

CS 466

Introduction to Bioinformatics

Lecture 24

Mohammed El-Kebir

Nov 18, 2020



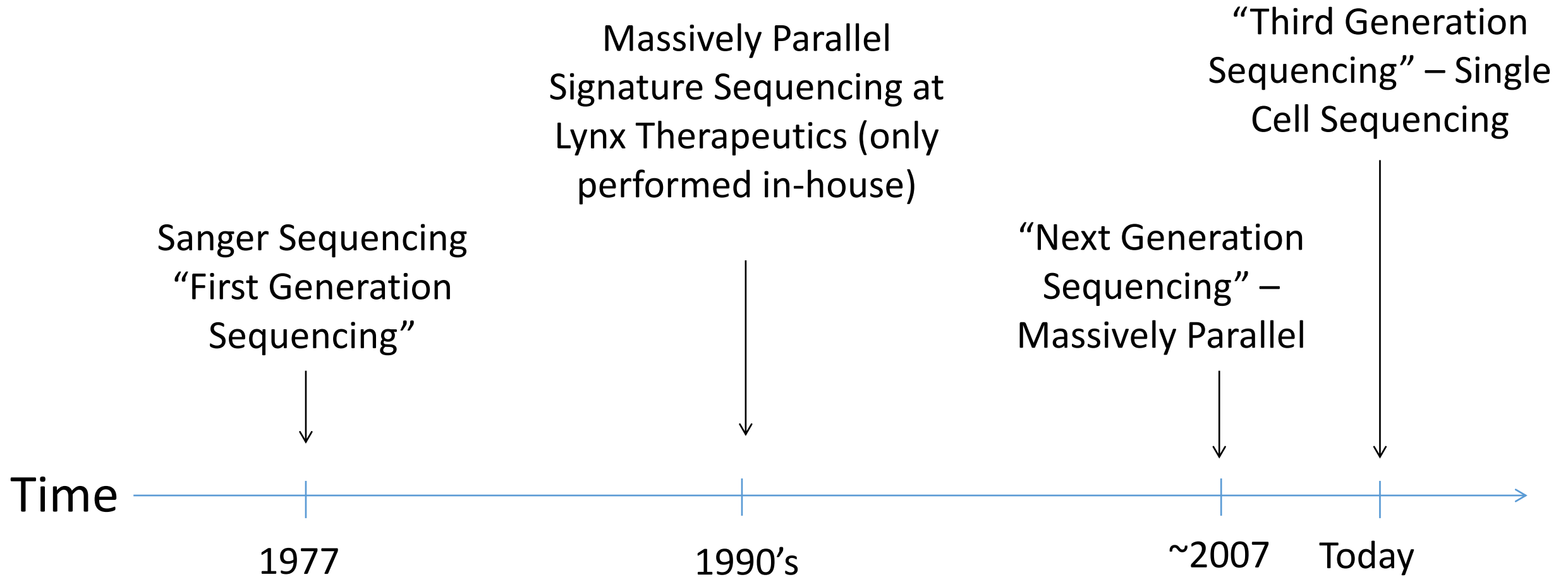
Course Announcements

HW4 due 11/21 by 11:59pm

Calendar:

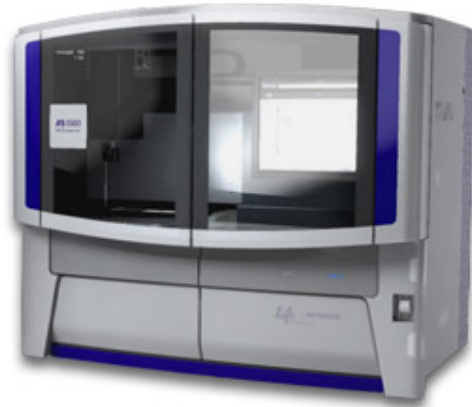
- Nov 18 (Wednesday) – Genome assembly I
- Nov 20 (Friday) – Genome assembly II
- Nov 24 (Wednesday) – Fall break
- Nov 26 (Friday) – Fall break
- Nov 29 (Sunday) -- HW5 released
- Dec 2 (Wednesday) – Project presentation
- Dec 4 (Friday) – Project presentation
- Dec 7 (Monday) – HW5 due
- Dec 9 (Wednesday) – Review
- Dec 9 (Wednesday) 7pm – Final exam
- Dec 20 (Sunday) – Project report

DNA Sequencing Technologies



Next-Gen Sequencing!

illumina®

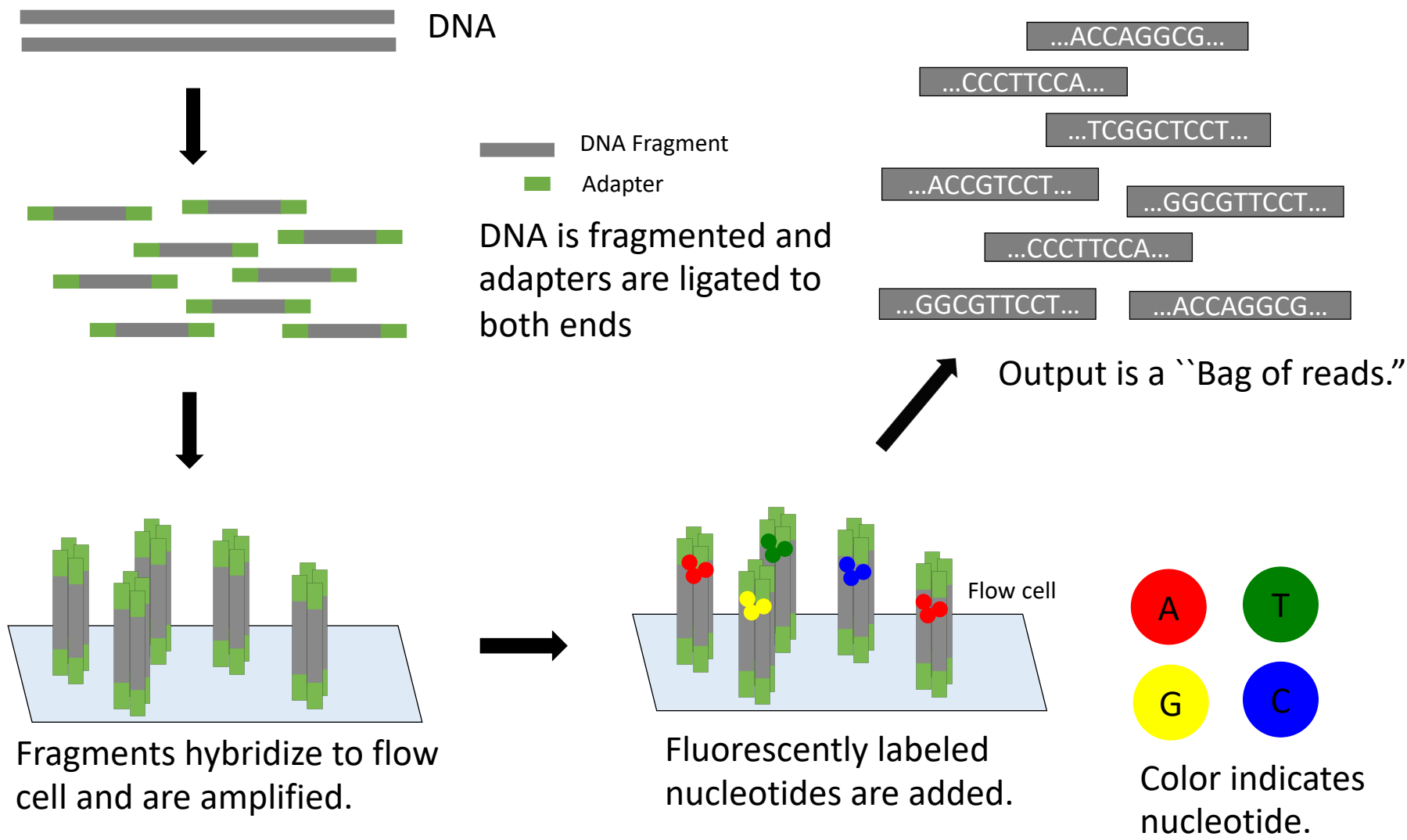


AB applied
biosystems™
by *life* technologies™



454
SEQUENCING

Next-Gen Sequencing



Two Different Protocols

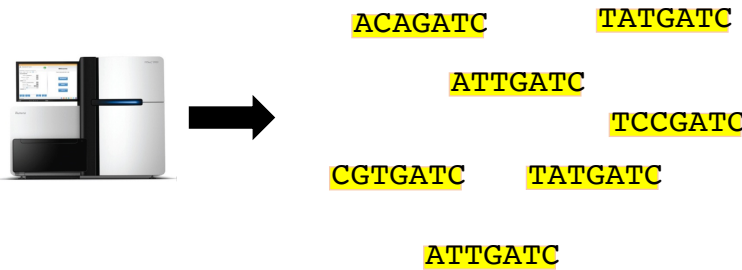
DNA
Fragment

Single Read Sequencing

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

→
ACTTCTATCTGATAGTCAATGTAG
TGAAGATAGACTATCAGTTACATC
 $n = 7$

All output reads have length n



Paired-End Sequencing

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

→
ACTTCTATCTGATAGTCAATGTAG
TGAAGATAGACTATCAG**TTACATC**
←
 $n = 7$

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



A Few Examples

	HiSeq 2500	NextSeq 500	MiSeq	HiSeq X Five	MiniSeq
Read Length	2 x 250 bp	2 x 150bp	2 x 250bp	2 x150bp	2 x 150bp
# Reads: 1 Flow Cell	~ 300 Million	~ 400 Million	24-30 Million	3 Billion	44-50 Million
Run Time:	60 hrs	~29 hrs	~39 hrs	< 3 days	~24 hrs
Output:	125 – 150 Gb	100 – 120 Gb	7.5-8.5 Gb	800-900 Gb	6.6 – 7.5 Gb

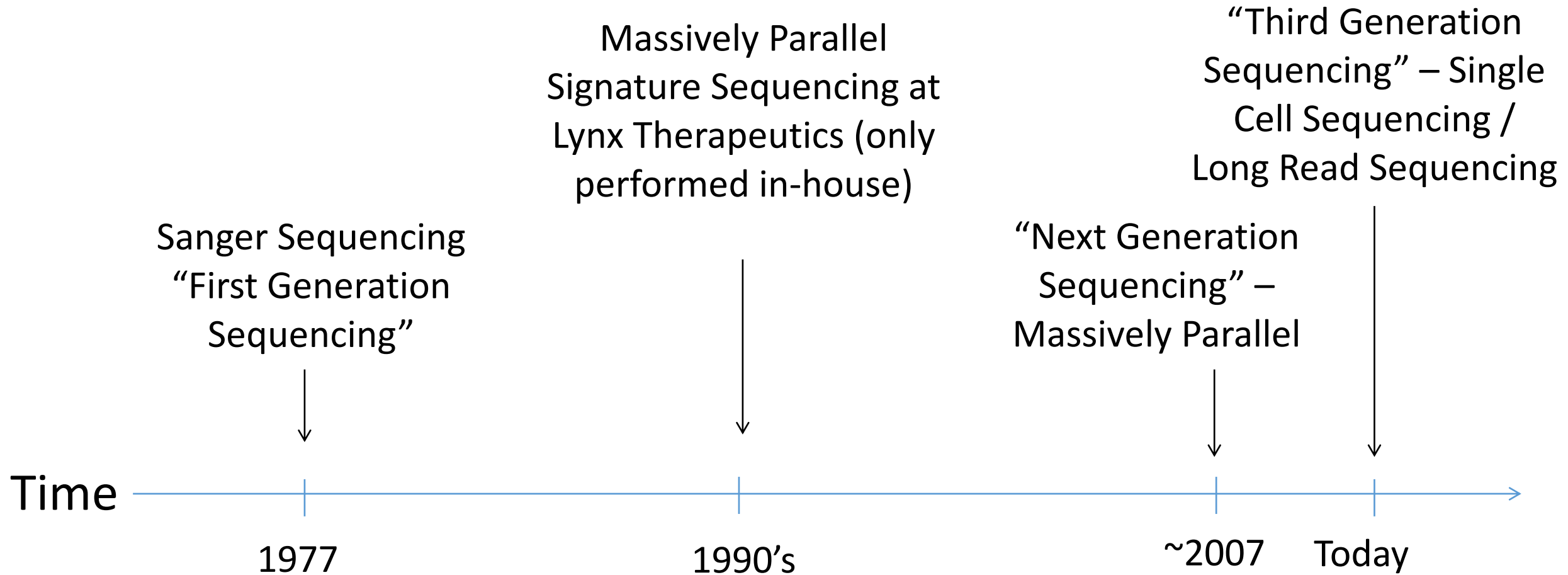
illumina®



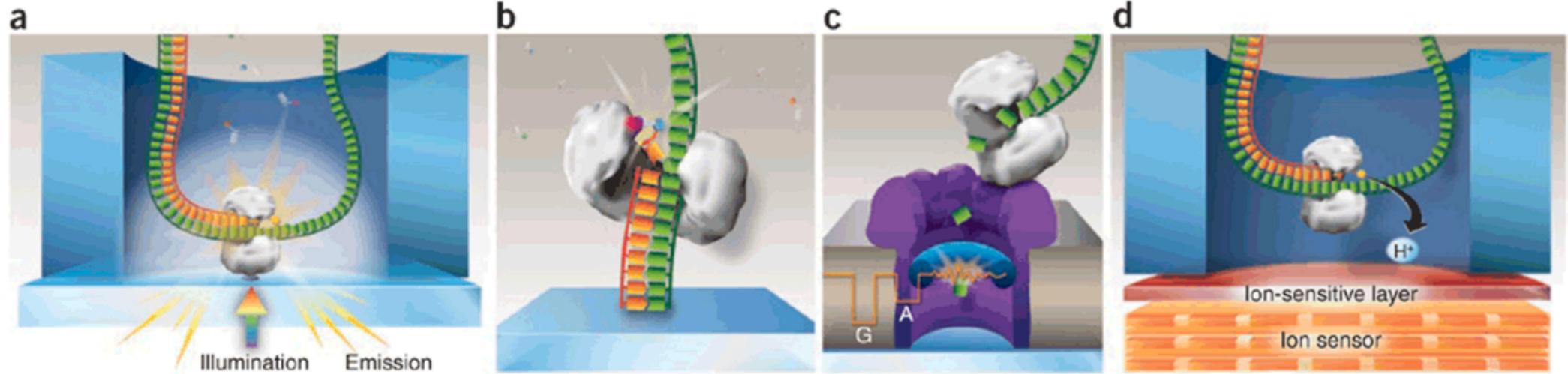
~ 1% error rate

** Numbers updated from: http://www.illumina.com/content/illumina-marketing/amr/en_US/systems.html on 2/4/2016 **

DNA Sequencing Technologies



Single Molecule Real Time (SMRT)



[Video](#)

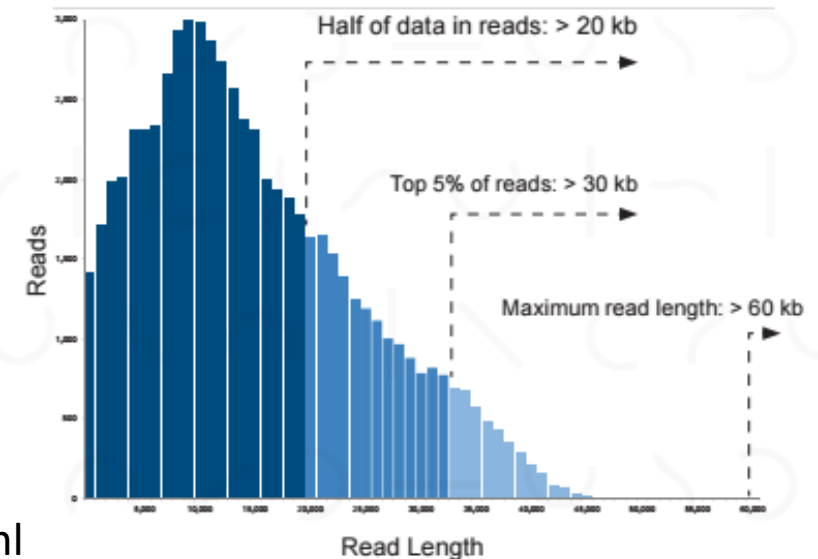


PACBIO®

~ 15% error rate

Longest Read Lengths

Read lengths > 20 kb
Data per SMRT Cell: 500 Mb - 1 Gb



Nanopore Technologies



MinION

Portable, real-time biological analyses

Single Cell Sequencing

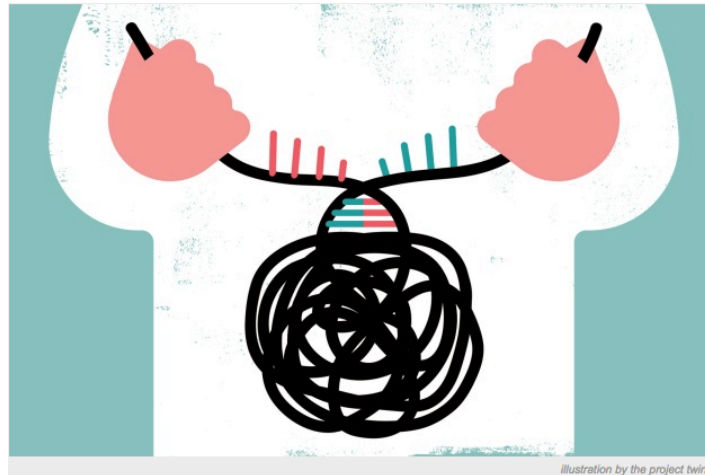
NATURE | TOOLBOX



Single-cell sequencing made simple

Data from thousands of single cells can be tricky to analyse, but software advances are making it easier.

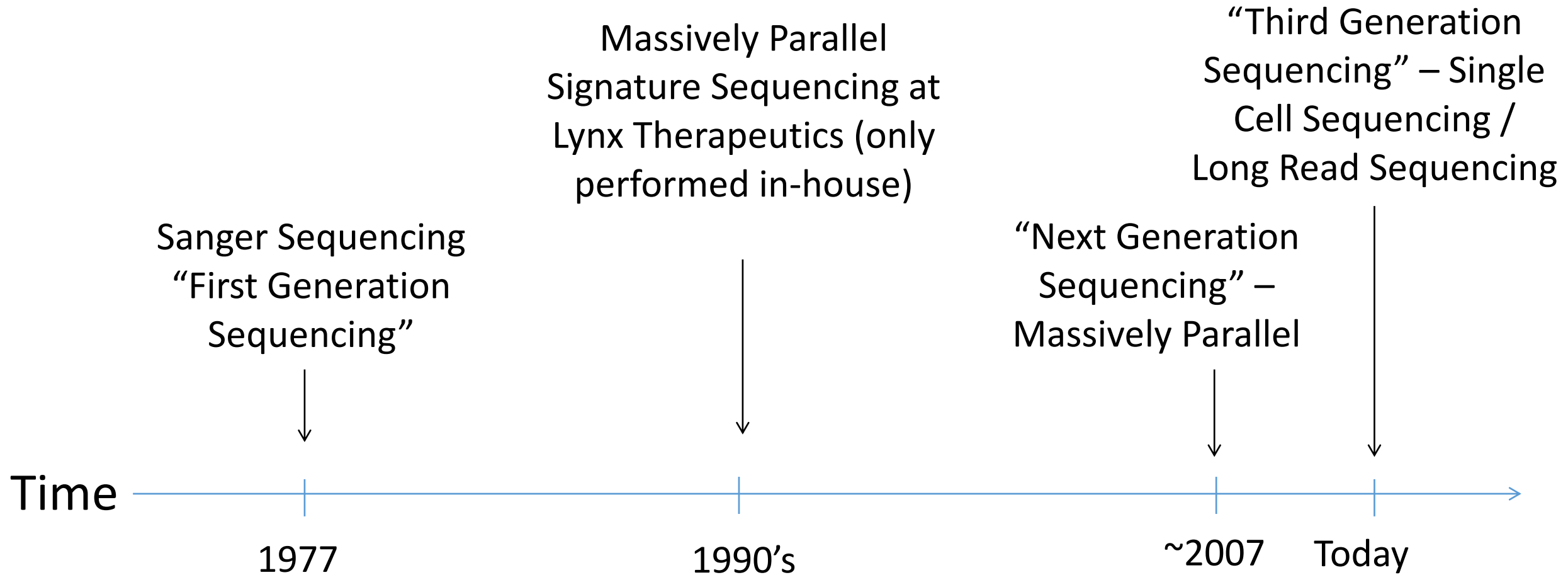
Jeffrey M. Perkel



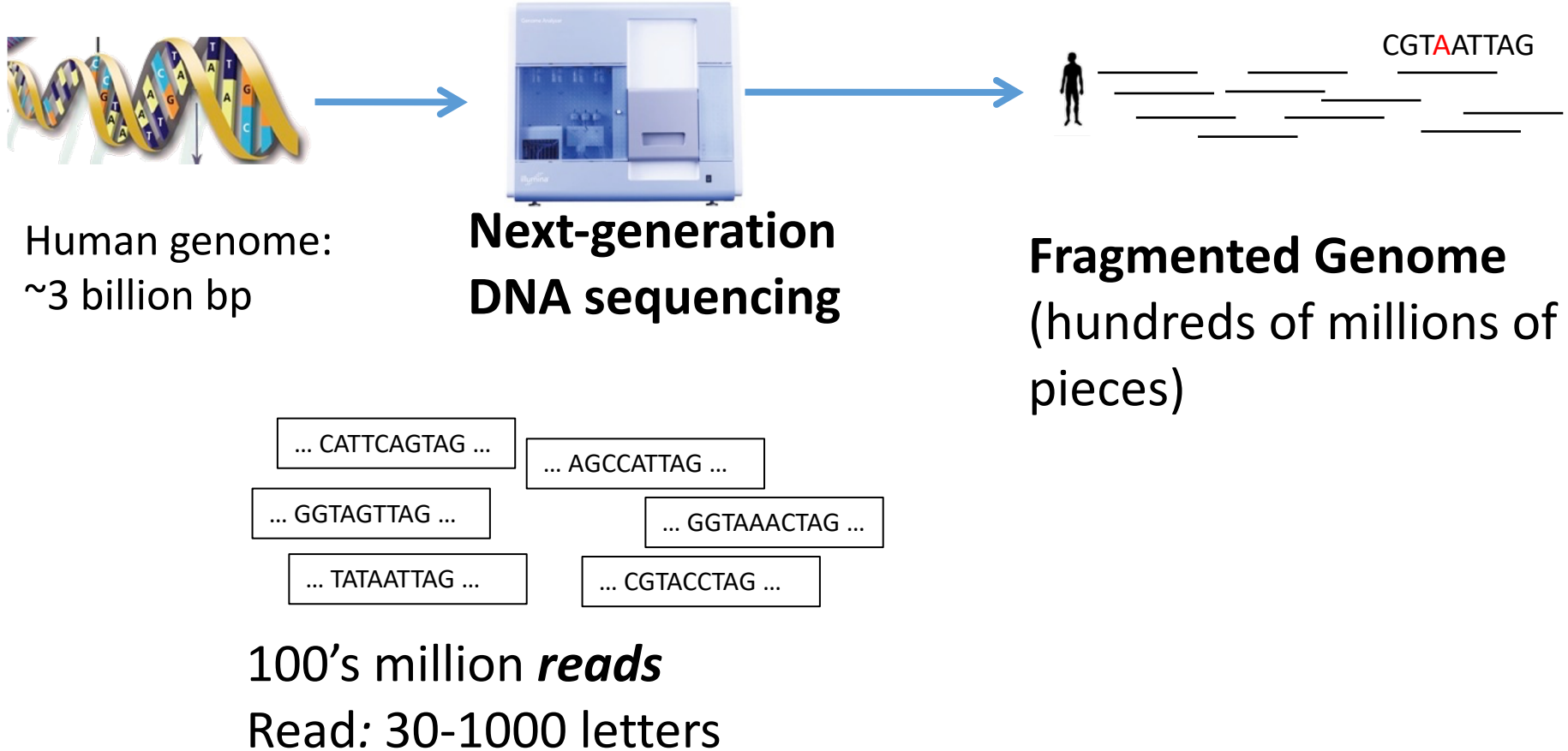
Single Cell Sequencing

- <https://www.youtube.com/watch?v=Dpq9gAg-5uA>

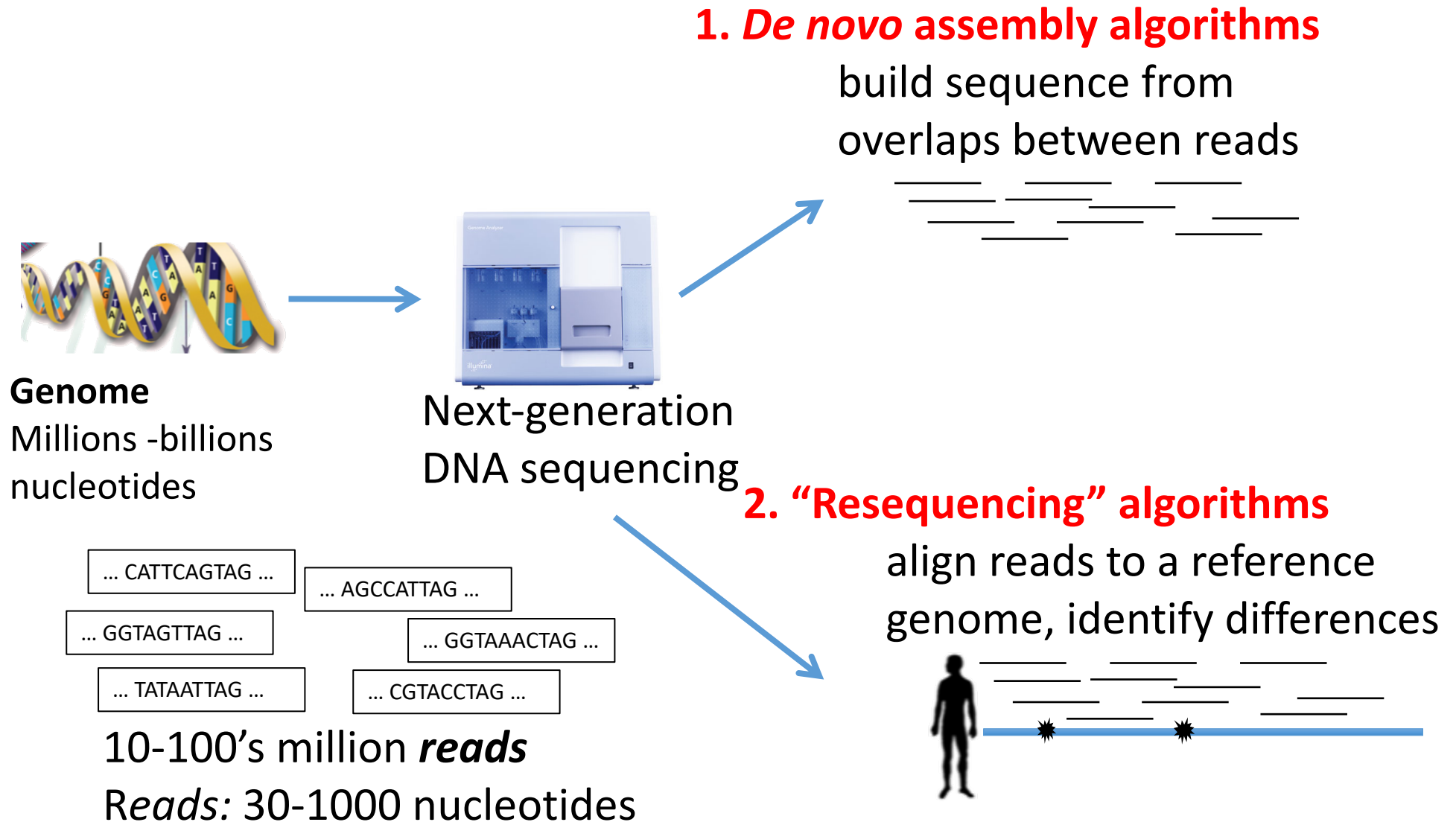
DNA Sequencing Technologies



Challenges in DNA Sequencing



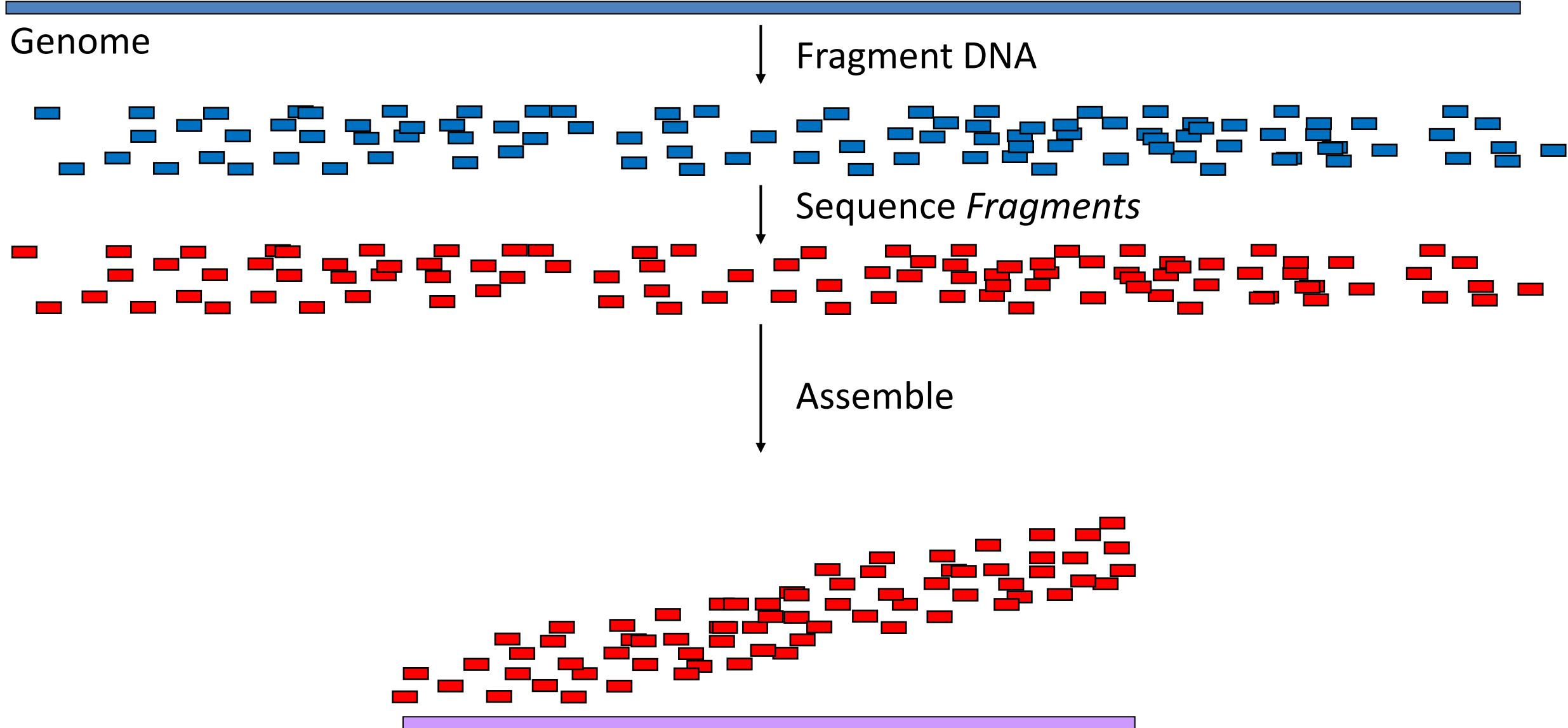
Algorithms for DNA sequencing



Issues with Reference-based Assembly

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

De novo Assembly



Let's try this out!

Let's try this out!

Four score and seven years ago our fathers brought forth, upon this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.

Let's try this out!

Four score and seven years ago our fathers brought forth, upon this continent, a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.

it is fun to have fun but you have to know how. i can hold up the cup and the milk and the cake! i can hold up these books! and the fish on a rake! i can hold the toy ship and a little toy man!