Improvements in Next Generation Sequencing

Haley Fiske

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, and HiScan are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

illumina®

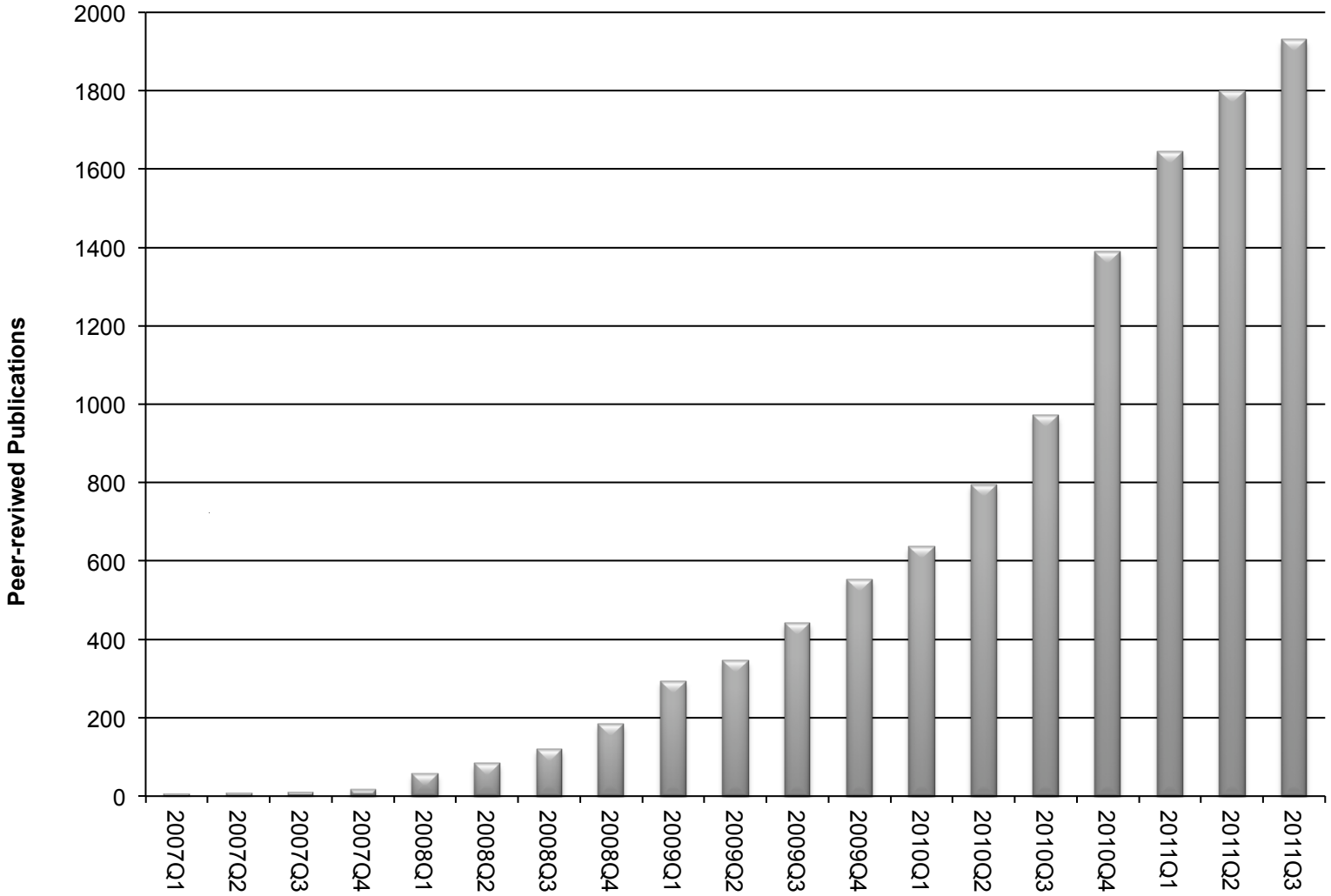
Topics

- MiSeq
- Nextera XT
- HiSeq

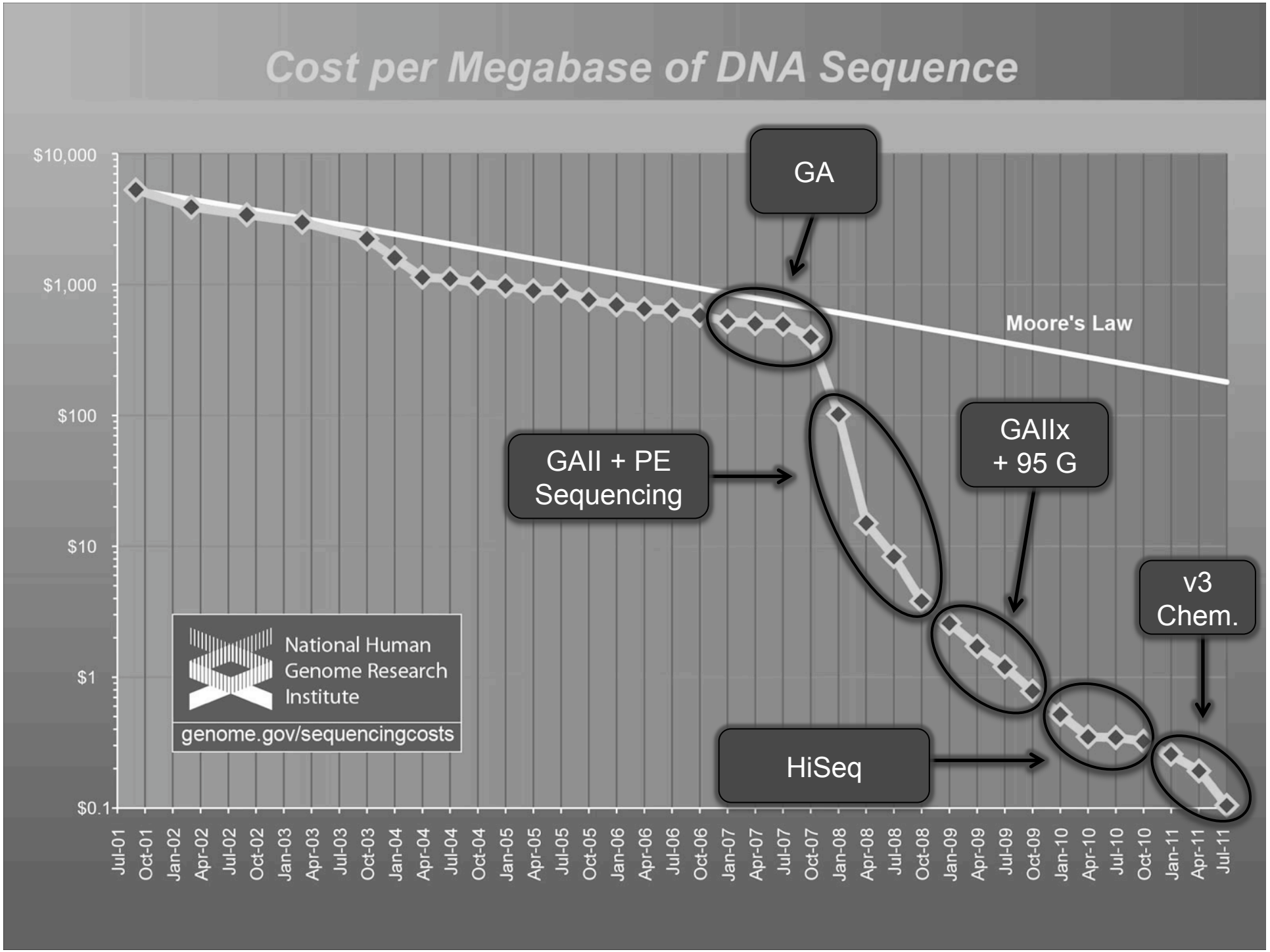
illumina®

Unparalleled Publications Ramp

Fastest NGS publication rate, >2500 peer reviewed publications



illumina®


Cost per Megabase of DNA Sequence
$10,000
$1,000
$100
$10
$1
$0.1
Jul-01
Oct-01
Jan-02
Apr-02
Jul-02
Oct-02
Jan-03
Apr-03
Jul-03
Oct-03
Jan-04
Apr-04
Jul-04
Oct-04
Jan-05
Apr-05
Jul-05
Oct-05
Jan-06
Apr-06
Jul-06
Oct-06
Jan-07
Apr-07
Jul-07
Oct-07
Jan-08
Apr-08
Jul-08
Oct-08
Jan-09
Apr-09
Jul-09
Oct-09
Jan-10
Apr-10
Jul-10
Oct-10
Jan-11
Apr-11
Jul-11
GA
Moore's Law
GAII + PE Sequencing
GAIIx + 95 G
v3 Chem.
HiSeq
National Human Genome Research Institute
genome.gov/sequencingcosts

Genomics Tools Are Turning the Vision Into Reality

A Sequencer for Every Need. Every Budget. Every Lab.

Redefining the trajectory of sequencing.	Powerful. Flexible. Scalable.	Two proven technologies. One powerful platform.	The most widely cited platform, now at half the price	My Samples. My Study. MiSeq
HiSeq 2000/2500	**HiSeq 1000/1500**	**HiScanSQ**	**GA*IIx***	**MiSeq**

illumina®

MiSeq Applications Portfolio

Integrated. Optimized. Simplified.

Amplicon Sequencing

Custom Amplicon

Targeted Resequencing

Custom Enrichment

Small RNA sequencing

Clone checking

ChIP-Seq

Library QC

Plasmid

Regulation

RNA-Seq

Resequencing

Small genome

RNA sequencing

De novo sequencing

16S Metagenomics

illumina

MiSeq System

Proven Pedigree – Bench top Friendly

Single Instrument Workflow

illumina

MiSeq Instrument

Next-Gen Made Simple: Load & Go

DESIGNED FOR THE WAY YOU WANT TO WORK

Preloaded single use reagent cartridge

Positive consumables tracking

Auto flow cell positioning

Walkaway automation

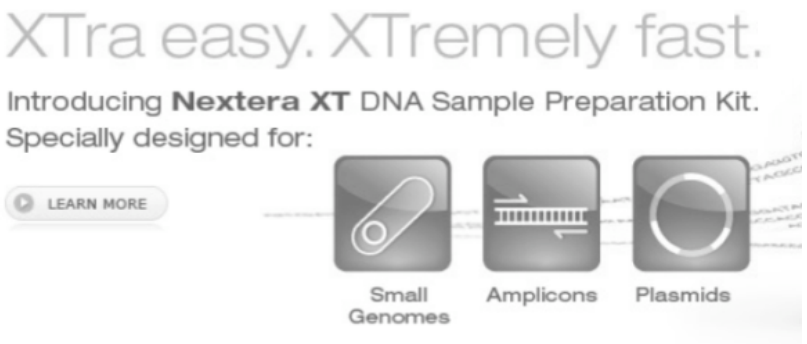
Go

illumina

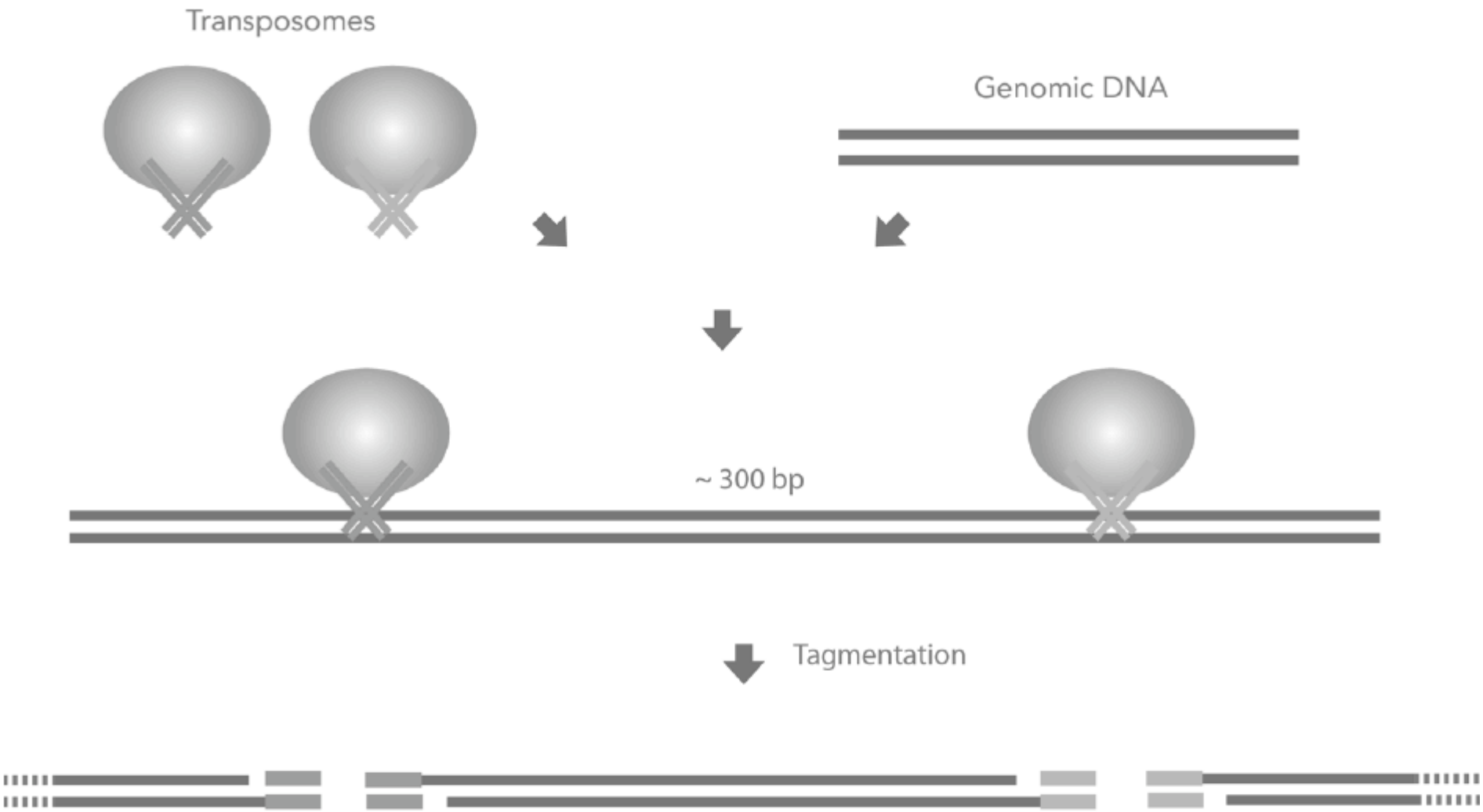
Nextera XT DNA Sample Prep

The fastest & easiest prep for small genomes, PCR amplicons and plasmids

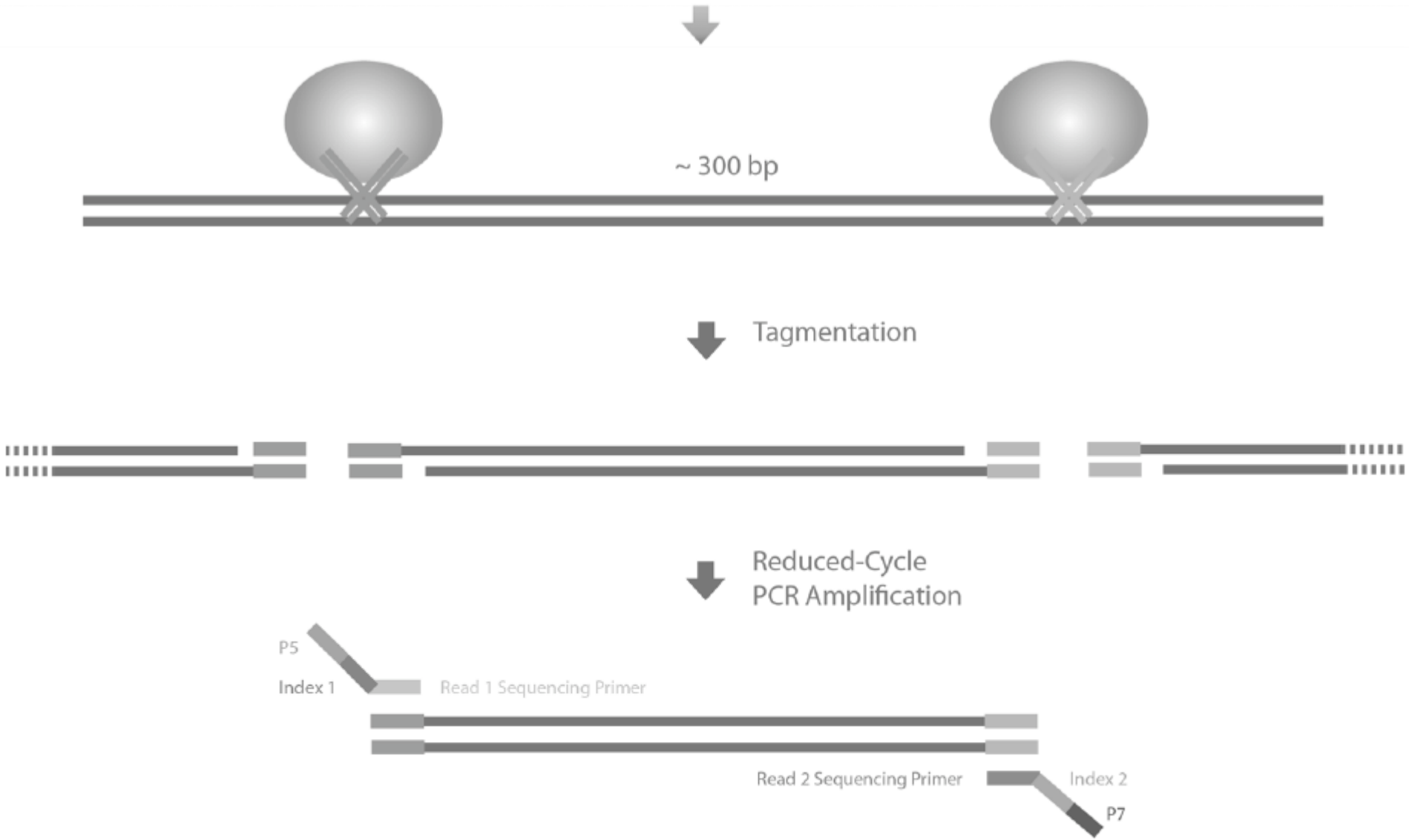
- Rapid Prep
  - 90 min prep, only 15 min of hands on time
- Optimized for small genomes, PCR amplicons and plasmids
- Innovative sample normalization
  - No library quantification needed
- Fastest time to results
  - DNA to analyzed data in <8 hours with MiSeq
- Ultra low input
  - only a single nanogram of input DNA needed



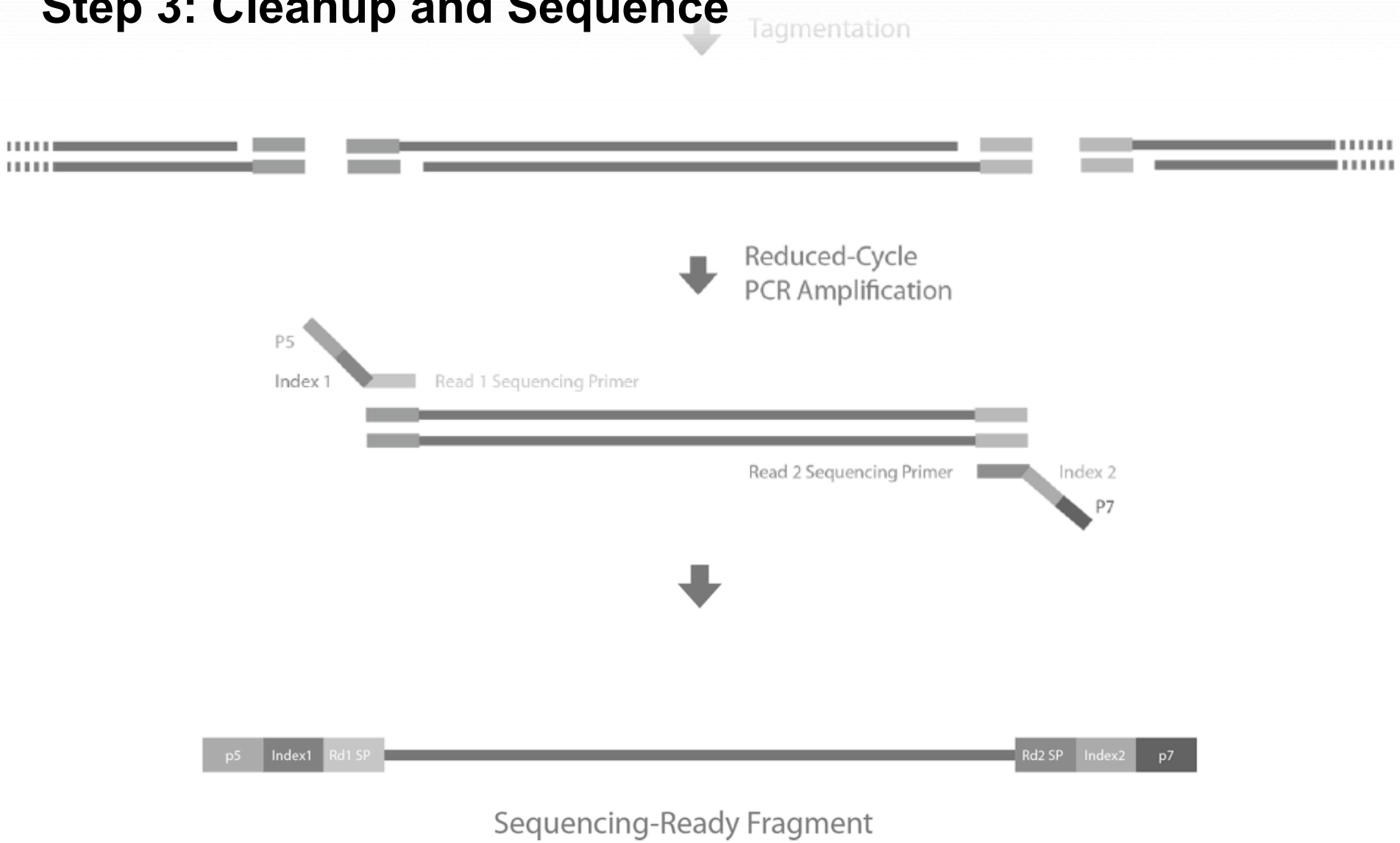
Step 1: Tagmentation of template DNA



Step 2: PCR to add adapters and indices



Step 3: Cleanup and Sequence

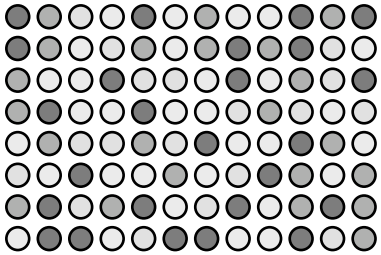


illumina

Sample Normalization is included

No library quantification or qPCR is required – go straight to MiSeq!

Completed libraries, range of yields



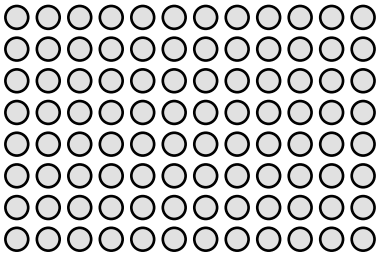
Quant in triplicate with qPCR → **Calculate dilutions** → **Manually dilute and pool** →

Index CV for 20-sample pools

Pool A	Pool B
15.8%	18.2%

↓

Bead-based Normalization: Bind, Wash, Elute →



→ *5 µl of each desired library*

→

Pool A	Pool B
13.5%	**15.5%**

Nextera XT sample pooling is as simple as pipetting 5 µl!

illumina®

Amplicon Workflow for Nextera XT

Sample Prep	Run Set-up	Sequence	Data Analysis

Feature
- New Nextera XT kit

Benefits
- Fastest sample prep
- Fewest hands-on steps and minimal hands on time

Feature
- New Nextera XT supported workflow
- Automated manifest file generation

Benefits
- Minimal informatics skills to setup analysis
- Reduce errors in run setup

Feature
- All-in-one sequence and analysis instrument
- Wizard driven UI
- Directed connected to BaseSpace

Benefits
- No additional amplification/ computer hardware needed
- Easy to use
- Automated bioinformatics

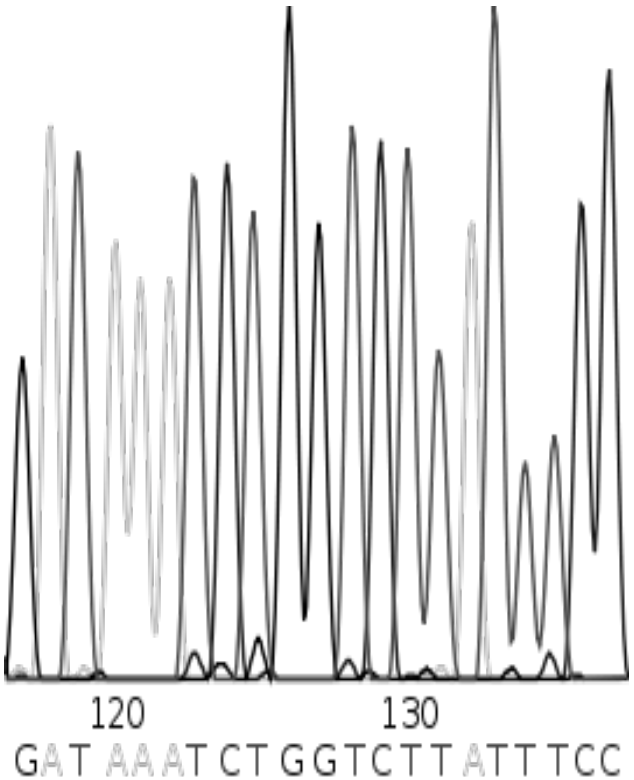
Feature
- New Nextera XT PCR amplicon bioinformatics analysis workflow
- Using BWA/GATK

Benefits
- Automated data analysis with min bioinformatics skills
- Easy to read, biologically relevant reports
- Industry standard algorithms

illumina®

CE for Amplicons? Time to try NGS!

Nextera XT enables easy transition to MiSeq

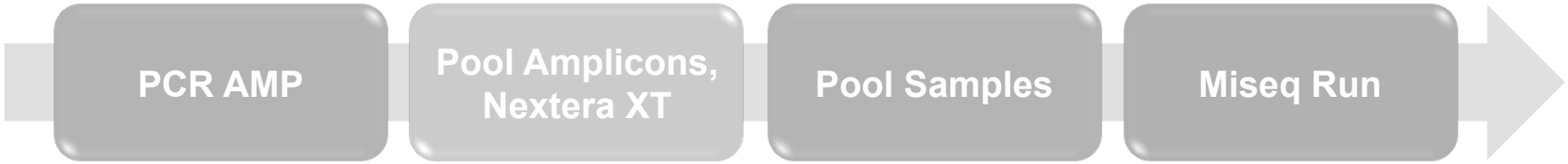


- **FAST**
  - Prep a sample in <90 min with 15 min hands on time
  - Rapid way to add sequencing adapters for ILMN sequencers
- **EASY**
  - Prep existing amplicons, including long range >1 kb!
  - Simple sample normalization for easy pooling
  - Barcoding supports pooling up to 96 samples per lane
- **ENABLING PRICE**
  - Single Nextera XT prep @ ~$30 for a pool of many amplicons

illumina®

Nextera for PCR Amplicons

Quick workflow for long or short range PCR



Short Range - *CFTR*

7 Exons 255-416 bp range	90 min total 15 min hands on	12 patient samples	2x36 (6.5hrs) 95%>Q30

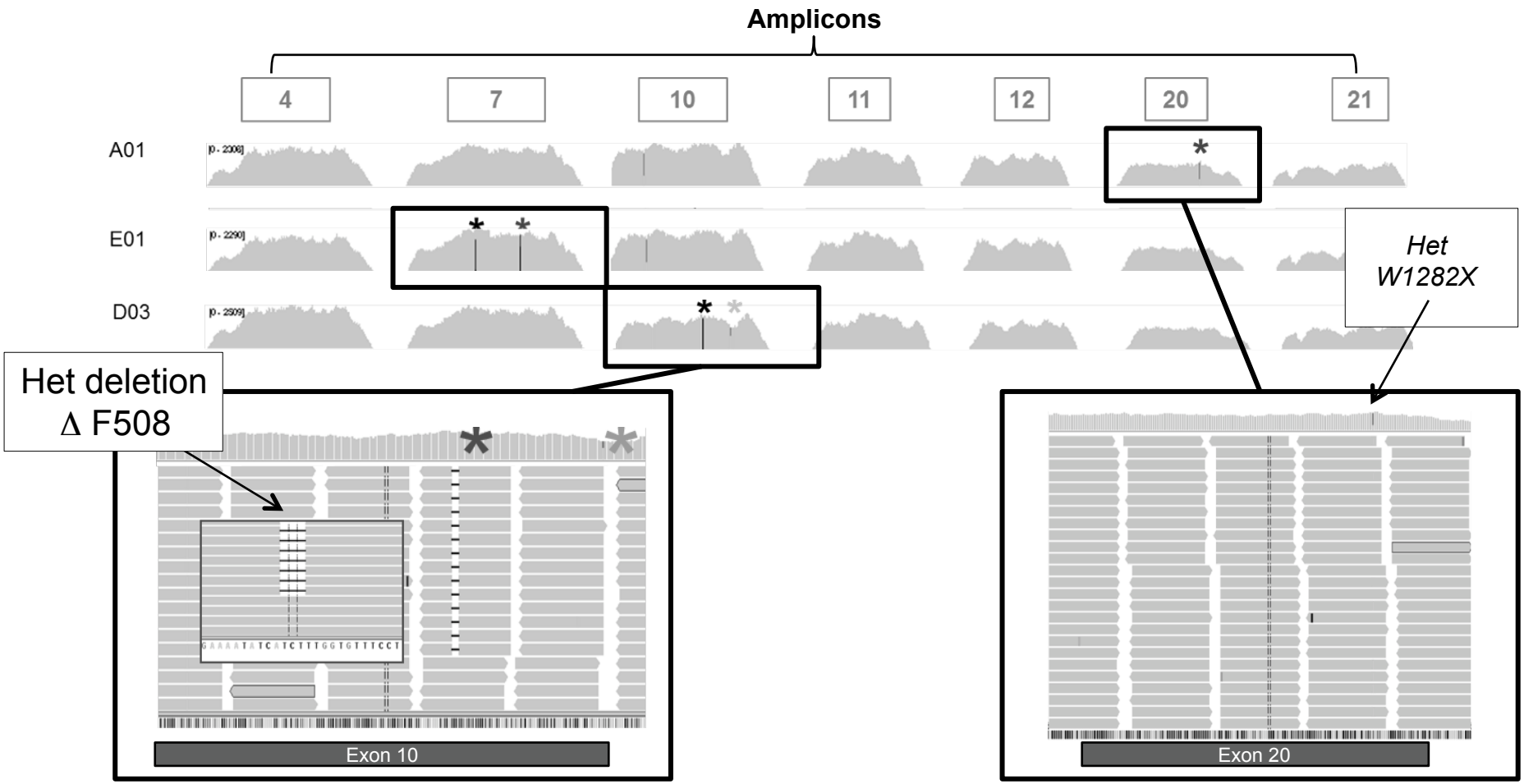
Long Range – *BRCA1/2*

22 Amplicons 1-6kb range 68.5kb total	90 min total 15 min hands on	5 patient samples	2x76(15hrs) 95%>Q30

illumina®

Nextera for PCR Amplicons

*Detection of mutations in CFTR with **short range** PCR*



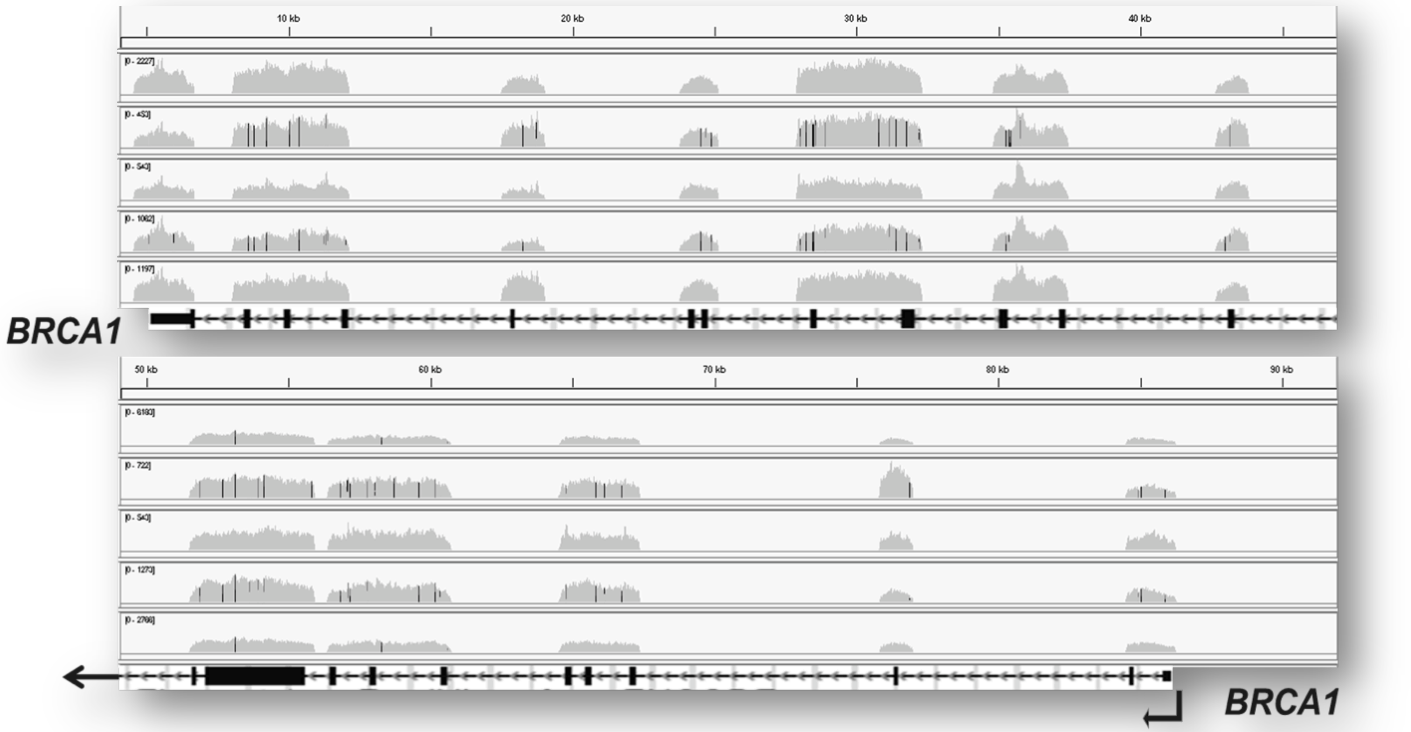
Complete coverage of target regions at >300X depth
Complete variant concordance

Collaboration with Steven Abbs and Michale Yau, London

illumina

Nextera for PCR Amplicons

Detection of mutations in BRCA1 & 2 *with* ***long range*** *PCR*



Complete coverage of target regions at >50x depth
Complete variant concordance

Collaboration with Graham Taylor, Leeds, UK

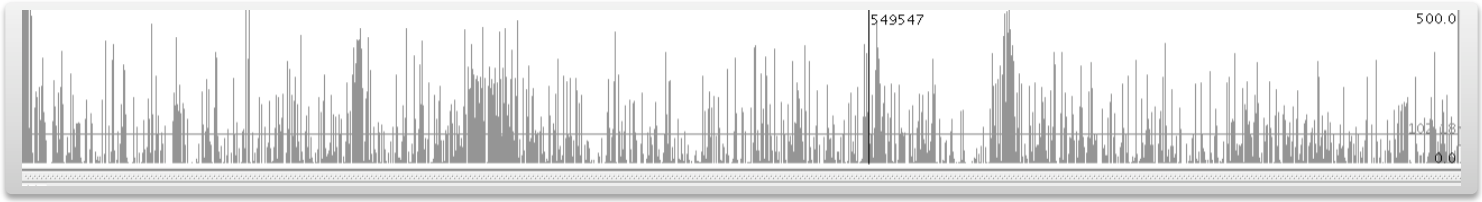
Nextera XT is perfect for microbial genomes

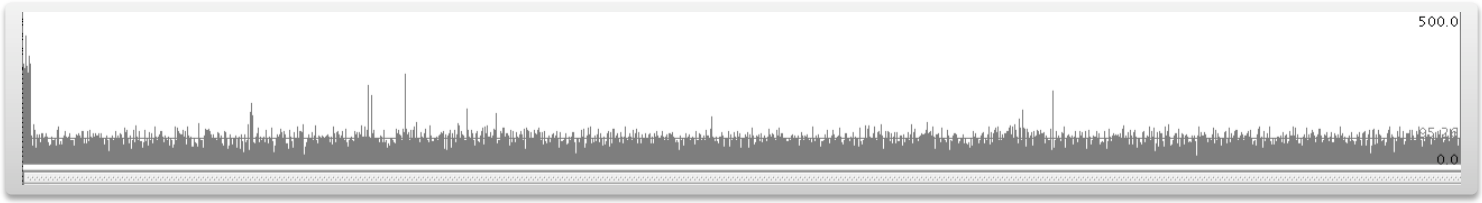
- **FAST**
  - Prep a sample in <90 min with 15 min hands on time
  - Sequence genomes in a single day with MiSeq
- **EASY**
  - Simple prep makes it easy to prepare many samples at the same time
  - Simple sample normalization for easy pooling
  - Barcoding supports pooling up to 96 samples per lane
- **ENABLING PRICE**
  - Sequence a bacterial genome for **<$100** with XT and MiSeq!

Superior *de novo* bacterial sequencing with MiSeq

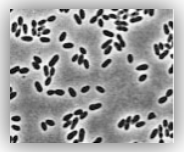
Even coverage across challenging high GC genomes

PGM exhibits poor coverage uniformity





Nextera and MiSeq provide even coverage across entire genome



Rhodobacter sphaeroides
4.6 Mb genome, 70% GC

Data courtesy of Dr. Tim Stinear, University of Melbourne, and Dr. Torsten Seemann, Monash University

illumina

Extending MiSeq performance

	MiSeq
Yield	1.5-2G
Read Length	2x150
Number of Reads	5 Million clusters PF
Run time	27 Hours for 2x150
Quality at 2x150bp	75% bases ≥Q30

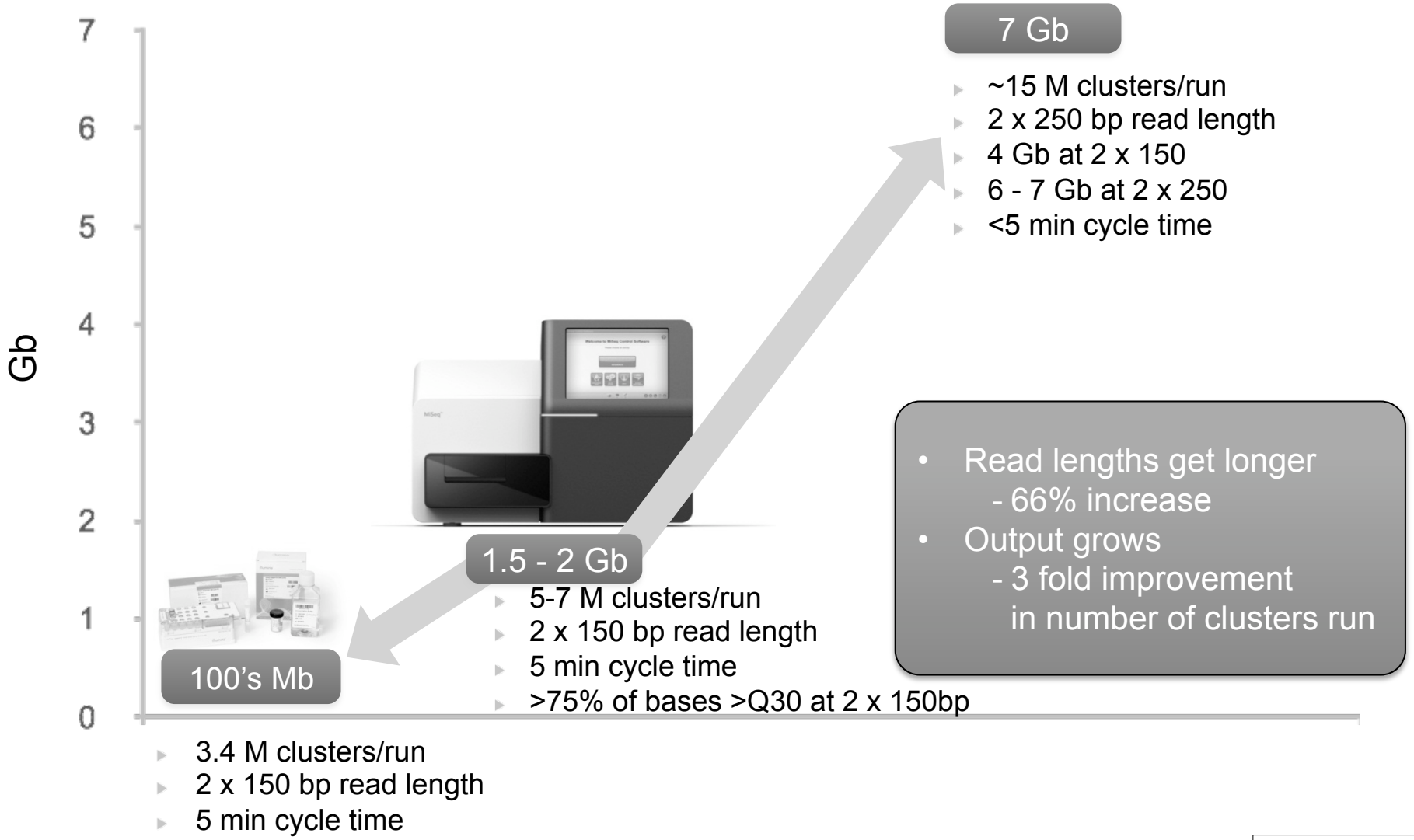
Improvements in Chemistry and Imaging

- Increased imaging area
  - Access second surface
- Improved SBS polymerase allowing faster kinetics
  - Chemistry cycle time of approx 2 mins (from 3.5mins)
  - Novel reagent formulation optimized for the MiSeq platform

illumina®

MiSeq Performance Improvements

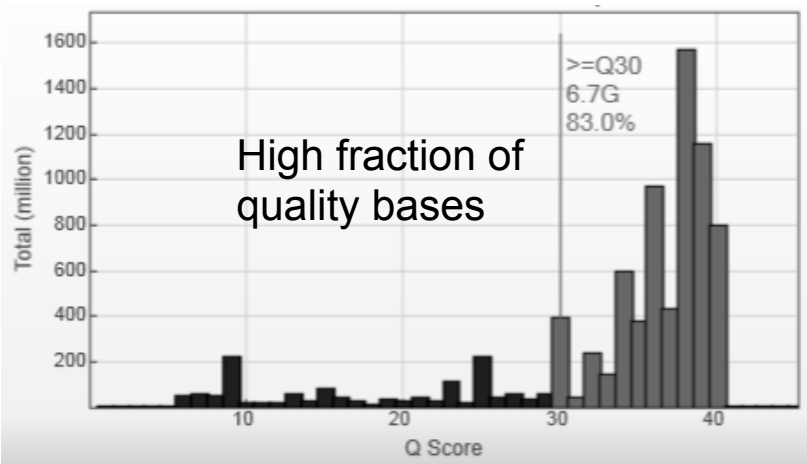
Next Six Months

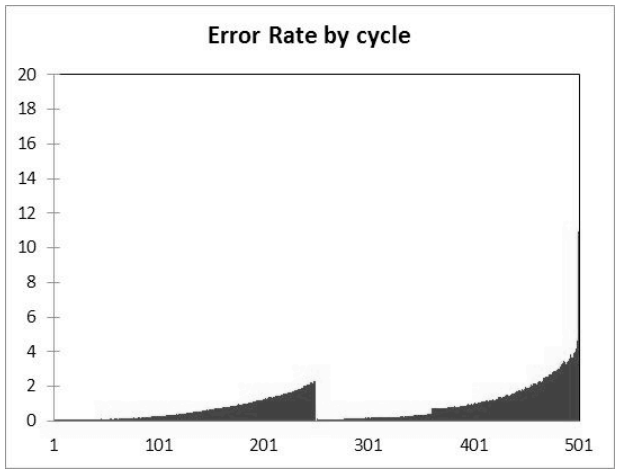


illumina

MiSeq data: Extending read lengths to 2x250bp

	2x250bp
Yield	8.1G
Number of Reads	16.1M PF
of which perfect	9.34M
Error rate (100 cycles)	0.14%, 0.19%
Error rate (250 cycles)	0.69%, 1.19%



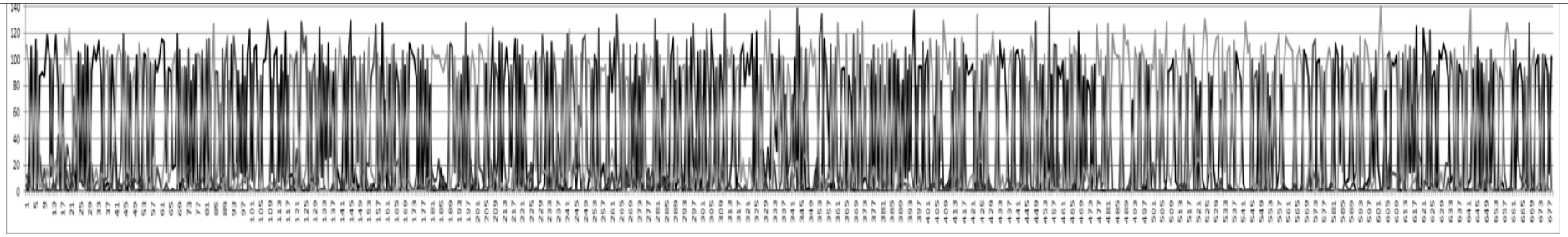


4.7G of perfect reads

Extending read length to 2x400 (in Research)

	2x400bp
Yield	3.4G
Number of Reads	3.8M PF
%≥Q30 (1x400; 2x400)	80.9%; 71.8%
Error rate at 100 cycles	0.29%; 0.61%
Error rate at 400 cycles	2%; 4.39%

Longest perfect overlapping read (so far): 678bp



500bp human library

illumina

Ion Torrent, episode #3, scalability, simplicity, and speed.

iontorrent | Subscribe | 93 videos

illumina

HiSeq2500
Combining innovation

MiSeq

HiSeq2500

HiSeq2000

- Clustering on-board
- Fast Chemistry
- Longer Reads

- 120 Gb runs in 27hr
- Clustering on-board
- Complete walk away workflow
- Longer 2x150 reads

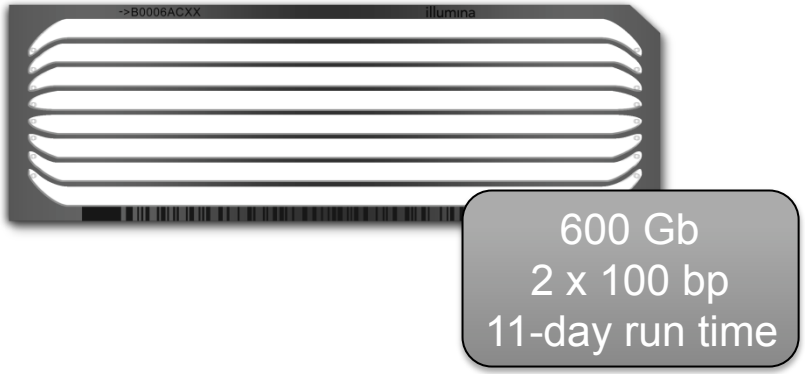
- Data rate
- TDI scanning
- Larger flow cell

HiSeq 2500 Sequencing

Standard run vs. rapid turnaround

Standard run configuration:

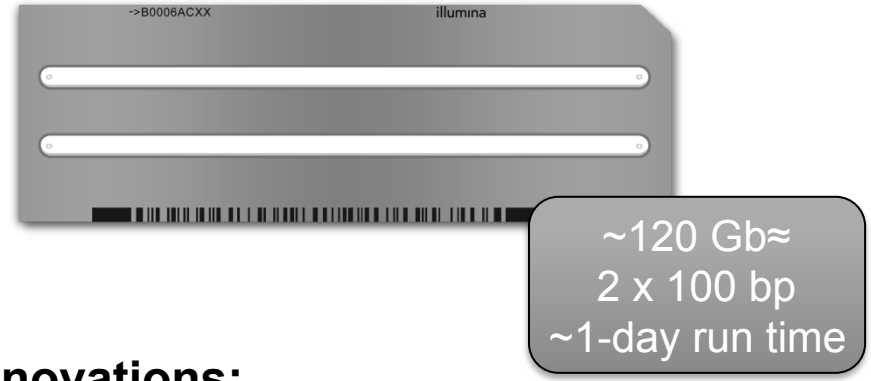
- HiSeq Cluster and SBS v3 reagents
- HiSeq Flow Cell v3 (8-channels)



- cBot clustering required

Rapid run configuration:

- *New* Cluster and SBS reagents based on MiSeq chemistry
- *New* HiSeq Flow Cell (2-channels)

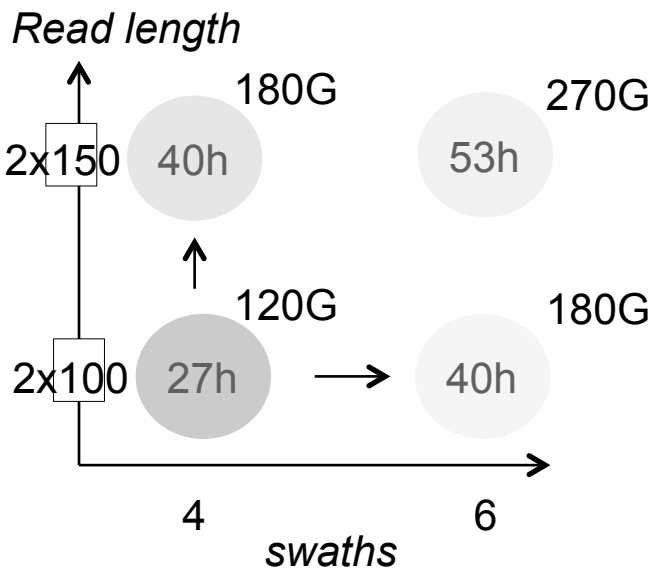


Innovations:

- Faster imaging / on-instrument clustering
- Longer read (2 x 150 bp supported)

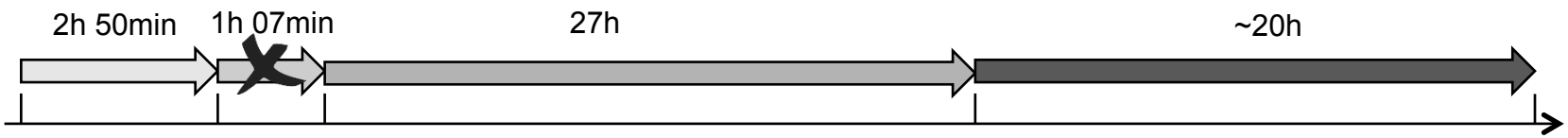
Product configurations and extensibility

Metric	HiSeq 2000	HiSeq 2500	HiSeq 2500
Mode	High Yield	As HiSeq2000 *plus*	And Short read
Read length	2x101	2x101	2x51
Yield	600G	120G	30G
Run time	11 days	27h	8.5h
Raw density (k/mm^2)	750-850	850-1000	850-1000
Reads (M)	1000	200	200
% bases ≥ Q30	>80%	same	>90%



100ng fast workflow (R&D)

Sample Prep → Sequencing → Analysis



1ug (Qubit)
Diluted 1:10

↓

100ng Starting material
'single shot' library

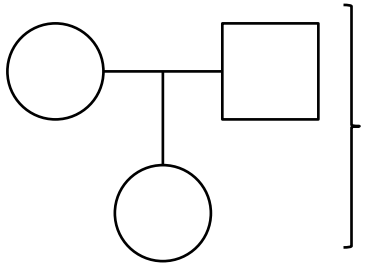
Load it all

HiSeq2500

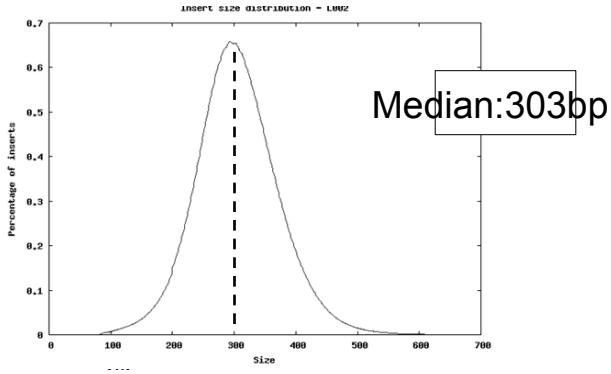
Annotated variants

Trio sample prep data

YRI trio



All prepped from
100ng input DNA



	Median (bp)	Diversity
NA18507	307	10Bn
NA18506	294	12Bn
NA18508	303	11Bn

Similar sample preps with good gaps size distribution and high diversity

illumina®

Trio analysis

	Gb	% ≥Q30	% align
NA18507	140	93	89
NA18508	162	93.5	90
NA18506	182	92	89

HiSeq 2500

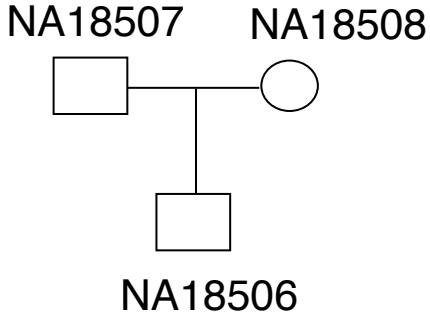
6 swaths of data/lane
2x100bp runs

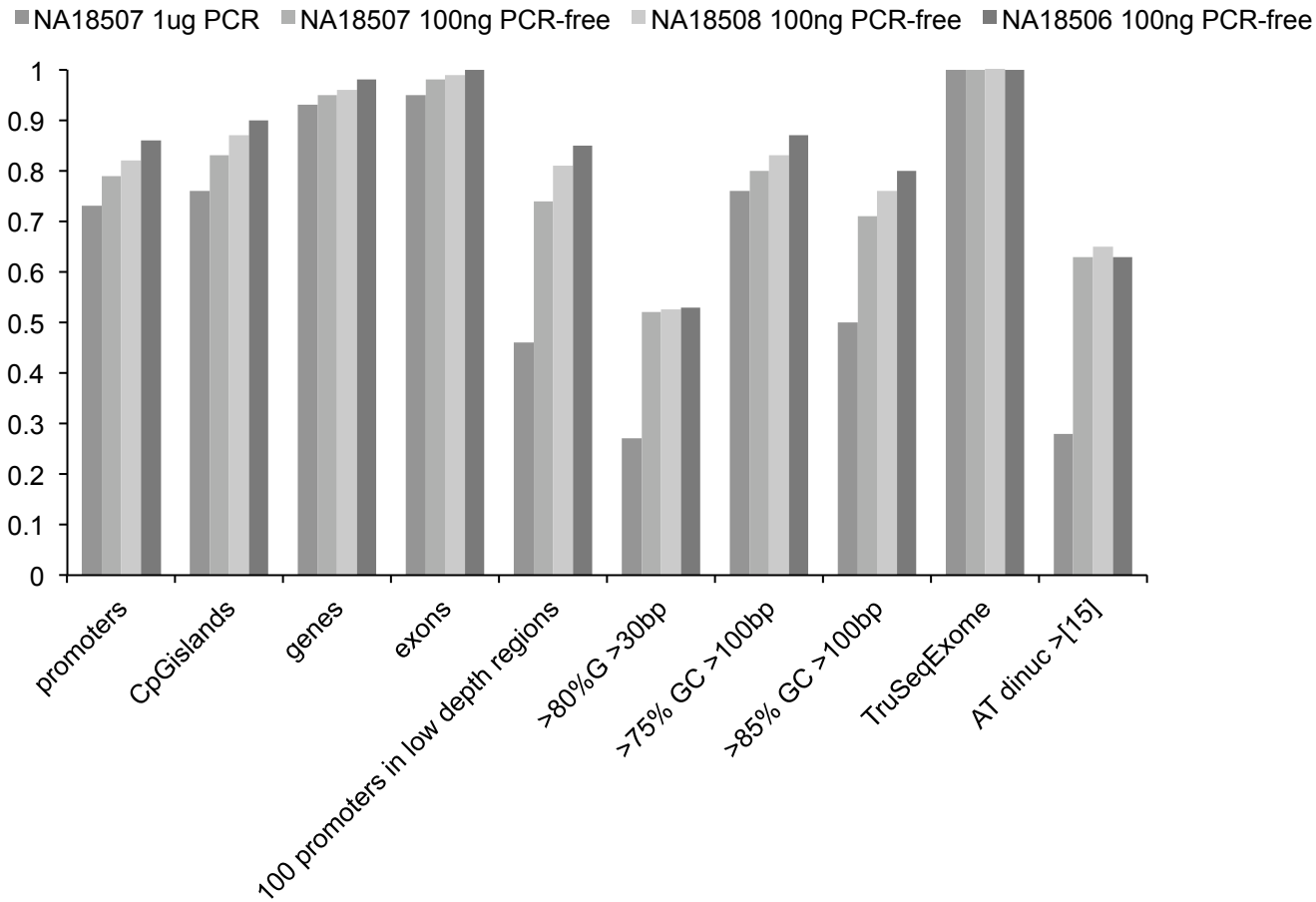
High depth and good concordance for all preps

Sample	NA18507	NA18508	NA18506
Depth	40.5x	49.2x	53.0x
TotalSNPs	4.54M	4.63M	4.57M
Genotype Coverage	99.6%	99.72%	99.76%
Genotype Concordance	99.41%	99.45%	99.45%

Y009 Trio Build Coverage

1ug PCR and 100ng PCR-Free





- Equivalent or better for low input preps on HiSeq2500

illumina

Informatics "In a Day"

Cloud support of HiSeq 2500

- BaseSpace cloud storage and computing for HiSeq data
  - Whole genome applications supported through build-out of BaseSpace 'appstore'
  - Minimized transfer time to the cloud through compression and serialization technologies
  - No need for customer to manage software installation, upgrades, and versioning
- Real-time alignment and variant calling (like MiSeq but for HiSeq)
  - Variant calling finished automatically upon run completion
- Near-zero incremental investment in computing infrastructure

illumina®

Complete solution

MiSeq

7 Gb*
2 x 250 bp
> 35 hrs

Few samples
Very fast
Longest reads

HiSeq2500

120 Gb**
2 x 150 bp
27 hrs

More samples
Very fast
Long reads

HiSeq2000

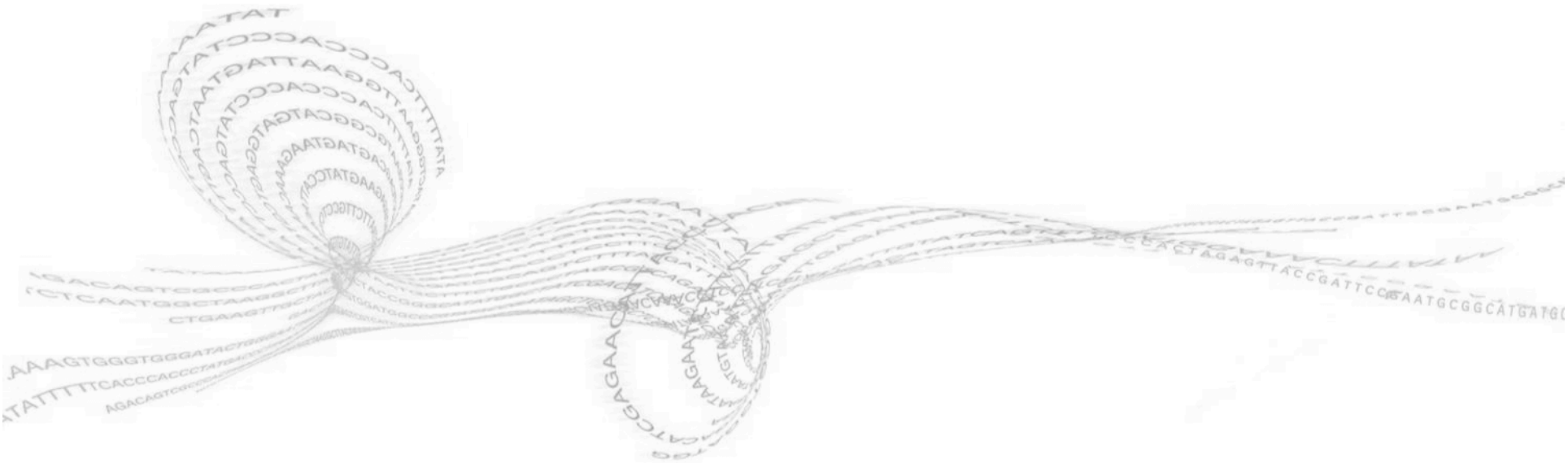
600 Gb
2 x 100 bp
9-11 days

Many samples
Large genomes
Lowest cost per base

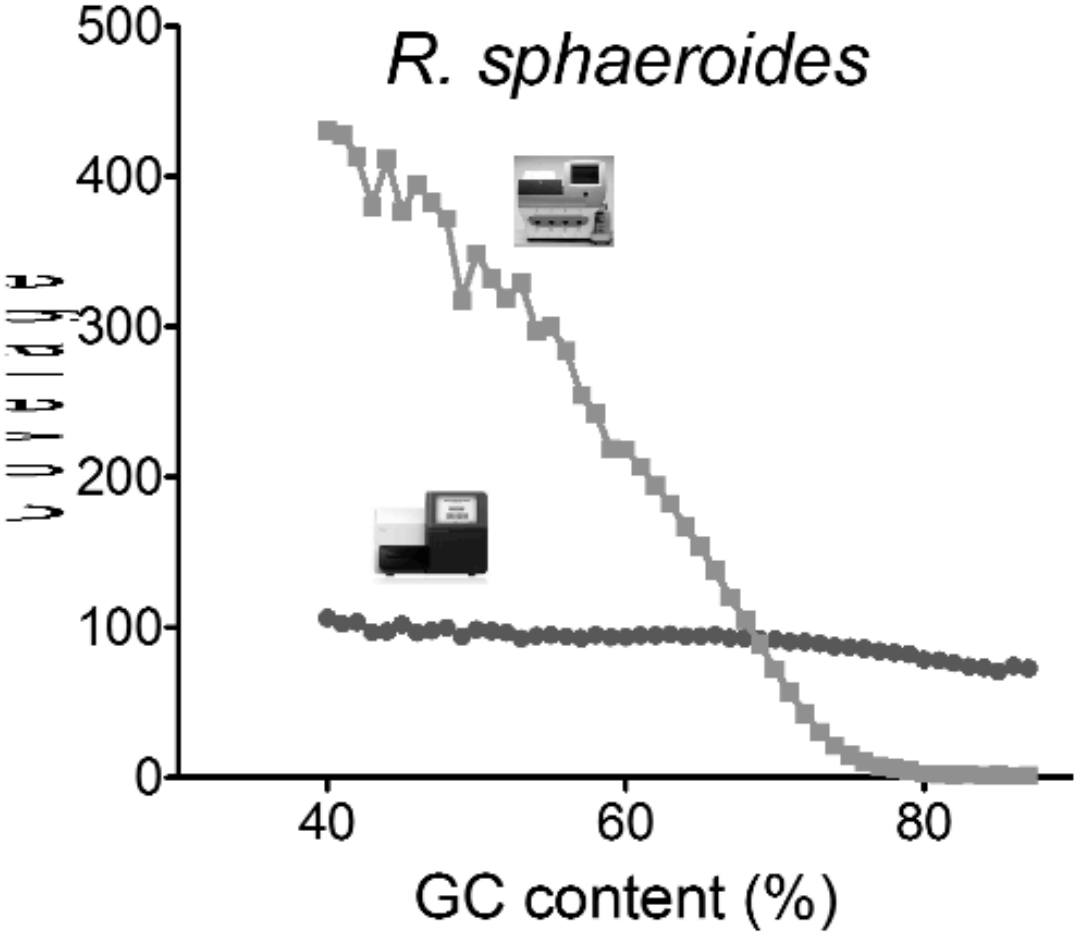
***De novo* Assembly** **Metagenomics** **RNA-Seq** **Amplicon Sequencing**
16S rRNA **Genotyping-by-Sequencing** **Whole Genome Sequencing**
Resequencing **Small RNA Discovery** **Library QC**

* Commercially available the 3rd Quarter of 2012 ** Commercially available the 4th Quarter of 2012

illumina®



Questions?

illumina®

Coverage vs %GC