

유전자 칩 및 다변량 분석방법을 이용한 사상체질 유전자 선별에 관한 연구

김판준 · 서은희 · 이정환 · 하진호 · 최홍식* · 정태영** · 구덕모

대구한의대학교 사상체질의학교실, *대구한의대학교 내과학교실
**대구한의대학교부속 제한의원

Abstract

A Study on Sasang Constitutional Gene Selection Using DNA Chips by Multivariate Analysis

Kim Pan-Joon, Seo Eun-Hee, Lee Jung-Hwan, Ha Jin-Ho,
Choi Hong-Sik*, Jung Tae-Young**, Goo Deok-Mo

Dept. of Constitutional Medicine, College of Oriental Medicine, Daegu Haany Univ.

*Dept. of Internal Medicine, College of Oriental Medicine, Daegu Haany Univ.

**Jaehan Hospital, College of Oriental Medicine, Daegu Haany Univ.

1. Objectives

This research uses the DNA chip, which includes 16,383 gene code, and various statistic prediction way that shows objectification index for the objectification of constitution diagnosis.

2. Methods

Drawing blood whose constitution is confirmed, and analyze its gene information by using 1.7k DNA chip to find the gene correlation through multivariate statistical method.

3. Results and Conclusions

Distinctive genes such as AK001919, U09384, NM_001805, X99962, NM_004796, AK026738, AL050148, BC002538, AK027074, AK026219, AF087962, AL390142, NM_015372, AL157466, NM_002446, AK024523, NM_014706, NM_014746 and AL137544 were related to Taeumin; AL157448, NM_005957, NM_005656, NM_017548, AK027246, NM_003025, NM_012302 and NM_005905 were represented in Soeumin, while AK026503, AF147325, NM_002076, AF147307, AK001375, NM_003740, NM_005114, AB007890, NM_005505, NM_015900, NM_014936, Z70694, AB023154, U52076, NM_004360, NM_005835, NM_017528, AF087987, NM_014897, AK021720, NM_006420, AJ277915, AK002118 and AK021918 were for Soyangin.

This study figured out the possibility to develop the prediction system by sorting each constitution's gene, and research each constitution's distinctive character of manifestation pattern.

Key Words : Sasang Constitution, Gene, DNA Chip, Multivariate Statistical Method

I. 緒 論

20세기 후반에 접어들면서 생명과학의 발달로 암이나 질병의 원인으로 유전자가 관여한다는 것이 밝혀진지 오래되었고 병원성 세균의 감염여부 및 항생제 내성검사, 신약개발, 유전자의 기능연구, 범죄자확인 및 친자확인 등 많은 부분에 DNA와 유전자감식기술이 이용되고 있다. 지금까지는 이러한 사실을 밝히기 위하여 한 번에 여러 개의 유전자발현 변화를 확인한다는 것이 쉽지 않았으나 DNA Chip 기술은 우리가 밝히고자 하는 검사대상에서 추출한 DNA sample을 반응시켜 그 결과를 컴퓨터로 처리하는 방식으로 기존의 방식보다 빠르게, 그리고 한 번에 많은 정보를 얻어낼 수 있게 되었다.

DNA Chip은 slide glass상에 수백, 수천 개의 DNA 유전자를 정렬, 고정화한 것으로 해석하고자 하는 세포에서 추출한 RNA로 조제한 형광 표식 cDNA를 hybridization하여 각 유전자의 발현 변화를 측정하는 방법이다. 한 번의 실험으로 매우 많은 유전자의 발현 정보를 얻을 수 있으며, 게놈해석에서 얻은 유전자 구조정보를 바탕으로 체계적인 발현 해석을 위한 강력한 screening tool로 이용되고 있다.

1894년 동무 이제마에 의해 창안된 사상의학은 체질적 차이를 감안하여 동일한 병이라 하여도 치료방법을 다르게 적용하는 사람중심의 의학적인 특성 뿐 아니라, 병의 원인을 심신 양면에 있다고 보고 외적인 요인만 경계하여 약물에만 의존하는 치료는 옳지 않다고 보고 정신적 요인을 다스리는 치료를 중시하는 의학이다. 즉 체질적인 知行을 통한 非藥物的 正氣를 통해 인체내부의 체질정기를 잘 관리하는 것이 더욱 중요하다고 인식하고 연령 및 체질에 따른 다른 양생법을 제시하고 있다¹.

이러한 사상의학의 실제적용을 위해서 무엇보다 중요한 것은 체질의 정확한 진단이라고 할 수 있다. 기존의 체질진단은 『東醫壽世保元』 「四象人辨證論」에 나타난 體形氣像, 性質才幹

容貌詞氣, 病症藥理의 네 가지를 종합적으로 판단하여 진찰자의 주관적인 판단으로 이루어지는 경우가 많았다. 그러므로 이를 극복하기 위한 체질진단의 정확성과 객관화를 위한 연구가 계속 진행되어오고 있다.

그간 안면진단법², 신체계측법³, 음성진단법⁴, 설문에 의한 진단방법⁵ 등 여러 분야에 걸쳐 체질진단의 객관화를 위한 연구가 진행되어 왔으나 아직까지는 만족할 만한 결과를 얻지 못했다.

최근 들어서는 주관적 정보를 객관화된 정보로 변환하고 체질의 근거를 찾기 위해 개체의 모든 정보를 지니며 인간의 몸에서 가능하고 있는 것, 즉 유전자를 중심으로 한 과학적 연구방법 또한 요구되고 있는 실정이다.

사상체질에 관한 지금까지의 유전자 특성에 관한 연구로는 이식항원 HLA-DR 유전자 polymorphism에 관한 연구⁶, Amp-FLP를 이용한 사상체질의 유전적 분석 연구⁷, 유전자 분석법에 의한 사상체질의 연구⁸, 유전자지문법을 이용한 사상체질의 유전적 분석 연구⁹, ACE 유전자와 사상체질과의 상관성을 살펴 본 연구¹⁰, HLA typing을 이용한 체질유전자 분석에 관한 연구¹¹, MTHFR의 단일염기다형성을 중심으로 한 연구¹², DNA chip을 이용한 사상체질의 연구¹³ 등이 있었으나, 유¹³를 제외한 기존의 연구들은 제한된 유전자의 특성만을 보았다.

이에 저자는 체질진단의 객관화를 위해 생명정보학에 대한 접근을 시도해 보았고, 16,383개의 유전정보가 들어 있는 DNA chip과 다양한 통계적 방법을 이용하여 체질을 예측할 수 있는 시스템과 분류방법에 대한 객관적인 지표를 제시해 보고자 본 연구를 시행하였다.

II. 研究對象 및 方法

1. 연구대상 및 total RNA 추출방법

1) 연구대상

정확한 체질진단에 근거한 대상군의 선정을 위해 ○○대학교 부속 ○○한방병원에 1년 이

상 근무한 교직원 중 체형기상, 용모사기, 성질 재간을 기본으로 사상체질 전문의에 의해 체질을 진단받은 결과 체질이 전형적인 것으로 진단된 자들을 우선 선택하였다. 이후 사상 처방을 4주 이상 복용했으며, 사상처방 복용후의 체질증이나 체질병증이 호전된 자 중 태양인을 제외한 태음인, 소양인, 소음인 각각 5명을 대상으로 하였다. 실험은 임의화의 법칙에 따라서 분류, 성별, 연령에 무관하게 무작위로 인원을 선택하여 3일간에 걸쳐서 실험을 수행하였다.

2) total RNA 추출방법

체질 감별에 의해 분류된 사람의 정맥혈액 35 ml을 채취하여 PAXgene™ Blood RNA Tube (QIAGEN 사)에 각각 2.5 ml씩 나누어 담은 뒤 상온에 2시간 두어 세포를 완전히 용해시켰다. PAXgene™ Blood RNA Tube에 담긴 혈액을 PAXgene™ Blood RNA Kit (QIAGEN 사)를 이용하여 사상체질별 각 5sample씩 분류하여 total RNA 50 μ g을 얻었다. 분리한 RNA 샘플을 DNA microarray에 적용하기 위하여 Agilent Bioanalyzer 2100 chip으로 RNA quality와 양을 측정하였다.

2. DNA Chip 실험

1) Hybridization

50 μ g의 total RNA와 oligo dT primer (1.5 μ g)을 3 μ l 넣어준 뒤 DEPC-H₂O를 이용하여 최종 20.5 μ l의 부피로 맞췄다. 70 $^{\circ}$ C에서 5분간 가열하여 RNA를 denaturation 시킨 뒤 빨리 얼음에 넣고 RT 반응 (AMV Buffer, low dT NTP, Cy3 or Cy5-dUTP, RNase inhibitor, AMV reverse transcriptase)을 준비하여 42 $^{\circ}$ C에서 1시간 반응시켰다. 반응 종료 후, 1N NaOH 10 μ l를 첨가한 뒤 37 $^{\circ}$ C에서 10분간 denaturation 시키고, 1M Tris-HCl (pH 7.5) 25 μ l를 첨가하여 중화시킨 후, Spin column 또는 sephacryl S-100을 이용하여 probe를 정제하였다. 1/10부피의 3M NaOAc와 2.5배 부피의 100% ethanol을 넣고 -80 $^{\circ}$ C에서 15분간 반응 한 후 4 $^{\circ}$ C에서 12,000rpm으로 15분간 원심 분리시키고, 상층액을 조심스럽게 제거한 뒤 1ml 70%

ethanol을 넣고 4 $^{\circ}$ C에서 5분간 12,000rpm으로 원심 분리하였다. DNA 침전물을 공기 중에 말린 뒤 15 μ l의 hybridization buffer에 녹인 뒤 준비된 probe를 95 $^{\circ}$ C에서 2분간 denaturation 시킨 후 얼음에 보관하였다. 10분간 14,000rpm에서 원심 분리하여 상층액의 probe만을 취하고 Probe를 DNA chip 위에 떨어뜨린 후 cover glass로 덮었다. 62 $^{\circ}$ C에 맞추어 놓은 humid chamber에서 12~16시간동안 hybridization을 수행하고 Hybridization이 끝나면 slide를 60 $^{\circ}$ C washing buffer (2X SSC, 0.2% SDS)에서 30분간 2번 닦아내고, Washing이 끝난 후 슬라이드를 실온에서 말렸다.

2) Scanning and Image Analysis

Scanarray 5000(GSI Lumonics, Billerica, MA, USA)을 이용해서 혼성화 반응이 끝난 microarray 슬라이드의 영상을 얻었다. Slide를 Scanarray 5000에 넣은 후 정밀도가 높은 상을 얻기 위하여 laser power와 PMT 값을 조절하여 Cy3와 Cy5영상이 비슷한 형광 intensity를 갖게 맞춘 뒤 영상분석을 위해서 Cy3와 Cy5 영상을 각각의 TIFF File로 저장하였다.

TIFF file로 저장된 영상은 Image 4.1(BioDiscovery, Los Angeles, CA) program을 이용해서 분석하여, Excel File을 얻은 다음 GeneSight 1.0(BioDiscovery) program을 이용하여 결과를 분석하였다.

3) Data Base Search

분석한 유전자들은 미국국립생물공학정보센터(National Center for Biotechnology Information, NCBI) <http://www.ncbi.nlm.nih.gov> 와 Stanford Microarray Database (<http://smd.stanford.edu/cgi-bin/source/sourceResult>)를 이용하여 gene의 기능적으로 분류하였다. Genes Ontology 별 분류 (Molecular Function, Biological Process, Cellular Component)는 <http://www.geneontology.org>를 이용하였다.

3. DNA Chip 분석

DNA Chip 발현 데이터를 분석하기 위해서 우선 Normalization과 filtering을 하여 데이터에

나타날 수 있는 오차(Error)와 잡음(Noise)을 제거하였다.

먼저 다중비교는 분산분석을 이용하여 각 유의수준별 유의하게 차이가 나는 유전자를 추출하였고, 다음으로 각 유의확률별 추출된 유전자를 단계적 분류분석 방법 중 스텝와이즈 선택법(Stepwise Selection)을 이용하여 유전자를 추출하였다. 이후 체질분류를 위한 판별분석을 하고 각 체질에 대하여 의미 있는 유전자를 파악하기 위해서 군집분석 방법을 사용하였다.

Ⅲ. 分析結果

1. 일반화(Normalization)

DNA 마이크로어레이 분석 실험에서는 보통의 실험 자료에 비해 잡음이 많이 포함되어 있으며 또한 자료에 일정한 패턴을 보이는 경우가 많다. 잡음이 추가 될수록 자료의 품질은 떨어지기 마련이며 실험자의 숙련도와 실험에 사용된 화학물질 등에 의해 그 정도가 달라질 수 있다. 특히 일정한 패턴을 지닌 잡음은 분석 결과에서 치명적인 효과를 발휘할 수도 있기 때문에 잡음을 제거하는 과정인 표준화(Normalization)를 수행하였다. 이 표준화 방법은 통계적인 회귀모형(Regression)을 이용하여 체계적으로 변이를 제거하는 방법

을 사용하였다. Normalization 후의 Plot을 보면 평균과 분산이 안정화된 상태가 되었음을 확인할 수 있었다. 각 슬라이드에 대하여 표준화한 자료를 기반으로 유전자를 선택하였다(Fig. 1).

2. 유전자 선택

16,383개의 유전자 중에서 사상체질을 가장 잘 분류할 수 있는 유전자를 선택하기 위해서 다중비교 방법을 이용하여 체질별 유의하게 차이가 나는 유전자를 선택하였다. 우선 Normalization 된 데이터에 대해서 각 실험 군을 통제변수로 설정한 후 설정된 통제변수를 이용하여 각 체질별 F통계량과 유의확률(P-Value)을 구하여 가장 유의한 차이가 나는 확률을 기초로 변수를 추출하였다.

여기서 각 유의확률 별 데이터에 대한 1차 선택 후 단계적 분류분석 방법을 이용하여 2차적으로 유전자를 선택하고 각 유의확률 별 추출된 유전자 중에서 동시에 추출된 유전자에 대해서는 분류력이 높은 유전자로 선택하고 차후 분석을 하였다.

1) 다중비교(Multiple Comparisons)결과

여러 집단에 대한 비교를 할 때 사용하는 방법으로 기본 선형모형에 대한 가정 사항을 설정한 후 오차에 대한 평균제곱오차와 모형에

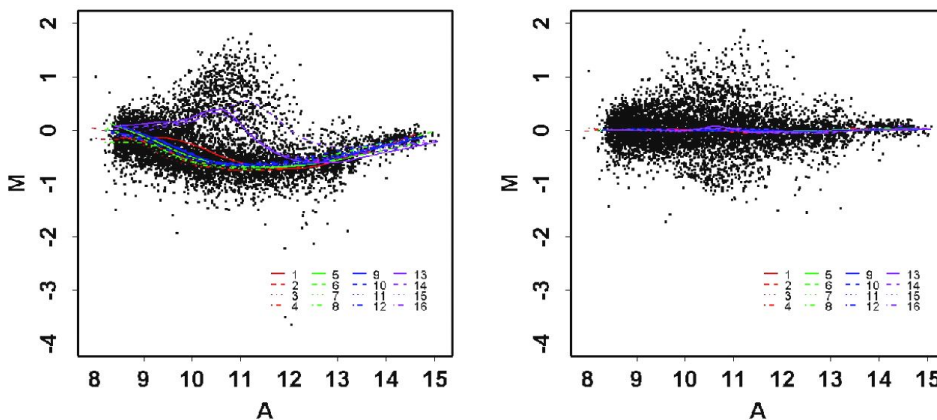


Fig. 1. M-A Plot Before and After Normalization

Table 1. Selected Gene Count which It Follows at Level of Significance

Prob	1%	5%	10%	15%	20%
Number	68	487	1,192	1,989	2,853

대한 평균제곱을 나는 값으로 F통계량을 계산하였다. 각 유의수준별 유의하게 차이가 나는 유전자를 보면 1%에서는 68개의 유전자가 5%에서는 487개의 유전자가 선택되었고 10%이상에서는 1,000개 이상이 선택됨을 보였다. 10%에서는 1,192, 15%에서는 1,989, 20%에서는 2,853개의 유전자가 선택되었음을 확인할 수 있었다 (Table 1).

2) 단계적 분류방법 결과

위의 각 유의확률별 추출된 유전자를 이용하여 단계적 분류방법 중 스텝와이즈 선택법 (Stepwise Selection) 을 사용하여 유전자를 추출하였다. 총 선택된 유전자는 58개로 그 중 5개의 유전자가 공통적으로 추출되었다 (Table 2).

각 유전자에 대해서 체질별 평균을 이용하여 유전자들이 체질을 구분할 수 있는지에 대한 신뢰도를 보기 위해 평균을 이용한 그림을 그려보면 유전자별 도식화된 그림이 세세하게 나누어짐을 확인할 수 있었다. 선끼리의 간격이 넓게 분포되어 있는 유전자들이 체질을 잘 분류할 수 있다고 할 수 있으며 각 체질별 높은 값 또는 낮은 값은 해당 체질을 대표할 수 있는 유전자라고 지칭할 수 있다 (Fig. 2).

위와 같은 도식화된 결과로 분류함수에 대한 개괄적인 설명력을 확인하였고 추출된 유전자에 대한 다변량 분석기법을 적용하여 분류함수 및 체질별 유전자를 선별하였다.

3. 체질 분류를 위한 판별분석

선택된 51개의 유전자를 이용하여 새로운 데이터가 들어올 때 체질을 예측할 수 있는 함수를 만들기 위해서 판별함수를 이용하였다. 세 집단에 대한 오분류율이 가장 적게 되도록 오분류비용의 기댓값 (ECM: Expected Cost of Misc-

lassification) 이 적게 되는 가중치 값을 구하도록 한다. 공분산행렬에 대한 유의성을 검정한 결과 유의하므로 선형판별분석을 사용하여 각 유전자에 대한 계수값을 표시하였다 (Table 3). Table 3의 Constant는 절편 항을 의미하는 것으로 각 유전자에 대한 값이 0일 경우 즉 시작점을 의미한다. 개별 체질에 따른 계수 값이 큰 순서를 보면 태음인의 경우 X99962가 태음인을 분류할 때 가장 높은 가중치를 주고 있고 순차적으로 NM_006420, AK027246, NM_004796 AL050148 등의 순으로 나타났다. 소음인의 경우는 AK027246의 유전자가 소음인에게 가장 높은 가중치를 부여하고 있고 다음으로 순차적으로 보면 X99962, NM_014706, NM_004796, NM_006420 등의 순으로 나타났다. 소양인에 대해서 볼 때 AK026503가 소양인에게 가장 높은 가중치를 할당하였고 그 다음 순차적으로 보면 NM_014706, X99962, AK001919, AF087987 등의 순으로 나타났다.

4. 체질별 유전자에 대한 군집분석

각 체질에 대하여 의미 있는 유전자를 파악하기 위해서 군집분석 방법을 사용해 보았다. 선택된 51개의 유전자의 특징을 가장 잘 설명할 수 있도록 하기 위해서 유전자별 체질에 따른 결과보다 유전자별 묶이는 정도를 보았다.

위에서 선별된 유전자에서 각 체질별 의미 있는 유전자를 선택하기 위해 Clustering 방법을 이용하여 체질별 설명력이 높은 유전자를 선택하고 유의수준 별로 가장 높은 차이를 보이는 그룹끼리 묶어보았다 (Table 4).

태음인에 대한 설명력이 높은 유전자에 대한 값을 볼 경우 각 유전자에서 태음인을 가장 잘 설명하면서 반대로 태음인과 가장 상반적인 결과를 나타내는 유전자는 2번으로 표시하였다.

Table 2. Selected Gene List which It Follows at Level of Significance

Gene	1%	5%	10%	15%	20%
NM_017528				NM_017528	
NM_004796					NM_004796
AF087962	AF087962				
AJ277915				AJ277915	
NM_005835		NM_005835			
AL157466				AL157466	
NM_014746	NM_014746	NM_014746			
NM_014706		NM_014706			
AK027246				AK027246	
AK001919			AK001919		
NM_005114					NM_005114
AL157448				AL157448	
U09384			U09384		
NM_003025					NM_003025
NM_005656					NM_005656
AB023154	AB023154				
Z70694			Z70694		
AK021720				AK021720	
AK024523					AK024523
AL050148		AL050148			
NM_002076		NM_002076			
AF147307		AF147307			
NM_001805			NM_001805		
NM_015900	NM_015900				
AK026219	AK026219				
AF147325		AF147325			
AK027074			AK027074		
AK026738	AK026738				
AL390142	AL390142				
NM_002446		NM_002446			
AK021918			AK021918		
AF087987			AF087987	AF087987	
AK002118	AK002118				
U52076	U52076				
AL137544				AL137544	
NM_012302					NM_012302
NM_014897		NM_014897			
NM_006420			NM_006420	NM_006420	NM_006420
X99962	X99962	X99962			
NM_003740				NM_003740	
AK001375	AK001375				
NM_005957	NM_005957				
NM_014936					NM_014936
NM_015372			NM_015372		
NM_004360	NM_004360				
AB007890					AB007890
BC002538		BC002538			
AK026503			AK026503	AK026503	AK026503
NM_005505	NM_005505				
NM_017548			NM_017548		
NM_005905		NM_005905			

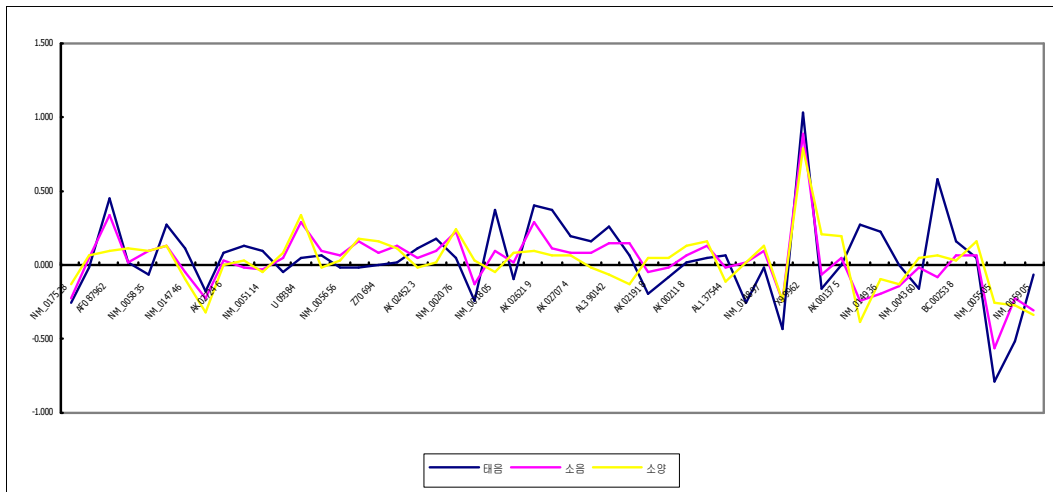


Fig. 2. The Plot against the Average of the Individual Test Objects against Each Gene

Table 3. The Coefficient against Each Gene which it Follows in Discrimination Analysis

Variable	Taeum	Soeum	Soyang	Variable	Taeum	Soeum	Soyang
Constant	-150	-73	-100	Constant	-150	-73	-100
NM_017528	-40	-17	-3	AK027074	56	29	28
NM_004796	60	58	49	AK026738	58	27	2
AF087962	23	22	7	AL390142	21	5	-16
AJ277915	42	14	45	NM_002446	6	0	-35
NM_005835	-35	-5	8	AK021918	-3	-3	-3
AL157466	36	13	10	AF087987	-48	7	52
NM_014746	12	-15	-23	AK002118	7	3	51
NM_014706	-51	-61	-80	U52076	-27	9	44
AK027246	68	69	51	AL137544	-8	-9	-19
AK001919	54	31	53	NM_012302	-1	-1	-8
NM_005114	2	5	8	NM_014897	-50	-19	-1
AL157448	29	38	35	NM_006420	-94	-51	-49
U09384	13	6	8	X99962	97	66	62
NM_003025	20	36	-16	NM_003740	6	10	11
NM_005656	-32	-11	-20	AK001375	9	13	33
AB023154	-25	0	23	NM_005957	-4	1	-1
Z70694	-14	10	48	NM_014936	-8	-8	-3
AK021720	-51	-17	-14	NM_015372	21	-2	-3
AK024523	6	-16	-20	NM_004360	-29	-6	5
AL050148	58	35	11	AB007890	-3	1	4
NM_002076	13	26	29	BC002538	56	50	33
AF147307	12	14	17	AK026503	29	36	119
NM_001805	6	2	2	NM_005505	-4	-3	-2
NM_015900	-8	6	29	NM_017548	-57	-29	-29
AK026219	23	17	5	NM_005905	-10	-9	-13
AF147325	23	27	31				

단위 : 10,000,000

Table 4. The Gene List which is Classified with the Taeumin, Soeumin and Soyangin

	태음	소음	소양
AK001919	1	2	
U09384	1	2	
NM_001805	1	2	
X99962	1		2
NM_004796	1		2
AK026738	1		2
AL050148	1		2
BC002538	1		2
AK027074	1		2
Taeumin AK026219	1		2
AF087962	1		2
AL390142	1		2
NM_015372	1		2
AL157466	1		2
NM_002446	1		2
AK024523	1		2
NM_014706	1		2
NM_014746	1		2
AL137544	1		2
AL157448	2	1	
NM_005957	2	1	
NM_005656	2	1	
Soeumin NM_017548	2	1	
AK027246		1	2
NM_003025		1	2
NM_012302		1	2
NM_005905		1	2
AK026503	2		1
AF147325	2		1
NM_002076	2		1
AF147307	2		1
AK001375	2		1
NM_003740	2		1
NM_005114	2		1
AB007890	2		1
NM_005505	2		1
NM_015900	2		1
NM_014936	2		1
Z70694	2		1
Soyangin AB023154	2		1
U52076	2		1
NM_004360	2		1
NM_005835	2		1
NM_017528	2		1
AF087987	2		1
NM_014897	2		1
AK021720	2		1
NM_006420	2		1
AJ277915		2	1
AK002118		2	1
AK021918		2	1

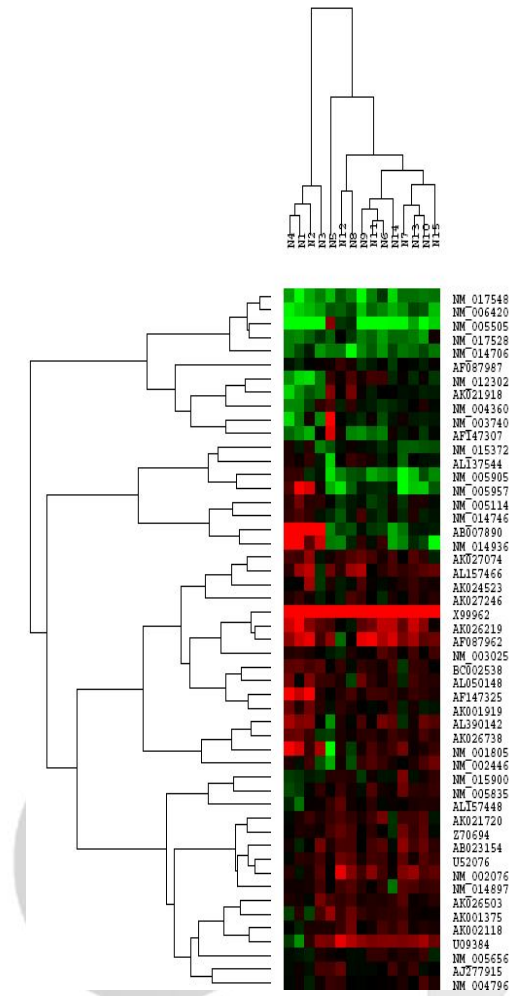


Fig. 3. Image Plot of Hierarchical Clustering

태음인과 가장 상반된 결과를 보이는 유전자에서 소음인의 경우 3개의 유전자 AK001919, U09384, NM_001805의 유전자이고, 태음인의 경우 상반된 결과를 보이는 유전자로는 X99962, NM_004796, AK026738, AL050148, BC002538, AK027074, AK026219, AF087962, AL390142, NM_015372, AL157466, NM_002446, AK024523, NM_014706, NM_014746, AL137544로 나타났다

소음인의 경우는 AL157448, NM_005957, NM_005656, NM_017548의 네 가지의 유전자가 태음인과 상반된 값을 보이고 있고 AK027246,

Table 5. The Gene List which is Selected with Duplication

Gene	Class1	Class2
NM_014706	Tacumin	Tacumin
NM_005905	Soeumin	Soeumin
NM_005957		
AB023154	Soyang	Soyang
AF147325		
AJ277915		
AK001375		
AK002118		
AK021720		
AK026503		
NM_002076		
NM_005835		
NM_014897		
NM_015900		
U52076		
Z70694		

NM_003025, NM_012302, NM_005905 의 네 개의 유전자는 소양인과 상반된 값을 나타내고 있다.

소양인의 경우 AK026503, AF147325, NM_002076, AF147307, AK001375, NM_003740, NM_005114, AB007890, NM_005505, NM_015900, NM_014936, Z70694, AB023154, U52076, NM_004360, NM_005835, NM_017528, AF087987, NM_014897, AK021720, NM_006420 유전자가 태음인과 상반된 값을 보이고 있고 AJ277915, AK002118, AK021918 유전자의 경우 소음인과 상반된 값을 나타내고 있다.

각 유전자에서 다변량 분석기법인 Clustering 방법을 이용하여 각 유전자별 체질을 설명할 수 있는 유전자를 선택해 보았다. 사용할 방법은 Hierarchical Clustering 방법을 사용하였다(Fig. 3).

각 유전자에 대하여 그룹별로 묶어서 그 설명력을 해석해 보면, 첫 번째로 선택된 유전자는 NM_017548, NM_006420, NM_005505, NM_017528, NM_014786, AF087987, NM_012302, AK021918, NM_004360, NM_003740, AF147307 로 11개의 유전자가 한 그룹으로 묶여져 있다.

두 번째로 선택된 유전자는 NM_015372, AL

137544, NM_005905, NM_005957, NM_005114, NM_014746, AB007890, NM_014936 로 8개의 유전자가 한 그룹으로 구성되었다.

세 번째로 선택된 유전자는 AK027074, AL157466, AK024523, AK027246, X99962, AK026219, AF087962, NM_003025, BC002538, AL050148, AK001919, AF147325, AL390142, AK026738, NM_001505, NM_002446, NM_015900, NM_005835, AL157448, AK021720, Z70694, AB023154, U52076, NM_002076, NM_014897, AK026503, AK001375, AK002118, U09384, NM_005656, AJ277915, NM_004796로 32개 유전자가 선택되었다.

위에서 선별된 유전자와 비교해 보면 태음인의 경우 NM_014706의 유전자가 공통적으로 선택되었고, NM_005905, NM_005957은 소음인에서 선택되었다.

소양인의 경우 공통적으로 선택된 유전자는 AB023154, AF147325, AJ277915, AK001375, AK002118, AK021720, AK026503, NM_002076, NM_005835, NM_014897, NM_015900, U52076, Z70694의 유전자가 선택되었다(Table 5).

IV. 考察 및 結論

체질진단의 정확성과 객관성을 확보하기 위한 노력은 사상의학계의 큰 관심이자 반드시 해결해야 할 과제이다.

동무 이제마에 의해 창안된 사상의학은 인체의 생리, 병리, 질병치료 및 양생까지 포괄하는 의학으로 사람을 네 가지 체질로 나눠 각 장기의 불균형과 그에 따른 성정의 차이에 따라 질병이 달리 발생된다고 한다. 그러므로 체질적 차이를 감안하여 동일한 병이라 하여도 치료방법을 달리 적용하고 있다¹⁴.

체질에 따라 다른 치료와 양생이 적용되기 위해선 우선 정확한 체질진단이 이루어져야 한다. 그러나 지금까지의 체질 구분은 대부분 진찰자의 주관적인 판단에 의하는 경우가 많으므로 체질감별의 객관성과 정확성이 요구되

고 있는 실정이다.

사상체질의학에서의 체질 진단은 심성, 체형과 용모, 체질증과 체질병증 등의 여러 기준이 종합되어야 진단될 수 있으며 주관적인 정보들을 대상으로 진단하게 된다. 체질 진단을 객관화하기 위해서는 주관적 정보들을 객관화된 정보로 변환할 수 있으며 사람의 여러 가지 정보가 포함된 것을 중심으로 체질의 단서를 찾는 연구가 필요하다. 체질의 근거를 밝히는 것은 사상체질의학을 현대 과학의 범주에서 해석하고 규명하는 연구의 가장 기초적인 작업이다. 그러므로 체질 근거를 찾기 위해 개체의 모든 정보를 지니며 인간의 몸에서 기능하고 있는 것, 즉 유전자를 중심으로 연구하는 것은 필연적이다.

이에 저자는 DNA chip을 통한 다양한 통계적 방법의 연구를 통해 체질을 구분할 수 있는 유전자를 찾는 것이 중요하다고 생각하여 본 연구를 하게 되었다.

본 연구에서는 체질이 확정된 사람의 혈액을 채혈하여 1.7k DNA chip을 이용해 발현된 유전자 정보를 전체적으로 분석하여 체질에 따른 유전적 상이성 및 상관성을 다변량 분석방법을 통해 찾아보았다.

우선 16,383개의 유전자 중에서 사상체질을 가장 잘 분류할 수 있는 유전자를 선택하기 위해 다중비교 방법을 이용하여 체질별 유의하게 차이가 나는 유전자를 선택하였다. Normalization 된 데이터에 대해서 각 실험 군을 통제변수로 설정한 후 설정된 통제변수를 이용하여 각 체질별 F통계량과 유의확률(P-Value)을 구하여 가장 유의한 차이가 나는 확률을 기초로 변수를 추출한 후 기본 선형모형에 대한 가정 사항을 설정하고 오차에 대한 평균제곱오차와 모형에 대한 평균제곱을 나눈 값으로 F통계량을 계산하였다. 그 결과 1%에서는 68개의 유전자가 5%에서는 487개의 유전자가 선택되었고, 10%에서는 1,192, 15%에서는 1,989, 20%에서는 2,853개의 유전자가 선택되었음을 확인할 수 있었다.

다음으로 각 유의확률별 추출된 유전자를 이용하여 단계적 분류방법을 이용하여 유전자를 추출하였다. 그 결과 총 선택된 유전자는 58개로 그 중 5개의 유전자가 공통적으로 추출되었으며 추출된 총 유전자 51개를 이용하여 예측함수를 구현하도록 하였다.

또한 각 체질에 대하여 의미 있는 유전자를 파악하고 선택된 51개의 유전자의 특징을 가장 잘 설명할 수 있도록 하기 위해 군집분석 방법을 사용하여 유전자별 묶이는 정도를 보았다.

그 결과 체질에 따른 유전자의 발현 패턴의 특이성을 갖는 유전자를 발굴하고 그 패턴에 따른 유전자의 세부 분류를 통하여 각 체질별 특이 유전자를 선별할 수 있었다. 크게 분류된 결과로는 태음인의 경우 AK001919, U09384, NM_001805, X99962, NM_004796, AK026738, AL050148, BC002538, AK027074, AK026219, AF087962, AL390142, NM_015372, AL157466, NM_002446, AK024523, NM_014706, NM_014746, AL137544가 선택되었고, 소음인의 경우 AL157448, NM_005957, NM_005656, NM_017548, AK027246, NM_003025, NM_012302, NM_005905가 선택되었고, 마지막으로 소양인의 경우 AK026503, AF147325, NM_002076, AF147307, AK001375, NM_003740, NM_005114, AB007890, NM_005505, NM_015900, NM_014936, Z70694, AB023154, U52076, NM_004360, NM_005835, NM_017528, AF087987, NM_014897, AK021720, NM_006420, AJ277915, AK002118, AK021918가 각각 선택되었다.

각 유전자에 따른 세부 분류의 결과로는 태음인의 경우 NM_014706, 소음인의 경우 NM_005905, NM_005957 소양인의 경우 AB023154, AF147325, AJ277915, AK001375, AK002118, AK021720, AK026503, NM_002076, NM_005835, NM_014897, NM_015900, U52076, Z70694가 밀도가 높은 유전자로 선택되었다.

집단별 표본수의 부족으로 각 체질에 대하여 정 분류되는 특이 유전자가 소양인에 많이 치우친 경향을 보이고 있으나 앞으로 실험집단에 대한 관측치가 누적 된다면 각 체질에 대

Table 6. The List of Genes that Over Express in Taeumin

Gene ID	Gene name	Locus
NM_001805	CCAAT/enhancer binding protein (C/EBP), epsilon	14q11.2
X99962	RAB39, member RAS oncogene family	11
NM_004796	Neurexin 3	14q31
AK026738	CDNA: FLJ23085 fis, clone LNG06916	18
AI050148	Sorting nexin 1	15q22.31
BC002538	Serpin peptidase inhibitor, clade B (ovalbumin), member 9	6p25
AK027074	Zinc finger protein 451	6p12.1
AK026219	SID1 transmembrane family, member 2	11q23.3
AF087962	HLA-B associated transcript 4	6p21.3
AL157466	Regulatory factor X, 3 (influences HLA class II expression)	9p24.2
NM_002446	Mitogen-activated protein kinase kinase kinase 10	19q13.2
AK024523	CDNA: FLJ20870 fis, clone ADKA02524	13
NM_014706	Squamous cell carcinoma antigen recognised by T cells 3	12q24.1
NM_014746	Ring finger protein 144	2p25.1
AL137544	Chromosome 6 open reading frame 157	6q14.1

Table 7. The List of Genes that Over Express in Soeumin

Gene ID	Gene name	Locus
AL157448	Protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), gamma isoform	4p16.1
NM_005957	Chromosome 1 open reading frame 167	1p36.3
NM_005656	Transmembrane protease, serine 2	21q22.3
NM_017548	Hypothetical protein H41	3q22.1
AK027246	Sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like	11q23.3
NM_003025	SH3-domain GRB2-like 1	19p13.3
NM_012302	Latrophilin 2	1p31.3
NM_005905	SMAD, mothers against DPP homolog 9 (Drosophila)	13q12-q14

한 분류력이 높은 유전자 선별의 정도를 높일 수 있을 것이다.

분석결과에 따른 각 체질별 특이유전자들의 생물학적 특성을 살펴보면 우선 태음인에게 관찰되는 특이적 유전자인 CCAAT/enhancer binding protein (C/EBP), epsilon 유전자는 DNA regulatory regions에서 homodimer로써 결합하여 transcription factor로써 작용한다. RAB39 유전자는 GTP와 binding 하여 세포내 단백질 수송에 관여

한다. Neurexin 3 유전자는 척추동물 신경계에 서 cell adhesion로써 작용한다. Sorting nexin 1 유전자는 phox(px) domain을 포함하고 있으며 이 domain은 phosphoinositide binding domain으로써 protein transporter activity에 관여한다. Serpin peptidase inhibitor, clade B 유전자는 granzyme b inhibitor로 작용하여 serine-type endopeptidase inhibitor를 활성화 시킨다. Zinc finger protein 451 유전자는 steroid receptors의 coactivator로써 작용하여

Table 8. The List of Genes that Over Express in Soyangin

Gene ID	Gene name	Locus
AK026503	Chromosome 8 open reading frame 51	8q24
NM_002076	Glucosamine (N-acetyl)-6-sulfatase (Sanfilippo disease IIID)	12q14
AF147307	PDZ domain containing, X chromosome	Xp11.23
AK001375	Hypothetical protein FLJ14640	19q13.11
NM_003740	Potassium channel, subfamily K, member 5	6p21
NM_005114	Heparan sulfate (glucosamine) 3-O-sulfotransferase 1	4p16
AB007890	Limkain b1	16p13.11
NM_005505	Scavenger receptor class B, member 1	12q24.31
NM_015900	Phospholipase A1 member A	3q13.13-q13.2
NM_014936	Ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative function)	6p21.1
Z70694	Hypothetical protein LOC153561	5q13.2
AB023154	Deltex 4 homolog (Drosophila)	11q12.1
U52076	Retinoic acid receptor beta 1 (RARbeta1) mRNA, exon 3, non-coding region	3
NM_004360	Cadherin 1, type 1, E-cadherin (epithelial)	16q22.1
NM_005835	Solute carrier family 17 (sodium phosphate), member 2	6p21.3
NM_017528	Williams Beuren syndrome chromosome region 22	7
AF087987	MAM domain containing glycosylphosphatidylinositol anchor 1	6p21
NM_014897	Zinc finger protein 652	17q21.32
AK021720	NS5ATP13TP2 protein	11q23.3
NM_006420	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)	20q13.13
AJ277915	LIM homeobox 9	1q31-q32
AK002118	RNA binding motif, single stranded interacting protein 1	2q24.2
AK021918	G protein-coupled receptor 172A	8q24.3

transcriptional regulation에 관여한다. Mitogen-activated protein kinase kinase kinase 10 유전자는 serine, threonine kinase family로 MAPK8, JNK, MKK4, SEK1을 인산화 시키는 역할을 한다. Squamous cell carcinoma antigen recognised by T cells 3 유전자는 RNA 대사에 관여하여 neuron 형성과 유지에 필요한 단백질이다. Ring finger protein 144 유전자는 Ring finger motif를 포함하는 단백질로 단백질-DNA, 단백질-단백질의 상호작용에 관여하는 모티프로 알려져 있다.

소음인에게 특이하게 관찰되는 유전자의 생물학적 기능을 살펴보면 Protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), gamma isoform 유전자는 Ser/Thr 인산화 및 세포성장 과 분열에 negative control로써 작용한다. Transmembrane protease, serine 2 유전자는 type II transmembrane domain을 포함하는 단백질로 cleavage 와 secreted에 의해 활성화 된다. Sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like 유전자는 콜레스테롤 생합성 대사에 관여하는

것으로 알려져 있다.

소양인에게 특이하게 관찰되는 유전자의 생물학적 기능을 살펴보면 PDZ domain containing, X chromosome 는 배아 발생과정 중에 신경계의 형성과 착상에 관여하는 도메인을 포함하는 것으로 알려져 있다. Potassium channel, subfamily K, member 5는 신장에서 많이 발현되어 칼륨 (K⁺)수송에 관여하는 것으로 알려지고 있다. Heparan sulfate (glucosamine) 3-O-sulfotransferase 1는 조직 내 뇌와 신장에서 많이 발현되며 심장과 폐, 췌장에서는 그 발현양이 적은 것으로 나타나며, 항응고 heparan sulfate 생합성에 관여한다. Scavenger receptor class B, member 1는 DNA가 RNA로 만들어지는 전사과정에서 보조인자로 관여하는 것으로 알려져 있다. Phospholipase A1 member A는 지방대사에 관여하여 불용해성의 기질의 활성을 높이는 역할을 한다. Hypothetical protein LOC15356는 keratan 단백질이 형성과 탄수화물 대사과정에 관여하는 것으로 알려져 있다. Solute carrier family 17 (sodium phosphate), member 2는 세포내에서 Na⁺이온 등의 수송에 관여한다. Zinc finger protein 652는 주로 핵내에서 존재하여 전사 조절대사에 관여한다.

본 연구에서는 체질이 확정된 사람의 혈액을 채혈하여 사상체질별 특이 유전자를 선별하여 체계적이고 객관적으로 체질을 분류하는 기준을 제시하고자 하였다. 유전자의 기능을 이해하고 각 체질별 유전자 발현양상을 DNA Chip을 이용하여 유전체 수준에서 각 유전자의 기능을 밝혀 사상체질 진단에 응용함을 연구에 목표로 하였다. 하지만 그 기능이 알려지지 않은 유전자가 많으며 단시일 내에 수백 명의 피험자를 확보해야하는 어려움 등으로 인하여 재현성이 이루어지지 않은 점 등으로 체질을 대표할 수 있는 모든 유전자를 밝힐 수는 없었지만 부분적으로 각 체질에 따른 특이 유전자를 검색할 수 있었다. 추후에 좀 더 많은 실험대상군과 여러 차례의 실험을 통한 자료의 축적과 분석을 통하여 체질 분류의 객관성을 확

립할 수 있을 것으로 사료된다.

V. 參考文獻

1. 송일병. 동의보감과 동의수세보원에 나타난 우리민족의 의학정신. 사상체질의학회지. 2004; 16(3):1-7.
2. 고병희, 송일병, 조용진, 최창석, 김종원, 홍석철, 이의주, 이상룡, 서정수. 사상체질별 두면부의 형태학적 특징. 사상체질의학회지. 1996;8(1): 101-186.
3. 허만희, 고병희, 송일병. 체간측정법에 의한 체질판별. 사상체질의학회지. 2002;14(1):51-66.
4. 김달래. 오음의 사상의학적 음성분석과 고찰. 사상체질의학회지. 2003;15(1):50-59.
5. 김영우, 이의주, 최선미, 김중화, 정성일, 이현민, 김종원. 사상체질진단을 위한 사상체질분류 검사II(QSCCII)의 연구. 사상체질의학회지. 2003; 15(3):11-21.
6. 권택부, 황치원, 설인찬, 이경옥, 황규원, 양동호, 홍세용. 한의학에서 분류하는 사상체질에 따른 이식항원 HLA-DR 유전자 polymorphism에 관한 연구. 대한내과학회지. 2000;58(1): 91-101.
7. 조동욱, 안선경, 김도균, 김대원, 지상은, 이의주, 홍석철, 고병희, 조황성. Amp-FLP을 이용한 사상체질의 유전적 분석 연구. 사상체질의학회지. 1997;9(2):163-173.
8. 김민희, 김경석, 지상은, 최선미, 조동욱. 유전자 분석법에 의한 사상체질의 연구. 사상체질의학회지. 1999;11(1):169-183.
9. 조동욱, 이창수, 고병희, 조황성. 유전자지문법을 이용한 사상체질의 유전적 분석 연구. 사상체질의학회지. 1996;8(2):151-163.
10. 최승훈, 임용빈, 이준우, 김홍열, 강철훈. 사상체질유형과 ACE 유전자 Type과의 상관관계. 사상체질의학회지. 1998;10(2):283-290.
11. 한성규, 지상은, 최선미. HLA Typing을 이용한 체질유전자 분석에 관한 연구. 사상

- 체질의학회지. 2001;13(1):97-105.
12. 이수경, 송일병. 사상체질의 유전적 다형성 연구. 사상체질의학회지. 2001;13(2):177-181.
 13. 유희룡, 임종순, 김윤식, 설인찬. DNA chip 을 이용한 사상체질의 연구. 대한한의학회지. 2004;25(1):72-84.
 14. 전국한외과대학사상의학교실. 개정증보 사상 의학. 집문당, 서울, 2004.

K C I