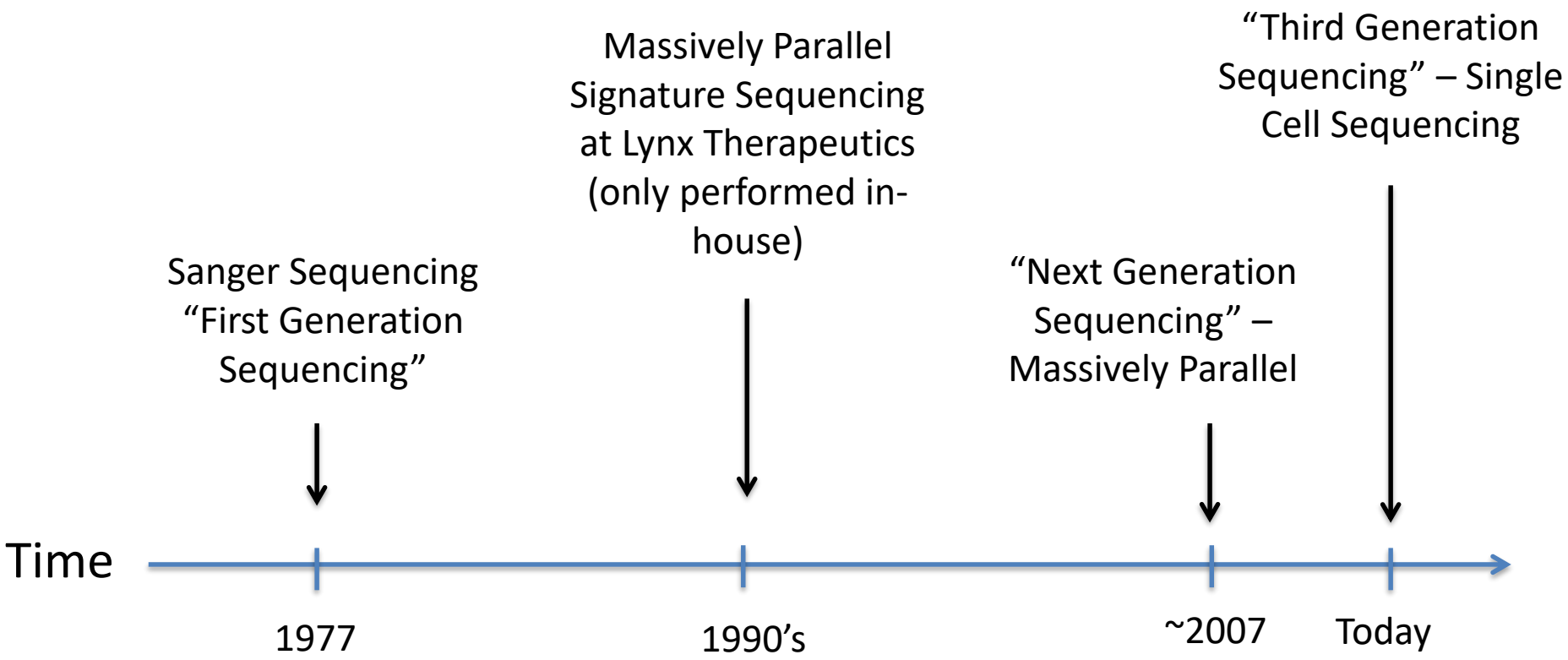
CS342: Bioinformatics Assembling a Genome

DNA Sequencing Technologies



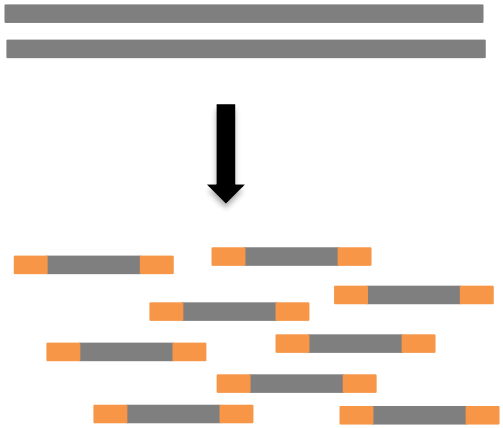
Next-Gen Sequencing!

illumina®

454 SEQUENCING

applied biosystems
by life technologies

Next-Gen Sequencing

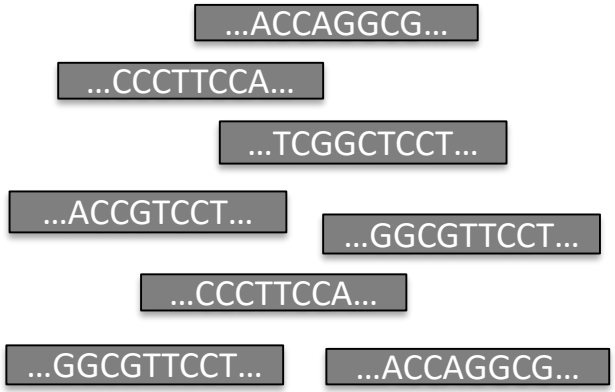


DNA

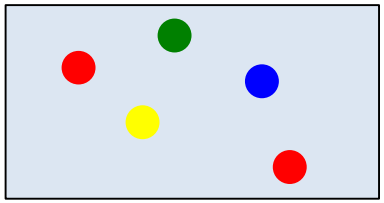
DNA Fragment

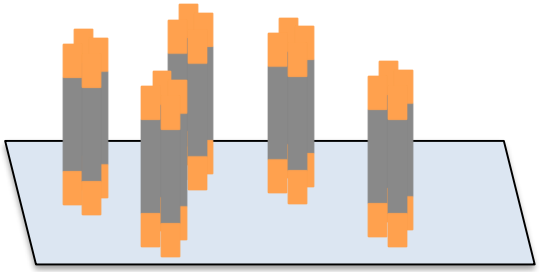
Adapter

DNA is fragmented and adapters are ligated to both ends

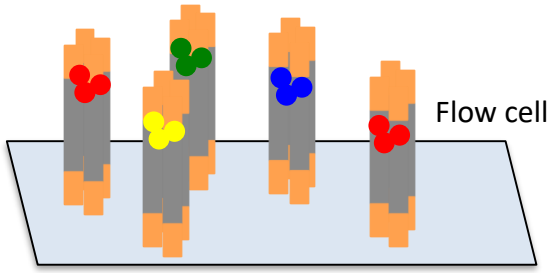


Output is a ``Bag of reads."

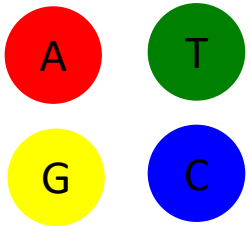




Fragments hybridize to flow cell and are amplified.



Fluorescently labeled nucleotides are added.

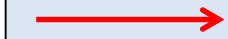


Color indicates nucleotide.

Two Different Protocols

Single Read Sequencing

Length *n* reads taken from one end of a DNA fragment.



ACTTCTATCTGATAGTCAATGTAG
TGAAGATAGACTATCAGTTACATC

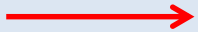
n = 7

All output reads have length *n*

ACAGATC TATGATC
ATTGATC
TCCGATC
CGTGATC TATGATC
ATTGATC

Paired-End Sequencing

Length *n* reads taken from both ends of a DNA fragment.



ACTTCTATCTGATAGTCAATGTAG
TGAAGATAGACTATCAGTTACATC

n = 7

All output reads have length *n* and are part of a pair of reads

Unknown length

ACTTCTA??????????.......
.......??????????TTACATC

AGGACTA????????.......
.......????????GTACCTC

CTTCCGC??????????.......
.......??????????ACAGATC

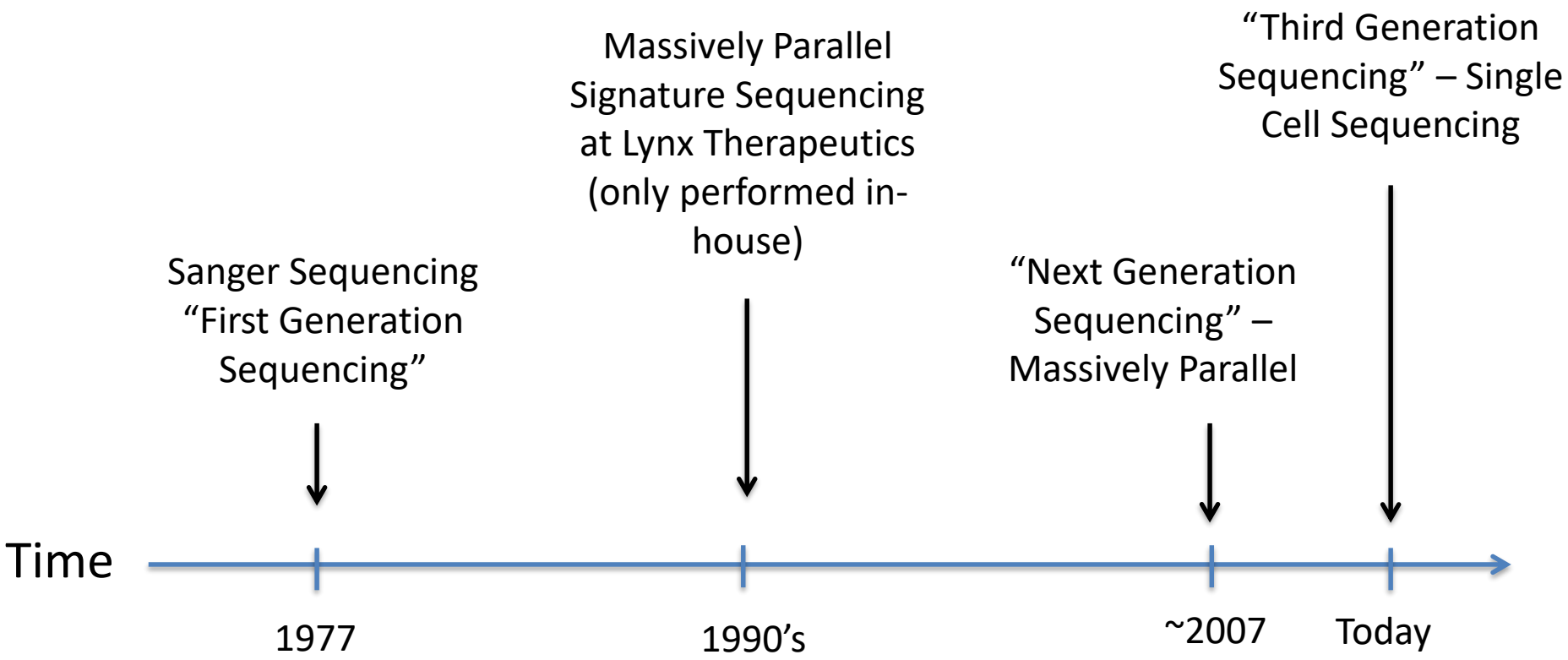
A Few Examples

	MiSeq	HiSeq 3000	HigSeq 4000	HiSeq X	NovaSeq
Max Read Length	2 x 300 bp	2 x 150 bp	2 x 150bp	2 x150bp	2 x 250bp
# Reads per run	1-25 Million	2.1 Million – 5 Billion	Up to 10 Billion	5.3-6 Billion	32-40 Billion
Run Time:	4-56 hrs	< 1-3.5 days	< 1-3.5 days	< 3 days	13-44 hrs
Output:	540 Mb - 15Gb	650 – 750 Gb	1300 – 1500 Gb	1.6 – 1.8 Tb	4800 – 6000 Gb

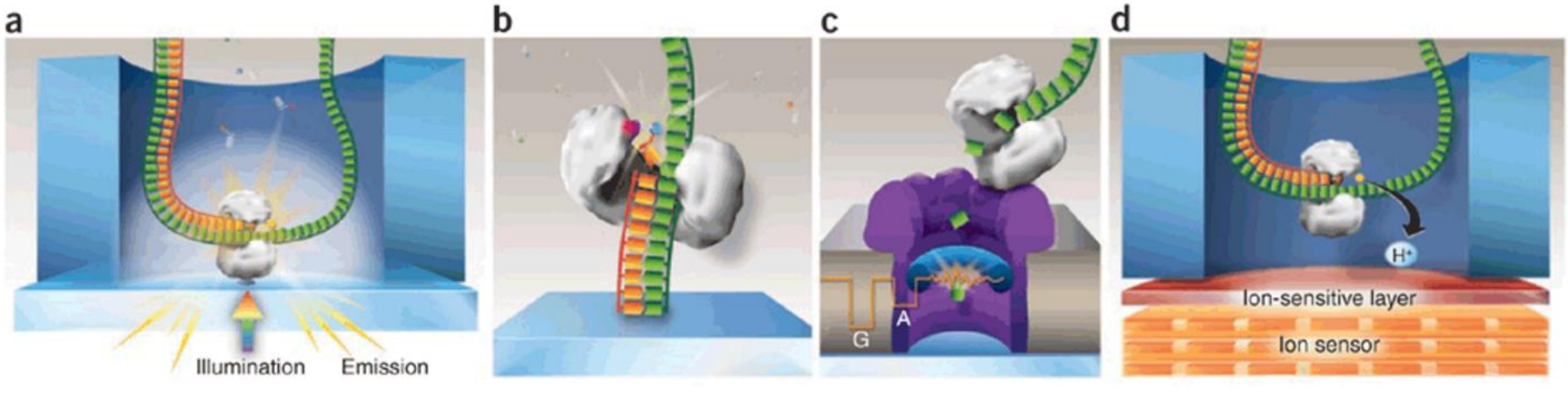
~ 1% error rate

** Numbers updated from: https://www.illumina.com/systems.html on 10/6/2019 **

DNA Sequencing Technologies



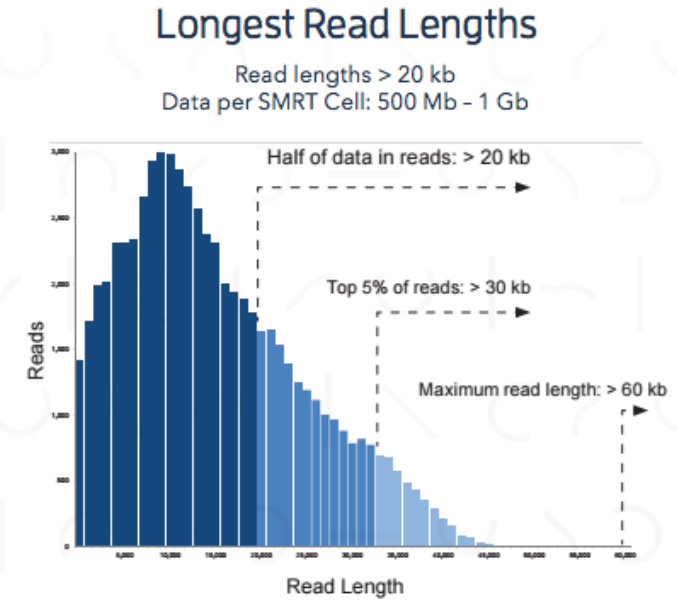
Single Molecule Real Time (SMRT)



Video

PACBIO®

~ 15% error rate



http://www.nature.com/nbt/journal/v28/n5/fig_tab/nbt0510-426_F1.html

DNA data from Africans reveals sequences that we'd missed

One reference genome doesn't capture the huge variation in human DNA.

CATHLEEN O'GRADY - 11/24/2018, 4:00 PM

https://arstechnica.com/science/2018/11/our-human-reference-genome-is-missing-a-lot-of-material/

Let’s try this out