CS 466
Introduction to Bioinformatics
Lecture 24

Mohammed El-Kebir

Nov 20, 2019

# Course Announcements

HW4 due 11/22 by 11:59pm

Calendar:
- Nov 20 (Wednesday) – Genome assembly I
- Nov 22 (Friday) – Genome assembly II
- Nov 27 (Wednesday) – Fall break
- Nov 29 (Friday) – Fall break
- Dec 2 (Monday) -- HW 5 released
- Dec 4 (Wednesday) – Project presentations
- Dec 6 (Friday) – Project presentations
- Dec 11 (Wednesday) – Review session
- Dec 16 (Monday) – Final exam
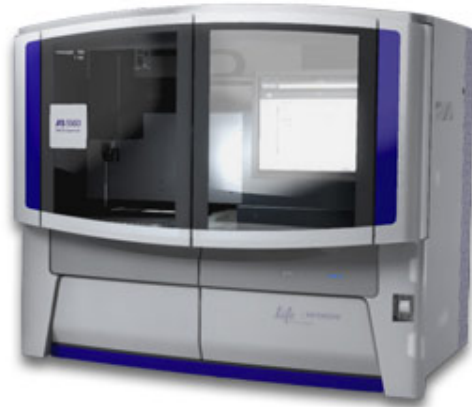- Dec 22 (Sunday) – Project report

Single Cell Sequencing
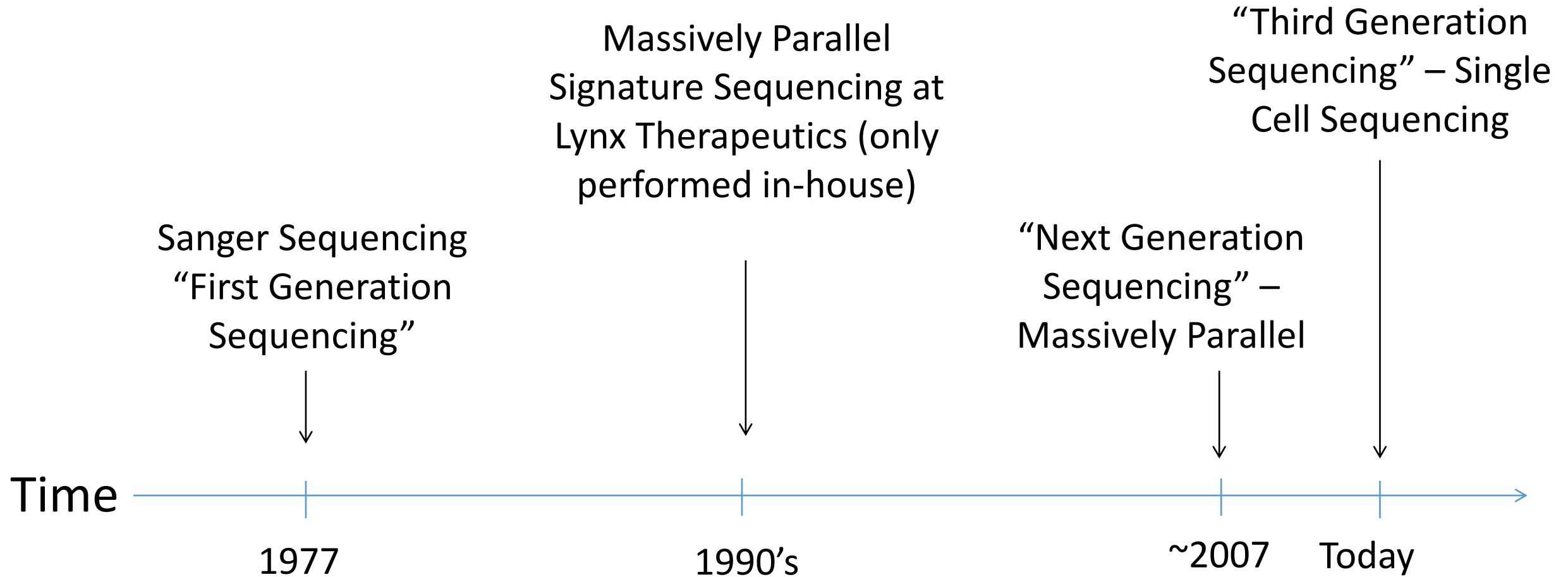
- https://www.youtube.com/watch?v=Dpq9gAg-5uA

DNA Sequencing Technologies

Next-Gen Sequencing!
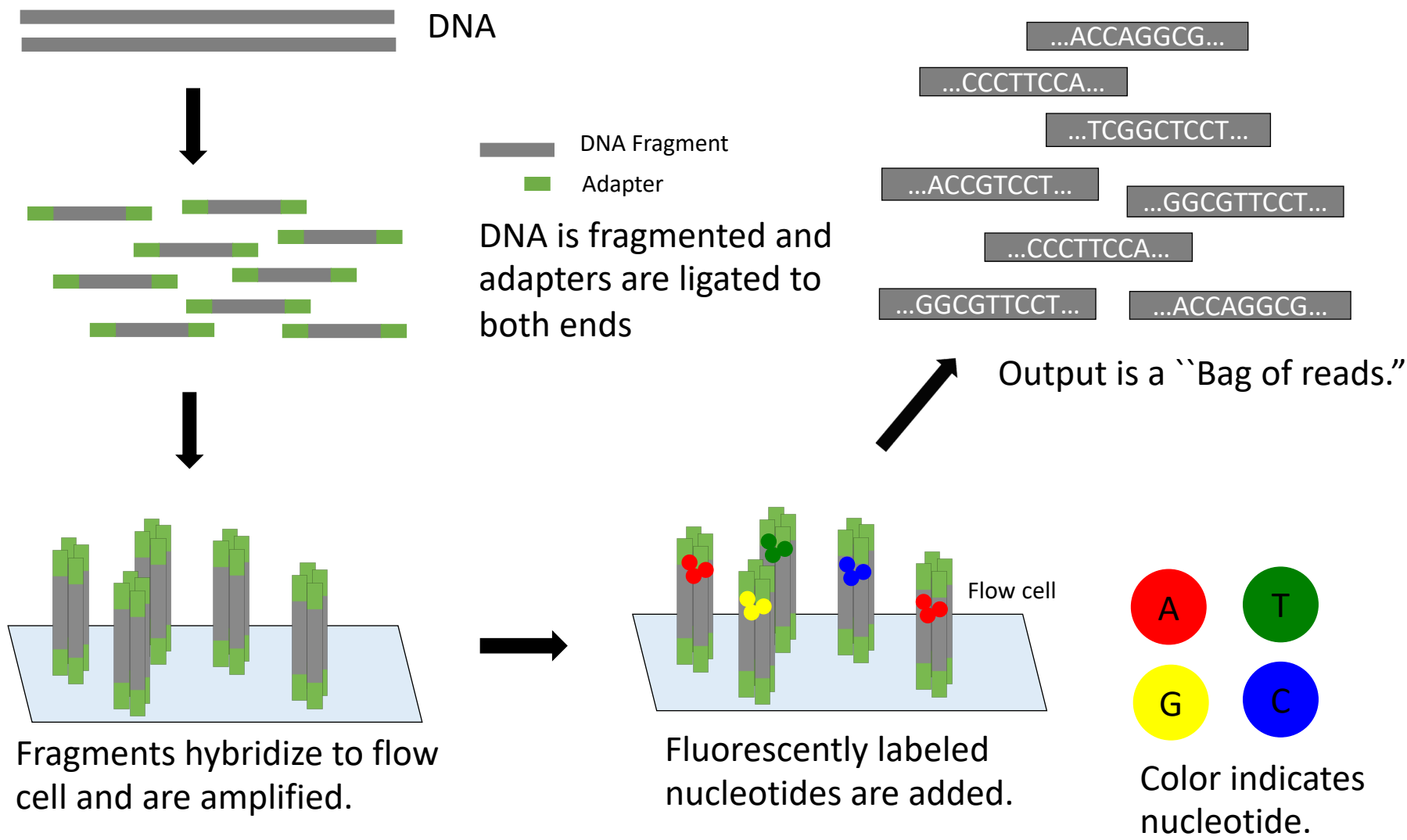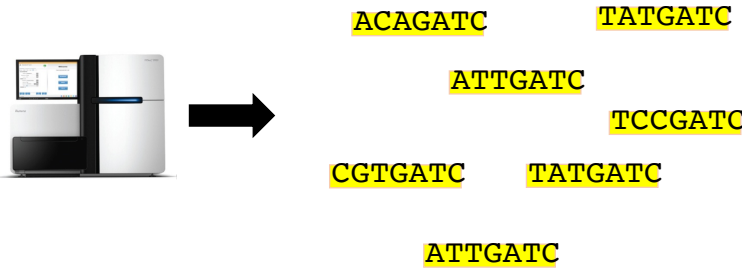
Two Different Protocols

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*

ACAGATC   TATGATC

ATTGATC

TCCGATC

CGTGATC   TATGATC

ATTGATC

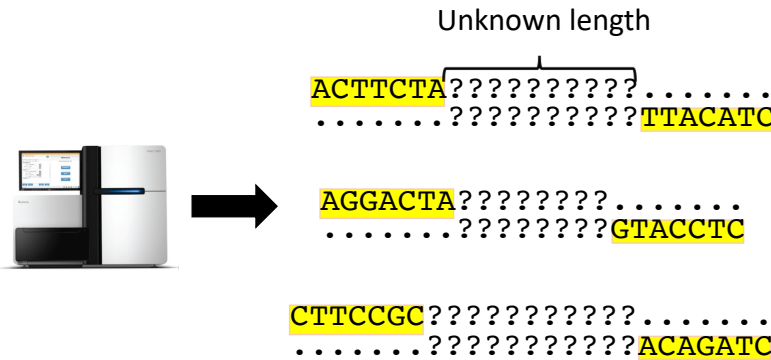## Paired-End Sequencing

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

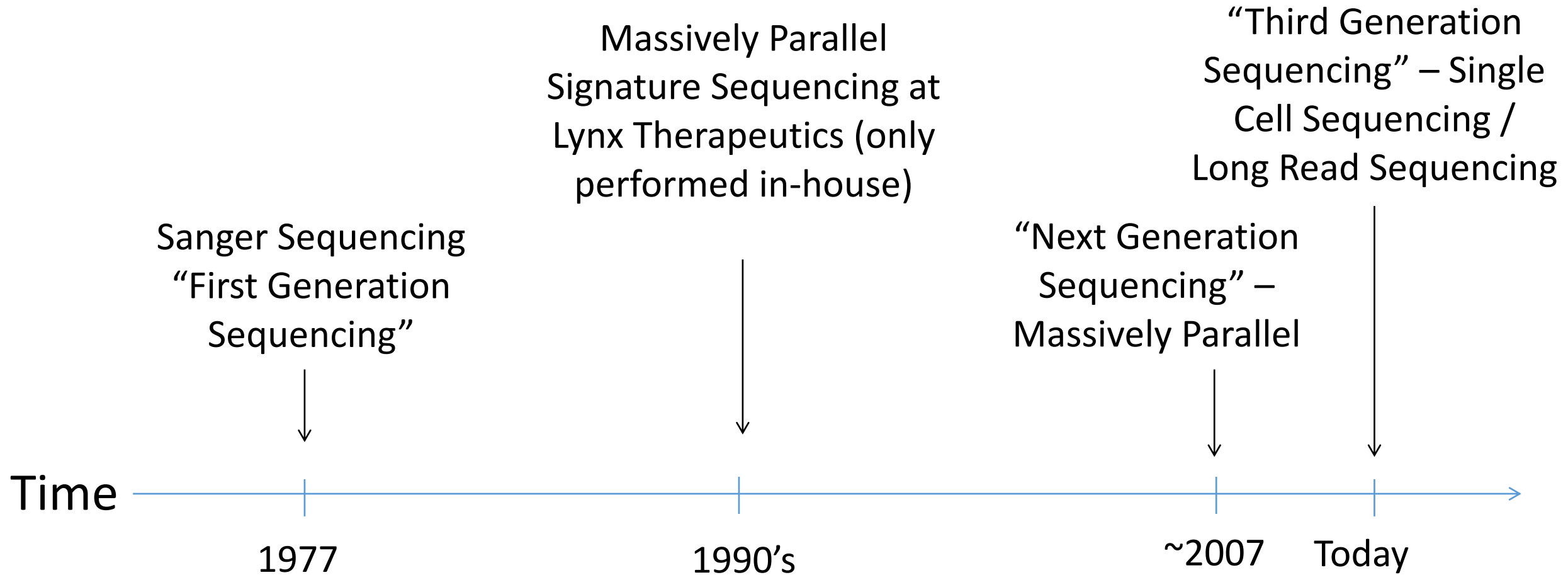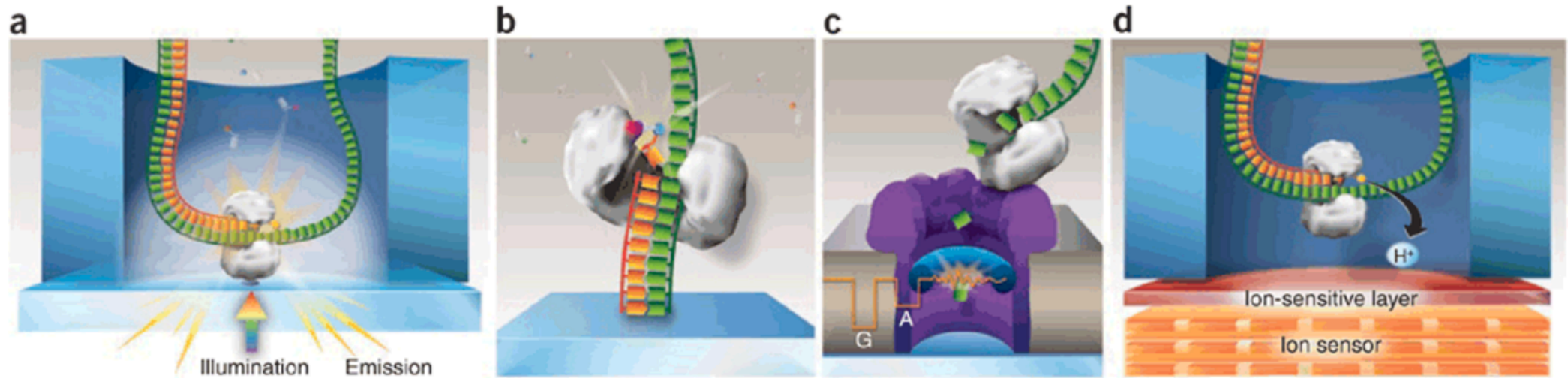|	HiSeq 2500	NextSeq 500	MiSeq	HiSeq X Five	MiniSeq
Read Length	2 x 250 bp	2 x 150bp	2 x 250bp	2 x150bp	2 x 150bp
# Reads: 1 Flow Cell	~ 300 Million	~ 400 Million	24-30 Million	3 Billion	44-50 Million
Run Time:	60 hrs	~29 hrs	~39 hrs	< 3 days	~24 hrs
Output:	125 – 150 Gb	100 – 120 Gb	7.5-8.5 Gb	800-900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

illumına®

~ 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

Massively Parallel Signature Sequencing at Lynx Therapeutics (only performed in-house)

"Third Generation Sequencing" – Single Cell Sequencing / Long Read Sequencing

Sanger Sequencing "First Generation Sequencing"

"Next Generation Sequencing" – Massively Parallel

Time

1977   1990's   ~2007   Today

Single Molecule Real Time (SMRT)

Video

PACBIO®

~ 15% error rate

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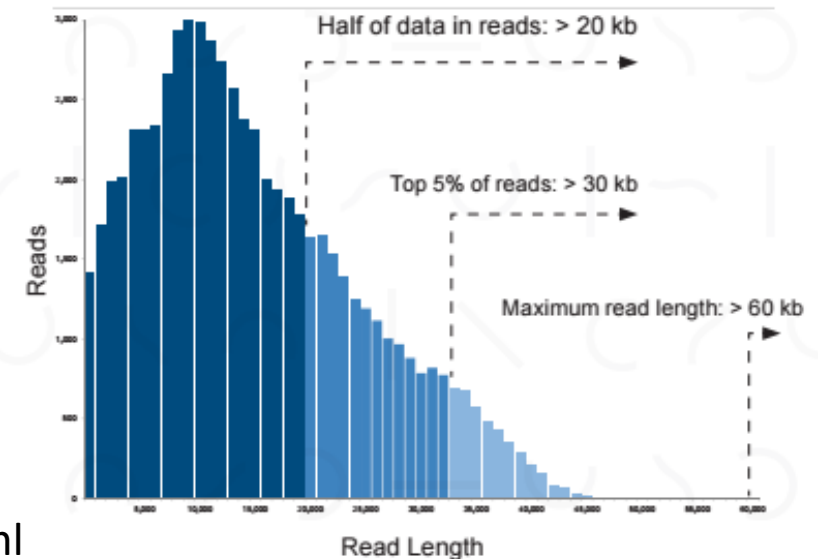
Nanopore Technologies

MinION

Portable, real-time biological analyses
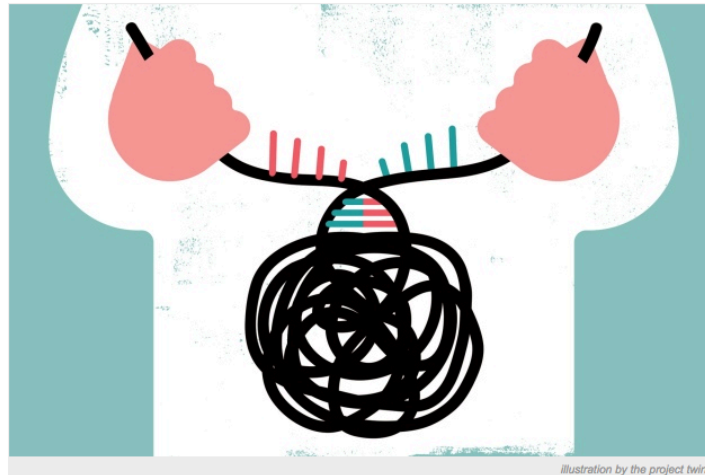
Single Cell Sequencing

NATURE | TOOLBOX

Single-cell sequencing made simple

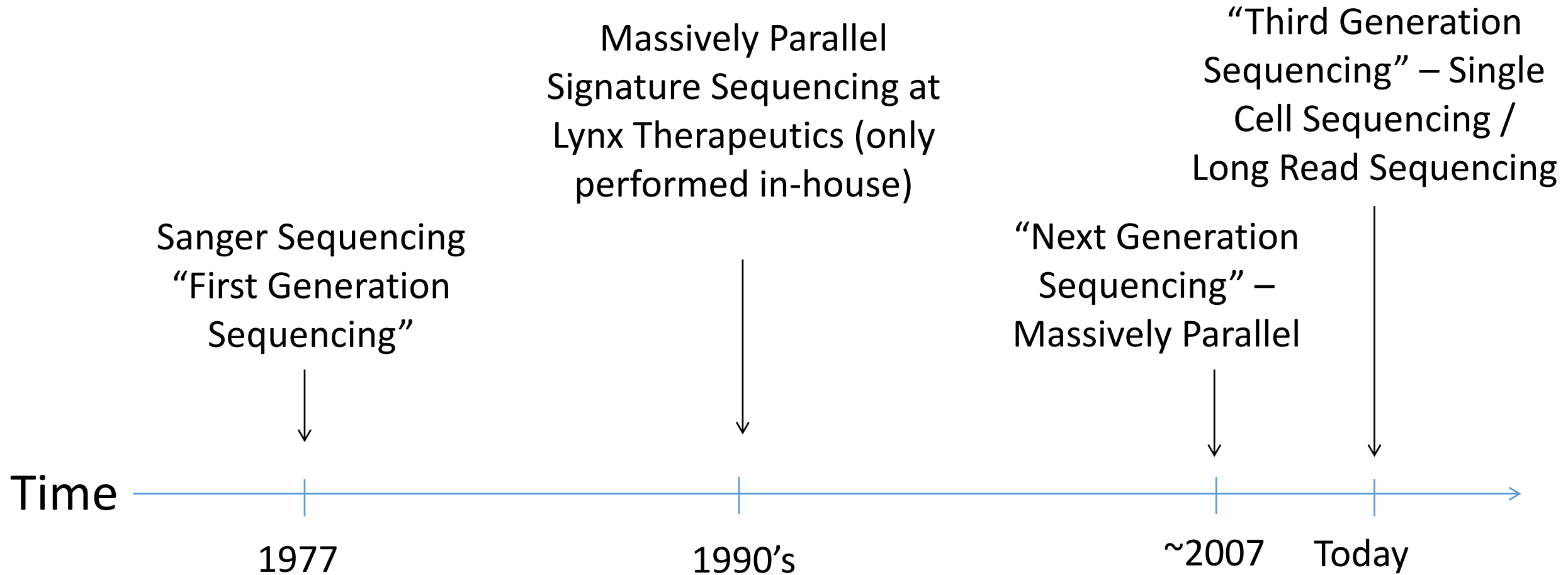Data from thousands of single cells can be tricky to analyse, but software advances are making it easier.

Jeffrey M. Perkel

illustration by the project twins

Single Cell Sequencing

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

DNA Sequencing Technologies

Challenges in DNA Sequencing

Human genome:
~3 billion bp

**Next-generation
DNA sequencing**

Fragmented Genome
(hundreds of millions of
pieces)

CGTA**A**TTAG

… CATTCAGTAG …

… AGCCATTAG …

… GGTAGTTAG …

… GGTAAACTAG …

… TATAATTAG …

… CGTACCTAG …

100's million *reads*
Read*: 30-1000 letters
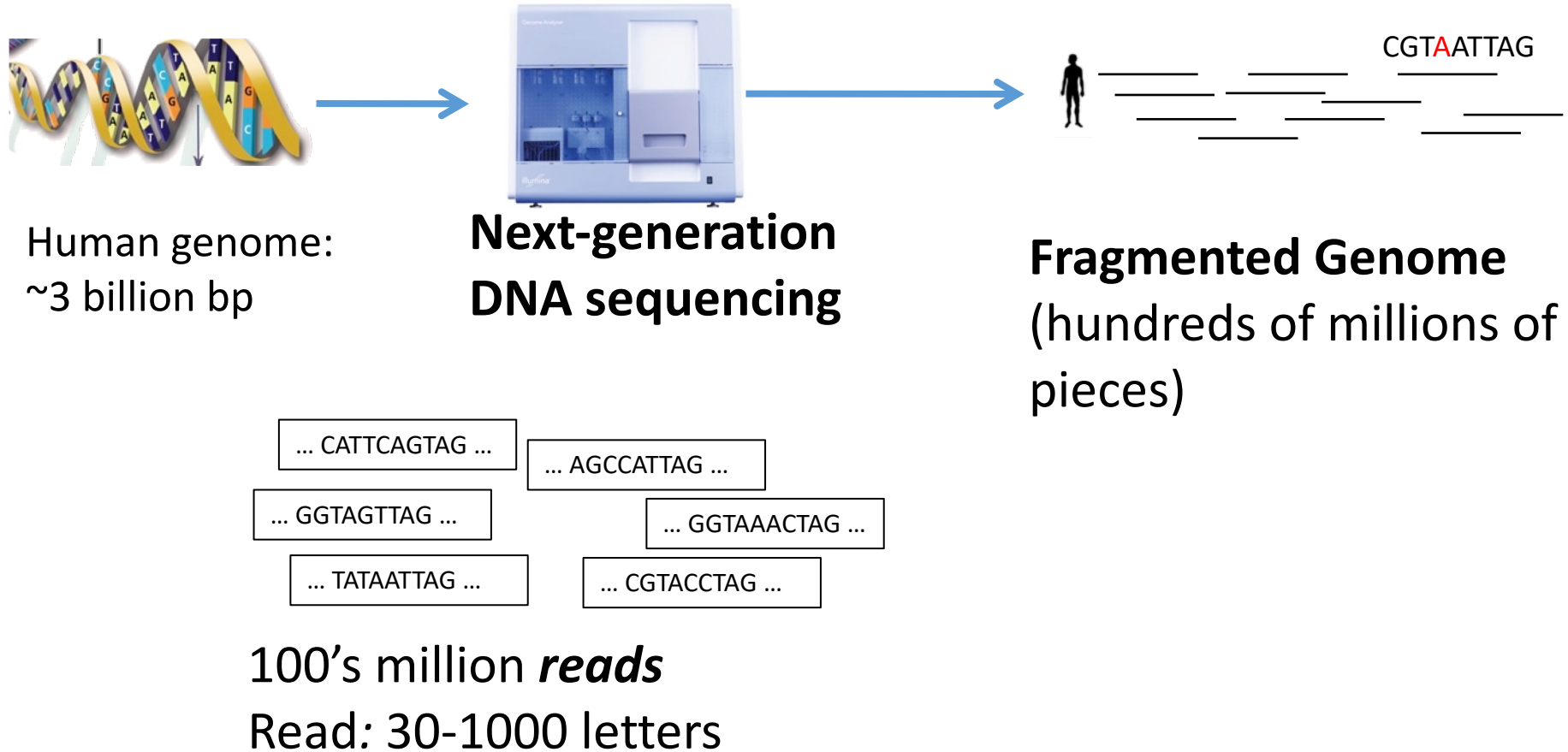
Algorithms for DNA sequencing

1. *De novo* assembly algorithms
build sequence from
overlaps between reads

Genome
Millions -billions
nucleotides

Next-generation
DNA sequencing

2. "Resequencing" algorithms
align reads to a reference
genome, identify differences

... CATTCAGTAG ...

... AGCCATTAG ...

... GGTAGTTAG ...

... GGTAAACTAG ...

... TATAATTAG ...

... CGTACCTAG ...

10-100's million ***reads***
R*eads:* 30-1000 nucleotides
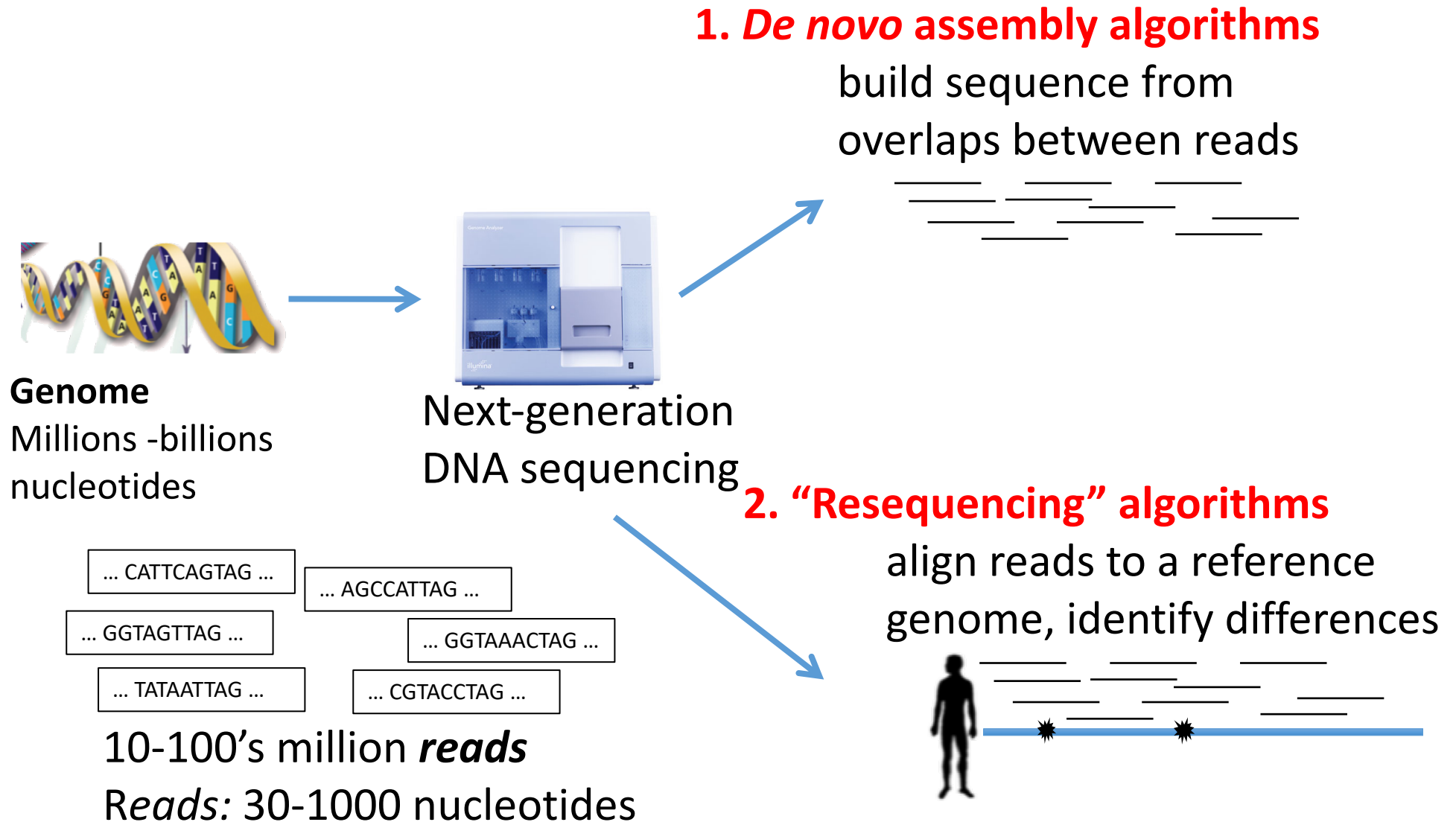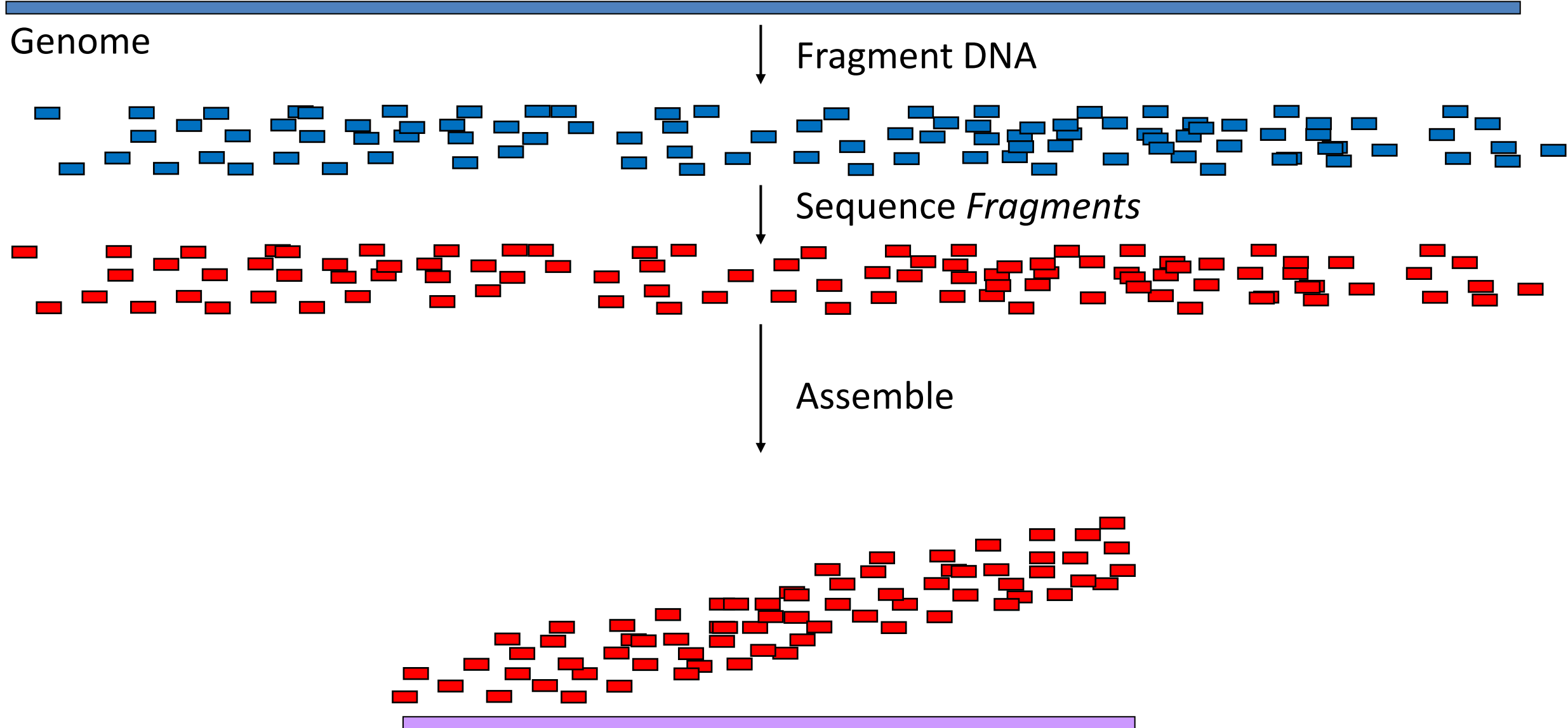
Issues with Reference-based Assembly

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

De novo Assembly

Genome

Fragment DNA

Sequence *Fragments*

Assemble

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!