Taiwanese Journal of Obstetrics & Gynecology 56 (2017) 210–216

Contents lists available at ScienceDirect

Taiwanese Journal of Obstetrics & Gynecology

journal homepage: www.tjog-online.com

Original Article

Gene set-based analysis of mucinous ovarian carcinoma

Chia-Ming Chang [a, b, c], Peng-Hui Wang [a, c, d, e], Huann-Cheng Horng [a, c, *]

[a] Department of Obstetrics and Gynecology, Taipei Veterans General Hospital, Taipei, Taiwan
[b] Institute of Oral Biology, National Yang-Ming University, Taipei, Taiwan
[c] Department of Obstetrics and Gynecology, National Yang-Ming University, Taipei, Taiwan
[d] Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan
[e] Department of Medical Research, China Medical University Hospital, Taichung, Taiwan

ARTICLE INFO

Article history:
Accepted 27 December 2016

Keywords:
function
gene expression
mucinous ovarian carcinoma

ABSTRACT

Objective: Mucinous ovarian carcinoma (MOC) is an uncommon subtype of epithelial ovarian cancers, and the pathogenesis is still poorly understood because of its rarity. We conducted a gene set-based analysis to investigate the pathogenesis of MOC by integrating microarray gene expression datasets based on the regularity of functions defined by gene ontology or canonical pathway databases.

Materials and methods: Forty-five pairs of MOC and normal ovarian tissue sample gene expression profiles were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus database. The gene expression profiles were converted to the gene set regularity indexes by measuring the change of gene expression ordering in a gene set. Then the pathogenesis of MOC was investigated with the differences of function regularity with the gene set regularity indexes between the MOC and normal control samples.

Results: The informativeness of the gene set regularity indexes was sufficient for machine learning to accurately recognize and classify the functional regulation patterns with an accuracy of 99.44%. The statistical analysis revealed that the GTPase regulators and receptor tyrosine kinase erbB-2 (ERBB2) were the most important aberrations; the exploratory factor analysis revealed phosphoinositide 3-kinase-activating kinase, G-protein coupled receptor pathway, oxidoreductase activity, immune response, peptidase activity, regulation of translation, and transport and channel activity were also involved in the pathogenesis of MOC.

Conclusion: Investigating the pathogenesis of MOC with the functionome provided a comprehensive view of the deregulated functions of this disease. In addition to GTPase regulators and ERBB2, a plenty of deregulated functions such as phosphoinositide 3-kinase, G-protein coupled receptor pathway, and immune response also participated in the interaction network of MOC pathogenesis.

Introduction

Primary mucinous ovarian carcinoma (MOC) is an uncommon subtype of epithelial ovarian cancers, accounting for 3–4% of all ovarian carcinomas [1,2]. Currently, the carcinogenesis of MOC is still poorly understood because of its rarity. Two genetic aberrations, KRAS and receptor tyrosine kinase erbB-2 (ERBB2), are known to be involved in the pathogenesis of MOC [3]; besides, knowledge about the function regulation and pathogenesis of this cancer is limited. Microarray gene expression [4–6] is the primary tool of investigation of the pathogenesis of complex diseases such as mucinous ovarian cancer. To further understand the pathogenesis of MOC, we conducted an integrative analysis of mucinous ovarian cancer with the microarray gene expression datasets downloaded from publicly available databases.

The workflow of this gene set-based analysis was introduced before [7–10]. In brief, it consisted of two steps. First, microarray gene expressions were converted to a gene set regularity (GSR) index by computing the expression ordering change among genes in a given gene set defined by the gene ontology (GO) or the canonical pathway database downloaded from the Molecular Signature Database [11]. Each gene set contains a group of genes, defining biological process, molecular function, or cellular component; for simplification, we called them “function” in this

* Corresponding author. Department of Obstetrics and Gynecology, Taipei Veterans General Hospital and National Yang-Ming University, 201, Section 2, Shih-Pai Road, Taipei 112, Taiwan.
E-mail address: hchorng@vghtpe.gov.tw (H.-C. Horng).

http://dx.doi.org/10.1016/j.tjog.2016.12.016

study. By measuring the change of gene expression level orderings between cancerous and normal states, regularity of the function defined by that gene set could be quantified. In the second step, the pathogenesis of MOC was investigated with the 1454 GO term- or 1330 canonical pathway-defined functions. We utilized exploratory factor analysis (EFA) to discover important deregulated functions and the interaction network involved in the pathogenesis of MOC.

Materials and methods

Microarray datasets, gene set definition, and data processing

We downloaded gene expression microarray datasets in the SOFT format from the National Center for Biotechnology Information Gene Expression Omnibus database. MOC and normal ovarian tissue were used for comparison. The common genes among all the datasets and the associated gene expression data were included. Datasets and gene sets were discarded if the number of the common genes and gene elements was less than 8000 and 3, respectively.

Computing GSR indexes

Figure 1 shows the workflow for computing the GSR indexes, which was modified from the Differential Rank Conservation [12]. The Differential Rank Conservation is designed to measure the perturbation of a given gene set by quantifying the change of gene expression orderings of gene elements in that gene set. Instead of measuring the perturbation, the GSR index quantifies the change of gene expression orderings between two phenotypes in a gene set. For this purpose, the GSR indexes of the MOC and normal ovarian tissue groups were computed by comparing the sample's gene expression orderings with the baseline gene set ordering template, defined as the most common gene expression ordering in a gene set among all the normal ovarian tissue samples. Subsequent analyses of MOC and normal ovarian tissue GSR indexes were carried out based on this baseline. The baseline gene set ordering template for each gene set was established by pairwise comparison between the expression levels of two genes for all possible combinations of gene pair. Establishment of the baseline gene set expression ordering template and computation of the GSR indexes were executed in R environment. The code and test datasets can be obtained from the GitHub (https://github.com/carlzang/GSR-model.git).

Statistical analysis

We used Mann–Whitney *U* test to test the differences between the MOC and control GSR indexes, and the data were corrected by

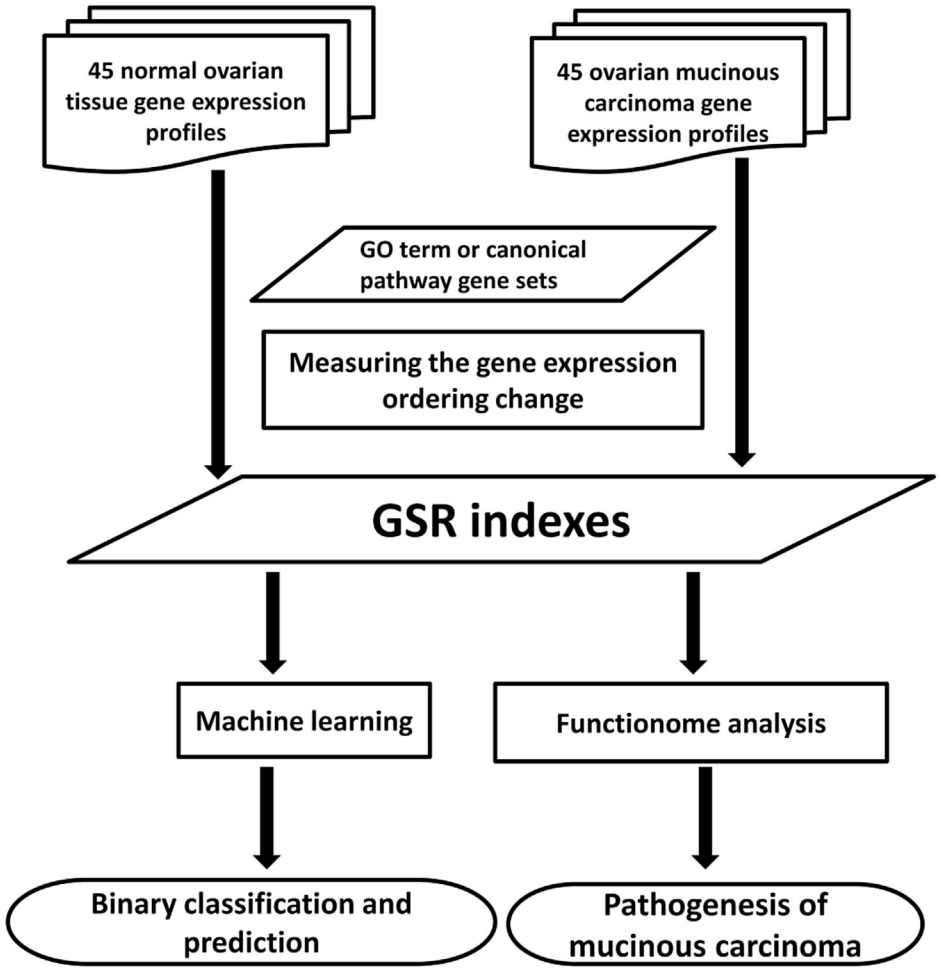

Figure 1. Workflow of the GSR model. The MOC or control GSR index was computed by converting the gene expression orderings of MOC or normal ovarian control sample through each GO term or canonical pathway gene set. Machine learning algorithm was trained to recognize the patterns consisting of the GSR indexes and then execute the binary (case–control) classifications. Functionome analysis was carried out to investigate the pathogenesis of MOC. GO = gene ontology; GSR = gene set regularity; MOC = mucinous ovarian carcinoma.