

# Genetic and Computational Analysis of Tgfb1 & Fgfr2 Polymorphism in Correlation to Breast Cancer Susceptibility in Pakistani Women

Uzma Shaukat<sup>1\*</sup>, Mehwish Toor<sup>2</sup>, Bilal Ahmad<sup>3</sup>, Sahar Fazal<sup>2</sup> and Nasir Mehmood<sup>4</sup>

<sup>1</sup>Institute of Biomedical and Genetic Engineering Islamabad, Pakistan

<sup>2</sup>Department of Bioinformatics, Muhammad Ali Jinnah University Islamabad, Pakistan

<sup>3</sup>International Islamic University Islamabad, Pakistan

<sup>4</sup>University of Health Sciences Lahore, Pakistan

## Abstract

Worldwide genome-wide association studies have proven very helpful in analyzing the association of susceptible breast cancer genes. In the study presenting here association of susceptible genes FGFR2 and TGFB1 to breast cancer development was explored in Pakistani population which will be helpful in determining the choice of therapy and genetic counseling for ultimately cure of the disease and better treatment. The outcome is that we analyzed 100 samples including follow up patients and found that T-C base substitution of FGFR2 rs1219648 and rs2981582 was strongly associated with the risk of breast cancer whereas the T-C base substitution of TGFB1 rs1800470 was found in some cases though not in all patients suggesting its association within specific caste Awan and Rajput belonging to Mianwali and Khushab district area and thereby elucidating weak association with breast cancer risk. In follow ups we have analyzed that patient with T nucleotide at base position 88 had showed no therapeutic response to chemotherapy i.e. taxanestamoxifen, whereas individuals with C nucleotide chemotherapy was effective. In Silico analysis revealed that mutations of FGFR2 are in intronic region. We subsequently scanned its intronic region. We subsequently scanned its intronic region and mutant structure comparison of TGFB1 with native showed drastic change in its structure having stemmed -loop which blocks the receptor site. Detailed medical history of patients revealed that in our population breast cancer is strongly associated with risk factors such as post menopause, high tea consumption and same caste of spouse and previous breast biopsy. Statistical analysis showed that the frequency of breast cancer at age greater than 40 is higher. It has also been seen that women who had children after age 30 were more susceptible to breast cancer. P value was greater than 0.1 which is not significant showing its high prevalence in our society similar to western population. Phylogenetic analysis has also been done on both genes. In FGFR2 mouse and humans are in same clusters showing close relation whereas in TGFB1 horse and chimpanzee show convergence to human. This analysis will help in choice of economical organism for pharmacogenomics to check drug response.

**Keywords:** rs1800470; SNP; TGFB1; Biopsy; Mutant; Chemotherapy

## Introduction

Understanding the genetic mechanism of a disease often initiates with uncovering alterations that directly or indirectly causes cells to malfunction or due to various environmental factors.

Breast cancer is the most commonly occurring cancer and the leading cause of cancer death among women worldwide, accounting for 27% of all cancers [1]. Breast cancer is the second most common cause of death among women in Western countries [2]. For the year 2009, it was estimated that in the United States approximately 192,370 female patients would be diagnosed with breast cancer and 40,170 would die from it [3]. In developing countries such as China, the increasing trend for the Western lifestyle has been accompanied by a sharp rise on the rate of breast cancer, especially in urban areas and for those of their prime age [4]. In China, the incidence rate of breast cancer has increased dramatically in recent years, with a sharp rise of 38.5% between 2000 and 2005 [5]. There is an alarming increase in the incidence of breast cancer in Pakistan as well. In Pakistan out of every nine women is on risk of the disease accounting for 38.4% in country (Omar Aftab, coordinator of Pink ribbon research).

There are some recognized risk factors that may contribute to the development of breast cancer including age, ethnicity, reproductive events, such as earlier menarche, null parity, later age at first birth and later menopause, exogenous hormones, environmental and lifestyle factors, bone density, as well as genetic factors [6]. The exact etiology of breast cancer is still poorly understood.

In recent years lot of effort has been made in understanding major risk factors that contributes to the development of Breast cancer. The penetrance of mutations in highly susceptibility alleles i.e. BRCA1 and BRCA2 have been studied in detail but to define the complete breast cancer risk some low to moderate susceptibility alleles are also equally important. Family-based studies indicates that only small percentage of cases are associated with known high penetrance alleles and it accounts for only 5-10% of breast cancer cases while the etiology of rest still remains to be explained. Common variants in other lower penetrance genes may be more important and may account for higher attributable risks [7]. Low-penetrance susceptibility genes combining with environmental and heritable factors have been indicated to be important for carcinogenesis. There are also many candidate low-penetrance breast cancer genes and many more are likely to be identified. In addition alterations, sequential and epigenetic changes in

\*Corresponding author: Uzma Shaukat, Senior Scientific Officer, Institute of Biomedical and Genetic Engineering, 24 Mauve Area, Sector G-9/1, Islamabad, PO Box: 2891, 44000-Pakistan, Tel: 0321-9424903; E-mail: [uzma.shaukat283@hotmail.com](mailto:uzma.shaukat283@hotmail.com)

Received August 21, 2014; Accepted September 30, 2014; Published October 03, 2014

Citation: Shaukat U, Toor M, Ahmad B, Fazal S, Mehmood N (2014) Genetic and Computational Analysis of Tgfb1 & Fgfr2 Polymorphism in Correlation to Breast Cancer Susceptibility in Pakistani Women. J Cancer Sci Ther 6: 433-439. doi:10.4172/1948-5956.1000305

Copyright: © 2014 Shaukat U, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

many genes play an important role in the pathology of breast cancer. Easton [8] discovered five breast cancer susceptibility genes, including TNRC9/TOX3 at 16q12, fibroblast growth factor receptor 2 (FGFR2) on chromosome 10q26, LSP1 at 11p15, MAP3K1 at 5q11, and a locus on 8q. Lot of effort is necessary in understanding the etiology of low susceptibility genes and the purpose of this work is to highlight the major low susceptibility genes and figure out its SNP variants in Pakistani population as lot of work has been carried out in Europeans, scarcely in Asians i.e. in Chinese population. Common variants in lower penetrance genes may be more important and may account for higher attributable risks [9]. Although low penetrance genes have very small effect on Breast cancer risk but in combination with other genes or environmental factors, making it significant Breast cancer risk factor. These type of variations are known as polymorphisms and are becoming a keen area of research in Medical genetics. Association studies of polymorphisms have been very helpful in understanding the etiology of disease as well as determining response of therapy in patients and choice of therapy to be adopted. Genome wide association studies have been conducted populations other than some in Asian population (in Chinese) so it is a key need to screen Pakistani population to check the frequency of polymorphism in low penetrance genes. Thereby association studies are very promising in identifying common, low-penetrance susceptibility alleles for many complex diseases.

Fibroblast Growth Factor Receptor 2 (FGFR2) is a protein that is encoded by FGFR2 gene. Fibroblast growth factor receptor 2 is a tyrosine kinase receptor that is a member of the family of individually distinct fibroblast growth factor receptors (FGFR) involved in cell proliferation, invasiveness, motility, and angiogenesis. It is located on the long (q) arm of chromosome 10 at position 26. Genome-wide association studies have identified FGFR2 as a breast cancer (BC) susceptibility gene in populations of European and Asian descent. After that, a number of studies reported that the rs2981582, rs1219648, and rs2420946 polymorphism in FGFR2 has been implicated in breast cancer risk [10].

Transforming growth factor beta 1 (TGF- $\beta$ 1) is a polypeptide member of the transforming growth factor beta superfamily of cytokines encoded by TGFB1 gene. It controls proliferation, differentiation and other functions in many cell types. It is located on the long (q) arm of chromosome 19 at position 13.1. Studies on Chinese population reported that SNP polymorphism rs1982073 contribute to the more aggressive breast cancer risk factor [11].

## Materials and Methods

### Extraction of genomic DNA

Blood samples were drawn with informed consent by 5 ml syringes in 10 ml Vacutainers. Genomic DNA was extracted from blood samples using standard organic method known as Phenol chloroform method

### Primer designing

Primers were designed for the selective SNPs of FGFR2 (rs2981582, RS1219648) and of TGFB1 (rs1800470) using tool Primer 3.

### Polymerase chain reaction (PCR)

Standard PCR reactions were carried out by the use of Bioline products to amplify the DNA. Standard constraints for cycling consisted of an initial denaturation step for 2min at 94°C, followed by 35 cycles comprising of denaturation at 94°C for 45seconds, an annealing step for 45 seconds again and extension at 72°C for 45

seconds with a final extension step at 72°C for 10 minutes. The specific annealing temperature for each pair of the primers was then optimized to give a maximum yield of the required PCR product.

### Agarose gel electrophoresis

For sequencing we need good amplification so to check this PCR product run on agarose gel. The gel was loaded with 5 $\mu$ l orange dye with 100bp ladder and run in Maxi cell Electrophoretic gel systems. The gel was then placed on a UV Trans illuminator and then gel was interpreted.

### Sequencing

The PCR product was first purified and then sequencing reaction was carried out under standard conditions i.e. 94°C for 10 sec and 60°C for 10 min for 25 cycles.

### SNP analysis

Amino acid variants that are directly affecting protein function were analyzed using SIFT server. The structural changes in the protein structure were depicted by Polyphen server. Untranslated motifs were scanned by UTRscan server.

### Modeling mutant structure

The mutant sequences were submitted to I-Tasser server for modeling and then compared it to its native structure. Structural effects of mutation of interest were predicted by Hope server.

### Statistical analysis

Statistical analysis was done on different variables including age, caste of spouse, age at Menopause, first live Birth, family history, medications, education level and tea consumption.

### Phylogenetic analysis

Protein sequences of FGFR2 and TGFB1 from different species were retrieved from Uniprot and build the tree to find out the founder effect and this will help to analyze the species related to human in which we can be able to do further gene analysis in those species using ClustalW and ClustalW2 phylogenetic tree.

## Results

For Genetic association of candidate SNPs and Breast cancer susceptibility the primers of FGFR2 gene for SNP rs2981582 and rs1219648 were optimized at 55°C and of TGFB1 gene for SNP rs1800470 were at 58°C. Gel pictures of FGFR2 and TGFB1 gene are shown in Figures 1a and 1b respectively.

The sequencing reactions were performed on an automated ABI PRISM 3100 Genetic Analyzer using Big Dye Terminator. Sequencing was performed in 25 cycles under conditions 96°C for 10sec and 60°C for 4min. Sequencing was performed using designed optimized primers of FGFR2 (rs1219648, rs2981582) and TGFB1 (rs1800470) genes on breast cancer cases and controls. The primer sequences are shown in Table 1.

We found out substitutions from T-C at base position 88 and in some cases C-T base substitution at position 201 showed that in our Pakistani population FGFR2 (rs1219648, rs2981582) is strongly associated with Breast cancer. We observed T-C substitution of other gene TGFB1 (rs1800470) in approximately 25% of our population in specific ethnic group belonging to northern Punjab showing its weak association with Breast cancer. The breast cancer patients belonging

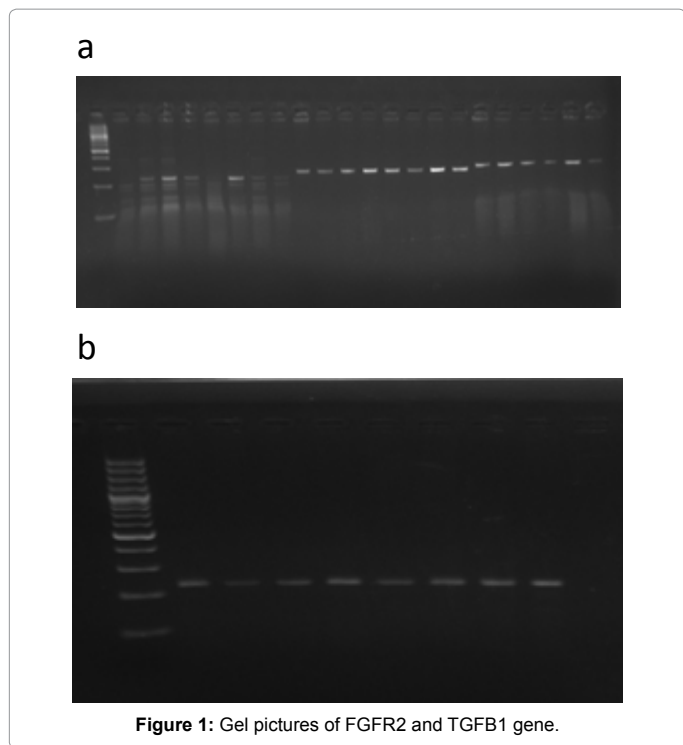


Figure 1: Gel pictures of FGFR2 and TGFB1 gene.

Sr.No	Primer name	5'-3' sequence	Length (bases)
1	rs 2981582 forward	AGCCAAGCCTCTACTTGGTG	20
	rs 2981582 reverse	CTCTGTCTCTCCCAGCCTA	20
2	rs1219648 forward	TCTCTTCTCACTGTGATTTG	22
	rs1219648 reverse	CAATGGCGCAGAATTACTTAC	21
3	rs1800470 forward	CACCACACCAGCCCTGTTC	19
	reverse	CCGCAGCTTGGACAGGAT	18

Table 1: Primer sequences of rs2981582, rs1219648 and rs1800470.

to Mianwali and Khushab district, of Awan and Rajput caste showed association with TGFB1 gene polymorphism. Furthermore patient's follow-ups highlighted the fact that patient having C-T substitution had showed no therapeutic response to chemotherapy taxanes and tamoxifen whereas individuals having T-C substitution chemotherapy was effective. The sequencing results of FGFR2 (rs1219648 and rs2981582) and TGFB1 (rs1800470) are shown in Figures 2 and 3 respectively.

For further analysis computational approaches were applied on selected SNPs of both genes that is rs1219648 and rs2981582 of FGFR2 gene and rs1800470 of TGFB1 gene. From sequencing results we found that mutations in FGFR2 lie in intronic region and that of TGFB1 in exonic region. So we performed intronic analysis of FGFR2 to scan its region for various functional aspects so the gene is scanned in UTRscan server to search the UTR motifs in the region. We found 8 motifs in the scanned region of rs1219648 and rs2981582 of FGFR2 shown in Table 2.

As mutations of TGFB1 are in exonic region therefore sift and Polyphen analysis was applied to predict its functional and structural damages. The protein sequences was submitted to the SIFT program. Mutation of Valine to Methionine at position 110 found to be deleterious affecting protein function. Proline to Leucine at position 20 does not affect protein function. Polyphen predicted that the two submitted mutations were found to be damaging the structure.

To model the mutant structures I-tasser server was used. The sequences of SNP IDs that were found to be structural damage were sent to I-tasser server for modeling. The modeled structures were refined by using FG\_MD server and validated by using Prosa-web and RamPage ramachandran plot shown in Figure 4. By comparing the mutant structure with the native structure of TGFB1, it has been clearly seen that a loop is being formed in the mutant structure leaving no space to bind with ligand shown in Figure 5.

### Statistical analysis

A questionnaire was structured which included different environmental factors like age, consanguinity of parents, marital status, caste of spouse, age at menarche, age at marriage, age at first

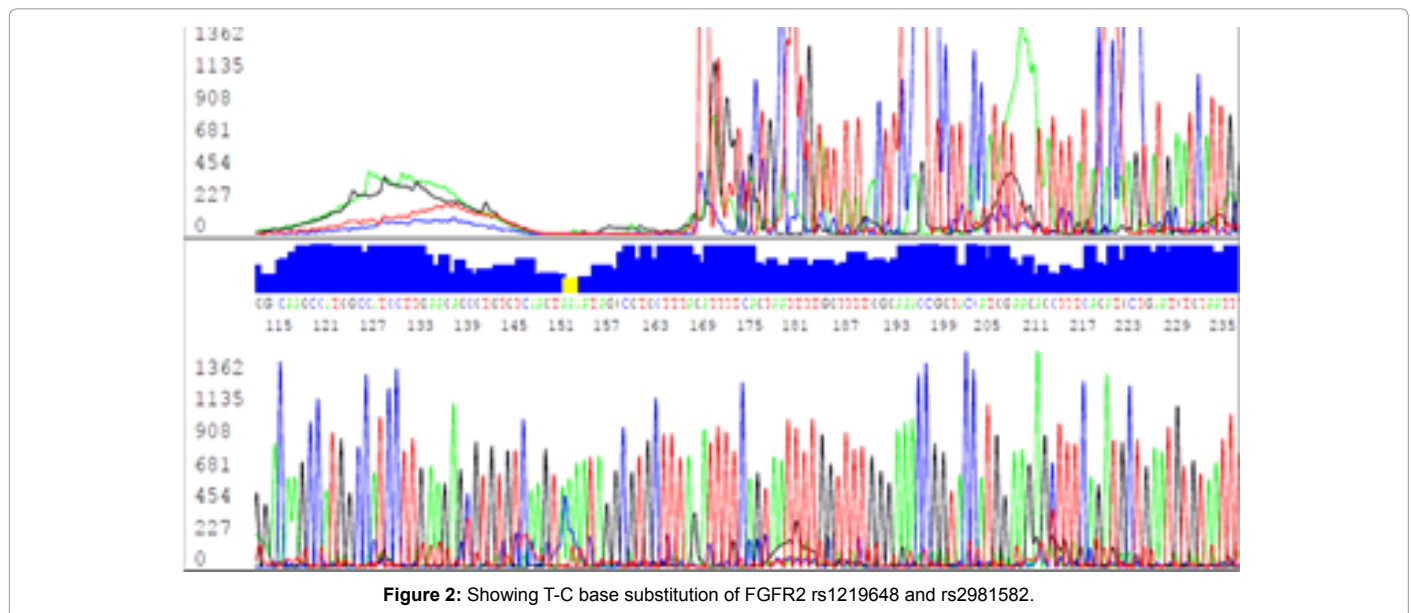
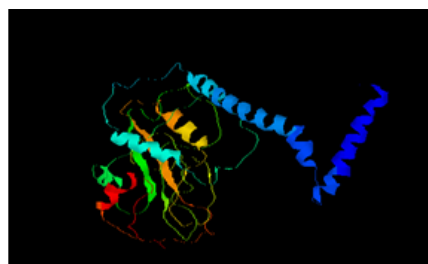
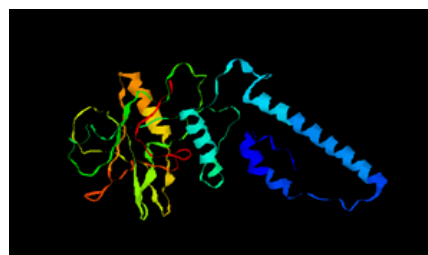


Figure 2: Showing T-C base substitution of FGFR2 rs1219648 and rs2981582.





Native



Mutant

Figure 5: Comparison of mutant structure of TGFB1 with its native structure.

live birth, age at menopause, number of pregnancies, abortions, miscarriages, education level, breast feeding, history of breast cancer, lump in breast, previous breast biopsy, anxiety, consumption of tea and any other disease. Interviewing patients revealed that in our population breast cancer is strongly associated with postmenopausal risk factor, high tea consumption and same caste of spouse and previous breast biopsy. Other factors like early menarche and strong family history also have strong association with breast cancer. In our population survey also showed that patients affected as breast cancer are mostly illiterate and was not aware of cancer awareness programs. Statistical findings showed that the estimated prevalence rate of breast cancer increases with increasing age.

It was concluded from the statistical analysis of our data that the incidence of breast cancer is higher in blood relations. Usually it is highly prevalent in same caste. Furthermore women who give birth in later ages (i.e. after age 30) are at greater risk of Breast cancer. Medications are also cause of breast cancer. Different types of radiations increase the risk of breast cancer shown in Figure 6. The different types of menopause frequency in different medications are shown in Tables 3 and 4. Chi square tests showed that minimum expected count is .08 shown in Table 4.

The variations related groups related to humans were selected including mouse, rat, chimpanzee, horse, chick and zebra fish. In FGFR2

We got three main clusters and two sub clusters. Mouse and humans are in same cluster showing close convergence and rat is also showing similarity as it is less distant from the root whereas chick and zebra fish showing divergence shown in Figure 7. In TGFB1 chimpanzee and horse falls in same cluster to humans and mouse and rat showing convergence shown in Figure 8. This analysis will be helpful for choice of organism which is easy to handle and economical.

## Discussions

There is an alarming increase in the occurrence of breast cancer and is common cause of death worldwide and in Pakistan. There

are number of complex factors that are involved in the behaviour or etiology of breast cancer involving genetic and environmental factors. In the past, it had been assumed that breast cancer is due to inherited mutations of common highly penetrant gene Brca1 and Braca2 but various studies have suggested that these genes are contributing only 5-10 percent of the disease therefore about remaining 95%. Needs to be explored then through various genome-wide association studies it has been found out that moderate to low penetrant genes together with various environmental factors are also important and may be contributing more in breast cancer susceptibility. In our study we checked the association of FGFR2 and TGFB1 with breast cancer susceptibility, as lot of studies have been reported which suggest the association of these genes in different communities of Europeans and Asian descent.

The FGFR2 belongs to a family of fibroblasts growth factor receptor (FGFR) and have tyrosine kinase (TK) activity and is involved in number of cell processes including cell proliferation, wound healing, embryonic development and angiogenesis [12]. A Study in European women in postmenopausal group that Four SNPs in intronic region of FGFR2 are highly associated with risk of breast cancer. In Arab



Figure 6: At age of 40, breast cancer is highly occurred in Pakistani population.

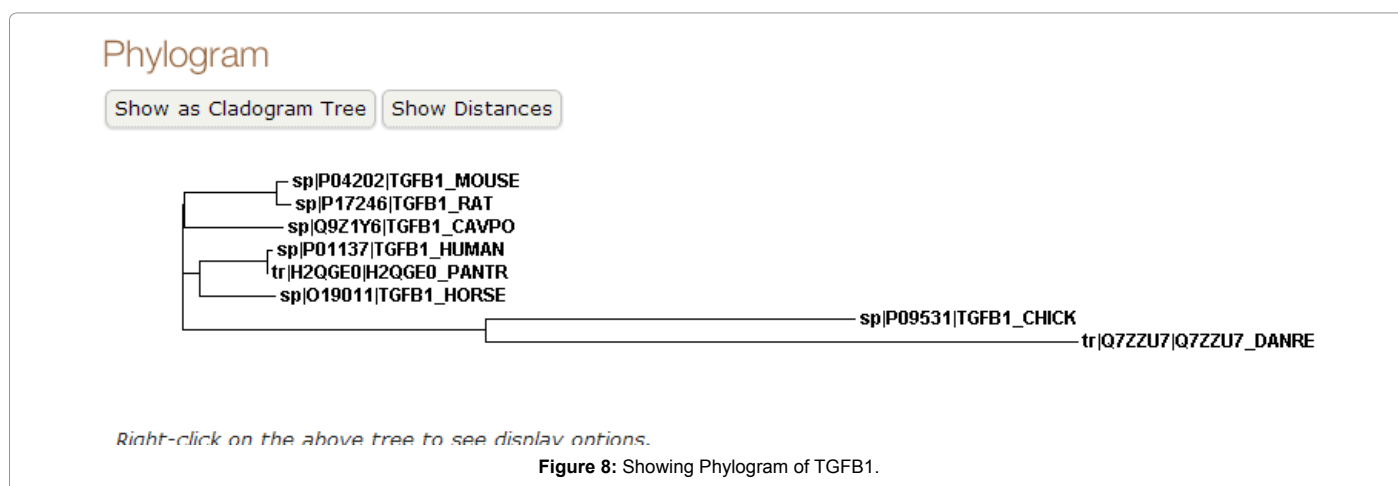
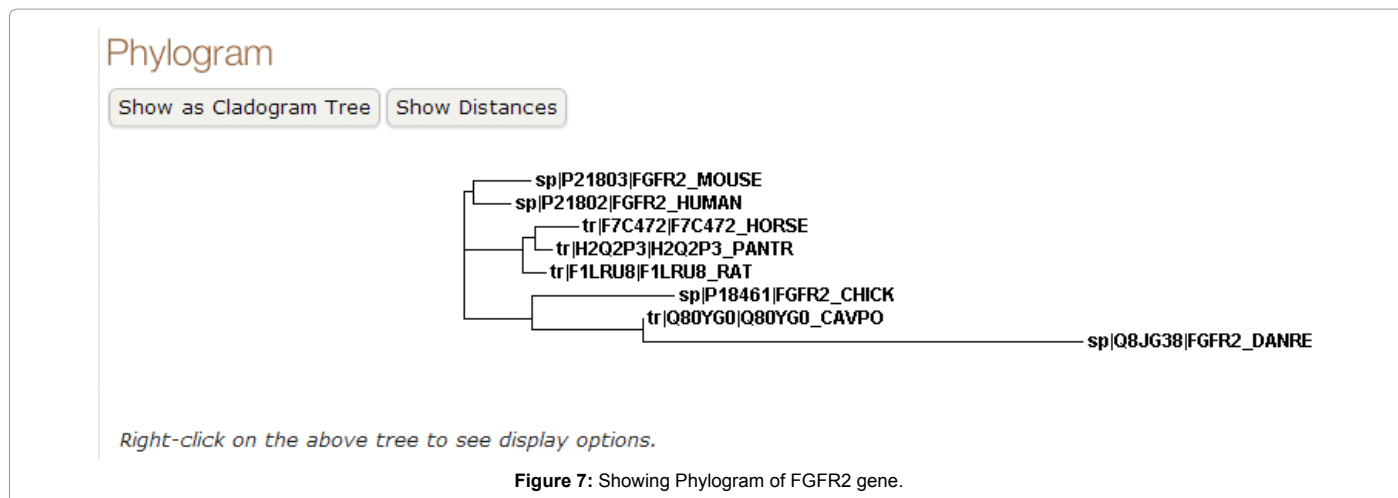
Types of Menu pause	Frequency	Percent	Valid Percent	Cumulative Percent
	4	7.7	7.7	7.7
Medications	4	7.7	7.7	15.4
Natural	19	36.5	36.5	51.9
Nil	15	28.8	28.8	80.8
Radiation	6	11.5	11.5	92.3
Surgical	4	7.7	7.7	100.0
Total	52	100.0	100.0	

Table 3: Comparison of different type of Menu pause.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	52.121 <sup>a</sup>	4	.000
Likelihood Ratio	17.071	4	.002
N of Valid Cases	52		

a. 5 cells (55.6%) have expected count less than 5. The minimum expected count is .08.

Table 4: Chi-Square Tests.



population, the genetic variations in rs1219648 show strong association with breast cancer. A study conducted on Chinese women to check the polymorphism of FGFR2 revealed that SNP polymorphism including rs2981582 in which C/T frequency is higher, rs 1219648 A/G and rs2420946 was significantly associated with breast cancer [13]. In the recent year the effect of FGFR2 and B7-H4 polymorphism association with breast cancer was investigated.

Transforming growth factor beta is polypeptide member of superfamily of cytokines and help cells to proliferate, differentiate and migrate [14]. There is strong evidence which suggest that TGFB1 has dual role acting as tumor suppressor and initiator of tumor progression. The polymorphism of TGFB1 has significant association with breast cancer susceptibility [15]. In Studies conducted in Chinese women to determine the association of TGFB1 with breast cancer, it was concluded that SNPs of TGFB1 rs1800470 associates with risk of breast cancer [6]. Large scale analysis from 12 different studies showed that TGFB1 rs1800470 presented the best significance.

In our study, we investigated the association of FGFR2 rs1219648, rs2981582 and TGFB1 rs1800470 in our Pakistani population and in addition also checked the response of chemotherapy. Our results reflect the variations among various geographic areas. The Analysis showed that the association of TGFB1 rs1800470 in our population is not statistically significant as it is detected in some cases specific to Mianwali and khusab district showing its weak association with

breast cancer risk. The T-C base substitution of FGFR2 rs1219648 and rs2981582 was strongly associated with breast cancer risk. The frequency of T allele is higher in breast cancer cases as compared to North China [6]. The structural analysis result of TGFB1 is found associated with findings in women of Han nationality. In the protein sequence we found the same leu10Pro base substitution and additional Val20Met base substitution association with an enhanced rate of breast cancer [6]. We extended our work as compared to the previous research on association study in Han women by analyzing the effect of chemotherapy in follow-up patients and we have observed that individuals having C-T substitution had showed no therapeutic response to chemotherapy whereas T-C substitution showed effective chemotherapy response.

In silico analysis of both genes helps us a lot in understanding the genetics of FGFR2 and TGFB1. The analysis revealed that leu10Pro and val20Met mutations in the protein sequence of TGFB1 are found to cause drastic changes in the structure forming stem-loop as compared to native. These mutations support the evidence that TGFB1 is associated with increased incidence of persistent breast cancer [16].

The survey conducted in population by interviewing the patients and investigation of different environmental factors including age, marital status, caste of spouse, family history, age at menarche, age at menopause, age at first live birth showed that the risk factors like postmenopausal condition, high tea consumption and same caste

of spouse are highly associated with the risk of breast cancer. These findings correlate with postmenopausal women of western world. Statistical analysis also revealed that risk of breast cancer is associated with women of higher ages of 40 plus as in American population [17].

## Future Work

Two genes FGFR2 and TGFB1 were screened in our local population and we found that FGFR2 gene is strongly associated with breast cancer etiology but TGFB1 gene association is found in 25% samples not in all. Also we found that these 25% of cases belong to Mianwali and Khushab district of Pakistan and were Awan and Rajput by caste. Future work will encompass increase in sample size and focus on above mentioned geographical areas and ethnicities to establish strong genetic association of TGFB1 gene polymorphism with breast cancer susceptibility causing severe damages to structure thereby to use the knowledge for better treatment of cancer patients.

## Acknowledgement

We are thankful of the participating institutes, hospital and physicians who provided us with samples. Dr. Uzma Shaukat, for conceiving the project and supervising the research work and Mehwish Toor who conducted the experimental procedures. We also wish to thank Dr. Nasir Mehmood, Dr. Sahar Afzal and Bilal Ahmad, whose useful suggestions and kind guidance helped in the conductance of this research.

## References

1. Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M, et al. (1991) BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* 6: 659-663.
2. Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, et al. (2010) SEER Cancer Statistics Review, 1975-2007. Bethesda: National Cancer Institute.
3. Brandt B, Hermann S, Straif K, Tidow N, Buerger H, et al. (2004) Modification of breast cancer risk in young women by a polymorphic sequence in the egfr gene. *Cancer Res* 64: 7-12.
4. Bhurgri Y, Bhurgri A, Hassan SH, Zaidi SH, Rahim A, et al. (2000) Cancer incidence in Karachi, Pakistan: first results from Karachi Cancer Registry. *Int J Cancer* 85: 325-329.
5. Bray F, McCarron P, Parkin DM (2004) The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res* 6: 229-239.
6. Chen XH, Li XQ, Chen Y, Feng YM (2011) Risk of aggressive breast cancer in women of Han nationality carrying TGFB1 rs1982073 C allele and FGFR2 rs1219648 G allele in North China. *Breast Cancer Res Treat* 125: 575-582.
7. Dumitrescu RG, Cotarla I (2005) Understanding breast cancer risk -- where do we stand in 2005? *J Cell Mol Med* 9: 208-221.
8. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087-1093.
9. Liang J, Chen P, Hu Z, Zhou X, Chen L, et al. (2008) Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis* 29: 2341-2346.
10. Jia C, Cai Y, Ma Y, Fu D (2010) Quantitative assessment of the effect of FGFR2 gene polymorphism on the risk of breast cancer. *Breast Cancer Res Treat* 124: 521-528.
11. Chen YC, Hunter DJ (2005) Molecular epidemiology of cancer. *CA Cancer J Clin* 55: 45-54.
12. Katoh M (2008) Cancer genomics and genetics of FGFR2 (Review). *Int J Oncol* 33: 233-237.
13. Liang J, Chen P, Hu Z, Zhou X, Chen L, et al. (2008) Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis* 29: 2341-2346.
14. Derynck R, Akhurst RJ, Balmain A (2001) TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet* 29: 117-129.
15. Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, et al. (2007) Cancer incidence in five continents. 9 ed. Lyon: IARC.
16. Dunning AM, Ellis PD, McBride S, Kirschenlohr HL, Healey CS, et al. (2003) A transforming growth factorbeta1 signal peptide variant increases secretion in vitro and is associated with increased incidence of invasive breast cancer. *Cancer Res* 63: 2610-2615.
17. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, et al. (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39: 870-874.