Original Article

DOI: 10.17354/ijss/2015/473

Impact of Mutations in Medical Science: A Focus on ErbB2 Gene

K P S Adinarayana[1], R Karuna Devi[2], P Ajay Babu[3]

[1]Associate Professor, Department of Anatomy, Andhra Medical College, Visakhapatnam, Andhra Pradesh, India, [2]Associate Professor, Department of Physiology, NRI Medical College, Visakhapatnam, Andhra Pradesh, India, [3]Visiting Professor, Department of Biotechnology, St. Peters University, Chennai, Tamil Nadu, India

Abstract

Introduction: Epidermal growth factor receptor family of receptor tyrosine kinases plays important roles in the development and severity of many cancers across human populations. Single-nucleotide polymorphisms (SNPs) play a major role in understanding the genetic basis of many complex human diseases. It is still a major challenge to identify the functional SNPs in a disease-related gene.

Purpose: To explore possible relationships between genetic mutation and phenotypic variations of ErbB2 gene.

Materials and Methods: In present study, different bioinformatics algorithms such as sorting intolerant from tolerant (SIFT), polymorphism phenotyping (PolyPhen), and I-mutant server to predict the impact of these amino acid substitutions on ErbB2 were employed.

Results: SIFT analysis resulted in 9 of 109 non-synonymous SNPs (nsSNPs) were predicted to be "damaging" and "possibly damaging," "probably damaging" and "benign" by PolyPhen program. I-mutant 3.0 results demonstrated that all respective mutations would decrease the overall stability of the protein. The orthologous multiple alignments of nsSNPs with ids namely rs28933368 (E914K), rs193171026 (L46F), rs149937802 (R34W), rs140980495 (R536Q), and rs144533600 (E1244K) showed that all mutations were found to be either conserved or the flanking amino acids showed a low degree of conservation except rs149937802 (R34W) where a high rate of mutation was observed among orthologs, which needs further investigation.

Conclusion: Current analysis represents the application of computational tools in understanding functional variation from the perspective of structure, expression, evolution, and phenotype.

Key words: Amino acid, Breast cancer, ErbB2, Receptor tyrosine kinase, Single-nucleotide polymorphism

INTRODUCTION

Breast cancer is the most common cause of cancer in women. The gene is involved in low-level susceptibility to breast cancer is ERBB2 (Heregulin 2 [HER2]). This gene is present on chromosome 17q12-q21, spans 38 kilobases, and comprises 27 coding exons.[1,2] ErbB2 is a member of the epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases, which in humans includes EGFR (ERBB1), ERBB2, ERBB3, and ERBB4. ErbB receptors are vital in facilitating proliferation and differentiation of cells in the developing embryo as well as in adult tissues. Further, it was conceived that an inappropriate activation might result in the development and callousness of many cancers.[3] Over expression of HER2 is found in 20-30% of human breast cancers and correlates with more aggressive tumors and a poorer prognosis. It was identified that ErbB2/ErbB3 heterodimer represents an important oncogenic unit[4] in breast cancer cell proliferation.

Over expression of the ErbB1 and ErbB2 proteins contributes to the aggressive behavior of malignant tumors originating from the endometrium. The expression levels are considerably higher in malignant ones when compared to benign tumors.[5] Anti-cancer therapies involving a monoclonal antibody targeting HER2, Herceptin (also

Access this article online

www.ijss-sn.com

Month of Submission : 08-2015
Month of Peer Review : 09-2015
Month of Acceptance : 10-2015
Month of Publishing : 10-2015

Corresponding Author: Dr. K P S Adinarayana, Department of Anatomy, Andhra Medical College, Visakhapatnam, Andhra Pradesh, India. Phone: +91-7032305740. E-mail: kpsanarayana@rediffmail.com

known as trastuzumab), is currently prescribed for breast cancer. Herceptin binds to the juxtamembrane region of HER2, identifying this site as a target for anticancer therapies.[6]

Single-nucleotide polymorphisms (SNPs) are modifications of a single nucleotide (adenine, thymine, cytosine, or guanine) in the genome. Around 90% of all human genetic variations constitute SNPs and the probability reaching every 100-300 bases in the human genome.[7] SNPs were found in both coding and noncoding regions of the genome. Non-synonymous SNPs (nsSNPs) is responsible for nearly half of the known genetic variations related to human disease.[8,9] Functional SNP analysis reported for BRCA1, ABL1, ERBB2, CFTR, and EGFR genes.[10-14]

Although several articles reported the association of SNPs in the ErbB2 gene, computational analysis describing the functional consequences of SNPs presented here. We applied different publicly available computational algorithms, such as sorting intolerant from tolerant (SIFT),[15] polymorphism phenotyping (PolyPhen),[16] and I-mutant 3.0 for protein stability analysis[17] and to identify likely deleterious SNPs which could affect protein function. Almost 80% success achieved with SIFT and PolyPhen in benchmarking studies employing amino acid substitutions[18-20] and the "false negative" and "false positive" error rates of SIFT and PolyPhen[21] is 31%, 20%, and 31%, 9%, respectively. The rationale behind the work is to study the importance of mutations in breast cancer target, ErbB2, in particular.

MATERIALS AND METHODS

Data Analysis

In this study, it was observed that many variations exist for ErbB2 gene and demarcation of choosing the correct SNPs was a precarious one.[22] One method was to arrange SNPs as per their structural and functional significance. Instead, gene cards (www.genecards.org) was accessed to identify SNPs, and we compared whether it represents a novel or an existing mutational event using an SNPs-database server.[23,24] Therefore, to check the overall effect of such mutations on structure and functional aspects of protein, SIFT and PolyPhen-2 software were employed.

SIFT

The SIFT[25] program was used to perform protein conservation analysis and predict the phenotypic effect of amino acid substitutions. SIFT was constructed on the principle that protein evolution is correlated with protein function. Variants that occur at conserved alignment positions are tolerated less than those that occur at diverse positions.[23] The algorithm constructs a multiple sequence alignment of proteins along with the query sequence of same group. The output comprises alignments of homologous sequences and scores that range from 0.0 to 1.0 to each residue are assigned. The SIFT scores[19] were classified as intolerant (0.00-0.05), potentially intolerant (0.051-0.10), borderline (0.101-0.20), or tolerant (0.201-1.00). The lower the tolerance index (TI) of a particular amino acid substitution, the larger is its likely impact. An nsSNP with a TI score of ≤0.05 is considered to be deleterious, and a score of >0.05 is considered as tolerant.

PolyPhen-2

PolyPhen-2[26] is a computational tool that identifies functionally potential nsSNPs in the coding region. The prediction is based on combined features involving phylogenetic, structural, and sequence annotations. For a positional variation of an amino acid, PolyPhen-2 performs the following: (a) The program extracts sequence-based features of the variation from the UniProt database, (b) calculates profile scores for two amino acid variants, (c) calculates the structural parameters, and substituted residue contacts. Based on PolyPhen-2 analysis, the scores represent "benign" (0.00-1.50), "possibly damaging" (1.50-1.99), or "probably damaging" (>2.0). The query was submitted as a single mutational event with a chromosome co-ordinate. PolyPhen-2 analyzes several protein structure databases and performs multiple alignments of homologous sequences, and reports the amino acid contact information. Further, the difference between two variants is calculated. High differences in scores signify higher functional impact of a particular amino acid substitution.[27]

Protein Stability Prediction Analysis by I-mutant

I-mutant version 3.0 was used to predict the changes in protein stability on single-site mutations. The program evaluates the stability change starting from the protein structure or sequence. This program was trained on a dataset derived from ProTherm,[17] the most comprehensive database of protein mutations derived from experimental data. I-mutant is a suite of support vector machine 2 (SVM2) based predictors, integrated into a unique web server[23] at http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi

RESULTS AND DISCUSSION

Gene ErbB2 with a potential role in breast cancer selected for the study. Out of a total of 1038 SNPs, 109 nsSNPs selected from gene cards database (www.genecards.org). Analysis concerning the amino acid conservation in a protein was performed using the SIFT algorithm that

predicts amino acid substitution and possible impact on protein function. The program does by aligning similar proteins and calculates a score that determines the evolutionary conservation status of an amino acid. All 108 nsSNPs submitted to SIFT and the PolyPhen servers, respectively. From the result, nine nsSNPs were predicted to be damaging by SIFT and "possibly damaging," "probably damaging," and "benign" by Polyphen program. The Polyphen program report the position-specific independent count (PSIC) score above of 1.0 is considered to be damaging, the results shown in Table 1. The validity of these algorithms based on the benchmarking studies carried out on "known" deleterious substitutions annotated in databases, such as Swiss-Prot, resulted in successful prediction of over 80% of amino acid substitutions. Experimental studies pertaining to individual proteins have confirmed the accuracy of SIFT and PolyPhen.[19,28,29]

From SIFT, the output data represents that higher the TI, the less will be the functional impact of a particular amino acid substitution and vice versa. From Table 1, it is clear that except SNP rs1801201, all other nsSNPs were classified as "damaging" and showed a deleterious TI score of 0.01-0.04 which possibly could affect the protein function of ErbB2 gene.

The nine nsSNPs that resulted from SIFT were submitted to the PolyPhen server and the amino acid variations at the structural level was determined. Table 1 presents the distribution of the variants by PolyPhen score. PolyPhen scores in this dataset ranged from 1.0 to 0.03. An SNP in a nucleotide sequence changes the respective amino acid and they possibly impact the folding patterns, interaction sites, solubility, or stability of proteins. Therefore, to assess the relationship between genetic and phenotypic variation, it is indeed necessary to verify the structural features of the respective non-synonymous mutations in proteins. The results obtained by the SIFT was found to be correlated well with the results obtained by PolyPhen, as seen in Table 1. Hence, we mapped known disease mutations onto known three-dimensional structure of ErbB2 protein based on PolyPhen score. The nsSNPs with IDs namely rs28933368 (E914K), rs193171026 (L46F), rs149937802 (R34W), rs140980495 (R536Q), and rs144533600 (E1244K) showed a PSIC score >0.9 were selected to perform multiple alignments of mutated amino acids on orthologous ErbB2 family of protein sequences. From analysis, evidence suggests that all mutations are either conserved or the flanking amino acids showed a low degree of conservation (Figure 1).

The mutation region corresponding to E914K as well as flanking regions conserved in all orthologous sequences. However, L46F region showed variation with amino acids V (*Mesocricetus auratus*) and M (*Notophthalmus viridescens*), respectively, moreover, the flanking amino acids are not much conserved (Figure 1). The mutation R34W region, also observed in other organisms represents variation with amino acids Q, L, A, T, S, respectively. The R536Q and E1244K regions highly conserved in all species except few residue variations in flanking regions of 536 position.

Changes in Stability Due to Mutation

I-mutant 3.0 results obtained in the analysis demonstrated the change in protein stability with relative free energy due to mutation (Table 2). We submitted independently the protein sequence of nine nsSNPs which predicted to be damaging both using SIFT and PolyPhen programs. The second SVM2 based predictor for protein stability changes on single point amino acid mutation demonstrated that all respective mutations would decrease the overall stability of the protein.

Table 1: Prediction result of SIFT and PolyPhen programs

SNP ID	Amino acid substitution	SIFT prediction	TI score	PolyPhen prediction	PSIC score
rs28933368	E914K	Damaging	0.01	Probably damaging	1.000
rs1801201	I654V	Tolerated	0.17	Benign	0.303
rs1136201	I655V	Damaging	0.02	Benign	0.406
rs193171026	L46F	Damaging	0.01	Probably damaging	0.974
rs149937802	R34W	Damaging	0.03	Probably damaging	0.992
rs140980495	R536Q	Damaging	0.01	Probably damaging	1.000
rs55943169	A1216D	Damaging	0.01	Benign	0.028
rs144533600	E1244K	Damaging	0.04	Probably damaging	0.999
rs111611886	D1105N	Damaging	0.01	Possibly damaging	0.791

SIFT result: Score ranges from 0 to 1. The amino acid substitution is predicted damaging if the score is ≤0.05 and tolerated if the score is >0.05. PolyPhen-2 result: Probably damaging (more confident prediction)/possibly damaging (less confident prediction). SIFT: Sorting intolerant from tolerant, PolyPhen: Polymorphism phenotyping, TI: Tolerance index, PSIC: Position-specific independent count

Table 2: Prediction result of I-mutant software

ErbB2	SNP ID	Amino acid position	WT	MT	SVM2 stability	DDG value prediction Kcal/mol	RI
	rs28933368	914	E	K	Decrease	−0.70	9
	rs1801201	654	I	V	Decrease	−0.99	7
	rs1136201	655	I	V	Decrease	−1.01	7
	rs193171026	46	L	F	Decrease	−1.00	4
	rs149937802	34	R	W	Decrease	−0.11	3
	rs140980495	536	R	Q	Decrease	−0.72	8
	rs55943169	1216	A	D	Decrease	−0.58	4
	rs144533600	1244	E	K	Decrease	−0.78	8
	rs111611886	1105	D	N	Decrease	−0.82	3

For all the predictions, pH and temperature were selected as 7.0 and 25°C, respectively. WT: Wild type amino acid, MT mutant type amino acid, DDG: DG (new protein)-DG (wild type) in Kcal/mol, DDG<0: Decrease stability, DDG>0: Increase stability, RI: Reliability index, SVM: Support vector machine, SNP: Single-nucleotide polymorphisms, SNP: Single-nucleotide polymorphisms

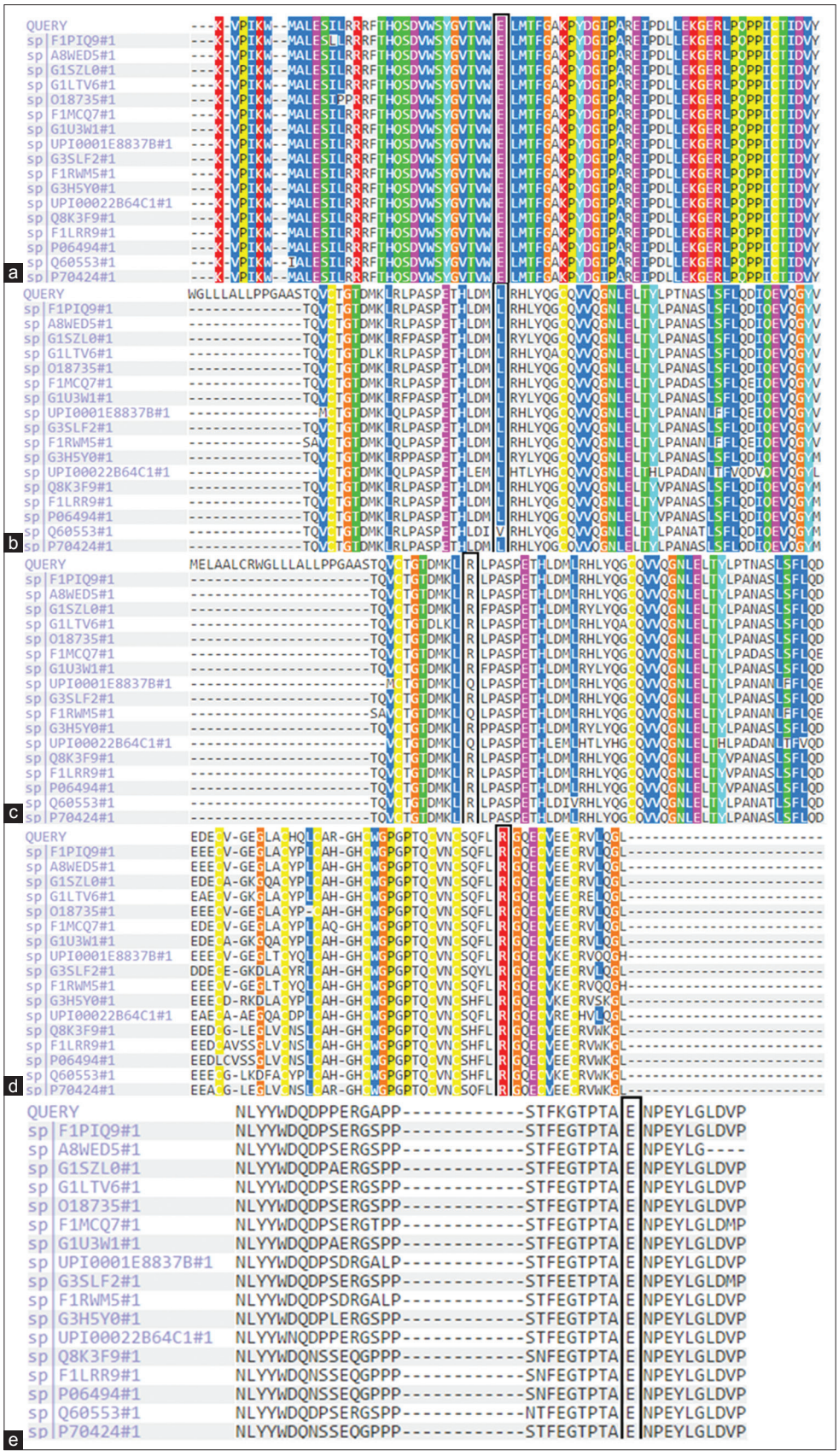

Figure 1: Multiple sequence alignments of non-synonymous single-nucleotide polymorphisms with position-specific independent count score >0.9, (a) E914K, (b) L46F, (c) R34W, (d) R536Q, (e) E1244K

Amino acid substitutions currently account for approximately half of the known gene lesions responsible for human inherited disease.[8] Therefore, the identification of nsSNPs that would probably affect protein function related to a disease is an imperative task in molecular biology. Assessment of nsSNPs based on phylogenetic information (residue conservation) as well as structural approaches, hence, much attention been focused on modeling by different methods. The possible phenotypic variations of SNPs modify amino acids at sequence level thereby affecting the structural parameters, where focus shifted on functional SNPs affecting regulatory regions. Moreover, because of widespread distribution of SNPs on the genome, they have become particularly important and valuable as genetic makers in the research for studying functional loss of proteins and their related pattern on disease susceptibility. Currently, several thousands of human SNPs found by high-throughput methods.

Most molecular studies focused on SNPs located in coding and regulatory regions, yet many of these studies are unable to detect significant associations between SNPs and disease susceptibility. We applied an evolutionary perspective followed by structural approach and mutation stability analysis to SNPs. Moreover, functionally significant amino acids conserved across species; hence SNPs that change the structure and functional features are more likely to be associated with cancer susceptibility. Overall, this analysis will provide useful information in selecting SNPs that are likely to have the potential functional impact on ErbB2.

CONCLUSION

Current analysis focused on SNPs in the coding regions of ErbB2 enzyme, and the outcome of the study could explain the cancer risk due to the significant fraction of mutational changes to the protein. SIFT analysis resulted in 9 nsSNPs being predicted to be "damaging" and "possibly damaging," "probably damaging" and "benign" by Polyphen program. These nsSNPs demonstrated a decrease in the overall stability of the protein by I-mutant 3.0 server. Multiple alignments of orthologous nsSNPs, rs28933368 (E914K), rs193171026 (L46F), rs149937802 (R34W), rs140980495 (R536Q), and rs144533600 (E1244K) showed that all mutations were either conserved or the flanking amino acids showed a low degree of conservation except rs149937802 (R34W). On the other hand, a high rate of mutation observed among orthologs. The analysis suggested the application of these software tools and utilizing publicly available databases such as NCBI, dbSNP for analysis resulted in efficient genetic association studies at the molecular level. Our analysis represents the application of computational tools in understanding functional variation from the perspective of structure, expression, evolution, and phenotype. Further, work is in advancement to evaluate the structural variations of amino acid positions based on PSIC score would highlight the functional and geometrical disturbances in protein structure.

REFERENCES

1. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR, *et al.* Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. J Natl Cancer Inst 2003;95:448-57.
2. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2001;2:127-37.
3. Ménard S, Pupa SM, Campiglio M, Tagliabue E. Biologic and therapeutic role of HER2 in cancer. Oncogene 2003;22:6570-8.
4. Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW Jr, *et al.* Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature 2003;421:756-60.
5. Brys M, Semczuk A, Rechberger T, Krajewska WM. Expression of erbB-1 and erbB-2 genes in normal and pathological human endometrium. Oncol Rep 2007;18:261-5.
6. Allen SD, Garrett JT, Rawale SV, Jones AL, Phillips G, Forni G, *et al.* Peptide vaccines of the HER-2/neu dimerization loop are effective in inhibiting mammary tumor growth *in vivo*. J Immunol 2007;179:472-82.
7. Lee JE, Choi JH, Lee JH, Lee MG. Gene SNPs and mutations in clinical genetic testing: Haplotype-based testing and analysis. Mutat Res 2005;573:195-204.
8. Krawczak M, Ball EV, Fenton I, Stenson PD, Abeysinghe S, Thomas N, *et al.* Human gene mutation database-a biomedical information and research resource. Hum Mutat 2000;15:45-51.
9. Prokunina L, Alarcón-Riquelme ME. Regulatory SNPs in complex diseases: Their identification and functional validation. Expert Rev Mol Med 2004;6:1-15.
10. Rajasekaran R, Sudandiradoss C, Doss CG, Sethumadhavan R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. Genomics 2007;90:447-52.
11. George Priya Doss C, Sudandiradoss C, Rajasekaran R, Purohit R, Ramanathan K, Sethumadhavan R. Identification and structural comparison of deleterious mutations in nsSNPs of ABL1 gene in chronic myeloid leukemia: A bio-informatics study. J Biomed Inform 2008;41:607-12.
12. Rajasekaran R, George Priya Doss C, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R. Effect of deleterious nsSNP on the HER2 receptor based on stability and binding affinity with herceptin: A computational approach. C R Biol 2008;331:409-17.
13. George Priya Doss C, Rajasekaran R, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R. A novel computational and structural analysis of nsSNPs in CFTR gene. Genomic Med 2008;2:23-32.
14. Rajasekaran R, Sethumadhavan R. In silico identification of significant detrimental missense mutations of EGFR and their effect with 4-anilinoquinazoline-based drugs. Appl Biochem Biotechnol 2010;160:1723-33.
15. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812-4.
16. Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet 2000;16:198-200.
17. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic disease associated to single point protein mutation with support vector machines and evolutionary information. Bioinformatics 2006;22:2729-34.
18. Sunyaev S, Lathe W 3rd, Bork P. Integration of genome data and protein structures: Prediction of protein folds, protein interactions and "molecular phenotypes" of single nucleotide polymorphisms. Curr Opin Struct Biol 2001;11:125-30.
19. 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.

20. 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.
21. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 2006;7:61-80.
22. Khalfalla AA, Hassan HE, Ahmed MO, Dowd AA, Homeida S, Hussain MA, *et al.* Novel deleterious non synonymous SNPs within HLA-H (HFE) gene can be used as diagnostic marker to predict hereditary hemochromatosis: Using bioinformatics analysis. Int J Comput Bioinform In Silico Model 2015;4:643-50.
23. Masoodi TA, Shammari SA, Al-Muammar MN, Almubrad TM, Alhamdan AA. Screening and structural evaluation of deleterious Non-Synonymous SNPs of ePHA2 gene involved in susceptibility to cataract formation. Bioinformation 2012;8:562-7.
24. Salih SS, Abdelhag IM, Abdalla WM, Mosad AS, Hassan MM, Hassan MA. Computational detection of deleterious single nucleotide polymorphisms in human adenomatous polyposis coli gene the gate-keeper of colorectal carcinoma. Int J Comput Bioinform In Silico Model 2014;3:531-7.
25. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001;11:863-74.
26. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: Server and survey. Nucleic Acids Res 2002;30:3894-900.
27. Doss CG, Sethumadhavan R. Investigation on the role of nsSNPs in HNPCC genes – A bioinformatics approach. J Biomed Sci 2009;16:42.
28. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002;12:436-46.
29. Brooks-Wilson AR, Kaurah P, Suriano G, Leach S, Senz J, Grehan N, *et al.* Germline E-cadherin mutations in hereditary diffuse gastric cancer: Assessment of 42 new families and review of genetic screening criteria. J Med Genet 2004;41:508-17.

How to cite this article: Adinarayana KPS, Devi RK, Babu PA. Impact of Mutations in Medical Science: A Focus on ErbB2 Gene. Int J Sci Stud 2015;3(7):180-185.

Source of Support: Nil, **Conflict of Interest:** None declared.