

Algorithm for identifying cancer gene expression signatures

Published date: Dec. 7, 2016

Technology description

Summary

Data mining algorithms and gene sequencing technologies allow scientists to build and analyze large data sets to identify the groups of genes associated with a given disease. However, existing algorithms cannot map specific genes to each of the specific phenotypic or biological anomalies that mark a given disease. This technology is an algorithm that identifies metagenes – linear combinations of individual genes – that are biomarkers for the specific underlying biological mechanisms of a disease. This algorithm is a powerful tool for mining large, publically available gene expression data sets and has been used to identify breast cancer prognostic biomarkers. This information can in turn be used to improve personalized diagnosis, prognosis, and treatment for cancer patients.

Data mining algorithm that simplifies identification of genetic biomarkers for specific disease phenotypes

This technology consists of an iterative algorithm that starts at a seed gene and converges on a metagene, which serves as a reliable biomarker for a specific biological mechanism that is characteristic of a larger disease. While existing algorithms can identify combinations of genes that represent a combination of phenotypes present in a given disease, this technology isolates the specific genes that represent a single, specific disease phenotype (e.g., cell transdifferentiation or the presence of an amplicon). As such, this technology may more precisely identify specific genetic biomarkers of anomalous biological mechanisms. The metagenes derived by this process may be used to develop more accurate tools for cancer diagnosis and treatment. Additionally, this technology may further clarify the biological mechanisms underlying such diseases and enable the more efficient discovery of relevant therapeutic targets.

The algorithm was validated using nearly 2,000 breast cancer samples and its predictions were shown to outperform those from current commercially available breast cancer genetic kits.

Publications

Cheng WY, Ou Yang TH, Dimitris Anastassiou. "Biomolecular Events in Cancer Revealed by Attractor Metagenes" PLoS Comput Biol. February 21, 2013; 9(2): e1002920.

Ou Yang TH, Cheng WY, Zheng T, Maurer MA and Anastassiou D, "Breast Cancer Prognostic Biomarker Using Attractor Metagenes and the FGD3-SUSD3 Metagene," Cancer Epidemiol Biomarkers Prev. December 2014; 23(12): 2850-6

Cheng WY, Ou Yang TH, Dimitris Anastassiou. "Development of a Prognostic Model for Breast Cancer Survival in an Open Challenge Environment" Sci Transl Med. April 17, 2013; 5(181): 181ra50.

Application area

Method to mine data for genetic biomarkers of specific disease phenotypes

Identification of biomarkers that can assist in the personalized diagnosis, prognosis, and choice of treatment for a given disease

Research tool to help understand the biological mechanisms behind a disease and improve drug target discovery

Advantages

Even large numbers of metagenes are easily viewable in a linear format

Metagenes group together only the most pertinent individual genes extracted from large data sets of gene expression profiles

No prior knowledge of gene function or gene interaction is needed

Metagenes can accurately portray a single phenotype or biological mechanism that occurs during the course of a given disease

Institution

[Columbia University](#)

Inventors

[Dimitris Anastassiou](#)

联系我们



叶先生

电话 : 021-65679356

手机 : 13414935137

邮箱 : yeyingsheng@zf-ym.com