

## Performance Comparison of Benchtop Next-generation Sequencing Systems

Hee Sam Na\*

*Department of Oral Microbiology, School of Dentistry, Pusan National University, Yangsan Korea*

With fast development and wide applications of next generation sequencing (NGS), genomic sequence information is within reach to various research fields. Three benchtop NGS instruments are now available. The 454 GS Junior (Roche), Ion PGM (Life Technologies) and MiSeq (Illumina) are laser-printer sized and offer modest set-up and running costs. By reviewing 2 studies that compared the performance of these instruments, the major characteristics of each benchtop platforms are compared to enable direct comparisons. The 454 GS Junior generated the longest reads and most contiguous assemblies but had the lowest throughput. The Ion Torrent PGM had the highest throughput and fastest run time. The MiSeq had the highest throughput per run and lowest error rates. The Ion Torrent PGM and 454 GS Junior both produced homopolymer-associated indel errors. Although all the platforms allow multiplexing of samples, details of experimental design, library preparation and data analysis may constrain the options. The features of the platforms provide opportunities both to conduct groundbreaking studies and to waste money. Thus, careful considerations should be made before purchasing or using any of them.

**Key Words:** Next generation sequencing, 454 GS Junior, Ion torrent, MiSeq

저자는 최근 개인 실험실에 보급되고 있는 Next-generation sequencing (NGS) 플랫폼 가운데 454 GS Junior, Ion torrent PGM과 MiSeq을 비교 발표한 논문을 읽고 이를 중심으로 각 플랫폼의 장단점을 아래와 같이 전달하고자 한다. 주요 참고 문헌은 'Comparison of Next-Generation Sequencing Systems (1), Performance comparison of benchtop high-throughput sequencing platforms (2, 3), Performance Comparison of Bench-Top Next Generation Sequencers Using Microdroplet PCR-Based Enrichment for Targeted Sequencing in Patients with Autism Spectrum Disorder (4)'이다.

유전체 연구는 NGS 기법의 도입과 함께 눈부시게 발전하고 있다. 특히 최근 시장에 연구소 단위의 커다란

기관뿐만 아니라 개인 연구실에서도 운영할 수 있을 정도로 부피가 작고 가격이 비교적 저렴한 NGS 기기들이 소개되고 있다. 가장 대표적인 NGS 장비로 454 GS Junior, Ion torrent PGM과 MiSeq가 있다. 454 GS Junior의 경우 가장 긴 read length를 보인 반면 가장 작은 데이터를 생산하였다. Ion PGM은 가장 빠른 데이터 생산과 가장 짧은 운용 시간을 보인 반면 동종 중합체에 대해 상당한 오류를 나타냈다. MiSeq는 가장 많은 데이터를 생산하였고 Mb당 가장 낮은 비용이 소모되었다. 이 글에서는 최근에 발표된 논문을 살펴봄으로써 각 NGS 장비의 특징과 장·단점을 비교하고자 한다.

지난 10여년간 유전자 염기 서열 분석의 발전은 의학 연구에 많은 변화를 가져왔다. 특히 최근 몇 년간 NGS

Received: March 21, 2014/ Revised: April 7, 2014/ Accepted: April 9, 2014

\*Corresponding author: Hee Sam Na. Department of Oral Microbiology, School of Dentistry, Pusan National University, Yangsan, 626-810, Korea. Phone: +82-51-510-8252, Fax: +82-51-510-8246, e-mail: heesamy@pusan.ac.kr

\*\*This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Ministry of Education, Science and Technology (NRF-2011-0013215). The authors have no financial conflicts of interest.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).

기술은 세균과 인간의 유전체 염기 서열 분석에 커다란 변화를 가져왔다 (5, 6). NGS가 시장에 소개되면서 많은 기술적인 개선이 지속적으로 이루어졌고, 다른 플랫폼과 경쟁함에 따라 유전자 염기 서열 분석에 커다란 혁신을 가져오게 되었다. 이러한 변화는 우리가 무엇을 할 수 있고, 어떻게 진행하고, 얼마에 가능한지를 결정하기에 이러한 기술의 흐름에 보조를 맞추는 게 필요하다. 기관들과 연구자들은 매년 기계 구매에 상당한 금액을 투자하고 있다. 그러한 구매는 실험실과 연구소에서 소모하는 소모품과 기술적 지원에 매년 상당한 연구비 지출로 연결된다. 일단 기계를 구입하면 기계적 특징, 화학 작용 기전, 연구비 등이 연구의 내용과 틀을 제한하게 된다.

이 글의 목적은 최근 시장에 소개된 benchtop NGS 기기들의 실험 결과를 비교한 논문을 살펴봄으로써 사용자의 목적에 맞는 플랫폼을 선택하는데 도움을 주고자 하는 것이다.

## 각 플랫폼에 대한 소개

### 1. Roche 454 GS Junior

2010년 초반에 출시된 454 GS Junior는 이전에 소개되었던 GS FLX Titanium system의 소형, lower-throughput 버전이다. GS Junior는 library 준비와 데이터 분석이 보다 간소화되었고 한번에 14 G까지 데이터를 생성할 수 있다. Roche 454는 최초의 성공적인 NGS로 pyrosequencing 기술을 기반으로 한다 (7). 2008년 출시된 454 GS FLX 시스템은 read length가 700 bp까지 이르고 정확도가 99.9%이며 1회 운영으로 0.7 G 데이터를 생산한다. Roche의 가장 큰 장점은 비교적 빠른 실험 진행과 read length이다. 실험은 염기 서열 분석 시작 후 10시간 후면 얻을 수 있고 read length도 다른 NGS 보다 탁월하게 길다. 그러나, 운영에 소모되는 시약의 가격은 상당한 단점으로 지적되고 있다. 또 다른 단점으로는 6 bp 이상 되는 poly-base의 경우 발생하는 에러율이다.

### 2. Ion Personal Genome Machine (PGM)

Ion PGM은 Ion Torrent사에서 2010년 말에 출시한 제품으로 반도체 염기 서열 분석 기술을 사용한다. Nucleotide가 중합 효소에 의하여 DNA에 결합하면 수소 이온이 유리된다. 이때 발생하는 pH의 변화를 측정함으로써 PGM은 nucleotide가 결합하였는지 여부를 알 수 있다. 매번 칩

에 nucleotide가 한 가지씩 제공되어 맞는 염기 서열인 경우 결합하여 pH 변화가 관찰되고 맞지 않다면 아무 변화도 관찰되지 않을 것이다. 만약 2개의 nucleotide가 결합하는 경우 변화는 2배가 된다 (8). PGM의 가장 큰 특징은 최초로 형광과 카메라 scanning을 필요로 하지 않는 염기 서열 분석 장비로 빠르고, 저렴하고 기계가 작다는 것이다. 현재 2시간 이내에 200 bp read가 가능하고 8개의 시료를 동시에 6시간 이내에 준비하는 것이 가능하다. 형광 염료를 사용하는 경우 50 cycle 이후에는 read length가 길어짐에 따라 형광이 약해지는 단점이 있는 반면 PGM은 형광 염료를 사용하지 않기 때문에 50 cycle 이후에도 품질이 우수한 장점이 있다.

### 3. MiSeq system

Illumina사에서 출시한 MiSeq는 sequencing by synthesis (SBS) (9), cluster generation, 데이터 분석 기능을 통합한 기계로 하루 안에 분석 결과를 얻을 수 있다(짧은 경우 8시간 이내). SBS 기술은 library에 adaptor를 결합시킨 후 flowcell에 부착시켜 bridge amplification을 통하여 각 DNA 단편에 대한 clone을 형성한다. 염기 서열 분석은 linearization 효소의 도움을 받아 단가닥(single strand)로 먼저 만든 후 각각 다른 형광을 띄는 네 가지 nucleotide (ddATP, ddGTP, ddCTP, ddTTP)와 blocking group이 한번에 하나씩 상보적 결합을 형성하고 그 신호를 CCD 카메라로 측정하여 염기 서열을 분석하는 방법이다. MiSeq의 가장 큰 특징은 가장 높은 데이터 무결성과 다양한 활용 범위이다. MiSeq는 amplicon 염기 서열 분석, clone 확인, ChIP-Seq, 작은 genome 염기 서열 분석 등에 사용될 수 있다. 단일 36 bp read (120 Mb 산출)에서부터 2×150 paired-end read (1~1.5 Gb 산출)까지 폭넓은 분석이 가능하다. Read length의 개선으로 HiSeq에 비하여 contig assembly에 보다 나은 결과를 보인다.

## 플랫폼 간 비교 실험 논문 고찰

### 1. Escherichia coli O104:H4 genome sequencing 결과

Nicholas 등은 454 GS Junior, MiSeq, Ion torrent PGM 세 가지 플랫폼을 사용하여 *Escherichia coli* O104:H4의 유전체 염기 서열 분석 결과를 비교하였다 (2). 각 플랫폼의 가격과 기본적인 특징을 Table 1에 비교하여 정리하였다.

**Table 1.** Comparison of benchtop instruments and sequencing runs (modified from reference # 2)

Platform	List price	Cost per run	Throughput (Read length)	Run time	Cost/Mb	Mb/h
454 GS Junior	\$108,000	\$1,000	35 Mb (400 bp)	8 hr	\$31	4.4
Ion PGM						
(314 chip)	\$80,490	\$225	10 Mb (100 bp)	3 hr	\$22.5	3.3
(316 chip)		\$425	100 Mb (100 bp)	3 hr	\$4.25	33
(318 chip)		\$625	1,000 Mb (100 bp)	3 hr	\$0.63	333
MiSeq	\$125,000	\$750	1,500 Mb (2 × 150 bp)	27 hr	\$0.5	55.5

**Table 2.** Run and alignment metrics for benchtop sequencers (modified from reference # 2)

Platform	Number of reads	Total bases	Modal read length in bases	Mean read length in bases (s.d.)	Alignment coverage		
					Chromosome	Large plasmid	Reads aligned (%)
454 GS Junior	135,992	70,999,968	518	522 (46)	11.50	5.66	99
Ion PGM (316 chip)	2,154,577	260,017,346	123	121 (16)	39.33	43.80	89
MiSeq	11,708,156	1,652,529,000	150	141 (22)	-	-	-
MiSeq demultiplexed strain 280	1,766,516	250,356,566	150	141 (21)	22.11	625.46	99

Metrics for each sequencing run are shown as well as results of alignment against the reference sequence. Depth of coverage for the chromosome and two large plasmids (pESBL and pAA) are shown with the percentage of read that align. For the MiSeq run, the sequence metrics are shown for the entire run as well as the results of de-multiplexing *E. coli* O104:H4 strain 280.

유전체 염기 서열의 재구성은 유전체 심도(genome depth), 분석의 균등도, read length, read quality에 의해 결정된다. 분석 결과는 서로 상이한 특징을 보였다(Table 2). 454 GS Junior는 평균 522 bp의 가장 긴 reads를 보였지만 세 플랫폼 가운데 가장 작은 데이터(70 Mb)를 생산하였다. Ion PGM은 454 GS Junior보다 4배 이상의 데이터를 만들어냈지만 평균 121 bps의 가장 짧은 reads를 만들었다. MiSeq는 가장 큰 데이터(1.6 Gb)와 함께 Ion PGM보다 약간 더 긴 141 bp의 read를 만들었다. MiSeq의 경우 각 유전체의 40배 coverage를 목표로 하였을 때 7개의 *E. coli* 종을 동시에 분석할 수 있었다. MiSeq는 DNA를 양 쪽으로 모두 분석하였다. 참고 유전체(reference genome)에 대하여 3가지 플랫폼은 비슷한 정도의 coverage를 보였다. 그러나, MiSeq의 경우 Shiga-toxin producing phage와 관련된 피크가 관찰되었고 Ion PGM에서도 약한 피크가 동일하게 관찰되었다. 이러한 피크는 DNA 준비 과정에

서 발생할 수 있는 phage lysis에 의한 것으로 여겨지고 있다.

각 제조사마다 염기 서열의 질을 관리하는 프로그램이 각각 다르기 때문에 직접적인 비교를 하는 것은 어렵다. 그래서, 참고 유전체를 기준으로 이에 대한 유전체 분석 결과를 비교하여 alignment quality를 비교하였다. 이렇게 측정된 alignment quality 점수는 prediction score와 유사한 결과를 보였다. 즉, Ion PGM은 비교적 quality score를 과소평가하는 경향이 있었고 다른 두 플랫폼은 과대평가하는 경향이 있었다. MiSeq가 가장 좋은 질(quality)을 나타냈는데 이는 낮은 대체 오류율(substitution error rate) (100 bp당 0.1 대체)와 indel 오류의 부재에 의한 것으로 여겨진다. Ion PGM은 100번째 염기 서열까지 지속적으로 정확도가 감소하는 경향을 보였다.

Indel의 빈도를 비교한 경우 Ion PGM의 경우 100 bp당 1.5의 indel이 발생하였다. 454 GS Junior의 경우 100 bp당

**Table 3.** Insertion/deletion and substitution errors on read level for benchtop sequencer (modified from reference # 3)

Platform	Sequencing kit	Indels per 100 bp	Indels per read	Substitution per 100 bp	Substitution per read
454 GS Junior	GSJ titanium	0.4011	1.8351	0.0543	0.2484
Ion PGM	100 bp	0.3520	0.3878	0.0929	0.1024
	200 bp	0.3955	0.6811	0.0303	0.0521
	300 bp	0.7054	1.4457	0.0861	0.1765
MiSeq	Nextera	0.0009	0.0013	0.0921	0.1318

0.38의 indel이 발생한 반면 MiSeq의 경우 거의 발생하지 않아 100 bp당 0.001 이하로 발생하였다. Indel의 가장 흔한 원인은 동종 중합체(homopolymers)였다. 가장 흔한 오류는 deletion으로 동일 염기 서열이 6개 이상인 경우 60% 이하의 정확도를 보였다. 이러한 결과는 Table 3에 정리하였다 (3).

고속 대량(high-throughput) 염기 서열 분석을 통한 유전자의 비교는 정확한 데노보 조합 (*de novo assembly*)이 중요하다. 다양한 작시법(metrics)를 사용하여 데노보 조합을 비교하였다. 작시법 가운데 총 조합 크기와 N50은 조합의 완벽성 또는 과편화 정도를 비교하기에는 좋은 반면 정확성에 대해서는 잘 알기 힘들다. 이상적인 조합 결과는 각 레플리콘(replicon)에 대해 하나의 콘택(contig)를 만들어야 하지만 반복적인 염기 서열은 이러한 것을 현실적으로 어렵게 한다. 세 가지 플랫폼을 비교한 결과 두 가지 특징이 발견되었다. 심한 과편화가 Ion PGM, 454 GS Junior (single run), MiSeq에서 나타났다. Coverage의 심도를 증가시키기 위해 454 GS Junior를 2회 반복하여 결합한 경우와 MiSeq 데이터에서 생성된 콘택 결과와 paired-end 정보로 보정한 경우 과편화가 보다 적게 나타났다.

참고 유전체에 대해 정확히 일치하는 콘택의 수는 유전체 coverage를 나타내는 좋은 자료이다. 각 플랫폼 간에 coverage의 차이를 보였다. 454 GS Junior의 경우 가장 넓은 coverage 결과를 보여 3.72%의 결합율을 보인 반면 Ion PGM은 4.6%, MiSeq은 3.95%의 결합율을 보였다.

Ion PGM은 다른 플랫폼에 비하여 가장 많은 수의 공백(gap)을 만들었다. 두 대의 Ion PGM의 결과를 결합하여 coverage를 증가시킨 경우 공백의 수를 감소시킬 수 있었으나, 동종 중합체에 의한 오류는 여전히 남아 있었다. Ion PGM 결과를 결합하여 조합하여 만들어진 결과에서

1/3에서 1/4의 공백은 콘택의 말단이나 비기록 염기 서열 (unmapped sequence)에 의한 것이었고 이외는 동종 중합체에 의한 것이었다. 비록 동종 중합체에 의한 공백의 증가는 일어나기 마련이지만 짧은 동종 중합체(2~3개)에 의한 공백은 454 GS Junior보다 유의하게 많이 발생하였다.

실험 결과 세 플랫폼 모두 *E. coli*의 유전체 염기 서열 분석에 크게 문제가 없었고 각 플랫폼 간의 장 단점이 존재하였다. MiSeq는 가장 큰 데이터 결과와 가장 낮은 오류율, indel이 거의 없고 가장 낮은 대체 오류를 보였다. 그러나 MiSeq는 454 GS Junior보다 read length가 현저히 짧아 조합의 질이 떨어지는 결과를 보였다. Paired-end 염기 분석을 시행하여도 단일 분석으로는 공백을 감소시키기 어려웠다. 또한 가장 긴 운영 시간이 소요(27시간 이상)되었다.

## 2. Autism spectrum disorder (ASD)에서 표적 염기 서열 분석 비교 결과 (4)

NGS 기술을 gene enrichment과 결합하여 사용함으로써 혼합된 질환들과 관련된 유전자의 염기 서열 분석이 가능하게 되었다. ASD는 수백 개의 관련 유전인자와 연관된 복잡한 질환으로 소유진 형태(polygenic mode)로 유전된다 (10). ASD는 1.1%의 유병율을 보이는 비교적 흔한 질환으로 어린아이 때 사회적 상호작용 장애, 의사소통 결함, 반복적인 행동 등을 특징으로 한다 (11). ASD의 유전에 대한 연구 가운데 단일 뉴클레오타이드 변형체 (single nucleotide variants, SNPs)가 보고되었고, 약 40% 유전되는 것으로 보고되었다 (12). 그러나, 이 외에 *de novo* 또는 유전되는 희귀 변형체들도 상당수 보고되고 있다 (13). 이에 Koshimizu 등은 ASD 환자에서 희귀 변형체를 탐색하기 위하여 Ion PGM과 MiSeq를 사용하여 비교하였다 (4).

**Table 4.** Comparison between Ion PGM and MiSeq sequencing performance in 10 positive controls (modified from reference # 4)

	Ion PGM (TMAP*)	MiSeq
Total read (Mb)	295.97	469.42
Average read length (bp)	116	150
% mapped on human genome	96.8%	75%
% on target regions	26.7%	22.7%
Mean depth of coverage	63	95
% of target regions at > 10-fold coverage	93.7%	96.8%
% of target regions at > 20-fold coverage	85.9%	93.2%

\*TMAP (Torrent Mapping Alignment Program) is a customized mapping tools for sequencing data generated by PGM, ignoring the indel calls around homopolymer stretch to reduce the hundreds of false negative calls.

Library는 환자의 말초 혈액에서 채취한 DNA를 RainDance ASD Research Screening Panel를 사용하여 제작하였다. RainDance ASDSeq panel은 RainDance Technology 사 (Lexington, MA, USA)에서 제공한 것으로 ASD와 관련되어 알려진 62개 유전자의 92%를 포함하는 유전자 스크리닝 키트이다. Library는 167에서 600 bp 길이의 2349 amplicon을 포함하며 1,034 kb 구역을 포함하였다. Coverage는 intron/exon splice junction과 1 kb 길이의 5' promoter 구역과 3' UTR을 포함하여 각 유전자의 exon과 그 앞, 뒤 50 bp를 포함하고 있다.

실험 결과 62개의 엑손이 모두 증폭되었다. RainDance의 효율과 각 플랫폼의 성능을 확인하기 위하여 10개의 양성 대조군을 사용하였다(Table 4). Ion PGM과 MiSeq는 각각 296 Mb와 469 Mb의 데이터를 생성하였고 96.8%와 75%가 유전체와 일치하였다. 또 각각 26.7%와 22.7%가 표적 구역에 해당하였다. Depth of coverage 평균은 각각 63과 95였다. Base-call quality score를 분석한 결과 Ion PGM은 초기에는 >25를 기록하였으나 100 bp 근처에서는 20으로 감소하는 경향을 보였다. MiSeq는 전체적으로 >30을 기록하였다. Read length는 Ion PGM은 60에서 150 bp까지 다양한 반면 MiSeq는 151 bp로 일정하였다. 다음으로 무작위로 특정 유전자를 선택하여 indel의 수를 비교한 결과 Ion PGM은 9,685개의 SNPs와 indel이 보고되었고 이 가운데 5,544 (57.2%)가 indel 이었다. Indel의 빈도는 1 kb당 1.34였다. MiSeq는 3,818개의 SNPs와 indel이 보고되었고 이 가운데 394 (10.3%)가 indel 이었다. Indel의 빈도는 1 kb당 0.096였다.

유전 변이를 측정 효율을 비교하기 위하여 Sanger법으

**Table 5.** Validation for mutation detection (modified from reference # 4)

	Ion PGM (TMAP)	MiSeq
Detection (+/total)	7/10	10/10
Coverage (range)	36.8 ± 21.8 (8~77)	92.4 ± 53.2 (46~223)
Mutant allele (%) (range)	50.1 ± 17.1 (33~83)	55.7 ± 23.1 (38~100)

로 확인된 유전자를 대상으로 비교하였다(Table 5). 확인된 변이는 최소한 x8 reads의 coverage를 보였고 이형접합자(heterozygote)의 경우 33~62%, 동형접합자(homozygote)의 경우 83~100%의 변이 형질률(mutant allele percentage)을 보였다. 변이 유전자를 검출률은 Ion PGM과 MiSeq 각각 70%와 100%였다. Ion PGM의 경우 동종 중합체 주변에 위치한 변이의 경우 동종 중합체 염기 분석 오류로 인하여 검출하지 못한 것으로 여겨진다.

두 플랫폼을 이용하여 유전자 변이를 검사한 결과 40 이하의 mapping quality를 보이는 read의 비율이 Ion PGM이 MiSeq에 비하여 월등히 높았다. Indel에 대해서도 MiSeq의 경우 고전적인 capillary sequencing의 오류율(1 kb당 0.11에서 0.88)과 비슷한 정도의 낮은 오류율을 보였다.

앞서 살펴본 두 개의 논문과 리뷰 논문을 종합하여 보면 454 GS Junior의 경우 가장 긴 read length를 보인 반면 가장 작은 데이터를 생산하였고, 동종 중합체에 의한 오류가 많이 발생하였다. Ion PGM은 가장 짧은 read length를 보였고 동종 중합체에 대해 가장 불량한 결과를 나타

냈으나, 가장 빠른 데이터 생산(80~100 Mb/h)과 가장 짧은 운용 시간(약 3시간)을 보였다. 또한 Ion PGM은 최근 가장 빨리 성능 개선이 이루어지고 있다. MiSeq는 가장 많은 데이터를 생산하였고 Mb당 가장 낮은 비용이 소모되었다. 또한 오류율이 고전적인 Sanger 분석법과 비슷한 정도로 낮게 발생하는 특징을 보였다.

플랫폼을 비교함에 있어 속도, 셋업, 운영비, 운영의 단순함 또한 중요하다. Ion PGM은 세 플랫폼 가운데 가장 저렴하다. 염기 서열당 운영비는 454 GS Junior가 다른 두 플랫폼에 비하여 한 단위 더 많이 드는 편이다. MiSeq의 운영은 기기에서 증폭이 바로 이루어지기 때문에 다른 플랫폼에 비하여 비교적 더 단순하다. Ion PGM은 3가지 칩을 제공하고 있어서 자신의 목적에 맞게 선택할 수 있는 장점이 있다.

## REFERENCES

- 1) Liu L, Li Y, Li S, Hu N, He Y, Pong R, *et al.* Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012;2012:251-364.
- 2) Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012;30:434-9.
- 3) Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, *et al.* Updating benchtop sequencing performance comparison. *Nat Biotechnol* 2013;31:294-6.
- 4) Koshimizu E, Miyatake S, Okamoto N, Nakashima M, Tsurusaki Y, Miyake N, *et al.* Performance comparison of bench-top next generation sequencers using microdroplet PCR-based enrichment for targeted sequencing in patients with autism spectrum disorder. *PLoS One* 2013;8:e74167.
- 5) Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010;11:31-46.
- 6) Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* 2011;11:759-69.
- 7) Products - Technology: 454 Life Sciences, a Roche Company 2014. <http://my454.com/products/technology.asp>.
- 8) Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 2010;7:461-5.
- 9) Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53-9.
- 10) Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011;70:863-85.
- 11) Beecham J. Annual Research Review: Child and adolescent mental health interventions: a review of progress in economic studies across different disorders. *J Child Psychol Psychiatry* 2014.
- 12) Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, *et al.* Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* 2012;3:9.
- 13) Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012;485:237-41.