Characterization of Small Genetic Variants in Breast Cancer Cell Line Under Tamoxifen Therapy

Mahnaz Nezamivand-Chegini1, Hamed Kharrati-Koopaee1, Seyed Taghi Heydari2, Hassan Giahi1, Fatemeh Sabahi3, Ali Dehshahri4, Kamran Bagheri Lankarani2

1 Institute of Biotechnology, Shiraz University, Shiraz, Iran
2 Health Policy Research Center, Institute of Health, Shiraz University of Medical Sciences, Shiraz, Iran
3 Department of Plant Protection, College of Agriculture, Shiraz University, Shiraz, Iran
4 Department of Pharmaceutical Biotechnology, Shiraz University of Medical Science, Shiraz, Iran

Abstract

Background: Tamoxifen (TAM) is an effective hormone therapy that reduces the risk of cancer recurrence. According to evidence, TAM contributes to the alterations of genetic variants background and plays a role in the effectiveness of treatments via alteration of the genetic variants. The effects of TAM on genomic features were investigated in the current study by discovering genetic variants and finding the answer to the following question: “Is there any association between the alterations of genetic variants under TAM consumption and an effective treatment process?”

Materials and Methods: Whole-transcriptome (RNA-seq) dataset from four investigations including 10 TAM-treated samples and 9 untreated samples as the control groups were derived from European Bioinformatics Institute (EBI). Using the process of variants calling, the differential genetic variants between and gene ontology enrichment analysis were detected by CLC Genomics Workbench (12).

Results: Current study reported about 5.8 million genetic variants. The outcomes of the chi-square test showed that distributions of genetic variants between control and treated samples were significant (p<0.05). The genetic variants comparison between the control and TAM-treated samples indicated that there were 67 differential genetic variants. Gene ontology enrichment analysis indicated that differential genetic variants were associated with several tumor suppressors and oncogenes including IL6ST, GEN1, FNTA, HSPA5, NSMCE2, and DDX11.

Conclusion: Most of the candidate genes with different genetic variants had dual roles as oncogenes or tumor suppressors. Therefore, it can be argued that TAM does not play a significant role in an effective treatment through alteration of the genetic variants. In other words, it cannot be concluded that the TAM therapy-resulted alterations of genetic variants play a positive or negative role in the treatment process.

[GMJ.2023;12:e2598] DOI:10.31661/gmj.v12i0.2598

Keywords: Tamoxifen; Breast Neoplasms; RNA-seq

Correspondence to: Seyed Taghi Heydari, Biostatistics, Health Policy Research Center, Institute of Health, Shiraz University of Medical Sciences, Shiraz, Iran. Telephone Number: +989173034420 Email Address: heydari.st@gmail.com
Introduction

Cancer is one of the leading causes of mortality worldwide and breast cancer is the second most common disease in women [1]. Hormonal therapy is a medical strategy for treating breast cancer [2]. Tamoxifen (TAM) is considered the main non-steroidal drug for the treatment of breast cancer in postmenopausal women [3], which competitively inhibits estrogen activity through binding to the estrogen receptor [4].

Several investigations have been carried out to illustrate the hormonal therapy effects and it will allow us to better understand the drug response mechanism and select an effective strategy for the therapeutic period [5-7]. The appropriate drug response is a complex interdependent procedure that is highly dependent on multiple factors, including genetic variants background, lifestyle, climate, smoking, and alcohol consumption [8].

Genetic variants refer to the genetic differences between individuals within a population [9]. DNA is a vulnerable molecule against various mutagens including ultraviolet, toxins, chemical agent, and free radicals [10]. Recently, high-throughput sequencing platforms have been applied as powerful tools to investigate the association between a large number of genetic variants and drug response [11,12].

It is shown that TAM has a mutagenic effect on the endometrium cells and increases the incidence of endometrial tumors [13]. Results from an evaluation of rat hepatic tissue showed that activated TAM could bind to the guanine N2-position of DNA and consequently, produce pro-mutagenic lesions [14]. More importantly, it was found that the TAM mutagenicity effect induced DNA damage in human endometrial cells [15]. Emons et al. (2020) showed that TAM may play a key role in tumor progression. It may increase the risk of uterus cancers, such as endometrial cancer and uterine sarcoma [16].

In vitro conditions, TAM would lead to gene mutations and increased incidence of abnormal chromosomal structures in rat liver tissues [17]. All of the mentioned literature reviews indicated that TAM could play a critical role in the alterations of genetic variants' background. Furthermore, vaginal dryness, sleep problems, weight gain, hot flashes, and depression have been reported as common TAM side effects [18]. There are several examples regarding the role of genetic variants in drug response. To achieve a therapeutic effect, there has to be an interaction between the drug and its target. DNA variations can both increase and decrease a drug binding affinity to its target. As an example, genetic variations can change the antagonist role of the drug into an agonist one; therefore, the most common problem of treatment procedures is resistant mutations in drug targets. TAM blocks estrogen receptor (ER-positive cancer) in the breast cancer treatment procedure and consequently, reduces the risk of cancer recurrence. It is an anti-estrogen hormone that inhibits the estrogen receptors; however, its efficiency would be decreased as a result of mutations in estrogen receptors and leads to the conversion of ER-positive into progesterone-positive cancer (PR-positive cancer). Consequently, it causes drug resistance development and a lack of response to treatment [19].

It is noteworthy that genetic variations may contribute to drug metabolism and affect drug response. For instance, if the drug is rapidly metabolized, its concentration will reduce due to the weaker drug action or side effects. Considering slower metabolism procedures, higher drug levels would result in stronger or longer-lasting effects and side effects [20].

The current study investigates the effect of TAM consumption on genetic variants background in the breast cancer cell line (MCF7). Since TAMs are mutagenic agents, there may be a link between genetic variants alterations and TAM treatment; therefore, it can affect the treatment process. It can also provide new insights to improve chance of survival, reduce side effects, and select appropriate strategies for treatment duration.

Materials and Methods

1. Data Collection

In the current study, the 19 whole-transcriptome (RNA-seq) datasets of four investigations were derived from European Bioinformatics Institute (EBI) (https://www.ebi.ac.uk/). The treatment group includes...
Alterations of Small Genetic Variants Under Tamoxifen Therapy

10 MCF7 cell lines treated with TAM and 4-hydroxytamoxifen (4-OHT), as well as 9 untreated MCF7 cell lines considered as the control groups. More details of collected samples were provided in Table-1. An overview of the genetic variants analyses for collected samples is showed in Figure-1.

2. Quality Control and Trimming
Quality control features of CLC Genomic Workbench (12) including length distribution, GC content, ambiguous base content, Phred score, nucleotide contribution, and duplicate sequences were applied to achieve proper quality control of the collected data [21]. Since adaptor sequences were cleaned in the achieved datasets, the adaptor trimming was not formed.

3. Genetic Variants Analysis

3.1. Reference Genome and Alignments Analysis
The reference genome (hg38) and all annotations were downloaded from the Ensembl database (www.ensembl.org). Mapping of short reads against the reference genome was performed by CLC Genomics Workbench 12 based on the following parameters: masking track=mRNA sequence, mismatch cost=2, cost of insertions and deletions=linear gap cost, insertion cost=3, deletion cost=3, length fraction=0.7, and similarity fraction=0.8 [22].

3.2. Variant Calling and Statistical Analysis
CLC genomics workbench 12 was applied for variant detections; also, there was no constant

![Figure 1. The summary of differential genetic variants analysis between treated breast cancer cell line (MCF7) by TAM and control group (untreated MCF7).](image-url)
The variant calling procedure was performed with a low-frequency algorithm based on the following parameters: required variant probability (%)=95.0 ignore broken pairs=yes, minimum coverage=10, minimum count=2, minimum frequency (%)=30, base quality filter=Yes, neighborhood radius=15, minimum central quality=30, and minimum neighborhood quality=25. A Chi-square test was performed to explain the differences in genetic variants distribution between control and treated samples.

3.3. Comparing the Variants and Gene Ontology (GO) Enrichment Analysis

After performing the variants calling process, genetic variants of TAM-treated samples were compared with the reads from control samples to remove the common genetic variants between treated and control samples. The file of gene ontology association, including the gene names and associated gene ontology terms, was downloaded from the gene ontology consortium and imported to CLC Genomic Workbench 12. Moreover, differential genetic variants were applied to perform GO enrichment analysis at the levels of biological process, molecular function, and cellular component. The significance of the level of GO analysis was determined to be 0.01.

Results

Genetic Variants Detection

Results of quality control indicated that there was no special trimming strategy required for RNA-seq datasets. The average quality control factors (per read) of the studied samples were reported as the following parameters, length distribution=131.5 bp, GC content=52.35%, ambiguous base content=0.2%, Phred score=18.12, nucleotide contribution=0.5% and duplicate sequences=2.10%. However, trimming was carried out according to the Phred score and the nucleotide contribution to minimize the mapping errors. At least ten primary bases were trimmed from the 3′ side of short reads, and 5% of reads containing the lowest Phred scores were also ignored. Results of alignments of short reads against the reference genome (hg 38) are provided in Table-2. Furthermore, 66%-89% was reported for the mapping percentage.

Current research has identified almost 5.8 million genetic variants including single nucleotide variations (SNVs), multi nucleotide variations (MNVs), insertion, deletion, and replacement. The highest and lowest frequencies among detected genetic variants were respectively related to SNVs and replacement. More details of genetic variant frequencies are depicted in Figure-2.

To investigate the effect of TAM on genetic variants distribution within control and treated samples, a statistical analysis was performed separately for each genetic variant based on a chi-square test for total genetic variants in the control and treated samples. Results showed that the genetic variants distribution between control and treated samples was significant (P≤0.05) (Table-3), which indicated the possible effects of TAM on the genetic variants frequency.

A comparison of genetic variants in control and treated samples indicated that there were 67 differential genetic variants. Among all
of the differential variants, 16 genetic variants were located in the coding regions and 10 variants led to the change of amino acid sequence within the protein structure. Table 4 shows more details of differential genetic variants.

The process of gene ontology enrichment analysis of different genetic variants was carried out at three levels of biological process, cellular component, and molecular function; therefore, a total number of 77 significant GO terms was reported (Table 5). At the biological process level, the most repetitive of reported overlapping gene names were GEN1, HSPA5, NSMCE2, AURKA, and DDX11 candidate genes.

Results achieved from molecular function analysis indicated that the most frequently enriched candidate genes in significant GO terms were IL6ST, COX15, and FNTA. The cellular component analysis showed that the nucleus and nucleoplasm were the most important cellular parts that may contribute to hormonal therapy.

**Discussion**

Breast cancer is a heterogeneous disease, which is divided into three groups of ER-positive, PR-positive and triple-negative breast cancer (TNBC). Hormone therapy may be used for ER and PR positive tumors; howev-

**Table 2. The Mapping Summary of Short Reads against the Reference Genome.**

<table>
<thead>
<tr>
<th>Accession number</th>
<th>Samples</th>
<th>Total reads</th>
<th>Mapped reads%</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-GEOD-59536</td>
<td>T1</td>
<td>89713168</td>
<td>68.80</td>
</tr>
<tr>
<td>E-GEOD-62613</td>
<td>T2</td>
<td>112247072</td>
<td>85.91</td>
</tr>
<tr>
<td>T3</td>
<td>34981408</td>
<td></td>
<td>81.22</td>
</tr>
<tr>
<td>T4</td>
<td>36012214</td>
<td></td>
<td>81.20</td>
</tr>
<tr>
<td>T5</td>
<td>40160428</td>
<td></td>
<td>82.10</td>
</tr>
<tr>
<td>T6</td>
<td>41384146</td>
<td></td>
<td>82.05</td>
</tr>
<tr>
<td>T7</td>
<td>39870210</td>
<td></td>
<td>81.72</td>
</tr>
<tr>
<td>T8</td>
<td>41063128</td>
<td></td>
<td>81.74</td>
</tr>
<tr>
<td>E-GEOD-78199</td>
<td>T9</td>
<td>10069398</td>
<td>87.85</td>
</tr>
<tr>
<td>T10</td>
<td>12018685</td>
<td></td>
<td>83.10</td>
</tr>
<tr>
<td>E-GEOD-59536</td>
<td>C1</td>
<td>97511228</td>
<td>66.10</td>
</tr>
<tr>
<td>E-GEOD-62613</td>
<td>C2</td>
<td>103822108</td>
<td>87.37</td>
</tr>
<tr>
<td>C3</td>
<td>39172180</td>
<td></td>
<td>82.47</td>
</tr>
<tr>
<td>C4</td>
<td>40347538</td>
<td></td>
<td>82.44</td>
</tr>
<tr>
<td>C5</td>
<td>44328838</td>
<td></td>
<td>80.16</td>
</tr>
<tr>
<td>C6</td>
<td>45695050</td>
<td></td>
<td>80.15</td>
</tr>
<tr>
<td>C7</td>
<td>36382948</td>
<td></td>
<td>82.53</td>
</tr>
<tr>
<td>C8</td>
<td>37484422</td>
<td></td>
<td>82.50</td>
</tr>
<tr>
<td>E-MTAB-822</td>
<td>T9</td>
<td>10069398</td>
<td>87.85</td>
</tr>
<tr>
<td>E-MTAB-822</td>
<td>C9</td>
<td>8569125</td>
<td>89.25</td>
</tr>
</tbody>
</table>

**Table 3. Results of Statistical Analysis of Genetic Variants Distribution between Control and Treatment Samples.**

<table>
<thead>
<tr>
<th>Genomic variants</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNV***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MNV***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insertion***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Deletion***</td>
<td>0.001</td>
</tr>
<tr>
<td>Replacement***</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**SNV:** single nucleotide variations, **MNV:** multi nucleotide variations

Results achieved from molecular function analysis indicated that the most frequently enriched candidate genes in significant GO terms were IL6ST, COX15, and FNTA. The cellular component analysis showed that the nucleus and nucleoplasm were the most important cellular parts that may contribute to hormonal therapy.

Breast cancer is a heterogeneous disease, which is divided into three groups of ER-positive, PR-positive and triple-negative breast cancer (TNBC). Hormone therapy may be used for ER and PR positive tumors; howev-
er, TNBC could not respond to common hormone therapy [24]. TAM is a type of hormonal therapy implemented to treat ER-positive breast cancer; it may also reduce the risk of invasive cancer development.

Our hypothesis regarding the role of TAM in the treatment process has not been fully confirmed. It was found that most of the candidate genes with differential genetic variants had dual roles as oncogenes or tumor suppressors; moreover, their exact contribution to breast cancer has not been investigated precisely. For example, the results of the genetic variant analysis revealed that differential genetic variants between control and treated samples (under TAM therapy) were overlapped with GEN1, HSPA5, NSMCE2, AURKA, and DDX11. GEN1 (Flap endonuclease GEN homolog 1) encoded a member of Rad2/xeroderma pigmentosum group G nuclease family. As it was observed for BRCA1 and BRCA2, GEN1 contributed to resolve the Holliday junction in the homologous recombination. It is noteworthy that the Holliday junction may play a vital role in the cancer chemosensitivity [25]. Somatic truncating GEN1 mutations have been reported in breast cancers; therefore, it would indicate the fact that GEN1 may be a predisposition gene in breast cancer. However, it was shown that although it plays a critical role in the double-strand DNA break repair, GEN1 would not make any appreciable contribution to breast cancer susceptibility through acting as a high- or intermediate-penetrance breast cancer predisposition gene, such as BRCA1, BRCA2, CHEK2, ATM, BRIP1, and PALB2 [26]. Sun et al. (2014) suggested that GEN1 would play a vital role in DNA damage response; therefore, its alteration could lead to breast cancer [27].

Heat-shock protein 5 (HSPA5) is considered a marker of poor prognosis in patients with
breast cancer, which plays a critical role in promoting drug resistance and metastasis [28]. A close association was observed between the cancer behaviors of the heat shock proteins (HSP) family; however, all members of the HSP family have not been studied completely [29]. NSMCE2 is an E3 SUMO ligase and a subunit of the SMC5/6 complex that could be associated with DNA repair [30].

Although SMC5/6 complex functions were not described precisely, reports indicated that it could act as a tumor suppressor in mice [31]. Aurora Kinase A (AURKA) is a serine/threonine kinase contributing to the regulation of cell cycle progression; therefore, it could be a potential cancer susceptibility gene [32]. Furthermore, it is considered a promising target in the treatment processes of patients with cancer [33].

DDX11 is a DNA helicase that plays a role in DNA replication, sister chromatid cohesion establishment, and general chromosome structure. The effects of DNA helicases among patients with cancer are dependent upon their genetic background and tumor type; however, it has not been illustrated precisely and there are various reports of their activities. For example, it was suggested that DNA helicase may have a tumor suppressor function, and the expression level of several DNA helicases at pre-cancerous stages would increase significantly [34].

At the molecular function level, results of GO analysis indicated that different genetic variants were associated with FNTA, IL-6, and COX15 candidate genes. FNTA is located on chromosome 8 and encodes the subunit alpha of the protein farnesyltransferase (FTase) enzyme (UniProtKB: P49354). It was found that FNTA could be a key gene for tumor progression; moreover, its abnormal copy numbers were associated with pathological transformations of breast cancer. Therefore, it could be considered as the main target of developing drugs [35]. Interleukin-6 (IL-6) as a cytokine released by various cells including cancerous cells contributed to the expansion and differentiation of tumor cells [36]. It was also shown that IL6ST may respectively act as a main factor and a tumor suppressor gene in TNBC progression, and diagnosis and treatment procedures [37]. Additionally, the IL6ST candidate gene was reported as a specific candidate gene for TNBC [38]. COX15 gene encodes cytochrome C Oxidase subunit 15 and contributes to the mitochondrial respiratory chain (UniProtKB: Q7KZN9). Gao et al. (2017) reported that the high-level expression of the COX5B gene was associated with a poor prognosis in breast cancer [39]. It was suggested that the level of COX5B protein may be related to the tumor size; also, its up-regulated form showed a worse disease-free survival. However, there was not enough evidence to illustrate the clinical implications of COX5B in breast cancer.

Conclusion

Results of differential genetic variants analysis between control and treated samples indicated that the most reported candidate genes had dual roles as oncogenes or tumor suppressors. Therefore, it was suggested that TAM could not have any significant role in an effective treatment through changing the genetic variants’ background.

Acknowledgments

Current study was financially supported by National Institute for Medical Research Development (NIMAD, grant number: 982563).

Conflict of Interest

The authors declare no conflict of interest.

References

4. Geisler J, Sasano H, Chen S, Purohit A. Ste-
21. Chang YW, Tseng CF, Wang MY, Chang WC, Lee CC, Chen LT, Su JL. Deacetylation of HSPA5 by HDAC6 leads to GP78-mediated HSPA5 ubiquitination at K447 and


