



## 생정보학 기술을 이용한 유전체학 동향: 오믹스(Omics)

박종화

(국가생물자원정보관리센터, 한국생명공학연구원)

### 서론

오믹스(Omics: 체학)는 생명학(Biology와 Life science)의 새로운 방법론이다. 이것은 새로운 기술, 새로운 종류의 데이터와 지식의 축적에 의해 최근에 부각된 것이다. 과거 생명학자들은 생명의 근본원리는 복잡한 네트워크의 분석을 통해서만 가능하리라고 생각해왔다. 프란시스 크릭이 DNA의 구조를 밝힌후 생명연구에서의 핵심이 정보의 흐름이라는 것을 깨닫고, 코돈(codon)의 연구에 매진하였다. 그러다가, 점차 뇌신경 네트워크에 관심을 가지게 되고, 죽을 때까지 뇌의 기능을 연구하게 되는 것은 생명학의 주요 문제 인식의 단계를 한 인생으로서 잘 보여주는 것이다.

생명학의 핵심은 에너지를 최대한 활용하기 위한 효율적이고 동력학적인 정보처리에 있다. 그러므로, 현대 생명학의 핵심분야로서 생정보학(Bioinformatics)가 부각되는 것은 매우 당연스런 결과이다.

이 글에서는 생명학의 핵심분야인 유전체학, 생정보학, 그리고, 유전체, 단백질, 발현체, 상호작용체들을 아우르는 체학의 기본 개념과 세계적 동향을 논한다.

### 유전체학

유전체를 연구하는 학문이다. 유전체학은 유전자들의 총합을 체학(omics)적인 관점에서 연구하는 과학이다. 통합적, 관계적으로 유전자들을 분석하고 서로간의 조절을 모델링하는 것을 목적으로하며, 서열해석, 유전자해석, 유전자 네트워크 지도화, 유전체간의 다양성비교 등의 분야를 가지고 있다.

유전체학은 현대 생명학의 핵심분야이다. 유전체

학은 영국 케임브리지의 MRC센터에서 시작되었고, 현재 창시자 생어의 이름을 딴 생어센터에서 활발하게 계속 연구가 되고 있다. 미국에서는 1980년대 중반 영국을 추월하기 위해 전략적으로 국회입법을 통해 국가생명공학정보센터(NCBI)가 세워졌고, 하버드대 교수등이 참여한 로비를 통해 인간 유전체프로젝트가 시작되었다.

### 유전체학의 역사

영국 케임브리지의 프레드 생어(Fred Sanger)에 의해 확립되었다. 생어(Frederick Sanger), 톰슨(E. O. P. Thompson) 과 한스 터피(Hans Tuppy)는 1953년 인슐린(Insulin)의 A와 B 체인의 단백질 서열을 화학적으로 해석하고 나서 노벨상을 받게된다. 그후 1970년대에 deoxy DNA 서열해석법을 만들었다. 그 서열해석 방법의 첫번째 목표는 phi X174 바이러스의 유전체를 해석하는 것이었다. 이것이 최초의 유전체로, Journal of Molecular Biology에 1970년대에 발표되었다 (Sanger, F. & Coulson, A. R. (1975) J. Mol. Biol. 94, 441-448). phi X174 바이러스는 5386 개의 염기를 가졌다. 최초의 바이러스 유전체를 다음으로, 최초의 세포소기관(Organelle)의 유전체인 마이코콘드리아의 유전체도 곧이어 해석이 되었다. 이러한 큰 서열들이 나오자, 자연히, 그것을 시각화, 저장, 분석을 할 컴퓨터 소프트웨어들이 필요해졌다. 같은 랩(MRC 센터의 LMB)의 로저 스타던(Roger Staden)이 1977년에 NAR학회지에 컴퓨터로 유전체 서열을 처리하는 최초의 패키지에 대한 논문을 발표하였다. 1995년대에 들어와 생어법을 사용한 최초의 박테리

아 유전체인 *Haemophilus influenzae*의 유전체가 발표되었는데, 이 때를 기점으로 유전체학은 세계적으로 각광을 받기 시작했다. 한 생명체를 운용하는 모든 유전자의 리스트를 얻고, 이를 변형할수 있다는 기대감이 넘쳐났다. 또, 이때쯤에 옥스포드대학의 서던(Southern)이 개발한 반응(hybridization)기술이 바이오칩으로 응용되면서, 기능유전체학이 생겨났다.

## 유전체학의 분야들

유전체학의 대표적인 분야들은 다음과 같다.

- 1) 비교유전체학(Comparative genomics) : 서로 다른 종들의 유전체를 생정보학적으로 비교하여, 특이한 유전자, 기능적 상동성등을 찾아내는 유전체학
- 2) 기능유전체학(Functional genomics) : 1990년대 중반에 생긴 말로, 전사체와 발현체(단백체포함)의 정보를 유전체 서열과 연관시켜서 대규모 기능유전자 스크리닝을 하는 유전체학.
- 3) 구조유전체학(Structural genomics) : 1997년경 생긴 말로, 유전체내의 모든 3차원 단백질의 구조를 예측, 실험적 해석을 통해 3차원 구성요소 모두를 확인하겠다는 유전체학.

유전체학과 필연적으로 연결된 학문이 생정보학이며, 유전체학은 이론생명학적만으로 볼때는 생정보학의 한 분야이기도 하다. 생정보학은 일반적으로 실험을 포함하지 않는다. 그러나, 미래에는 실험도 많은 부분이 자동화 되고, 컴퓨터에서 가상의 실험을 하는 경우가 늘어나게 될 것이므로, 점차 이론과 실험생명학의 차이가 줄어들 것이라고 예측된다.

## 이론 생명학의 핵심: 생정보학

생정보학의 최종목적은 생명의 모든 현상을 정보적 모사하고 이해하는 것이다. 그럴려면, 지구상의 모든 생명체들이 어떤 여정을 걸어와서 현재와 같은

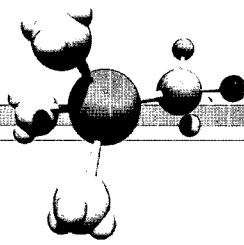
복잡한 정보처리객체들을 이루었는가를 이론적으로 분석하는 것이 필요하다. 생명체가 걸어온 길을 현재까지 우리는 진화라는 개념으로 과학적, 사회적으로 이해하고 있다. 그래서, 진화는 생명학과 생정보학의 핵심 원리이다. 생명학에서 진화는 증명된, 법칙 수준의 이론이다. 진화의 중요성은 그것이 상태의 변화들을 시간의 축에서 잘 설명이 되기 때문이다.

## 생정보학과 유전체학에서의 시간의 중요성

시간을 생명학을 연구하는 데 매우 중요한 도구이고, 시간개념에서 진화의 이론적 전개가 가능하고, 진화는 생명학에서 거의 유일하다시피한 법칙이 되었다. 이 우주에 만약 신이라는 존재가 꼭 있어야 된다면, 그것의 가장 가까운 이름은 '시간' 이다. 왜냐하면, 시간은 상태변화를 뜻하고, 정보를 생산하며, 이 우주의 존재 전체에 의미를 주는 동력 역할을 하기 때문이다. 시간은 의미의 동력, 혹은 정보의 동력이라고 말할수 있다. 시간이, 똑딱 똑딱 뛰기 시작하면, 상태변화를 보게 되는데, 이런 상태변화는 그 구성입자들의 상호작용의 변화를 말한다. 무엇인가, 조금 전의 시간에서의 상호작용이 변화를 해서, 이제는 다른 상호작용을 하기 때문에 상태변화가 일어난다. 상호작용은 여러 레벨에서 일어날 수 있다. 인간들 사이의 상호작용 변화는 인간 사회의 상태변화를 초래한다. 생정보학은 필연적으로 생명객체들의 상호작용을 다루게 된다. 이런 상호작용하는 생명객체들의 정보를 종합적으로 모아놓은 것이 유전체이다.

## 생정보학적 상호작용

생명학에서 가장 궁극적으로 중요한 데이터의 근원은 각종 분자 상호작용나, 상호작용의 네트워크이다. 생명학 맨 밑바닥의 정보는 분자나 원자들의 상호작용에서 나온다. 상호작용이 없으면 네트워크도 없고 복잡함도(complexity) 없다. 그것은 고등한 유전체나 생명의 부재를 뜻한다.



## 생정보학과 주역, 태극기, 사주팔자

상호작용이 현상을 촉발하는 것의 본질이라는 것은 철학적 관심이 있는 사람이면 인생에서 한번쯤 생각하게 된다. 한국에서는 이런 상호작용에 관한 철학이 발달했다. 주역의 근본은 음과 양이라는 짝을 이루는 완전치 못한 두 조각이 상호작용을 함으로서 변화가 발생하고 우주가 진행되어 나간다는 것을 기본으로 한다. 우주의 근본이 음과 양의 합성을 통해 생긴 정보 처리적인 객체라고 가정한다. 이진법적 음과 양이 상호작용을 할때, 최초의 신호가 발생하고, 정보들이 유전체내에 쌓인다. 이 정보들은 여러가지의 궤로 표현되어질 수 있으며, 모든 우주의 현상, 특히 인간과 관계된 생명현상은 이러한 정보처리적인 표에 의해 분석이 된다는 것이다. 그것의 한 결과로 고등 생명체의 발생, 발달 과정도 대략 4개의 큰 기둥과 8개의 글자들의 조합으로 분류가 된다고 설명을 해왔다. 이것은 현대와 미래 생정보학에서, 큰 범위의 연구를 매우 추상적으로 한 것이다. 한국의 국기인 태극기가 이러한 상징을 가지는 것은 이상하리만한 우연이다. 이런 측면에서, 생정보학의 기원은 인생을 표로 만들어 예측하고 싶어했던 고대인들에게도 있는 보편적인 학문이라 주장할 수도 있다. 그 중에서도, 한국, 중국등의 사주팔자와 관계된 주역등이 치밀한 통계학적 표를 만든 경우이다. 이 표는 유전체를 간단한 시간좌표화한 것이라고 할수 있다. 인간 사회의 변화를 도표화할때 사주팔자 시간을 쓴 것이다. 이 표의 장점은 시간 좌표가 매우 깨끗하고, 규칙적인 것이기 때문이다. 시간이 아닌 공간 좌표를 쓸수도 있다. 그러나, 공간 좌표는 지구상에서 계속 변하는 것이라서, 시간이 지나면 통계자료가 맞지 않게 된다. 인간 유전체도 이러한 일종의 좌표속의 데이터 덩어리이며, 그 좌표는 시간, 공간, 기타 환경들이고, 축적된 유전데이터는 그 속에서 지속적으로 적응해온 네트워크덩어리이다.

## 생정보학의 두가지 중요한 변수: 에너지와 정보

유전체라는 네트워크 덩어리의 존속에 있어서, 생

정보학적으로 가장 중요한 두가지 변수는 에너지와 정보이다. 세포내에서 에너지의 양이 어떻게 유전적으로 조절되는가와 어떤 정보가 조절되는가가 생명체의 생존을 결정짓는다. 비유적으로 말하면 먹을 에너지가 있어야 되고 그것을 활용코자 할때 지속적으로 에너지를 구할수 있는 정보가 있어야 한다. 예를 들면 대장균이 에너지를 냄새맡고(정보) 그쪽을 향해가기 위해 ATP라는 분자속의 화학에너지를 경제적으로 사용하는 것이다. 생물세포내에서 이런 에너지와 정보를 처리하는 가장 핵심 부품은 단백질이다. 그래서 지구에서 생명현상을 연구한다는 것은 단백질을 연구한다는 것과 근본적으로 같은 것이다. 단백질체는 유전체의 일종의 도구체 혹은 기능체라고 할수 있다.

## 에너지, 정보 그리고 체학(Omics)

에너지를 가장 효율적으로 쓴다는 것은 조그만 박테리아 뿐만아니라, 거대한 국가생명체인 대한민국에서도 중요한 문제이다. 그것을 가장 경제적으로 쓰기 위해, 경제학이 필요하고, 교통망, 통신망등이 필요하다. 여러 사회간접자본시설 중에서 가장 중요한 것이 정보이다. 어디에서 어떻게 에너지를 활용할수 있는가의 정보가 적시적소의 에너지 활용을 결정짓는다. 이러한 에너지 활용의 밑바닥에는 어떤 원칙이 있을까? 그러한 원칙을 연구하는 것이 체학(오믹스)라고 할수 있다. 오믹스는 각각의 체의 단위에서 정보객체들이 어떻게 상호작용을 해서, 그 레벨에서 가장 안정적이면서도, 효율적으로 정보처리를 하는가를 연구하는 것이라고 할수 있다. 예를 들면, 유전체라는 거대한 유전 정보덩어리는 어떠한 조직적 구조를 가지고 있으면서, 정보를 효율적으로 저장하고, 보존하고, 다음 단계로 넘겨주는가의 문제를 가지고 있다. 단백질체는 단백질들이 어떠한 종류를 가지고, 세포의 기능들을 효율적으로 처리하는가를 궁극적으로 연구한다. 그러기 위해서 먼저 어떤 단백질들이 과연 발현이 되는지 동정(identification)을 해야하고, 그들이 어떤 기능을 가지고 있고, 어떤 상호작용을

하는지등을 도표화 해야 한다. 이런 과정에서 단백질은 주어진 에너지를 가지고 인간이 정한 그 영역에서 정보처리를 하는가를 연구하는 것이다.

**진화하는 오믹스**

오믹스는 각종 체학을 아우르는 말이며, 각각의 체학은 연구자들이 필요에 따라서 정의를 내릴 수 있다. 각 체학은 다른 체학들과 중복영역이 있을 수도 있고, 또, 독립성이 떨어질수도 있다. 현재 많은 오믹스 용어들이 생성되고 있는데, [Http://omics.org/](http://omics.org/)에 가면, 그런 용어들의 정의를 잘 볼수가 있다. 오믹스중에서 중요한 것은 미래에 살아남을 것이고, 그렇지 않은 것을 사라질 것이다. 다음의 그림 1은 그런 오믹스의 많은 영역을 분자생명학적인 측면에서 보여준다.

**오믹스 매트릭스**

위의 오믹스의 영역들을 현재 주요한 몇개의 영역과 그들이 다루는 정보의 종류로 나누면 다음의 오믹스 매트릭스를 만들수가 있다. 오믹스 매트릭스의 첫 번째 층은 유전체이다. 역사적으로 가장 오래되었고, 가장 기초적인 정보가 저장되어있기 때문이다. 그 뒤로, 단백질, 전사체, 발현체, 상호작용체, 기능체, 문헌체등이 올수가 있는데, 기능체는 OMIM 과 같이 질병이나, 특수한 기능(효소기능) 데이터베이스 등을 말한다. 문헌체는, 인간이 축적한 많은 생물지식도, 하나의 생명학적 체로 간주하고, 그것을 어떻게 잘 활용하는가 하는 것이 한 이슈이다. 인간은 진화를 통해 유전체에 엄청난 양의 정보를 저장해 왔는데, 최근에는 용량이 너무 커져서, 컴퓨터에다 유전정보 등의 생명정보를 저장하게 되었다고 보편된다. 각각

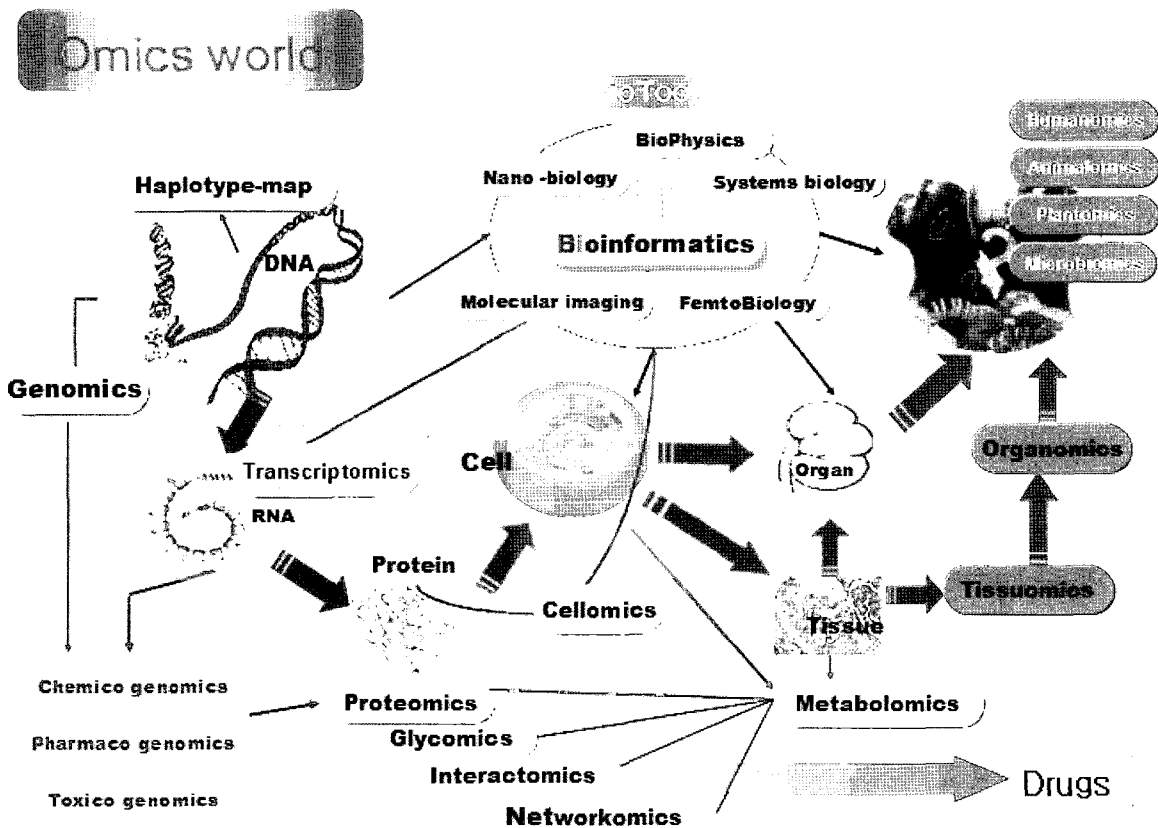
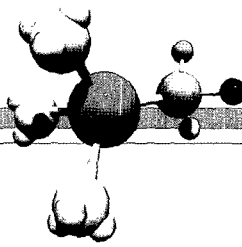


그림 1. 각종 오믹스(체학)의 영역



의 체계분야는, 서열, 구조, 발현, 상호작용 등의 정보 타입으로 분류되어서 저장 및 활용 될수가 있다. 정보 타입의 가장 꼭대기에 있는 것은 네트워크이다. 많은 오믹스 객체들이 네트워크형태로 기능을 하고 있기 때문이다. 예를 들면 유전체학은 유전자들의 네트워크를 연구하는 학문이라고 해도 틀리지 않는다.

그러므로, 이런 체들을 통합하는 오믹스를 통합체학(Integromics)라고 한다. 최근에 생정보학과 관련된 분야인 생명계학(systems biology)은 이런 통합체학의 동의어 또는 한 분야라고 볼 수 있다. 그림2 참조.

## 통합체학

오믹스는 구성요소들을 유기적으로 네트워크로 통합함으로써 생기는 기능을 연구하는 학문이므로, 오믹스내와 다른 오믹스간의 통합을 통해서 생명현상을 더 체계적으로 연구를 한다. 유전체와 상호작용체를 잘 통합하면, 유전자들의 조절과 단백질들의 상호작용을 통한 새로운 유전자 기능을 찾을 수가 있다.

## 네트워크오믹스 (Networkomics)

과거에는 유전자들간의 상동성을 비교하여서 계통도를 그렸다. 그러나, 계통도는 근본적으로 문제가 많은 가정들을 가지고 있고, 유전자나 생명개체는 계통도로는 한계가 많은 특성들을 많이 가지고 있다. 1995년경 부터는 이러한 점에서, 계통도가 아닌 계통네트워크(Phylogenetic network)를 사용하기 시작하게 된다. 같은 경향으로, 생명을 이해하려면, 생명객체들간의 네트워크를 이해하는 것이 가장 중요한 것

# The Omics Matrix and Integromics

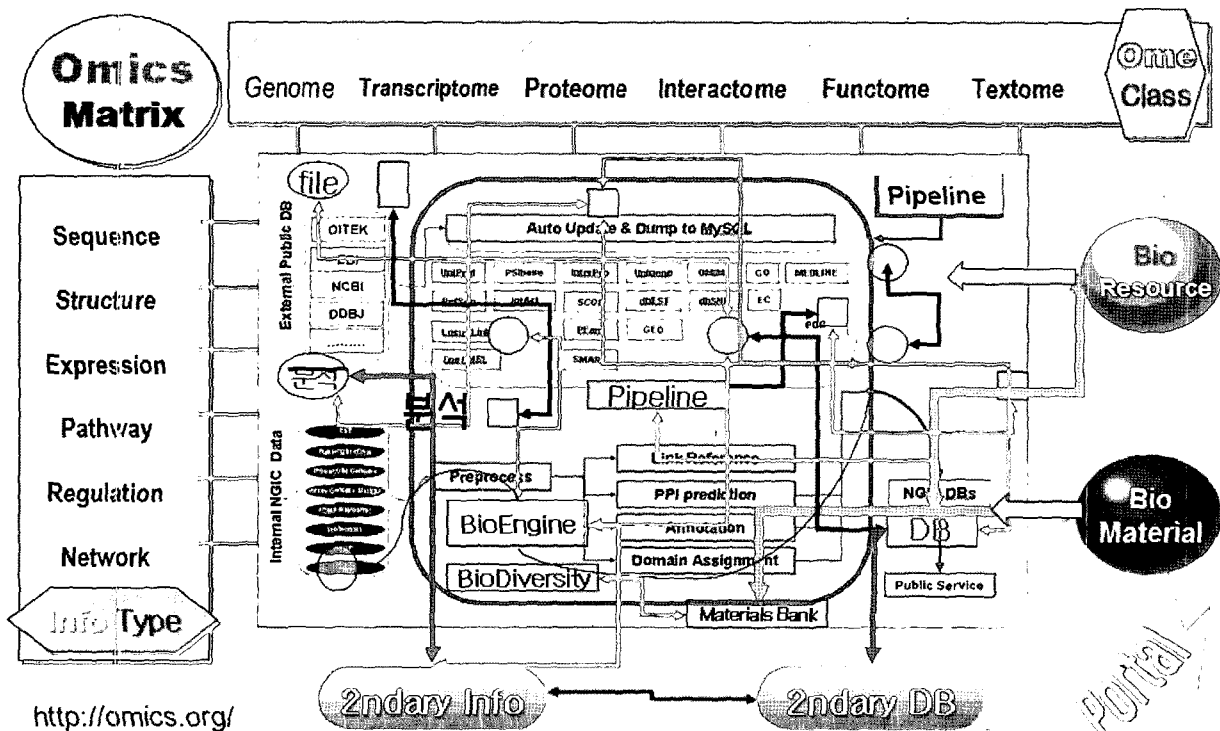


그림 2. 오믹스 매트릭스와 통합체학

으로 부각이 되고 있다. 앞으로 생명학의 모든 영역에서 네트워크분석, 그래피론(Graph theory), 베이저안 네트워크등이 이용될 것이라고 예측된다.

### 유전체학의 최근동향

유전체학은 2003년에 인간유전체의 완성을 두고 후기 유전체학(Post genomics)이 도래할 것이라고 하는 학자들이 많았다. 그러나, 앞으로 생명학이 존재하는 유전체학이후의 유전체학은 존재할수가 없다는 것이 저자의 생각이다. 이것은 유전체가 생명정보의 가장 기초적인 네트워크 정보덩어리이고, 모든 지구상의 생명체는 이런 유전체를 정점으로 진화를 해왔기 때문이다.

1990년대 중반과 2000년대 중반의 유전체 연구의 가장 큰 차이는, 많은 양의 유전체 수가 증가함으로써, 과거에는 불가능했던 비교유전체학이 급격히 발달했다는 것이다. 이제는 유전자들을 비교하는 것이 아니라, 유전체들을 비교하는 시대가 도래했다.

### 변이체학의 대두

매우 비슷한 아종이나, 같은 인종사이의 미세한 유전체를 비교하는 변이체학(Variomics: <http://variome.org>)이 부각되고 있다. 자동으로 수십만개의 변이염기를 뽑아낼수 있기 때문에, 이제는 컴퓨터를 이용한 막대한 양의 생명정보 분석이 필수적인 시대가 왔다. 변이체학은 기존의 STR, RFLP, micro satellite 등의 마커에서 이제는 단일염기다형성(SNP)로 추세가 변하고 있고, 따라서 바이오칩의 활용이 증대되고 있으며, 머지 않아, 막대한 바이오칩의 시장이 형성될 것이다.

### 유전체학의 대중화 기업화

유전체학은 개개인의 차이를 판별할 정도의 해상도를 최근에 가지게 되었다. 인간유전체도 이제는 매우 적은 비용으로 해석을 할수 있고, 아피메트릭스의 500k 혹은 1 mega SNP 칩을 쓰면, 개개인의 유전체의 특징을 찾아낼수 있게 되었다. 이것은 개인유전체학(Personal Genomics <http://personalgenome.net>) 시대가 왔음을 말한다.

미국 하바드 의과대학에서는 현재 자원자를 받아서 개인유전체학 프로젝트가 진행되고 있다 (<http://www.thepersonalgenome.com/>). 개인유전체학은 다른 여러개의 체학들과 융합 분화 하면서, 맞춤의학에 필요한 모든 단계의 생명정보처리 역할을 하게 될 것이다.

### 결론

유전체학은 체학의 발전과 더불어 개인화, 산업화, 자동화, 대량화를 겪으면서 미래 생명학의 핵심분야로 발전할 것이다. 체학적인 접근 방법은 생명학의 연구방법을 혁신화 할 것이며 그와 관련된 많은 파생된 분야를 형성할 것이며, 생명학의 폭발적인 분화를 촉진케 할 것이다.

### 참고 사이트들

<http://omics.org/> 오믹스 홈페이지

<http://personalgenome.net/> 개인유전체학 페이지

<http://www.thepersonalgenome.com/> 개인유전체 프로젝트 홈페이지

<http://variome.net/> 변이체학 페이지

<http://www.kobic.re.kr/> 국가생물자원정보관리센터 홈페이지