Transcription factors and target genes relationship based method reveals potential thermal injury related genes

Bo Li, Yuxin Song, Hui Zhang and Wenyuan Luo*

Department of Trauma surgery, Gansu provincial Hospital, No.204 Donggang west road, Lanzhou city, Gansu province 730000, P.R. China.

Accepted 6 July, 2011

Thermal injury to the skin can induce local and systemic perturbations that are costly in terms of human suffering as well as in strains on the health care system. We used the GSE8056 microarray data to identify potential genes related to burn wound to construct a regulation network. In the network, some of transcription factors (TFs) and target genes have been proved that are related to burn wound in previous study. We also find some new TFs and target genes response to thermal injury, such as EGLN3 and CREM. It is demonstrated that transcriptome network analysis is useful in identification of the candidate genes in thermal injury.

Key words: Transcription factors, genes, thermal injury, transcriptome network.

INTRODUCTION

Thermal injury is a kind of severe trauma which causes damage to the partial or entire physical function. The recovery undergoes several stages like inflammation and hypermetabolism (Dasu et al., 2004). In the process of response to the thermal injury, more body heat would be lose through the impaired skin and the liver will synthesize higher levels of plasma proteins (Baumann and Gauldie, 1994), both resulting in changes in energy expenditure. Thence, physical stress caused by the thermal injury has great influence on the metabolism in liver, the organ critical to the modulation of immune function, inflammatory process and the restoration of homeostasis (Yang et al., 2007).

It had been hypothesized that the change of liver gene expression patterns for the thermal injury recovery will reflect the liver’s role in the response. The description of the left molecular fingerprint, expression profile for instance, may be helpful for us to clarify these intrigued mechanisms and to find out new therapeutic intervention. Models have been suggested to figure out the transcriptional characteristics of hypermetabolism through measuring the appropriate gene expression response (Chinnaiyan et al., 2001; Vemula et al., 2004) in live animals. If the generated data can provide enough information, the mechanism activated by the thermal injury will be finally unveiled with the help of transcriptional profiling that monitors responses of the inflammation (Spies et al., 2002; Dasu et al., 2004; Vemula et al., 2004). We hold that relevant genes act in concert and incline to correlate highly together as clusters. Thus, it is feasible for a selection step based on their population to isolate them to acquire genes in highly correlated sets.

DNA microarray analysis as a global approach is applied to investigate physiological mechanisms in health and disease (Spies et al., 2002). A high-throughput microarray experiment was designed to analyze genetic expression patterns and identify potential genes to target for thermal injury (Greco et al., 2010).

We developed a transcriptome network in which the differentially expressed genes, regulated by a set of transcription factors, are induced by thermal injury and response to burn wound. Further, analysis of the genes and pathways in the network to identify potential mechanisms that response to the thermal injury should proceed.

*Corresponding author. E-mail: dopaleepo@hotmail.com.
DATA AND METHODS

Affymetrix microarray data

Transcription profiles of thermal injury of human skin GSE8056 were obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/) which are based on the Affymetrix GPL570 platform data (Affymetrix Human Genome U133A Array) (Wachi et al., 2005). The study compares gene expression from burn wound margins at various times following thermal injury to expression observed in normal skin. Total 12 slides are divided into 4 sets, and in each set, three samples serve as replicates.

Pathway data

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (Kanehisa, 2002). The PATHWAY database records networks of molecular interactions in the cells, and their variants specific to particular organisms (http://www.genome.jp/kegg/). Total 130 pathways, involving 2287 genes, were collected from KEGG.

Regulation data

There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of these are presumed to function as transcription factors (Wachi et al., 2005). The combinatorial use of a subset of the approximately 2000 human transcription factors easily accounts for the unique regulation of each gene in the human genome during development (Brivanlou and Darnell, 2002).

These transcription factors are grouped into 5 super class families, based on the presence of conserved DNA-binding domains. TRANSFAC database contains data on transcription factors, their experimentally-proven binding sites, and regulated genes (Wingender, 2008).

Transcriptional regulatory element database (TRED) has been built in response to increasing needs of an integrated repository for both cis- and trans- regulatory elements in mammals (Jiang et al., 2007). TRED has done the curation for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The curation is currently focusing on target genes of 36 cancer-related TF families.

774 pairs of regulatory relationship between 219 transcription factors (TFs) and 265 target genes were collected from TRANSFAC (http://www.gene-regulation.com/pub/databases.html). 5722 pairs of regulatory relationship between 102 transcription factors (TFs) and 2920 target genes were collected from TRED (http://rulai.cshl.edu/TRED/).

Combining the two regulation datasets, total 6328 regulatory relationships between 276 TFs and 3002 target genes were collected (Table 1).

### Table 1. Regulation datasets.

<table>
<thead>
<tr>
<th>Source</th>
<th>Regulation</th>
<th>TFs</th>
<th>Targets</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRED</td>
<td>5722</td>
<td>102</td>
<td>2920</td>
<td><a href="http://rulai.cshl.edu/TRED/">http://rulai.cshl.edu/TRED/</a></td>
</tr>
<tr>
<td>Total</td>
<td>6328</td>
<td>276</td>
<td>3002</td>
<td></td>
</tr>
</tbody>
</table>

Differentially expressed genes (DEGs) analysis

For the GSE8056 dataset, the Limma method (Smyth, 2004) was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA method with the default settings implemented in Bioconductor, and then construct the linear model. The DEGs with the fold change value larger than 2 and p-value less than 0.05 were selected.

Co-expression analysis

For demonstrating the potential regulatory relationship, the Pearson correlation coefficient (PCC) was calculated for all pair-wise comparisons of gene-expression values between TFs and the DEGs. The regulatory relationships whose absolute PCC are larger than 0.6 were considered as significant.

Gene ontology (GO) analysis

The BiNGO analysis (Maere et al., 2005) was used to identify over-represented GO categories in biological process. The result of gene ontology analysis was shown in Table 2.

Regulation network construction

Using the regulation data that have been collected from TRANSFAC database and TRED database, we matched the relationships between differentially expressed TFs and its differentially expressed target genes. Based on the aforementioned two regulation datasets and the pathway relationships of the target genes, we build the regulation networks by Cytoscape (Shannon et al., 2003). Based on the significant relationships (PCC > 0.6 or PCC < -0.6) between TFs and its target genes, 46 putative regulatory relationships were predicted between 7 TFs and 41 target genes. The result of regulation network is shown in Figure 1.

Significance analysis of pathway

We adopted an impact analysis that includes the statistical significance of the set of pathway genes but also considers other crucial factors such as the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions, etc. (Draghici et al., 2007). In this model, the impact factor (IF) of a pathway Pi is calculated as the sum of two terms:

\[
IF(Pi) = \log\left(\frac{1}{pi}\right) + \sum_{g \in \pi} \left| \frac{PF(g)}{\Delta E \cdot \text{N}_s(Pi)} \right|
\]

The first term is a probabilistic term that captures the significance of the given pathway Pi, from the perspective of the set of genes contained in it.
Table 2. Gene ontology biological process analysis.

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>Description</th>
<th>Count</th>
<th>p-Value</th>
<th>Correlate p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>42221</td>
<td>Response to chemical stimulus</td>
<td>28</td>
<td>1.28E-17</td>
<td>1.55E-14</td>
</tr>
<tr>
<td>42127</td>
<td>Regulation of cell proliferation</td>
<td>19</td>
<td>1.37E-12</td>
<td>8.30E-10</td>
</tr>
<tr>
<td>50896</td>
<td>Response to stimulus</td>
<td>33</td>
<td>7.20E-12</td>
<td>2.90E-09</td>
</tr>
<tr>
<td>6950</td>
<td>Response to stress</td>
<td>24</td>
<td>2.19E-11</td>
<td>6.62E-09</td>
</tr>
<tr>
<td>9611</td>
<td>Response to wounding</td>
<td>15</td>
<td>3.19E-11</td>
<td>7.73E-09</td>
</tr>
<tr>
<td>9719</td>
<td>Response to endogenous stimulus</td>
<td>14</td>
<td>1.24E-10</td>
<td>2.50E-08</td>
</tr>
<tr>
<td>6952</td>
<td>Defense response</td>
<td>15</td>
<td>2.17E-10</td>
<td>3.74E-08</td>
</tr>
<tr>
<td>9725</td>
<td>Response to hormone stimulus</td>
<td>13</td>
<td>4.15E-10</td>
<td>6.28E-08</td>
</tr>
<tr>
<td>9605</td>
<td>Response to external stimulus</td>
<td>14</td>
<td>5.33E-10</td>
<td>7.16E-08</td>
</tr>
<tr>
<td>48545</td>
<td>Response to steroid hormone stimulus</td>
<td>10</td>
<td>1.18E-09</td>
<td>1.43E-07</td>
</tr>
</tbody>
</table>

Figure 1. Regulation network in thermal injury of human skin.

It is obtained by using the hyper geometric model in which $p_i$ is the probability of obtaining at least the observed number of differentially expressed gene, $N_{de}$, just by chance (Tavazoie et al., 1999; Draghici et al., 2003).

The second term is a functional term that depends on the identity of the specific genes that are differentially expressed as well as on the interactions described by the pathway (that is, its topology). The second term sums up the absolute values of the perturbation factors (PFs) for all genes $g$ on the given pathway $P_i$.

The PF of a gene $g$ is calculated as follows:

$$PF(g) = \Delta E(g) + \sum_{u \in U_{g}} \beta_{ug} \cdot \frac{PF(u)}{N_{u}(u)}$$

In this equation, the first term $\Delta E(g)$ captures the quantitative information measured in the gene expression experiment. The factor $\Delta E(g)$ represents the normalized measured expression change of the gene $g$. The first term $\Delta E(g)$ in the equation is a sum of all PFs of the genes $u$ directly upstream of the target gene $g$. 
normalized by the number of downstream genes of each such gene $N_{ds}(u)$, and weighted by a factor $\beta_{ug}$, which reflects the type of interaction: $\beta_{ug} = 1$ for induction, $\beta_{ug} = -1$ for repression (KEGG supply this information about the type of interaction of two genes in the description of the pathway topology). $US_p$ is the set of all such genes upstream of $g$. We need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, $N_{de}(P)$. In order to make the IFS as independent as possible from the technology, and also comparable between problems, we also divide the second term in Equation 1 by the mean absolute fold change $\Delta E$, calculated across all differentially expressed genes.

RESULTS

Regulation network construction in thermal injury of human skin

To get pathway-related DEGs of thermal injury of human skin, we obtained publicly available microarray data sets GSE8056 from GEO. After microarray analysis, the differentially expressed genes with the fold change value larger than 2 of GSE8056 and p-value less than 0.05 were selected. GSE8056 are separated in 3 groups by time course. At last, the overlap of 3 groups DEGs was calculated, and 1856 genes were selected as DEGs from GSE8056. To get the regulatory relationship, the co-expressed value (PCC $\geq 0.6$) was chosen as the threshold. Finally, we got 46 regulatory relationships between 7 TFs and their 41 differentially expressed target genes. By integrating the regulatory relationships aforementioned, a regulation network of burn wound was built between TFs and its target genes (Figure 1). In this network, EST1, ESR1 and H1F1A with higher degrees form a local network suggesting that these genes may play an important role in burn wound. Besides, transcription factor ETS1 regulates FOSL1 and both of them regulate the target gene PLAU. ETS1 regulate ADH1B and target gene IL6 through the ANPEP.

GO analysis of the regulation network in burn wound

Several gene ontology (GO) categories were enriched among these genes in the regulatory network, including defense response, response to stimulus, response to wounding and response to stress (Table 2, list of top 10).

<table>
<thead>
<tr>
<th>Pathway name</th>
<th>Impact factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukocyte transendothelial migration</td>
<td>621.256</td>
</tr>
<tr>
<td>Cell adhesion molecules (CAMs)</td>
<td>467.021</td>
</tr>
<tr>
<td>Phosphatidylinositol signaling system</td>
<td>27.306</td>
</tr>
<tr>
<td>Cytokine-cytokine receptor interaction</td>
<td>25.907</td>
</tr>
<tr>
<td>Allograft rejection</td>
<td>15.453</td>
</tr>
</tbody>
</table>

Significant pathway analysis in burn wound

To identify the relevant pathways changed in burn wound, we used a statistical approach on pathway level. Significance analysis at single gene level may suffer from the limited number of samples and experimental noise that can severely limit the power of the chosen statistical test. Pathway can provide an alternative way to relax the significance threshold applied to single genes and may lead to a better biological interpretation. So, we adopted a pathway based impact analysis method that contained many factor including the statistical significance of the set of differentially expressed genes in the pathway, the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions and so on. The impact analysis method yields many significant pathways contained Leukocyte transendothelial migration, cell adhesion molecules (CAMs), phosphatidylinositol signaling system and so on (Table 3).

DISCUSSION

From the result of regulation network construction in thermal injury of human skin, we could found that many TFs and pathways closely related with thermal injury have been linked by our method. The gene ESR1, HIF1, CEBP, CD13, IL8 TIMP1, PLAU are also a hub nod in our transcriptome network and have a close relationship with thermal injury proved by previous study. Although the role of CREM and EGLN3 in thermal injury has not been investigated to date; some evidence also suggests that CREM and EGLN3 may play an important role in the function of skin.

CREM encodes a bZIP transcription factor that binds to the cAMP responsive element found in many viral and cellular promoters. It is an important component of cAMP-mediated signal transduction during the spermatogenetic cycle, as well as other complex processes. Alternative promoter and translation initiation site usage allows this gene to exert spatial and temporal specificity to cAMP responsiveness. Multiple alternatively spliced transcript variants encoding several different isoforms have been found for this gene, with some of them functioning as activators and some as repressors of transcription. Transcriptional regulator that binds the cAMP response element (CRE), a sequence present in many viral and
cellular promoters. Isoforms are either transcriptional activators or repressors, plays a role in spermatogenesis and is involved in spermatid maturation. Though CREM is not directly related to burn, some evidence also suggests that CREM may play an important role in the function of skin. For example, transcriptome analyses revealed a discriminating gene expression profile in human CD34+ progenitor-derived dendritic cells (DC) after exposure to skin sensitizers versus non-sensitizers. By comparing their responsiveness towards a non-sensitizing danger signal and a sensitizer, CREM appears to display a specific response. Research indicate that CREM may be functionally involved in sensitizer-induced DC activation (Lambrechts et al., 2011).

ESR1 encodes an estrogen receptor, a ligand-activated transcription factor composed of several domains important for hormone binding, DNA binding, and activation of transcription. The protein localizes to the nucleus where it may form a homodimer or a heterodimer with estrogen receptor 2. Estrogen and its receptors are essential for sexual development and reproductive function, but also play a role in other tissues such as bone. Estrogen receptors are also involved in pathological processes including breast cancer, endometrial cancer, and osteoporosis. A study investigated the effect of cutaneous bromine vapor, an industrial chemical that causes severe cutaneous burns, exposure on gene expression using a weanling swine burn model by microarray analysis found that the transcripts encoding ESR1 were identified using IPA as common potential therapeutic targets for Phase II/III clinical trial or FDA-approved drugs (Price et al., 2011).

Hypoxia-inducible factor-1 (HIF1) is a transcription factor found in mammalian cells cultured under reduced oxygen tension that plays an essential role in cellular and systemic homeostatic responses to hypoxia. In a study, mice that were heterozygous (HET) for a null allele at the locus encoding the HIF1 subunit (HET mice) and their wild-type (WT) littermates were subjected to a thermal injury involving 10% of the body surface area. HIF1 protein levels were increased in burn wounds of WT but not of HET mice on day 2. Researches delineate the impaired angiogenesis and mobilization of circulating angiogenic cells in HIF1 heterozygous-null mice after burn wounding (Zhang et al., 2010). Researchers found that cell apoptosis and expression of hypoxia-inducible transcription factor-1 alpha in kidney tissue after severe burn with delayed fluid resuscitation in rats in areas of different altitude (Jiang et al., 2008). Schwacha provided evidence proved that burn injury- induced alterations in wound inflammation and healing are associated with suppressed hypoxia inducible factor-1 alpha expression (Schwacha et al., 2008).

CEBP protein encoded by this intronless gene is a bZIP transcription factor which can bind as a homodimer to certain promoters and enhancers. It can also form heterodimers with the related proteins CEBP-beta and CEBP-gamma. Study shows that a decrease in C/EBPalpha mRNA levels occurs in response to thermal injury and suggests the C/EBP transcription factor genes respond to thermal injury (Gilpin et al., 1996).

CD13 was thought to be involved in the metabolism of regulatory peptides by diverse cell types, including small intestinal and renal tubular epithelial cells, macrophages, granulocytes, and synaptic membranes from the central nervous system (CNS).

The protein IL8 is a member of the CXC chemokine family. This chemokine is one of the major mediators of the inflammatory response. It functions as a chemoattractant, and is also a potent antigenic factor. Ulrich measured the concentration of cytokine IL8 in plasma from 27 patients with large burns and found that the mean patient plasma concentration of IL-8 was about 60 times higher than that of healthy controls. The increased IL-8 concentrations seem to be related to burn size and to have a role in the pathophysiology of sepsis in patients with large burns. The large amounts of circulating IL-8 following thermal injury may contribute to the strong and sustained activation of neutrophils reported earlier in patients with large burns (Vindenes et al., 1995).

TIMP1 belongs to the TIMP gene family. The proteins encoded by this gene family are natural inhibitors of the matrix metalloproteinases (MMPs), a group of peptidases involved in degradation of the extracellular matrix. In addition to its inhibitory role against most of the known MMPs, the encoded protein is able to promote cell proliferation in a wide range of cell types, and may also have an anti-apoptotic function. Transcription of this gene is highly inducible in response to many cytokines and hormones. The elevated systemic TIMP-1 concentration might contribute to tissue fibrosis, leading to pathological scar formation. The increase of PIIINP after thermal trauma indicates a fibrogenic component of wound healing (TIMP-1, MMP-2, MMP-9, and PIIINP as Serum Markers for Skin Fibrosis in Patients following Severe Burn Trauma). Ulrich found that the elevated systemic TIMP-1 concentration might contribute to tissue fibrosis, leading to pathological scar formation. The increase of PIIINP after thermal trauma indicates a fibrogenic component of wound healing (Ulrich et al., 2003).

PLAU encodes a serine protease involved in degradation of the extracellular matrix and possibly tumor cell migration and proliferation. A present study identifies specific gene PLAU may be useful as potential therapeutic targets to promote improved wound healing (Price et al., 2009). Though CREM and EGLN3 are not directly related to burn, some evidence also suggests that CREM and EGLN3 may play an important role in the function of skin. EGLN3 is involved in the regulation of hypoxia-inducible factor (HIF) through a negative feedback loop (del Peso et al., 2003) and HIF-1 promotes angiogenesis during burn wound healing (Zhang et al., 2010). Egln3 as a key factor of HIF activity in epidermal keratinocytes that it might be related to the thermal injury.
The impact analysis method yields many significant pathways including, leukocyte transendothelial migration pathway, cell adhesion molecules (CAMs) pathway, phosphatidylinositol signaling system pathway, cytokine-cytokine receptor interaction pathway and Allograft rejection pathway. Previous study proves leukocyte transendothelial migration pathway and cell adhesion molecules (CAMs) pathway (Ley et al., 2007). The observed inflammatory response of soluble cell adhesion molecules could be useful in monitoring endothelial activation immediately following thermal injury (Rassoul et al., 2009). That proved the CAM pathways might be activated by the burn wound. Leukocyte migration from the blood into tissues is vital for immune surveillance and inflammation. Thermal injury alters the expression of leukocyte adhesion molecules (intercellular adhesion molecule 1 (ICAM-1), E-selectin, and leukocyte CD11a) (Dressler et al., 1997).

The basic understanding of the mechanisms underlying the functioning of burn related gene is important. A deeper understanding of transcription factors and their regulated genes remain an area of intense research activity in futures. Our regulation network is useful in investigating the complex interacting mechanisms of transcription factors and their regulated genes. We also predicted that CREM and EGLN3 was burn related gene. However, further experiments are still needed to confirm the conclusion.

REFERENCES