

유전학을 이용한 사상체질 감별법 개발 모델의 연구

안황용, 이용훈**, 임남규**, 김동희, 박종오*

Model study for genetic research of Sasang constitution confirmation method

Hwang yong An, Yong heun Lee**, Nam kyu Lim**, Dong hee Kim, Jong oh Park*

*Dept. of Pathology, College of Oriental Medicine, Daejeon University

**Pharmacodesign (Co. Ltd.)

The Sasang constitutional medicine, a traditional Korean medical typology, was initially proposed to the oriental medical field in 1894 by the Korean Confucian scholar Jae-Ma Lee. The key principle of his theory is that all humans can be categorized into four groups (Tae-Yang, Tae-Eum, So-Yang, and So-Eum) based on the function of major organs such as the heart, liver, kidney and lungs as well as other characteristics of the body. Now there are many scientists that try to confirm the method to classify the four group using genetics.

In this study, we hypothesized that some genes are may associated with Sasang constitution on the basis of the theory of Jae-Ma Lee who insist everybody can be divided from the birth and can not be changed during the life. Before the genetic research, we first suggest the genetic research model. It may be effective to carry out genetic research concerned with Sasang constitution.

I. 서론

사상의학은 심신의 조화를 치병(治病)의 우선으로 두며, 개인의 체질에 따라 동일한 질병, 동일한 증후라도 그 치료나 예방법등을 차등적으로 적용하는 의학으로, 만성병이나 난치병, 심신증 등의 질병군들에 대해 효능을 인정받고 있다^{1,2)}.

또한 사상의학은 유전학적 원리가 바탕이 되므로 서구에서도 대체의학(alternative medicine)으로서 관심대상이 되고 있다³⁾. 현대 의학은 맞춤의학(Personal Medicine) 등, 개별 환자별로 개개인에

맞는 최적화된 처방을 내리기 위한 노력이 계속되고 있는데, 이런 측면에서 사상의학은 현대의학의 맞춤의학과 함께 유전학적 연결고리를 가지고 있음을 알 수 있다^{4,5)}.

유전체학(Genomics)이란 인간의 유전체(Genome)를 폭넓게 연구함으로써 인간의 유전적 다형성을 연구하는 학문으로서 최근 미국에서 HGP(Human Genome Project)를 성공적으로 수행함으로써 세계 각국에서 이를 바탕으로 비교 유전체학, 기능 유전체학, 화학 유전체학 등 여러 분야에서 다양한 연구를 활기차게 진행하고 있다.

이러한 과정에서 DNA에 존재하는 단일 유전자 변이(Single Nucleotide Polymorphism : SNP)는 유전자 지놈에서 1,000개의 염기 당 1개의 빈도로 약 3,000,000개의 단일 유전자 변이가 존재할 것으

* 대전대학교 한의과대학 병리학교실

** 파마코디자인(주)

· 교신저자 : 박종오 · E-mail : davpark@dju.ac.kr

· 채택일 : 2004년 6월 25일

로 알려져 있다. 특히 단백질 해독과 직접 관련되어 있는 해독 단일 유전자 변이(cSNP : coding SNP)는 약 120,000개로 추정되어 가장 흔한 형태의 유전자 변이라 할 수 있음. 예를 들어 해독 단일 유전자 변이의 경우 단백질 기능에 영향을 주어 개인에 따라 특정 질병에 걸릴 수 있는 소인(predisposition)이나, 특정 약물의 민감성(sensitivity) 및 독성(toxicity) 반응에 영향을 미치기 때문에 최근 이와 관련된 연구가 활발하게 진행되고 있다.

인간의 체질을 각기 다르게 구분하는 사상의학과 유전적 다형성을 통해 인간의 다형성을 이해하는 유전체학은 서로 일맥상통함으로 유전체학을 통한 사상의학의 규명은 한의학이 세계로 약진할 수 있는 좋은 기반이 될 수 있을 것이다. 사상체질에 대한 과학적 규명이 없이는 사상에 기반을 둔 약의 사용, 치료 효과의 측정 등 추가적인 작업이 객관적으로 이루어지기 힘들고, 따라서 사상의학 등 한국 한의학의 세계 진출을 위해서는 서구에서 인정할 수 있는 유전학적 방법을 통한 사상체질의 규명이 시급하다고 할 수 있으며, 유전학과 사상의학과의 관련성을 규명하려는 연구가 다각적으로 진행되어 왔다^{6,7,8)}. 본 연구의 목적은 유전학과 최신통계학을 이용한 유전학을 이용한 사상체질 감별법의 개발을 위한 모델을 제시하는데 있다.

II. 본 론

1. 국내·외 관련기술의 현황 및 문제점

1) 현재 주류를 이루는 체질 진단의 방법은 체형검사, 비만도, 혹은 두면부 및 체간부의 형태학적 특징을 종합하여 판단하는 방법을 사용하고 있으나, 한의사의 주관적인 판단에 상당 부분 의지하여 객관성이 떨어진다는 지적을 받고 있다.

2) 기계적인 방법을 사용하여 맥파를 분석하는 맥진기, 지문을 이용한 사상 분류기, 체열진단기, 생기능진료기 등 진단기기를 통한 사상감별법이 연구되고 있으나, 아직까지 그 정확성 및 사상의학에 대한 과학적 규명의 토대를 마련하지 못하고 있다.

3) 실제 임상에 적용하기 위해 설문 형식으로 사상체질을 진단하는 시스템(QSCC)이 연구 중이지만, 역시 아직 설문 응답자의 주관적 작성에 의존하므로 정확성, 객관성의 측면에서 부족하다고 할 수 있다.

4) 현재 사상체질에 관한 유전학적 연구가 시도되고 있으나, 사상체질이 단순한 몇 개의 유전자에 의해 결정되기 보다는 전체 유전자인 유전체적 접근이 필요하므로 향후 보다 다양한 유전자에 대한 연구와 보다 많은 수의 샘플들을 통한 모형 검증 이루어져야 정확한 사상체질의 유전학적 특성이 규명이 가능하므로 많은 공동연구가 필요할 것이다.

2. SNPs를 이용한 감별모델의 요약

1) 1차 단계

① 표본군(sample)의 구성

- 사상체질의 판별이 다수의 한의사의 소견이 일치하는 대상을 선정

- 대상자의 동의하에 채혈

② 사상관련 SNP 도출

- 전달과 번역에 관련된 유전자 중 사상체질별로 RNA 발현량이 각각 다르게 나타난 유전자 중 체질별 RNA 발현량의 차이 정도와 기존의 한국 사상의학에서 임상적으로 차이가 나타난 표현형을 고려하여 5개 이상의 유전자를 선택하여 각각의 유전자에 존재하는 SNP를 공개된 DB에서 검색하여 단백질로의 해독부위(Coding region)와 발현량 조절 부위(Regulation region)에 존재하는 SNP만을 일차적으로 선택하여 Genotyping을 수행함

- 위에서 수행된 표본군의 SNP scoring과 사상체질과의 상호연관성을 χ^2 -test와 fisher's exact test 방법을 통하여 분석하여 사상체질과 유의한 관계(P-value < 0.1)를 갖고 있는 SNPs 도출

③ 사상관련 일배체형(Haplotype) 도출

- 근접한 SNPs들 간의 상호 연관불균형을 고려하여 근접한 다수의 SNP 간의 일배체형을 도출함

- DNA 해독을 토대로 도출된 염기서열에서 Heterozygoty 한 부분을 EM(expectation maximization) 혹은 MCMC(Markov Chain Monte

Carlo) 알고리즘 등을 사용하여 전체에서의 각각의 SNP의 빈도비율을 추정한 후에 이 추정량에 근거하여 Diploid된 염기서열을 두 개의 염기서열로 분해하여 각 Chromosome 별 염기서열로 변환한 후, SNPs의 LD(linkage disequilibrium)값과 Haplotype frequency를 산출함으로써 다량의 SNP의 연관관계와 haplotype들과 사상체질과의 연관관계를 χ^2 -test와 fisher's exact test 방법을 통하여 분석하여 사상체질과 유의한 관계(P-value < 0.1)를 갖고 있는 Haplotype 도출

④ 일배체형의 Key-SNPs 분석

- 향후 판별모형에 사용될 Haplotype들에서 비용절감의 효과를 갖기 위하여 구성된 일배체형 중에서 Key-SNPs를 일배체형 내에 있는 SNP들 간의 분석을 통하여 (1~2)개 정도만을 도출하여 Key-SNP만을 Genotyping 함으로써 전체 SNP를 Genotyping 하는 효과와 비교적 근사한 효과를 도출함.

⑤ 일배체형-일배체형 상호작용 분석

- 사상체질과 관련된 유전자들의 Haplotype을 도출한 후에 이런 유전자들 간의 network를 구성하기 위하여 Haplotype과 Haplotype 간의 상관관계를 알아보기 위한 여러 가지 데이터마이닝 통계적 기법들을 이용한다.

⑥ 사상판별 모형의 생성

- Key-SNP 분석과 일배체형-일배체형 상호작용의 분석을 통하여 과학적이며 저렴하고 정확한 사상판별 모형을 생성함.

⑦ 사상관련 유전자의 정보, SNPs 정보 및 Haplotype 정보의 데이터베이스화

- 연구를 통하여 획득된 SNP 정보와 Haplotype의 정보를 XML기반의 DB로 구축하여 사상의학기반의 한국 한의학의 표준화 연구에 기여하고, 세계적 표준화 추세에 맞는 데이터베이스 구축 경험을 공유함으로써 학적인 발전을 기대할 수 있음.

3. 모델 개발의 예

1) 샘플 모집

대전대학교 병리학교실과 협력관계의 한의원 및

병원 방문 환자들을 대상으로 약물반응, 설문지 검사 및 맥진 등의 방법을 사용하여, 감별 결과가 다수의 한의사에 있어서 모두 일치하는 사람을 대상으로 약 1000명의 피실험자를 선택한다. 선택된 피실험자를 대상으로 200명의 모델 생성 그룹을 임의로 선택하여 모델 생성시 사용하고, 나머지 피실험자는 모델의 검증 과정시 사용한다. 일단 모델생성에 선택된 그룹은 보다 안전하고 안정적으로 DNA를 확보하기 위하여 채혈을 통해 샘플을 획득하고 검증 그룹으로 분리된 사람들은 모근을 통하여 샘플을 획득한다.

2) 관련 유전자 추출

사상관련 후보 유전자군을 대상으로 일차적으로 전달(transport) 및 면역(immune response) 관련 유전자를 선택하여 일차적으로 DNA상에서 해독부위(Coding Region)와 발현량 조절 부위(Regulation Region)에 존재한다고 알려진 SNPs를 대상으로 선정한다.

3) SNP Genotyping.

이를 위하여 채혈한 혈액중 1cc로부터 추출 시약 및 kit를 이용하여 genomic DNA를 분리한 후 Pyrosequencing 방법을 통하여 SNP를 해독한다.

① Pyrosequencing은 4개의 효소(enzyme)를 이용하는 SNP Genotyping 분석방법으로, nucleotide 중합과정에서 방출된 PPi를 형광 검출하는 sequencing-by-synthesis 방법임.

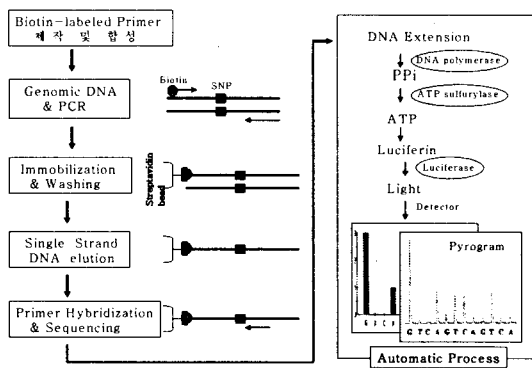
② DNA polymerase에 의해 DNA가 합성되는 동안 4개의 dNTP(deoxynucleotide triphosphates)를 한번에 하나씩 첨가하여 반응시킴. 이 반응이 일어나는 동안 dNTP에 붙어있는 PPi가 떨어져 나오며, 방출된 PPi가 ATP sulfurylase에 의해 ATP로 전환되는 반면 매 반응 과정에서 반응하지 않은 dNTP는 apyrase에 의해 분해 됨. 이렇게 만들어진 ATP는 luciferase라는 효소에 의해 luciferin을 oxi -luciferin으로 전환하는 데 사용되어 형광이 생성되며, 이에 해당하는 peak를 만드는 데, 이는 dNTP가 반응한 수에 비례하여 CCD(charge -coupled device) 카메라에 의해 실시간으로 측정됨.

③ 이 시스템은 특히 96개 또는 368 샘플을 빠른 시간(예를 들면, 96sample의 경우 5분)자동 분

석할 수 있기 때문에 대량 유전자 단일 변이를 스크리닝(Screening) 하고자 할 때 적합함.

④ 또한 이 시스템은 SNP 스크리닝 외에도 haplotyping, multiplex sequencing 등이 가능함.

⑤ Pyrosequencing은 TDGS와 마찬가지로 프로그래밍을 통해 특정 유전자 서열에 부합하는 프라이머(primer)를 디자인하고 특정 유전자 변이에 맞는 peak 패턴을 예측함으로써 실제 실험에서 얻은 결과와 비교할 수 있다는 장점을 가짐. 이는 비용 절감 및 분석을 용이하게 함.



【그림 1】 파이로 시퀀싱 과정의 이해

4) 일배체형의 추정(Haplotype Inference)

파이로시퀀싱을 이용한 Genotyping은 이형집합체(Heterozygosity) 부분의 해석에 모호함이 있으므로, Genotyping을 통해 해독된 개인별 염기서열에서 이형집합체(Heterozygosity) 부분을 EM (Expectation Maximization) 알고리즘(Excoffier

and Slatkin, 1995; Clark, 1990)을 사용하여 전체에서의 각각의 SNPs의 빈도 비율을 추정하였다. 이 추정량에 근거하여 이배체(Diploid) 염색체를 각기 분리하여 염기서열로 변환한 후, 각각의 SNPs를 분석하였으며, 가능한 경우 각 유전자별로 Haplotype의 분석까지 시도함.

5) 사상체질과의 관련성 연구

Haplotype과 사상체질과의 관련성을 분석하기 위한 통계적인 분석방법들로는 다음과 같은 것들이 있다.

* 분포 검정 : 각 변수에 대한 특정분포에 검정 방법

- 기존에 많이 사용되어 지고 있는 분포들의 분포함수(Distribution Function)와 각 변수(spot)에 대한 관측치들의 분포함수를 비교함.

* χ^2 -test : 범주형 자료에서 각 범주들간의 동질성 검증 방법

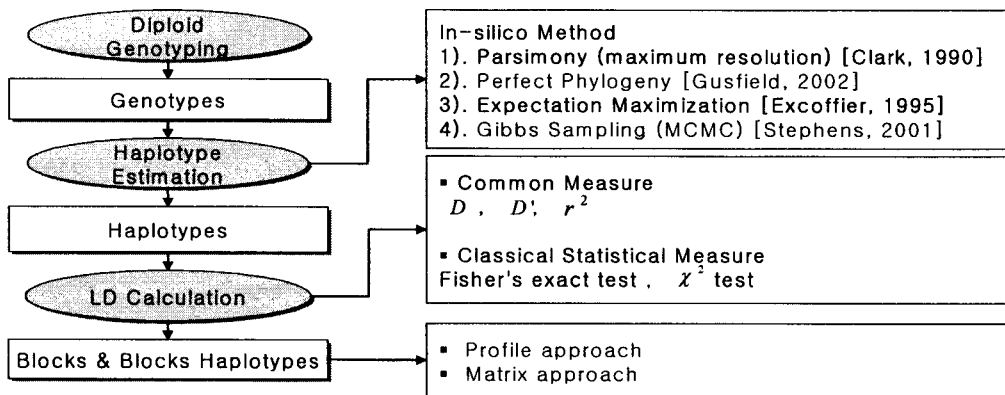
- 범주형 자료에서 각 범주들 간의 동질성을 검증하는 가장 기본적인 방법이며 이 외에 fisher's exact test등이 사용된다.

* ANOVA Analysis : 여러집단에서 단일 종속 변수의 대한 집단간 차이를 검정하는 통계적 방법

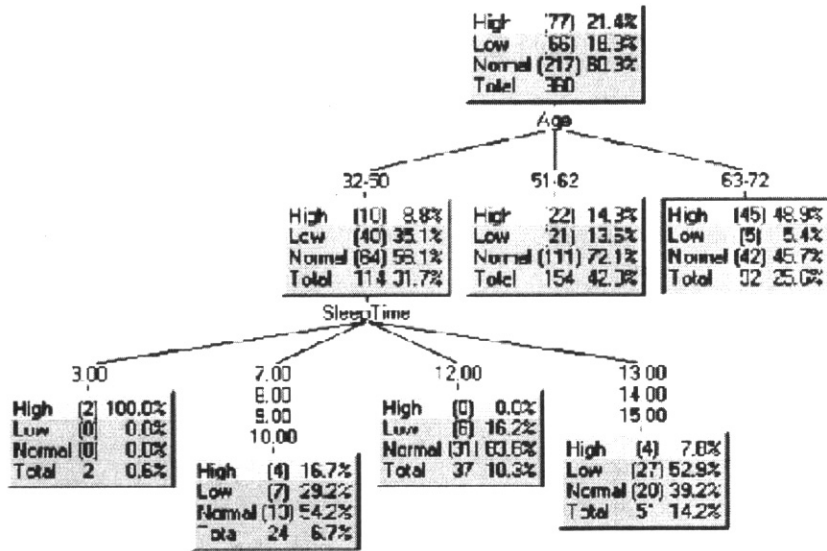
- 종속 변수에 대한 전체 분산과 각 집단에서의 분산의 합의 관계를 비교함으로써 집단간의 차이를 검정.

6) 사상판별 모형의 생성

- 사상체질과 관련된 Haplotype들을 바탕으로



【그림 2】 Haplotype Inference



【그림 3】 Decision tree 예제 - Age, Sleep 에 의한 세집단(High, Low, Normal) 분류

사상별 판별이 가능한 모델 생성함.

- 변수 형태

* 종속변수(Output Variable) : 태양인, 태음인, 소양인, 소음인(4 Categorical variable)

* 독립변수(Input Variable) : 5개의 이상의 Locus에 나타난 SNPs (5개 이상의 변수, 각 변수는 A,C,G,T 4개의 값을 가짐)

- 사상별 판별 모델 생성을 위한 데이터 분석

- 이러한 판별분석을 가능하게 해주는 기존의 통계학적 분류기법과 인공지능 기법들을 다음과 같은 것들이 있다.

① Decision Tree : 이산형 종속변수에 대하여 독립변수의 조건에 따른 분류 분석

* Decision Tree는 예측과 분류를 위해 보편적이고 강력한 방법임. 신경망구조 분석과는 달리 tree 구조로 규칙을 표현하기 때문에 이해하기가 쉬움.

* Decision Tree의 tree 생성 알고리즘은 다양하지만, 가장 보편적인 것으로 CART (classification and regression trees)와 CHAID(chi-squared automatic interaction detection)이고, 좀 더 새로운 알고리즘은 C5.0 등이 있음.

* Decision Tree의 장점 : 이해하기 쉬운 규칙을

형성, 많은 컴퓨팅 작업 없이 분류과정 형성, 연속 변수와 범주형 변수에 모두 사용가능, 예측과 분류부분에서 가장 효과적인 방법.

* Decision Tree의 약점 : 몇몇 Decision Tree 알고리즘이 이진분리를 하기 때문에 분리 가치의 수가 너무 많고 error-prone 발생, 가지 형성 시 컴퓨팅 비용이 많이 든다. 가지치기 시, 분리기준 설정 시 각각의 경우를 다 고려해야 되므로 조합의 경우를 모두 고려할 경우 그리고 가지치기를 할 경우 상당한 컴퓨팅이 제공되어야 함.

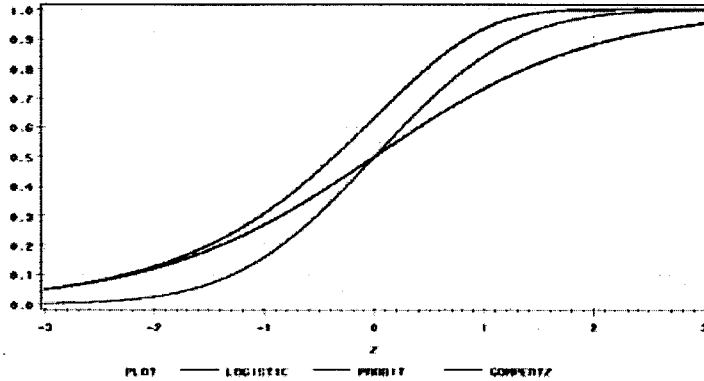
② 다항 로지스틱 회귀 모형 : 종속변수가 이산형인 경우, 종속변수의 특정한 값에 대한 확률값을 적당한 변환을 통하여 새로운 종속변수 변환 후 회귀분석 모형 적합

$$* \log \frac{p(y=\text{특정값}|x)}{1-p(y=\text{특정값}|x)} = \alpha + \beta x : \text{로}$$

지스틱 회귀분석 모형

* 로지스틱 회귀 모형은 장점은 원인과 결과와의 관계를 정확히 이해할 수 있음.

* 로지스틱 회귀 모형의 단점은 종속 변수의 특정값에 대한 확률값의 분포가 특정 모형을 따라야만 잘 적합됨.



【그림 4】 Link Function

③ Discriminant analysis : 이산형인 종속변수에 대해 독립변수들이 다변량 정규분포를 따르는 경우에 독립변수들을 바탕으로 판별 함수를 생성하여 판별기준을 확립하는 모형

* 판별함수의 도출이 가능하므로 로지스틱 회귀 모형과 같이 정확한 원인 규명이 가능함

* 다항 로지스틱 회귀모형보다 엄격한 변수 제약 조건이 필요함.

* SNP 데이터 분석에서는 독립변수들이 이산형 변수들 이므로 변수변환을 통해 연속형 독립변수로 치환)

④ 신경망 분석(Neural Network) : 신경망 분석은 사람 뇌의 구조를 모방한 데이터 모델링 기법

* 사람의 뇌 처럼 신경망은 Input set에서 학습을 하고, 패턴을 찾기 위해 학습결과를 모델의 파라미터를 조정하는 데 사용

* 은닉마디라고 불리는 독특한 구성요소에 의해서 일반적인 통계 모형과 구별

은닉마디(hidden unit) : 입력변수들의 결합(combination)을 수신하여 목표변수에 전달

연결강도(synaptic weights) : 결합에 사용되는 계수(coeffcient)

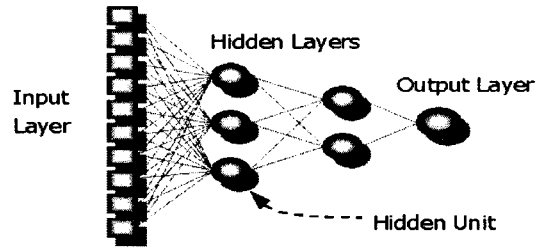
활성함수 : 입력값을 변환하고 이를 입력으로 사용하는 다른 마디로 출력

* 특정한 확률모형이나 가정이 필요없고 데이터의 형태에 대한 제한이 없으므로 대용량의 데이터에서 판별 모형으로 쉽게 적합이 가능.

* 원인과 결과의 관계를 이해하기가 난해하고 적은 용량의 데이터에서 Over Fitting의 위험이

존재함.

* 데이터에 대한 제약 조건이 없고 쉽게 사용이 가능하므로 현재 생물학 자료 처리에 있어서 가장 부각되어지고 있는 방법론임.



【그림 5】 신경망 분석

⑤ Genetic Algorithm : 진화라는 주제에 고무되어 나타난 계산 모델 알고리즘

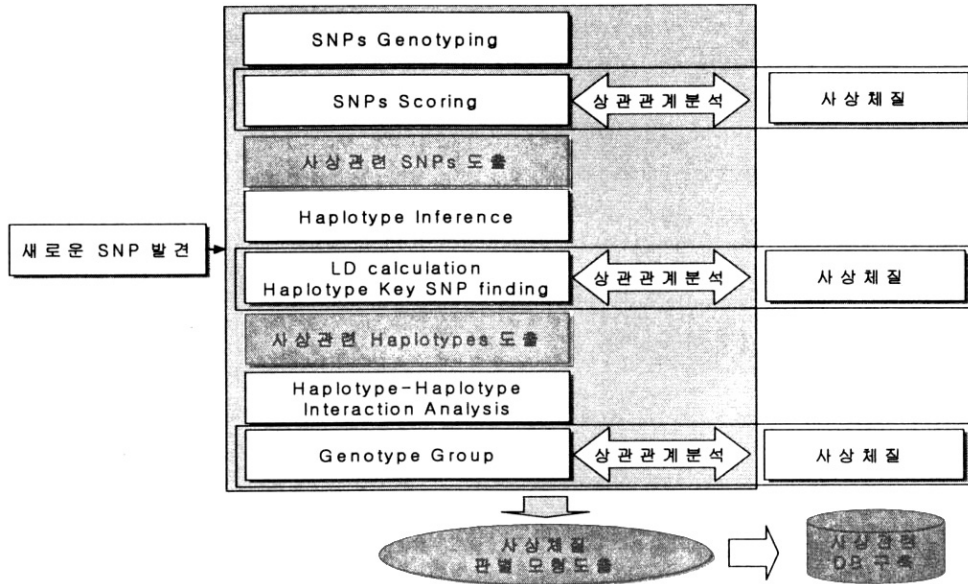
* Genetic Algorithm는 염색체 비슷한 자료구조(Chromosome-like Data Structure)를 사용하여 잠재적인 해를 특정한 문제의 해로 부호화하며, 이러한 자료구조에 재조합 연산자(Recombination Operator)를 적용하여 결정적인 정보를 보존한다.

7) 사상판별 모형의 생성

- 사상판별 모형 생성에 사용되지 않은 800명의 그룹을 통해 모형의 정확성을 검증한다.

4. 최종목표

- 사상체질 판별 시 기준으로 활용할 수 있는



【그림 6】 감별 모델개발의 최종목표 SNPs의 도출.

- 도출된 SNPs를 바탕으로 동일 일배체형 구간 (Haplotype Block) 안에 있는 SNP들을 통한 Haplotype 구축.
- 사상체질과 관련된 다수의 Haplotype을 통한 효율적이고 정확한 판별 알고리즘을 구현하여 객관적인 사상 판별 기준 및 기술 개발.
- 사상관련 유전자에 대한 정보의 종합적인 데이터베이스(DB) 구축.

III. 결론

이상의 연구모델을 통해 다음과 같은 효과를 기대할 수 있음.

- ① 한방사상진단과 특히 단일유전자변이(SNP)와의 연관 관계 규명을 통하여 한방 의학의 객관화 작업에 일조함.
- ② 이는 다음 연구에 도움을 주고 한방과 현대 의학의 접점을 찾는 데 중요한 역할을 할 것으로 기대됨.
- ③ 프로젝트를 통하여 얻은 한방 데이터와 유전자 데이터, 단백질 데이터를 데이터 베이스화 하여 한방에 관한 추후에 데이터 분석에 자료로 사용될 수 있도록 함.

④ 한방 데이터의 구축은 한방 자료의 표준화 연구에 기여. XML 등 세계적 표준화 추세에 맞는 데이터베이스 구축 경험을 공유함으로써 학적인 발전을 기대함.

⑤ 분자생물학(Molecular Biology)적 측면에서 사상 진단을 객관화하여 과학적인 체질진단의 지표를 제공함으로써 한국의 한의학을 세계화할 수 있음.

IV. 참고문헌

1. 송일병 외 : 사상의학과 난치성질환, 사상의학회지 14(3) pp. 1-6, 2002.
2. 김경용 외 : 난치병과 면역 그 사상의학적 접근, 사상의학회지 7(2) pp. 113-128, 1995.
3. 김종원 외 : 뇌졸중에 대한 한방치료법 연구 (증치의학과 사상의학) 및 한방, 양방, 양·한방 협진치료 효과에 관한 연구, 사상의학회지 10(2) pp.351-429, 1998.
4. 조동욱 외 : 사상의학의 객관화를 위한 유전적 분석 연구, 한의학연구원 2(1) pp. 402-406, 1996.
5. 하만수 외 : 유전자를 이용한 체질유형감별의 방법론에 관한 고찰, 사상의학회지 11(2) pp. 185-194, 1999.

6. 한성규 외 : 체질유전자 분석에 관한 연구, 사상의학회지 15(1) pp.109-117, 2003.

7. 박종오 외 : 태음인, 소양인, 소음인별 Cytochrome P450 유전자의 2D6, 2C9, 1A2 DNA 부위에 대한 SNPs과 Haplotype에 관한 연구, 동의생리병리학회지 16(6) pp. 1201-1206, 2002.

8. 김선형 외 : 사상체질과 아토피성 질환 유전자의 다형성(IL-4 polymorphism)에 관한 연관성 연구, 사상의학회지 14(2) pp.98-105, 2002.