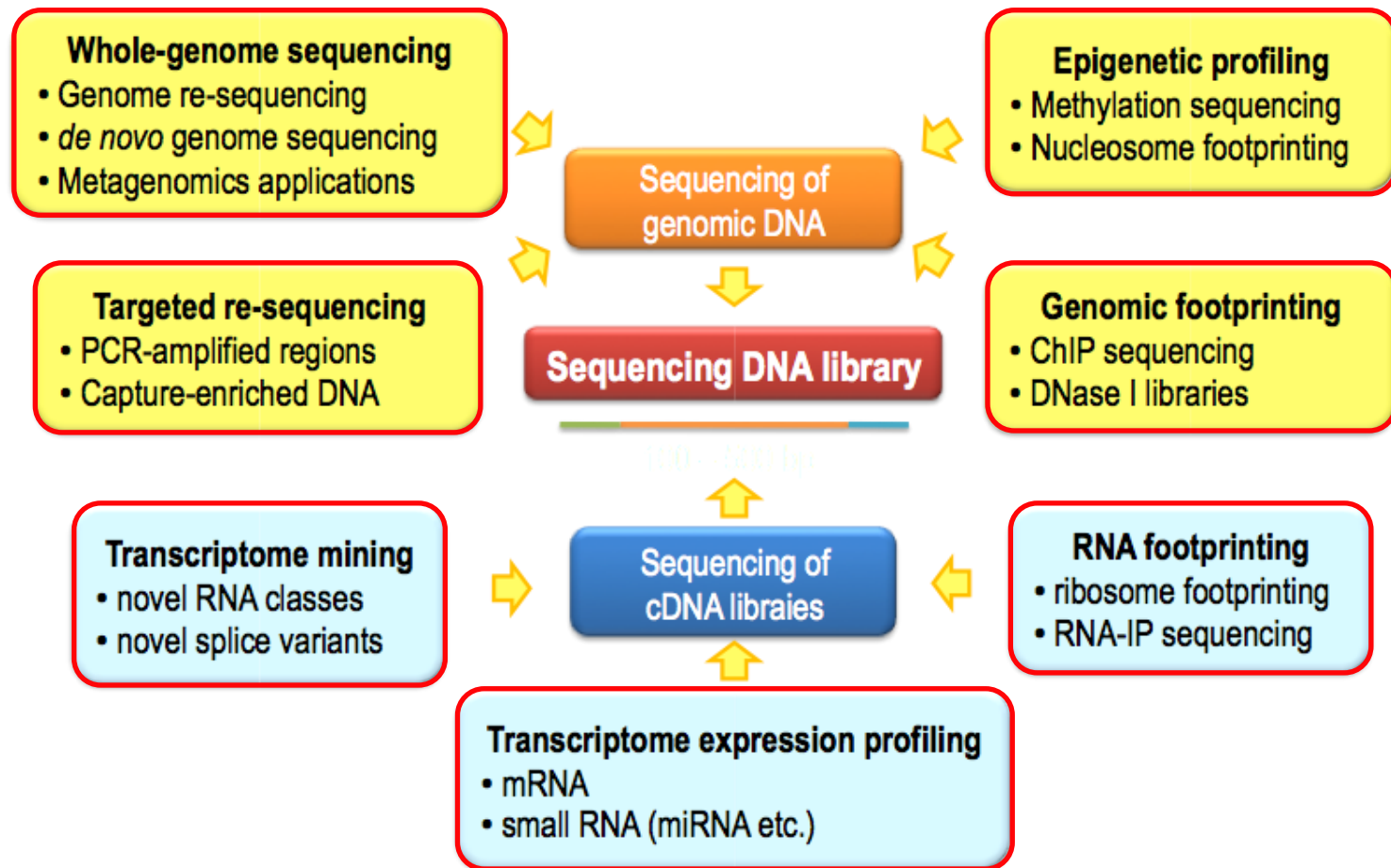


ION PROTON VS ILLUMINA HISEQ2500 VS SOLID 5500

Applications of Next-Generation Sequencing



Next-Generation Sequencing Workflow



1. Sample Preparation



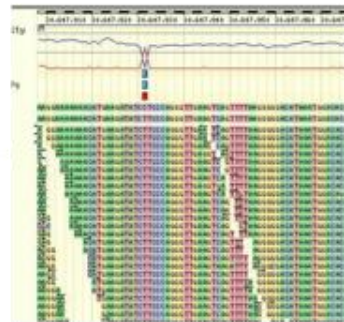
2a. Flow Cell



2b. Cluster Generation



6. Publication



5. Data Analysis



4. Initial Image Analyses



3. Sequencing & Imaging

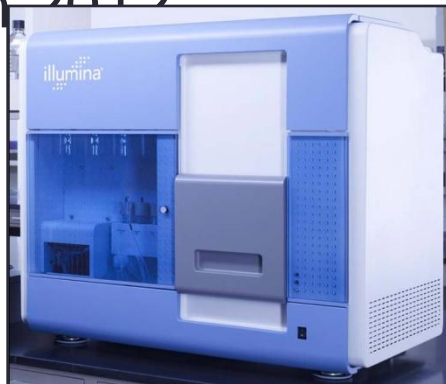
Illumina, Roche 454 or ABI SOLiD?

NEXT-GENERATION SEQUENCING (DEEP SEQUENCING) PLATFORMS

- Short reads
 1. Genome Analyzer Ix (GAIx), HiSeq2000, HiSeq2500, MiSeq – Illumina
 2. SOLiD 5500xl System – Applied Biosystem
 3. HeliScope™ Single Molecule Sequencer - Helicos
- Long reads
 1. Genome Sequencer FLX System (454) – Roche
 2. PacBio RS - Pacific Bioscience
 3. Personal Genome Machine, Ion Proton - Ion Torrent
 4. GridION – Oxford Nanopore
- Mapping sequences to large DNA fragments
 1. NABsys
 2. Bionanomatrix

Sequencing Platforms

- Roche/454 FLX: 2004
- Illumina Solexa Genome Analyzer: 2006
- Applied Biosystems SOLiD™ System: 2006
- Helicos Heliscope™ : 2009
- Pacific Biosciences SMRT: 2010
- Ion Torrent: 2011
- Ion Proton: 2012



Strengths (Ion-Torrent)

- rapid sequencing speed and low upfront and operating costs. This has been enabled by the avoidance of modified nucleotides and optical measurements.
- sequencing can occur in real-time. In reality, the sequencing rate is limited by the cycling of substrate nucleotides through the system. Ion Torrent Systems Inc., the developer of the technology, claims that each incorporation measurement takes 4 seconds and each run takes about one hour, during which 100-200 nucleotides are sequenced.
- The cost of acquiring a pH-mediated sequencer from Ion Torrent Systems Inc. at time of launch was priced at around \$50,000 USD, excluding sample preparation equipment and a server for data analysis

Limitations (Ion-Torrent)

- If homopolymer repeats of the same nucleotide (e.g. GGGGG) are present on the template strand (strand to be sequenced) then multiple introduced nucleotides are incorporated and more hydrogen ions are released in a single cycle. This results in a greater pH change and a proportionally greater electronic signal. This is a limitation of the system in that it is difficult to enumerate long repeats. This limitation is shared by other techniques that detect single nucleotide additions such as pyrosequencing. Signals generated from a high repeat number are difficult to differentiate from repeats of a similar but different number; e.g., homorepeats of length 7 are difficult to differentiate from those of length 8.
- Longer read lengths are beneficial for de novo genome assembly. The read length achieved by Ion Torrent semiconductor sequencing is currently 500 base pairs per run.

Table 1 Technical specifications of Next Generation Sequencing platforms utilised in this study

Platform	Illumina MiSeq	Ion Torrent PGM	PacBio RS	Illumina GAIIx	Illumina HiSeq 2000
Instrument Cost*	\$128 K	\$80 K**	\$695 K	\$256 K	\$654 K
Sequence yield per run	1.5-2Gb	20-50 Mb on 314 chip, 100-200 Mb on 316 chip, 1Gb on 318 chip	100 Mb	30Gb	600Gb
Sequencing cost per Gb*	\$502	\$1000 (318 chip)	\$2000	\$148	\$41
Run Time	27 hours***	2 hours	2 hours	10 days	11 days
Reported Accuracy	Mostly > Q30	Mostly Q20	<Q10	Mostly > Q30	Mostly > Q30
Observed Raw Error Rate	0.80 %	1.71 %	12.86 %	0.76 %	0.26 %
Read length	up to 150 bases	~200 bases	Average 1500 bases**** (C1 chemistry)	up to 150 bases	up to 150 bases
Paired reads	Yes	Yes	No	Yes	Yes
Insert size	up to 700 bases	up to 250 bases	up to 10 kb	up to 700 bases	up to 700 bases
Typical DNA requirements	50-1000 ng	100-1000 ng	~1 µg	50-1000 ng	50-1000 ng

Advantage and mechanism of sequencers

Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	99.999%
Reads	1 M	3 G	1200~1400 M	—
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput

Review Article

Comparison of Next-Generation Sequencing Systems

Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law

Components and cost of sequencers

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Instrument price	Instrument \$500,000, \$7000 per run	Instrument \$690,000, \$6000/(30x) human genome	Instrument \$495,000, \$15,000/100 Gb	Instrument \$95,000, about \$4 per 800 bp reaction
CPU	2* Intel Xeon X5675	2* Intel Xeon X5560	8* processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1 GB
Hard disk	1.1 TB	3 TB	10 TB	280 GB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REM e system	cBot system	EZ beads system	No
Cost/million bases	\$10	\$0.07	\$0.13	\$2400

Application of sequencers

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Resequencing		Yes	Yes	
<i>De novo</i>	Yes	Yes		Yes
Cancer	Yes	Yes	Yes	
Array	Yes	Yes	Yes	Yes
High GC sample	Yes	Yes	Yes	
Bacterial	Yes	Yes	Yes	
Large genome	Yes	Yes		
Mutation detection	Yes	Yes	Yes	Yes

Comparing Different Technologies

Sanger Sequencing

Advantages	Disadvantages
Lowest error rate	High cost per base
Long read length (~750 bp)	Long time to generate data
Can target a primer	Need for cloning
	Amount of data per run

Comparing Different Technologies

454 Sequencing

Advantages	Disadvantages
Low error rate Medium read length (~400-600 bp)	Relatively high cost per base Must run at large scale Medium/high startup costs

Comparing Different Technologies

Ion Torrent Sequencing

Advantages	Disadvantages
Low startup costs	New, developing technology
Scalable (10 – 1000 Mb of data per run)	Cost not as low as Illumina
Medium/low cost per base	Read lengths only ~100-200 bp so far
Low error rate	
Fast runs (<3 hours)	

Comparing Different Technologies

Illumina Sequencing

Advantages	Disadvantages
Low error rate	Must run at very large scale
Lowest cost per base	Short read length (50-75 bp)
Tons of data	Runs take multiple days
	High startup costs
	De Novo assembly difficult

Costs

Criteria	Ion Torrent Proton	Illumina HiSeq 2500
System Price	\$243,000	\$740,000
Annual Service cost (yr 2 & 3)	\$19,400	\$59,200
Per Gb cost	\$16.67	\$46.00
Per Run Gb	~ 60 Gb (est.)	120 Gb
Per Run cost	\$1,000	\$5520 (est)
Three year TCO @ 30% utilization	\$731,800	\$2,100,400
Three year TCO w/ HiSeq Upgrade	\$731,800	\$1,410,400
Time from library to data	8 hours	27 hours
Throughput per 40h workweek	600 Gb	480 Gb
Accuracy	99%	99%
Expected readlength at launch	200 base-pairs	150 base-pairs
Potential readlength improvement	400 base-pairs	250 base-pairs
Ease of use	+++	+++