Material and Methods

Data Collection and Preprocessing on Expression Profiles

All public expression profiling data were searched from GEO data sets according to the keywords “psoriasis,” “lesion,” and “*Homo sapiens*” on April 9, 2016. The data used in the latter analyses had a distinguishing feature of definite gene expression profiles of psoriatic skins, including lesion and nonlesion samples. We found 4 chip data sets which we adopted for subsequent analysis. The data details are shown in Table 1.

Raw data were downloaded in CEL format from the GEO database. The background correction and normalization of the expression profile data were processed by using Affy packets in R language, including the conversion of the original data format, the missing value supplement by using the median method, the background correction by using the MAS method, and the data standardization by using the quantile method [16]. Each probe corresponded to a gene through the annotation platform of each group of expression profiles after elimination of the unloaded ones. As to a single gene corresponding to many probes, the mean value served as its initial expression value.

Detection of Candidate Feature Factors Using Meta-Analysis

In order to detect credible genes and correct the bias from diverse test platforms, objective quality controls were first processed by using the quality control standard in

the MetaQC package [17], and then evaluation and selection of data sets were further made by using principal component analysis and standardized mean rank. There were 4 established quality control specifications in MetaQC package: (1) the internal quality control, which was used to examine the structural homogeneity of gene expression values among the objects in data sets; (2) the external quality control, which was used for a consistency check of expression structures of the pathway database; (3) the accuracy quality control, which was responsible for the accuracy exploration of significantly informative genes and pathway recognition; (4) the consistency quality control of distinguishing feature genes or signal pathways.

By using MetaDE.ES in the MetaDE package, genes with significantly differential expression were selected from the obtained authentic data under the strict examination of MetaQC quality control. The MetaDE.ES method performed tests for heterogeneity of gene expression values drawn from different platforms, and then statistical magnitude parameters for heterogeneity judgment were received, such as tau2, Q value and Qpval. First of all, when the tau^2 value is zero, or the Q value follows the χ^2 test of the degrees of freedom equal to K--1 while Qpval is more than 0.05, the research samples are confirmed to be homogeneous and unbiased by the software. Genes of samples are identified to present significantly differential expression on condition that the threshold value of the false discovery rate is set at 0.05.

Screening of Disease Associated-Modules and Genes

With the help of weighted gene coexpression network analysis (WGCNA) [18, 19] based on meta-analysis, we carried out network construction and module

discriminatory analysis of expression values of the aforementioned remarkable feature genes to search target modules and genes related to psoriasis. The details of confluence analysis are as described below.

Consistency Analysis among Data Sets

We acquired expression values of the prior significantly expressed genes incorporated into data sets and then compared the correlation of these values between 2 groups of data sets. Excellent consistency of gene expression values among data sets was shown by high correlation, which was the prerequisite for construction of a WGCNA network in meta-analysis.

Definition of Correlation Matrix of Gene Coexpression

The index of pairwise correlation between genes is the element in the correlation matrix of gene coexpression. The correlation coefficient of a gene pair (m and n) is computed by the formula $S_{mn} = |cor_{(m,n)}|$, and then multiple coefficient values constitute the correlation matrix of gene coexpression.

Definition of Adjacency Function

The immediate function divides gene pairs into the relevant or irrelevant groups by assigning threshold values of gene correlation coefficients, such as $S_{mn} = 0.8$. This defining method is simple and practicable, but will cause a massive loss of information. For this reason, the power adjacency function a_{mn} is used as an indicator to measure the relationship between genes in the WGCNA algorithm. The equation is $a_{mn} = power_{(S_{mn},\beta)}$, meaning the exponential weighting of correlation coefficients.

Parameter Determination of Adjacency Function

In this study, the weighting coefficient β was determined according to the principle of scale-free networks. There existed 2 values $\log_2 k$ and $\log_2 p(k)$, meaning number of join nodes and probability of node appearance were taken by logarithm, respectively. During the determination process, the correlation coefficient of the 2 aforementioned logarithm values was at least 0.9.

Dissimilarity Degree among Nodes

After the determination of parameter β，the following was the transformation of the correlation matrix S_{mn} into the adjacency matrix a_{mn}. Considering the relationship of one gene and all the others in an analysis, a_{mn} was transformed into the topology matrix Ω ($\Omega = w_{mn}$). Elements in the matrix are shown below:

$$w_{mn} = \frac{l_{mn}+a_{mn}}{\min\{k_m,k_n\}+1-a_{mn}}.$$

In this equation, l_{mn} is the area sum of adjacency coefficients of nodes connective to genes of m and n; k_m is the sum value of adjoining numbers of nodes in single connection with gene m. If genes m and n were connectionless to each other, even to any other genes, the w_{mn} value would be zero. Meanwhile, the dissimilarity degree of nodes (dmn) was measured by the equation: dmn = 1 – w_{mn}, which is the foundation of network construction.

Identification of Gene Modules

Gene modules were identified using the dynamic hybrid cutting method [20], which is a “bottom-up” algorithm, and constituted of identification and test procedures. In this algorithm, there were 4 types of modules to be recognized: first, meeting of

number of module genes with the preset minimum value; second, deletion of branches with hypertelorism from settled gene modules; third, significant differences among gene modules; fourth, firm link within key genes of modules. The unallocated genes were clustered into preliminary modules that finally participated in the formation of the gene network structure. In the present study, we used this algorithm to build hierarchical clustering trees on the basis of gene discrepancy coefficients. Different branches of cluster numbers represented corresponding gene modules.

Correlation Identification between Network Module and Disease Status

As to grouping phenotypic data (such as disease status), we evaluated the expression significance of genes (p value) from different analysis groups such as disease group and normal group, then defined gene significance as a value of $\log_{10} P$, and finally defined module significance as the mean value of gene significance.

Network Construction of Interaction with Genes in Disease-Related Modules

The interaction information of human genes from databases of BioGRID, HPRD, and BIND was first obtained, and then the distinguishing genes mentioned above were mapped to the interaction network of human proteins. The interaction network was constructed by the relation edges of interaction.

Searching for Drug Molecules Related to Interactive Genes

In this section, we uploaded feature factors in network modules which were significantly associated with disease to a connectivity map, found drug molecules

correlated with gene performance and disease, using differential gene expression of human cells stimulated by drugs, and selected the molecules with high correlation scores (>0.8). Moreover, we mapped those drug molecules and their target genes to our previously built network, constructed drug-targeted interaction networks, and visualized them by Cytoscape.

GO Analysis and Pathway Enrichment

The biological processes and KEGG pathways of module genes which were mapped to functional GO nodes were analyzed by using a hypergeometric distribution algorithm. The significance threshold *p* value was set at 0.05, and the significantly enriched functional genes were retained with $p < 0.05$. The equation is shown below:

$$\mathrm{p} = 1 - \sum_{i=0}^{H-1} \frac{\binom{\mathrm{M}}{\mathrm{i}}\binom{\mathrm{N-M}}{\mathrm{K-H}}}{\binom{\mathrm{N}}{\mathrm{K}}},$$

with N equaling the number of genes with GO function annotation, K the number of genes of significantly differential expression, M the number of genes annotated to special GO function nodes, and i representing an annotated gene.