

HOSTED BY



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: <http://www.elsevier.com/locate/jtcms>

Consistency between traditional Chinese medicine constitution-based classification and genetic classification

Ruoxi Yu ^a, Xiaohong Zhao ^b, Lingru Li ^c, Cheng Ni ^c, Yin Yang ^c,
Yuanyuan Han ^d, Ji Wang ^c, Yan Zhang ^c, Qi Wang ^{b,*}

^a Health Cultivating Institute, Beijing University of Chinese Medicine, Beijing 100029, China

^b Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100864, China

^c Center for Studies in Constitution Research of Traditional Chinese Medicine, School of Basic Medical Science, Beijing University of Chinese Medicine, Beijing 100029, China

^d The Fifth People's Hospital of Datong, Shanxi 037006, China

Received 15 June 2015; accepted 20 August 2015

Available online 15 March 2016

KEYWORDS

Yin-deficient constitution;
Yang-deficient constitution;
Gene expression;
Constitution identification;
Genetic classifier

Abstract *Background:* We studied the consistency between two classification systems for categorizing patients: traditional Chinese medicine (TCM) constitution-based methods, versus genetic clustering. Genetic classification in constitutional identification was also evaluated.

Methods: A TCM physician evaluated the constitution of each patient, according to four examinations (inspection, auscultation-olfaction, interrogation, and palpation). Those who met the criteria for Yang-deficient, Yin-deficient, and balanced constitutions were enrolled in the study. Peripheral blood samples were obtained from the participants, and peripheral blood mononuclear cells were separated from the samples within 2 hours. Total RNA extraction from the white blood cells was performed; and an Affymetrix HG-U133 Plus2.0 array was used to determine the peripheral blood gene expression profiles. The samples were classified using a support vector machine genetic classifier, and the "leave-one-out" method was used for validation.

Results: The global gene expression profiles of 32 samples were grouped into three categories, and the samples in each of the gene categories corresponded with the three constitution categories. The three constitution types were distinguished using the genetic classifier with 165 genes. The accuracy of the prediction classification was greater than 95% using mathematical method.

Conclusions: Participants with Yin-deficient, Yang-deficient, and balanced constitutions have varying physical characteristics and gene expression patterns. Additionally, the results from

* Corresponding author. Tel.: +86 10 64286989.

E-mail address: wangqi710@126.com (Q. Wang).

Peer review under responsibility of Beijing University of Chinese Medicine.

TCM constitution classification matched those obtained by genetic classification. Finally, our preliminary gene classifier distinguishes among Yin-deficient, Yang-deficient, and balanced constitutions, and provides a methodological basis for identifying the different constitutions. © 2015 Beijing University of Chinese Medicine. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Medicine's guiding model is anticipated to change in the twenty-first century from "disease-oriented" to "health-oriented" or "holistic" medicine, with an emphasis on the prevention, versus the treatment of diseases. Further, group treatment approaches are expected to be replaced by individualized medicine and more specific treatment approaches. Concurrent with modern medical science developments, the global medical community commonly believes that the best medicine is not curative, but rather is preventive.

TCM constitution in individuals can be changed when the inherited and acquired characters are affected after birth. Thus, an entity with relatively stable physiology, psychology and mentality is formed.¹ Chinese medicine constitution approaches for studying the status of an individual's health are well-developed. Individuals can be classified into nine constitution types according to an epidemiological study of the Chinese Han population²: balanced constitution, Qi-deficient constitution, Yang-deficient constitution, Yin-deficient constitution, phlegm-dampness constitution, damp-heat constitution, stagnant blood constitution, stagnant qi constitution, and inherited special constitution. Constitution classification facilitates the path toward personalized medical care. Constitution identification takes an individual's physical fitness as an observational object, assesses physical state and classifies different physical characteristics to elucidate the overall health factors and individual differences in human participants, which can then be used to develop preventative medicine principles. Thus, intervention can be based on the qualities presented by different people.

The biological basis and clinical relevance of the constitution system in TCM have been intensively investigated over the years.^{3–5} TCM constitution may reflect genetic features, and different constitutions might result from differential gene expressions, an area that needs to be confirmed by further research. In previous studies, genetic classifiers have been assigned to classify participants using a mathematical model. Classifiers were assigned to known categories of unknown participants. Through a novel research approach, we aim to determine whether the classification schemes of TCM constitution conform to genetic differences, using a gene classifier system.

In TCM theory, when Yin and Yang are in equilibrium, positive health is achieved and sustained. People with Yang-deficient constitution are marked by aversion to cold while those with Yin-deficient constitution by aversion to heat. The study of these two constitutions is the foundation of all other TCM constitution-based research. Various

studies have demonstrated that the Yin-Yang phenomenon can be explained at least partially through molecular biology.^{6,7} In one study, the nuclear transcription factor Yin-Yang1 (YY1) gene was shown to have similar regulatory functions to traditional Chinese Yin and Yang. The YY1 transcription factor was identified as targeting a plethora of potential target genes important for cell proliferation and differentiation which is similar to Yin and Yang in TCM.⁶ Another study showed that YY1 possessed a wide range of regulatory functions, including a role as a critical negative regulator of excitatory amino acid transporter (EAAT) 1 in one of two glial glutamate transporter pathways.⁷ So Yin and Yang can be partly explained by molecular biology.

Support vector machine (SVM) is used as a new approach for data mining. It is based on information provided by the limited sample in a complex model (i.e. the specific learning accuracy of training samples, and accuracy) and learning ability (i.e. any samples without error identification) to seek the best compromise and to obtain optimal generalization ability.⁸ SVM can better solve the problems of small samples, nonlinear, high dimension and local minimum points. It is an important method in the field of machine learning to solve problems and estimation of nonlinear functions, and it is widely used in pattern recognition, signal processing and time series prediction.⁹ It is an excellent classifier owing to its robust theoretical background, strong adaptability, and extensive generalized functions. Because of its suitability with small sample sizes and its generalized performance, SVM has recently been internationally used in the pattern recognition field.¹⁰ Furthermore, its scalability and generalization capabilities make it very suitable for classification study.^{11–13}

Our research aims to evaluate whether a relationship exists between TCM constitution-based classification and genetic clustering classification, and to explore the application of a genetic classifier to constitution identification. Genomics is the essence of biophysical research. This study has selected the Yin and Yang constitutions to carry out the biophysical genomics research. The results will provide a strong basis for further study of the TCM theory and clinical application.

Methods

Participants

Volunteer participants were chosen from students without compensation at Beijing University of Chinese Medicine and from staff at Beijing New Era Company, between February 17 and May 26, 2012. Participants were 18–28 years old,

and were assessed by TCM constitution identification by using a standardized questionnaire (Wang Qi's Body Constitution Classification Questionnaire, Chinese version, frequent code: ZYYXH/T157-2009), and confirmed by a TCM practitioner. We confirmed that none of the participants presented with any disease symptoms, via overall health checkups. Those determined to possess a Yang-deficient constitution, a Yin-deficient constitution, or a balanced constitution were recruited to the study, and other participants were excluded.

Inclusion and exclusion criteria

Inclusion criteria specified: (1) participants must be citizens of China, of Han ethnicity; (2) aged 18–28 years, as they are young and have less physical disorders, of either sex; (3) they must match the determinant standard for a balanced constitution, or a Yang-deficient constitution or a Yin-deficient constitution, described in the TCM constitution guidelines in 2008 (Table 1) and issued by the Chinese Association of Traditional Chinese Medicine (frequent number: ZYYXH/T157-2009). The study protocol was approved by the Ethics Committees of the Center for Studies in Constitution Research of Traditional Chinese Medicine, School of Preclinical Medicine, Beijing University of Chinese Medicine, Beijing, and all the participants signed informed consent (No. 2012BZHLL0301).

Exclusion criteria specified: (1) participants must not match the criteria for constitutions other than Yang-deficient, Yin-deficient, or balanced constitutions; (2) they must not possess simultaneous characteristics of more than one constitution; (3) they must not possess any viral or bacterial infection or other infectious disease; (4) they must not possess any mental illness that could interfere with informed consent.

Experimental methods

Peripheral blood was taken from the participants. Peripheral blood mononuclear cell isolation was carried out within 2 hours, and then total RNA was extracted from the white blood cells. We used Affymetrix HG-U133 Plus2.0 array expression profile chips from the Boao Biological Company (Beijing, China) for the peripheral blood gene expression profiling.

Data analysis

We processed the original cell file with Robust Multi-array Average pretreatment, which was used for background correction and data standardization. We used an unsupervised hierarchical clustering method to assign the data to all 32 chip probe samples. We classified the sample prediction using Array Tools analysis.

Four forecasting methods were used: (1) the mean number of genes in the classifier; (2) diagonal linear discrimination analysis, capable of avoiding data overfitting and ignoring the correlation between genes; (3) a1-nearest neighbor predictor; (4) a3-nearest neighbor predictor. Nearest neighbor methods were based on a distance function for pairs of observations, such as the Euclidean distance, or one minus the correlation. For the gene expression data considered here, the distance between two mRNA samples, with gene expression profiles $x = (x_1, \dots, x_p)$ and $x_0 = (x'_1, \dots, x'_p)$, was based on a correlation between the two gene expression profiles.

The four classifiers were verified using the leave-one-out method, which takes any of the 31 samples of a model system to predict the 32nd sample of the group. In this way, the classification was checked 32 times to ensure complete validation. In addition, we checked the sensitivity, specificity, positive predictive value, and negative predictive value of the four parameters, to illustrate the classification efficiency of the classifier.

Classification prediction method

There were significant differences in the gene expression profiles between the selected participants based on the known category of the samples, after which the category model was established. Through this model, it was possible to determine the category of the unknown samples.

Results

Baseline data

Thirty-two participants (7 males and 25 females, aged 18–28 years) were recruited according to the study design. The Yang-deficient constitution group, and the

Table 1 Diagnostic standards for Yang-deficient, Yin-deficient, and balanced constitution.

Characteristics	Yang-deficient constitution	Yin-deficient constitution	Balanced constitution
Main Characteristics	Cold intolerance Cold hands, feet, stomach, and waist Prefer hot food and drinks Susceptible to cold	Heat intolerance Hot body and face Dry eyes Prefer cold food and drinks Susceptible to heat	Energetic Without any symptoms or characteristics of above two constitutions
Secondary Characteristics	Watery stool Obesity Whitish skin Nocturia Tender and pale tongue	Constipated Emaciation Red lips, and dry skin Feel parched and need to drink water Dry and red tongue	Good sleep Cold tolerance Good memory

Yin-deficient constitution group each contained 12 participants (samples) and were designated as the experimental groups; the remaining 8 participants in the balanced constitution group were designated as the control group. A chi-squared test of the three groups for sex distribution gives $\chi^2 = 0.3048$, $P = 0.8587$ ($P < .05$); thus, the three groups showed no significant difference in sex distribution.

Clustering of results

All 54,675 probes in the chip were used for the unsupervised clustering analysis of the samples. The result showed that these 32 samples could be divided into three categories, which was highly consistent with the classification by TCM constitution. The samples obtained from the 32 recruited participants were assorted by clustering (Fig. 1).

Balanced constitution sample numbers were: 206, 207, 208, 209, 210, 211, 212, and 213. Yin-deficient constitution sample numbers were: A1, A9, A13, A14, A15, A16, A17, A18, A19, A20, A21, and A22. Yang-deficient constitution sample numbers were: B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B13, and B14.

Multidimensional scaling analysis results

We also performed multidimensional analysis on the samples to independently validate the results obtained from hierarchical clustering (Fig. 2). Multidimensional scaling (MDS) analysis is based on the similarity or distance between the objects of a study. In our case the research objects are images in a low-dimensional space, clustering by a graphical method of dimensional analysis, and presented by the multidimensional scaling analysis on a spatial location map. This provides a simple and clear description of all the relative relationships between the research objects. MDS allowed us to confidently distinguish the three kinds of samples observed, based on the relationships that could be visualized from the expression profile chip values.

The MDS analysis clearly demonstrated that three different kinds of samples existed in the gene expressions of these chips.

Four classifier classifications

All of the sample labels had three categories, namely, the Yang-deficient constitution (Yang), Yin-deficient constitution (Yin), and balanced constitution (N). The four classifiers were checked by the leave-one-out method, which takes any of the 31 samples for a model system to predict which group the last sample belongs to. The only mis-assigned sample was N211, which was a balanced constitution sample assigned to the wrong group according to the TCM constitution classification (Table 2). The accuracy of prediction classification was greater than 95% (Table 3).

Classifier effect assessments

Our statistical analysis showed sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the classifier categories all to be above 85% of them. So the classifier effect was very good. (Table 4)

Application

Using the predictor category, a model was set up. By establishing a good model, it was possible to determine the category of the unknown samples. This method only requires knowledge of some specific gene expression values to be able to judge the category of the unknown samples. It is also simple and feasible. After building the model, we found 165 genes from our samples whose expression profiles relate to the characteristics of the three physical classifications. After obtaining the nearest centroid coordinates of the 165 genes in each category, we made a classification prediction for the subsequent sample, and could do so as long as the centroid distance to the three categories was known.

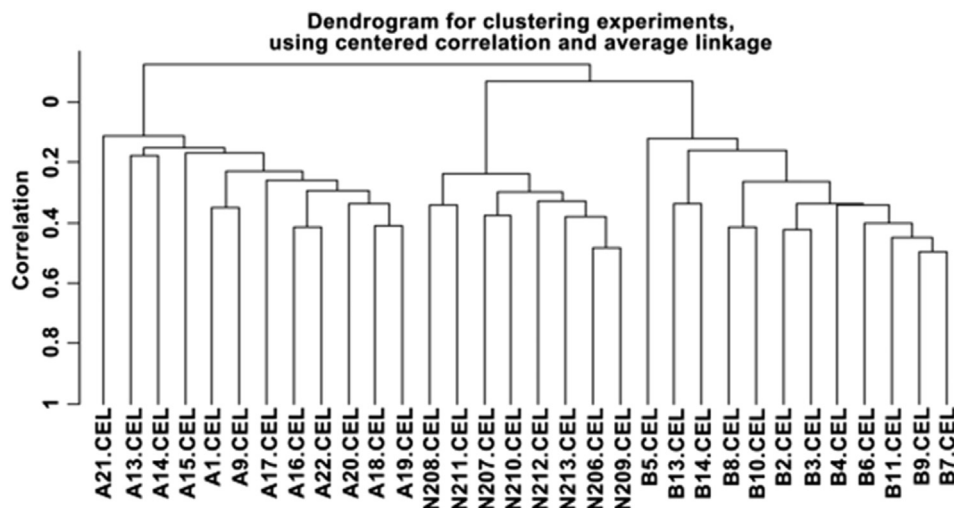


Fig. 1 Clustering figure.

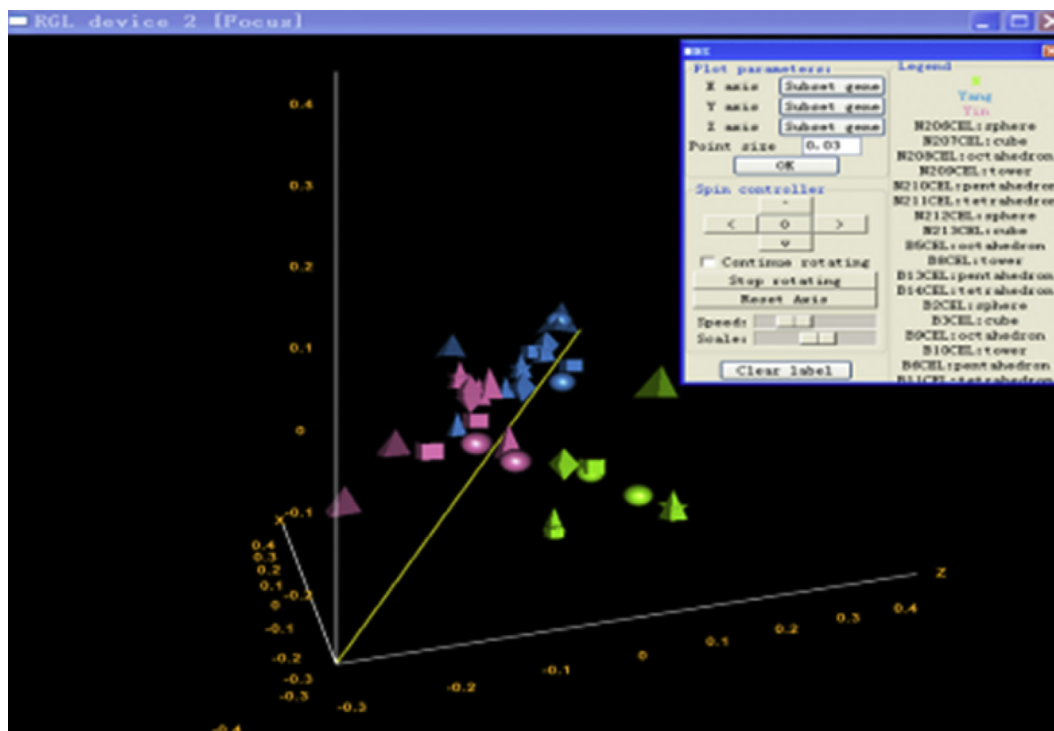


Fig. 2 MDS Analysis Diagram.

Note: Blue represents Yang-deficient constitution, green represents balanced constitution, and purple represents Yin-deficient constitution.

Method to detect the category of unknown samples is showed as the followings: (1) detect the expression of these 165 genes of the participants under test values, such as X1:34 (gene1 34),, X165 (gene 165, 87), with X referring to the gene date; (2) then calculate the distance of the gene to each centroid category. The following formula can then be used: $D(X, Y) = (X1 - Y1) \times (X1 - Y1) + \dots + (X165 - Y165) \times (X165 - Y165)$, with Y representing the centroid value of Yang-deficient or Yin-deficient or balanced constitutions. This calculation was repeated three times, once for each constitution, and the three centroid values were obtained to derive the Yang-deficient or Yin-deficient or balanced constitution assignment. Samples most likely belong to the constitution with the minimum value.

Discussion

The results show that the Yin-deficient constitution, Yang-deficient constitution, and balanced constitution are all associated with corresponding gene expression patterns. Furthermore, TCM constitution classification and genetic classification results were identical. Our preliminary gene classifier can distinguish between Yin-deficient constitution, Yang-deficient constitution, and balanced constitutions, and can provide a methodological basis for studying the health status of people with different constitutions.

Sound health is regarded as a balance between Yin and Yang in TCM. Daily life activities are possible because of the coordination of Yin and Yang, which are two aspects of

keeping a balanced body in coordination with growth and decline. Modern studies have demonstrated that pathological constitutions, i.e. disturbances in Yin and Yang, differ from a balanced constitution on a molecular level.⁵ These studies have outlined the genetic and metabolic components of the Yang-deficient and Yin-deficient constitutions.¹⁴ A previous study has shown that the expression of thyroid hormone receptor beta and several key nuclear receptor coactivators, including steroid receptor coactivator 1, steroid receptor coactivator 3, cAMP-response element-binding protein, and some mediators, were significantly decreased in Yang-deficient constitution individuals.⁵ The same study also explored susceptibility genes associated with Yang-deficient constitution using single nucleotide polymorphism genotyping. This pilot study suggests that polymorphisms in RGS6, mGluR5, GAPDHL19, and IKZF1 are associated with changes in cyclic adenosine monophosphate and cyclic guanosine monophosphate levels, memory, metabolic energy status, and immune function in people with Yang-deficient constitution.

Investigators are presently paying increasing attention to TCM-based constitution classifications, mainly because of increasing concerns regarding the prevention of disease, and suboptimal health conditions.¹⁴ TCM constitution is an effective tool for biological research because assigning constitution in TCM can improve disease prevention. If we can find a sign of disturbance before a disease occurs, we can take preventive approaches to block the onset of disease.

Constitution identification is mainly determined using a TCM constitution identification questionnaire and the "four diagnoses" (inspection, auscultation-olfaction,

Table 2 Gene classifier results.

No.	Array ID	Class label	Mean number of genes in classifier	Diagonal linear discriminant analysis correct	1-Nearest neighbor	3-Nearest neighbors correct	Nearest centroid correct
1	N206.CEL	N	226	YES	YES	YES	YES
2	N207.CEL	N	229	YES	YES	YES	YES
3	N208.CEL	N	229	YES	YES	YES	YES
4	N209.CEL	N	227	YES	YES	YES	YES
5	N210.CEL	N	229	YES	YES	YES	YES
6	N211.CEL	N	238	NO	NO	NO	NO
7	N212.CEL	N	2285	YES	YES	YES	YES
8	N213.CEL	N	227	YES	YES	YES	YES
9	B5.CEL	Yang	230	YES	YES	YES	YES
10	B8.CEL	Yang	229	YES	YES	YES	YES
11	B13.CEL	Yang	231	YES	YES	YES	YES
12	B14.CEL	Yang	223	YES	YES	YES	YES
13	B2.CEL	Yang	2285	YES	YES	YES	YES
14	B3.CEL	Yang	226	YES	YES	YES	YES
15	B9.CEL	Yang	225	YES	YES	YES	YES
16	B10.CEL	Yang	225	YES	YES	YES	YES
17	B6.CEL	Yang	224	YES	YES	YES	YES
18	B11.CEL	Yang	222	YES	YES	YES	YES
19	B7.CEL	Yang	224	YES	YES	YES	YES
20	B4.CEL	Yang	226	YES	YES	YES	YES
21	A13.CEL	Yin	2285	YES	YES	YES	YES
22	A14.CEL	Yin	2285	YES	YES	YES	YES
23	A15.CEL	Yin	226	YES	YES	YES	YES
24	A16.CEL	Yin	225	YES	YES	YES	YES
25	A17.CEL	Yin	227	YES	YES	YES	YES
26	A18.CEL	Yin	225	YES	YES	YES	YES
27	A19.CEL	Yin	227	YES	YES	YES	YES
28	A20.CEL	Yin	227	YES	YES	YES	YES
29	A21.CEL	Yin	230	YES	YES	YES	YES
30	A22.CEL	Yin	226	YES	YES	YES	YES
31	A1.CEL	Yin	226	YES	YES	YES	YES
32	A9.CEL	Yin	227	YES	YES	YES	YES
Mean percent of correct classification:				97	97	97	97

interrogation, and palpation) of TCM. We propose that further research is needed to confirm whether there are actual biophysical differences between the constitutions. We were able to investigate the relationship among these three constitutions based on our study's centroid concept, and distinguish between them. Using our method, we can determine the constitution of a patient who wishes to receive individualized diagnosis and treatment. However,

presently we can only use the current method to identify the above three types of constitutions.

Development of preventative and individualized medicine could certainly be enhanced by establishing a reliable system to identify a patient's TCM constitution. However, although there are various research studies on TCM constitution, in the fields of immunology, molecular biology, genomics, and metabolomics, these collective approaches have been insufficient. Further their research results have not provided an effective objective basis to identify constitution on an individual level. We need to find an integrated subjective and objective method to better apply constitution identification to health care. Only when macro and micro approaches are combined used, can Chinese and western medicine fully utilize their respective advantages and complement each other. The complete health status of an individual can only be determined when all health status with subjective and objective information phenomenon can be obtained.

Table 3 Effect evaluation.

Class	Sensitivity	Specificity	PPV	NPV
Balanced constitution (N)	0.875	1	1	0.96
Yang-deficient constitution (Yang)	1	0.95	0.923	1
Yin-deficient constitution (Yin)	1	1	1	1

Table 4 The centroid coordinates of 165 genes of the three constitutions.

Probe ID	Gene symbol	Balanced constitution (N)	Yang-deficient constitution (Yang)	Yin-deficient constitution (Yin)
203504_s_at	ABCA1	5.3548	6.317	4.9863
207275_s_at	ACSL1	8.00805	9.1052	7.4246
202912_at	ADM	7.4032	8.0578	5.9304
215783_s_at	ALPL	3.5893	5.3104	3.1654
233011_at	ANXA1	6.847	8.2914	8.2055
205568_at	AQP9	7.1327	8.7493	6.3109
205239_at	AREG	6.65	4.3428	3.3646
202672_s_at	ATF3	6.634	6.1036	4.8659
225612_s_at	B3GNT5	7.0237	7.283	5.5871
202391_at	BASP1	9.0853	10.3873	8.6151
205681_at	BCL2A1	9.9159	10.1677	8.4462
203140_at	BCL6	7.661867	8.1904	6.798233
1568768_s_at	BRE	5.8148	5.1606	4.1114
214696_at	C17orf91	8.0923	8.9449	7.6108
229899_s_at	C20orf199	6.5326	7.2674	5.6047
1556072_at	C22orf37	6.8676	7.9517	6.6044
1553158_at	C3orf34	4.9779	6.70115	5.4576
220088_at	C5AR1	8.8344	9.499	7.8603
205476_at	CCL20	8.2493	3.4758	3.112
204103_at	CCL4	11.0101	8.9828	8.8948
204440_at	CD83	9.1088	7.9399	6.5142
209395_at	CHI3L1	4.4416	7.10365	4.33295
229967_at	CMTM2	6.6578	9.4988	6.4267
223796_at	CNTNAP3/// LOC643827	2.2849	3.7749	2.3345
218610_s_at	CPPED1	6.8578	7.5301	6.501
205931_s_at	CREB5	6.198867	7.376033	5.9652
1553297_a_at	CSF3R	7.61675	8.97895	7.38585
225557_at	CSRNP1	7.528	7.9911	6.3795
204470_at	CXCL1	6.5755	6.9209	4.5445
209774_x_at	CXCL2	7.53075	4.38515	3.905
207850_at	CXCL3	7.0659	3.3828	3.0279
209201_x_at	CXCR4	11.1109	10.0866	9.2737
206515_at	CYP4F3	4.3857	7.5091	4.332
226064_s_at	DGAT2	5.41075	6.7315	4.97155
216260_at	DICER1	4.2014	4.6932	3.3942
244840_x_at	DOCK4	4.6518	4.9459	3.5587
230263_s_at	DOCK5	7.5335	8.5654	7.3558
226817_at	DSC2	5.1045	6.1178	4.5443
201044_x_at	DUSP1	7.1354	9.0824	7.8039
204794_at	DUSP2	8.6174	7.2046	6.5616
218660_at	DYSF	7.106	7.8364	6.4868
210724_at	EMR3	4.3581	6.2485	4.2687
205767_at	EREG	6.7111	4.9714	3.6252
213506_at	F2RL1	6.0156	7.4858	5.9538
217966_s_at	FAM129A	8.3577	9.5287	8.1245
226811_at	FAM46C	7.54685	6.7476	6.29645
1553798_a_at	FBXL13	2.0184	2.947	1.9239
207674_at	FCAR	5.5256	6.1456	4.72025
203561_at	FCGR2A	8.2824	9.2683	8.0226
204007_at	FCGR3B	9.7316	11.6887	10.0039
221345_at	FFAR2	7.2908	8.6041	5.8323
241627_x_at	FLJ10357	5.9021	8.0932	6.2049
209189_at	FOS	9.5647	11.707	11.0737
202768_at	FOSB	7.2397	9.3754	8.0034
205118_at	FPR1	7.2141	8.03525	6.66575
210773_s_at	FPR2	6.21995	7.6254	5.9002

Table 4 (continued)

Probe ID	Gene symbol	Balanced constitution (N)	Yang-deficient constitution (Yang)	Yin-deficient constitution (Yin)
213524_s_at	GOS2	9.6817	10.66	6.9642
215977_x_at	GK	4.69335	6.0589	4.5983
229770_at	GLT1D1	7.1582	8.1242	6.6823
200648_s_at	GLUL	6.4752	7.863	6.5075
221958_s_at	GPR177	4.921833	6.0432	4.436133
205419_at	GPR183	8.3364	6.6969	7.009
1553723_at	GPR97	4.44215	5.5603	3.90845
206643_at	HAL	6.049	6.8395	5.3715
214455_at	HIST1H2BC	4.52025	5.976	4.4298
202637_s_at	ICAM1	7.7379	7.0739	5.8964
201631_s_at	IER3	10.3555	8.4295	7.5133
242903_at	IFNGR1	6.1827	5.3071	4.472
220704_at	IKZF1	4.3546	5.4346	6.0253
210118_s_at	IL1A	5.0945	2.3225	2.1614
39402_at	IL1B	11.05035	8.09235	6.5172
202948_at	IL1R1	3.5847	5.0251	3.4627
211372_s_at	IL1R2	4.94035	7.2222	4.28465
212657_s_at	IL1RN	8.1657	7.6877	6.3134
211506_s_at	IL8	10.4488	11.10225	8.1208
207094_at	IL8RA	6.5308	8.2206	6.3926
207008_at	IL8RB	7.2761	10.0419	7.6469
201465_s_at	JUN	7.748	9.303	8.2263
210119_at	KCNJ15	3.9367	6.159667	3.9064
206765_at	KCNJ2	7.7667	7.1934	5.7117
218963_s_at	KRT23	3.8679	7.2113	3.8802
206440_at	LIN7A	4.1162	5.199	3.886
1555847_a_at	LOC284454	7.7955	7.0786	5.8307
1560058_at	LOC399900	5.023	6.2127	5.1946
206522_at	LOC642103///MGAM	5.6723	8.1529	5.1545
165412_at	LOC643072	3.8439	5.1863	3.8252
244065_at	LOC643827	1.8799	3.1389	2.0023
225955_at	LOC653506///METRNL	8.0564	6.8559	6.6609
165648_at	LRG1	4.8511	6.5389	4.3771
36711_at	MAFF	8.1411	8.0344	6.1474
220945_x_at	MANSC1	4.869	6.9585	4.6192
209179_s_at	MBOAT7	7.0665	8.0125	6.6904
1556873_at	MEX3C	4.295	5.99625	5.2552
203435_s_at	MME	3.78625	6.58835	3.5685
207890_s_at	MMP25	5.8263	7.758	5.6059
203936_s_at	MMP9	5.5168	6.9896	5.1521
1554906_a_at	MPHOSPH6	4.3326	2.7497	2.8477
165846_at	MXD1	7.127233	8.202333	6.2886
217738_at	NAMPT	8.8418	9.848825	7.499775
205147_x_at	NCF4	7.2878	8.1872	6.8508
214657_s_at	NCRNA00084	4.7529	6.2555	5.9244
237591_at	NCRNA00173	5.4817	6.6547	5.3659
209930_s_at	NFE2	6.0508	7.4438	6.6519
1560527_at	NF-E4	3.0781	4.5128	2.7735
203574_at	NFIL3	8.4519	8.6381	6.9806
223217_s_at	NFKBIZ	10.8438	10.0969	8.8418
205220_at	NIACR2	7.9392	9.6234	6.7112
216016_at	NLRP3	6.752533	5.296467	4.592333
214321_at	NOV	2.1575	3.8467	2.1259
204622_x_at	NR4A2	8.134	7.85555	6.3976
214963_at	NUP160	6.1361	4.8125	4.6703

(continued on next page)

Table 4 (continued)

Probe ID	Gene symbol	Balanced constitution (N)	Yang-deficient constitution (Yang)	Yin-deficient constitution (Yin)
205040_at	ORM1	4.1613	5.3258	3.8915
230170_at	OSM	5.6894	5.5958	4.0032
209791_at	PADI2	6.6471	7.25	5.9869
232304_at	PELI1	8.9908	9.489	7.9447
202861_at	PER1	7.4921	6.1927	5.6084
202464_s_at	PFKFB3	8.5692	6.9885	6.4278
217996_at	PHLDA1	5.7819	4.5717	3.9376
203691_at	PI3	4.44245	6.54935	4.1471
211924_s_at	PLAUR	8.46415	8.0218	6.798
37028_at	PPP1R15A	8.4937	8.51195	7.0396
222662_at	PPP1R3B	5.4997	6.9612	5.5988
232629_at	PROK2	5.7173	8.1611	5.9806
1554997_a_at	PTGS2	7.3188	8.2003	4.64065
1559018_at	PTPRE	6.3697	6.9328	5.5514
206157_at	PTX3	8.2768	5.6571	4.6725
1554999_at	RASGEF1B	7.2219	5.1585	4.0996
205645_at	REPS2	5.2897	6.4961	4.8862
216834_at	RGS1	7.6377	6.5212	5.38325
216515_at	RLIM	6.6135	5.3204	4.7011
204669_s_at	RNF24	6.9795	8.0992	6.935
1555878_at	RPS24	5.5965	4.1408	4.5507
1569599_at	SAMSN1	6.6265	5.509	4.0717
202083_s_at	SEC14L1	6.4132	7.7718	5.7518
217977_at	SEPX1	8.6024	9.6337	8.3166
201739_at	SGK1	10.1568	9.7038	8.5696
205896_at	SLC22A4	5.8746	6.661	5.2022
222528_s_at	SLC25A37	5.667043	7.553529	5.645543
225597_at	SLC45A4	6.3093	7.4325	6.1292
203021_at	SLPI	4.9251	6.2416	4.3877
1565703_at	SMAD4	5.2845	3.7644	4.7363
227697_at	SOCS3	7.087	6.5462	5.3323
215078_at	SOD2	8.81465	8.5355	6.6791
1554114_s_at	SSH2	5.7978	6.9324	5.5747
217104_at	ST20	2.4913	3.9221	3.0363
204542_at	ST6GALNAC2	5.4115	6.5312	5.2487
225987_at	STEAP4	4.63245	5.96105	4.32325
209238_at	STX3	5.18975	6.6962	5.22775
218750_at	TAF1D	6.8475	4.913	5.5385
235086_at	THBS1	6.89215	4.06815	4.032
226489_at	TMCC3	5.4706	6.0311	4.7302
235798_at	TMEM170B	8.1688	6.7007	6.9328
207113_s_at	TNF	9.1365	5.904	5.3948
202643_s_at	TNFAIP3	11.22655	10.3819	9.1021
206026_s_at	TNFAIP6	6.24895	7.155	4.72605
206222_at	TNFRSF10C	4.7478	7.4201	4.795767
224836_at	TP53INP2	6.1607	6.7992	5.1125
219434_at	TREM1	7.8535	8.5927	6.1994
202241_at	TRIB1	7.9989	9.2007	7.2704
224412_s_at	TRPM6	3.1853	4.2378	3.019
231990_at	USP15	7.5371	7.8791	6.5665
205922_at	VNN2	8.4091	9.6114	8.0625
1553514_a_at	VNN3	4.7093	5.95575	4.15935
212860_at	ZDHHC18	7.0442	8.0505	6.8802
219854_at	ZNF14	6.6961	4.9773	5.0998

Conclusions

Our study clearly shows that TCM constitution classification and genetic classification are closely connected. It confirms that TCM constitution classifications have a genetic basis. We identified 165 genes that differed between Yin-deficient, Yang-deficient, and balanced constitutions. We were able to evaluate the relationships among the three constitutions according to our centroid data, and then determine the constitutions of unknown populations. Using our preliminary gene classifier we were able to differentiate between Yin-deficient, Yang-deficient, and balanced constitutions. This study should provide a strong methodological basis for further research of constitution identification in healthy individuals. Moreover, it provides a methodological basis for other constitution analyses applications by using a support vector machine and centroid data. However, the specific methods will require significant improvement with more constitutions and more people to verify. The key aim of this study was to determine whether a relationship exists between TCM constitution-based classification and genetic clustering classification, and to explore the application of a genetic classifier to constitution identification. We believe that the study will enrich Yin-Yang theory, and help to inform people about the effectiveness of constitution identification as a tool for understanding overall health.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We thank Beijing Institute of Genomics for designing the study and for performing the statistical analysis, especially professor Qianfei Wang. This work was supported by the National Key Basic Research Program of China (973 Program, No. 2011CB505400).

References

1. Wang J, Li Y, Ni C, et al. Cognition research and constitutional classification in Chinese medicine. *Am J Chin Med.* 2011;39:651–660.
2. Q W, Zhu Y. Chinese constitution epidemiological investigation of TCM in the general population – based on 21948 cases of 9 provinces epidemiological data. *J Chin Med.* 2009;1:7–12.
3. Wong W, Lam CLK, Wong VT, et al. Validation of the constitution in Chinese medicine questionnaire: does the traditional Chinese medicine concept of body constitution exist? *Evid Based Complement Altern Med.* 2013:481491.
4. Wang Q, Ren X, Yao S, et al. Clinical observation on the endocrinal and immune functions in subjects with yin-deficiency constitution. *Chin J Integr Med.* 2010;16:28–32.
5. Wang Q, Yao S. Molecular basis for cold-intolerant yang-deficient constitution of traditional Chinese medicine. *Am J Chin Med.* 2008;36:827–834.
6. Boucherat O, Landry-Truchon K, Berube-Simard F, et al. Epithelial inactivation of Yy1 abrogates lung branching morphogenesis. *Development.* 2015;142:2981.
7. Karki P, Kim C, Smith K, et al. Transcriptional regulation of the astrocytic excitatory amino acid transporter 1 (EAAT1) via NF-kappa B and Yin Yang 1 (YY1). *J Biol Chem.* 2015;290:23725–23737.
8. Cherkassky V. The nature of statistical learning theory. *IEEE Trans Neural Netw.* 1997;8:1564.
9. Goh ATC, Goh SH. Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. *Comput Geotech.* 2007;34:410–421.
10. Zhang M, Cheng W, Guo P. Intelligent recognition of mixture control chart pattern based on quadratic feature extraction and SVM with AMPPO. *J Coast Res.* 2015;73:304–309.
11. Zhang G, Piccardi M. Structural SVM with partial ranking for activity segmentation and classification. *IEEE Signal Proc Lett.* 2015;22:2344–2348.
12. Sun L, Liu H, Zhang L, et al. IncRScan-SVM: a tool for predicting long non-coding RNAs using support vector machine. *PLoS One.* 2015, 10(e013965410).
13. Czarniecki WM, Podlowska S, Bojarski AJ. Robust optimization of SVM hyperparameters in the classification of bioactive compounds. *J Cheminformatics.* 2015;7.
14. Lee C, Li T, Tsai C, et al. Yang deficiency body constitution acts as a predictor of diabetic retinopathy in patients with type 2 diabetes: Taichung diabetic body constitution study. *Evid Based Compl Alt Med.* 2015:940898.