

동물학논단

인간유전체 프로젝트의 진행상황 및 전망

송 규 영

현재 울산대학교 의과대학 (교수)

1. 인간유전체프로젝트의 진행상황

1953년 제임스 왓슨과 프란시스 크릭에 의하여 생명체 유전의 신비를 간직한 DNA의 이중나선구조가 밝혀진 이래 인간을 위시한 여러 생명체들의 모든 유전정보를 규명하는 것은 생명과학의 숙원이었다. 생명체가 가지고 있는 유전자를 포함한 모든 DNA를 통 털어서 유전자(**gene**)와 염색체(**chromosome**)의 합성어인 “유전체(**genome**)”라고 부른다. 인간의 유전체를 모두 규명하겠다는 대담한 계획이 바로 “인간유전체 프로젝트(Human Genome Project; HGP)”이다.

인간유전체 프로젝트는 1985년 미국 에너지부(Department of Energy, DOE)가 제출했던 “인간의 유전적 돌연변이를 검출하는 기술들(Technologies for Detecting Heritable Mutations in Human Beings)”라는 보고서로부터 유래하였다. 이 보고서는 대규모 예산지원을 받는 거대하고, 복잡적이고, 획기적인 계획 없이 기존의 방법만으로는 돌연변이 측정에 한계가 있다고 지적함으로써 인간유전체 프로젝트라는 착상을 낳게 하는 계기가 된다. 1986년 에너지부는 Santa Fe에서 국제회의를 열어 인간유전체 프로젝트의 실행 가능성을 타진하였다. 그 결과 그것이 실현가능하며 생물학의 뛰어난 업적이 될 것이라는 데는 이견이 없었다. 그러나 그 프로젝트를 시행하는 것에 대한 회의론도 있었다. 인간유전체 3×10^9 bp 중 발현되는 유전자는 약 2% 정도로 예상되었는데, 그렇게 많은 돈을 투자하여 유전체의 염기서열을 모두 규명하는 것을 무모하게 여기는 학자들도 있었고, 발현되는 모든 유전자를 찾더라도 각각의 기능 연구가 상당히 어려운데 우선 그 모든 염기서열

부터 규명하려는 대단위 생물학 연구를 달가워하지 않던 학자도 상당히 있었다. 그럼에도 불구하고 미국내 과학자들과 심도 있는 토론을 거친 에너지부는 1987년에 “인간유전체 계획 추진에 관한 보고서(Report on the Human Genome Initiative)”를 발표하였다. 그때 제시된 세 가지 목표는 1) 인체염색체의 정교한 물적지도(physical map) 작성, 2) 인간유전체 연구를 위한 제반 기술 및 설비 개발, 3) 정보통신망과 데이터베이스 수용량의 확장이었다.

1988년에 미국 과학기술분석과(US Office of Technology Assessment)와 국립연구위원회(National Research Council)의 보고가 나왔고, 미국립보건원(NIH)은 국내의 다른 기관들과 유전체 연구를 통합조정하기 위하여 인간유전체연구과(Office of Human Genome Research; 나중에 National Human Genome Research Institute가 됨)를 개설하였다. 그 해 미 국회는 1990년 10월 1일 시작하는 총 예상 비용 30억 달러가 소요되는 15년 계획의 유전체 프로젝트를 승인하였다. 이 프로젝트의 주요 목표는, 1) 인체 및 소수 모델생명체의 유전체 염기서열 규명, 2) 연구결과의 수집, 저장, 분배 및 분석 능력 개발, 3) 상기목표 달성에 필요한 제반기술개발이었다.

이 프로젝트가 시작된 지 9년째 되는 현재 이미 *E. coli*, *Helicobacter pylori*를 비롯한 20종의 미생물과 효모, 그리고 다세포 생물인 *Caenorhabditis elegans* 등 모델 생물체들의 전체 염기서열이 밝혀져서 그 결과가 공공 데이터베이스에 등장하였다. 뿐만 아니라 '99년 9월 9일 셀레라 유전체사(Celera Genomics Corp)는 초파리 전체 유전체에 해당하는 18억 염기쌍의 서열 작업을 완료했다고 발표했다. 앞으로 몇 달 내에 그 서열들을 모아서 정확하게 연결된 전체 초파리 유전체 염기배열로 짜 맞춰질 것이다. 인간유전체 연구도 개발된 제반기술을 바탕으로 급속하게 추진

되어서 제1차(1991~1995), 2차(1993~1998) 5개년 계획에서 제시된 목표들을 모두 달성, 혹은 초과달성하기에 이르렀다. 이미 인체염색체 24개에 대한 염색체 및 물적지도 작성이 완료되었으며, 제반기술들도 모두 갖추어서 현재 이 프로젝트는 절정에 다다른 상태이다. 예를 들자면 1998년까지 염기서열규명 예상목표는 주로 다른 생물들의 유전체 8천만 bp였으나, 그때 이미 인간유전체 염기서열 1억8천만 bp를 비롯하여 대략 그 3배에 달하는 염기서열이 데이터베이스에 등장하였다. 1998~2003년까지 제3차 5개년 계획은 민족들간의 인체 DNA염기서열의 차이가 특정 질병의 발병 및 보호에 미치는 영향, 개체 단위에서의 유전자 기능연구를 위한 신기술 및 전략 개발, 개인의 신원과 민족적 배경과 유전학의 연결, 그리고 그런 사항들의 철학적 종교적 의미 연구 등의 새로운 방향들이 포함된다.

인간유전체 프로젝트에 급속한 변화는 1998년 5월에 미국 유전체연구소(TIGR) 소장 크레이그 벤터(J. Craig Venter) 박사와 퍼킨 엘머사(Perkin-Elmer Corporation)가 셀레라 유전체사를 설립하여(Celera Genomics Corporation, Web://www.celera.com) 인간유전체 염기서열규명을 3년 안에 완료하겠다는 담대한 계획을 발표함으로써 시작되었다. 이에 영국의 웰컴사(Wellcome Trust)도 생거연구소(Sanger Center)에 대한 지원액을 당초 계획의 두 배로 늘려서 인간유전체 중 1/3의 염기서열을 규명하겠다고 발표하였다. 그러자 미 국립 인간유전체연구소(NHGRI)도 프로젝트 완료시기를 예정보다 2년 앞당긴 2003년으로 수정 발표하였다가, 최근에는 다시 이것을 18개월이나 앞당긴 2000년 3월까지 90%를 완료할 계획이라고 발표했다. 새로운 계획을 위하여 전에는 유전체 염기서열을 대부분 10번씩 sequencing하던 것을 이제는 5번씩만 하여 90% 초안을 완성하고자 한다. 찬성하지 않는 연구자도 있었지만 공개적인 반대는 없었으므로 '99년 Cold Spring Harbor 회의에서 인간염색체를 염기서열센터 별로 나누는데 합의하였는데 그 내용은 다음과 같다: Baylor 대학(3, 12, X), MIT(17, 나머지 필요한 부위), 생거연구소(1, 6, 9, 10, 13, 20, 22, X), 미국 에너지부(5, 16, 19), 워싱턴대학(2, 3, 7, 11, 15, 18, Y), 프랑스(14),

독일(8, 21), 일본(8, 18, 21, 22). '99년 10월 3일 현재 인간유전체는 14.8 %가 완료되었다.

인간유전체 염기서열규명 완료목표인 서기 2003년은 제임스 왓슨과 프란시스 크릭이 DNA의 이중나선구조를 발견한 50주년이 되는 해이기도 하다. 왓슨은 1989년부터 1992년까지 국립보건원의 인간유전체 프로젝트 첫 책임자로 근무하기도 하였다. 미국과 영국은 바로 그때에 맞추어 인간의 유전체 염기배열규명을 완성함으로써 그 기념비적 업적을 축하하고 역사적으로 한 획을 그으려고 서두르고 있는 것이다.

2. 인간유전체프로젝트 이후 전망

2.1. 유전체 기능 및 생리학적 연구(Functional Genomics)

인간유전체 프로젝트 완료, 즉 인간유전체 전체염기서열 결정 완료 후의 나아갈 길을 크게 두 가지로 나누어 볼 수 있다. 하나는 functional genomics(기능유전체학)로 유전자의 발굴 및 기능연구이고 또 다른 하나는 comparative genomics(비교유전체학)로 인간의 표준염기서열결과를 바탕으로 인종, 질병발병, 의약품에 대한 개개인의 차이를 규명하는 것이다. Functional genomics는 유전자와 유전자산물의 기능연구로 유전자가 발굴되어야하는데 임의적으로 전체 cDNA를 분리 염기서열을 결정하거나 중요한 유전자의 발굴을 위하여 기능이 잘못된 동물로부터 관련유전자를 분리할 수 있다. 대부분의 사람들이 인간유전체의 전체염기서열을 얻음으로서 10만개로 추정되는 인간의 모든 유전자를 찾을 수 있을 것으로 예상하는데 실은 그렇지가 않다. 인간유전체는 다른 모델생명체의 유전체보다 훨씬 복잡하며 유전자로 추정되는 부분은 전체의 2% 정도에 지나지 않는데다가 거의 모든 유전자가 exon과 intron으로 이루어져있어 genomic DNA의 염기서열결과로부터 유전자를 발굴하는 것이 쉬운 일이 아니다. Exon trapping과 direct cDNA selection으로 유전자를 찾을 수 있고 혹은 컴퓨터 소프트웨어를 이용하여 유전체의 염기서열부터 유전자의 존재를 예측하는 것도 가능하나 어떤 경우이던지 완

벽하지 않고 설사 유전자를 찾았다하더라도 그 유전자에서 여러 개의 전사물이 만들어지는 경우 모두 발굴이 어렵다. 따라서 아무리 인간유전체의 전체염기서열을 가지고 있다고 하여도 발현되는 유전자를 찾는 cDNA 프로젝트의 도움 없이 모든 유전자 발굴은 어려운 것으로 예상된다. 이런 이유로 미국, 일본 등 선진국은 인체의 full-length cDNA 발굴 및 염기서열 결정에 주력하고 있다. 눈부신 분자생물학 기술 발달에도 불구하고 cDNA 프로젝트의 문제점은 시료 확보 및 방법에 모두 존재한다. 우선 인체 조직에서 발현되는 전사물을 얻은 후 역전사효소 (reverse transcriptase)를 이용하여 cDNA를 합성하기 때문에 구하기 어려운 조직이나 혹은 특정 조직에서 발생과정에 잠시 발현되는 유전자의 발굴은 어렵다. 게다가 cDNA 합성은 RNA 분리, 클로닝 벡터 및 고열에 안전한 역전사효소 등 여러 단계에 의존하므로 길이가 긴 전사물의 full-length cDNA 합성 효율이 낮다. 따라서 선진국에서는 현재 우선 크기가 작은 인체 cDNA 발굴 및 염기서열 결정에 주력하고 있다. 인체뿐 아니라 유전체 구조가 복잡한 동물의 유전자 발굴은 유전체 전체 염기서열에만 의존할 수 없고, 발현유전자 발굴을 목표로 하는 cDNA 프로젝트와 병행해 나아갈 것으로 예상된다. 무작위적인 cDNA 염기서열결정 유전자 발굴 방법은 유전자목록을 작성하는데는 좋으나 기능연구에 어려움이 따른다. 일부에서는 포유류 생물학의 중요한 유전자를 발굴 동정하기 위하여 쥐, 초파리, zebrafish에 화학적돌연변이 (chemical mutagenesis)와 gene trapping 기술을 이용하여 유전체 전체에 돌연변이를 (genome-wide saturation mutagenesis) 일으키고 이들 돌연변이 동물들 결합에 관련된 유전자를 위치적클로닝 (positional cloning)하기도 한다. 현재까지 이 방법으로 진행된 가장 좋은 예는 zebrafish의 ENU를 이용한 돌연변이 양산으로 zebrafish 연구자들에게 귀중한 자료를 제공하였다. 초파리와 쥐에도 이 방법으로 돌연변이를 양산하는 일이 시도되고 있다.

모든 유전자를 얻게 되면 그 다음은 각각의 기능을 연구 할 차례이다. 유전자의 기능연구는 곧 생물학 자체로 많은 노력과 시간을 요구하는데, 유전체의 기능적 연구 (functional genomics)란 한

유전자의 기능연구를 지칭하는 것이 아니라 신기술과 전략을 이용한 대규모의 유전자 기능 연구를 말한다. 여러 방법의 접근이 가능한데 크게 나누면 여러 상황에서 유전자의 발현양상을 DNA chip을 이용하여 대단위로 비교 분석하거나 생물체생물학로 인간보다 단순한 모델생명체를 이용하여 유전자 돌연변이 혹은 소실된 생명체를 만들어 그 표현형을 분석함으로써 생명체내에서의 유전자 기능을 연구하는 것이다. 전자는 생리적 유전체학 (physiological genomics)으로 조직이나 기관의 생리학적 상태의 기능 연구로 정상과 질병 상태에서의 유전자의 역할을 조사하는 것으로 예로 미국 정부의 CGAP (Cancer Genome Anatomy Project)과 BMAP (Brain Molecular Anatomy Project)을 들 수 있다. 이들 프로젝트의 두 가지 결정적인 공통 요인은 유전자발굴과 유전자 발현 양상 조사이다. 유전자발굴 (gene discovery)은 정상과 암 조직 (CGAP), 성인과 발달과정의 뇌의 여러 부위에서 (BMAP) 발현하는 모든 유전자의 목록을 만드는 것이 목표이다. cDNA library 구축, 3'과 5' EST sequencing과 EST clone들을 중복되지 않는 유일한 클론으로 clustering하여 주어진 조직에서 발현되는 Unigene의 목록을 만들 수 있을 것이다. 그 다음은 유전자발현 양상을 결정하고 발생, 전이 및 다른 병리 상태에서의 유전자 발현 윤곽 (profile)의 변화를 추적하는 것이다. Laser-Capture Microscopy, cDNA microarray, 대단위 염기서열결정 기술은 이들 프로젝트의 진행을 돕는데 모두 중요한 도구이며 기술이다. 이들 연구를 통하여 10만개에 달하는 인간유전자 각 각의 기능은 상세히 알 수 없어도 화학의 주기율표 같이 인간 유전자의 분류가 가능할 것으로 예상된다. 실제로 cDNA microarray 기술 발전으로 효모 뿐 아니라 인간의 암 진단에 이 방법을 적용하는 시도가 이미 있었고 현재 이 기술의 문제점이 해결되는대로 이들 분류는 질병의 예후를 예측하거나 치료 방법을 선택하는데 유익한 정보를 제공할 것이다.

2.2. 비교유전체학 (Comparative Genomics)

인간유전체 전체 염기서열이 결정되면 또 다른

중요한 과제는 비교유전체학 (comparative genomics) 으로 표준염기서열결과를 바탕으로 인종, 질병발병, 의약품에 대한 개개의 차이를 규명하는 것이다. 유전학의 최고의 목표는 표현형이 다른 것을 DNA상의 차이와 연결하는 것이다. 즉 사람마다 모습이 다른 것은 어떤 유전자 때문인가? 또한 각종 유전병들은 어떻게 발생하고 유전되며, 장수하는 집안과 단명하는 집안 사이의 차이점은 무엇인가? 현재까지 알려진 유전질환은 5000여종에 이르지만, 그 중에서 관련 유전자가 분명히 밝혀진 것은 15%에 지나지 않는다. 나머지 대부분은 유전성향은 의심되지만 관련 유전자가 여러 개이거나 아직 밝혀지지 않은 질환들이다. 어째서 생명과학이 눈부시게 발달하고 있는 요즘에도 아직 그렇게 모호한 부분이 많이 남아 있을까? 그것은 생물학의 중요한 기본 개념인 중복성과 다양성을 고려하면 쉽게 이해할 수 있다.

1997년 9월에 4.6 Mb에 이르는 대장균의 전체 유전체 염기서열이 밝혀졌다. 비교적 단순한 미생물인 대장균은 모두 4,288개의 유전자를 갖고 있다. 그중 38%는 그 기능이 전혀 알려지지 않았다. 유전자를 찾는 것보다 그 기능을 밝히기가 더 어려운 것이다. 어쨌든 적게는 4000여 개의 유전자로도 생명체를 유지할 수 있는데, 무엇 때문에 인간에겐 10만 개나 되는 유전자가 필요한 것일까? 물론 단세포생물인 박테리아에 비해서 다세포동물인 인간에게는 세포들 사이의 관계를 형성하고 조절하는데 많은 유전자들이 필요할 것이다. 그뿐이 아니다. 고등생물에는 한가지 기능을 수행하는 유전자가 여러 개 존재하는 경우가 많이 있다 (중복성). 또한 한 유전자가 여러 기능을 수행하기도 한다 (다양성). 그러면 한 유전자에 이상이 생겨도 그 기능을 다른 유전자들이 떠맡게 된다. 생명체는 자연의 보이지 않는 손으로 보호받고 있는 것이다. 그러므로 대부분의 질병들에 관련된 유전자들이 여러 가지라는 것도 놀랄 일이 아니며, 그래서 그 유전자들을 모두 발굴하는 것이 쉽지 않은 것이다.

현재까지 유전질환의 유전자를 찾는 방법은 두 가지로 기능적 클로닝 (functional cloning)과 위치적 클로닝 (positional cloning)이 있는데 전자는 질환유전자의 기능을 예측하고 유전자를 발굴하는

것이고 후자는 질환유전자에 관한 정보가 전혀 없이 전 염색체에 존재하고 멘델유전하는 개인 구별이 가능할 정도로 다양성을 가진 지표 (polymorphic marker)를 이용하여 관련유전자의 대개 위치를 찾고 그 부위에 존재하는 여러 유전자들 각각 분리하고 돌연변이 여부를 분석 조사하여 질환유전자를 찾는 것이다. 실제로 현재까지 발굴된 유전질환 유전자는 대개 위치적 클로닝 방법을 이용하였고 따라서 유전자가 발굴되었다라도 그 기능에 관한 정보가 없는 것이 단점이었다. 위치적 클로닝에 사용된 지표 (marker)를 유전적 지표 (genetic marker)라 하고 이들의 위치를 보여주는 지도를 유전적 지도 (genetic map)하는데, 현재 가장 최근에 발표된 지도 (Dib et al., 1996)는 5,264개의 microsatellite 지표를 이용한 것으로 평균 간격은 1.6 cM으로 상당히 자세한 것이었다. 그러나 microsatellite 지표는 특정부위에 존재하는 두 개염기의 반복수의 다양함으로 개인을 구별하는 것인데 그 빈도 및 안정성에 대한 정보가 부족할 뿐 아니라 시료 DNA를 각 지표로 PCR하고 그 산물의 크기를 비교 분석하여 개인을 구별한다. 최근에 DNA chip 기술이 개발되자 이를 이용하여 돌연변이 genotyping과 다양성 발굴이 가능한데, 유전자 발현을 조사하는 cDNA microarray와는 달리 genotyping chip이라 한다. 원리는 cDNA microarray와 유사하게 핵산의 hybridization에 의존하는데 다양성 발굴에 이용되는 genotyping chip은 1.28 cm² 단위 면적에 25-mer 크기의 oligonucleotide를 고밀도 (400,000개)로 놓고 분석할 유전체의 DNA로 hybridization하여 염기서열변이를 발굴한다. 고등동물 유전체의 복잡성 때문에 cDNA와는 hybridization 역학이 달라 유전체 DNA를 그대로 이용하려면 기술의 발전이 요구된다 (Lander, E., 1999). 더욱이 반복되는 두 염기의 수가 다른 microsatellite marker를 chip에 얹어 분석할 수 없다. 따라서 DNA chip이 개발된 이후 단일염기다형성에 관심이 모아지고 있다. 염기서열 변이는 모든 유전체에 존재하는데 인간유전체에서 가장 흔히 발견되는 다형성 (polymorphism)은 단일염기다형성으로 single nucleotide polymorphisms (SNPs)이라고 불리운다. SNP는 1,000 b마다 하나씩 발견되는데 그 빈도가 높고, 안정하며 유전체 전체에

끌고루 잘 분포되어 있어 자동화로 대단위 발굴이 가능하다. SNP는 발현되는 유전자뿐 아니라 유전자가 아닌 부위에도 존재하는데, 유전자의 SNP를 cSNP (coding SNP)이라 한다. 유전자의 cSNP를 기능과 관련지을 수 있다면 훨씬 더 유용할 것이지만 유전자에 따라 SNP의 수가 몇에서 몇백 개까지 존재하기도 하고, 변이에 따라 단백질 기능에 영향을 미치는 것과 아닌 것이 있는데 그 구별이 어려워, 인간유전자 약 200개 cSNP 발굴 논문에 의하면 cSNP 이용하는 일이 만만치 않을 것으로 밝혀지고 있다 (Cargill et al., 1999; Halushka et al., 1999). 또한 유전자는 진화적으로 selection pressure를 견딘 만큼 cSNP 발굴보다는 전체유전체에 일정간격으로 존재하는 SNP를 발굴하여 촘촘한 지도를 작성하면 이는 인간유전체에서 개개인의 차이 (variation)을 발견하기 위한 유용한 단일염기차원의 유전적지표로 genome-wide association study에 이용되어 복합 형질인 암, 당뇨, 정신질환 등 질병 관련 유전자 발굴에 사용될 뿐 아니라 학제간연구를 통하여 유전적변이와 의약품에 대한 개개인의 차이를 예견하는데도 사용될 것으로 예상된다. 이 분야의 선두주자는 미국 MIT의 Whitehead 연구소의 Eric Lander박사 팀으로 이미 인간유전체 전체에 3,000여 개의 SNP를 발굴하고 2,000여 개의 SNP 지표로 이루어진 유전적지도도 작성 발표하였다 (Wang et al., 1998). 일찍이 DNA 염기서열 변이에 관한 정보가 질병 분석, 진단, 치료 및 예방 등 다양한 분야에 이용될 수 있는 잠재력을 인정한 미국립보건원은 '98년부터 3년간 100,000개의 SNP발굴을 목표로 3,000만 달러를 투자하였고 '99년 4월에는 전세계 우수 제약회사 10개가 SNP Consortium (TSC)을 구성하여 3,000만 달러를 투자하여 다음 2년간 300,000개의 SNP 데이터베이스 구축을 시도하고 있다. '99년 7월에는 일본정부도 다음 2년간 150,000개의 SNP 발굴을 위하여 5,000만 달러 투자하겠다는 계획을 발표하였다.

인간유전체 염기서열결정의 완성을 앞두고 그

결과가 가져올 파급효과를 몇 가지로 나누어 예측해보고자 한다. 첫째, 유전자에 이상이 생겨서 발생하며 자손에게 유전되는 것이 유전질환들의 경우, 유전자검색을 통하여 이상 유전자의 보유 여부를 미리 알아서 자손에게 나타날 수 있는 유전병의 발생빈도를 줄일 수 있게 될 것이다. 또한 태아와 신생아의 유전자검색을 통하여 초기에 유전자 이상 여부를 조사하여 가능한 한 유전병을 예방하고 장래에 일어날 상황에 미리 대비할 수 있다. 뿐만 아니라 유전 성향이 있는 것으로 보이는 질환들도 그 보인자를 검색하여 발생 빈도를 줄일 수 있다. 현재까지 알려진 5,000여 가지의 유전질환 중 검색 가능한 유전질환의 수는 아직 100가지를 넘지 않지만, 인간유전체 프로젝트가 완료되면 검색 가능한 질환의 숫자는 기하급수적으로 늘어날 것으로 기대된다. 같은 기능을 수행하는 유전자들이 다수인 경우 (유전자의 중복성)가 종종 있는데, 그럴 때에는 관련 유전자만으로는 질병의 발병시기와 증상을 예측하는 것이 불가능하다. 이럴 때 모든 유전자의 성향을 살펴보는 전체유전체 분석이 가능할 것이다. 둘째로 프로젝트가 완료되면 인간의 표준염기서열 결과를 바탕으로 인종, 질병발병, 의약품에 대한 개개인의 차이를 규명하는 것이 가능해 질 것이다. 현재까지의 서양 의학은 사람은 다 같다는 전제하에 특정 증상에 대하여 일률적으로 처방하였으나 앞으로는 SNP 발굴을 바탕으로 인간의 분류가 가능해 질 것이고 이는 치료에 영향을 미치고 약리유전체학 (pharmacogenomics)가 발달하면서 개인별 의학시대가 열릴 것이다. 셋째로 21세기는 유전체 서열연구에 이어 본격적으로 유전체 기능연구가 활발해져서 의료 및 산업분야 응용에 기여할 것이다. 인간유전체 프로젝트가 완료되면 신기술과 새로운 정보를 접목한 대단위 유전자검색을 가능하게 되어 머지않아 사람들마다 각각 유전자 신분증을 소지하며, 그것을 바탕으로 배우자도 앞으로 태어날 아이가 유전병의 영향을 받지 않도록 선택할 수 있게 될 것이다.