



Identification of marker genes in diabetic wounds by DNA microarray study

T. Ni¹, N. Wang¹, Z.G. Mao¹ and M. Yao^{1,2}

¹Department of Burn and Plastic Surgery,
No. 3 People's Hospital and Institute of Traumatic Medicine,
School of Medicine, Shanghai Jiao Tong University, Shanghai, China

²Wellman Center for Photomedicine, Harvard Medical School,
Massachusetts General Hospital, Boston, MA, USA

Corresponding author: M. Yao
E-mail: yaominyaomin268@gmail.com

Genet. Mol. Res. 12 (4): 5348-5355 (2013)

Received November 27, 2012

Accepted April 16, 2013

Published November 7, 2013

DOI <http://dx.doi.org/10.4238/2013.November.7.9>

ABSTRACT. This study aimed to identify marker genes in diabetic wounds using a dataset based on a DNA microarray of dermal lymphatic endothelial cells, and our results provide a basic understanding of diabetic wounds through further study of these differentially expressed genes (DEGs). From the Gene Expression Omnibus database, we downloaded a gene expression microarray (GSE38396) that includes 8 samples: 4 normal controls and 4 disease samples (type II diabetes). We then identified genes that were differentially expressed between normal and disease samples using packages in R language, constructed a protein-protein interaction (PPI) network and analyzed modules in the network. In addition, phylogenetic analysis was performed by MEGA to find the most conserved genes. Two hundred and thirteen genes were identified as being differentially expressed between normal and disease samples, and we constructed a PPI network that included 213 pairs of proteins. We then identified a module including 20 genes, the function of which was

significantly enriched in wounding response. Lastly, the most conserved genes, CD44 and CCL5, were identified through phylogenetic analysis. In summary, we found differentially expressed marker genes, a wounding response-related module, and the most important genes CD44 and CCL5. Our findings suggest new approaches to therapies for diabetic wounds.

Key words: Diabetic wounds; Module analysis; Phylogenetic analysis

INTRODUCTION

With the development of a global economy, living standards have improved markedly at the same time as eating habits have clearly deteriorated. The global prevalence of diabetes has increased as a consequence of this situation (Unsal et al., 2012). Nowadays, more and more researchers are focusing on the diabetes, which is a complex disease with multiple symptoms, including the sensitivity to injury of patients' skins. Wound healing is a complex biological process involving several overlapping phases including inflammation, proliferation, and cellular remodeling (Baker and Leaper, 2003). Diabetic ulcers are refractory wounds because of their long disease duration and easy recurrence (Pierce, 2001). Moreover, diabetic wounds often deteriorate into intractable ulcers. Therefore, diabetic ulcer wounds are clinical problems which need to be solved (Ackermann and Marrero, 2007). With the development of biological science, research on diabetic wound healing has become more in-depth. Current studies with animal models show exogenous neuropeptides (sP) can significantly accelerate healing of refractory diabetic wounds and improve the healing quality (Gibran et al., 2002). Nilsson et al. (1985) confirmed that substance P can stimulate tissue cell growth and promote DNA synthesis. In addition, increasing expression of advanced glycation end-products in the membranes of fibroblasts can cause their injury and dysfunction (Goova et al., 2001). However, studies screening differentially expressed genes (DEGs) in diabetic wounds using bioinformatics methods remain rare. Nevertheless, genome-wide approaches have reshaped the landscape of the problem, and meta-analysis of multi-study data has allowed the identification of differentially expressed genes that may have biomarker potential (Chan et al., 2008).

In this study, 213 genes were identified as being differentially expressed between normal and disease samples using a DNA microarray from dermal lymphatic endothelial cells, following which a protein-protein interaction (PPI) network was constructed. Cytoscape and its plugin, Mcode, were used to visualize the network and identify modules within it. In addition, phylogenetic analysis was performed by the Molecular Evolutionary Genetics Analysis software (MEGA) to find the most conserved genes, which were CD44 and CCL5. Many earlier studies have shown that CD44 and CCL5 are involved in diabetic wound healing (Falanga, 2005; Gallagher et al., 2007; Dulmovits and Herman, 2012).

MATERIAL AND METHODS

Affymetrix microarray profile

Gene expression profiles of dermal lymphatic endothelial cells (GSE38396) were downloaded from the Gene Expression Omnibus database. These comprised four normal controls and four disease samples (type II diabetes). Platform information: GPL6244 [(HG-

U133_Plus_2) Affymetrix Human Genome U133 Plus2.0 Array]. Annotation information for all ATH1 (25 K) probe sets is provided by Affymetrix.

Data processing and differentially expressed gene analysis

Package Affy (Troyanskaya et al., 2001; Fujita et al., 2006) in R language was used to pretreat the datasets. The limma package (Smyth, 2004) was then used to identify DEGs between disease and normal samples ($P < 0.05$, $|\logFC| > 1$). Multiple test correction was performed using Bias methods (Benjamini and Hochberg, 1995).

Selection of hub gene product

String 8.3 (Szklarczyk et al., 2011) was used to predict the interaction of these DEGs. Pairs with a combination score larger than 0.5 were recorded. The interaction network was analyzed, and the degree of each node calculated.

Analysis of modules in the interaction network

Gene products in the same module often have the same or similar roles (Table 1), working together to carry out their biological function. Therefore, Cytoscape and its plugin Mcode were used to visualize the network and identify the modules within (Bader and Hogue, 2003; Smoot et al., 2011). Parameters: Degree cutoff ≥ 2 (each node in the module was larger than 2 degrees), K-core ≥ 2 (more than 2 subgraphs of each node in the module). Then Bingo (Maere et al., 2005) was used to determine which gene ontology (GO) terms are significantly overrepresented in the genes in our module (adj $P < 0.01$).

Phylogenetic analysis of module genes

Phylogenetic analysis was performed using the MEGA software (Tamura et al., 2011) to find the most conserved genes in the module. These are regarded as the most important genes in our study. The MEGA software is multifunctional, providing many convenient facilities for the assembly of sequence data sets from files or web-based repositories. It includes tools for visual representation of the results in the form of interactive phylogenetic trees and evolutionary distance matrices. The maximum likelihood algorithm was used to construct a phylogenetic tree.

RESULTS

DEG selection

After data processing (Figure 1), differential expression analysis was performed. Two hundred and thirteen genes were selected as being differentially expressed ($P < 0.05$, $|\logFC| > 1$): 89 were downregulated and 124 upregulated. The 213 DEGs are listed in Table 1. As show in Figure 1, the medians of each set of data are similar, demonstrating that the degree of standardization is optimal.

Table 1. Function list of genes in module.

GO-ID	P	corr P-value	Description
9611	1.25E-07	1.02E-04	Response to wounding
32502	1.06E-06	1.85E-04	Developmental process
48513	1.58E-06	1.90E-04	Organ development
42221	2.49E-06	2.43E-04	Response to chemical stimulus
10033	4.55E-06	2.63E-04	Response to organic substance
48731	3.65E-06	2.63E-04	System development
50896	4.76E-06	2.63E-04	Response to stimulus
9653	5.48E-06	2.81E-04	Anatomical structure morphogenesis
48856	9.94E-06	4.30E-04	Anatomical structure development
6950	1.42E-05	5.56E-04	Response to stress
7275	3.31E-05	9.72E-04	Multicellular organismal development
48522	2.99E-04	3.67E-03	Positive regulation of cellular process
32501	3.31E-04	3.88E-03	Multicellular organismal process
16043	3.59E-04	4.10E-03	Cellular component organization
48518	6.28E-04	6.30E-03	Positive regulation of biological process

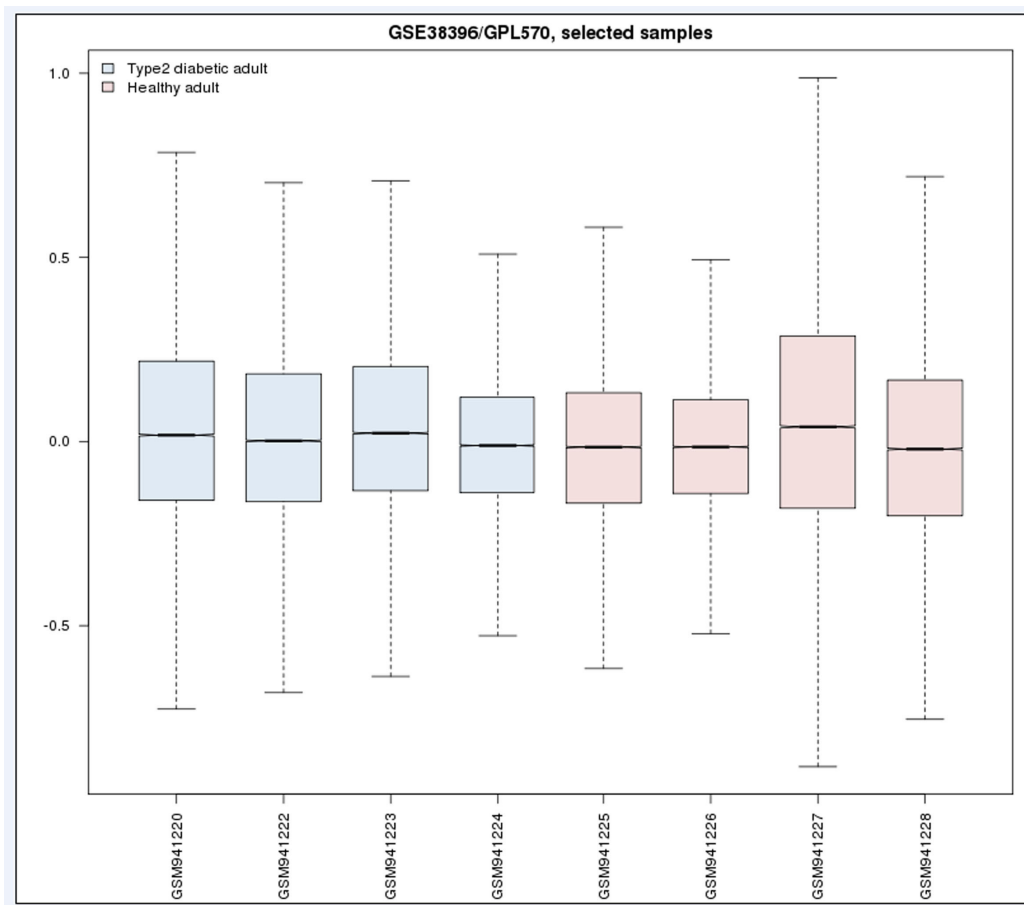


Figure 1. Boxplot of expression data. The blue and pink boxes represent the normal and disease samples, respectively. The black lines in the box represent the median of each set of data.

Construction of a protein interaction network

Combination score was calculated by STRING 8.3 for all differentially expressed genes, and pair records with combination scores larger than 0.5 were chose as DEGs. An interaction network was created which includes 213 pairs of proteins. In addition, we calculated the degree of distribution of the whole network, as shown in Figure 2A.

Analysis of modules in the interaction network

Cytoscape and its plugin, Mcode, were used to visualize the network and identify the modules within. (Parameters are listed in Material and Methods). One module was identified in which there were 20 differentially expressed genes: six upregulated and fourteen downregulated. Bingo was performed to assess the functional enrichment of the module based on the GO pathway annotation terms (adj $P < 0.01$). Fifteen significantly enriched functions were found of which response to wounding was the most significant in our study.

Phylogenetic analysis of module genes

Phylogenetic analysis was performed and a phylogenetic tree constructed using MEGA, as shown in Figure 3. From Figure 3, we learn that genes in the module are associated with each other in the evolutionary processes. In this study, they were divided into several parts according to the distances between branches. Among the genes, CCL5 was closest to the root of the phylogenetic tree. Considering CCL5 still persists after a long genetic evolutionary process, we concluded that the function and the structure of CCL5 are important.

The degree of distribution of module genes is shown in the Figure 2B. Nodes with a high degree are less common, but most of the nodes in our module are 7 degrees or higher, with CD44, at 24, being the highest.

From the phylogenetic tree branches, we can conclude that the phylogenetic relationship of genes CD44 and CCL5 is very close. Furthermore, they are in the same branch. All the evidence suggests there is functional and structural similarity between CD44 and CCL5.

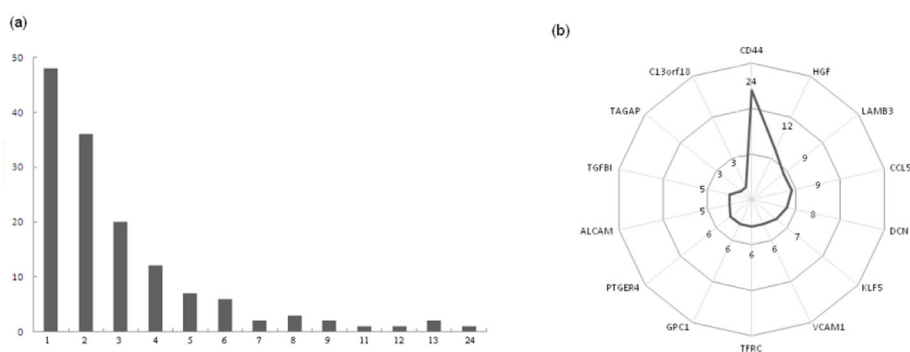


Figure 2. **A.** Degree distribution of interaction network. X coordinate axis refers to degree, and Y coordinate axis refers to number of genes. **B.** Degree of genes in the module. Names around the circle are gene names, and numbers are degree of genes.

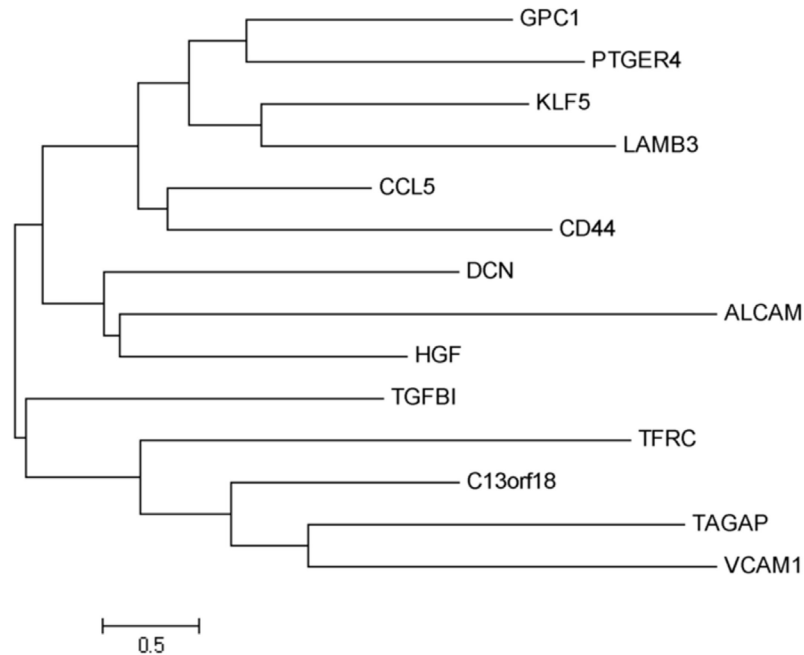


Figure 3. Phylogenetic tree for module genes. The length scale is shown left corner. Bar = 0.5 my (million year).

DISCUSSION

In this study, using a dataset based on a DNA microarray from dermal lymphatic endothelial cells, 213 genes were identified as being differentially expressed between normal and disease samples. Considering that single genes cannot act in isolation (Gauderman, 2002), we looked at how these DEGs execute their regulatory functions through interactions with other genes by constructing a PPI network. Modules in the network were identified by Cytoscape.

To identify the most conserved genes in the module, phylogenetic analysis was performed using MEGA. The MEGA software is a desktop application designed for comparative analysis of homologous gene sequences either from multigene families or from different species. It places a special emphasis on inferring evolutionary relationships and patterns of DNA and protein evolution (Kumar et al., 2008). This analysis identified CD44 and CCL5, which have been shown in many earlier studies to be involved in diabetic wound healing

Thus, through construction of an interaction network and phylogenetic analysis, functional modules were identified along with the most important genes, CD44 and CCL5. From the module function, we concluded that CD44 and CCL5 are most likely to be related to diabetic wound healing (Gallagher et al., 2007; Dulmovits and Herman, 2012).

The CD44 antigen is a cell-surface glycoprotein involved in cell-cell interactions, cell adhesion and migration. In humans, the CD44 antigen is encoded by the CD44 gene on chromosome 11 (Haynes et al., 1989) and plays important roles in oocyte maturation, fertilization and preimplantational embryo development (Luz et al., 2012). CD44 can also affect the regulation of the wound repair process. In particular, CD44 expression is upregulated in non-insulin-de-

pendent diabetes mellitus, and the increased expression is considered to affect the properties of fibroblasts. Specific properties of fibroblasts may be involved in the increased susceptibility of patients with diabetes to chronic ulceration (Yevdokimova and Podpryatov, 2005). When considered with other research (Cha and Falanga, 2007; Schwarz et al., 2013), our study indicates that fibroblasts could potentially become a therapeutic cell population in diabetic wounds.

CCL5 is an 8 kDa protein recognized as a kind of chemotactic cytokine for T cells and basophils, and it often plays an important role in recruiting leukocytes into inflammatory sites (Donlon et al., 1990). CCL5 is expressed in healing wounds, and abnormal expression or activity of CCL5 is associated with impaired healing (Gillitzer and Goebeler, 2001). Some previous studies reported that CCL5 might recruit leukocytes to execute their function in healing wounds. Sustained induction of chemokines such as CCL5 in diabetic mice often leads to wounds becoming refractory (Wetzler et al., 2000). These findings may be very valuable in the design of new drugs.

Molecular biological, genetic, and other experimental techniques are continually being updated. Progress in these fields can provide us with new methods and new ideas for the treatment of refractory diabetic wounds. In this study, 213 DEGs were identified. Through PPI network construction and phylogenetic analysis, CD44 and CCL5 were identified as the genes most likely to play an important role in diabetic wound healing. Although more experiments are needed to verify these results, our findings increase the likelihood that improved treatments will be developed for diabetic and other chronic wounds.

REFERENCES

- Ackermann RT and Marrero DG (2007). Adapting the Diabetes Prevention Program lifestyle intervention for delivery in the community: the YMCA model. *Diabetes Educ.* 33: 69, 74-69, 78.
- Bader GD and Hogue CW (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2.
- Baker EA and Leaper DJ (2003). Profiles of matrix metalloproteinases and their tissue inhibitors in intraperitoneal drainage fluid: relationship to wound healing. *Wound Repair Regen.* 11: 268-274.
- Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Society B* 289-300.
- Cha J and Falanga V (2007). Stem cells in cutaneous wound healing. *Clin. Dermatol.* 25: 73-78.
- Chan SK, Griffith OL, Tai IT and Jones SJM (2008). Meta-analysis of colorectal cancer gene expression profiling studies identifies consistently reported candidate biomarkers. *Cancer Epidemiol. Biomarkers Prev.* 17: 543-552.
- Donlon TA, Krensky AM, Wallace MR, Collins FS, et al. (1990). Localization of a human T-cell-specific gene, RANTES (D17S136E), to chromosome 17q11.2-q12. *Genomics* 6: 548-553.
- Dulmovits BM and Herman IM (2012). Microvascular remodeling and wound healing: a role for pericytes. *Int. J. Biochem. Cell Biol.* 44: 1800-1812.
- Falanga V (2005). Wound healing and its impairment in the diabetic foot. *Lancet* 366: 1736-1743.
- Fujita A, Sato JR, Rodrigues LO, Ferreira CE, et al. (2006). Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7: 469.
- Gallagher KA, Liu ZJ, Xiao M, Chen H, et al. (2007). Diabetic impairments in NO-mediated endothelial progenitor cell mobilization and homing are reversed by hyperoxia and SDF-1 alpha. *J. Clin. Invest.* 117: 1249-1259.
- Gauderman WJ (2002). Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* 155: 478-484.
- Gibran NS, Jang YC, Isik FF, Greenhalgh DG, et al. (2002). Diminished neuropeptide levels contribute to the impaired cutaneous healing response associated with diabetes mellitus. *J. Surg. Res.* 108: 122-128.
- Gillitzer R and Goebeler M (2001). Chemokines in cutaneous wound healing. *J. Leukoc. Biol.* 69: 513-521.
- Goova MT, Li J, Kislinger T, Qu W, et al. (2001). Blockade of receptor for advanced glycation end-products restores effective wound healing in diabetic mice. *Am. J. Pathol.* 159: 513-525.

- Haynes BF, Telen MJ, Hale LP and Denning SM (1989). CD44 - a molecule involved in leukocyte adherence and T-cell activation. *Immunol. Today* 10: 423-428.
- Kumar S, Nei M, Dudley J and Tamura K (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* 9: 299-306.
- Luz JV, Alcantara-Neto AS, Batista RI, Souza JM, et al. (2012). Expression of CD44 in sheep oocytes and preimplantation embryos. *Genet. Mol. Res.* 11: 799-809.
- Maere S, Heymans K and Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449.
- Nilsson J, Von Euler AM and Dalsgaard CJ (1985). Stimulation of connective tissue cell growth by substance P and substance K. *Nature* 315: 61-63.
- Pierce GF (2001). Inflammation in nonhealing diabetic wounds: the space-time continuum does matter. *Am. J. Pathol.* 159: 399-403.
- Schwarz F, Jennewein M, Bubel M, Holstein JH, et al. (2013). Soft tissue fibroblasts from well healing and chronic human wounds show different rates of myofibroblasts in vitro. *Mol. Biol. Rep.* 40: 1721-1733.
- Smoot ME, Ono K, Ruscheinski J, Wang PL, et al. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431-432.
- Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3: Article3.
- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids. Res.* 39: D561-D568.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525.
- Unsal A, Koc Y, Basturk T, Akgun AO, et al. (2012). Risk factors for progression of renal disease in patient with diabetic nephropathy. *Eur. Rev. Med. Pharmacol. Sci.* 16: 878-883.
- Wetzler C, Kampf H, Stallmeyer B, Pfeilschifter J, et al. (2000). Large and sustained induction of chemokines during impaired wound healing in the genetically diabetic mouse: prolonged persistence of neutrophils and macrophages during the late phase of repair. *J. Invest. Dermatol.* 115: 245-253.
- Yevdokimova NY and Podpryatov SE (2005). Hyaluronic acid production and CD44 expression in cultured dermal fibroblasts of patients with non-insulin-dependent diabetes mellitus with and without chronic ulcers on the lower extremity. *Wound Repair Regen.* 13: 181-188.