Genome Medicine

RESEARCH Open Access

Sample-level enrichment analysis unravels shared stress phenotypes among multiple cancer types

Gunes Gundem[1,2*] and Nuria Lopez-Bigas[1,3*]

Abstract

Background: Adaptation to stress signals in the tumor microenvironment is a crucial step towards carcinogenic phenotype. The adaptive alterations attained by cells to withstand different types of insults are collectively referred to as the stress phenotypes of cancers. In this manuscript we explore the interrelation of different stress phenotypes in multiple cancer types and ask if these phenotypes could be used to explain prognostic differences among tumor samples.

Methods: We propose a new approach based on enrichment analysis at the level of samples (sample-level enrichment analysis - SLEA) in expression profiling datasets. Without using *a priori* phenotypic information about samples, SLEA calculates an enrichment score per sample per gene set using z-test. This score is used to determine the relative importance of the corresponding pathway or module in different patient groups.

Results: Our analysis shows that tumors significantly upregulating genes related to chromosome instability strongly correlate with worse prognosis in breast cancer. Moreover, in multiple tumor types, these tumors upregulate a senescence-bypass transcriptional program and exhibit similar stress phenotypes.

Conclusions: Using SLEA we are able to find relationships between stress phenotype pathways across multiple cancer types. Moreover we show that SLEA enables the identification of gene sets in correlation with clinical characteristics such as survival, as well as the identification of biological pathways/processes that underlie the pathology of different cancer subgroups.

Background

Complex genetic diseases such as cancer are characterized by phenotypic heterogeneity reflected at the molecular level in the form of variations in the activity of certain signaling pathways. In support of this notion, recent cancer genome studies point to the idea that distinct types of alterations in different genes tend to accumulate in pathways central to the control of cell growth and cell fate determination [1-4]. It has been proposed that expression signatures indicative of activity status of pathways can be used to define specific molecular phenotypes that characterize individual tumors [5]. A number of methods have been developed to analyze the transcriptomic changes specific to tumor samples and identify patterns of pathway deregulation that differentiate distinct patient subgroups [6-12]. These methodologies are based on the idea that analysis of pathway-level differences among samples could have an advantage of reflecting the true oncogenic phenotypes achieved through consistent expression of a set of genes compared with the acute expression of a single gene. However, each of these methods has been designed to address specific questions and, thus, have limited use for a more general application. For instance, that of Xia and Wishart is specific to metabolomic data [9], and that of Bild *et al.* [6] requires cell line perturbation data in a platform comparable to that of the tumor data. The methodologies developed by Edelman *et al.* [7], Verhaak *et al.* [8] and Yi *et al.* [10] require *a priori* information of phenotypic classification of the samples. In this manuscript, we propose a new methodology, sample-level enrichment analysis (SLEA), that overcomes these limitations and has a more general use for enrichment analysis (EA) at the level of samples. The pathways or modules are represented as lists of genes, which can be obtained from literature or online repositories such as

* Correspondence: gg10@sanger.ac.uk; nuria.lopez@upf.edu
[1]Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, Barcelona, Spain
Full list of author information is available at the end of the article