Sequencing Strategies

Lucas Bishop, M.S.
Nevada Bioinformatics Center

Sanger Sequencing

- First developed in 1977 by Fred Sanger et al. which they used to sequence the first full genome: that of PhiX174.
- Early on, 4 reactions were set up, with one of the modified nucleotides and 3 normal ones. Sequence viewed on gel.
- Method is based on modified nucleotides (ddNTP) that are missing an -OH group. This stops elongation directly after addition.
- Was then developed into dye-terminator sequencing where each ddNTP had a unique fluorescent label and could be done in one reaction, sequence visualized by chromatogram.

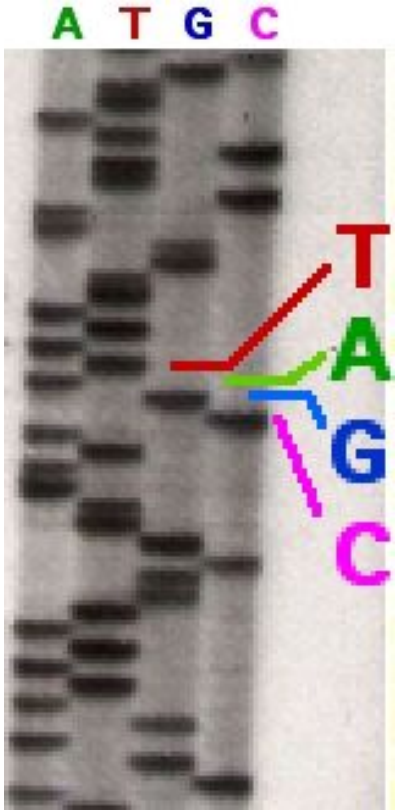
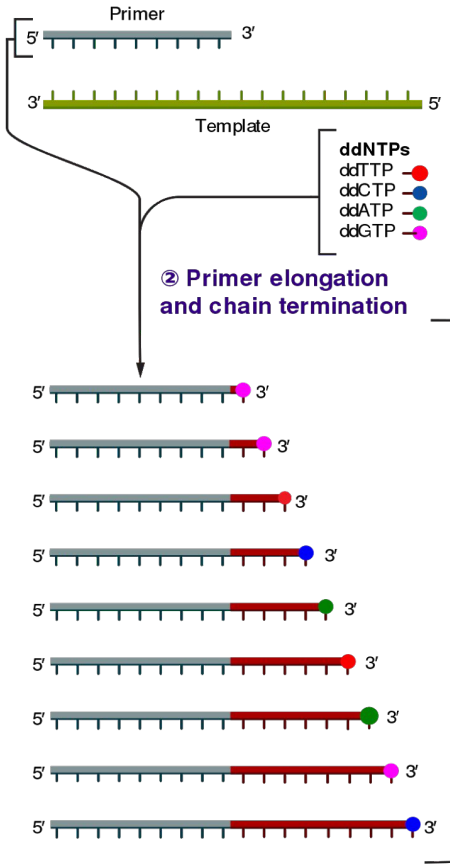
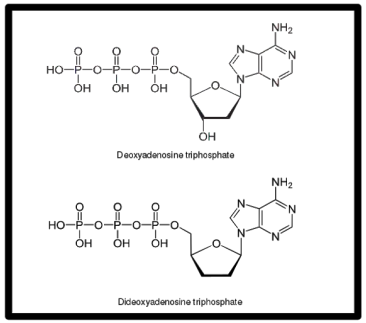


Image courtesy of John Schmidt, https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Sequencin

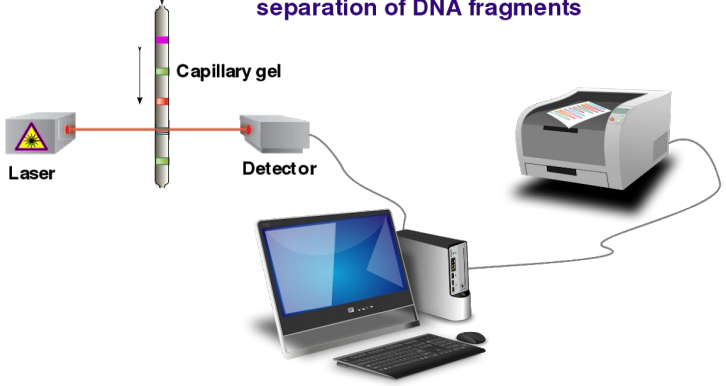
① Reaction mixture

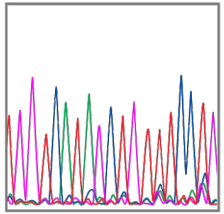
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)





③ Capillary gel electrophoresis separation of DNA fragments





④ Laser detection of flourochromes and computational sequence analysis

Sequencing of individual genes

- As researchers begin to implement sanger sequencing, whole genomes are still unattainable, but individual, gene-level differences can be seen!
- SNP’s and gene-level mutants (insertions/deletions)
- Problem: the fluorescence of the same dNTP in a row can’t always be seen. example) the difference in fluorescence between AAAA and AAAAA is not discernible.
- Problem: The importance of sequencing depth and read length.
    - Depth/coverage= the percentage of reads that map to a reference sequence

Human Genome Project (HGP)

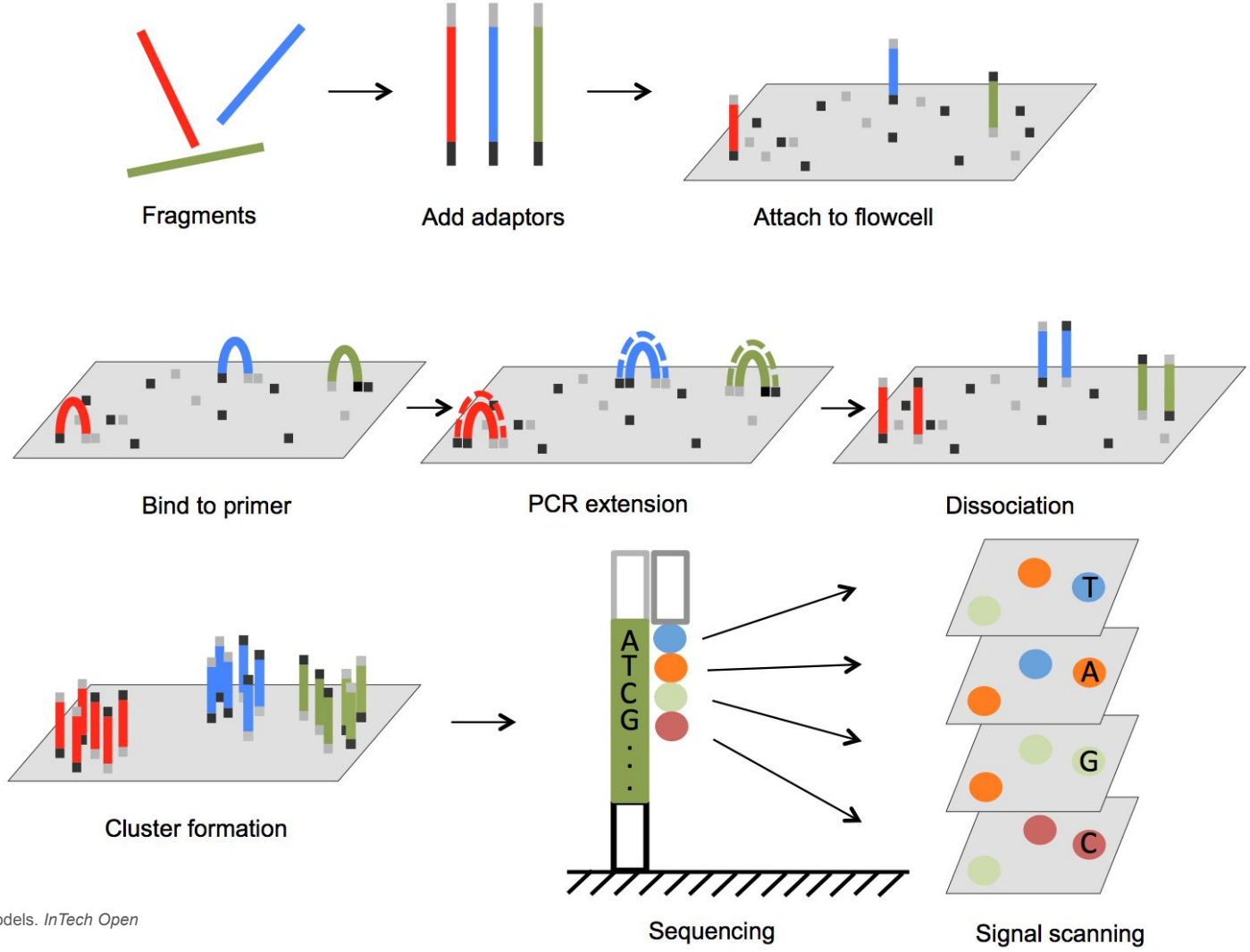
- Human Genome Project: International (US, Japan, UK, China to name a few), multi-year project with aim of mapping all nucleotides in *Homo sapiens* genome.
- Shotgun sequencing: sheering genome(150k bp), ligated into BAC, replicated in Bacteria, sequenced, aligned and assembled.
- Draft completed in 2001(83%); Project deemed complete in 2004.
- Marred by controversy: Craig Venter and Celera Genomics
- Paved way for Human Genetics research; genetic variants underlying genetic disease, etc.

(www.genome.gov)

NGS - short read / Illumina

- By massively parallelizing a sequencing-by-synthesis approach onto a chip (Illumina) or beads (454), sequencing 500-600 million bp with read lengths ~500 became possible.
- Advantages:
  - it offers single-nucleotide resolution, allows for more complex workflows: detecting related genes (or features), alternatively spliced transcripts, allelic gene variants, SNP's, etc.
  - Doesn't have the repeated nucleotide issue.
  - Much less DNA input. Technically and financially efficient.
  - Reproducibility.

Sequencing-By-synthesis: Illumina Platforms

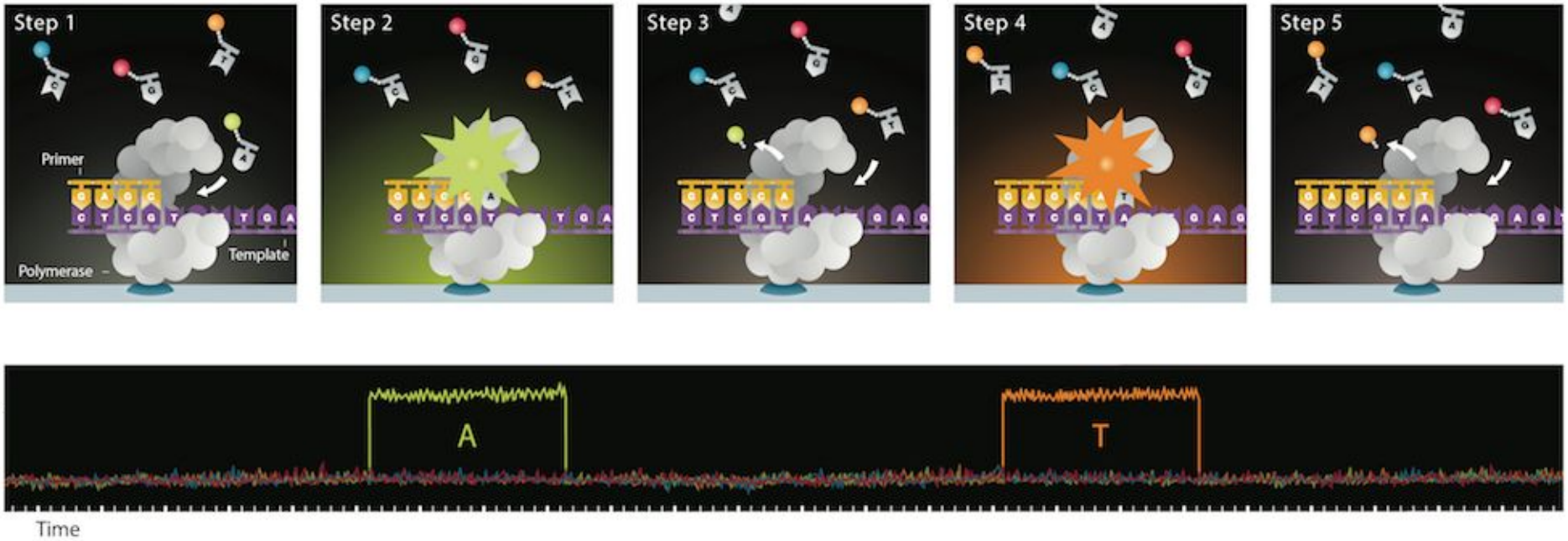


Yu et al. 2016. Next Generation Sequencing in Aquatic Models. *InTech Open*

NGS - long read / PacBio & Oxford Nanopore

- Illumina is really good at sequencing small amplicons (<500bp), many times.
- Systems were needed to sequence long reads, particularly for whole genome sequencing.
- How? Both use variations of a polymerase bound to a small pore.
- PacBio SMRT system: good for genomes, longest reads possible. 25,000+ bp reads. Measures fluorescence within the pore.
- MinION/GridION: Measures miniscule pH changes within pore.No fancy fluorometer, so much cheaper. Portable, can be used in field work or traveling diagnostics.

Single Molecule Real Time (SMRT) sequencing



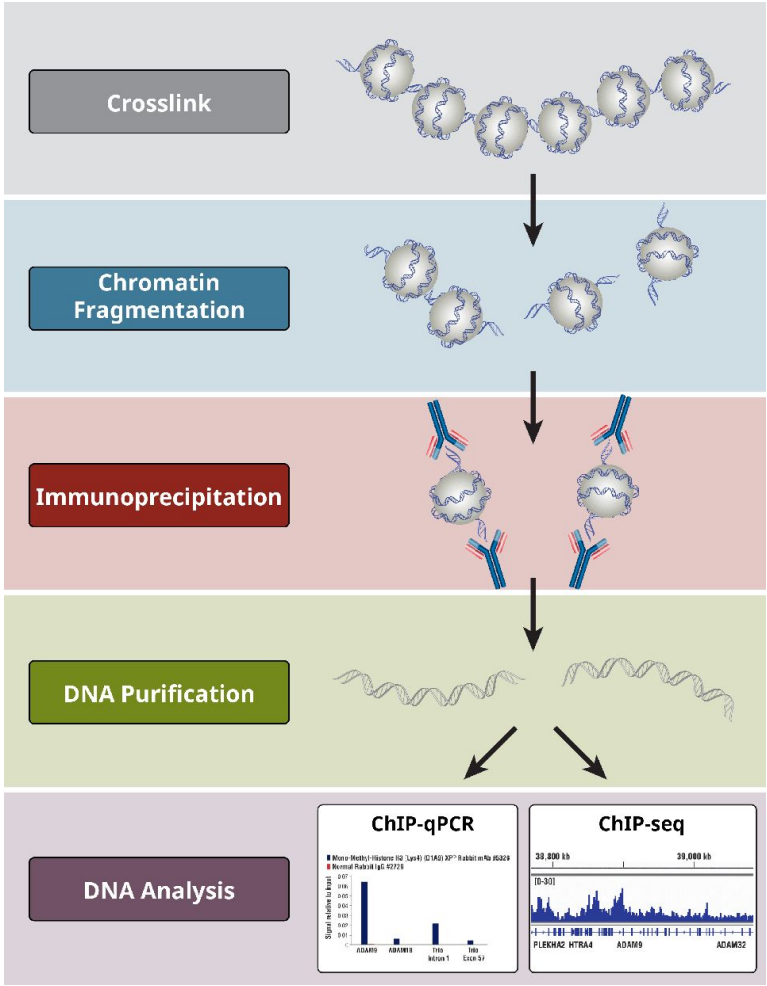
RNA-seq

- RNA-seq : Using NGS to see of the transcribed parts of an organism's DNA, was first published in 2007 by Emrich et al. when they sequenced a partial transcriptome of maize (corn) on the Roche 454 platform.
- Applications:
  - Alternative gene spliced transcripts
  - Post-transcriptional modifications
  - Different populations of RNA: total RNA, small RNA, such as miRNA, tRNA, and ribosomal profiling
  - Exon/intron boundaries
  - Single cell sequencing (scRNA-Seq)
  - **Gene expression over time and/or across different groups/treatments**
- RNA-seq analysis must be question-specific: are you interested in transcription of one gene in case vs. control or the whole transcriptome of a species?

ChIP-seq

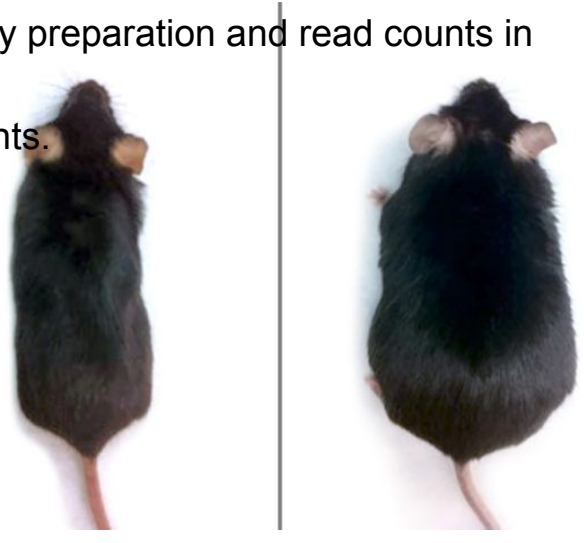
- ChIP-seq = Chromatin Immunoprecipitation sequencing. Used to identify protein-DNA interactions and sequence those portions of DNA.
- Immunoprecipitation: using antibody to pull specific protein antigen out of solution
- First example: Promoter regions bound to RNA polymerase
- Now: Using histone modification ChIP-seq to examine regulatory networks

https://www.abcam.com/epigenetics/studying-epigenetics-using-chip

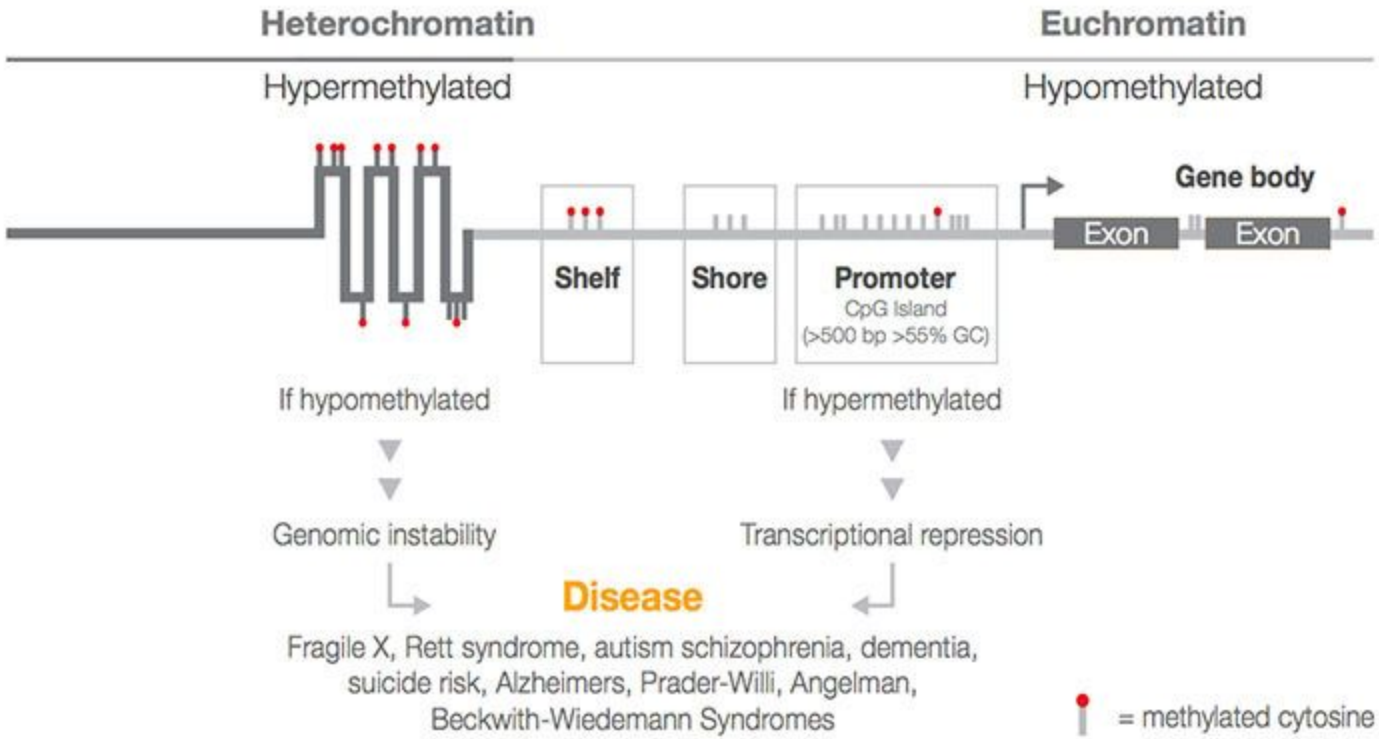


Epigenetics/methylome

- Epigenetics: Genetic expression changes acquired after birth that are passable to offspring.
  - Example calorie high- and low- diet in mice leads to metabolic phenotype in offspring
- Methylation: epigenetic modification where a methyl (CH3) group is added to DNA, often modifying function of gene/change in expression.
- How? Bisulfite sequencing
  - Bisulfite converts methylated cytosine's to uracil during library preparation and read counts in sequence data determine % of methylation.
  - Further bioinformatics determine location of methylation events.



Perturbation of Methylation



Future of Sequencing

- Significant cost reduction: in 2009 when Illumina launched a personal genome sequencing service, cost was >$40,000. Now companies like Helix and 23andMe sequence human genome for $200.
- Precision medicine: diagnoses and treatments specific to a person's genome/transcriptome/etc.

ONEOME - Analytics suite provided to healthcare providers

BillionToOne - 'omics based prenatal screening

VIOME - personalized diet/health advice based on gut transcriptome