



## Identification of deleterious mutation in CHEK2 gene an insilico approach

Punyadeep<sup>1</sup>, Priyanka singh<sup>2</sup>, C. Sudandiradoss\*

School of Biological science and Technology, VIT University, Vellore- 632014, Tamil Nadu  
Phone: +91 9705449862, Email ID: deep.punya@gmail.com

Received: May 28<sup>th</sup> 2012, Accepted: June 20<sup>th</sup> 2012, Published: July 1<sup>st</sup> 2012

### Abstract:

Single-nucleotide polymorphisms (SNPs) play a major role in the understanding of the genetic basis of many complex human diseases. The current study reports, of the total of 682 SNPs in the CHK2 gene, 10 non-synonymous SNPs (nsSNPs) were identified as functionally significant. SIFT, PolyPhen and I-Mutant programs were used to analyze the nsSNPs. Out of total 10 nsSNPs, 8 were found to be deleterious by SIFT and 5 were damaging as per the PolyPhen program. The complete analysis of these mutations has been the focus of the paper. The structural stability of the mutated structures was studied using I-Mutant program. From a comparison of the stabilizing residues of the native and mutant proteins, we propose that the nsSNP rs121908706, rs121908705, rs121908701, rs112032633 and rs111292981 could be important candidates for the various cancers caused by the mutation in CHEK2 gene.

**Keywords:** *CHEK2 gene, SIFT, PolyPhen, I-Mutant.*

### Introduction:

CHEK2 is the symbol for the human gene CHK2 checkpoint homolog. It is located on the long (q) arm of chromosome 22 at position 11[3]. The *CHEK2* gene give signals for making a protein termed checkpoint kinase 2 (CHK2). This protein acts as a tumor suppressor, which regulates cell division by keeping cells from growing and dividing too rapidly or in an uncontrolled way. Damaged DNA or the breakage in the DNA strands activates the CHK2 protein. In response to DNA damage, the CHK2 protein interacts with several other proteins, including tumor protein 53. These proteins halt cell division and determine whether a cell will repair the damage or undergo for apoptosis. This process keeps cells with mutated or damaged DNA from

dividing and helps in prevention of the tumor development [5]. Activation by phosphorylation of upstream kinase Ataxia Telangiectasia Mutated (ATM) activates the oligomerization and auto-phosphorylation of CHK2 gene [6] which spreads the damage signal by phosphorylating multiple substrates including the Cdc25A and Cdc25C phosphates, BRCA1, and p53. This triggers different pathways involved in growth arrest or apoptosis. Mutations in CHEK2 have been identified in a broad range of human tumor types like breast, colon, and prostate cancer [10]. Mutations in the *CHEK2* gene have also been found in other hereditary and nonhereditary (sporadic) cancers affecting many of the body's organs and tissues. Although the full range of cancers associated with *CHEK2* mutations has not been determined, studies have found associations between the mutations in this gene with prostate, lung, colon, kidney, thyroid, and ovarian cancers. *CHEK2* mutations have also been found in some brain tumors and in a type of bone cancer called osteosarcoma.

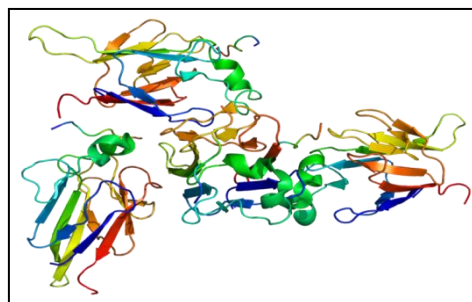


Fig: 1 Structure of chek2 protein

**Materials and Methods:**

**Datasets**

The SNPs and their related protein sequence for the CHEK2 gene were retrieved from the dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) for computational analysis. The human CHEK2 gene SNP IDs are rs121908706, rs121908705, rs121908704, rs121908703, rs121908702, rs121908701, rs121908694, rs113947614, rs112032633 and rs111292981.

**Analysis of functional consequences of coding nsSNPs by SIFT program**

SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution in a protein will have a phenotypic effect. It is (<http://blocks.fhcrc.org/sift/SIFT.html>) used to detect the deleterious coding non synonymous SNPs. SIFT dbSNP tool used, provides SIFT predictions a list of non synonymous SNPs (rsIDs) retrieved from NCBI. A query sequence is submitted and using multiple sequence alignment information the tolerated and deleterious substitutions for every position of the query sequence were determined. It is a multistep procedure works on four basic steps.

- (1) Searches for similar sequences
- (2) Chooses closely related sequences that may share similar function to the query sequence
- (3) Obtains the alignment of these chosen sequences
- (4) Calculates normalized probabilities for all possible substitutions from the alignment.

**Simulation for functional change in coding nsSNPs by PolyPhen program:**

**PolyPhen** (*Polymorphism Phenotyping*) is an automatic tool used for prediction of possible impact of an amino acid substitution on the structure and function of a human protein. Analyzing the damaged, coding non synonymous SNPs at the structural level is very important to understand the functional activity of the protein. The query were submitted to PolyPhen server either as protein sequence or SWALL database ID or accession number together with sequence position and two amino acid variants. It calculates PSIC scores for each

of the two variants and then computes the PSIC score difference between them which shows higher the PSIC score difference, higher is the functional impact on a particular amino acid substitution.

**Predicting protein stability upon mutation by I-Mutant 2.0 program:**

I-Mutant2.0 is a Support Vector Machine-based web server for the automatic prediction of protein stability changes upon single-site mutations [24]. I-Mutant2.0 is optimized to predict the protein stability change upon mutation either starting from the protein structure or the protein sequence. When tested with a cross validation procedure, I-Mutant2.0 correctly predicts whether the protein mutation stabilizes or destabilizes the protein.

- (1) **Protein Sequence:** The protein sequence in raw format and one letter code
- (2) **Position:** The position number in the sequence of the residue that undergoes mutation
- (3) **pH:** The negative logarithm value of H<sup>+</sup> concentration [0- 14]
- (4) **Temperature:** The Temperature value in Celsius degrees [0-100]

**Result and Discussion:**

**Table 1: Showing SIFT Results:**

Gene name	SNP IDs	Nucleotide change	Amino acid change	Tolerance index
CHEK2	rs121908706	G - A	R- H	0.00
	rs121908705	A - G	N- D	0.01
	rs121908704	A - G	T- A	0.00
	rs121908703	C - T	S - L	0.00
	rrs121908702	G - A	E- K	0.00
	rrs121908701	G - A	R- H	0.04
	rrs121908694	C- T	S- F	0.01
	rrs113947614	T - C	I- T	0.00
	rrs112032633	G - A	G- D	--
	rrs111292981	C - T	S- S	--

independently to the SIFT program. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and those greater than or equal to the cutoff are predicted to be tolerated. The cutoff value in the SIFT program is the tolerance indicator. Higher tolerance index indicates less functional impact of particular amino acid substitution.

**Inference:**

Out of 10, 8 were found highly deleterious by SIFT algorithm.

**Table 2: Showing POLYPHEN Result:**

Gene name	SNP IDs	Nucleotide change	Amino acid change	PSIC	Status
<b>CHEK2</b>	rs121908706	G - A	R- H	2.146	Damaging
	rs121908705	A - G	N- D	2.146	Damaging
	rs121908704	A - G	T- A	1.380	Damaging
	rs121908703	C - T	S - L	1.504	Benign
	rs121908702	G - A	E- K	2.091	Possibly Damaging
	rs121908701	G - A	R- H	1.507	Possibly Damaging
	rs121908694	C- T	S- F	1.795	Damaging
	rs113947614	T - C	I- T	1.538	Benign
	rs112032633	G - A	G- D	1.575	Damaging
	rs111292981	C - T	S- S	1.575	Damaging

**Damaged nsSNP found by the PolyPhen server:**

The structural levels of alteration due to mutation were determined by applying the PolyPhen program. 27 nsSNPs investigated in this work were submitted to the PolyPhen server. A position-specific independent count (PSIC) score difference of 1.5 and above is considered to be damaging.

**Inference:**

Out of 10, 8 were found in damaging condition showing PCID score is greater than 1.5 by polyphen.

**Table 3: Showing Mutations found to deleterious by POLYPHEN, SIFT and MUTANT:**

Gene name	SNP IDs	Position	Stability after mutation
<b>CHEK2</b>	rs121908706	517, 474,445	Decreases
	rs121908705	489, 446,417	Decreases
	rs121908704	444, 401	Decreases
	rs121908703	399, 356	Decreases
	rs121908702	282, 239	Decreases
	rs121908701	224, 181	Decreases
	rs121908694	41	Decreases
	rs113947614	264, 221	Decreases
	rs112032633	30	Decreases
	rs111292981	67	Decreases

**Inference:**

All nsSNP were showing decreased stability after mutation at various positions.

**Conclusion:**

The human CHEK2 gene was investigated in this work. The work was done by evaluating the influence of functional SNPs through computation methods and the most deleterious non-synonymous SNPs (nsSNPs) among the CHEK2 gene coding for CHK2 protein kinase were predicted. Out of total 682 SNPs in the CHEK2 gene, 10 were found to be non synonymous. From among these 10 nsSNPs, 8 were found to be deleterious by SIFT and 5 were found damaging as per the PolyPhen Program and 5 nsSNPs were found to be common in both the SIFT and the PolyPhen program. PolyPhen showed that rest of the SNPs is in probably damaging or benign conditions while I - Mutant program showed the structural instability after mutations at these sites. It was found that mutations in CHEK2 gene at the positions 517,404,445 (in rs121908706), 282,239 (in rs121908705), 224,181(in rs121908701), 30(in rs112032633) and at 67(in rs111292981) show a potential deleterious effect on the function of this protein which is probably the reason for causing breast cancer, Li-Fraumeni syndrome, and other cancers like prostate, lung, colon, kidney, thyroid and ovarian cancer. The results of this study have therefore provided novel evidence of the correlation using human data, in turn facilitating genotyping efforts in future molecular epidemiological studies and providing targets for phenotypic analysis of genetic variants.

**Acknowledgement:**

We take this golden opportunity in expressing our gratitude and sincerity to Dr. C. SudandiraDoss, Assistant professor, VIT University. We would also like to give our heart full thanks to our program manager Dr. Pragasaam Vishwanathan whose valuable supports are really helpful in the completion of

this project on time. We thank the Management of VIT for providing good facility for carrying out this work.

**References:**

- [1]H Nevanlinna and J Bartek, The CHEK2 gene and inherited breast cancer susceptibility, *Oncogene* (2006) 25, 5912–5919.
- [2]Cargill M, Altshuler D, Ireland J et al (1999) Characterization of single-nucleotide polymorphisms incoding regions of human genes. *Nat Genet* 22:231–238. doi:10.1038/10290
- [3]Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7:85–97. doi:10.1038/nrg1767
- [4] [5] Ng CP, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11:863–874. doi:10.1101/gr.176601
- [5]Ng CP, Henikoff S (2002) Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 12:436–446. doi:10.1101/gr.212802.

\*\*\*\*\*