNF	H52N	Benign	1.222	64	0	sequence annotation	Substitution is in Transmembrane region
	P84L	Possibly damaging	0.839	90	12	structure	Hydrophobicity change at buried site
	A94T	Benign	0.695	93	30	alignment	
	I194N	Probably damaging	2.476	93	30	alignment	Hydrophobicity change at buried site
IL6	P32S	Benign	0.225	14	0	alignment	
	L119P	Probably damaging	2.517	39	2	alignment	Hydrophobicity change at buried site
	D162V	Benign	1.026	35	2	alignment	Charge change at exposed site
	D162E	Benign	0.549	35	2	alignment	

Of the 18 SNPs examined, 3 were predicted as “probably damaging”, 3 were predicted as “possibly damaging” and the remainder were predicted to be “benign.”

A strength of the PolyPhen program is its incorporation of structural data into its predictions. Of the 18 SNPS examined, one prediction was made on the basis of structure and another was made based on sequence annotation, as the variant amino acid was located in a transmembrane region. The remaining 16 predictions were based on the analysis of aligned homologous sequences. This reliance on sequence for predictions may be related to the lack of available structural data. Only 7 of the 18 SNPs had any structural data available and only 3 SNPs had more than 3 homologous structures for comparison. The lack of sequence data was also seen in one gene, as the program found only 2 homologous sequences for the recently-described IRAK4. The program does not warn the user about the lack of data available for prediction.

Table 4. SIFT results

Gene	Reference amino acid, amino acid position, SNP amino acid	SIFT prediction	Score	Median sequence conservation	Sequences represented at this position
IL1R1	A124G	Tolerated	0.43	2.59	21
IRAK1	S532L	Tolerated	0.83	2.88	10
	R315G	Affect Protein Function	0.00	2.65	23
	C203S	Tolerated	0.74	2.73	20
	F196S	Affect Protein Function	0.03	2.69	20
	R194H	Tolerated	0.55	2.70	19
	T113I	Tolerated	0.12	2.83	13
IRAK4	S98R	Affect Protein Function	0.01	3.32	7
	H390R	Tolerated	0.41	2.70	26
	A428T	Tolerated	0.57	2.70	26
TNF	H52N	Tolerated	0.29	2.76	45
	P84L	Tolerated	0.20	2.74	44
	A94T	Affect Protein Function	0.02	2.64	48
	I194N	Affect Protein Function	0.00	2.64	55
IL6	P32S	Tolerated	0.47	2.88	29
	L119P	Affect Protein Function	0.00	2.61	38
	D162V	Tolerated	0.25	2.61	30
	D162E	Tolerated	0.52	2.61	30

The SIFT program returned an “Affect Protein Function” prediction for 6 SNPs with scores less than 0.05 and a “Tolerated” prediction for the remaining 12 SNPs. More than 10 homologous sequences were available for alignment for 17 SNPs. The remaining SNP, S98R in IRAK4, had only 7 homologous sequences. The program returned a warning for this SNP. The median sequence conservation of this SNP was 3.32, which was above the 3.25 recommended cutoff.

The PolyPhen and SIFT results were compared as shown in Table 5. The programs returned the same predictions, ‘probably damaging” and “affect protein function” or “benign” and “tolerated” for 11 SNPs, or 61% of the SNPs. They made opposite

predictions for 4 SNPs or 22%. The remaining 3 predictions, or 17%, were ambiguous due to the "possibly damaging" prediction from PolyPhen.

Table 5. Comparison of PolyPhen and SIFT predictions

Gene	Reference amino acid, amino acid position, SNP amino acid	PolyPhen Prediction	SIFT prediction
IL1R1	A124G	Benign	Tolerated
IRAK1	S532L	Possibly damaging	Tolerated
	R315G	Possibly damaging	Affect Protein Function
	C203S	Probably damaging	Tolerated
	F196S	Benign	Affect Protein Function
	R194H	Benign	Tolerated
	T113I	Benign	Tolerated
IRAK4	S98R	Benign	Affect Protein Function
	H390R	Benign	Tolerated
	A428T	Benign	Tolerated
TNF	H52N	Benign	Tolerated
	P84L	Possibly damaging	Tolerated
	A94T	Benign	Affect Protein Function
	I194N	Probably damaging	Affect Protein Function
IL6	P32S	Benign	Tolerated
	L119P	Probably damaging	Affect Protein Function
	D162V	Benign	Tolerated
	D162E	Benign	Tolerated

D. Discussion

The prioritization or selection of interesting SNPs to study in experimental situations is important due to the large number of nsSNPs that may affect a disease phenotype. Software programs such as PolyPhen and SIFT provide an estimate of the effect of an amino acid substitution on a protein. Both have web-based interfaces that are easy to use for the most novice user. The inputs are standard sequence formats or ID numbers as well as the SNP information.

With the five genes chosen in this analysis, it was difficult to compare the correctness of predictions. No references for the 18 nsSNPs were found in the literature

through searches of PubMed, and they were not described in the OMIM database. A recent study used PolyPhen and SIFT to examine SNPs in hemoglobin chains (Hb) and glucose-6-phosphate dehydrogenase (G6PD) as well as two less well-described genes.[20] For the Hb genes, the programs worked surprisingly well, yet failed to accurately predict the most common disease allele as damaging. The authors state that this is due to a failure to analyze supramolecular interactions. For G6PD, the authors track the predictions compared to severity of disease and can show trends in that the more SNPs causing more severe phenotypes correlated with a higher percentage of "probably damaging" or "affect protein function" predictions. The genes in the current study have not been described in the literature and thus are not applicable to evaluating the accuracy of the programs.

Difficulties associated with the use of such prediction programs are the limited number of protein sequences and structures available. This was seen with the recently-described gene, IRAK4. Only 2 homologous sequences and no structures were available to PolyPhen for this gene. SIFT found 7 and 26 homologous sequences. SIFT did provide a warning that the prediction based on 7 sequences was made with "low confidence." PolyPhen provided no such warning with its prediction. As additional sequences are found and structures are determined, predictions will improve.

Both PolyPhen and SIFT provide an automated prediction of an SNP's effect on a protein and may allows researchers to prioritize SNPs of interest. However, these programs are limited in their prediction capability, reemphasizing the need for ongoing phenotypic research.

E. Literature Citations

1. Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet 2001;27(3):234-6.
2. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405(6788):847-56.
3. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 2003;33(4):518-21.
4. Carlson CS, Newman TL, Nickerson DA. SNPing in the human genome. Current Opinion in Chemical Biology 2001;5(1):78-85.
5. Kwok PY. Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet 2001;2:235-58.
6. NCBI. "dbSNP Summary" http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi. 12/10/2004.
7. NCBI. "Functional Summary for sbSNP build 123" http://www.ncbi.nlm.nih.gov/SNP/snp_summary_byOrg.cgi?tax_id=9606&build_id=123. Volume 12/10/2004; 12/10/2004.
8. CLIFFORD RJ, EDMONSON MN, NGUYEN C, SCHERPBIER T, HU Y, BUETOW KH. Bioinformatics Tools for Single Nucleotide Polymorphism Discovery and Analysis. Ann NY Acad Sci 2004;1020(1):101-109.
9. Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. Genome Res. 2001;11(5):863-874.
10. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002;12(3):436-46.

11. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31(13):3812-4.
12. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucl. Acids Res. 2002;30(17):3894-3900.
13. Sunyaev S, Ramensky V, Koch I, Lathe W, 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet 2001;10(6):591-7.
14. Stitziel NO, Binkowski TA, Tseng YY, Kasif S, Liang J. topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association. Nucleic Acids Res 2004;32 Database issue:D520-2.
15. Riva A, Kohane IS. SNPper: retrieval and analysis of human SNPs. Bioinformatics 2002;18(12):1681-5.
16. 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(5):387-94.
17. NCBI. "BLink Help Document" http://www.ncbi.nlm.nih.gov/sutils/static/blinkhelp.html. 12/10/2004.
18. NCBI. "BLOSUM62 Matrix" http://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt. 12/10/2004.
19. NCBI. "NCBI FieldGuide Glossary" http://www.ncbi.nlm.nih.gov/Class/FieldGuide/glossary.html. 12/10/2004.
20. Tchernitchko D, Goossens M, Wajcman H. In Silico Prediction of the Deleterious Effect of a Mutation: Proceed with Caution in Clinical Genetics. Clin Chem 2004;50(11):1974-1978.
21. O'Neill LA, Dinarello CA. The IL-1 receptor/toll-like receptor superfamily: crucial receptors for inflammation and host defense. Immunol Today 2000;21(5):206-9.
22. Miller KM, Anderson JM. Human monocyte/macrophage activation and interleukin 1 generation by biomedical polymers. J Biomed Mater Res 1988;22(8):713-31.
23. Li S, Strelow A, Fontana EJ, Wesche H. IRAK-4: a novel member of the IRAK family with the properties of an IRAK-kinase. Proc Natl Acad Sci U S A 2002;99(8):5567-72.