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

K P S Adinarayana¹, R Karuna Devi², P Ajay Babu³

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

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 (Herregulin 2 [HER2]). This gene is present on chromosome 17q12-q21, spans 38 kilobases, and comprises 27 coding exons. 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. 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 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. 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: kpsananarayana@rediffmail.com

DOI: 10.17354/ijss/2015/473
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 analysis17 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 substitutions18-20 and the “false negative” and “false positive” error rates of SIFT and PolyPhen21 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 SIFT25 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 scores19 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-226 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 server25 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. 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, I, 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

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

**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

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

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, DDGico: Decrease stability, DDG++: 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. 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 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.

REFERENCES

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.
Adinarayana, et al.: Impact of Gene Mutations in Medical Science


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.