MOLECULAR ONCOLOGY XXX (2015) 1–13

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

Pathway-based personalized analysis of breast cancer expression data

Anna Livshits[a,1], *Anna Git*[b,c,1], *Garold Fuks*[a,1], *Carlos Caldas*[b,c], *Eytan Domany*[a,*]

[a]*Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel*
[b]*Cancer Research UK Cambridge Research Institute, Cambridge CB2 0RE, UK*
[c]*Department of Oncology, University of Cambridge, Cambridge CB2 2XZ, UK*

ARTICLE INFO

Article history:
Received 8 September 2014
Received in revised form
23 March 2015
Accepted 16 April 2015
Available online ■

Keywords:
Breast cancer
Pathway-based analysis
PKA pathways

ABSTRACT

Introduction: Most analyses of high throughput cancer data represent tumors by "atomistic" single-gene properties. Pathifier, a recently introduced method, characterizes a tumor in terms of "coarse grained" pathway-based variables.

Methods: We applied Pathifier to study a very large dataset of 2000 breast cancer samples and 144 normal tissues. Pathifier uses known gene assignments to pathways and biological processes to calculate for each pathway and tumor a Pathway Deregulation Score (PDS). Individual samples are represented in terms of their PDSs calculated for several hundred pathways, and the samples of the data set are analyzed and stratified on the basis of their profiles over these "coarse grained", biologically meaningful variables.

Results: We identified nine tumor subtypes; a new subclass (comprising about 7% of the samples) exhibits high deregulation in 38 PKA pathways, induced by overexpression of the gene *PRKACB*. Another interesting finding is that basal tumors break into two subclasses, with low and high deregulation of a cluster of immune system pathways. High deregulation corresponds to higher concentrations of Tumor Infiltrating Lymphocytes, and the patients of this basal subtype have better prognosis. The analysis used 1000 "discovery set" tumors; our results were highly reproducible on 1000 independent "validation" samples.

Conclusions: The coarse-grained variables that represent pathway deregulation provide a basis for relevant, novel and robust findings for breast cancer. Our analysis indicates that in breast cancer reliable prognostic signatures are most likely to be obtained by treating separately different subgroups of the patients.

Abbreviations: PKA, Protein Kinase A; PDS, Pathway Deregulation Score; ER, Estrogen Receptor; HER2, Epidermal growth factor receptor 2; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; TIL, Tumor Infiltrating Lymphocyte; KEGG, Kyoto Encyclopedia of Genes and Genomes; PAM, Prediction Analysis of Microarrays.

* *Corresponding author*. Tel.: +972 89343964.
E-mail addresses: annaliv@gmail.com (A. Livshits), Anna.Git@cruk.cam.ac.uk (A. Git), Garold.fuks@weizmann.ac.il (G. Fuks), Carlos.Caldas@cruk.cam.ac.uk (C. Caldas), eytan.domany@weizmann.ac.il (E. Domany).

[1] Equal first author contribution.

http://dx.doi.org/10.1016/j.molonc.2015.04.006
1574-7891/

Please cite this article in press as: Livshits, A., et al., Pathway-based personalized analysis of breast cancer expression data, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.04.006

1. Introduction

Breast cancer is one of the most common malignancies, with about one in nine women contracting it during their lives (DeSantis et al., 2011). It is a highly heterogeneous disease (Polyak, 2011), in terms of pathological and clinical parameters of the patients (age, tumor size, node status, and histological grade); heterogeneity is reflected also in immunohistochemically measured biomarkers (Estrogen Receptor (ER), Progesterone Receptor (PR), Epidermal growth factor receptor 2 (HER2)), in the tumors' molecular characteristics, patient response to therapy and clinical outcome. Considerable effort has been invested over several years to stratify breast cancer into clinically distinct groups on the basis of the tumors' molecular signatures (TCGA, 2012). For example, understanding the underlying molecular aberrations is essential for the design of personalized drugs, and clinically meaningful stratification is expected to serve as the basis for prediction of outcome and the choice of therapeutic strategy.

Some of the earliest seminal studies (Perou et al., 2000; Sorlie et al., 2001) identified several subtypes of breast tumors, based on gene expression patterns derived from microarray analysis of several hundred "intrinsic genes" (Sorlie et al., 2001). While there is consensus regarding robustness of the Luminal A and Basal intrinsic subtypes, reproducibility of the expression signatures of the others, e.g. Luminal B, HER2-enriched and Normal-like (Perou et al., 2000; Sorlie et al., 2001) has been questioned (Alexe et al., 2006).

An outcome predictor method based on these intrinsic subtypes – PAM50 (Parker et al., 2009) has recently gained FDA approval; nevertheless, it is clear that this stratification does not capture the full heterogeneity and complexity of the disease. In fact, a lot of effort has been invested in designing various molecular outcome predictors and in producing biologically and clinically relevant sub-classifications or new groupings of breast cancer (Bilal et al., 2013; Geiger et al., 2012; Sotiriou and Pusztai, 2009; van der Vegt et al., 2009). A recent study by Curtis et al. (Curtis et al., 2012) generated a very large dataset (*METABRIC*, Molecular Taxonomy of Breast Cancer International Consortium), of nearly 2000 tumors and 144 normal breast samples, for all of which expression and copy number were measured. Curtis et al. integrated these two types of data and using unsupervised clustering divided breast cancer into 10 new subtypes, to which they refer as iClusters. These subtypes are the representation of recurrent selection of specific somatic genomic aberrations, which in turn cause the over- or under-expression of driver oncogenes and tumor suppressors, respectively, to which different tumors are addicted.

The studies mentioned above used a priori existing biological information only in a very limited manner, if at all. The prognostic gene lists of most predictors were assembled using machine learning, either utilizing no biological knowledge at all (van 't Veer et al., 2002) or minimal information, such as treating ER-/+ tumors separately (Wang et al., 2005) or focusing on "intrinsic genes" for classification (Parker et al., 2009; Sorlie et al., 2001). A recently introduced method, *Pathifier* (Drier et al., 2013), advocates taking a different direction, which does make extensive use of available biological knowledge (Ideker et al., 2011). *Pathifier* uses the known association of each relevant biological pathway or process with a corresponding list of genes that were shown to play a role in it. First, for each individual tumor k and pathway P a Pathway Deregulation Score (PDS), denoted by $D(P,k)$, is derived (see Supplementary File 1, Supplementary Figure 1 for a schematic presentation of the method). Next, any kind of preferred method of analysis (e.g. clustering) is performed on *these variables*, rather than on the "raw" expression or copy number data. The approach is phenomenological and, unlike the method of Vaske et al., 2010, requires neither knowledge of the inter-relations between thousands of "biomolecular entities" nor measurement of their status. It was demonstrated (Drier et al., 2013) (for Glioblastoma and colon cancer) that by simple unsupervised analysis of these "coarse grained" biologically meaningful scores, *Pathifier* finds clinically relevant patient groups and relationships that were not captured by standard methods. Several recent studies adopted pathway-based approaches for the analysis of cancer expression data, going beyond the simplest enrichment analysis (Huang et al., 2014; Verhaegh and Van de Stolpe, 2014; Verhaegh et al., 2014).

In the present study we apply *Pathifier* on the *METABRIC* dataset to stratify breast cancer. Since copy number variations affect the biological state of a cell primarily via the corresponding transcriptome, and transcript levels reflect directly genes' association with the activity of a pathway or biological process, we used only the expression data of Curtis et al. (Curtis et al., 2012). Hence it is not surprising that the classes that we found do not overlap with the iClusters.

We find nine tumor types with distinct PDS profiles over 7 clusters of pathways. Our stratification reproduces known partitions (clearly – into ER+/-, and to some limited extent – into the intrinsic subtypes), but reveals also previously unreported subclasses. One of these is a group of Luminal tumors (mainly of type A) with high deregulation scores of a cluster of pathways associated with Protein Kinase A (PKA) activity (Francis et al., 2011; Johnson et al., 2001). Deregulation of the PKA pathways possibly plays a role in the malignant process (Beuschlein et al., 2014; Forlino et al., 2014); in that case this finding may have therapeutic implications.

Another interesting finding is a clear separation of basal (ER-/HER2-) tumors into two groups, which exhibit either high or low levels of deregulation of immune system related pathways. We first identify the biological basis of this partition in terms of the presence (absence) of Tumor Infiltrating Lymphocytes (TILs), and show that the two basal subgroups have different outcome. We have reasons to attribute this to different responses to therapy exhibited by the two groups. Association of outcome and response to therapy with the level of TILs has been noted previously for several other breast cancer subtypes: for HER2+ (Alexe et al., 2007), HER2+/ER- (Rody et al., 2009), and node + subjects (Loi et al., 2013). Additional papers that address association of TILs with response to therapy were discussed by S. Ganesan in the 2014 ASCO50 meeting. Our finding is in agreement also with (Calabrò et al., 2009; Teschendorff et al., 2007), and a most extensive study which established association of T-cell infiltration with survival in ER-breast cancer patients (Ali et al., 2014).

Please cite this article in press as: Livshits, A., et al., Pathway-based personalized analysis of breast cancer expression data, Molecular Oncology (2015), http://dx.doi.org/10.1016/j.molonc.2015.04.006