Integrative Computational Identifications of the Signaling Pathway Network Related to TNF-alpha Stimulus in Vascular Endothelial Cells

Jin Gu, Shao Li, Yang Chen, Yanda Li

MOE Key Laboratory of Bioinformatics and Bioinformatics Division
TNLIST / Department of Automation, Tsinghua University
Beijing 100084, China
Email: gujin_00@mails.tsinghua.edu.cn, {shaoli, yc, daulyd}@tsinghua.edu.cn

Abstract—Integrating multiple datasets becomes essential for systematically understanding complex biological processes. Here, we presented a new method to identify and annotate the downstream pathways related to TNF-alpha (TNF) stimulus in vascular endothelial cells (VECs). It is found that the downstream pathways of TNF can both cause cell proliferation and cell death, which may provide an explanation for the ambiguous roles of TNF in angiogenesis. By calculating the distances in protein-protein interaction (PPI) network between different signaling pathways, many possible cross-talks are predicted, which suggests complex downstream responses for TNF stimulus.

Keywords—Systems biology; pathway; TNF-alpha (TNF); vascular endothelial cells (VECs)

I. INTRODUCTION

Understanding the complex responses of cells to environmental stimuli is a promising task for systems biology. Here, we present a new integrative computational approach to identify the pathway-level responses to TNF-alpha (hereinafter referred as TNF) stimulus in vascular endothelial cells.

TNF, mainly secreted by macrophages, is an important cytokine, which is involved in a wide spectrum of biological and physiological processes. One of the TNF-involved processes is angiogenesis, a process involving the growth of new blood vessels from pre-existing vessels and essential for many tumor growth [1-4].

Vascular endothelial cells (VECs), such as human umbilical vein endothelial cells (HUVECs) and human dermal microvascular endothelial cells (HDMVECs), are usually used as in vitro model of angiogenesis. TNF stimulus can affect the proliferation and migration of VECs, which are essential steps for angiogenesis. But whether TNF promotes or inhibits angiogenesis remains ambiguous [5-7]. According to current knowledge, TNF binds to its receptors, TNFR1 and TNFBR, and stimulates downstream signaling pathways, such as TGF-beta signaling pathway, NF-kB signaling pathway and many other pathways [7, 8]. Previous studies mainly focused on single gene or pathway but thought little of the possible cross-talks between multiple pathways. Based on gene expression and protein-protein interaction (PPI) data, an integrative computational method was developed to analyze the pathways related to TNF stimulus in a systematical view.

One common computational strategy for systematically analyzing a specific biological process is to find the neighbor genes of a few “essential” genes in protein-protein interaction network (PPI network). For examples, Pujana et al. constructed a breast cancer gene network based on three genes and successfully found a new breast cancer related gene HMMR [9]; Bromberg et al. constructed a sub-network related to CB1R-induced neurite outgrowth by analyzing 23 activated transcription factors in PPI network [10, 11]. For angiogenesis, Abdollahi et al. constructed an angiogenic switch network: they first screened a set of anti-expressed genes by stimulating HUVECs using pro-angiogenic and anti-angiogenic factors; then the genes directly interacted with the anti-expressed ones were identified by literature mining; finally, the angiogenic network was established based on all those genes and their interactions [12].

In the first place, our method is similar to above approaches, the differentially expressed genes detected by microarray and their directly interacted genes in PPI network were identified. While in the following steps, our method differs: 1) we mainly studied the responses in pathway level rather than single genes; 2) a typical pro-angiogenic process induced by vascular endothelial growth factor A (VEGF) was used as “reference” to analyze the identified pathways.

II. METHODS AND MATERIALS

A. Identifications of the seed genes affected by TNF and VEGF in vascular endothelial cells (VECs)

Under environmental stimuli, the transcription of downstream genes will be activated or repressed. The differentially expressed genes provide important information of the induced process. Two microarray datasets, GSE2638 and GSE2639 (all the microarray datasets can be found in NCBI GEO database with the given IDs), were used to identify the defferentially expressed genes in VECs after TNF stimulus by SAM package with default parameters [13]. To reduce the noise in microarray data, additional evidences were used to filter the defferentially expressed genes [14]: the gene differentially expressed at least in one microarray dataset and also supported by additional evidences (at least three evidences in total) were left for the following analysis.
TABLE I. THE MULTIPLE DATASETS USED TO IDENTIFY SEED GENES

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Cut-off</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microarray</td>
<td>GSE2639</td>
<td>SAM</td>
<td>TNF</td>
</tr>
<tr>
<td>Microarray</td>
<td>GSE2639</td>
<td>SAM</td>
<td>TNF</td>
</tr>
<tr>
<td>Microarray</td>
<td>GSE837</td>
<td>SAM</td>
<td>VEGF</td>
</tr>
<tr>
<td>Microarray</td>
<td>GSE3299</td>
<td>SAM</td>
<td>VEGF</td>
</tr>
<tr>
<td>PubMed</td>
<td>Abstracts with key words (TNF AND (angiogenesis OR inflammation))</td>
<td>≥ 5 hits</td>
<td>TNF</td>
</tr>
<tr>
<td>PubMed</td>
<td>Abstracts with key words (VEGF AND (angiogenesis OR inflammation))</td>
<td>≥ 5 hits</td>
<td>VEGF</td>
</tr>
<tr>
<td>GO</td>
<td>Genes in “angiogenesis” (GO:0001525 with all child terms, 68 genes) and “response to inflammation” (GO:0006954 with all child terms, 180 genes)</td>
<td>≥ 1 hit</td>
<td>TNF VEGF</td>
</tr>
<tr>
<td>KEGG Pathway</td>
<td>Genes in the pathways containing TNF</td>
<td>≥ 1 hit</td>
<td>TNF</td>
</tr>
<tr>
<td>KEGG Pathway</td>
<td>Genes in the pathways containing VEGF</td>
<td>≥ 1 hit</td>
<td>VEGF</td>
</tr>
<tr>
<td>dbEST</td>
<td>Expressed sequence tags in endothelial cells</td>
<td>≥ 10 hits</td>
<td>TNF VEGF</td>
</tr>
</tbody>
</table>

The other four evidences are 1) literature information from PubMed; 2) GO annotations; 3) KEGG pathways; 4) expressed sequences tags (ESTs). Please see table 1 for detail. The identified genes were called “seed” genes in this study.

Finally, 70 seeds genes were identified for TNF stimulus in VECs. With the same procedure, another two microarray datasets were used for VEGF stimulus, GSE3299 and GSE837, and 78 seed genes were identified.

B. Validations of the seed genes by another time-course microarray dataset

To validate the seed genes, another independent time-course microarray dataset GSE9055 (HUVEC with 10ng/ml TNF treatment, 0~8h, 25 time points) was used. Using the clustering software CLUSTER and the tree viewing software TreeView [15], two tight clusters, including 25 and 15 genes with Pearson Correlation > 0.9 were identified and visualized (Fig. 1). The two clusters were significantly co-expressed than background (p-value 4.74E-30), which supports that the seed genes are closely related to TNF-induced process.

C. Identifications of the bridge genes in protein-protein interaction (PPI) networks

Comparing the lists of the 70 TNF seed genes and the 78 VEGF seed genes, only 8 genes are overlapped (Fisher’s exact test p-value > 0.05). It seems that the two processes are independent.

The differentially expressed seed genes may not contain all information about the two processes, because many genes are activated by post-transcriptional modifications, such as STAT3, a member of signal transducer and activator of transcription (STAT) gene family, is regulated by protein phosphorylation [16]. The genes involved in the same process are more likely to interact with each other than unrelated genes. So searching for the neighbors of seed genes in PPI network can cover an additional set of downstream genes, which may not have significant changes in mRNA expression level [9-12].

The PPI network was constructed according to annotated PPIs in HPRD [17-20]. Only the genes within the biggest connected sub-network, including 9,028 nodes and 34,803 edges, were left for the following analysis. Fifty-nine of the 70 TNF seed genes and 68 of the 78 VEGF seed genes can be mapped on to the network.

Then the genes directed interacted with the seeds (DIGs) were identified in the network. To avoid bias, the eight overlapped genes were removed from VEGF seeds, so in the following analysis, there are only 70 VEGF seed genes. We found 570 genes directly interacted with TNF seed genes (TNF-DIGs) and 521 genes directly interacted with VEGF seed genes (VEGF-DIGs). One hundred and thirty-five genes were overlapped (bridge genes, BGs) between the two sets of DIGs (Fisher’s exact test p-value, 8.09E-12), which indicates that TNF-induced process is indeed closely related to VEGF-induced process.
D. Two-Stage Pathway Enrichment Analysis

In the first stage, pathway enrichments were simply calculated in seed genes, TNF-DIGs, VEGF-DIGs and BGs using DAVID web-based tool [21, 22]. Significantly enriched pathways with EASE p-value < 0.001 were identified [23]. Several pathway annotations are available in DAVID. In this study, we mainly used KEGG signaling pathways and another two essential pathways, cell cycle and apoptosis.

In the second stage, also using EASE modified Fisher’s exact test, the pathways with significant higher enrichment in BGs against TNF-DIGs were identified (p-value < 0.05). These pathways with significant enrichment increase are predicted as pro-angiogenic, because these pathways are significantly overlapped with the downstream pathways in the typical pro-angiogenic process induced by VEGF. The Fisher’s exact test table was shown in Table 2.

E. Analysis of cross-talks between pathways

In biological systems, multiple pathways cooperate or “cross-talk” with each other instead of working separately. But in most pathway databases, pathways are annotated singly. One strategy to predict the possible cross-talks is to embed individual pathways into a genome-wide PPI network [24]. For example, Lu et al. identified a set of cross-talks simply by searching the neighbors and the PPIs of the genes in different pathways [24]. Making a step forward, Li et al. proposed a method by examining the significance of direct interactions between different pathways in PPI network based on Fisher’s exact test [25].

In our study, pathway distance was introduced to predict the possible cross-talks between two pathways. Denote the i-th pathway and the genes in the pathway as:

\[ p_i = \{ g_{i,1}, g_{i,2}, L, g_{i,n_i} \}, \]

and the shortest path length in PPI network between two genes as:

\[ l_{i,j}(g_{i,k_1}, g_{i,k_2}), \quad i_1 \neq i_2. \]

The distance between any pair of pathways can be calculated as the average of the shortest path lengths between all pairs of genes in the two pathways:

\[ d(p_i, p_{i'}) = \frac{1}{n_{i,k_1} \cdot n_{i',k_2}} \sum_{k_1} \sum_{k_2} l_{i,j}(g_{i,k_1}, g_{i,k_2}), \quad i_1 \neq i_2. \]

To estimate the significance of the distance, we randomly shuffled the genes in pathways to generate background datasets. Each time, 1) two different pathways are randomly selected, and then 2) two different genes, one for one pathway, are randomly selected. To reduce bias, the two genes are required to have equal degree in PPI network, because the genes with higher degree (means more neighbor genes) are more likely to have shorter shortest path length than other genes. Finally, 3) swap the two genes between the selected pathways. A background dataset is generated by repeating shuffling 1000,000 times.

According to the shuffling procedure, 1,000 background datasets were generated. By comparing the distance of any pair of pathways in original dataset and the background datasets, p-value was calculated as:

\[ p = \frac{\# \{ k \mid l_k \leq l_{i,j} \leq 1 \leq k \leq 1000 \} }{1000}. \]

The pairs with p-value < 0.005 are predicted as cross-talks.

III. RESULTS

A. Identifications of downstream pathways related to TNF stimulus in vascular endothelial cells (VECs)

Previous experimental studies show that TNF is an important cytokine involving in angiogenesis (see review in [6]), but its role in angiogenesis remains ambiguous. Identifying its downstream genes and pathways will help understand the complex process of angiogenesis. By combining multiple evidences, including microarray expression profiles, PubMed abstracts, GO / KEGG annotations and ESTs (Tab. 1), 70 seed genes related to TNF stimulus in VECs were identified. These genes were validated by another independent time-course microarray dataset.

Using DAVID web-based tool, Toll-like receptor signaling pathway (p-value 1.32E-08) and apoptosis (p-value 1.60E-09) were identified as significantly enriched in the 70 seed genes. Toll-like receptor signaling pathway is an important pathway in pro-inflammatory reaction and closely related to TNF stimulus [26, 27]. And years before, apoptosis in endothelial cells has been reported after TNF stimulus [28, 29]. So the computational results are consistent with the report in literatures.

The activity changes cannot be detected on mRNA expression level for many genes due to complex post-transcriptional regulations. But now, there is still no high-throughput method to profile protein activities. The genes involved in the same process are more likely to interact with each other in PPI network (see two examples in [30, 31]). So we extended the seed genes to their direct neighbors in PPI network to cover a broader set of downstream genes. A set of 570 genes directly interacted with TNF seeds (TNF-DIGs) were identified in PPI network. Then these TNF-DIGs were also mapped to pathways by DAVID enrichment analysis. Except the above two pathways, other six pathways were also significantly enriched (p-value < 0.001): ErbB signaling pathway, MAPK signaling pathway, TGF-beta signaling pathway, VEGF signaling pathway and cell cycle (Tab. 3).
TABLE III. ENRICHED PATHWAYS IN DIFFERENT SETS OF GENES.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>p-value</th>
<th>Foreground</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoptosis</td>
<td>1.60E-09</td>
<td>TNF seed genes</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>Toll-like receptor signaling path</td>
<td>1.32E-08</td>
<td>TNF seed genes</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>Toll-like receptor signaling path</td>
<td>5.30E-14</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>Cell Cycle</td>
<td>2.82E-12</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>ErbB signaling pathway</td>
<td>5.03E-10</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
<td>6.63E-10</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>5.46E-08</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>TGF-beta signaling pathway</td>
<td>3.79E-06</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>Jak-STAT signaling pathway</td>
<td>4.42E-06</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>VEGF signaling pathway</td>
<td>3.07E-04</td>
<td>TNF-DIGs</td>
<td>All genes in KEGG</td>
</tr>
<tr>
<td>ErbB signaling pathway</td>
<td>5.47E-04</td>
<td>BGs</td>
<td>TNF-DIGs</td>
</tr>
<tr>
<td>VEGF signaling pathway</td>
<td>2.09E-02</td>
<td>BGs</td>
<td>TNF-DIGs</td>
</tr>
</tbody>
</table>

a. The p-values for enrichments are calculated using EASE modified Fisher’s exact test.

B. Comparing to the seed genes in the VEGF-induced pro-angiogenic process

VEGF is the one of key growth factors which greatly promotes endothelial cells migration and proliferation, and induces angiogenesis. Comparing to VEGF-induced process would provide important information to understand the TNF stimulus in VECs.

Using the same method, 78 VEGF seed genes were identified. But only eight of them are overlapped with TNF seed genes (Fisher’s exact test p-value > 0.05). It seems that the two processes are irrelevant.

Then we tried to analyze the relationships between the two sets of genes in PPI network. It is found that they are close-related in the network: 1) the shortest paths from the VEGF seed genes to the TNF seed genes are significantly shorter than the other genes (average shortest path length: 1.98 for VEGF seeds, 2.41 for background genes, t-test p-value 9.08E-06); 2) in the 521 VEGF-DIGs (already excluding the genes interacted with the eight overlapped genes), 135 are overlapped with TNF-DIGs (p-value 8.09E-12). These results suggest that in VECs the TNF-induced process is highly related to VEGF-induced process or pro-angiogenic process.

C. Predictions of the pathways highly related to pro-angiogenic process

In PPI network, 135 bridge genes (BGs) were identified to be both directly interacted with TNF and VEGF seed genes. These genes are more likely to be involved in pro-angiogenic process than the other TNF-DIGs. The pathway enrichment was compared in BGs against TNF-DIGs using EASE modified Fisher’s exact test (Tab. 2).

Two pathways, ErbB signaling pathway (p-value 5.47E-04) and VEGF signaling pathway (p-value 2.09E-02) were predicted as pro-angiogenic, because they have significant enrichment increase in BGs than in all TNF-DIGs. According to KEGG annotation, the major role of the two pathways is to promote cell proliferation and differentiation. Additionally, the enrichments of MAPK signaling pathway (p-value 0.150) and cell cycle (p-value 0.183) are also slightly increased.

Except the four increased pathways, the other pathways are more related to apoptosis or cell death. For example, activation of TGF-beta signaling pathway will cause cell death of endothelial cells [32, 33].

These results indicate that the TNF induced downstream pathways have both roles in promoting cell proliferation (pro-angiogenic) and promoting cell death (anti-angiogenic). The different dynamic changes of the activities in these pathways may cause different effects on angiogenesis.

D. Network of the pathways related to TNF stimulus

According to previous biological results and above computational analysis, it is suggested that single pathway is not enough to explain the complex process in VECs after TNF stimulus. Put all the related pathway into a unified network will benefit future studies.

By analyzing the distance in PPI network, possible cross-talks between different pathways were identified. Many pathways are close-related in PPI network (Fig. 2), which suggesting complex interactions between them. MAPK signaling pathway and ErbB signaling pathway are hub nodes in the network, both of which connect with seven different pathways. While TGF-beta signaling pathway does not connect with any other pathway. Based on this network, other types of pathways, such as human diseases pathways, can also be easily added.
IV. DISCUSSIONS

Identifications of the downstream pathways are very useful for understanding a specific biological or physiological process. Here, we presented a new method to identify and analyze the downstream pathways mainly based on microarray data, genome-wide protein-protein interaction (PPI) data and KEGG pathway annotations. This method was applied on an important biological process, TNF stimulation in vascular endothelial cells (VECs), and successfully identified several downstream signaling pathways and the possible cross-talks between them. By comparing to a typical pro-angiogenic process induced by VEGF, the roles of the identified pathways in angiogenesis were also predicted.

In this study, we mainly focus on pathways instead of single gene. Genome-wide data, such as microarray or yeast two hybrid (for profiling protein-protein interactions), contain many noises. By mapping single genes to pathways, we can reduce the noises and provide clearer biological interpretations of the computational results. Integrating multiple data has two advantages for understanding the complex biological process: 1) reduce the noises by cross-validations between different data sources; 2) remedy undetected information in any single dataset, such as PPIs are helpful to study the genes without significant mRNA expression changes.

ACKNOWLEDGMENT

We thank Ningbo Zhang for providing PPI data and network analysis tool. We thank Bo Zhang, Lisha Li and Tao Ma for useful discussions. This work is supported by NSFC30873464 and 863 Project 2006AA02Z311.

REFERENCES


