Overview of Next Generation Sequencing (NGS) Technologies

Vivien G. Dugan
Office of Genomics and Advanced Technologies
NIAID/NIH

Timothy Stockwell
J. Craig Venter Institute

August 26th, 2013

National Institute of Allergy and Infectious Diseases

NIAID Genomic Sequencing Centers for Infectious Diseases

J. Craig Venter™ INSTITUTE

UNIVERSITY of MARYLAND SCHOOL OF MEDICINE INSTITUTE FOR GENOME SCIENCES

BROAD INSTITUTE

- Sample Processing Method Develop
- High Throughput Sequencing Pipelines
- Metagenomics Transcriptomics
- Bioinformatics Tools Data Analysis Pipelines
- Genomics Bioinformatics Training

What is ‘NextGen’ sequencing?

- Different chemistry from Sanger
- Sequences **everything** in a sample
  - Host, pathogen, cells, etc.
- Sequences clonally amplified molecule

- Sequencing occurs in parallel
  - Millions of sequences produced concurrently
- Gigabytes of sequences

NIH National Institute of Allergy and Infectious Diseases

What is ‘NextGen’ sequencing?

- Less time than Sanger
- Large capacity
- Multiplexing, variation detection, gene expression, metagenomics
- Address various biological questions

Sanger vs Next-generation sequencing

100 of these….

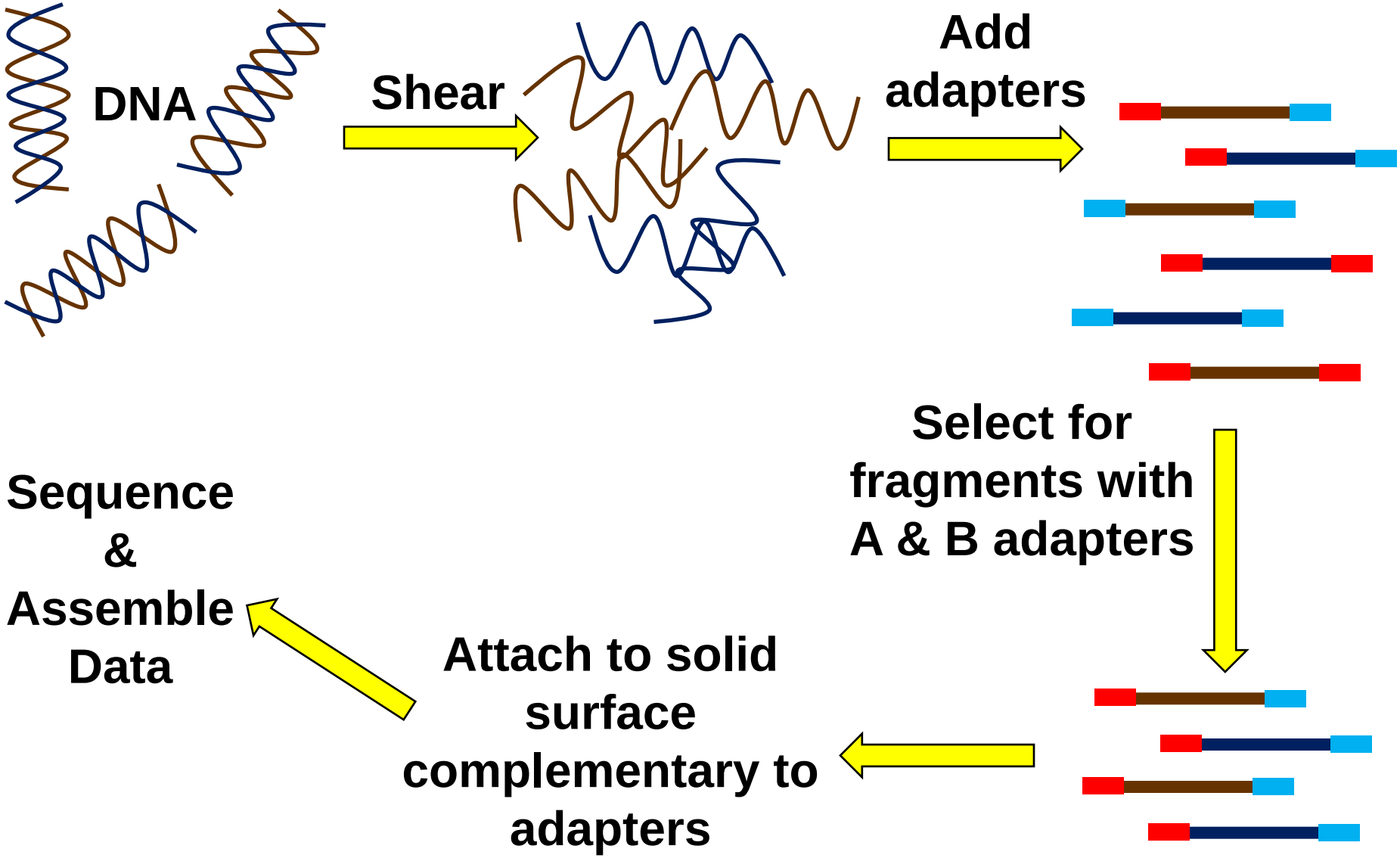
ABI 3730x

= 1 of these….

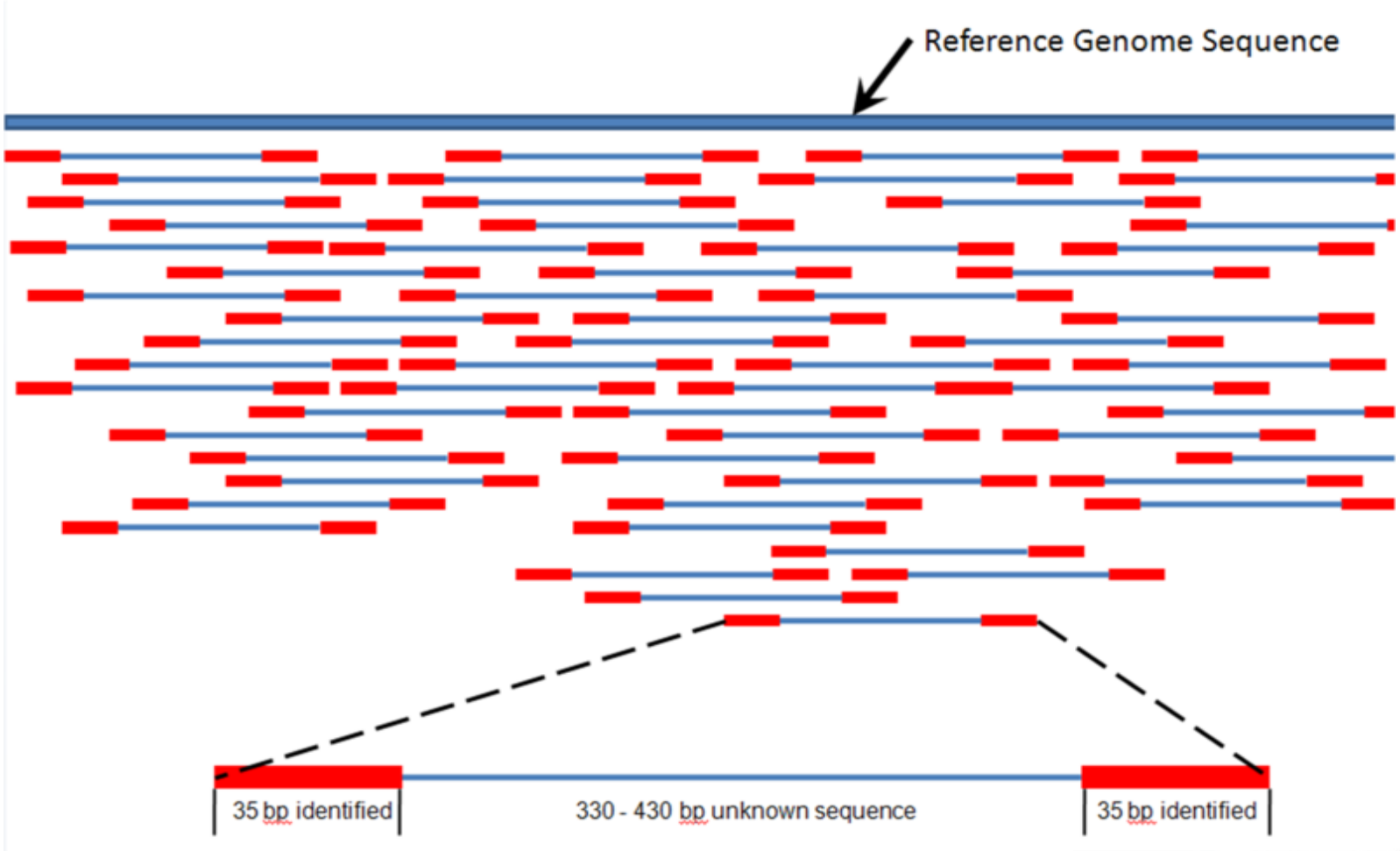
GS-FLX Roche/454

NIH National Institute of Allergy and Infectious Diseases

“Single Molecule Sequencing”



Mapping sequence reads to reference

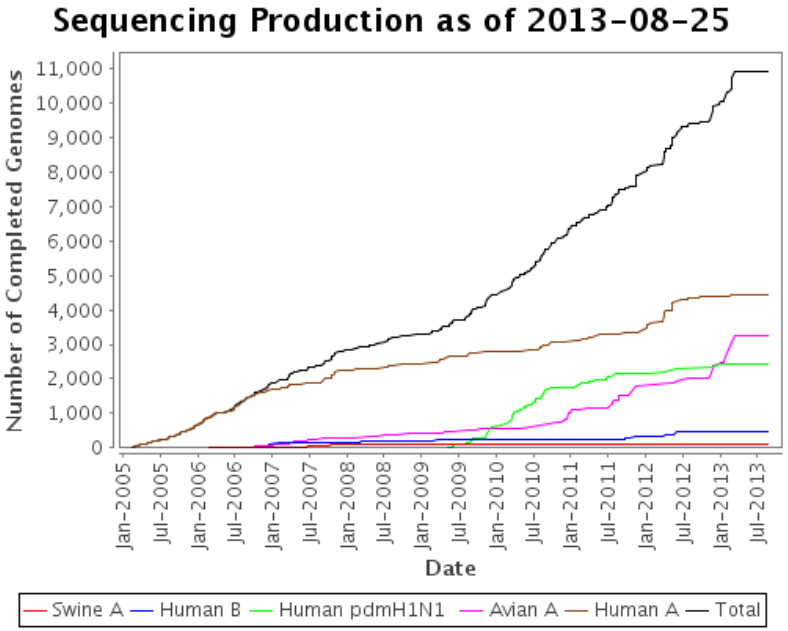


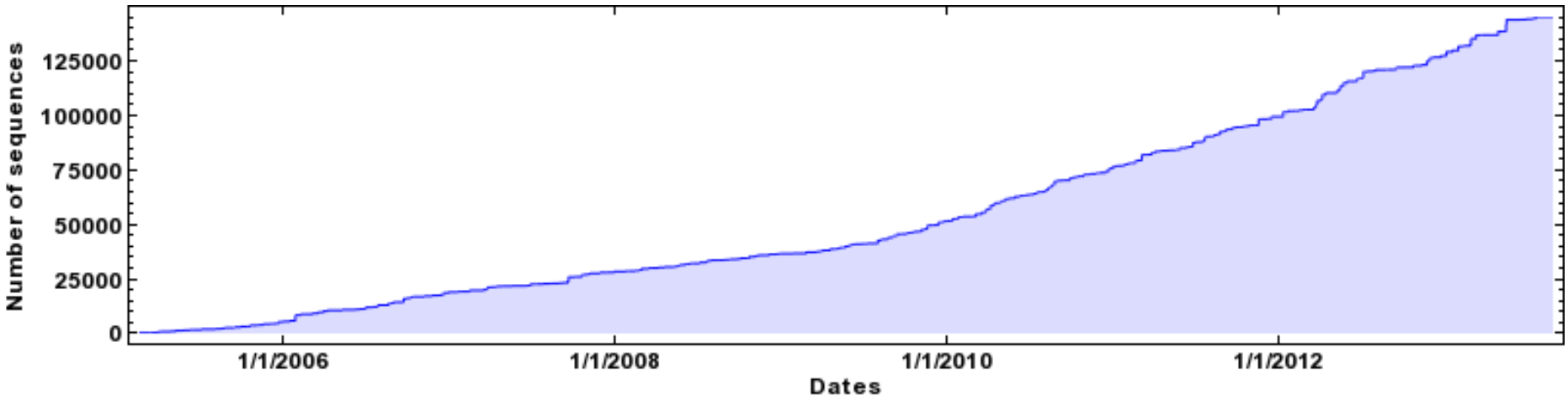
Why use NextGen?

- High rates of accuracy
- Many reads per sequencing run
- Faster time per sequencing run
- Multiplexing capabilities
- Decreased cost
- Useful for many different applications

Why use NextGen?

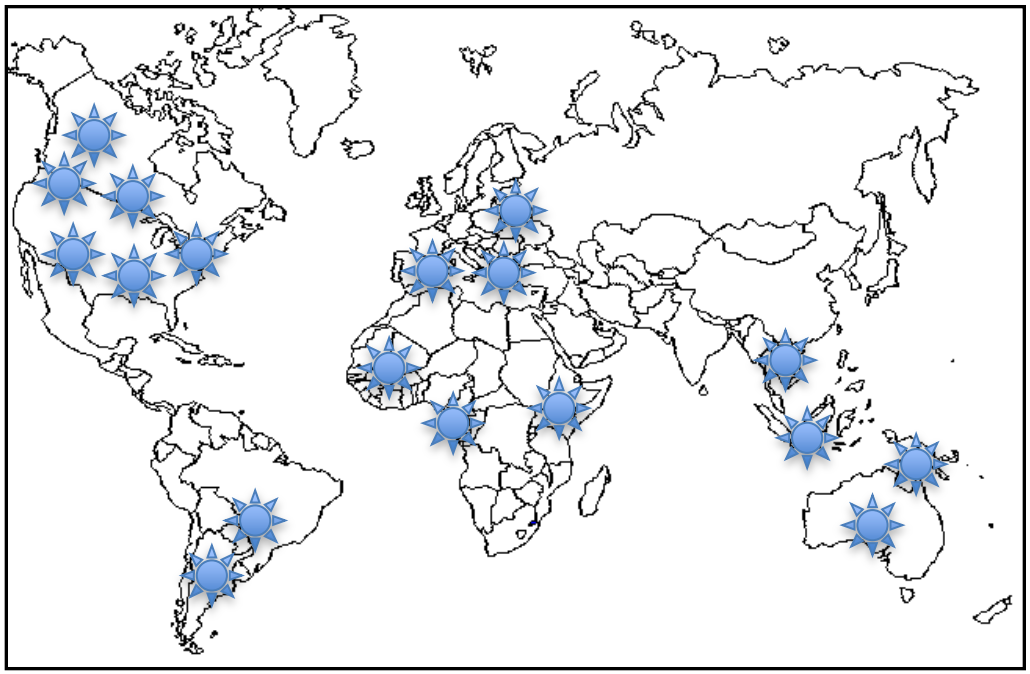
- 2004: < 100 influenza genomes in NCBI
- 2013: 14,000+ influenza genomes in NCBI



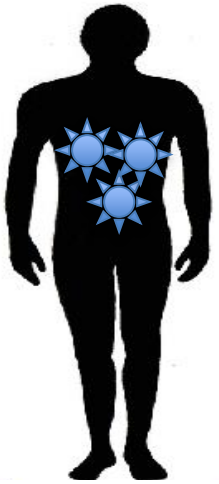


Genomics Analysis at the Population Level

Molecular Epidemiology

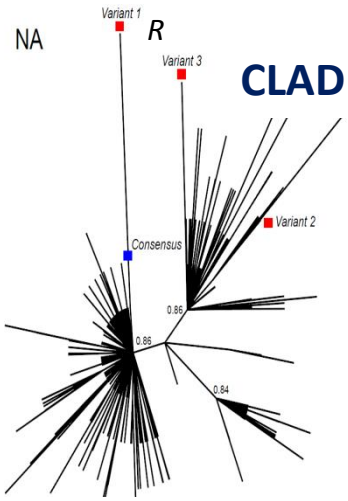


Consensus sequencing



Diversity

Deep sequencing



CLADE 2

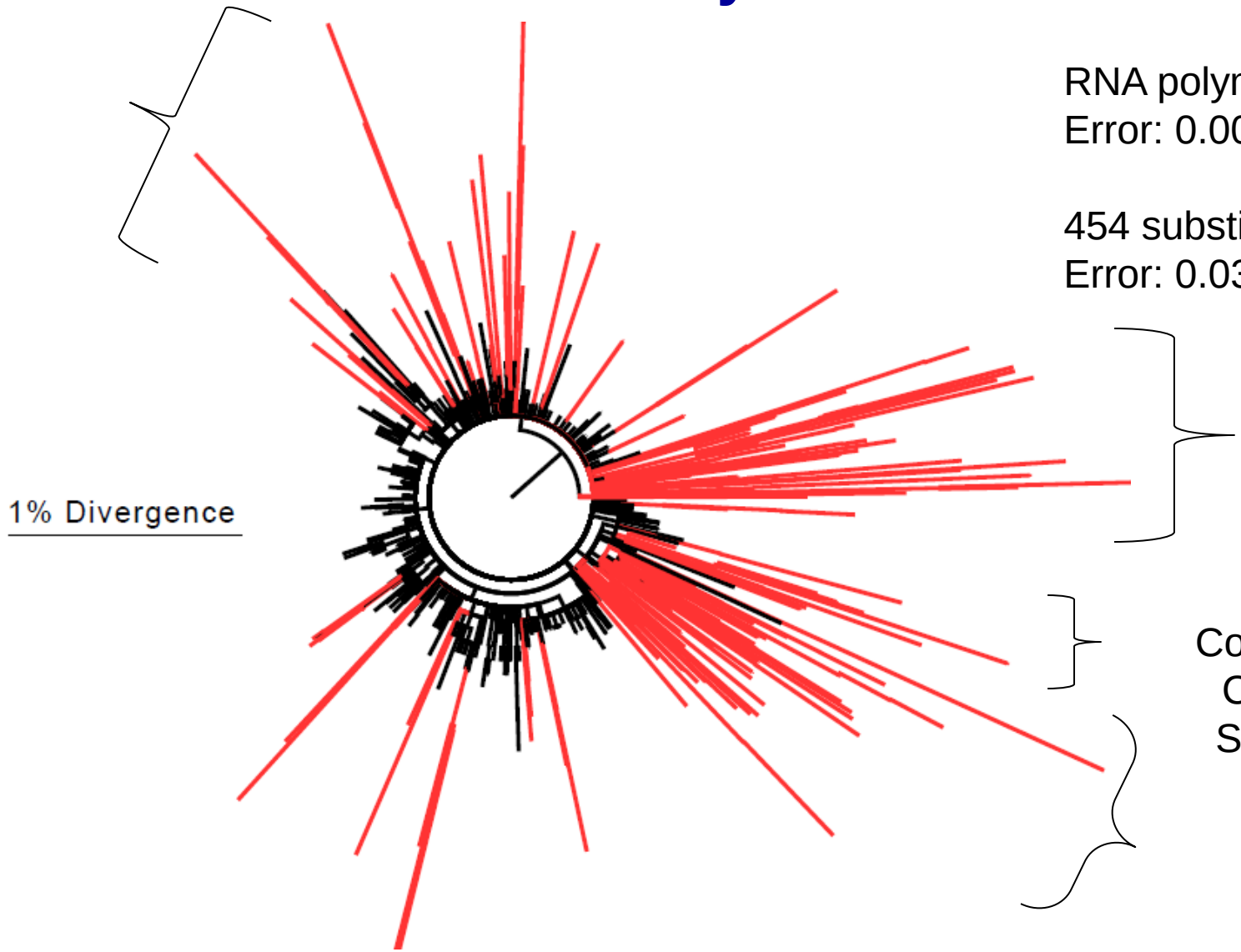
CLADE 7

Elodie Ghedin
Center for Vaccine Research
Dept. Computational & Systems Biology

NGS: Things to Consider

- Each platform has advantages & disadvantages
  - Read length, accuracy, reads per run, time, sequencing error rates
- Biology of the pathogen of interest
- What is your goal in sequencing?
  - Complete genome
  - Specific region or gene

True Diversity or Error?



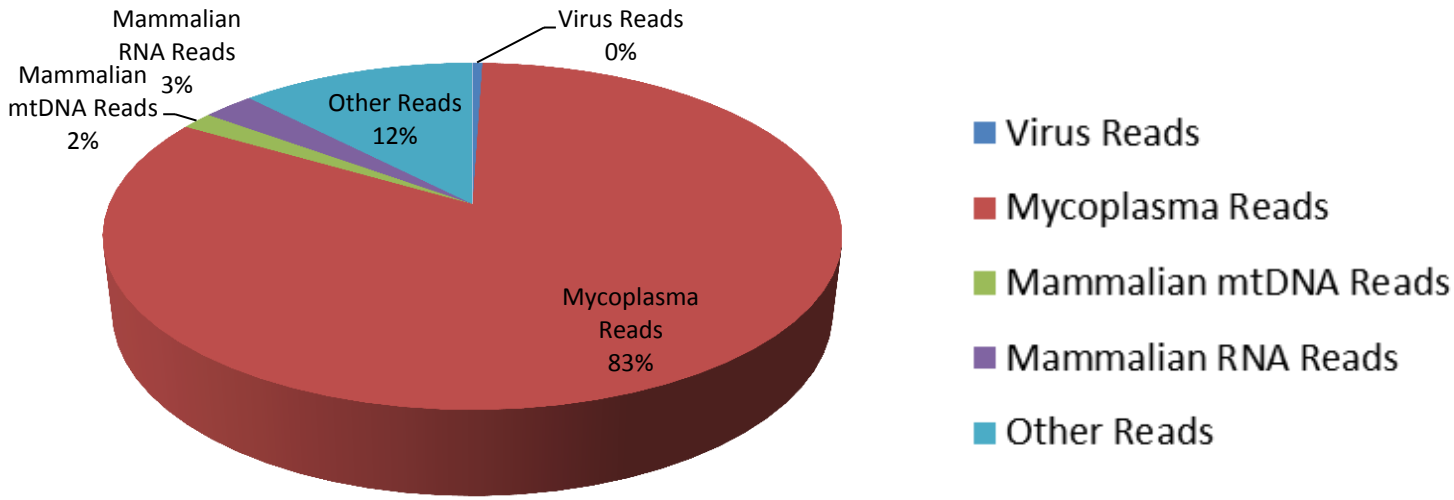
RNA polymerase
Error: 0.001%

454 substitution
Error: 0.03%

Consensus of Clusters to Smooth out errors

NGS: Things to Consider

- Sample preparation is important
  - Sequencing everything in the sample



National Institute of Allergy and Infectious Diseases

Summary

- Next Generation sequencing provides increasingly vital information not previously available
- NGS technologies becoming more commonly used in the field of infectious disease research
- Sequencing technologies, assembly and analyses tools rapidly improving

NGS Criteria to Consider

- Ultimate goal
- Sequencing platform(s)
  - Coverage level/depth
  - Read length
  - Error rates
- Sample preparation
- Confirmatory sequencing

Overview of Next Generation Sequencing (NGS) Technologies

Timothy Stockwell (JCVI)
Vivien Dugan (NIAID/NIH)

J. Craig Venter™
INSTITUTE

Outline

- Some history of DNA sequencing
- Overview of NextGen Sequencing Technologies at JCVI
- Roche/454 Pyrosequencing
- LifeTechnologies/IonTorrent Semiconductor Sequencing
- Illumina/Solexa Sequencing By Synthesis (SBS)
- Other technologies

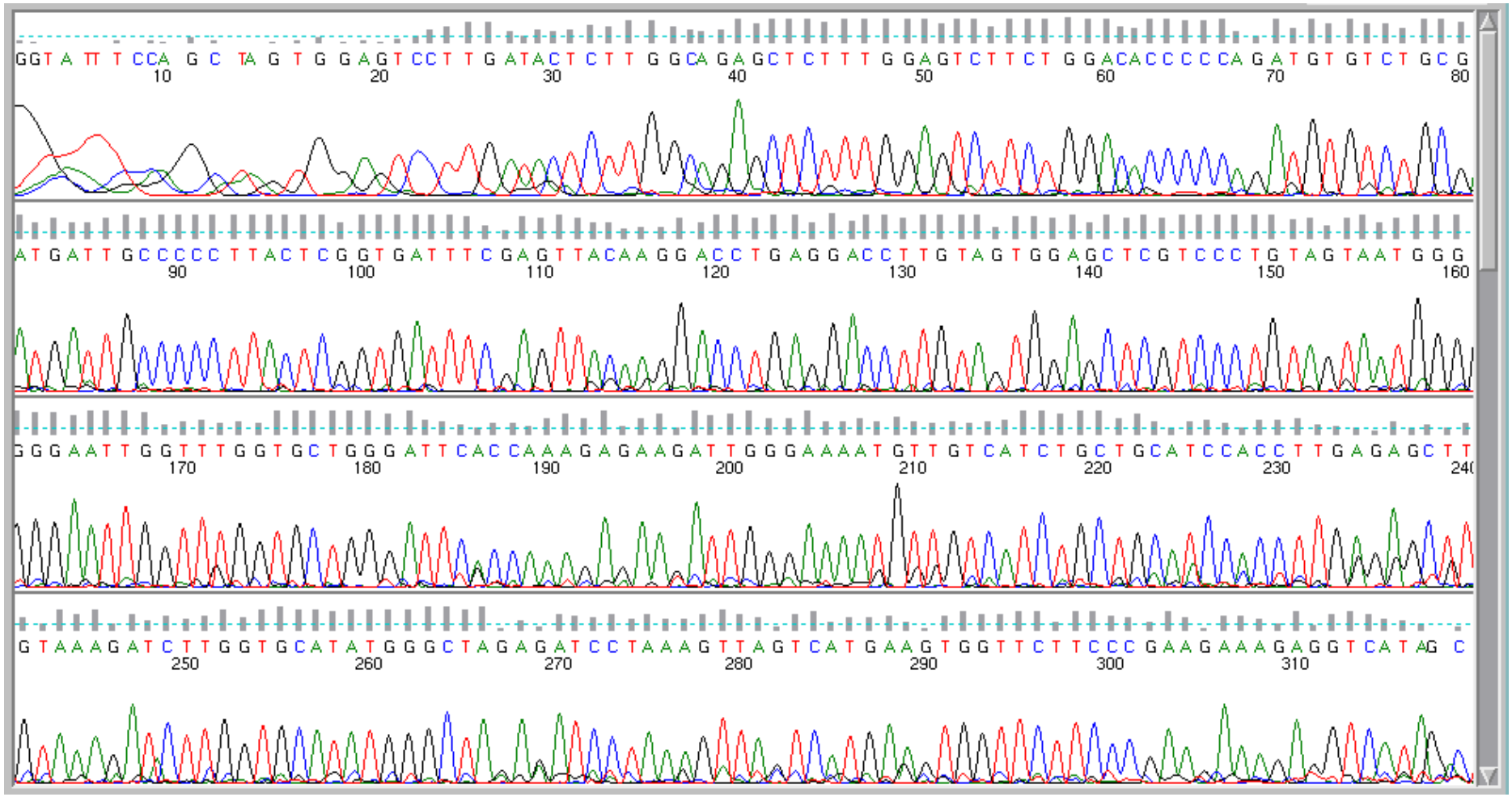
Review - Sanger Sequencing

- Randomly shear DNA, put it in a vector, and amplify with E. coli, or PCR amplify a region of a genome
- The Sanger sequencing reaction is like PCR, except there is only one primer, and in addition to regular nucleotides, there are also a small amount of dye labelled dideoxy nucleotides, with a distinct dye for each base
- As polymerase makes new ssDNA fragments, when a dye labelled dideoxy nucleotide is added, extension stops, and the fragment is labelled with a dye corresponding to the last base added.

Review - Sanger Sequencing

- Over many cycles, fragments of all the different lengths are formed, with each length fragment ending with the dye corresponding to the base at that position
- Capillary electrophoresis in polyacrylamide gel is used to separate the fragments by length and pass them by a laser and reader to interrogate the base at each position
- The result is a chromatogram, that is then “base called” using algorithms to output the most likely base at each position, usually with an indication of accuracy of the base call.

A chromatogram



J. Craig Venter INSTITUTE

Sanger Sequencing

- Think about the issues of scaling Sanger sequencing to obtain 1 million reads
- The E. coli clones or PCR reactions need separated wells – 2600 384-well plates
- To read the DNA from both ends, need double the number of wells, and have to keep track of mate pairs – 5200 384-well plates
- Also think about storage, pipet tips, labor required, etc.
- So then came along Next Generation Sequencing (NGS) Technologies

NextGen Sequencing Technologies

454 GS FLX

Illumina HiSeq 2000

Illumina MiSeq

Ion PGM

J. Craig Venter™
INSTITUTE

Sequencing Technologies in Use at JCVI

	Read length bp	Throughput /machine run	Run time	Throughput /day	Accuracy
ABI 3730xl	600-800	75,000bp	30-60 min	1-2 Mb	> QV 30
454	400-600	400 Mb	7 hr	800 Mb	QV 20
Illumina HiSeq	up to 100	up to 600 Gb	up to 12 days	50 Gb	~80% bases > QV30
Illumina MiSeq	up to 250	up to 8.5 Gb	up to 39 hours	5.2 Gb	~75% bases ➢QV30
Ion Torrent	~150	900 Mb	up to 4.5 hours		80%bases > QV20

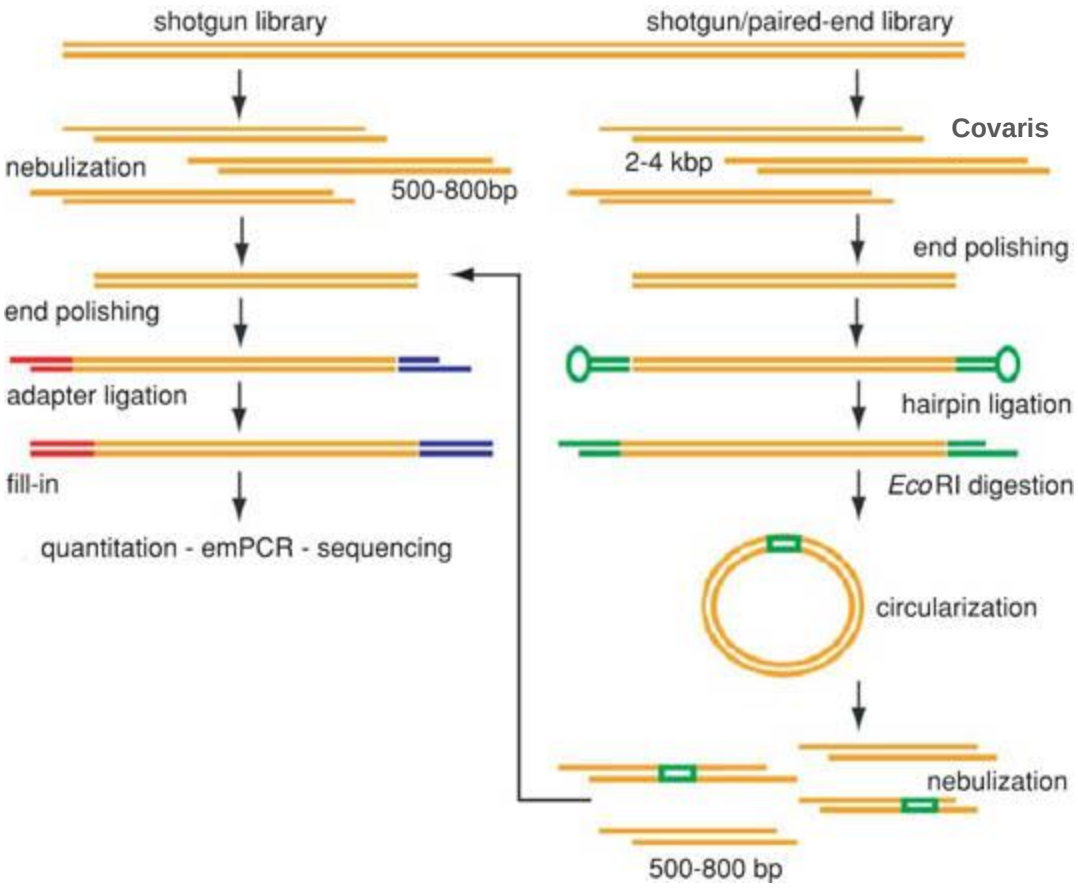
J. Craig Venter™ INSTITUTE

Roche 454 Sequencing

- Library Construction
- Sequencing Process Overview

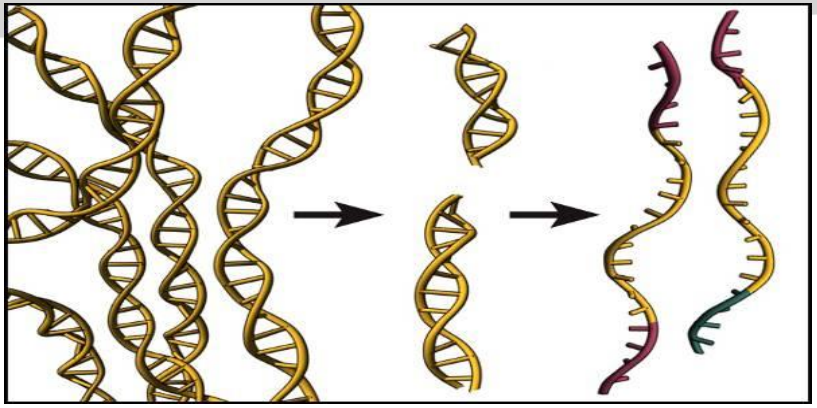
J. Craig Venter™ INSTITUTE

454 Library construction

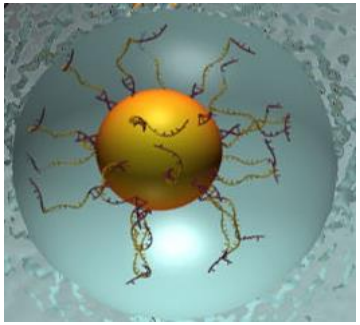


J. Craig Venter INSTITUTE

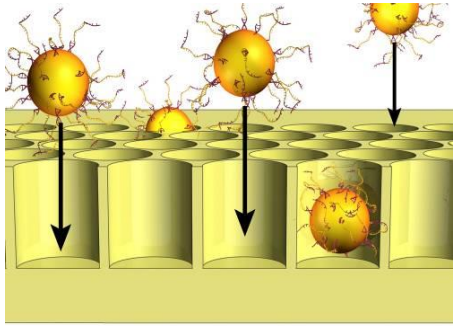
454 Massively Parallel Pyrosequencing
Process Overview



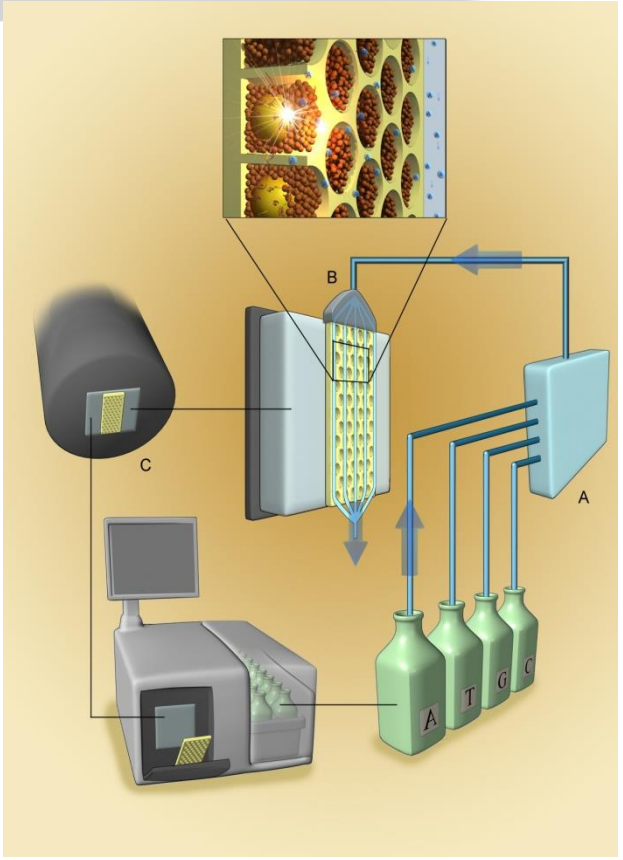
1) ssDNA library preparation



2) emPCR amplification



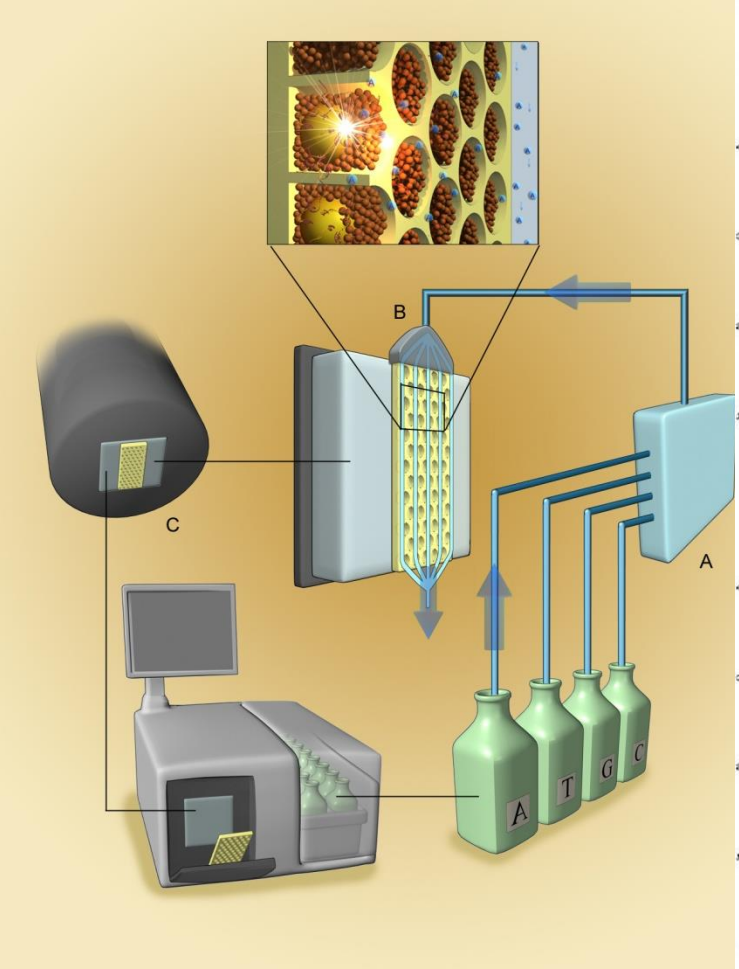
3) Load beads & enzymes in PicoTiter Plate™

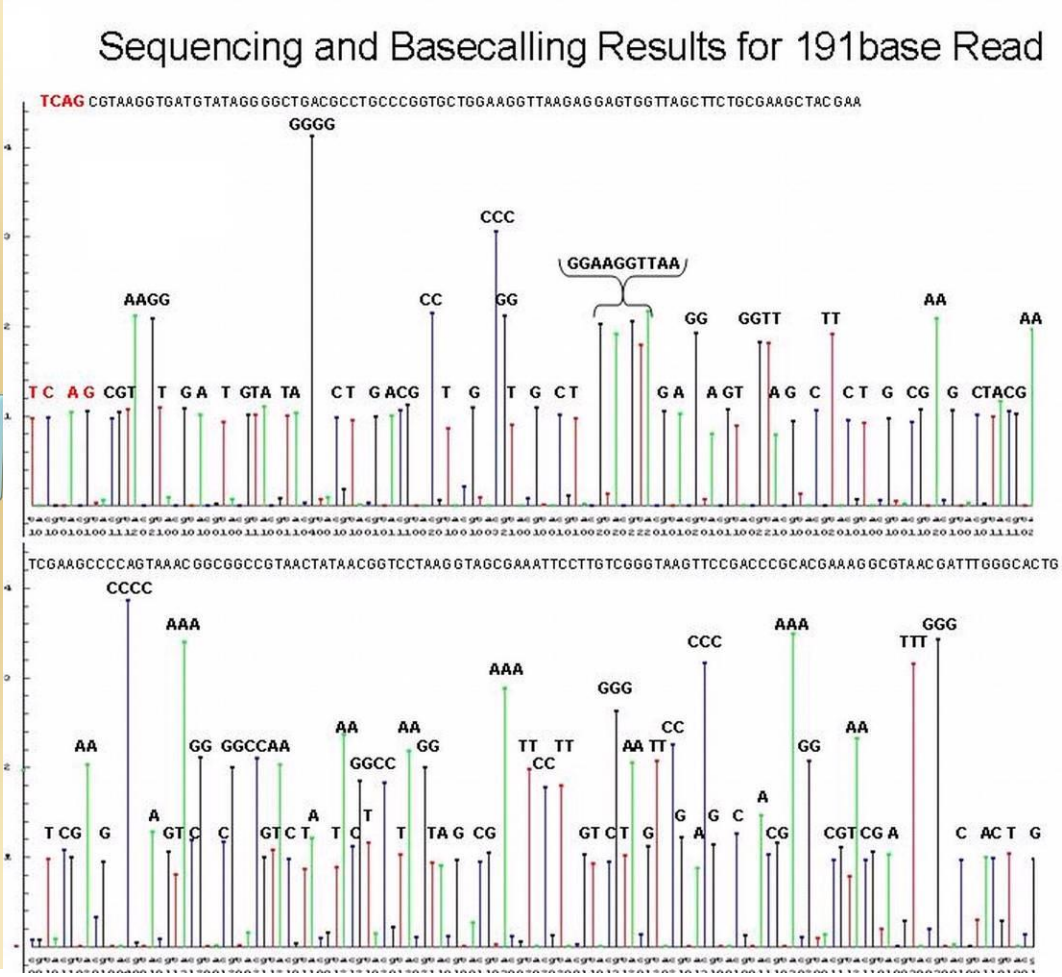


4) Perform sequencing by synthesis on the 454 instrument

J. Craig Venter™ INSTITUTE

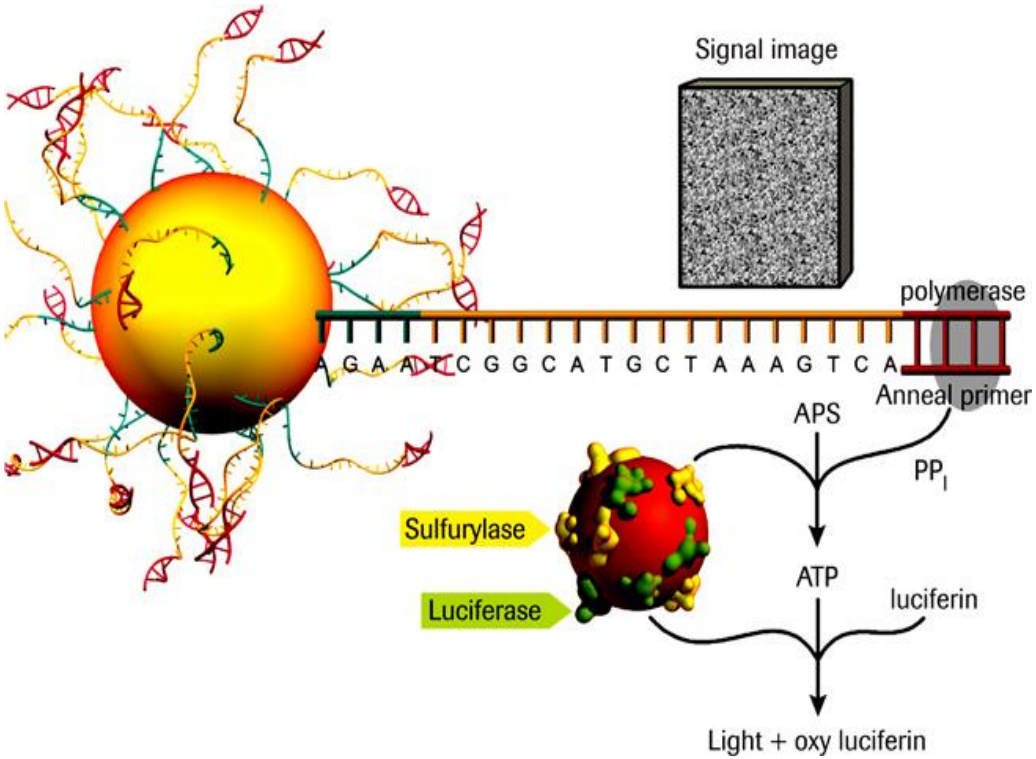
454 Instrument and Data Output





J. Craig Venter INSTITUTE

454 Sequencing Workflow
Sequencing by Synthesis

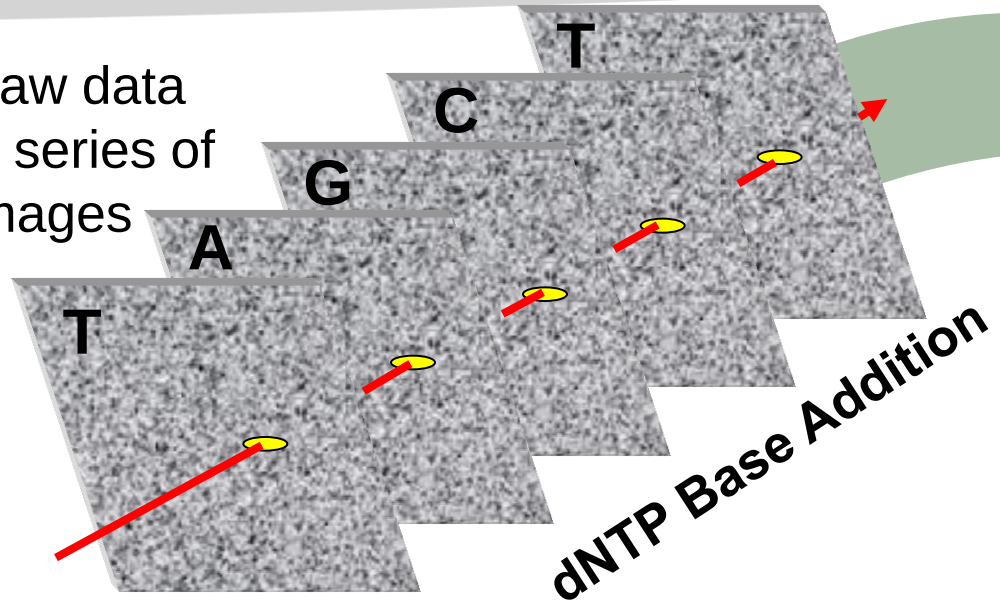


- Bases (TACG) are flowed sequentially and always in the same order (100 times for a large GS FLX run) across the PicoTiterPlate device during a sequencing run.
- A nucleotide complementary to the template strand generates a light signal.
- The light signal is recorded by the CCD camera.
- The signal strength is proportional to the number of nucleotides incorporated.

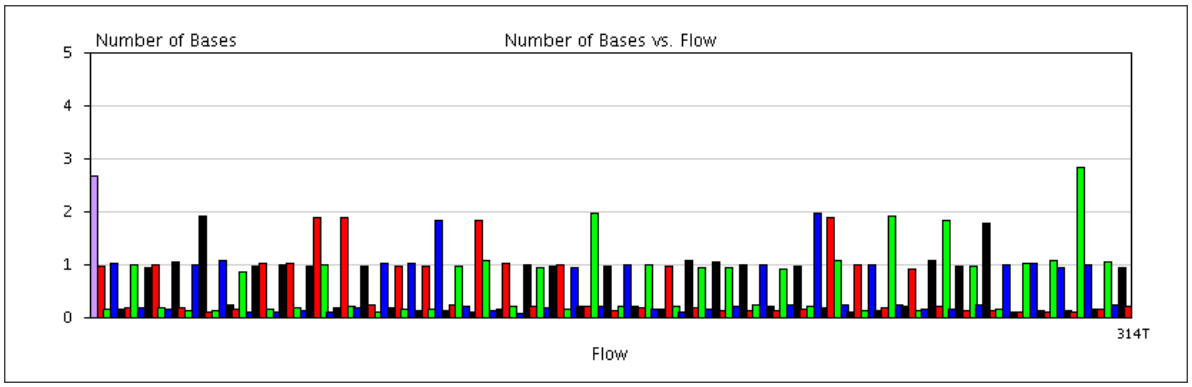
J. Craig Venter™ INSTITUTE

454 GS FLX Data
Image Processing Overview

1. Raw data is series of images



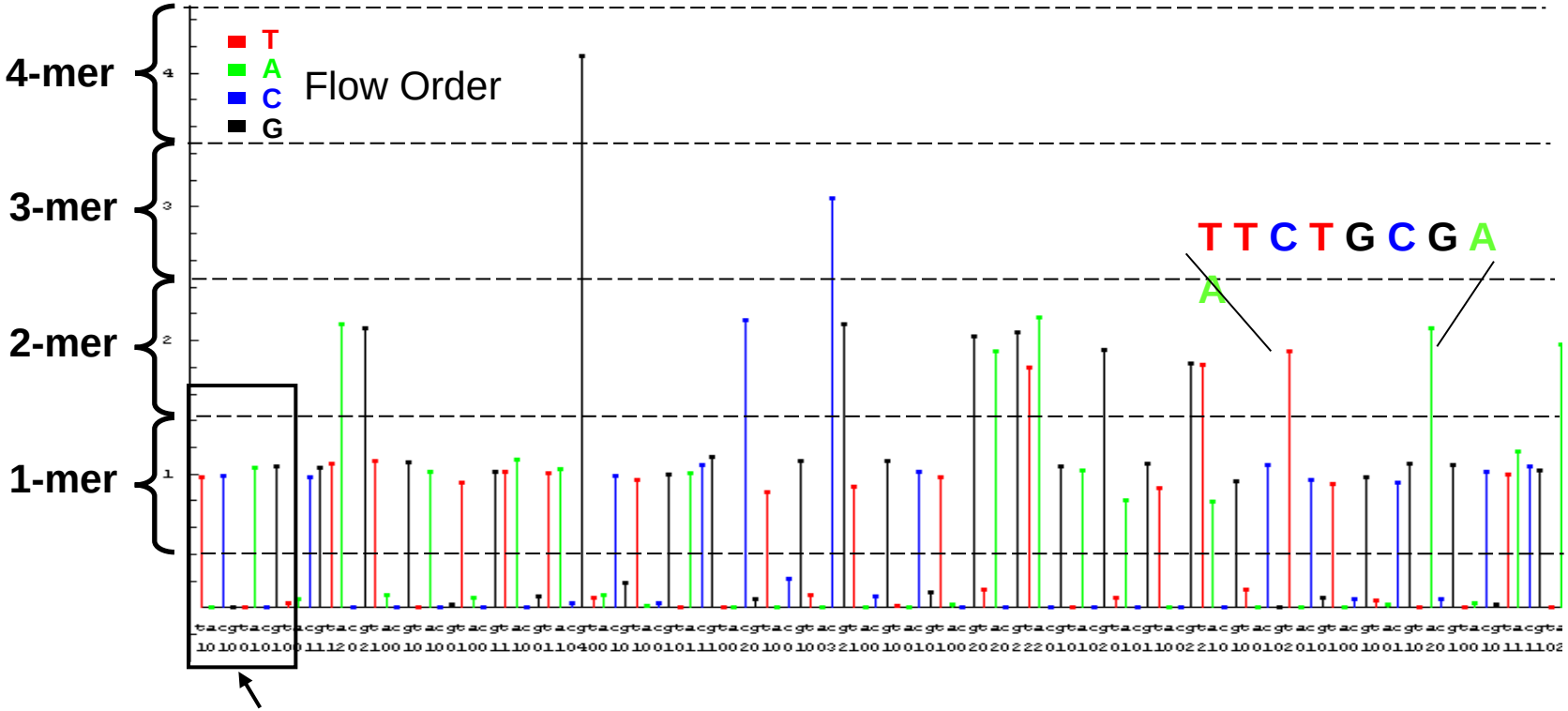
1. Each well's data extracted, quantified and normalized



1. Read data converted into "flowgrams"

J. Craig Venter INSTITUTE

454 GS FLX Data
Flowgram Generation



J. Craig Venter INSTITUTE

454 GS FLX Plate

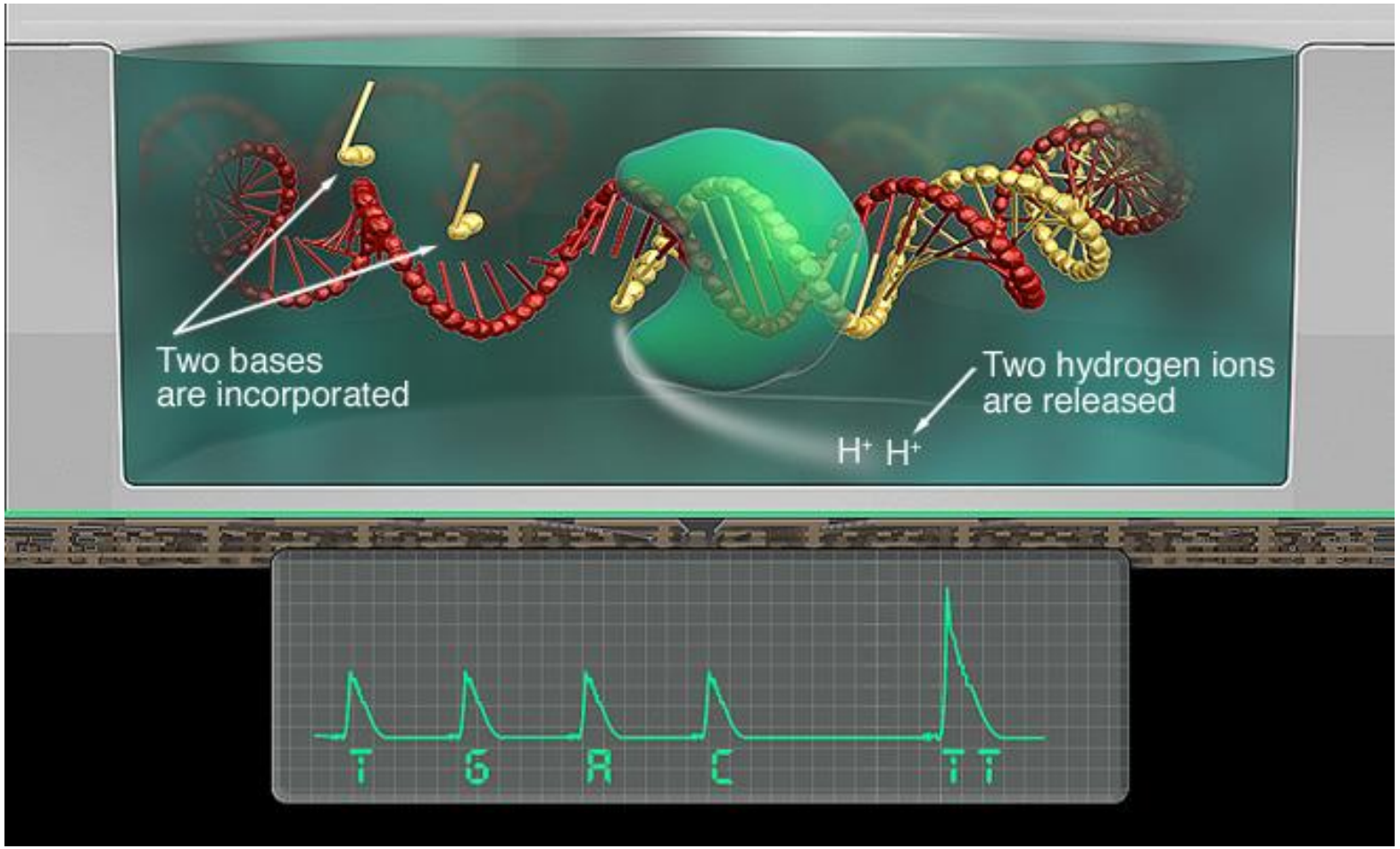
J. Craig Venter INSTITUTE

454 GS FLX Sequencer

Ion Torrent Sequencing

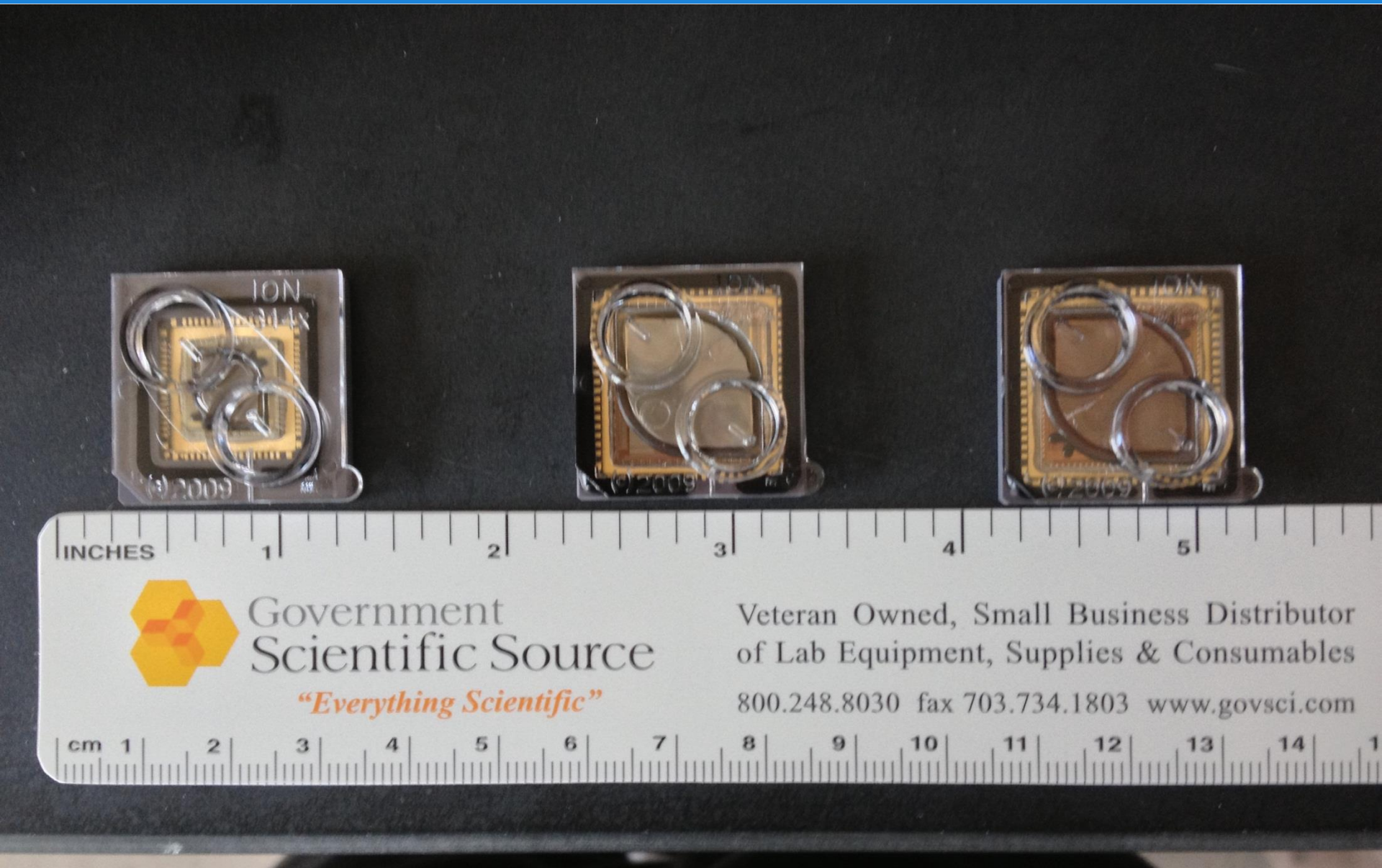
- Similar to 454, but rather than creating fluorescence and measuring light, Ion Torrent instead measures pH changes due to protons released during base incorporation
- The Ion Torrent chips are a massively parallel array of the world's smallest pH meters
- As a semiconductor device, Ion Torrent has been able to make there chips denser and denser (more and more wells), following the trend of the electronics industry

Ion Torrent Sequencing

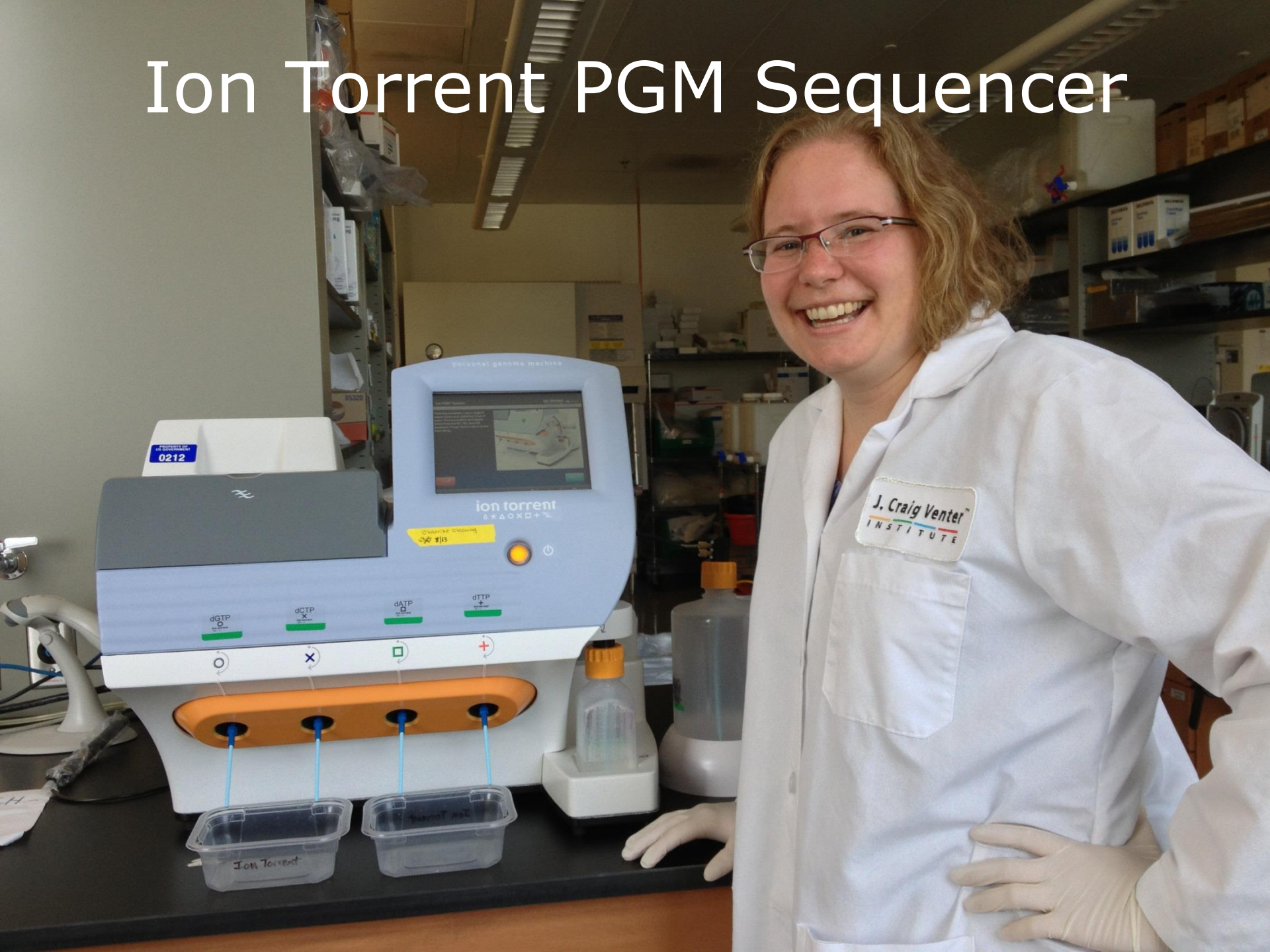


J. Craig Venter™
INSTITUTE

Ion Torrent Chips



Ion Torrent PGM Sequencer

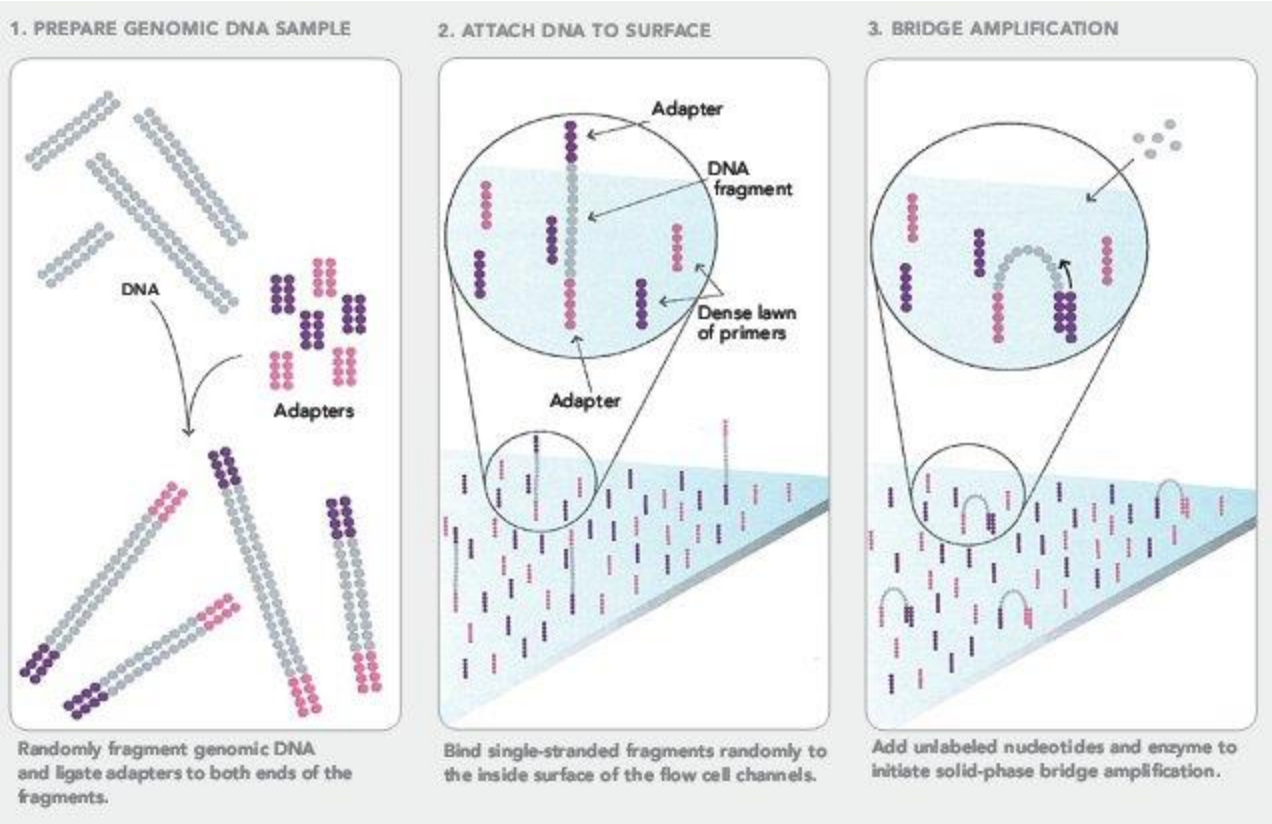


Illumina Sequencing

- Technology Overview
- Mate Pair Library Construction

J. Craig Venter™ INSTITUTE

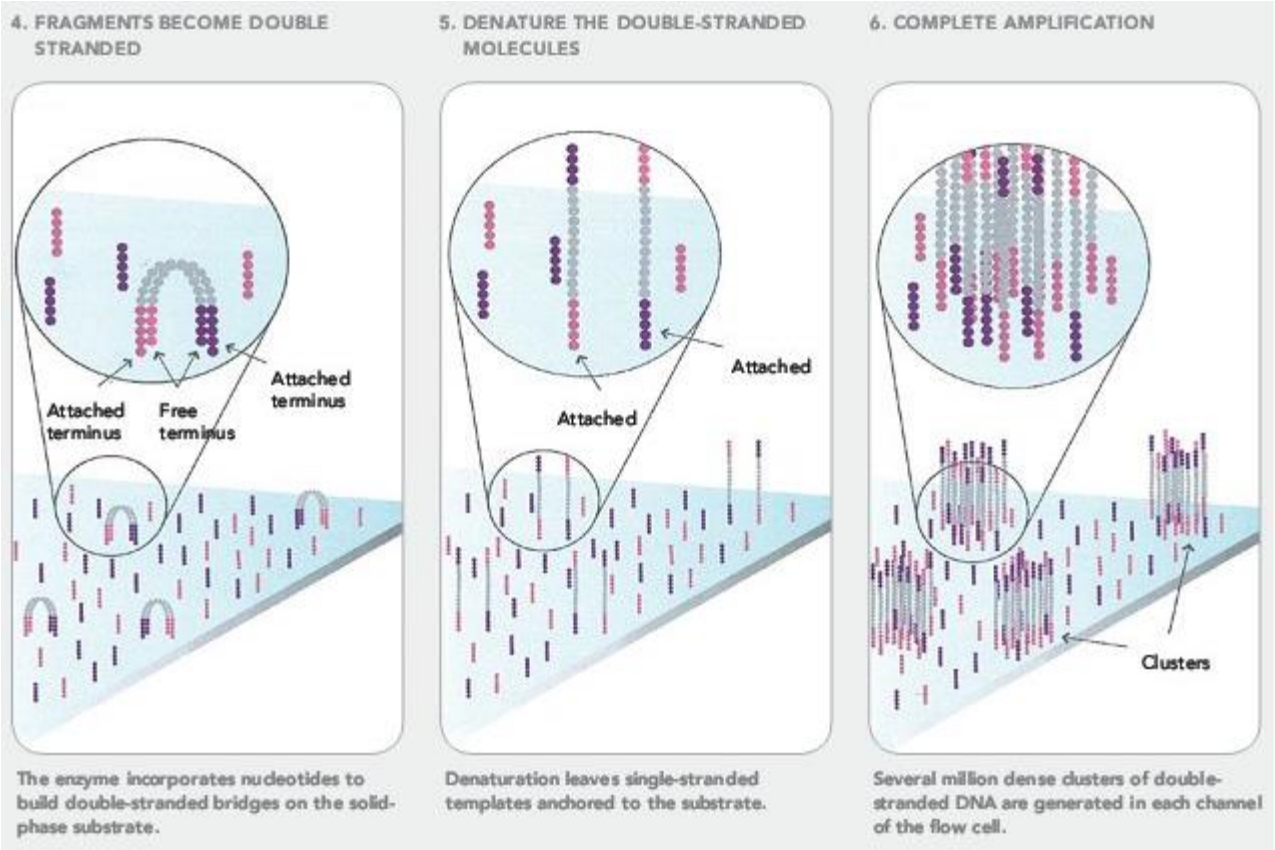
Illumina Technology Overview (1)



http://seqanswers.com/forums/showthread.php?t=21v

J. Craig Venter™ INSTITUTE

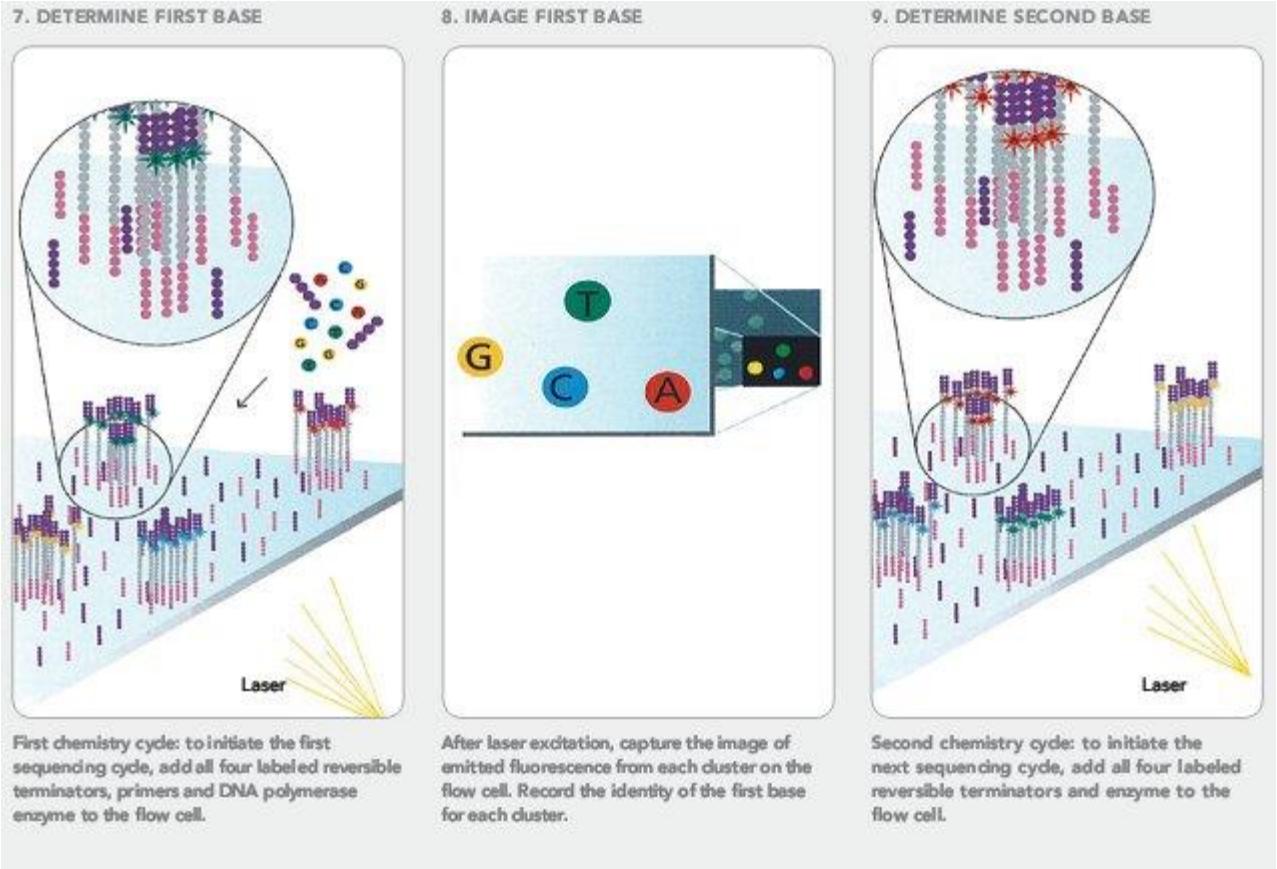
Illumina Technology Overview (2)



http://seqanswers.com/forums/showthread.php?t=21v

J. Craig Venter INSTITUTE

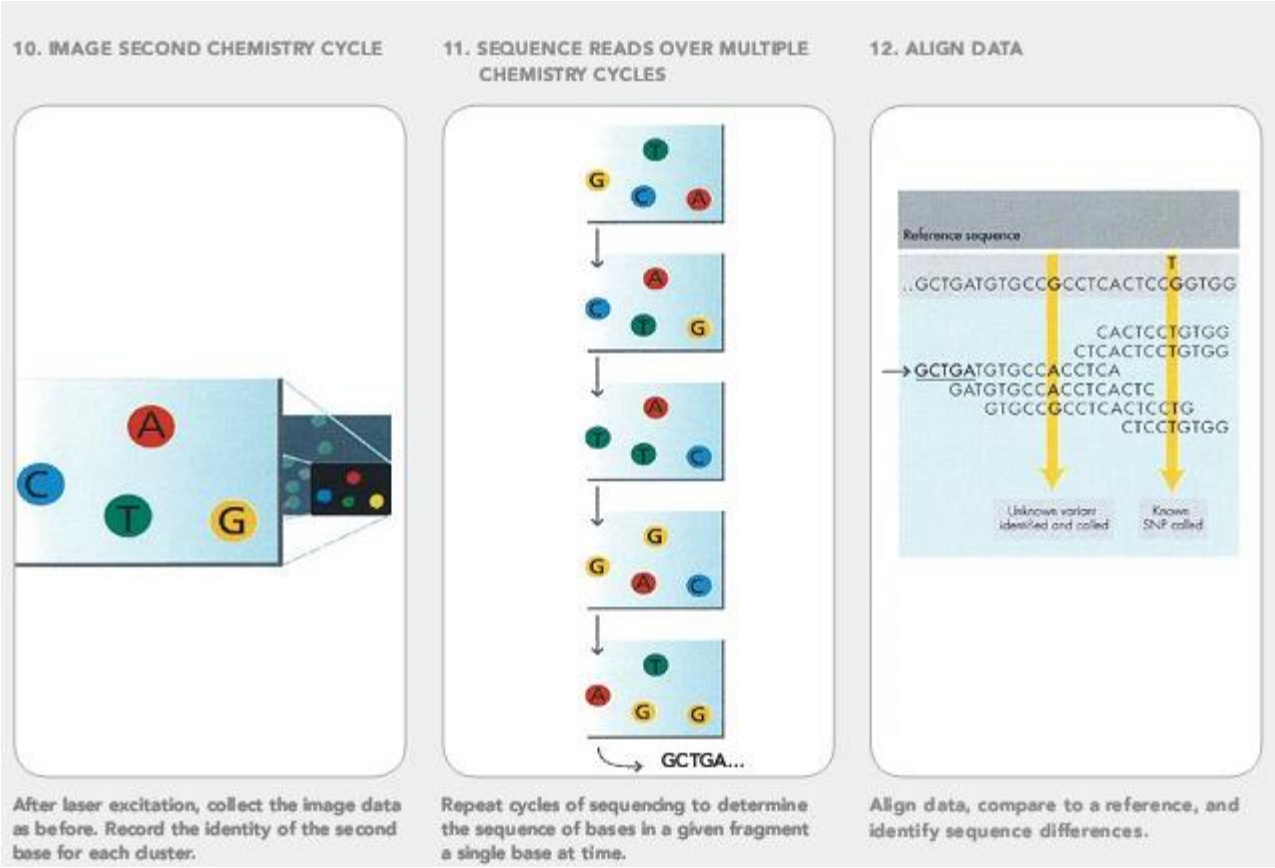
Illumina Technology Overview (3)



http://seqanswers.com/forums/showthread.php?t=21v

J. Craig Venter™ INSTITUTE

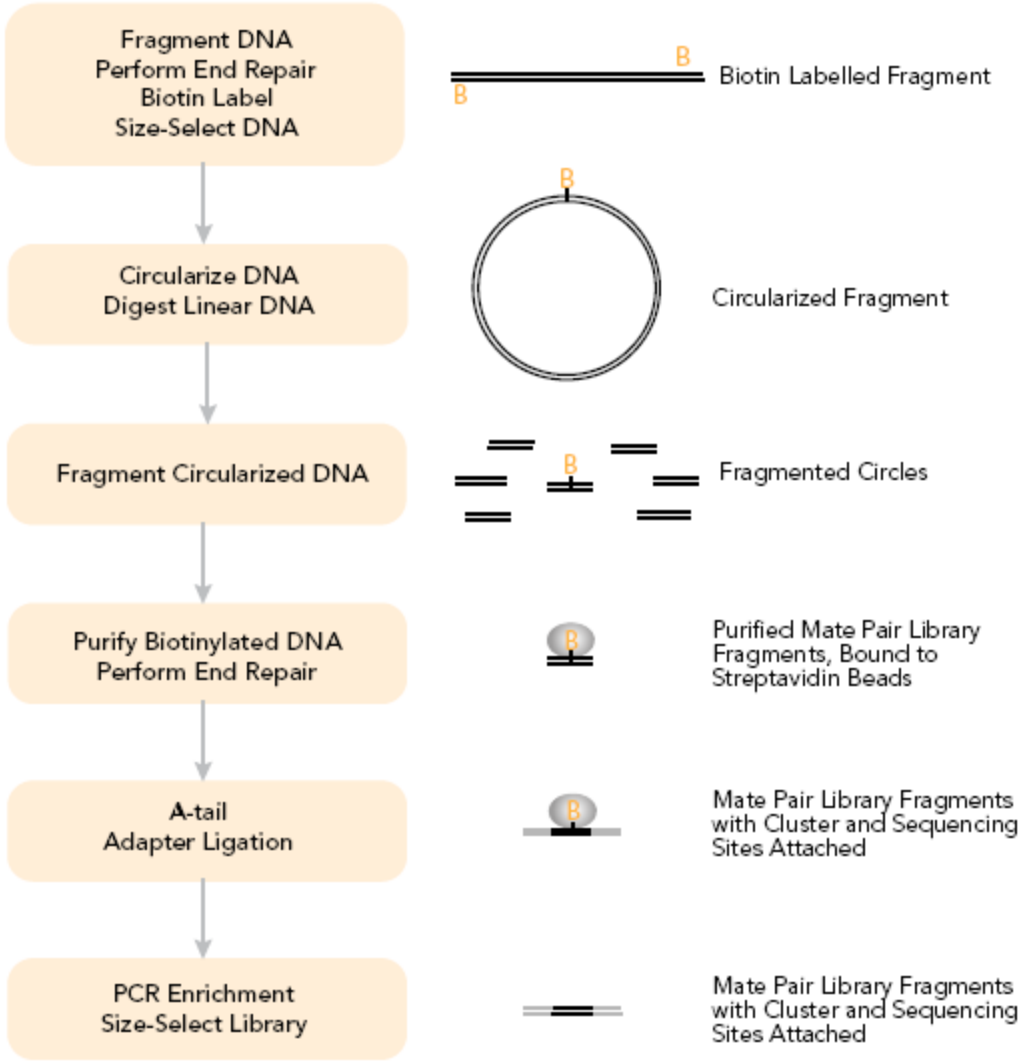
Illumina Technology Overview (4)



http://seqanswers.com/forums/showthread.php?t=21v

J. Craig Venter INSTITUTE

Illumina Mate Pair Library Construction



J. Craig Venter™ INSTITUTE

Illumina Flow Cells

Illumina MiSeq Sequencer

J. Craig Venter INSTITUTE

Illumina HiSeq Sequencer

Other Technologies

- **Pacific Biosciences** – single molecule sequencing, measures the incorporation of a single dye labelled base at a time, by laser-excitation of an *extremely* small volume that contains the polymerase and the DNA
- ***Oxford Nanopore*** – single molecule sequencing, measures the electrical changes in a pore that arise when bases enter and exit the pore.

Readings

- Zagori et al. (2012) **Read length versus depth of coverage for viral quasispecies reconstruction**. *PloS One 7(10):e47046*
- Quail et al. (2012) **A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers** *BMC Genomics 13:341*
- Liu et al. (2012) **Comparison of Next-Generation Sequencing Systems** *J. Biomed Biotechnol July 5.*
- Metzker (2010) **Sequencing technologies – The next generation**. *Nature Reviews Genetics 11:31*
- Harismendy et al. (2009) **Evaluation of next generation sequencing platforms for population targeted sequencing studies.** *Genome Biology 10:R32*
- Nagarajan and Pop (2010) **Sequencing and genome assembly using next-generation technologies.** *Computational Biology, Methods in Molecular Biology Vol. 673*