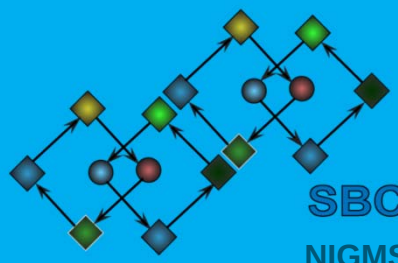


NIGMS funded Center

Experimental Methods in Systems Biology

Part of the Coursera Certificate in Systems Biology

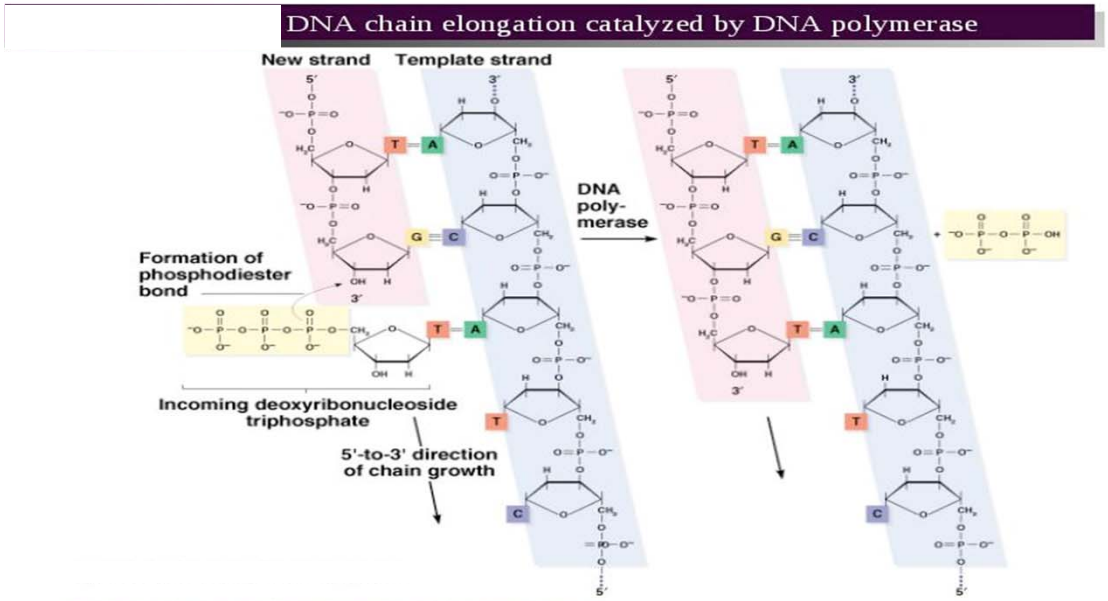
Marc Birtwistle, PhD
Department of Pharmacology & Systems Therapeutics
Fall 2014, Week 2, Deep mRNA Sequencing

Icahn School of Medicine at Mount Sinai

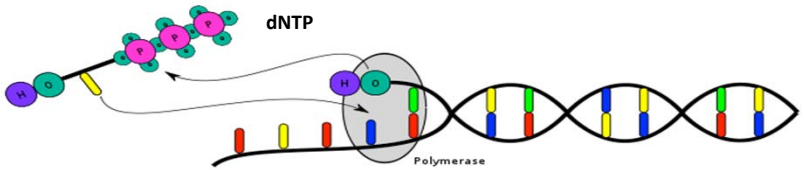
2nd Generation Sequencing

Sequencing by synthesis—2nd generation chemistry

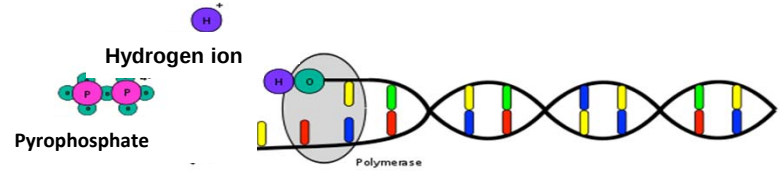
Sequencing by synthesis can solve the issue of DNA strand separation after addition of complimentary base



Products of base addition reaction



Polymerase integrates a nucleotide.



Hydrogen and pyrophosphate are released.

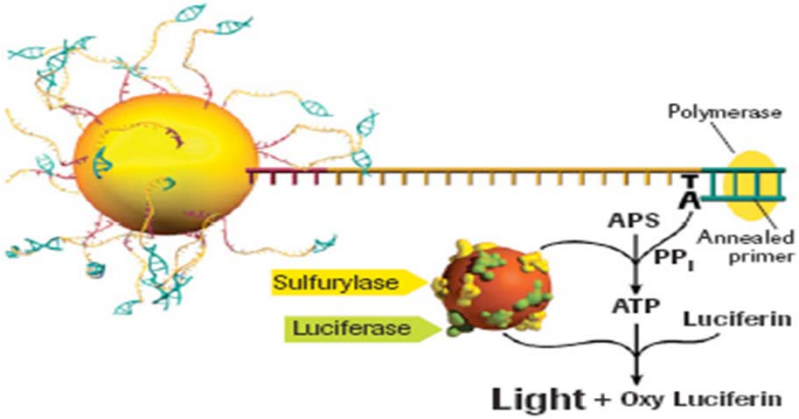
Second (next) generation sequencing technologies

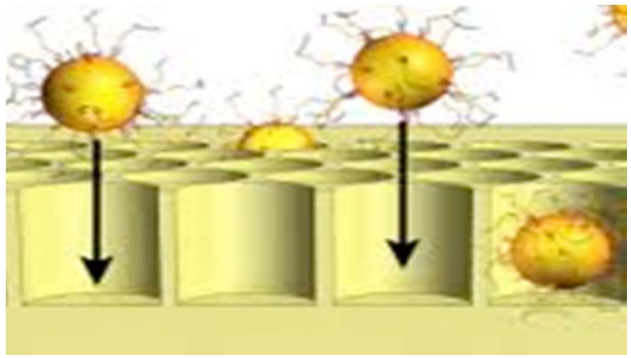
Company	Platform	Method	Detection	Length	Advantages	Disadvantages
Roche/454	FLX genome sequencer	Pyrosequencing Detecion of pyrophosphate release	Optical	0.4-1 Kb	Long read length	High cost; challenging sample prep.
Life Technologies	IonPGM IonProton	Sequencing by synthesis	Released H+ ions	200 bp	Rapid runs, low cost	Lower throughput compared to Ilumina; Maturing technology
Illumina	HiSeq 2500 MiSeq	Rev. terminator sequencing by synthesis	Fluorescence/ optical	2x150 or 2x250 bp	Very high throughput	Long run time for standard runs
Life technologies	5500 SOLiD W system	Sequencing by ligation	Fluorescence/ optical	1x75 or 2x60 bp	Very high throughput	Short read lengths; non-standard data analysis

Illumina platform is market leader – one 30x coverage human genome for $5-10k

Sequencing by synthesis:

Detection and estimation of pyrophosphate release

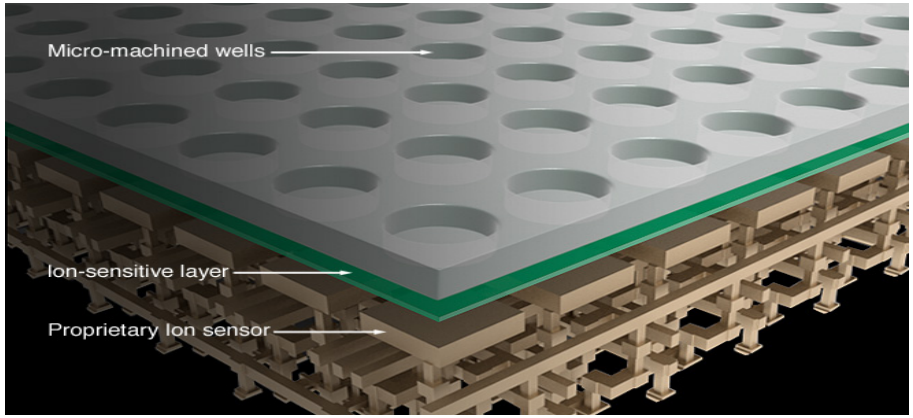


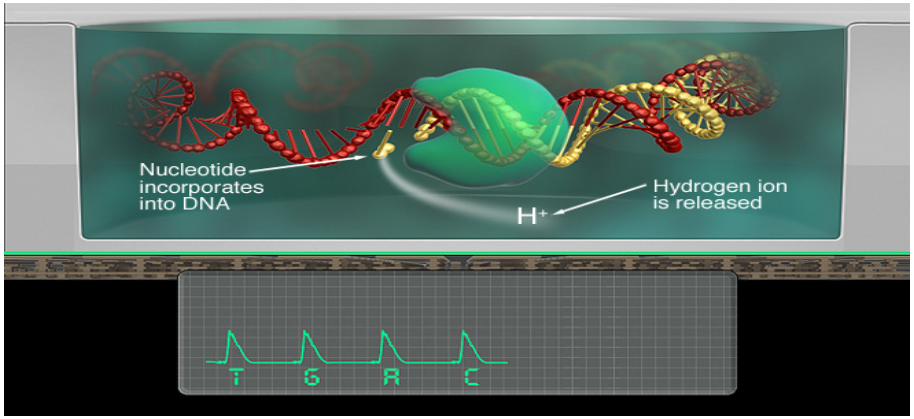


454/Roche: First NGS sequencer in market

Sequencing by detecting H^+ release after addition of the base

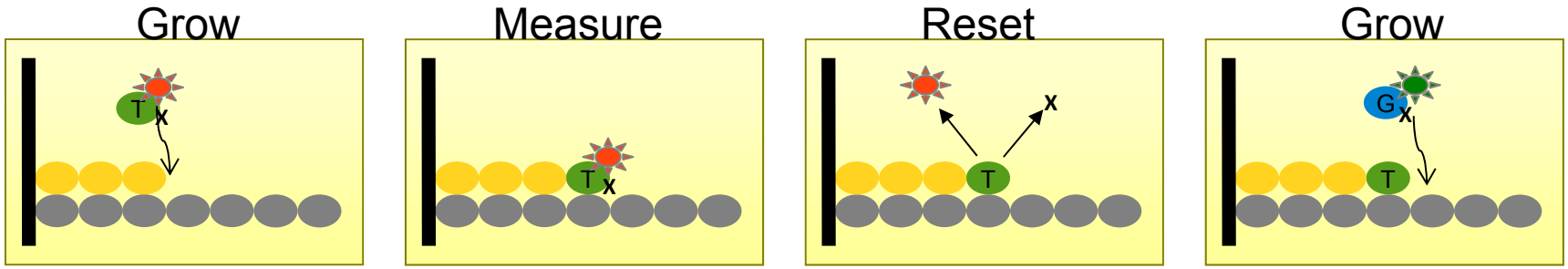
- Polymerase releases H+ during base incorporation
- Measured by semi-conductor wafer
- Essentially a massively parallel pH meter



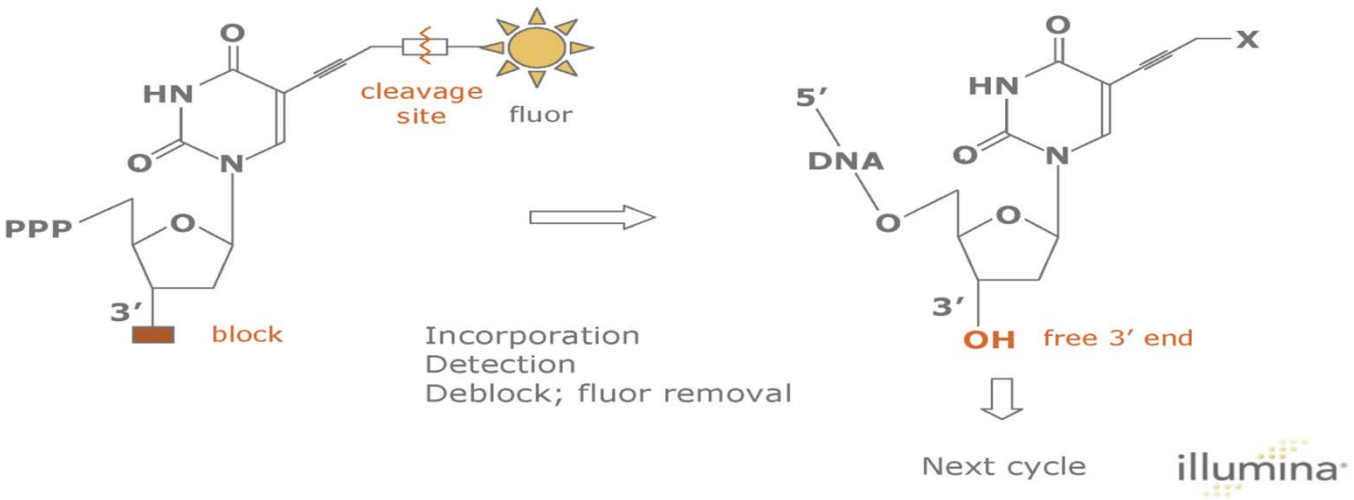


Life/IonTorrent 'Electrical' Sequencing

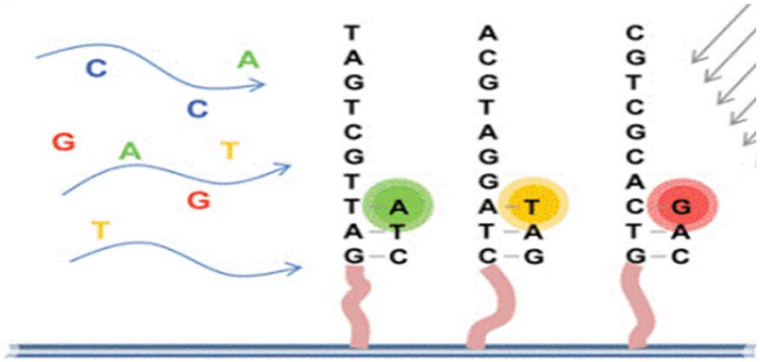
Dye sequencing by synthesis using reversible terminator

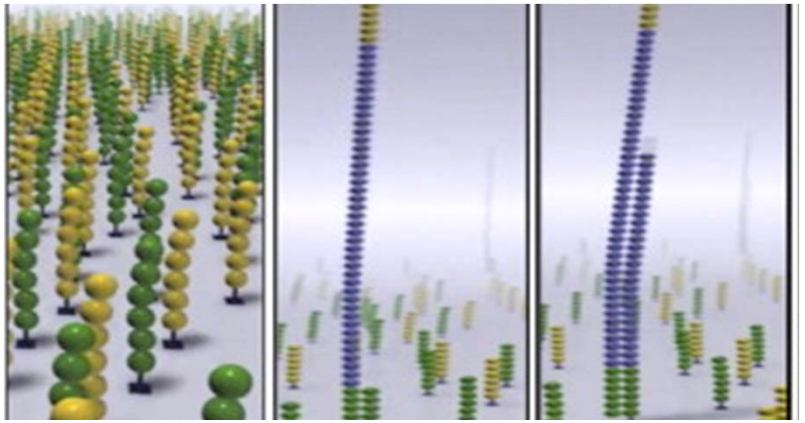


Reversible dye chemistry



Non-Well format massively parallel sequencing by Illumina





Limitations of second-generation sequencing

- Second generation sequencing requires amplification to get sufficient number of sequences to meet detection thresholds.
  - Coverage of GC rich sequences
  - Amplification bias
    - Inherently a problem for quantitation
    - Unique molecular identifiers however may solve this problem (Islam et al., Nat Methods, 2014; Kivioja et al., Nat Methods, 2012)—see Week 1

- Second Gen Seq technologies have practical limits in read length
  - Mapping of long repeat regions in the genome
  - Identification and mapping of duplicate genes and pseudogenes

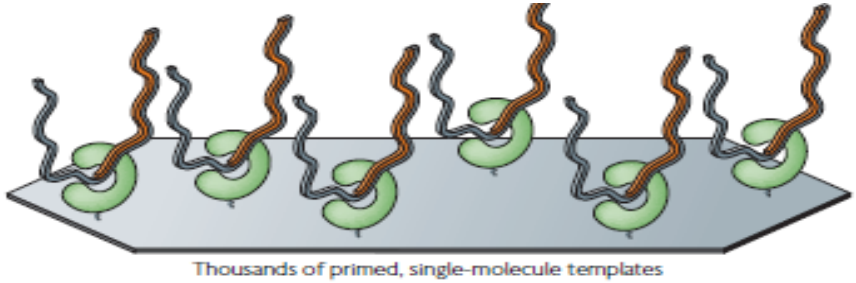
- Third-generation sequencing seeks to therefore have longer read length without amplification

Third-generation sequencing technologies

Company	Platform	Method	Detection	Length	Advantages	Disadvantages
Pacific Biosciences	PacBio RS II	Single-molecule real-time sequencing	Fluorescence/ optical	Up to 20Kb	Very long read length	High per-base error rate and cost; low throughput
Oxford Nanopore	GridION MinION	Nanopore sequencing	Voltage Sensing	>10kb?	Very long read lengths, Low cost and low error rates, fast run times?	???

PacBio is currently market leader

PacBio real-time sequencing

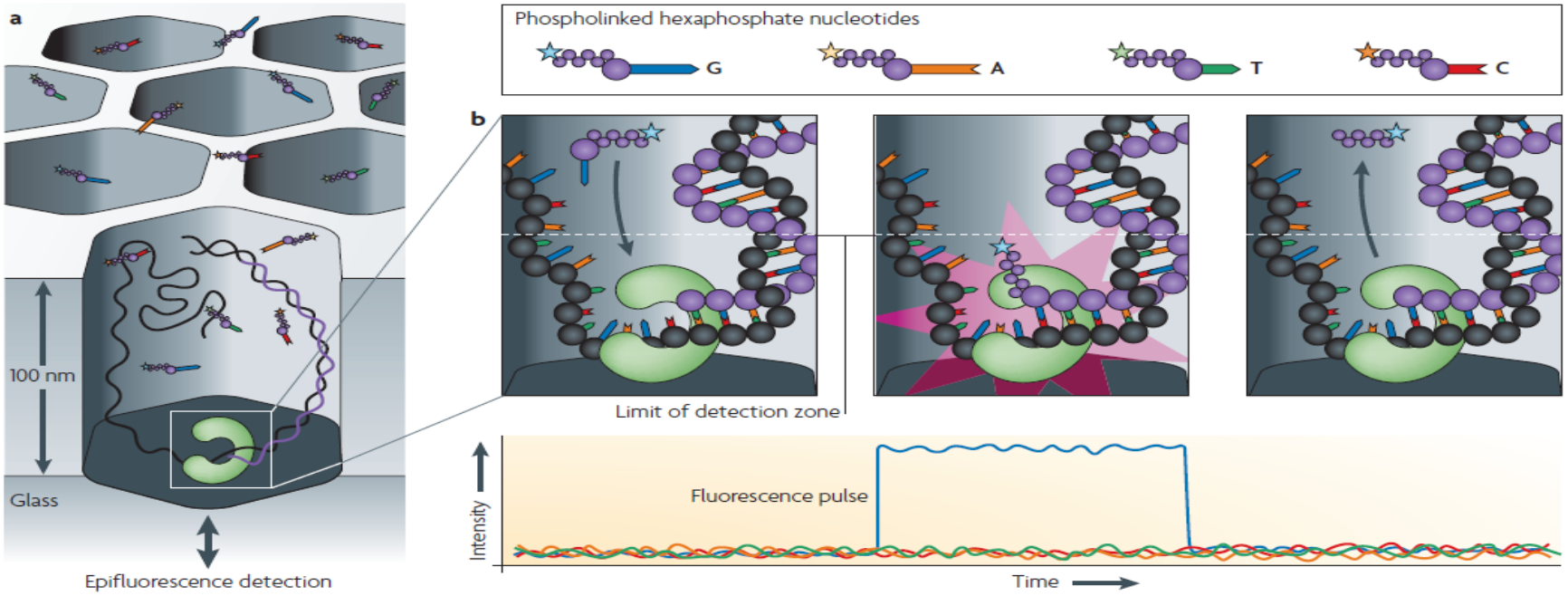


Pacific Biosystems RS
3rd Generation Single Molecule Sequencer

Pacific Biosciences RS
Immobilized Polymerases + fluorescent dNTP +
really, really good optics

The particular polymerase used (“displacing”) is on the order of bp/sec

Detection of base incorporation



Extension results in different fluorescent signal for each base

Detecting base modifications