

SL101 Human Genome Project & Post-Genome era

Yong Sung Kim
Genome Research Center, KRIBB

1. Genome Project

1) 생명의 비밀 열쇠: genome

현재 지구상에는 단세포 생물인 원핵생물로부터 수십조의 세포로 구성된 인간에 이르기까지 무수한 생물종이 서식하고 있다. 생물종들이 제각기의 생활방식을 유지하고 세대에 따라 생명을 영속하는 비밀은 어디에서 기인할까?

비밀의 열쇠는 각 생물체를 구성하고 있는 기본단위인 세포의 염색체에 있다. 원핵생물인 미생물로부터 고등한 동식물에 이르기까지 모든 생물체의 세포 속에는 염색체가 존재하며 이 염색체에 담겨 있는 유전정보에 의해 각각의 생물체는 생명현상을 발휘하게 된다. 염색체는 이중 나선형의 DNA(Deoxyribonucleotide acids)라는 생체고분자로 이루어져 있는데, 유전정보는 DNA를 구성하는 아데닌(A), 구아닌(G), 티민(T), 시토신(C)이라는 네가지 염기의 조합으로 구성된다. 유전자 발현과정에서 염기는 3개씩 짝을 이루어 아미노산을 하나씩 불러온다. 그러므로 염기서열에 따라 연결되는 아미노산의 종류도 달라진다. 이렇게 연결된 아미노산은 생명체의 모든 활동을 담당하는 단백질을 만든다. 결국 DNA의 유전정보는 단백질을 합성하는 암호문이라 할 수 있다. 이 암호문은 염기라는 글자로 이루어져 있고, 글자들은 아미노산이라는 단어를 만들어 마침내 단백질이라는 문장을 이루는 것이다. 결국 이렇게 만들어진 다양한 단백질의 기능에 의해 다양한 생명현상이 발휘하게 되는 것이다.

염색체의 염기배열은 하나의 단위체로서 자손에게 전해지게 된다. 이때 자손에게 전해지는 한 세트 염색체로부터의 완전한 전체 염기배열을 유전체 (게놈, genome)이라 정의한다. 예컨대 인간의 23쌍의 염색체로부터의 모든 염기배열을 일렬로 나열하면 약 32억쌍의 염기수가 되며 이 32억쌍의 염기를 인간유전체라 한다.

2) Human Genome Project의 배경 및 필요성

인류가 지구상에 등장하고 현재에 이르기까지 인류는 원인도 모른 채 많은 유전병에 의해 시달려 왔다. 오늘날 유전병의 수는 5,000여개로 추정되고 있으나 그 원인을 알고 있는 종류는 극소수에 불과하다. 유전학이 발전하면서 많은 학자들은 유전병의 원인이 염색체상 유전자이상으로부터 비롯된다는 사실을 깨닫고 원인이유전자를 동정하는 한편 염색체상의 위치를 파악하려는 연구에 노력을 기울여 왔으나, 하나의 질환에 대해 수십년의 시간이 필요하다는 사실을 깨닫게 되었다. 예컨대, 알츠하이머병(일명 치매)은 극도의 정신장애를 일으키는 중추신경계의 퇴행성 질환으로 가족을 단위로 중년기에 발병하여 인간을 황폐화시켜 결국 죽음에 이르게 하는 질병이다. 1930년대에 이병이 알려지게 되었으며 많은 노력의 결과 1980년대에 원인이유전자에 대한 대략적인 염색체상 위치를 파악하게 되었으나 원인이유전자를 분리하는 작업은 불가능한 것으로 생각하였다.

이때 유전학자들은 이 병의 성공적인 연구를 위해서는 인간 염색체상에 꼬리표유전자(STS marker)를 모아 자손으로 전해지는 염색체영역을 추적할 수 있게 하는 '유전체지도'의 필요성을 절실히 인식하게 되었으며, 유전체지도를 이용하면 알츠하이머병 뿐만 아니라 많은 유전병의 원인이유전자를 분리하는데 혁명적인 발전을 가져올 것으로 기대하였다. 실제로 이 같은 인식은 많은 학자들에 의해 제안이 되고 결국 인간유전체지도작성 및 염기서열의 완전해독을 골자로 하는 '인간유전체연구' (Human Genome Project)가 출범하게 된 것이다.

그 결과, 1996년 인간염색체 1번과 14번에 위치하고 있는 알츠하이머병의 원인이유전자가 분리되었으며, 이들 유전자로부터 만들어진 단백질은 신경세포벽을 구성하는 성분인 것으로 밝혀졌다. 따라서 이들 유전자의 이상이 신경세포벽 단백질의 이상을 초래하고 그럼으로써 신경세포의 기능이상을 야기하는 것이다. 이 유전체정보는 곧 가족력이 있는 가계에서 태아의 알츠하이머병 산전진단에 사용되고 있으며, 또한 이미 태어난 아이에게는 치료제 개발을 위한 정보로 활용되고 있다.

2. Human Genome Project 현황

1) HGP 출범

1985년 미국의 시세이머가 최초로 인간유전체지도, 유전자연관지도, 대규모염기서열분석기술개발의 중요성을 역설하였으며, 이어 1986년 미국 에너지성 및 국립보건원에서 인간유전체의 염기분석 필요성을 인식한 바 있으며, 1987년 미국의 길버트는 '게놈사' 기업을 설립하고 유전체지도작성과 체계적인 염기분석을 수행하여 데이터베이스를 구축하기 시작하였다. 이러한 움직임은 결국 1989년 미국 에너지성 및 국립보건원을 중심으로 공동으로 투자계획을 발표하고 이듬해인 1990년 인간유전체연구 5개년 계획을 수립하고 다음의 내용을 주요 골자로 하는 계획이 출범하기에 이른 것이다 : 인간유전체지도 작성, 유전자연관지도 작성, DNA 염기서열결정기술 개발, 모델생물 유전체연구 동시수행, 생물정보과학 개발, 윤리, 사회적 문제 대비책 마련, 연구, 훈련 계획의 수립, 관련기술의 개발 등.

1990년 당시 1차 5개년 계획은 출범당시 상당한 기술을 개발하는 전제하에서 수립된 것이었으나 실제로 많은 목표가 예상보다 일찍 결과를 얻게 되었다. 따라서 미국의 에너지성(DOE)과 국립보건원(NIH)은 1993년 새로운 5개년계획(1994-1998)을 수립하였다. 이어서 1997년에는 약 100 kb 간격의 유전체지도를 작성하기 위해 필요한 약 30000개의 코리표DNA 중 27000개가 개발되어 인간유전체지도작성에 사용되었으며, 이에 따라 인간염색체 24종 염색체 중 20종에 대한 물리적 지도작성이 완료되었다. 또한 약 60만개의 cDNA에 대한 부분적 염기서열이 결정되었으며 종류수로는 약 60%에 달하는 유전자가 분리하게 되었다. 이러한 결과는 염기서열결정 기술의 발달과 함께 다시 2003년으로 최종목표를 다시 수정하기에 이르렀다. 실제로 기술적으로 목표달성이 가장 어려웠던 부분은 2003년까지 인간유전체의 전체염기서열을 완성하는 것이다. 이를 위해서는 염기서열결정의 자동화, 고속화, 저렴화 등을 중심으로 한 기술노력이 추진되어 왔다. 그 결과 1990년대 하나의 염기의 서열결정에 드는 비용이 5불에서

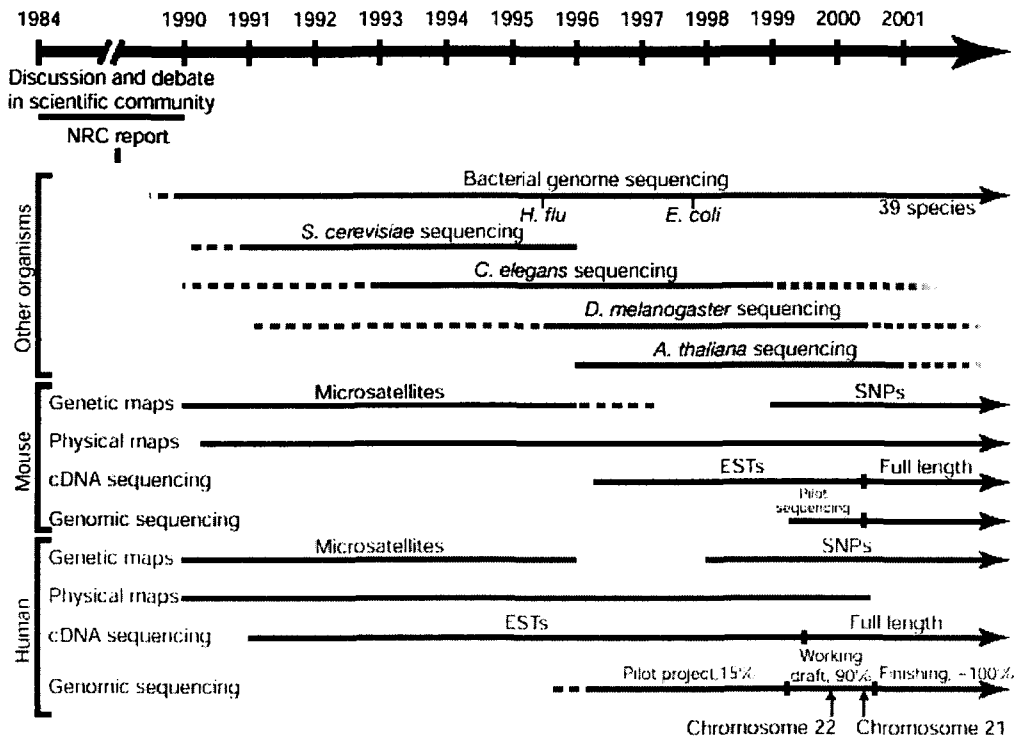


그림 1. 1990년에서부터 대규모 유전체 해독연구의 연도별 진척도

1997년에는 0.5 볼로 낮아졌으며, 2000년에는 0.1볼로 낮아졌다. 염기서열결정 속도도 모세관을 이용한 자동염기서열분석기의 개발로 1990년도에 비해 수백배 빨라지게 된 것이다.

2) HGP 진척도

개발된 이들 기술은 인간 뿐만 아니라 모델 생물의 유전체분석에 활용되어 2001년 3월 현재 50종의 미생물종 (유전체 크기 0.6~4.5 Mb)에 대한 유전체가 완전 해독되었으며, 13 Mb의 효모, 100 Mb의 선충, 130 Mb의 초파리 유전체 등이 완전 해독되었다. 최근 생쥐의 유전체 해독도 완료되었다.

인간유전체의 경우 1999년 12월에 22번 염색체의 염기서열이 완료된 이후 마침내 2001년 2월 약 94%에 이르는 인간유전체 초안이 발표되었다. 위와 같은 성공적인 성과는 미국내 17개 유전체전담센터 등 18개국 350개 연구기관, 1,000여명의 연구진이 참여하여 진행되었으며 각 국가별 염색체를 분담하여 국제 협력연구를 수행함으로써 가능하게 된 것이다.

이렇게 성공적으로 유전체연구가 성공하게 된 배경은 유전체프로그램 창안자들이 일종의 개혁운동으로서 추진되었으며, 분자생물학이 21세기 생물공학의 중심이라는 공감대가 형성되었으며, 결과에 대한 경제적 타당성 인식 및 유전체연구가 모든 분자생물학의 기초토대를 포괄하고 있다는 판단등에 근거를 두고 있다. 또한 정부차원의 대규모 투자계획과 함께 민간 주도의 투자도 유전체연구가 성공적으로 출범하게 된 원동력이 되었던 것이다.

3) 인간유전체연구의 성과

2001년 2월에 발표된 인간유전체의 해독결과는 여러 가지 측면에서 흥미롭다. 지금까지 해독된 것 중에서 가장 큰 유전체로서, 규모로서는 인간을 제외하고 지금까지 해독한 규모를 전부 합한 양의 약 8배에 해당된다. 또한 첫 번째 완성된 척추동물의 유전체라는 것과 무엇보다도 중요한 점은 바로 우리 자신의 유전체 염기서열이라는 것이다.

2월에 인간 유전체 염기서열이 공개되면서 지금까지 밝혀진 사실을 요약하면, 1) 인간유전체 염기서열에서 유전자(genes)를 포함하여 전이성 인자(transposable elements), CpG

islands 등 유전체 구성요소의 분포 및 염기서열에서 개체사이에서 뚜렷한 변이를 발견할 수 있다는 것이다. 이러한 변이는 유전자의 기능을 규명하는데 매우 유용하게 사용될 수 있다. 2) 인간 유전체로부터 약 4만개의 단백질 생산할 수 있는 유전자를 확인할 수 있다. 이러한 수치는 선충이나 초파리의 전체 유전자에 비해 2배에 지나지 않는다는 사실에 의아스럽기도 하지만, 한편으로는 인간유전체로부터의 유전자들이 선충이나 초파리의 유전체에 비해 보다 복잡한 기작에 의해 4만개 유전자로부터 배 이상의 단백질이 생산할 수 있을 것으로 판단하고 있다. 3) 무척추동물의 유전체와 비교하여 인간 유전체가 생산하는 전체 단백질 (프로테오姆, proteome)이 훨씬 복잡한 것으로 분석되었다. 이 사실은 척추동물만이 특이적으로 보유하게 된 단백질의 특수영역 (domain 및 motif)에서 비롯되었다고 볼 수 있으며 이것은 무척추동물 단백질의 구성요소들이 척추동물로의 진화과정을 통해 훨씬 풍부한 특수영역으로 배열된 것으로 해석되고 있다. 4) 인간 유전체에서 확인된 유전자 중에서 수백개 정도는 척추동물의 진화과정 중 어느 시점에서 미생물로부터 수평적 전달(horizontal gene transfer)에 의해 유래된 것으로 보이며 이중 수십개는 전이성 인자에 의해 도입된 것으로 밝혀지고 있다. 5) 인간 유전체의 절반정도는 전이성 인자에 증폭된 것으로 생각되지만 그러나 유인류의 진화과정에서 이들 전이성 인자들의 전이성이 현저하게 상실되어 가는 경향이 보인다. 예컨대, 한 종류의 전이성 인자인 DNA transposons의 전이성이 인간 유전체에서 완전히 소실된 것으로 보이며, 또 다른 종류의 전이성 인자인 LTR(long terminal repeat, retroposon)도 전이성이 점차 상실되어 온 것으로 보인다. 6) 인간 염색체의 중심체 및 말단체 부위는 비교적 근자에 유전체의 부분 절편이 반복적으로 복제되면서 채워진 것으로 보인다. 이러한 판단은 이러한 효모나 초파리, 그리고 선충의 유전체에 비해 훨씬 빈번하게 발견되고 있기 때문이다. 7) 감수분열 과정에서 돌연변이율이 남자의 경우가 여자에 비해 두배 가량 높은 것으로 보인다. 8) 인간 유전체에서 대부분 GC%가 낮은 부위는 염색체 핵형의 G-밴드와 일치한다. 9) 일반적으로 염색체의 말단부위(약 20 Mb 부근)와 작은 염색체 종류들에서 재조합율이 훨씬 높은 발생하는 경향이

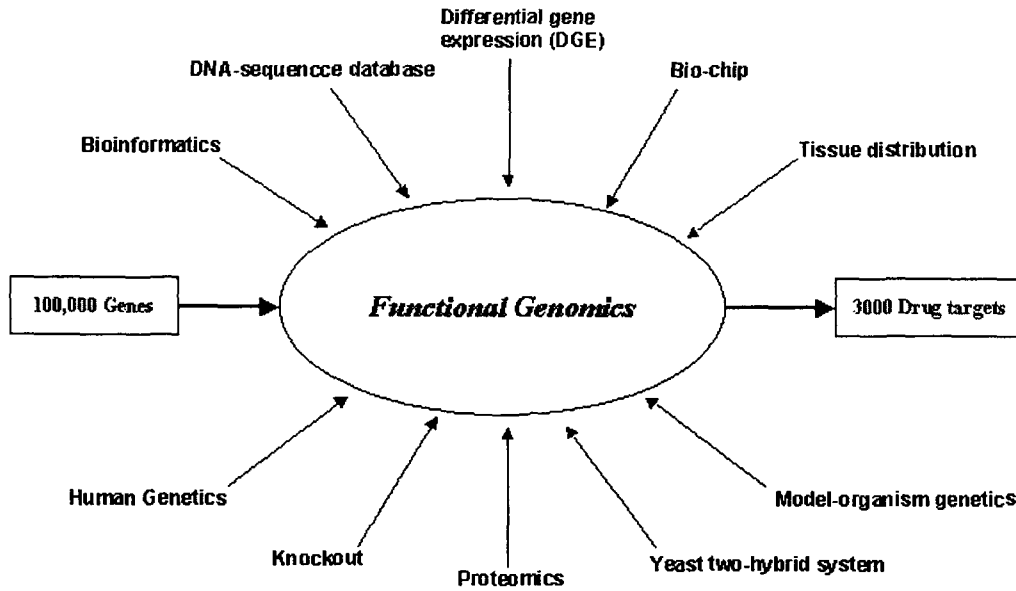


그림 2. 인체유전체기능연구의 목표

있다. 이러한 현상은 감수분열시 염색체마다 적어도 한번의 교차의 기회를 부여하는 것으로 보인다. 10) 현재까지 2백 만개 이상의 단일염기변이 (SNPs, single nucleotide polymorphisms)가 발굴되었는데, 이중 1% 이하만이 실제로 단백질 구조에 영향을 미치는 것으로 보이며 이러한 변이정보는 인류집단에서 많은 유전자들의 대규모 연관분석에 활용될 것으로 전망된다.

3. Post-Genome era

인간유전체가 해독되고 많은 모델생물의 유전체가 완성된 지금, 각 생물유전체에서 유전자에 대한 기능을 규명하는 유전체기능연구시대(Functional Genomics)로 전환되고 있다. 우리는 이를 지금까지의 유전체구조해독시대(Structural Genomics)와 구분하며 한편으로는 포스트유전체시대(Post Genome era)라고 부르기도 한다. 인간유전체으로부터 각 유전자들의 생체기능을 밝히는 일이 본격적으로 추진될 때 누가 먼저 기능을 밝혀 이를 산업적으로 이용하는냐에 따라 지적재산권뿐만 아니라 산업적

재산권까지도 획득하게 되는 시대가 된 것이다. 또한 21세기는 과학기술 정보화와 전 세계적인 정보 네트워크의 구축이 중심이 되므로 얻어진 유전체 정보는 곧 경제적 가치를 창출하게 된다. 유전자 진단등에 활용을 통한 각종 암의 조기 예방과 각종 난치병 치료제의 개발 그리고 각 개체별 특성을 고려한 최적의 치료제 개발 등에 활용 될 수 있어 의료복지의 증진을 도모할 수 있는 무한한 공익적 가치를 지니고 있을 뿐만 아니라 최신기술을 종합적으로 활용하여 확고한 신약 개발 프로세스로서 감히 그 가치를 산정할 수 없는 막대한 산업적 잠재 가치를 갖고 있는 것이다.

따라서 유전체연구는 생명공학산업의 지적재산권 확보를 통한 장기간의 독점적 지위 유지가 가능하며, 또한 미래의 시장 지배력이 매우 클 뿐만 아니라 의료이외에도 농업 및 환경분야로의 기술적 파급 효과가 막대할 것으로 평가하고 있기 때문에 21세기 유망 성장산업의 하나로 지목하고 있는 것이다. 또한 생명공학분야의 중요성과 잠재력을 일찍 간파한 대다수 선진국의 다국적 화학 기업들과 거대 제약 기업들은 21세기 화학 및 생명 과학 산업의 주도

권 확보와 기업 생존을 위하여 연구개발 구조 재편 및 사업 구조 전환에 박차를 가하고 있으며, 상업적 가치를 지닌 유전자 및 단백질의 선점을 위하여 천문학적인 자금을 투입하여 과감한 연구개발을 진행하고 있는 상황이다.

SL102 Comparative Proteome Analysis Using Two-Dimensional Electrophoresis

Yung Joon Yoo

Department of Life Science, Kwangju Institute of Science and Technology(K-JIST), Kwangju 500-712, yjyoo@kjist.ac.kr

Proteome research (proteomics) is gaining more and more attention as a core of functional genomics because the ultimate functional molecules in living organisms are proteins. Proteome is originally defined as all proteins expressed by a genome, which is continuously changed in an organism in space and time. In order to analyze this dynamic state of proteomes, protein array or protein map is prerequisite and two-dimensional polyacrylamide gel electrophoresis (2D PAGE) is currently only method to separate thousands of proteins simultaneously. Although 2D-PAGE was introduced more than two decades ago, the reproducibility and resolving power are recently improved mainly by immobilized pH gradient (IPG) in isoelectric focusing (IEF). Using homemade IPG strips with various ranges of pH gradient and mini gels for SDS-PAGE, we demonstrate that the production of reference protein map is rather simple and rapid.

Using two-dimensional electrophoresis and MALDI-TOF mass spectrometry, we performed the comparative analyses of proteomes from cultured cells, patient tissues, model animals, cellular organelles, or molecular complexes. I will introduce some

experiences of our proteome studies to illustrate the scope of proteomics technology.

In first case, we found that the expression of peroxiredoxin 1 was increased in malignant lung epithelial cells (A549) in relation to normal ones (BEAS-2B) through the comparative analysis between two cells. This augmented expression of Prx-I was confirmed in tissues from lung cancer patients. In second case, it was found that cellular retinoic acid binding protein II (CRBP-II) was down-regulated with differentiation by the proteome analysis during chondrogenesis. It should be closely related with the effect of retinoic acid on chondrogenesis.

This work was supported by the grant from Life Phenomena and Function Research Group Program-2001 from the Ministry of Science & Technology, and the BK 21 Program.

SL103 Structural Genomics/ Proteomics as the Next Frontier in Biotechnology and an Example of Obtaining Insights into Biochemical Function from Three-dimensional Structure

Se Won Suh

National Research Laboratory on Structural Proteomics, School of Chemistry and Molecular Engineering, Seoul National University, Seoul 151-742, Korea
e-mail: sewonsuh@snu.ac.kr

Genome sequencing in the past decade has provided massive genetic information on many microorganisms and a smaller number of eukaryotic organisms. One of the next frontiers in the 21st century biotechnology is structural genomics/proteomics, the protein structure determination at the genome-wide scale. Its ultimate goal is to provide all the