Ten Putative Dysregulated Genes and Potential Mechanism Pathways in Breast Cancer Cell Line Treated with Epigallogallocatechin Gallate from Prismatomeris Glabra

Latifa Hanim Shamshul Bahari, Ninie Nadia Zulkifli, Wan Rohani Wan Taib

School of Biomedical Science, Faculty of Health Sciences, Universiti Sultan Zainal Abidin, 21300, Kuala Nerus, Terengganu, Malaysia.

Corresponding author: wanrohani@unisza.edu.my

Received: 14th August 2023    Accepted: 12th December 2023    Published: 24th December 2023

Abstract
World Health Organization estimates that around 48,000 new instances of cancer were diagnosed in Malaysia every year and 17.3% of it is breast cancer. Some treatments such as surgery, radiation, chemotherapy, hormone therapy, targeted therapy, and immunotherapy are frequently used with their side effects. Therefore, using natural products from the *Prismatomeris* genus such as *P. glabra* to treat breast cancer is safer and has been the subject of much inquiry as it is believed to have anticancer substances. This study was conducted to clarify the molecular mechanisms underlying the action of the Epigallogallocatechin gallate (EGCG) identified from *P. glabra* on MCF-7 the cell line. The ten most significant associated expressed genes with their expression level were identified based on Transcriptomic Profile in EGCG-treated MCF-7. Based on these data, the mechanism of the signalling pathway and the role of the putative genes were evaluated using the Reactome database. Then, the protein interaction was determined using STRING software. The associated genes with their expression level and the mechanism of the signalling pathway were determined based on the p-value. The result from the Transcriptomic Profile in EGCG-treated MCF-7 showed differentially expressed aberrant genes with significant association (p-value < 0.05). The dysregulated expressed genes were classified according to upregulated and downregulated (volcano plot) genes including the *ASS1* (2.75E-23), *ID1* (9.70E-18), *LCP1* (3.20E-15), *TAGLN* (1.14E-14), *ELOVL6* (2.70E-13), *ID2* (1.20E-12), *DDIT3* (3.76E-12), *ACSL3* (5.05E-11), *SREBF1* (1.70E-09) and *SNORA109* (2.24E-08). The study revealed the molecular mechanism of differentially expressed genes in EGCG from *P. glabra* in treating breast cancer such as the MAPK signalling pathway and interaction among the proteins which provide insight into the potential usage of medicinal plants in developing personalized medicine for better treatment for the patient with breast cancer in ensuring better patient healthcare.

Keywords: Expressed Genes, Breast Cancer, Epigallogallocatechin Gallate

Introduction
The World Health Organization (WHO) estimates that 48,639 new instances of cancer were diagnosed in Malaysia every year, and that number is likely to quadruple by 2040 (World Health Organization, 2021). Men are more likely to develop lung, prostate, colorectal, stomach, and liver cancer than women, who are
more likely to develop breast, colorectal, lung, cervix, and thyroid cancer. However, breast cancer is the most prevalent malignancy in women in Malaysia.

The American Cancer Society states that the course of treatment for breast cancer will depend on the patient's kind, stage, and general health. Surgery, radiation, chemotherapy, hormone therapy, targeted therapy, and immunotherapy are frequently used with their side effects (Solowey et al., 2014). Given that breast cancer is characterized by a genetic mutation, gene therapy has been theorized to be beneficial against the disease. However, there are currently no methods or products for gene therapy that have been approved. Safety concerns exist when mutagenesis is introduced into healthy tissue (McCruden & McCarthy, 2014). Therefore, using natural products like herbs to treat breast cancer is safer. The importance of these products in cancer prevention and treatment is highlighted by mounting evidence that plant-derived chemicals may act as inhibitors of various phases of carcinogenesis and related inflammatory processes (Solowey et al., 2014).

Due to their widespread usage as traditional medicine by the local population, the Prismatomeris genus has been the subject of much inquiry, particularly by scientists in China and Thailand. Typically, Prismatomeris tetrandra, connata, and malayana are the species being researched. These species have antibacterial, antifungal, anticancer, and antimalarial properties. Locals and indigenous Malaysians have employed a different species, Prismatomeris glabra, for health, wellness, and hypertension (Mohd Tajuddin, 2016). P. glabra demonstrated that it could affect mice's non-spatial memory and lengthen their time spent swimming forcibly. Breast cancer cell lines for the P. glabra species have not yet been studied. This genus mostly contains compounds from the anthraquinone, iridoid, and pentacyclic triterpenoid groups. Since natural plant products have shown promise as anti-tumor and anti-cancer agents, these problems have prompted a multimodal and interdisciplinary approach that uses them to improve survivorship. Their effectiveness is also indicated by reports of decreased use toxicity and fewer recurrent resistances to anti-cancer medications that target hormones. The results of this project may provide insight of knowing the genetic pathways that may help us find herbs with fewer side effects to use in treating this disease and help identify target genes and their highly related pathways. By focusing on the differentially expressed genes, the data will offer potential biomarkers for personalised therapy.

Materials and Methods

Study design
The study design that used for this research was an experimental study design with the aim to unravel the molecular mechanism of differentially expressed genes in EGCG identified from P. glabra leaves using aqueous extract in treating breast cancer. Three RNA samples obtained from treated-MCF-7 cell line with EGCG and three RNA samples obtained from non-treated MCF-7 cell lines as control were sent for RNA sequencing for differentially expressed genes identification. The procedure of Phase 1 was already being processed by PhD's Student. After the P. glabra was extracted in ethanol extract, the bioactive compound identification from the extract was being done by using Liquid Chromatography – Mass Spectrometry (LC-MS). Then, the purchased compound which is EGCG was treated directly on MCF-7 as compared to non-treated cell lines. These samples were extracted for RNA before being sent to Neoscience Sdn Bhd for Transcriptomic Profile. Approximately 200 ng of total RNA was used for cDNA library construction with the MGI Easy RNA. The MGI DNBSEQ-G400 platform was used for RNA sequencing.

Data collection
Differently expressed genes were identified based on Transcriptomic Profile in EGCG-treated MCF-7 as compared to non-treated MCF-7. The top 10 most significant associated differentially expressed genes (DEGs) were chosen based on the p-adjusted value of the RNA sequencing output such as expression level according to the volcano plot and related pathways from generated KEGG terms with human genome as model provided by the bioinformation of the Neoscience Sdn Bhd. The mechanism of the signalling pathway
and role of the putative genes were evaluated using Reactome database and the protein interactions were determined by using in-silico analysis which is the STRING software.

Results

Transcriptomic Profile
The data portrayed by the Transcriptomic Profile for the ten differentially expressed aberrant genes. The results that were displayed were the significant ones with only adjusted p-value of less than 0.05. The highly significant associated expressed genes identification with the smallest adjusted p-value were chosen as shown in Table 1. A 0.05 p-value indicates that there will be false positives in 5% of the tests. A FDR adjusted p-value of 0.05 indicates that there will be false positives in 5% of significant tests. There will be fewer false positives with the latter. Adjusted p-value is chosen as the indication for the result as it is more accurate than raw p-value. In addition, the dysregulated expressed genes were classified according to upregulated and downregulated (volcano plot) genes. In a volcano plot (Figure 1), the genes that are most upregulated are on the right, the genes that are most downregulated are on the left, and the genes that are most statistically significant are at the top.

Table 1: Top 10 differentially significant expressed genes identification based on Transcriptomic Profile in EGCG-treated MCF-7 as compared to non-treated MCF-7

<table>
<thead>
<tr>
<th>REGULATION</th>
<th>GENES</th>
<th>P-VALUE</th>
<th>P-ADJUSTMENT</th>
<th>ENSEMBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downregulated</td>
<td>ASS1</td>
<td>2.63E-27</td>
<td>2.75E-23</td>
<td>ENSG00000130707</td>
</tr>
<tr>
<td>Downregulated</td>
<td>ID1</td>
<td>1.86E-21</td>
<td>9.70E-18</td>
<td>ENSG00000125968</td>
</tr>
<tr>
<td>Up-regulated</td>
<td>LCP1</td>
<td>9.19E-19</td>
<td>3.20E-15</td>
<td>ENSG00000136167</td>
</tr>
<tr>
<td>Up-regulated</td>
<td>TAGLN</td>
<td>4.36E-18</td>
<td>1.14E-14</td>
<td>ENSG00000149591</td>
</tr>
<tr>
<td>Downregulated</td>
<td>ELOVL6</td>
<td>1.29E-16</td>
<td>2.70E-13</td>
<td>ENSG00000170522</td>
</tr>
<tr>
<td>Downregulated</td>
<td>ID2</td>
<td>6.87E-16</td>
<td>1.20E-12</td>
<td>ENSG00000115738</td>
</tr>
<tr>
<td>Up-regulated</td>
<td>DDIT3</td>
<td>2.52E-15</td>
<td>3.76E-12</td>
<td>ENSG00000175197</td>
</tr>
<tr>
<td>Downregulated</td>
<td>ACSL3</td>
<td>3.86E-14</td>
<td>5.05E-11</td>
<td>ENSG00000123983</td>
</tr>
<tr>
<td>Downregulated</td>
<td>SREBF1</td>
<td>1.46E-12</td>
<td>1.70E-09</td>
<td>ENSG00000072310</td>
</tr>
<tr>
<td>Up-regulated</td>
<td>SNORA109</td>
<td>2.14E-11</td>
<td>2.24E-08</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Figure 1: The volcano plot for the indication of upregulated (green) and downregulated (red) expressed genes.
Reactome Pathway

The data portrayed by the Reactome Pathway Database for the potential mechanism for the expressed genes was based on KEGG comparison between EGCG treated and non-treated breast cancer cell lines. The top 10 KEGG categories potential mechanism for upregulated genes was analysed based on the adjusted p-value or False Discovery Rate (FDR) of less than 0.05 and the set size from the GSEA KEGG chart (Figure 2). The pathways that were highly significant exerted the minimum p-value and FDR value were chosen. When compared to the raw p-value, it had clean data. Therefore, it could be selected largely based on an entity’s FDR value. Thus, the FDR value provided greater assurance in addition to the p-value, which was important. The outcome is more highly significant when the entities’ p-value and FDR value are smaller. The label GeneRatio (as indicated in the x-axis) is the ratio of the number of differentially expressed genes annotated in a pathway (as indicated in the y-axis) to the number of all genes annotated in this pathway. The highly significant potential mechanism with the smallest adjusted p-value or FDR and bigger set size were chosen.

Table 2: Top 3 highly potential mechanisms from Reactome Pathway website.

<table>
<thead>
<tr>
<th>POTENTIAL MECHANISMS</th>
<th>SETSIZE</th>
<th>P-VALUE</th>
<th>P-ADJUST</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPK signalling pathway</td>
<td>283</td>
<td>6.65E-05</td>
<td>6.65E-05</td>
</tr>
<tr>
<td>Prion disease</td>
<td>236</td>
<td>0.00047</td>
<td>0.00047</td>
</tr>
<tr>
<td>Parkinson disease</td>
<td>235</td>
<td>3.18E-06</td>
<td>3.18E-06</td>
</tr>
</tbody>
</table>

Figure 2: The GSEA KEGG chart

Other than that, based on the KEGG chart shown in Figure 3, it indicates the pathways with the largest number of upregulated genes expressed differentially in Treated and Control sample. Based on this KEGG chart, the pathway with the largest number of upregulated genes expressed differentially in Treated and Control is being analysed by the gene number (as indicated in the x-axis).
Table 3: Top 3 upregulated genes expressed differentially in treated-MCF-7 cell lines

<table>
<thead>
<tr>
<th>Pathways of upregulated genes expressed differentially in treated-MCF-7 cell lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 P13K-Akt signalling pathways</td>
</tr>
<tr>
<td>2 Human Papilloma Virus Infection</td>
</tr>
<tr>
<td>3 MAPK signalling pathway</td>
</tr>
</tbody>
</table>

Figure 3: The KEGG chart of upregulated genes expressed differentially in Treated and Control sample.

**STRING Analysis**

The STRING database provided findings that included information on the interactions or connections between genes. In the displayed result, the edges or coloured lines denoted protein-protein relationships or interactions with other proteins, while the network nodes represented proteins. Each node in the output
was a representation of all the proteins made by a particular locus encoding a protein. The coloured nodes in the data indicated that the proteins being studied were the first shell interactors, which were the query proteins. The white nodes that were not included in the results, however, were identified as second shell interactors. The outcome also displayed whether the nodes were filled or vacant. While the empty nodes indicated that the proteins had an unknown 3D structure, the filled nodes, which were the nodes in the result, suggested that the proteins had known or predicted 3D structure. Additionally, it was intended for the linkages of the edges to be precise and significant. It was not necessary for the proteins to be physically attached to one another if they cooperatively contributed to a shared function. The edges with light blue and pink colours, which were taken from curated databases and experimentally determined, respectively, were the edges of known interactions. Gene co-occurrence, gene fusions, and gene neighbourhood were represented by the coloured edges of expected interactions in green, red, and blue, respectively. Other interactions included text mining, co-expression, and protein homology, each represented by the colours light green, black, and grey.

Figure 4: The result extracted from STRING website.
Discussion

Analysis of significantly associated expressed genes

In this study, ten highly significant associated expressed genes identification were chosen which include the ASS1, ID1, LCP1, TAGLN, ELOVL6, ID2, DDIT3, ACSL3, SREBF1 and SNORA109 obtained from the Transcriptomic Profile database. ASS1 gene is belongs to the Type 1 subfamily of argininosuccinate synthase family. In breast cancer, ASS1 can be found in low abundance in some breast tumours. Although the formation of these tumours is accelerated by ASS1 loss, these cells are vulnerable to exogenous arginine deprivation (Fuming Qiu et al.,2014). As for ID2, as stated by Yin Liu et al. (2019), it was discovered that ID2 increased the Ductal Carcinoma In Situ (DCIS) formation by increasing the cancer stemness of premalignant cells, identifying it as a crucial gene for DCIS initiation. Additionally, ID2 is essential for the survival of cancer cells that are aggressive. ELOVL6 gene, which controls the metabolism of intracellular lipid components, has been linked to the onset and spread of breast cancer (Li. H et al.,2019). Next, high aggressiveness and metastasis are linked to high ID1 expression in breast cancer cell lines, which is a major contributor to breast cancer patient mortality. The TAGLN gene has been recognised for the first time as a target of DNA hypermethylation in breast cancer as it was often and severely downregulated via promoter DNA hypermethylation in breast cancer cells (Sayar.N et al.,2015). In breast cancer tissues, SREBF1 or SREBP-1 gene mRNA and protein levels are increased when compared to precancerous tissues, following the pattern in other malignancies. SREBP-1 expression levels are positively linked with lymph node metastasis, metastatic stage, and tumour differentiation (Zhao Q. et al.,2022). ACSL3 has been demonstrated to be downregulated in triple-negative breast cancer, it has also been discovered to be elevated in women with ER-negative breast cancer. In fact, ACSL3 interacts with the pro-metastatic protein CUB domain-containing protein 1 (CDCP1) in triple-negative breast cancer. As for LCP1, it is interesting to note that the dual inhibition of ABCE1 and LCP1 had an additive effect on the invasion, migration, and proliferation of cancer cells. In addition, compared to the knockdown of either genes alone, the combination of ABCE1 and LCP1 significantly inhibited tumour development, reduced metastatic activity, and improved survival (Pillar N. et al.,2019). However, there is lack of study of DDIT3 related with breast cancer cells meanwhile, as for SNORA109 gene, there is lack of study about SNORA109 gene thus, makes the information about this gene is limited.

Analysis of significant pathways related with breast cancer

In this study, three highly significant pathways were chosen which were MAPK signalling pathway (283, 6.65E-05), Prion disease (236, 0.00047) and Parkinson disease (235, 3.18E-06) along with their respective set size and adjusted p-value. Thus, those three pathways were deemed as significant and highly associated with respective putative genes. Other than that, the pathway with the largest number of upregulated genes expressed differentially in Treated and Control is being analysed by the gene number. Therefore, the top 3 pathways with the largest number of upregulated genes expressed differentially in the Treated-MCF-7 cell lines were p13K-Akt signalling pathways, Human Papilloma Virus Infection, and MAPK signalling pathway. According to Jiang, W. et al., 2020, MAPK signalling pathway is an essential signalling mechanism implicated in the emergence of Triple Negative Breast Cancer (TNBC). Invasion, metastasis, and the prognosis of TNBC are all tightly tied to MAPKs, which are crucial in the production of ER, PR, and HER-2. Higher MAPK activity may indicate a worse prognosis because it is linked to lower survival times in TNBC patients. According to Yu. G. et al.,2012, excessive PrP expression in MCF-7 breast cancer cells prevents TNF- or Bax-induced cell death. When PrP is silenced, drug-resistant MCF-7 sublines become more susceptible to the apoptotic effects of paclitaxel or the TRAIL chemotherapy. In addition, PrP-expressing ER-negative breast tumours seem to react poorly to adjuvant chemotherapy which is chemotherapy administered after surgery. PrP has been overexpressed in breast cancer cells or silenced drug-resistant cell sublines in research using breast cancer cell models up to this point. In earlier investigations, it was discovered that women with PD had a higher death rate and an increased risk of breast cancer. However, despite the lack of research into the precise shared molecular processes, certain investigations have suggested that oestrogen is neuroprotective and offers neuroprotection against PD. Additionally, mutations linked to numerous genes,
such as ATM (ataxia telangiectasia mutated), PARKIN, and tumour suppressors are found in oestrogen and progesterone receptor-negative (ER/PR) breast cancers, as are elevated transcript levels of the genes linked to neurodegeneration, such as Seladin-1, APP, and PSEN1 (D. Advani et al., 2022). The normal cellular activities involved in cell growth, proliferation, metabolism, motility, survival, and death are significantly regulated by the PI3K/AKT signalling system. In many types of human cancer, abnormal PI3K/AKT pathway activation encourages tumour cell survival and growth. The pathway activation is most frequently seen in a variety of human malignancies, including breast, lung, ovarian, and prostate cancers. This pathway increased activity is frequently linked to the development of tumours and resistance to cancer treatments. According to a considerable number of recent research, high-risk HPV subtypes, particularly HPV subtypes 16, 18, or 33, were detected in human breast cancer tissues in about 29% of cases. Contrarily, several other analyses found no HPV subtypes in breast cancer tissue or normal breast tissue from breast cancer patients (Wang T. et al., 2012).

**Analysis of gene networking**

The STRING database was used to determine the gene-gene interactions of the ten putative genes. The STRING database findings displayed the interactions in accordance with the colour of the edges, interactions with which genes, and overall number of interactions. First off, for the known interactions that came from the channel that was determined through experiment, only pink edges were in the result. This pink edge, as defined by Damian S. et al. (2021), indicates experimentally determined gene-gene interaction, based on data from assays and laboratory studies such biochemical, genetic, and biophysical experiments. The most gene-gene interactions were from SREBF1 gene, followed by ACSL3 and TAGLN, with the least interactions. This suggested that SREBF1 gene had the most gene-gene interactions in terms of evidence from assays and experiments done in the laboratory, in which it interacted with ID2 and DDIT3 genes. In breast cancer tissues, SREBF1 or SREBP-1 gene mRNA and protein levels are increased when compared to precancerous tissues, following the pattern in other malignancies. According to Zhao Q. et al., 2022, SREBP-1 expression levels are positively linked with lymph node metastasis, metastatic stage, and tumour differentiation. As for blue edges were in the result for the predicted interactions which were gene co-occurrence channel. This co-occurrence gene-gene interaction only had two interactions which was between ACSL3 and SREBF1 genes which both interacted or associated with ELOVL6 gene. Damian S. et al. (2021) claim that this gene co-occurrence channel was calculated using whole-genome comparisons and the availability of a fully sequenced genome. In addition, Heiko M. and Francesco M. (2008) claim that co-occurrence of two genes can arise when they are dependent on one another and do not function when they are isolated, either directly as binding partners or as catalysts necessary in a certain metabolic pathway. So, this demonstrates and supports the STRING database result, which showed that the ACSL3 and SREBF1 genes, which both interacted with or were associated with the ELOVL6 gene, had co-occurrence gene-gene interactions, indicating that they had shared functions and may have been binding partners or catalysts in a particular metabolic pathway. Next, the green edges were in the result for the others interaction which were textmining channel. The most textmining gene-gene interactions were SREBF1 gene with three interactions, followed by TAGLN, and ACSL3 genes with one interaction each. According to Damian S. et al. (2021). The textmining channel is based on publishing of papers and literatures in which genes are cited in the same sentence, paragraph, or articles. Other than that, black edges were in the result for the others interaction which were co-expression channel. The co-expression gene-gene interactions were shown from ACSL3 and TAGLN genes have one co-expression interactions each. Damian S. et al. (2021) claim that evidence of gene expressions was gathered from multiple sources, normalised, and then the redundant or unneeded data was removed before the expression profiles were compared under various circumstances. Then, co-expression was defined as the constant similarity between the expression profiles of two genes. Finally, among of all the ten putative genes listed, ASS1 and ID1 genes does not displayed any edges or coloured lines denoted protein-protein relationships or interactions with other proteins, meanwhile the SNORA109 gene cannot be detected by STRING website.
Conclusion
In conclusion, this study had unravel the molecular mechanism of differentially expressed genes in EGCG from P. glabra in treating breast cancer. Based on the significant association of the genes in the pathways and the significant gene networking among the ten putative genes, this study discovered that the ten putative genes (ASS1, ID1, LCP1, TAGLN, ELOVL6, ID2, DDIT3, ACSL3, SREBF1, and SNORA109) were strongly and significantly associated expressed genes with the likelihood or disease progression of breast cancer. As these targeted genes are useful in providing the necessary information needed which serve as a baseline of personalized medicine for further investigations such as the precise loci of the biomarkers integrated and analyzed from the information of the significant biological mechanism, and gene networking, these targeted genes have the potential to be used as biomarkers in plans in the development of personalized medicine for breast cancer. Therefore, to fully understand the accurate and detailed information of the putative genes to develop personalized medicine as a better treatment for breast cancer patients and ensure better efficacy of the treatment and patient healthcare, the limitations of this study must be overcome with the suggestions mentioned.

Funding
This study was supported by Labmat Grant by Universiti Sultan Zainal Abidin (ref: UniSZA/LABMAT/2018/07)

Acknowledgements
This study was supported by Labmat Grant by Universiti Sultan Zainal Abidin (ref: UniSZA/LABMAT/2018/07)

Conflict of Interest Disclosure
None to declare

References


