

차세대 게놈 시퀀싱 기술 개발 동향

이성훈 (국가생물자원정보관리센터(KOBIC) 연구개발실장)

2003년 끝난 휴먼게놈프로젝트에 13년의 시간과 3조 5천억 달러가 사용되었다. 자동화된 Sanger DNA 시퀀서와 고성능 컴퓨터 시스템과 프로그램이 효율적으로 결합되어, 30억 염기쌍 서열을 해독을 성공적으로 마쳤다. 15년 계획으로 시작된 이 프로젝트는 DNA 시퀀싱 기술의 진보로 인해 2년의 시간이 단축되었다.

다시 한번 시퀀싱 시간과 비용을 단축시키기 위한 치열한 경쟁이 벌어지고 있다. 2004년 미국 NHGRI (National Human Genome Research Institute)가 시작한 혁신적인 시퀀싱 기술 개발을 위한 프로젝트의 목표는 "한 개인의 전장 유전체 서열해석을 1000 달러 이하의 비용으로 임상적 처방에 적용할 수 있을 정도로 짧은 시간 내에", 즉 수시간에서 수일 안에, 수행할 수 있는 기술을 개발하는 것이다. 수십 개의 단체나 회사가 참여하여 기술개발을 진행 중이다. 그 중 성공적으로 진행되고 있는 기술들의 현재 상황을 살펴보았다.

1. Illumina, Genome Analyzer II

Illumina Genome Analyzer (GA)는 진핵생물 게놈 re-sequencing에 가장 성공적으로 사용되고 있는 플랫폼이다. Sequencing by synthesis 기법을 사용하는데, 각 염기의 중합반응을 검출하기 위해 비가역적 형광 표식 terminator를 사용한다. 36~75bp의 서열을 98.5%의 정확도로 읽으며, 현재 약 20Gbp를 한번의 실험으로 읽을 수 있다.

Illumina사의 로드맵에 따르면 올해 말까지 paired-end로 90Gbp를 한번에 읽을 수 있도록 업그레이드할 계획이다. 즉 한번에 약 1만달러의 비용으로 30억 bp의 인간유전체를 30X depth로 1주일 만에 읽을 수 있다. 일반적으로 20X depth가 넘으면 heterozygous SNP까지 거의 정확하게 검출해 낼 수 있기 때문에, 한번의 실험만으로 개인유전체 분석을 완료할 수 있다. 올해 초와 비교하면, DNA 해독에 들어가는 시간과 비용이 모두 수십 분의 일로 낮아진 것이다. 현재까지 발표된 5편의 인간유전체 re-sequencing 논문 중 James Watson 박사 유전체분석을 제외한 4편에서 Illumina GA를 이용해 분석되었을 만큼 신뢰도가 검증된 방법이다. 한국의 김성진 박사 유전체분석에도 이 기법이 사용되었다.

2. Applied Biosystems, SOLiD

ABI SOLiD는 sequencing by Ligation이라는 시퀀싱 방법을 사용한다. 2 base encoding system이라는 방법을 쓰기 때문에 99.94% 이상의 정확도로 시퀀싱할 수 있는 장점이 있다. 이는 차세대 시퀀싱방법 중에서 가장 높은 정확도이다. 또한 기존의 ver.2는 35bp 길이밖에 읽지 못했지만, 최근에 공개한 ver.3은 50bp 길이를 mate pair로 읽을 수 있고, 한 실험당 40Gbp를 읽을 수 있다. 즉 2번의 실험으로 인간유전체 분석에 필요한 양의 서열을 얻을 수 있다.

Mate pair 사이 insert 길이는 600bp에서 10kbp로 상당히 길다. 이는 다른 방법들과 비교할 때, 장점이자 단점이 될 수 있다. 해독 후 맵핑이나 assembly 같은 분석작업에서 진핵생물의 반복서열때문에 일어나는 문제를 상당부분 해소해 줄 수 있는 장점과, 긴 insert로 인해 mate pair 사이 insert 길이의 오차가 크게 생겨 정확한 분석이 어려울 수 있다는 단점이 그것이다.

SOLiD는 높은 정확도를 무기로 GA의 긴 read 길이와의 치열한 경쟁이 예상된다.

3. Roche/454 Life Sciences, GS FLX

Illumina GA와 비슷하게, sequencing by synthesis 기법을 이용해 서열을 읽는다. GS FLX의 가장 큰 장점은, 차세대 시퀀싱 기법 중에서 가장 read 길이가 길다는 것이다. Roche가 최근 발표한 Titanium 시스템은 평균 400bp를 paired-end로 읽을 수 있다.

긴 길이의 paired-end read를 이용해 정확한 de novo assembly가 가능하며, 차세대시퀀싱기술로 미생물 게놈을 de novo sequencing한 논문의 상당 수가 GS FLX로 읽은 긴 서열을 이용한 것이다.

짧은 read 길이의 Illumina GA나 ABI SOLiD를 이용해서 발표된 de novo assembly 논문이 거의 없는 것을 보면, 미생물의 de novo assembly 분야에서 독보적인 위치에 있다고 할 수 있다. 단점은 한 실험당 읽을 수 있는 양이 약 0.5Gbp로서 GA나 SOLiD의 1/40~1/80 정도로 아주 낮고, 비용이 비싼 것이다.

Roche는 이런 단점을 극복하기 위해서, 시퀀싱 길이를

1kb로 늘리는 방향을 선택했다. 즉 de novo sequencing 과 RNA-Seq, metagenomics 등의 분야에 최적화된 플랫폼으로 개발하는 것이다. 또한 paired end read 사이의 insert 길이를 3kb, 8kb, 20kb까지 늘림으로써 고품질의 draft assembly를 만들 수 있는 것을 목표로 하고 있다.

4. Danaher Motion Polonator G.007

Danaher Polonator는 하버드 의과대학의 조지 처치 박사에 의해 개발된 것으로서, 전체적으로 ABI SOLiD와 유사한 sequencing by ligation 기법을 사용하여 서열을 읽는다.

현재 3~4 base gap을 둔 13bp의 짧은 paired-end read를 읽을 수 있기 때문에, 실험에 적용하기에는 아직 무리가 있다. 하지만, 10Gbp를 읽을 수 있는 한 실험에 드는 비용이 500~1,500 달러로 다른 플랫폼과 비교할 수 없이 싸다. Danaher Polonator의 가장 큰 특징은 open source idea에 기반한 사업모델이라는 것이다. 이는 개발자인 조지 처치 박사의 철학으로서, Danaher Polonator의 각 부분을 사용하거나 다른 회사가 마음대로 고치거나 바뀔 수도 있다. 사용되는 시약킷과 flow cell도 마찬가지이다. 유사한 전략의 IBM 컴퓨터가 Apple을 이기고 전 세계를 장악했듯이, Polonator가 다른 경쟁사제품을 이길 수 있을까? 창의성과 도전정신이 있는 세계의 여러 개발자/사용자에 의해 Polonator의 운명이 결정될 것으로 보인다.

5. Pacific Biosciences, Single-Molecule Read Time (SMRT) Sequencing

2010년도에 상업화가 계획된 단일분자 시퀀싱 기계를 개발 중이다. Sequencing by synthesis 기법을 사용하는 이 기계의 가장 큰 특징은 단일분자(Single Molecule)를 실시간(Real-Time) (SMRT)으로 시퀀싱한다는 것이다. 긴 서열을, 짧은 시간 내에, 높은 정확도로, 적은 노력과, 낮은 비용으로 시퀀싱하는 것이 목적이다.

SMRT가 가능하기 위해서는 단일 형광을 검출할 수 있어야 한다. 이는 Zero Mode Waveguides (ZMWs)라는 기술을 이용한 CCD 카메라에 의해 가능하다고 한다. 이 CCD 카메라가 분리된 단일가닥 DNA상에 중합되는 개개의 핵산이 발하는 형광을 검출하는 것이다.

1.5kbp를 분당 600bp의 속도로 해독하는데 성공했다. Pacific Biosciences사의 궁극적인 목표는 시간 당 100Gbp를 해독하는 것이다.

Pacific Biosciences사가 처음 이 아이디어를 발표할 당시, 많은 사람들이 획기적인 기술에 기대를 하면서도 실제로 가능할지 의구심을 가졌다. 그러나, 작년 7월, Intel이 포함된 투자단체로부터 1억달러의 투자를 받았다는 소식은 이런 의구심을 상당부분 해소시켜 주었다.

이 방법은 현재까지 나온 가장 이상적인 시퀀싱방법이다. 단일세포의 염색체 변이를 읽을 수 있을 뿐만 아니라 2배체, 4배체 등 다배체에서 각 염색체별 변이를 분석하는데 별도의 조작이 필요없다.

올 6월, 일루미나는 이 기술을 네 개의 유전정보 분석회사(23andMe, Knome, Navigenics, Decode Genetics)와 협력하여, 개인유전체를 48,000달러에 분석 서비스할 예정이라고 발표했다. 같은 시기에, Knome과 SeqWright는 SOLiD와 454 플랫폼을 이용한 서비스를 개시한다고 발표했다. 비용은 개인은 24,500달러, 가족에 속하면 19,500달러이다.

DNA 시퀀싱기술 개발의 성공에 힘입어 1000달러 개인유전체 분석이 가시권에 들어옴에 따라, Google, Microsoft, Intel, IBM 등 생명과학과 무관한 것처럼 보였던 거대 기업이 개인유전체 시장을 선점하기 위해 각축전을 벌이고 있다. 정보처리 및 서비스와 분석에 필요한 저장장치, 메모리 등의 하드웨어 등 부가산업의 개발이 동반될 것이기 때문이다.

생명과학의 대중화를 이끌어 나가기 위한 국가차원의 준비가 필요한 때다.

[참고 웹 사이트]

- [1] NHGRI Next-generation sequencing
<http://www.genome.gov/>
- [2] Illumina 사
<http://www.illumina.com>
- [3] Applied Biosystems 사
<http://www.appliedbiosystems.com>
- [4] Roche/454 Life Sciences 사
<http://www.454.com/>
- [5] Polonator R.007
<http://polonator.org/>
- [6] Pacific Biosciences
<http://www.pacificbiosciences.com/>
- [7] Intel의 Pacific Biosciences 투자 뉴스
<http://webreprints.djreprints.com/1991460243367.html>
- [8] Illumina와 4개의 유전정보 분석회사 제휴서비스 뉴스
<http://www.genomeweb.com/sequencing/illumina-partners-dtc-genomics-firms-offer-whole-genome-sequencing-48k>
- [9] Knome과 SeqWright 분석서비스 뉴스
<http://www.genengnews.com/news/bnitem.aspx?name=56032121&chid=2&taxid=20>