RESEARCH ARTICLE

Prediction and Analysis of Breast Cancer Related Deleterious Non-Synonymous Single Nucleotide Polymorphisms in the PTEN Gene

C Kumaraswamy Naidu, Y Suneetha*

Abstract

One of the most common cancer types faced by the women around the world is breast cancer. Among the several low, moderate and high penetrance genes conferring susceptibility to breast cancer, PTEN is one which is known to be mutated in many tumor types. In this study, we predicted and analyzed the impact of three deleterious coding non-synonymous single nucleotide polymorphisms rs121909218 (G129E), rs121909229 (R130Q) and rs57374291 (D107N) in the PTEN gene on the phenotype of breast tumors using computational tools SIFT, Polyphen-2, PROVEAN, MUPro, POPMusic and the GETAREA server.

Keywords: Breast cancer - PTEN - nSNPs - genetic factors - computational tools

Asian Pac J Cancer Prev, 17 (4), 2199-2203

Introduction

Breast cancer is one of the common cancer types faced by the women in the modern life. In countries like United States, one in 8 women develops breast cancer in her lifetime and its incidence rates is slightly increasing among African American women (De Santis et al., 2014). Along with colorectal cancer, it ranks high in all countries. Its incidence in the United States and Europe is twice as high as it is in Asian countries, and its incidence rates have been increasing in all countries (Saika and Sobue, 2013). Tumor heterogeneity, lifestyle factors including obesity, breastfeeding, and alcohol consumption are some of the traditional risk factors associated with breast cancer (Kwan et al., 2009). Some of the genetic and hormonal factors constitute risk to breast cancer (Martin and Weber, 2000). Phosphatase and tensin homolog deleted on chromosome ten (PTEN) is one of the frequent mutated gene found in many primary and metastatic malignancies including breast cancer (Kechagioglou et al., 2014). Human estrogen receptor-positive (ER+) breast cancer cell lines containing inducible PTEN short hairpin RNAs, result in the hyperactivation of the PI3K pathway and a concomitant change in gene expression similar to luminal B breast cancer types (Maggi and Weber, 2015). In breast cancer cells, reduced expression of PTEN is known to confer susceptibility to inhibitors of the PI3 kinase/Akt pathway (De Graffenried et al., 2004). Triple-negative breast cancers are aggressive forms, CIB1 plays a broad role in its cell survival, tumor growth and a low expression of PTEN is a key predictor of sensitivity to CIB1 depletion (Black et al., 2015). So, treatment with trastuzumab to improve the disease-free and overall survival has been a standard approach for HER2-overexpressing breast cancer patients and PTEN status was suggested to be one of the indicators (Adamczyk et al., 2015).

Variations on the promoter region of PTEN are known to affect the progression of breast cancer and the survival of the patients (Heikkinen et al., 2011). Previously several single nucleotide polymorphisms such as rs1234212, rs11202586, rs1234221, rs1903860, rs1234220, rs1234219, rs1903858, rs2299939, rs1234224, rs1234223, rs1234213, rs2673832 with a PTEN haplotype associated with breast cancer risk were predicted (Haiman et al., 2006). Previous study showed that a high level of discordance in PTEN level, PIK3CA mutations and receptor status between primary tumors and metastases influenced the patient selection and response to PI3K-targeted therapies (Gonzalez-Angulo et al., 2011). An in vivo study showed that non catalytic PTEN missense mutation predisposes the organ-selective cancer development (Caserta et al., 2015). In the present study, we aim to predict the breast cancer-associated nSNPs in PTEN and to further to analyze

Materials and Methods

SNP datasets used for the study

SNP datasets for PTEN were retrieved from the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/, Build 138; access date: August 22, 2015) (Sherry et al., 2001) for our study.
Prediction of Deleterious SNPs

SIFT and Polyphen-2 database servers to screen out the deleterious coding nSNPs from other SNPs of PTEN. ‘Sorting Tolerant From Intolerant’ (SIFT) (http://sift.jcvi.org/) uses a sequence homology based approach for predicting the amino acid substitution in a protein affecting the protein function (Kumar et al., 2009). It assigns 0-0.05 score for intolerant or deleterious amino acid substitutions and 0.05–1 scores for tolerant or neutral amino acid substitutions (Ng and Henikoff, 2003; Ng and Henikoff, 2006). PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) on the other hand predicts the functional significance of variation using Naïve Bayes classifier. We used WHESS.db a quick access for precomputed set in PolyPhen-2 predictions was used for our analysis (Adzhubei et al., 2010). We submitted our query in the

Table 1. Functionally significant SNPs Predicted Using the SIFT Server

| dbSNPID     | Amino acid change | Nucleotide change | Protein ID  | Tolerance Index Using orthologues in the Protein alignment | Using homologues in the Protein alignment |
|-------------|-------------------|-------------------|-------------|-----------------------------------------------------------|-----------------------------------------|
| rs121909218 | G129E             | A/G               | NP_000305   | 0                                                         | 0                                       |
| rs121909221 | S170R             | A/T               | NP_000305   | 0                                                         | 0                                       |
| rs121909222 | H123R             | A/G               | NP_000305   | 0                                                         | 0                                       |
| rs121909223 | C124R             | C/G/T             | NP_000305   | 0                                                         | 0                                       |
| rs121909225 | M35R              | G/T               | NP_000305   | 0                                                         | 0                                       |
| rs121909226 | L70P              | C/T               | NP_000305   | 0                                                         | 0                                       |
| rs121909229 | R130Q             | A/C/G             | NP_000305   | 0                                                         | 0                                       |
| rs121909230 | L112P             | C/T               | NP_000305   | 0                                                         | 0                                       |
| rs121909233 | D19N              | A/G               | NP_000305   | 0.18                                                      | 0.04                                   |
| rs121909235 | R234Q             | A/G               | NP_000305   | 0.31                                                      | 0.05                                   |
| rs121909236 | H61D              | C/G               | NP_000305   | 0                                                         | 0                                       |
| rs121909237 | A121G             | C/G               | NP_000305   | 0.08                                                      | 0                                       |
| rs121909238 | H93R              | A/G               | NP_000305   | 0                                                         | 0                                       |
| rs121909239 | D252G             | AG/G              | NP_000305   | 0.01                                                      | 0                                       |
| rs121909241 | G132V             | A/G/T             | NP_000305   | 0                                                         | 0                                       |
| rs57374291  | D107N             | A/G               | NP_000305   | 0.05                                                      | 0                                       |
| rs121913293 | R173C             | C/T               | NP_000305   | 0                                                         | 0                                       |
| rs121913294 | R173H             | A/C/G             | NP_000305   | 0                                                         | 0                                       |

Table 2. Functionally Significant SNPs Predicted Using the Polyphen-2 Server

| Nucleotide change | Amino acid change | Protein ID | HDivPred | HDivProb | HVarPred | HVarProb |
|-------------------|-------------------|------------|----------|----------|----------|----------|
| G/A               | G129E             | P60484     | Prob. dam | 1        | Prob. dam | 1        |
| T/A               | S170R             | P60484     | Prob. dam | 1        | Prob. dam | 0.999    |
| A/G               | H123R             | P60484     | Prob. dam | 1        | Prob. dam | 0.998    |
| T/C               | C124R             | P60484     | Prob. dam | 1        | Prob. dam | 0.999    |
| T/G               | M35R              | P60484     | Prob. dam | 0.996    | Prob. dam | 0.974    |
| T/C               | L70P              | P60484     | Prob. dam | 1        | Prob. dam | 1        |
| G/A               | R130Q             | P60484     | Prob. dam | 1        | Prob. dam | 0.998    |
| T/C               | L112P             | P60484     | Prob. dam | 1        | Prob. dam | 1        |
| G/A               | D19N              | P60484     | Prob. dam | 0.988    | Prob. dam | 0.815    |
| G/A               | R234Q             | P60484     | Prob. dam | 0.98     | possibly dam | 0.617 |
| C/G               | H61D              | P60484     | Prob. dam | 1        | Prob. dam | 0.998    |
| C/G               | A121G             | P60484     | prob. dam | 0.999    | prob. dam | 0.968    |
| A/G               | H93R              | P60484     | prob. dam | 1        | prob. dam | 0.998    |
| A/G               | D252G             | P60484     | prob. dam | 0.989    | prob. dam | 0.862    |
| G/T               | G132V             | P60484     | Prob. dam | 1        | Prob. dam | 1        |
| G/A               | D107N             | P60484     | Prob. dam | 1        | prob. dam | 0.999    |
| C/T               | R173C             | P60484     | Prob. dam | 1        | prob. dam | 0.966    |
| G/A               | R173H             | P60484     | prob. dam | 1        | prob. dam | 0.966    |
Prediction and Analysis of Breast Cancer Related Non-Synonymous SNPs in the PTEN Gene

form of dbSNP id for both SIFT and Polyphen-2.

Table 3. Phenotype of the Predicted Deleterious SNPs

| dbSNPID     | Amino acid change | Phenotype                          | References                          |
|-------------|-------------------|------------------------------------|-------------------------------------|
| rs121909218 | G129E             | Cowden disease, Breast cancer      | (Liaw et al., 1997)                  |
| rs121909221 | S170R             | Bannayan-Riley-Falz parallel       | (Marsh et al., 1997)                 |
| rs121909222 | H123R             | Cowden disease                     | (Nelen et al., 1997)                 |
| rs121909223 | C124R             | Cowden disease                     | (Olschwang et al., 1998)             |
| rs121909225 | M35R              | Cowden disease                     | (Marsh et al., 1998)                 |
| rs121909226 | L70P              | Cowden disease                     | (Kurose et al., 1999; Baig et al., 2011) |
| rs121909229 | R130Q             | Cowden disease, Breast cancer      | (Nelen et al., 1997)                 |
| rs121909230 | L112P             | Lhermitte-duclus disease           | (Staak et al., 2002)                |
| rs121909233 | D19N              | Malignant melanoma                 | (Poetsch et al., 2002)              |
| rs121909235 | R234Q             | Glioma                             | (Butler et al., 2005; Butler et al., 2005; Nagy et al., 2014) |
| rs121909236 | H61D              | Macraclephaly                      | (Butler et al., 2005)               |
| rs121909237 | A121G             | Squamous cell carcinoma            | (Butler et al., 2005)               |
| rs121909238 | H93R              | Autism                             | (Butler et al., 2005; Nagy et al., 2014) |
| rs121909239 | D252G             | Autism                             | (Staak et al., 2002)                |
| rs121909241 | G132V             | Hamartoma tumor syndrome           | (Sutphen et al., 1999)              |
| rs121913293 | R173C             | Hereditary cancer-predisposing      | Clinvar                             |
| rs121913294 | R173H             | Hereditary cancer-predisposing      | Clinvar                             |

Table 4. Impact of the SNPs Predicted Using the PROVEAN Server

| dbSNPID     | Amino acid change | PROVEAN score | Prediction (cutoff= -2.5) |
|-------------|-------------------|---------------|-------------------------|
| rs121909218 | G129E             | -7.772        | Deleterious             |
| rs121909229 | R130Q             | -3.858        | Deleterious             |
| rs57374291  | D107N             | -4.675        | Deleterious             |

Figure 1. Distribution of PTEN SNPs. Coding nonsynonymous SNPs (nSNPs), coding synonymous SNPs (sSNPs), 3’ UTR SNPs, 5’ UTR SNPs and intronic SNPs

SNPs was performed using the databases dbSNPedia (Caras and Lennon, 2012), (Schaefer et al., 2012), Hapmap (International HapMap et al., 2010), Pubmed (Sood and Ghosh, 2006). The SNPs with breast cancer phenotype were further cross checked for deleterious nature using the PROVEAN software (Choi et al., 2012).

Modeling nSNPs locations in protein structure

Crystal structure of PTEN downloaded from the protein databank (PDB ID: 1d5r, chain A) (Lee et al., 1999) was used for modelling the nSNPs in the protein structure. All water molecules and the TLA ligand were removed from the crystal structure and the mutants (MTs) G129E, R130Q and D107N were created by replacing the wild-type (WT) protein residue with its polymorphic residue using PyMOL (PyMol, 2006). Mutants were optimized and energy minimized using Nomad-Ref server (Lindahl et al., 2006) with conjugate gradient method.

Analysis of the impact of mutant on the PTEN protein product

The effect of amino acid changes on the stability of PTEN protein was analyzed by using the MUpro (http://mupro.proteomics.ics.uci.edu/) (Cheng et al., 2006) web servers. Thermodynamic stability of the mutants was analyzed using POPMusic server (Dehouck et al., 2009). The solvent accessibility information of the WT and MTs was analyzed using GETAREA server (http://curie.utmb.edu/getarea.html) (Fraczkiewicz and Braun, 1998) which considers the residues which exceeds the ratio value 50% to be solvent and buried if it is less than 20% marked as “o” and “i” respectively.

Results and Discussion

nSNPs from dbSNP database

A search for total SNPs in PTEN against dbSNP database resulted in a total of 18242 SNPs, out of which 5597 were found to be Human (active) SNPs (i.e., Active Human RS and not including those that have been merged). Among the 5597 Human (active) SNPs, 247 were coding non-synonomous SNPs (nSNPs), 83 were coding synonymous, 358 SNPs occurred in the mRNA 3'
Deleterious nSNPs in PTEN gene

Among the 247 coding nSNPs from dbSNP, 18 were found to be deleterious with a tolerance index score of less than or equal to 0.05 using SIFT server. Among these 18 deleterious nSNPs, 13 had a highly deleterious tolerance index score of 0.00 using orthologues and homologues in the protein alignment and the remaining 5 deleterious nSNPs had a tolerance index score of more than 0.00 using orthologues and homologues in the protein alignment respectively (Table 1). Among 18 nSNPs predicted to be deleterious using SIFT server, seven nSNPs showed a nucleotide change of A/G, three showed a change of C/T, two showed a change of C/G, one showed a change of A/T, one showed a change of G/T, two showed a change of A/C/G, one showed a change of C/G/T and one showed a change of A/G/T respectively. A/G and C/T changes occurred maximum number of times compared to the other nucleotide changes.

18 nSNPs that are predicted to be deleterious using SIFT server were submitted to Polyphen-2 to predict their respective functional significance of allele replacement. Results showed that among the 18 nSNPs, three SNPs rs121909218 (G129E), rs121909229 (R130Q) and rs57374291 (D107N) showed a phenotype in breast tumors (Table 3). Results from PROVEAN server also showed that these SNPs as deleterious (Table 4) these SNPs were considered for further analysis.

Deleterious nSNPs impact on PTEN protein

To analyze the impact of deleterious nSNPs on the PTEN protein product, we have analyzed the stability of each mutant. Stability analysis of the mutants using the MUpro server showed that the mutants R130Q and D107N showed a decrease in the stability whereas the mutant G129E showed an increase in the stability (Table 5). Their respective change in the solvent accessibility and the free energy were provided in the Table 6 given below. Results from the Total area/energy of each mutant showed the mutants G129E and D107N major change compared to the wild type PTEN protein (Table 7).

In conclusion, the results from our study indicate that three mutations R130Q, D107N and G129E in PTEN are associated with the breast cancer phenotype. Results showed that these three mutants showed a change in stability. Overall, the present computational approach reported in this study allowed elucidation of the role of deleterious mutations in PTEN thereby providing useful information for the design of PTEN mutant-based therapeutic strategies against breast cancer.

Acknowledgements

The present study was supported under UGC Research Award 2014-16 funded by UGC, New Delhi (No. F. 301/2014(SAH))

References

Adamczyk A, Niemiec J, Janecka A, et al (2015). Prognostic value of PIK3CA mutation status, PTEN and androgen receptor expression for metastasis-free survival in HER2-positive breast cancer patients treated with trastuzumab in
adjuvant setting. *Pol J Pathol*, **66**, 133-41.

Adzhubei IA, Schmidt S, Peshkin L, et al (2010). A method and sever for predicting damaging missense mutations. *Nat Methods*, **7**, 248-9.

Black JL, Harrell JC, Leisner TM, et al (2015). CIB1 depletion impairs cell survival and tumor growth in triple-negative breast cancer. *Breast Cancer Res Treat*, **152**, 337-46.

Cariaso M, Lennon G (2012). SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res*, **40**, 1308-12.

Caserta E, Egríboz O, Wang H, et al (2015). Noncatalytic PTEN missense mutation predisposes to organ-selective cancer development *in vivo*. *Genes Dev*, **29**, 1707-20.

Cheng J, Randall A, Baldi P (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, **62**, 1125-32.

Choi Y, Sims GE, Murphy S, et al (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, **7**, e46688.

DeGraffenried LA, Fulcher L, Friedrichs WE, et al (2004). Reduced PTEN expression in breast cancer cells confers susceptibility to inhibitors of the PI3 kinase/Akt pathway. *Ann Oncol*, **15**, 1510-6.

Dehouck Y, Grosfils A, Folch B, et al (2009). Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics*, **25**, 2537-43.

DeSantis C, Ma J, Bryan L, et al (2014). Breast cancer statistics, 2013. *CA Cancer J Clin*, **64**, 52-62.

Fraczkiewicz R, Braun W (1998). Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J Comp Chem*, **19**, 319-33.

Gonzalez-Angulo AM, Ferrer-Lozano J, Stemke-Hale K, et al (2011). PI3K pathway mutations and PTEN levels in primary and metastatic breast cancer. *Mol Cancer Ther*, **10**, 1093-101.

Haiman CA, Stram DO, Cheng I, et al (2006). Common genetic variation at PTEN and risk of sporadic breast and prostate cancer. *Cancer Epidemiol Biomarkers Prev*, **15**, 1021-5.

Heikkinen T, Greco D, Pelttari LM, et al (2011). Variants on the promoter region of PTEN affect breast cancer progression and patient survival. *Breast Cancer Res*, **13**, 130.

International HapMap C, Altshuler DM, Gibbs RA, et al (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52-8.

Kechagioglou P, Papi RM, Provatoappolou X, et al (2014). Tumor suppressor PTEN in breast cancer: heterozygosity, mutations and protein expression. *Anticancer Res*, **34**, 1387-400.

Kumar P, Henikoff S, Ng PC (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, **4**, 1073-81.

Kwan ML, Kushi LH, Weltzien E, et al (2009). Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast Cancer Res*, **11**, 31.

Lee JO, Yang H, Georgescu MM, et al (1999). Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell*, **99**, 323-34.

Lindahl E, Azuara C, Koechl P, et al (2006). NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res*, **34**, 52-6.

Maggi LB, Jr., Weber JD (2015). Targeting PTEN-defined breast cancers with a one-two punch. *Breast Cancer Res*, **17**, 51.

Martin AM, Weber BL (2000). Genetic and hormonal risk factors in breast cancer. *J Nutr Cancer Instr*, **92**, 1126-35.

Ng PC, Henikoff S (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, **31**, 3812-4.

Ng PC, Henikoff S (2006). Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet*, **7**, 61-80.

Saika K, Sobue T (2013). [Cancer statistics in the world]. *Gan To Kagaku Ryoho*, **40**, 2475-80.

Schaefer C, Meier A, Rost B, et al (2012). SNPdb: constructing an nsSNP functional impacts database. *Bioinformatics*, **28**, 601-2.

Sherry ST, Ward MH, Kholodov M, et al (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, **29**, 308-11.

Sood A, Ghosh AK (2006). Literature search using PubMed: an essential tool for practicing evidence-based medicine. *J Assoc Physicians India*, **54**, 303-8.