Methods 124 (2017) 89–99

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Protein network construction using reverse phase protein array data

Rency S. Varghese [a], Yiming Zuo [a,b], Yi Zhao [a,c], Yong-Wei Zhang [a], Sandra A. Jablonski [a], Mariaelena Pierobon [d], Emanuel F. Petricoin [d], Habtom W. Ressom [a,*], Louis M. Weiner [a,*]

[a] Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA
[b] Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA, USA
[c] Department of Biostatistics, School of Public Health, Brown University, Rhode Island, Providence, USA
[d] Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA, USA

ARTICLE INFO

Article history:
Received 4 April 2017
Received in revised form 22 May 2017
Accepted 17 June 2017
Available online 24 June 2017

Keywords:
RPPA
MANOVA
Network construction
Topology analysis
Breast cancer

ABSTRACT

In this paper, we introduce a novel computational method for constructing protein networks based on reverse phase protein array (RPPA) data to identify complex patterns in protein signaling. The method is applied to phosphoproteomic profiles of basal expression and activation/phosphorylation of 76 key signaling proteins in three breast cancer cell lines (MCF7, LCC1, and LCC9). Temporal RPPA data are acquired at 48 h, 96 h, and 144 h after knocking down four genes in separate experiments. These genes are selected from a previous study as important determinants for breast cancer survival. Interaction networks are constructed by analyzing the expression levels of protein pairs using a multivariate analysis of variance model. A new scoring criterion is introduced to determine relevant protein pairs. Through a network topology based analysis, we search for wiring patterns to identify key proteins that are associated with significant changes in expression levels across various experimental conditions.

1. Introduction

In cancer, genetic and epigenetic changes are often associated with disease development. Studying epigenetic changes such as protein phosphorylation will greatly aid in understanding the causes and determining effective treatment of cancers and other diseases. With the development of personalized therapeutics for oncology, the systematic and targeted analysis of selected proteins including phosphorylated proteins in tumor tissues is receiving increasing interest [1]. Reverse-phase protein arrays (RPPAs) have emerged as a useful tool for the large-scale analysis of protein expression and protein activation, allowing for the specific detection and quantification of proteins in a reproducible and highly parallelized manner [2]. Besides monitoring differential expression, RPPAs allow the profiling of differential protein modification. Due to the dependency on dot blot compatible antibodies, the number of detectable proteins is limited, but a large number of samples can be profiled due to the reverse array format. Therefore, RPPAs can be used in complex studies, where the impact of multiple experimental factors (e.g., multiple treatments, doses, and time points) on protein expression and cellular signaling is investigated [3].

RPPA data analysis is still a growing area of research. Studies utilizing RPPA data have employed a diverse range of data preprocessing and benchmarking methods, but no single protocol for processing RPPA data has been universally accepted [4–8]. Some of the tools include a web-based data analysis pipeline RPPApipe [3] and an R-package RPPanalyzer [9]. A review of the tools and software approaches developed for RPPA data normalization and data analysis can be found in Ref. [10]. In a typical RPPA data analysis, proteins are analyzed individually and expression values are considered to be of primary interest. However, since proteins function in networks and interact with many partners, a network-based approach is desired to analyze the RPPA-based protein expression data. For example, RPPA were used to analyze the expression of 203 proteins in cells taken from acute myeloid leukemia (AML) patients using a network-based approach [11]. Dominant overlapping protein networks between subtypes of AML patients were characterized using a paired *t*-test and lasso regression analysis. Signaling networks were constructed from the protein pairs that were significantly different. Predicted networks were also compared to known networks from public protein–protein interaction and signaling databases. In a recent study, the levels of 134 proteins measured in 21 breast cancer cell lines stimulated with IGF1 or insulin for up to 48 h were evaluated by network analysis [8]. Specifically, directed protein expression networks named as time translation models were constructed using lasso regression, conventional matrix inversion, and

* Corresponding authors.
E-mail addresses: hwr@georgetown.edu (H.W. Ressom), weinerl@georgetown.edu (L.M. Weiner).

http://dx.doi.org/10.1016/j.ymeth.2017.06.017
1046-2023/

entropy maximization. The inferred interactions were ranked by differential magnitude to identify pathway differences.

In this study, we introduce a method that performs statistical analysis of protein-pairs to construct protein networks, unlike most studies where proteins are analyzed individually. Instead of taking the ratio of the expression levels of the proteins, we analyze the protein pairs via a multivariate analysis of variance (MANOVA) to test for patterns, where the expression levels of a protein pairs are considered as a bivariate outcome. This approach allows a higher power of identifying significant changes than the ratio comparison. In addition, we introduced a new scoring criterion that utilizes the significance and the correlation of each protein pair, as well as the importance of each protein, for construction of a network. This differs with information-theoretic methods, which constructs networks based on the associations of protein pairs using metrics such as correlation, and mutual information [12,13]. These information-theoretic methods will not work well on this dataset since only limited samples exist for each condition while our proposed method will perform well under this restriction. Another typical network based method is weighted gene co-expression network analysis (WGCNA), which mainly focuses on identifying clusters [14]. This differs with our aim to identify protein markers. In addition, it is unclear how well WGCNA performs under small sample size restrictions. To explore the mechanisms of resistance to therapies targeting estrogen pathways in the treatment of estrogen receptor (ER) positive breast cancer, we previously performed a systemic biological screening of a library of siRNAs targeting an ER- and aromatase-centered network [15]. We identified 46 genes that are dispensable in the estrogen-dependent MCF7 cell lines, but are selectively required for the survival of estrogen-independent MCF7-derived cell lines (LCC1 and LCC9). Based on viability data, we selected four genes: Transducer of ERBB2, 1 (*TOB1*), Polymerase (RNA) II subunit B (*POLR2B*), Proteasome 26S Subunit ATPase 5 (*PSMC5*), and Cysteine Rich Angiogenic Inducer 61 (*CYR61*). These genes were knocked down individually in each cell line to explore their role in the survival of estrogen-independent ER positive cells and the impact on the signaling architecture of cancer-focused pathways.

We evaluated the proposed computational method using RPPA data derived from breast cancer cells with activation/phosphorylation of 76 key signaling proteins in MCF7, LCC1 and LCC9 cells. Interaction networks between the proteins and phosphorylated proteins were constructed to determine the proteins that significantly changed when a gene was knocked down. The networks were built through the recognition of protein-protein interactions by identifying those pairs of proteins whose expression levels changed in each knockdown at 48 h, 96 h, and 144 h compared to the negative control (without gene knockdown) for cell lines MCF7, LCC1 and LCC9 [11]. A topological analysis of the networks identified key proteins in each network that can play an important role in estrogen-dependent and estrogen-independent cell lines.

The rest of the paper is organized as follows. Section 2 describes the cell lines, the workflow for the analysis of RPPA data, the proposed MANOVA model for network construction, the protein-protein pair correlative analysis, and topological analysis of the resulting networks. Section 3 presents the networks generated by the analysis of temporal RPPA data and the key proteins identified for each network. Finally, Section 4 summarizes the work and discusses future goals.

2. Materials and methods

2.1. Reverse phase protein array (RPPA) expression data

To explore the roles of *POLR2B, TOB1, CYR61* and *PSMC5* genes in the survival of estrogen-independent cells and the impact on the signaling architecture of cancer-focused pathways, basal expression and activation of 76 key signaling proteins in MCF7, LCC1 and LCC9 cells were analyzed using RPPA. MCF7 is an ER positive and estrogen-dependent breast adenocarcinoma cell line and sensitive to treatment with the anti-estrogen (AE) reagents: tamoxifen and fulvestrant. LCC1 is derived from MCF7 and selected in vivo for estrogen-independence, which commonly reflects resistance to aromatase inhibition (AI), but remains sensitive to tamoxifen and fulvestrant. LCC9, further derived by selection from LCC1 cells, is resistant to both tamoxifen and fulvestrant [16,17].

The four genes for knockdown in each cell line were selected from a previous study [15] that aimed to identify new points of vulnerability in estrogen-independent, AE/AI-resistant breast cancers. The study identified a group of genes with action specifically required for the survival of estrogen independent cells. Tumor suppressor gene TOB1 was identified as a critical determinant of estrogen-independent ER-positive breast cell survival. In addition to *TOB1* gene, other 45 genes presented potential function in estrogen-independent growth of ER positive breast cancer. In order to broaden the understanding of the mechanisms of estrogen-independent growth, based on viability data, *TOB1* and three genes (*POLR2B, CYR61,* and *PSMC5*) were selected for knockdown experiments and RPPA based analysis. These knocked-down genes in the estrogen-independent breast cancer cell lines also demonstrated varying levels of apoptotic activity, and were chosen for RPPA based analysis to compare signaling pathways affected by knockdown of each gene. *TOB1, POLR2B CYR61, PSMC5*, or negative scrambled siRNAs with a final concentration of 10nM were reverse transfected into cells for 48 h, 96 h and 144 h. Triplicates of each transfection were collected for analysis by RPPA [18]. We used 76 antibodies listed in Ref. [15] for acquisition of RPAA data as described previously [19].

2.2. Protein-protein pair correlation analysis and protein network construction

Fig. 1 presents the overall workflow for the analysis of RPPA data. Following normalization of the RPPA data, protein pairs were created and the expression levels were analyzed for each protein pair.

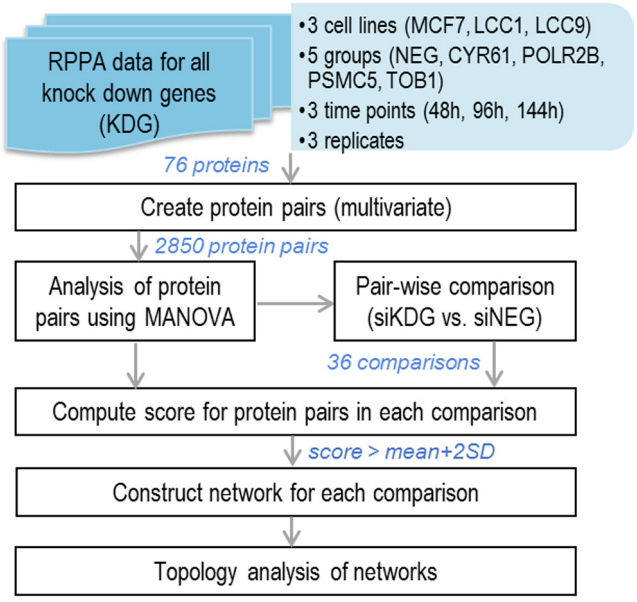

Fig. 1. An overview of the workflow for RPPA data analysis.