

차세대 게놈 시퀀서의 등장과 Personal Sequencing 시대의 도래

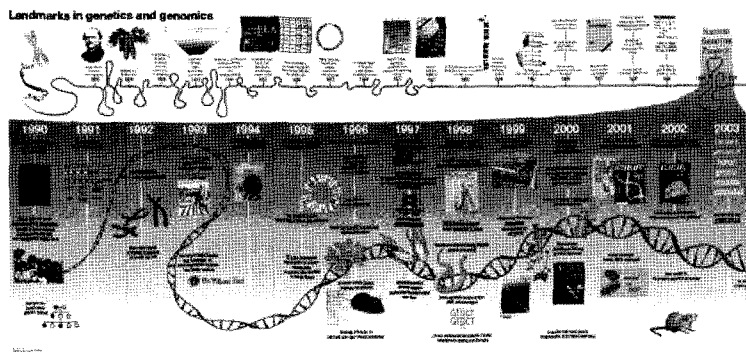
책임연구원 박근준(질병관리본부 국립보건연구원 유전체센터 바이오과학정보팀)

1990년 시작되었던 휴먼게놈프로젝트 (Human Genome Project, HGP)는 인간의 30억 염기쌍인 게놈 서열을 전부 읽어보자는 장대한 프로젝트였다. 처음에는 그 실현 가능성이 무모하게 여겨지기까지 했지만, 점차로 그 가치가 '생물학의 이플로 계획'으로 불릴 정도로 중요하게 인식되면서 추진되었고, 마지막 결과물만이 아닌 프로젝트 진행 과정에서 게놈 시퀀서를 중심으로 다양한 기술 발전이 이루어지는 과학적인 성과도 얻을 수 있었다. 그 결과 예상보다 빨리 2003년에 프로젝트의 완성이 선언되었고 전 세계 과학자들은 인간의 완전한 염기서열 정보를 함께 사용하여 생물정보학 연구를 진행하고 있다. 그리고 미래의 생명과학에서는 정말로 한 사람 한 사람의 염기서열을 해독하고 그것을 바탕으로 각 개인의 건강을 향상시킬 수 있는 정보를 얻으려는 생각까지 그려졌는데, 최근 예상보다 빨리 그러한 시대의 도래를 실현할지도 모를 움직임들이 나타났으므로 그에 대한 이야기를 적어보고자 한다.

개인의 게놈정보를 사용하게 된다는 미래의 생명과학 실현에서 가장 큰 장벽의 하나는 30억 염기쌍을 읽어내는 데 필요한 비용과 시간으로, 처음에 소개한 휴먼게놈프로젝트 (HGP)는 1990년부터 시작해 세계 석학들의 국제적인 프로젝트로 추진해서, 27억 달러에 달했다고 알려진 엄청난 비용을 투입하여 2003년에야 겨우 한 사람분의 게놈을 읽어

냈다. 하지만 현재 과학자들은 가까운 미래에 \$1,000 정도로 한 사람의 게놈 해독이 가능해질 것으로 예상하고 그에 대한 대비를 논하고 있다[1]. 그 이유는 새로운 기술발전으로 게놈 시퀀서의 성능이 비약적으로 발전하고 있기 때문이며, 실제로 최근에는 차세대 게놈 시퀀서라고 불리는 상품들이 계속해서 등장했다. 지난 2008년 1월 한국 유전체학회 동계심포지엄에서도 바로 이러한 차세대 게놈 시퀀서들이 소개되어 모두의 주목을 받았다. 현재 상용화된 차세대 시퀀서들은 2007년 1월 Illumina에 합쳐진 Solexa의 Genome Analyzer, Roche (LifeSciences)의 FLX system, Applied Biosystems의 SOLiD로 압축되는데, 한국의 유전체학회 동계심포지엄보다 앞서 열린 2007년 11월 도쿄에서의 제7회 국제게놈회의에서도 이 세 가지 차세대 시퀀서들의 관련된 성과들이 소개되면서 세계적인 화제를 모았었다.

Illumina Genome Analyzer (Solexa)의 경우는 1 run 당 해독량 1.5Gbp (2.5일)이며, 분석되는 게놈의 단편 길이는 35bp로 짧은 편에 속한다. 1 run 당 데이터는 수백 GB~1TB에 이른다. 다른 시퀀서들과 비교하면 가동 비용이 낮은 편에 속한다고 알려졌다. 생가센터나 북경게놈연구소 등을 비롯한 세계 각국의 연구기관들이 대량으로 보유하고 있으며, 현재 분석 길이를 75bp로 늘리고 1 run으로 15Gbp가 해독 가능한 제3세대장치를 개발 중이라고 한다.



[그림 1] 휴먼게놈프로젝트 중심의 연대표 (네이처, 2003년)



Roche의 GS-FLX 시스템은 1 run 당 해독량 100Mbp (7.5시간)이며, 특징은 게놈의 단편 길이가 평균 250bp로 상대적으로 길며, 현재 100대 이상의 활약을 통해 가장 많은 실적을 쌓았다는 점을 들 수 있다. 다른 기종에 비해 긴 단편을 읽을 수 있다는 것은 나중에 어셈블링이 훨씬 편하다는 걸 의미한다. 얼마 전에 화제가 되었던 DNA 이중나선 발견자인 왓슨(Watson)박사의 게놈을 해독한 것이 바로 Roche의 시퀀서였다.



[그림 2] 454 LifeSciences (Roche)의 설립자인 Jonathan Rothberg가 왓슨에게 그의 게놈정보를 건네는 모습

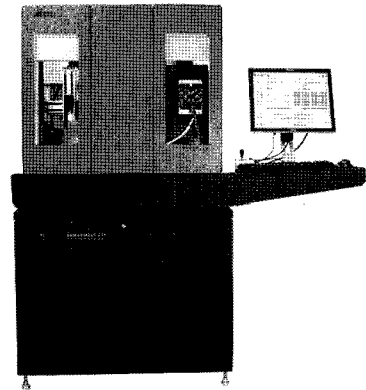
왓슨박사의 게놈해독에는 67일 동안 100만 달러가 들었고 알려지지만[2], 제품 선전을 위한 것이었으므로 아주 최소한의 필요경비였을 테고, 실제로는 장비 도입 외에 연구원 훈련 과정 등도 고려하면 적어도 1년 이상의 기간과 10배 가까운 비용이 들었을 것으로도 여겨진다. 왓슨에 앞서 개인 게놈 시퀀싱이 되었던 J. Craig Venter (head of Celera Genomics)의 경우는 HGP에서 사용했던 기술을 이용하여 약 1억 달러가 들었다고 한다[3].

따라서 왓슨의 게놈 해독 비용 100만 달러라는 것은 HGP에서 사용되었던 기술에서 필요로 한 1억 달러의 1%에 불과한 셈이다.

벤처박사의 게놈 시퀀싱 뉴스는 당시에 그다지 큰 화제가 되지 못했는데, 지금의 왓슨박사의 게놈 시퀀싱 뉴스가 화제가 되는 이유의 하나는 새로운 기술이 도입된 차세대

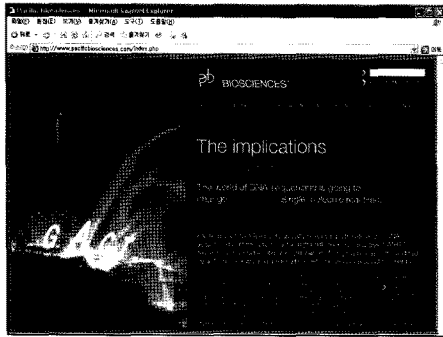
시퀀서 사용으로 이렇게 극적인 비용 변화가 있었기 때문이다.

Applied Biosystems의 SOLiD (Sequencing by Oligonucleotide Ligation and Detection)는 같은 염기를 2번씩 읽는 방법으로 1 run 당 해독량이 4Gbp (192시간)이라고 하는데, 1 run 당 해독량이 4Gbp 라는 것은 이론상으로 따진다면 1 run으로 인간 게놈을 다 읽어낼 수 있다는 뜻이 된다. 1 run 당 나오게 되는 데이터양은 수 TB에 이른다. SOLiD에 관련된 소개 글은 Nature Japan에서도 확인할 수 있다[4].



[그림 3] 차세대 SOLiD 시스템 (네이처, 2007년 10월 4일자 소개)

이미 상품화가 이루어져 여러 가지 결과물들을 내기 시작한 위의 세 가지 차세대 게놈 시퀀서들 외에도 다른 새로운 기술이 적용된 다양한 차세대 게놈 시퀀서들에 관한 소식도 관심을 끌고 있다. 닛케이 바이오테크놀로지 2008년 2월 20일 뉴스에 따르면, 2008년 2월 19일에 일본 도쿄에서 열렸던 게놈 네트워크 심포지엄에서 미국 Pacific Biosciences가 2010년에 초고속 DNA 시퀀서를 제품화한다는 계획을 발표해 화제였다고 한다[5, 6]. 실현 여부는 좀 더 지켜봐야 할지도 모르겠지만, 인간의 30억 염기쌍 (이배체이므로 실제로는 60억 염기)을 읽는 데 필요한 시간이 겨우 4분이라는 경이적인 내용이었다. 1시간에 1,000억 염기쌍이라는 해독능력을 갖췄다고 알려졌으며 관련 논문이 이미 발표되었다. (Proc Natl Acad Sci USA, 2008 Jan 29; 105(4): 1176-81)



[그림 4] www.pacificbioscience.com

이러한 방대한 게놈 데이터가 생산되게 된다면 당연히 그 데이터를 저장하고 처리할 수 있는 데이터베이스 기술의 발전도 필요해지는데, 관련된 뉴스로 오라클이 위와 같은 움직임을 예상해 생명과학 분야에 본격적으로 뛰어든다는 의사를 표명한 2001년 뉴스가 있었다. 위에서도 소개되었던 셀레라의 벤터박사는 인간 게놈 데이터만으로 250테라바이트가 필요하며 여기에 신약개발에 필요한 프로테오믹 데이터가 추가되면 페타 (peta)바이트 급, 더 나아가 개인의 게놈 정보로부터 약을 만들려 한다면, 필요한 데이터가 엑사 (exa)바이트 급으로 팽창한다고 언급했으며, 오라클에서는 페타바이트 이상의 데이터베이스를 시야에 넣고 Oracle Database의 기능강화를 해 나간다고 했었다[7].

지금도 계속 새로운 뉴스들이 보고되고 있는데, 그 중의 하나인 2008년 3월의 뉴스에 따르면 개인 게놈 시퀀싱 비용이 위에서 언급한 왓슨박사의 게놈 시퀀싱에 필요했던 100만 달러에서 또다시 급격하게 떨어져서 벌써 \$60,000가 언급되기 시작했다[9]. 심지어 일부 기업들은 \$1,000 게놈시대가 아닌 \$100 게놈시대라는 도전적인 목표를 향해 달리기 시작했다. MIT의 Technology Review 2008년 4월 17일 뉴스는 두 개의 회사인 Complete Genomics와

BioNanomatrix의 \$100 게놈해석을 향한 야심 찬 도전과 현재 성과들을 소개하면서 새로운 화학과 진보된 나노 제조 기술에 의해 5년 정도 후에 실현될 가능성을 소개했다[10].

위와 같은 움직임들을 종합해 봤을 때, 아직 \$1,000 게놈 해석 단계는 아니지만 게놈 분야에서 대단히 극적인 상황이 본격적으로 시작되었다고 할 수 있다. 이제 데이터 생산은 본격적인 시동이 걸렸고, 거기에 맞는 데이터베이스와 분석용 슈퍼컴퓨터 서버 구축과 소프트웨어 개발 등이 뒤따르면서 앞으로 2~3년 후에는 생물정보학 중심의 커다란 변화가 몰려올 것으로 예상된다. 아직은 차세대 시퀀서들이 고가의 장비인데다가 (일본 기준으로 7,500만 엔~1억 2,500만 엔), 운영비용도 비싸므로 일반 연구실들에서 구입해 운영하기에는 어렵다는 문제점이 존재한다. 하지만 위에서 언급한 것처럼 다양한 새로운 기술들이 적용되면서 자꾸 새로운 개념의 빠르고 저렴한 게놈 시퀀서들이 개발되고 있으므로 비용과 관련된 문제점은 점차로 해결될 것으로 예상된다.

이러한 초고속 게놈 시퀀서를 누가 구입하고 있느냐에 관한 뉴스 하나를 소개하자면, 일본 오키나와켄(현)에서 2007년도 보정예산으로 차세대 게놈시퀀서 3대를 구입하여 초고속 DNA시퀀스 센터를 설립하기로 했었는데[8], 일본의 지방자치단체에서 차세대 DNA 시퀀서를 구입한 최초의 예로 보도됐다. 국가별 생명과학연구 역량 강화를 위한 게놈 시퀀서 구입 경쟁만이 아니라 지역의 생명과학연구 활성화를 노린 움직임도 나온 것으로 해석된다. 일본 전체로는 2008년 3월 시점에 10여 대가 가동 중인 것으로 알려졌다. 현장에서 들리는 한 가지 문제점은 실제 분석에 들어가면 상당한 비용이 들기 때문에 아주 면밀하게 연구 계획을 세우기 전에는 좀처럼 본 연구에 들어가기 어렵다는 점을 들 수 있다고 한다.



Thursday, April 17, 2008

The \$100 Genome

Forget the \$1,000 genome. Some companies are looking far past that goal to create a *really* inexpensive sequencing technology.

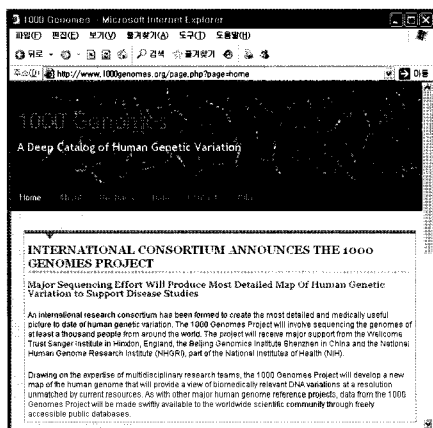
By [Emil Sinner](#)

[그림 5] \$100 Genome 해석을 전망한 기사

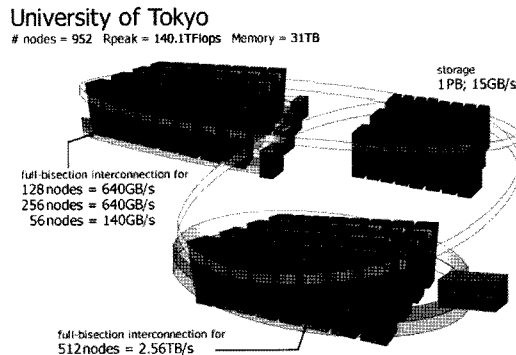
차세대 게놈 시퀀서의 공격적인 도입과 운영에 관련되어서는 최근 중국의 움직임이 주목을 모으고 있다. 예를 들어 모두가 잘 아는 영국의 생가센터는 현재 ABI 3730 45대, Roche 454 3대, Solexa 27대, SOLiD 5대를 보유하고 있으며 정보학 전문가만 100명 단위로 참가하고 있다고 한다.

그런데 중국의 북경게놈연구소 (BGI)에서도 최소한 Roche 454 3대, Solexa 17대, SOLiD 2대 (ABI 보유대수는 불확실)를 갖추고 있으며 100명 정도의 정보학 전문가들을 갖추고 있다고 한다. 이 연구소는 또한 최근 화제가 되었던 미국, 영국 중심의 1000 Genomes Project에 참가 증으로, 중국이 게놈 시퀀싱 연구분야에서는 일본보다 더욱 적극적인 자세로 돌입했음을 시사한다.

지금까지 게놈 시퀀서의 고속화와 비용의 저렴화를 다루었는데, 생물정보학 입장에서 따져본다면 사실은 또 하나의 근본적인 문제가 남는다. 이러한 초고속 차세대 게놈 시퀀서들이 쏟아내는 대용량 정보를 누가 무엇으로 분석하느냐는 것이다. 차세대 게놈 시퀀서만 많이 산다고 해서 모든 것이 저절로 해결되는 것이 아니라, 최신의 분석용 컴퓨터 서버와 그것을 다룰 수 있는 정보학 전문가 집단도 빨리 갖춰지지 않으면 개인 게놈 분석에 관련된 국제경쟁에서 낙오된다는 전망이 나올 수 있다는 사실이다. 2008년 6월 4일 Biotechnology Japan의 기사에서는, 6월 2일에 가동이 시작된 일본 최고성능의 T2K 오픈 슈퍼컴퓨터 (도쿄대학, 교토대학, 쓰쿠바대학 연계)마저도 시퀀서의 성능향상을 고려한다면 앞으로 1년을 건널 수 있겠느냐는 도쿄대학 모리시타 교수의 발언이 제목으로 나올 정도였다[12].



【그림 6】 1000 Genomes Project 웹사이트



【그림 7】 T2K Open Supercomputer (도쿄대학)

조사 시점이 약간 앞이었던 것으로 여겨지는데, 일본 과학 기술진흥기구 연구개발전략센터의 '생명과학분야 과학기술·연구개발의 국제비교 2008년 판'에서 Illumina로부터의 정보제공에 따른 Solexa의 국가별 도입상황을 비교하면 미국 60대 이상, 유럽 40대 이상, 중국 8대, 한국 2대, 일본 4대로 역시 중국이 적극적인 자세로 뛰어들었음을 짐작하게 한다. 일본에서는 이러한 인간 개인 게놈 서열분석의 실용적인 가치에 대해 회의적인 시각도 존재하는 듯하다. 각 개인에게 대응한다고는 하지만 이미 해독이 끝난 인간의 전체 게놈 서열을 다시 읽는 것에 앞으로의 연구 중점을 두기 어렵다는 의견이다[11].

이처럼 차세대 시퀀서들의 등장은 생물학에서 많은 양의 데이터 축적과 처리, 분석을 요구하기 때문에 일반 생물학 연구실에서도 이제는 컴퓨터 환경의 정비가 더욱더 중요한 위치를 차지하게 된다. 인간의 대규모 게놈 시퀀싱이 이렇게 가능해진다는 상황의 도래는 다른 동물이나 미생물의 소규모 게놈을 분석하는 것은 더더욱 쉬운 시대가 된다는 것을 의미하며 결국 모든 생물학 연구실에 영향을 끼치기 때문이다. 이러한 컴퓨터 환경이 정비되고 컴퓨터과학과 생물학을 이해할 수 있는 정보학 전문가들이 참여함으로써 실험연구와 컴퓨터 과학이 융합된 data-driven 연구나 시뮬레이션 등이 게놈 연구를 중심으로 한 생명과학 전반으

로 광범위하게 확립되어 나갈 것이다.

이제는 한국도 국가 차원에서 언제 어느 정도의 예산을 집중시켜 이러한 차세대 게놈 시퀀서 환경을 구축하고 관련 연구를 추진하는 것이 타당한지를 신속하게 결정해야 할 시기가 된 것으로 여겨진다. 게놈 역학, 질환 유전자 탐색 등의 연구를 활성화해 실제 개인 게놈 서열이 주어졌을 때 얻어낼 수 있는 정보를 충실하게 만들 필요도 있고, 국민의

건강과 복지에 가장 도움이 되는 형태로 연구전략을 세우기 위한 검토, 생명과학을 기반으로 학계와 산업계를 연결해 실질적인 산업발전과 국가경제에 이바지하는 방안, 신약개발과 미래 맞춤형의학 등에 관련된 지적재산권 문제 등에서 장래에 한국이 외국에 전부 의존하는 상황을 피하기 위한 전략 등이 진지하게 논의될 필요가 있을 것으로 여겨진다.

[참고 문헌]

- [1] \$1000 인간 게놈 시퀀싱이 가능하게 된다면?
<http://www.nature.com/ng/qoty/index.html>
- [2] 왓슨박사 게놈해석 뉴스
<http://www.asahi.com/science/update/0529/TKY200705290419.html>
- [3] Levy, S. et al. PLoS Biol. 5, e254 (2007)
- [4] Nature Japan의 SOLiD 소개 기사
http://www.natureasia.com/japan/advertising/advertorial/pdf/applied_bio.pdf
- [5] Pacific Biosciences의 초고속 시퀀서 2010년 상품화 뉴스
http://biotech.nikkeibp.co.jp/bionewsn/detail.jsp?id=20050654&newsid=SPC2008021953046&pg_nm=1&sail=0&new1=1&news1=1&icate=0&yunw=1
- [6] PacBio to Start Selling Next-Gen Sequencer To Early Users in 2010; Goal is 100 Gb/Hour
http://www.in-sequence.com/issues/2_7/webreprints/145257-1.html
- [7] 오라클의 생명과학 관련 데이터베이스 사업 강화 뉴스
http://www.oracle.co.jp/news_owa/NEWS/news.NEWS_PRINTER_VIEW?p_news_code=612
- [8] 일본 오키나와현 차세대 시퀀서 구입 뉴스 (2008년 2월 19일 보도)
http://biotech.nikkeibp.co.jp/bionewsn/detail.jsp?hit_cnt=895&newsid=SPC2008021953002
- [9] 인간 게놈해석 비용 60,000 달러 (2008년 3월 25일 보도)
<http://www.newscientist.com/article/mg19726482.900-genome-is-a-snip-at-60000.html>
- [10] The \$100 Genome (2008년 4월 17일 보도)
<http://www.technologyreview.com/Biotech/20640/>
- [11] 시퀀싱의 새로운 시대 (Nature Japan 2007년)
http://www.natureasia.com/japan/tokushu/pdf/feature_69_20071220.pdf
- [12] 6월 2일에 일본 최고성능의 슈퍼컴퓨터 가동
http://biotech.nikkeibp.co.jp/bionewsn/detail.jsp?id=20053550&newsid=SPC2008060455942&pg_nm=1&sail=0&new1=1&news1=1&icate=0&yunw=1