Paper Discussion Assignments

- Go to BBLearn Course Content homepage for this class
- Click on “Paper Presentation”
- Here is where you upload your Power Point for grading.
- Marc will go over the rubric

DNA Sequencing at 40: Past, Present, and Future

Shendure et al. (2017) *Nature*

Comparative Genomics

The goal of this paper:
To review the history of DNA sequencing

1. The underlying technologies
2. The breadth of problems for which it has proven useful

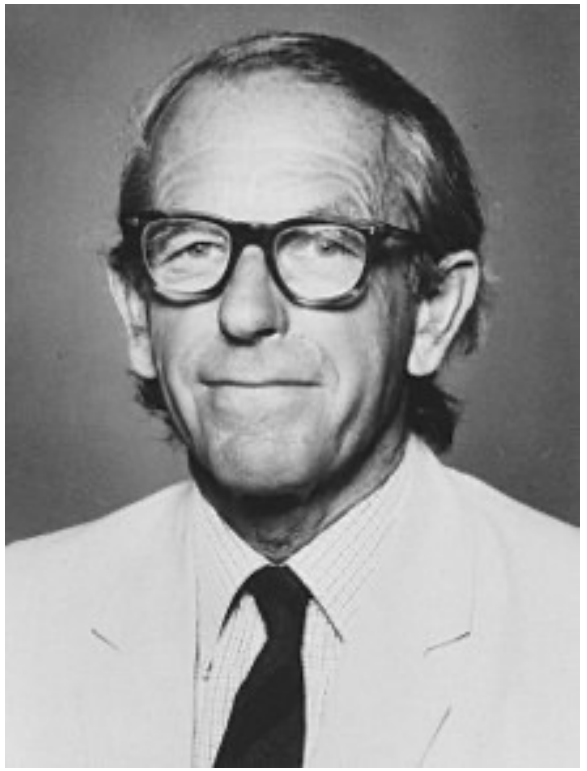
Early Protein Sequencing

In the early 1950s, Frederick Sanger determined the protein sequence for insulin.

Sanger showed that proteins had defined patterns of amino acid residues

By the late 1960s, the process of Edman degradation allowed scientists to decipher many proteins.

They had determined that ***each protein sequence varied across species and between individuals.***



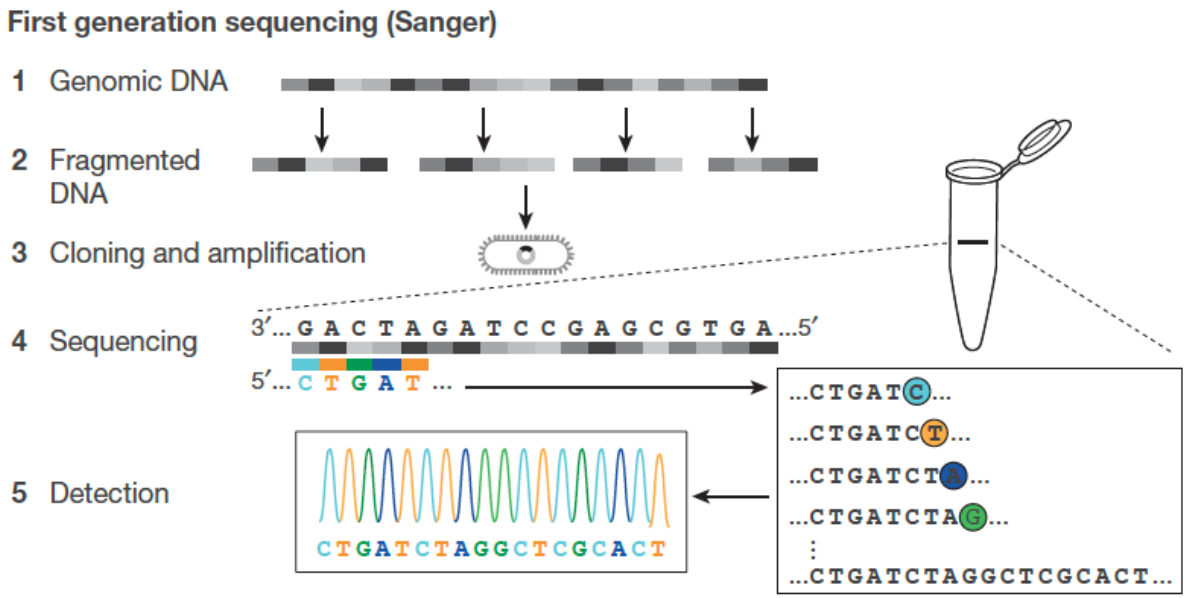
F. Sanger (1918-2013)
Nobel Prize in Chemistry (1958 and 1980)

DNA Sequencing

Initially very slow – in 1973 it took 2 years to sequence 24 bases of the lactose-repressor binding site.

In 1976, Sanger developed the ***chain terminator procedure*** which used distances along a DNA molecule from a radioactive label to positions occupied by each base to determine nucleotide order

Shotgun sequencing – sequencing of random clones followed by sequence assembly based on overlaps – was used to assemble bacteriophage-lambda in 1982.



The Human Genome Project

A ‘hierarchical shotgun’ strategy:

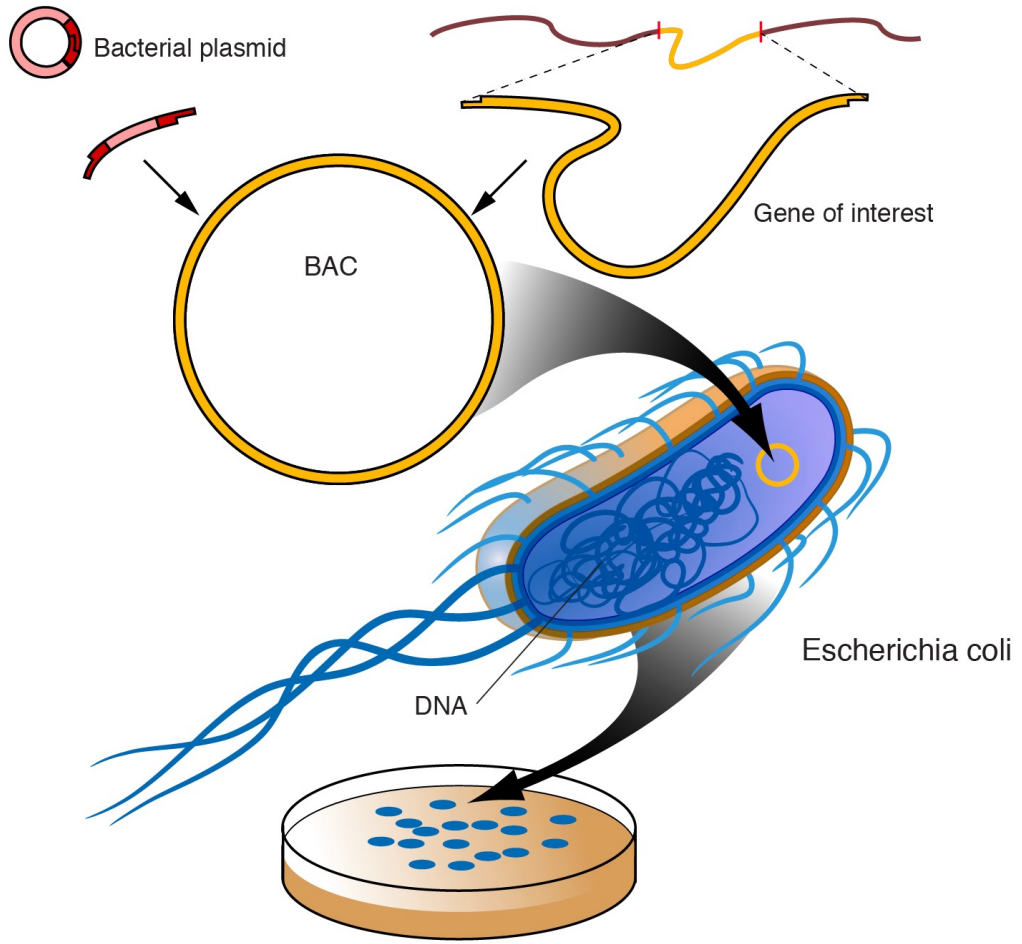
Large fragments of the human genome were cloned into bacterial artificial chromosomes (BACs)

Clones were picked and grown on a dish, and the DNA was isolated

Purified DNA was used as a template for Sanger sequencing.

The sequenced BAC clones were compared against genetic and physical maps to determine their order on the human chromosomes.

Fragments by this time were ‘paired end’ – sequenced from both sides so were able to be linked together.



The Race to Sequence the Human Genome

Craig Venter, Francis Collins, Bill Clinton

https://www.nature.com/articles/d41586-020-01849-w

Next Generation Sequencing: Illumina

Genomic DNA is fragmented

Adapters of known sequence are ligated (connected) the the fragments

Adapters are used to bind fragments to a flow cell

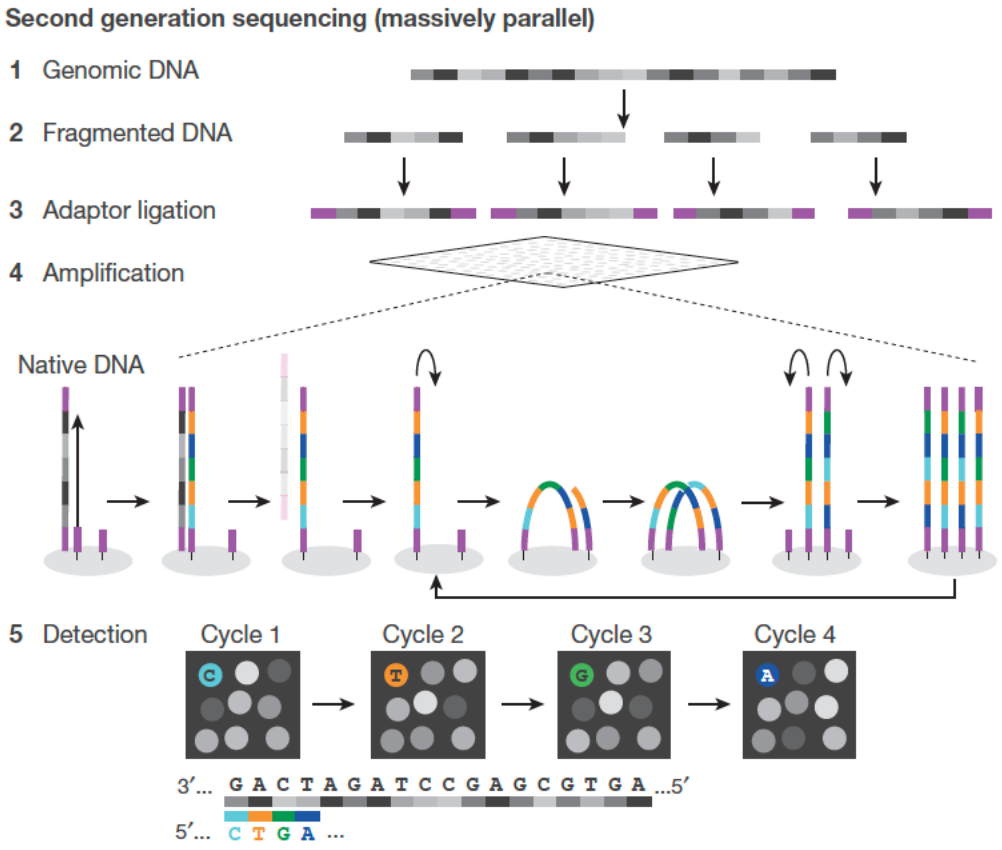
DNA polymerase copies the fragments

The fragments are amplified (copied many times over)

Sequencing by Synthesis: Dye-labeled terminal ends of fragments give off light

The wavelengths are used to determine the bases as they are synthesized by the polymerase

The key innovation here is ***multiplexing*** – the ability to pool together many fragments from different libraries and sequence them all at basically the same time.



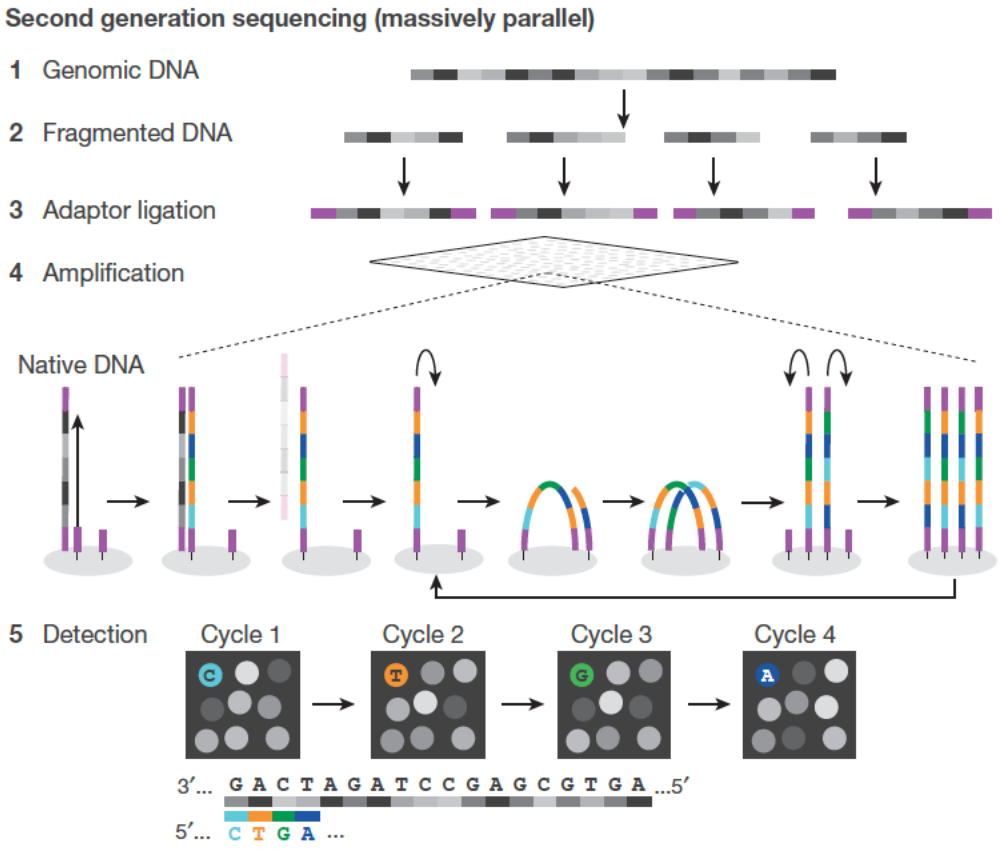
Next Generation Sequencing: Illumina

Read lengths are in the low hundreds of bases – still shorter than Sanger sequences

Base calls are ~99.9% accurate, based on phred scores

Over a billion bases can be generated by a single person running the instrument

Cost is brought down to a few thousand dollars



Single Molecule Sequencing

Previous methods rely on amplification of a template sequence.

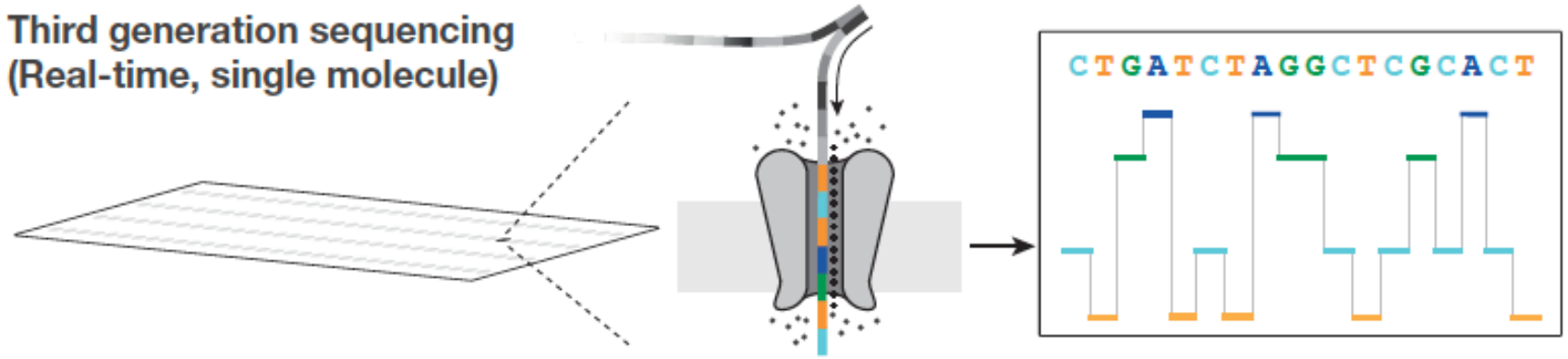
This leads to copying errors, sequence-dependent biases and information loss (such as methylation)

PacBio – optically observes polymerase mediated synthesis in real time.

A synthesized DNA strand is pulled through a zero-mode waveguide which is less than half a wavelength of light.

The growing DNA chain emits signals that are called in real time.

Very long reads – 10kb – 100kb.

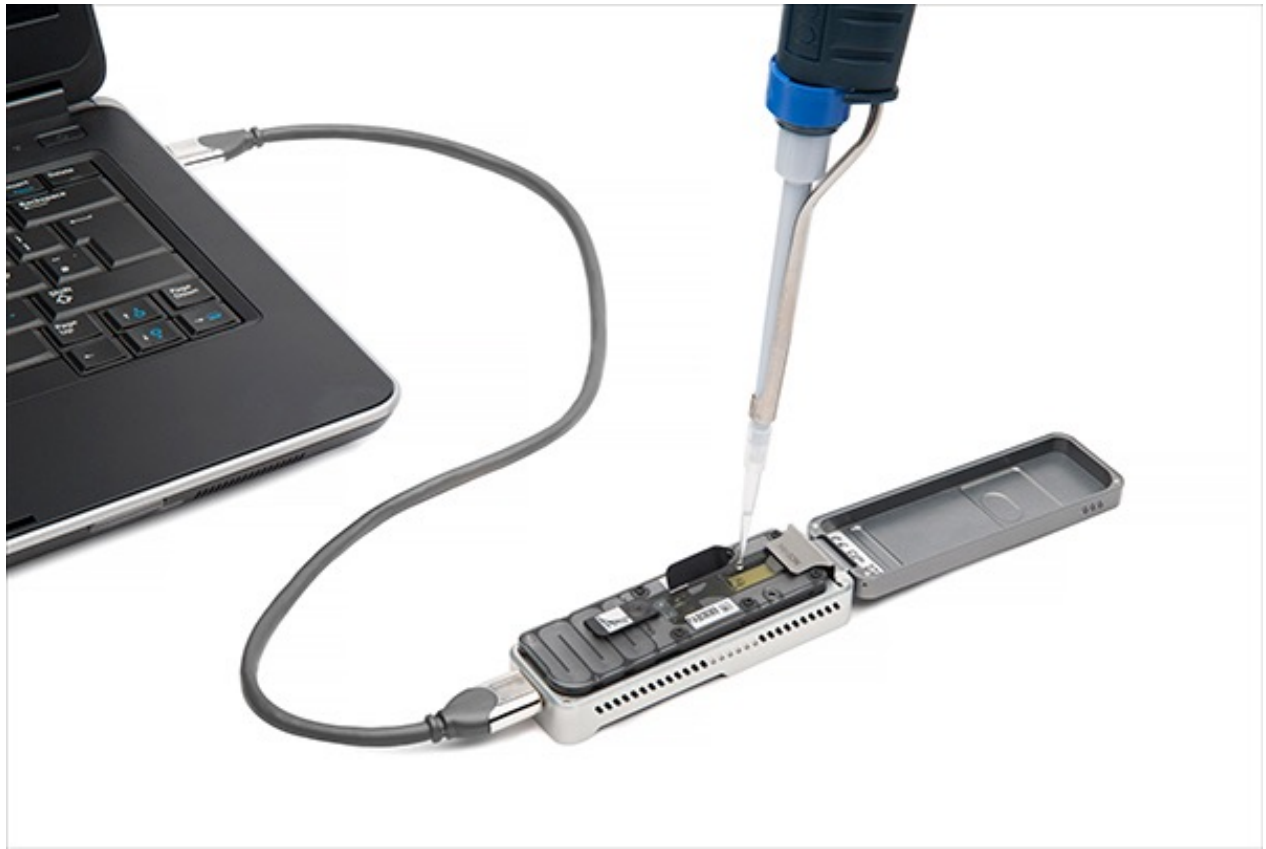


Single Molecule Sequencing

Oxford Nanopore

Reliance on electrical signal allows very small instruments

USB plugin sequencing!



Applications of DNA Sequencing

De novo genome assembly

Instead of relying on paired-end information, long reads will resolve repetitive regions

HMW fragments allow long-range mapping (Hi-C, optical mapping)

Resequencing

'$1,000 human genomes'

Higher resolution of copy number and structural variants

Clinical applications

Non-invasive prenatal testing

Diagnosing Mendelian and neurodevelopmental disorders

Molecule counting

Characterizing the amount of transcripts produced by gene expression

Metagenome sequencing

Environmental and human microbiomes

The Future of DNA Sequencing

Genome diversity

Chromosome-scale assemblies for 1000s of species

Population-scale resequencing

Millions of humans sequenced, discovery of new *de novo* mutations

Developmental biology

Single-cell sequencing and understanding spatial context

Real-time, portable sensors

Monitoring affects of environment on nucleic acids

Unconventional uses

Large amounts of data can be encoded in synthetic DNA