Drug–target–disease network analysis of gene–phenotype connectivity for genistein in ovarian cancer

Chen Zhang1,*
Fan Yang2,*
Suiqin Ni1
Wenbing Teng1
Yingxia Ning3

1Department of Clinical Pharmacy, Guangzhou First People’s Hospital, Guangzhou Medical University, Guangzhou 510180, China; 2Department of Health Examination, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China; 3Department of Gynaecology and Obstetrics, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China

*These authors contributed equally to this work

Purpose: Genistein belongs to the group of isoflavones, which include powerful anticancer agents. Its antitumor properties have been intensively described in many cancers, but related studies assessing ovarian cancer are scarce. The aim of this study was to develop a new method of the underlying mechanisms of genistein’s effects and broaden the perspective of targeted therapies in ovarian carcinoma.

Materials and methods: Genistein targets were searched in the DrugBank database. Prediction of drug interactions with targets (including secondary targets) was performed with STRING database. Interaction pairs with overall score above 0.9 were recorded for protein–protein interaction (PPI) network generation based on the Cytoscape software. Genes with intense interconnections were grouped into a module. Then, PPI network modules with significance were assessed using Molecular Complex Detection (MCODE) analysis tool. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed for the critical genes. Furthermore, disease targets were searched in Comparative Toxicogenomics Database (CTD). The overlapping targets were studied using a Kaplan–Meier analysis to evaluate ovarian carcinoma survival.

Results: A total of 13 direct targets and 372 secondary targets were identified for genistein and further analyzed with the MCODE analysis tool to identify critical genes. The top 72 genes were further assessed with KEGG. Then, the term “ovarian cancer” was searched in CTD, and 123 genes associated only with the marker “T” or “M” were recorded. Next, seven overlapping genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3C, and AKT1) resulting from the intersection of three pathways and 123 genes were obtained from CTD. Elevated CDKN1B amounts showed correlation with overall survival (log-rank $P=0.021$) according to Kaplan–Meier analysis.

Conclusion: The current findings indicated that drug–target–disease network analysis represents a useful tool in gene–phenotype connectivity for genistein in ovarian cancer. Our result also showed that CDKN1B is worthy of further research.

Keywords: protein–protein interaction, PPI, DrugBank, Comparative Toxicogenomics Database, CTD, CDKN1B, PI3K/AKT signaling pathway, FoxO signaling pathway

Introduction

Ovarian cancer has the highest mortality rate among gynecological malignancies in industrialized countries. With current protocols, a 70%–80% response to first-line chemotherapy can be achieved; however, most cases relapse and eventually succumb after metastasis.1 Successful ovarian cancer therapy depends greatly upon the effectiveness of cytotoxic anticancer drugs. Resistance to chemotherapeutics is
associated with ovarian cancer relapse. However, the related mechanisms remain largely undefined.

Genistein (4’,5,7-trihydroxyisoflavone) naturally occurs in fruits, nuts, and soybeans and potently inhibits cancers, including breast, prostate, liver, ovarian, bladder, gastric, and brain cancers as well as neuroblastomas and chronic lymphocytic leukemia.2 Genistein acts as a chemotherapeutic agent against different types of cancer, by inhibiting inflammation, promoting apoptosis, and regulating steroid hormone receptors and metabolic pathways.3–4 The antitumor properties of genistein have been intensively described in many cancers, but related studies evaluating ovarian cancer are scarce.

We have previously confirmed that a genistein analog can suppress PI3K/AKT signaling5 and the FoxO signaling pathway.6–8 With advances in genomics, network analysis could help dissect multiple human diseases. In this study, DrugBank database9 was used to broadly assess genistein and drug–target data. Then, the prediction of drug–target interactions was performed with STRING database.10 Protein–protein interaction (PPI) network modules with significance were assessed using the Molecular Complex Detection (MCODE) software.11 Then, associated genes were obtained for further exploration of genomic changes with the cBio Cancer Genomics Portal (cBioPortal) database.12,13 The Kaplan–Meier method14 was employed for survival assessment. Our findings demonstrated that drug–target–disease network (DTDN) analysis is valuable in exploring the mechanisms associated with genistein’s effects on ovarian cancer, broadening the perspective of targeted therapies.

Materials and methods
Drug–target search
The DrugBank database is a comprehensive, freely accessible, online database containing information on drugs and drug targets. As DrugBank is both a bioinformatics and a cheminformatics resource, it combines detailed drug (ie, chemical, pharmacological, and pharmaceutical) data with comprehensive drug target (ie, sequence, structure, and pathway) information.9 DrugBank (https://www.drugbank.ca/) was employed to identify interactions between genistein and its target molecules. Genistein was searched as a keyword under drug classification entry.

PPI network generation and module assessment
STRING is a database of known and predicted PPIs.10 The STRING resource is available online at http://string-db.org/. Prediction of drug interactions with targets (including secondary targets) was performed with STRING. Interaction pairs with overall score above 0.9 were recorded for PPI network generation based on the Cytoscape software (http://www.cytoscape.org/).15 Genes with intense interconnections were grouped into a module. Then, PPI network modules with significance were assessed using MCODE (http://baderlab.org/Software/MCODE).

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and overlapping genes of ovarian cancer
WebGestalt16 is a comprehensive web-based integrated data-mining system that provides the maximum flexibility for functional enrichment analyses. Biochemical pathways and functions related to genistein gene set were specifically queried with a KEGG pathway17 enrichment analysis in WebGestalt. Top 10 pathways with an adjusted P-value of <0.01 were selected. To identify disease targets for ovarian cancer, we searched “ovarian cancer” in Comparative Toxicogenomics Database (CTD, revision 15462).18 Genes of some pathways and targets of ovarian cancer were the selected overlapping genes for further analysis.

Exploring cancer genomics data linked to genistein
cBioPortal for cancer genomics is an open-access, open-source resource for interactively exploring multidimensional cancer genomics data sets.12,13 It was employed to assess the connectivity of genistein-related genes in all ovarian cancer reports. Using the portal search function, genistein-related genes in ovarian cancer publications were evaluated for genomic alterations across a set of patients, performing network analysis and identifying trends, eg, mutual exclusivity or co-occurrence between gene pairs of the same gene set.

Analysis of genistein-associated genes and ovarian cancer survival
The Kaplan–Meier method14 (http://kmplot.com/) was employed to evaluate the effects of 54,675 genes on survival in 10,461 cancer cases, including 1,816 ovarian cancer cases with mean follow-up periods of 69, 40, 49, and 33 months. Altered genes were then assessed by the Kaplan–Meier method for evaluating ovarian cancer survival according to gene expression.

Results
Characterization of genistein using DrugBank
DrugBank resources (version 5.1.0; released April 2, 2018) were used to identify the bioactivities of genistein and
targeted genes. This resulted in accession number DB01645 (EXPT01582) categorizing genistein among anticarcinogenic agents, enzyme inhibitors, phytoestrogens, and isoflavones. To date, genistein is under clinical investigation for prostate cancer and might kill cancer cells via inactivation of cell growth-associated proteins. Table 1 summarizes the 13 targets of genistein, including ESR2, TOP2A, PTK2B, NCOA1, ESR1, NCOA2, ESRRA, ESRRB, NR1I2, AKT1, GPER1, CYP1B1, and SHBG.

**PPIs identified with the MCODE analysis tool**

Proteins encoded by the 13 target genes associated with genistein were considered primary protein targets. A total of 372 target–protein interactions were detected using STRING and further analyzed with the MCODE analysis tool to determine critical genes. The top 72 genes are shown in Figure 1.

**Pathway enrichment of hub genes using KEGG analysis**

Hub genes were found in 80 molecular pathways in KEGG (WebGestalt) enrichment. The first 10 KEGG hits included pathways in cancer (48 genes), PI3K/AKT signaling pathway (36 genes), hepatitis B and prostate cancer (30 genes), Kaposi’s sarcoma-associated herpesvirus infection (28 genes), HTLV-I infection (27 genes), breast cancer, human papillomavirus infection and breast cancer (26 genes), and FoxO signaling pathway and proteoglycans in cancer (25 genes). The obtained KEGG enrichment pathways reflected functional features of genistein gene sets, and further assessment was performed. We have previously shown that a genistein analog can suppress PI3K/AKT signaling as well as the FoxO signaling pathway. To identify disease targets for ovarian cancer, we searched “ovarian cancer” in CTD (revision 15462). A total of 123 genes with curated associations to the disease with the marker “T” or “M” only were recorded. “M” means a gene that may be a biomarker of a disease or play a role in the etiology of a disease. “T” means a gene that is or may be a therapeutic target in the treatment of a disease. Seven overlapping genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1) resulted from the intersections of the three pathways alongside the 123 abovementioned genes were visualized using Venn diagrams (Figure 2). Further analysis of the seven selected overlapping genes was performed.

**Genomic alterations of genistein-associated genes in ovarian cancer**

Only two ovarian carcinoma studies were included in cBioPortal (Version 1.13.1). The seven genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1) associated with the ovarian cancer pathway were searched, and two reports were involved. The results showed that 107 (33%) of the 327 sequenced cases had alterations in one or more of these genes; alteration frequencies are shown in Figure 3. PTEN (8%) alterations mostly included deep deletions. The seven genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1) had 16 gene pairs showing mutually exclusive alterations (with no statistical significance), while five had concurrent alterations (with no statistical significance).

Next, cBioPortal was employed for interactive analysis and to generate networks of genes showing alterations in cancer; the neighbors of the abovementioned seven genes are shown in Figure 4. Only neighbors with high alteration frequencies are shown in Figure 4. The seven genes were found to be related to TP53, and a filter of $\geq 64.1\%$ alteration frequency was used. Meanwhile, nine genes, including TP53 and MYC, were identified with a filter of $\geq 32.6\%$ alteration frequency.

| No | Targets | Gene name | General function | Uniprot ID |
|----|---------|-----------|------------------|------------|
| 1  | Estrogen receptor beta | ESR2 | Zinc ion binding | Q97231 |
| 2  | DNA topoisomerase 2-alpha | TOP2A | Ubiquitin binding | P11388 |
| 3  | Protein tyrosine kinase 2-beta | PTK2B | Signal transducer activity | Q14289 |
| 4  | Nuclear receptor coactivator 1 | NCOA1 | Transcription coactivator activity | Q15788 |
| 5  | Estrogen receptor alpha | ESR1 | Zinc ion binding | P03372 |
| 6  | Nuclear receptor coactivator 2 | NCOA2 | Zinc ion binding | Q15596 |
| 7  | Steroid hormone receptor ERR1 | ESRR | Zinc ion binding | P11474 |
| 8  | Steroid hormone receptor ERR2 | ESRRB | Zinc ion binding | Q95718 |
| 9  | Nuclear receptor subfamily 1 group I member 2 | NR1I2 | Zinc ion binding | O75469 |
| 10 | RAC-alpha serine/threonine-protein kinase | AKT1 | Protein serine/threonine/tyrosine kinase activity | P31749 |
| 11 | G-protein-coupled estrogen receptor 1 | GPER1 | Steroid hormone binding | Q99527 |
| 12 | Cytochrome P450 1B1 | CYP1B1 | Oxygen binding | Q16678 |
| 13 | Sex hormone-binding globulin | SHBG | Androgen binding | P04278 |
frequency. Eleven gene clusters, among which TP53, MYC, AGO2, and PTK2, were identified with a 27.6% alteration frequency as cutoff; 12 genes, including TP53, MYC, AGO2, PTK2, and MAFA, were obtained at 26.8%.

Genistein-associated genes and survival in ovarian cancer
The seven selected genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1) identified were employed to perform survival analysis with clinical profiles in ovarian cancer. As shown in Figure 5, elevated CDKN1B amounts showed correlation with overall survival (log-rank $P=0.021$) according to Kaplan–Meier analysis, with cases grouped by mean mRNA levels.

Discussion
A PubMed search with “genistein and cancer” returned over 2,963 publications where Akiyama et al.\textsuperscript{19} reported that genistein inhibits the EGF receptor. Multiple biologically relevant effects have been described for genistein. For example, genistein is known to inhibit Glut receptors,\textsuperscript{20} which is overexpressed in cancer cells, which may be the probable reason for its anticancer activity.\textsuperscript{21} However, the mechanisms by which genistein exerts these beneficial effects are not fully understood. Therefore, novel methods or platforms that could...
bridge genistein to its targets are required for the evaluation of biological effect.

The present study used the functional/activity network (FAN) analysis and performed functional network analysis with multiple web-based tools. Using a system biochemistry approach integrating DrugBank and STRING and WebGestalt, associations of ovarian cancer and molecules with drug targets were globally detected. A total of 13 primary and 372 secondary target genes/proteins (Supplementary materials) were obtained. The top 72 genes in MCODE and 10 enriched KEGG pathways related to genes altered by genistein are summarized in Table 2. To obtain disease targets for ovarian cancer, we searched “ovarian cancer” in CTD and identified 123 genes showing associations only with the marker “T” or “M” only; in addition, seven overlapping genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1) were found in the intersection of the three pathways and the 123 genes in CTD.

![Figure 3](image.png)

**Figure 3** Alteration frequencies of seven genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1) in cBio cancer genomics portal. **Abbreviations:** TCGA, The Cancer Genome Atlas; MSKCC, Memorial Sloan Kettering Cancer Center.

![Figure 4](image.png)

**Figure 4** Interactive analysis to generate networks of genes showing alterations in cancer. **Notes:** (A) Network. (B) Network (26.8% + TP53 + MYC + AGO2 + PTK2 + Masf). (C) Network (27.6% + TP53 + MYC + AGO2 + PTK2). (D) Network (32.6% + TP53 + MYC). (E) Network (64.1% + TP53).
Figure 5 Survival analysis of seven selected genes (CDKN1B, PTEN, EGFR, MAPK1, MAPK3, PIK3CA, and AKT1): mRNA expression in ovarian cancer (A-G).

Notes: (A) CDKN1B (overall survival). (B) AKT1. (C) PIK3CA. (D) MAPK1. (E) MAPK3. (F) PTEN. (G) EGFR.
We previously demonstrated that a genistein analog significantly suppresses tumor stemness and promotes ovarian cancer cell apoptosis through FOXM1 downregulation and FOXO3a induction, \(^6\)\(^-\)\(^8\) suppressing PI3K/AKT signaling. \(^5\) Meanwhile, Li et al. \(^23\) identified the highly mutated super enhancer CDKN1B with concordant gene expression changes in their target genes and prognostic significance for ovarian cancer. In this study, high mRNA CDKN1B amounts showed correlation with overall survival in ovarian cancer. Guo et al. \(^24\) reported that G6PC silencing induces cell cycle-associated proteins and restores CDKN1B expression. Peng et al. \(^25\) described the small natural compound goniothalamin (GTN) as a CKI inducer. GTN stabilizes CDKN1B protein expression by degrading its specific E3 ubiquitin ligase (S-phase kinase-associated protein 2). Therefore, the mechanisms between genistein and CDKN1B protein expression should be studied in in vitro experiments.

**Conclusion**

We reported in this study a new method, DTDN analysis, to investigate the underlying mechanisms of genistein action and broaden the perspective of targeted therapies in ovarian carcinoma. We employed DTDN method to analyze DrugBank, STRING, and WebGestalt databases and found that MAPK, PIK3CA, and AKT1 are associated with ovarian cancer, which agrees with our experimental study result reported. Furthermore, using the reported method, we observed a correlation between high expression of mRNA CDKN1B and low overall survival rate in ovarian cancer, indicating a new gene target worthy of further in vitro experimental study. Our work showed that DTDN analysis could facilitate the interpretation of the disease mechanisms of ovarian cancer and the identification of new targets, hence could help accelerate the investigation of genistein’s anticancer effects by screening potential gene targets to be studied in in vitro experiments.
Ning Y, Li Q, Xiang H, Liu F, Cao J. Apoptosis induced by 7-fluoromethoxyl-5,4-di-octyl genistein in ovarian cancer cells. *Gynecol Oncol*. 2007;105(1):23–30.

Ning Y, Luo C, Ren K, Quan M, Cao J. FOXO3a-mediated suppression of the self-renewal capacity of sphere-forming cells derived from the ovarian cancer SKOV3 cell line by 7-fluoromethoxyl-5,4-di-octyl genistein. *Mol Med Rep*. 2014;9(5):1982–1988.

Ning Y, Li Q, Xiang H, Liu F, Cao J. Apoptosis induced by 7-fluoromethoxyl-5,4-di-octyl genistein via the inactivation of FoxM1 in ovarian cancer cells. *Oncol Rep*. 2012;27(6):1857–1864.

Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074–D1082.

Szkarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):D447–D452.

Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2.