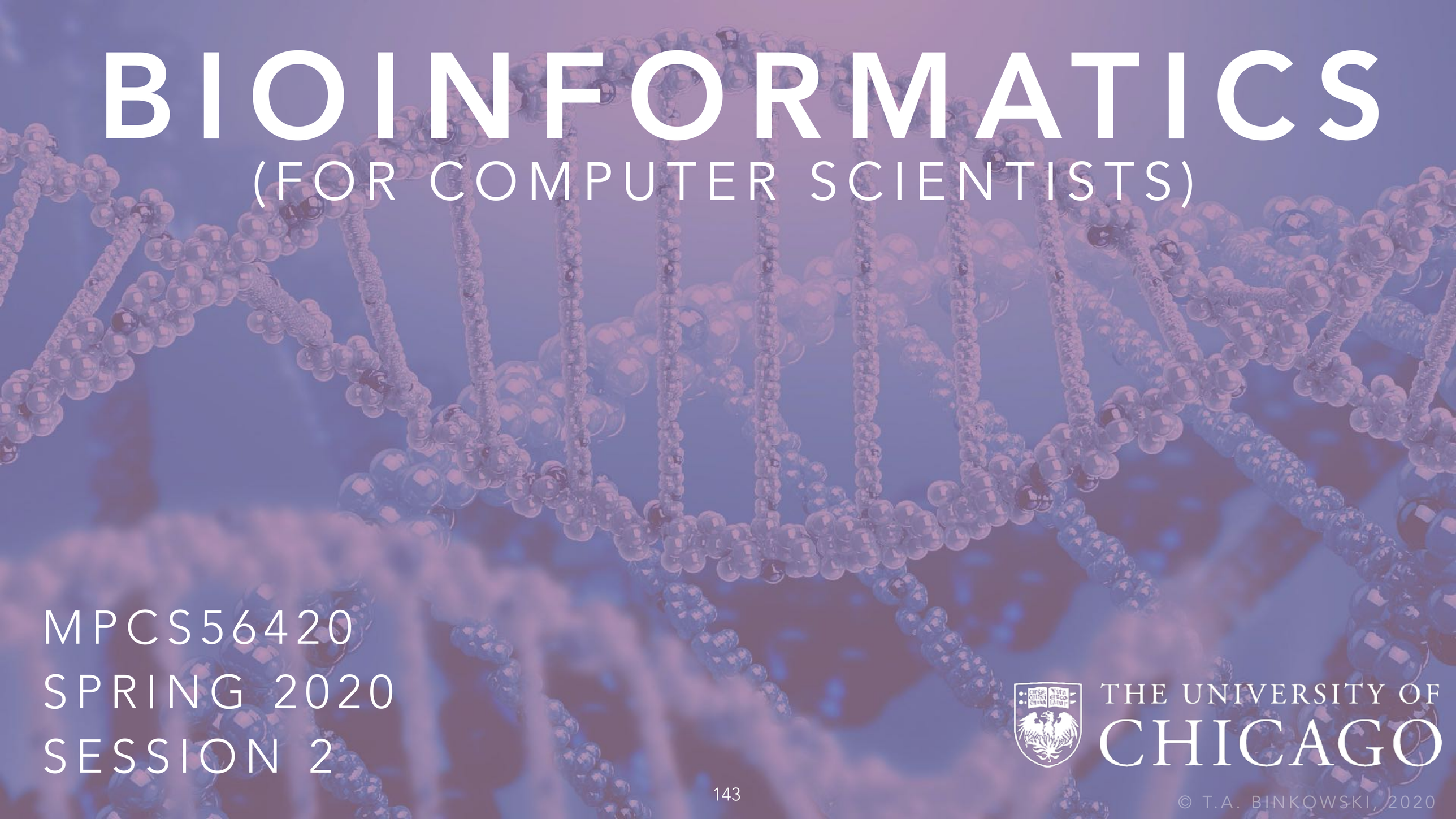


BIOINFORMATICS

(FOR COMPUTER SCIENTISTS)

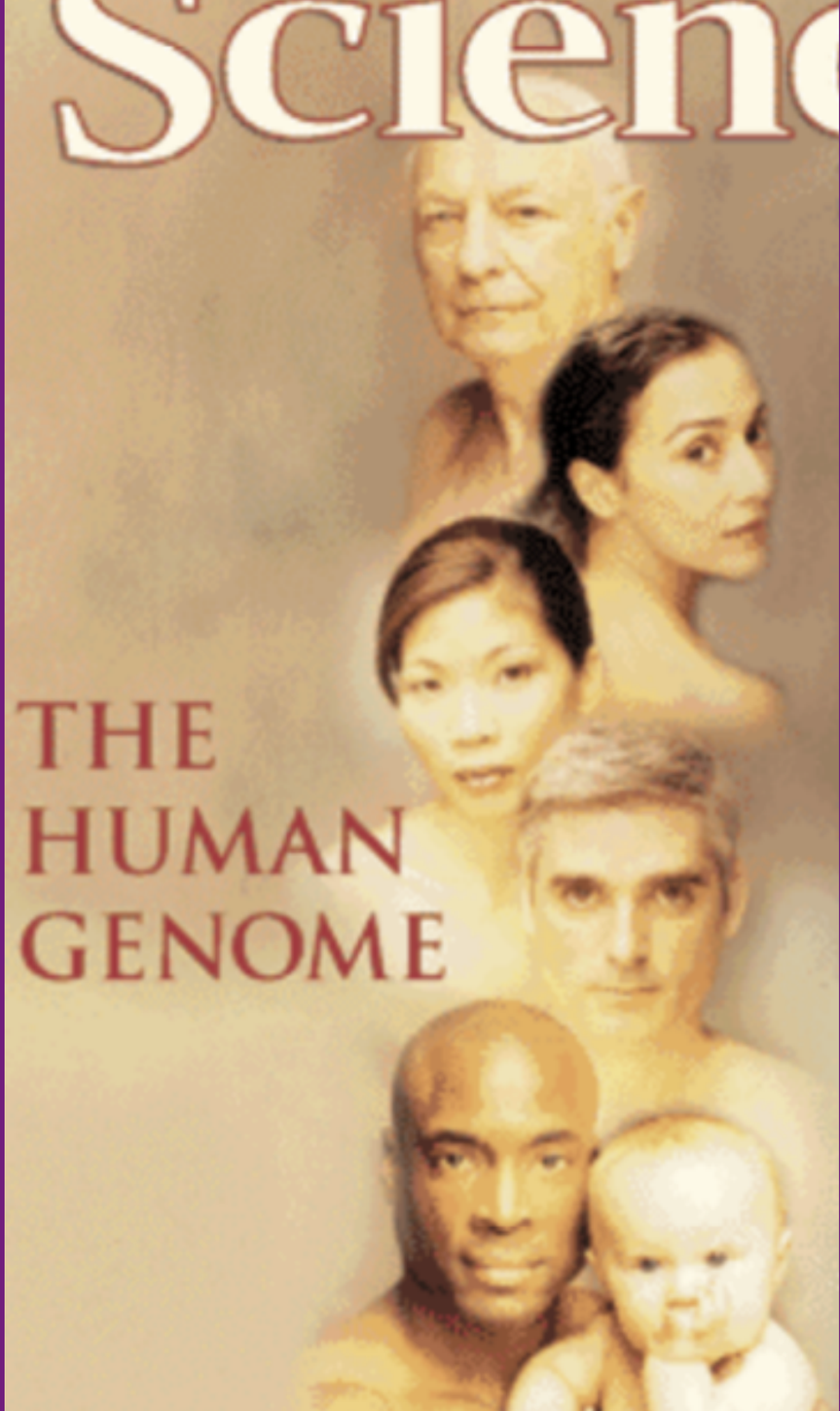
MPCS56420
SPRING 2020
SESSION 2

THE UNIVERSITY OF CHICAGO


DNA
SEQUENCING

DNA SEQUENCING

- A Brief history of sequencing
  - 1982, small viral genomes (e.g., lambda virus)
  - 1994, Haemophilus influenzae whole-genome shotgun sequencing
  - 2001, whole-genome shotgun sequencing of Human genome
    - Human Genome Project
    - Celera Genomics



DNA SEQUENCING

Scientists Finally Pronounce Human Genome

'It's Gatcaatgaggtggacaccagaggc...'

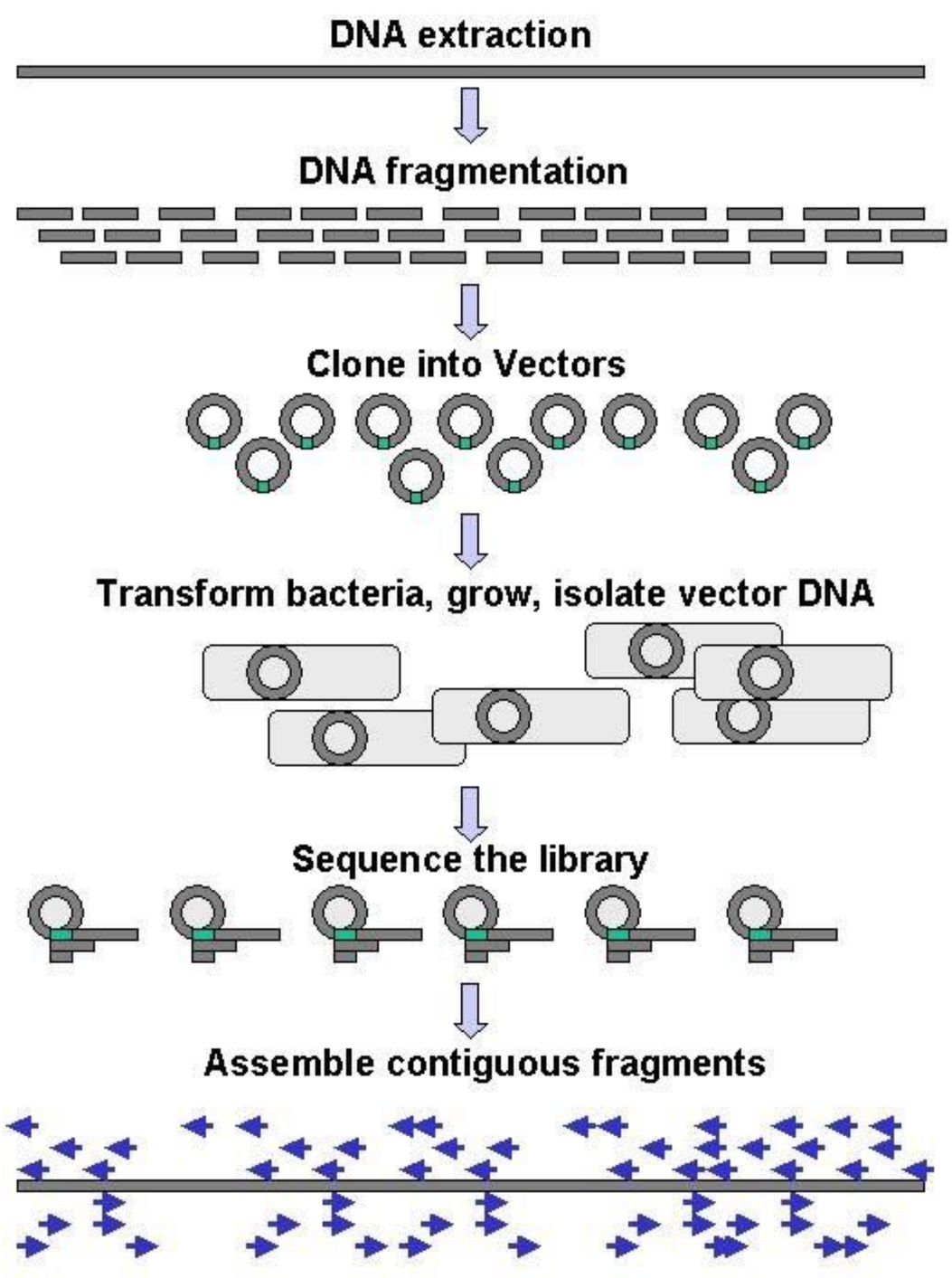
NEWS IN BRIEF · Science & Technology · Science · ISSUE 49·33 · Aug 15, 2013

Share on Facebook 3.2K Share on Twitter 961

105

DNA SEQUENCING

- Large-scale sequencing requires DNA to be broken into fragments
  - Cutting (with enzymes)
  - Shearing (with mechanical forces)
- DNA is duplicated into a vector
- Individually sequenced
- Assembled



DNA SEQUENCING

- Methods:
  - Chain termination or dideoxy method (Sanger)
  - Shotgun sequence method
  - Next generation sequence methods
  - Next-*next* generation

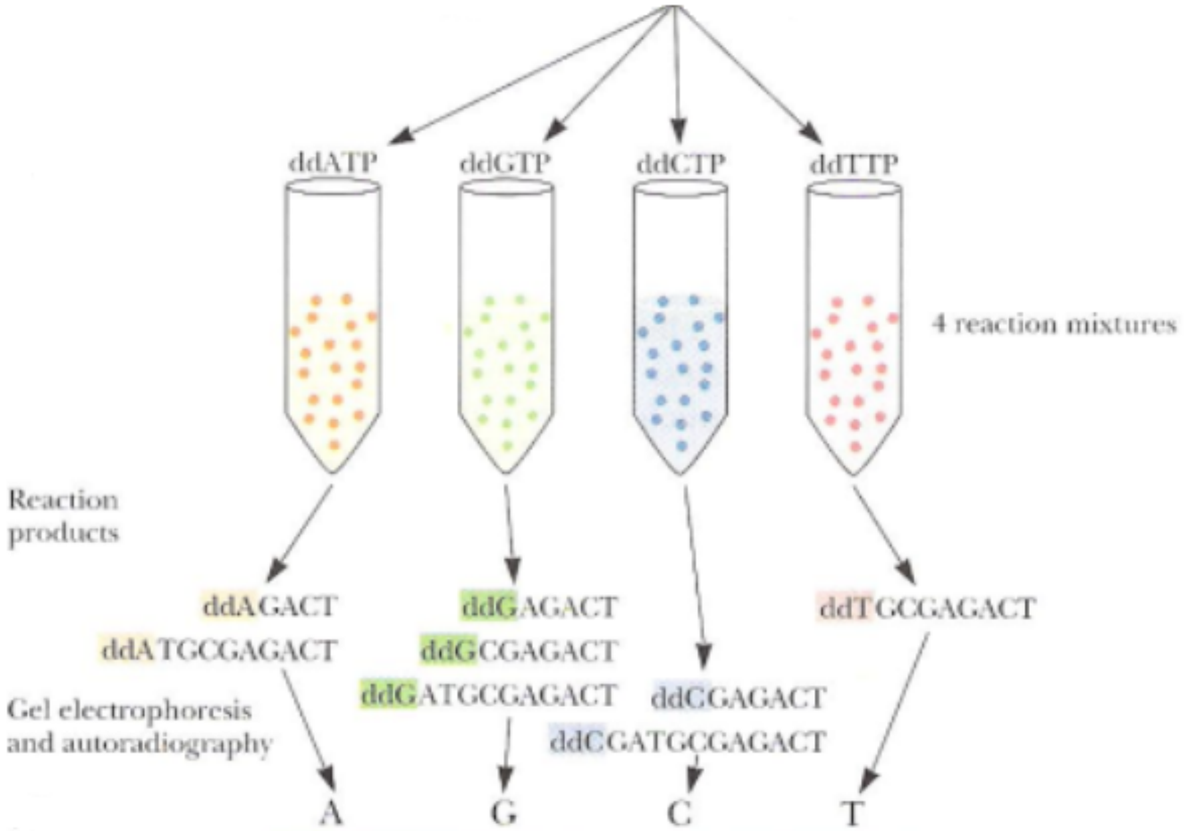
Sanger Sequencing

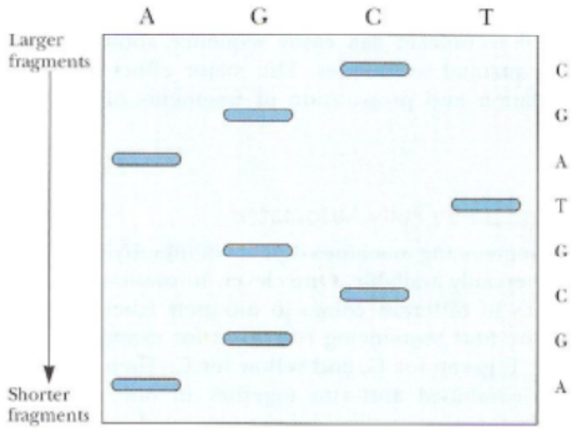
SANGER SEQUENCING

DNA Sequencing

SANGER SEQUENCING

- Dideoxy method
  - Run four separate reactions each with different ddNTPs
  - Run on a gel in four separate lanes
  - Read the gel from the bottom up with UV light
- Example from image
  - CGATGCGA





SANGER SEQUENCING

- The dideoxy method is good only for 500-750bp (human genome is about 3 billion bp)
- Expensive and time consuming

Shotgun Sequencing

SHOTGUN SEQUENCING

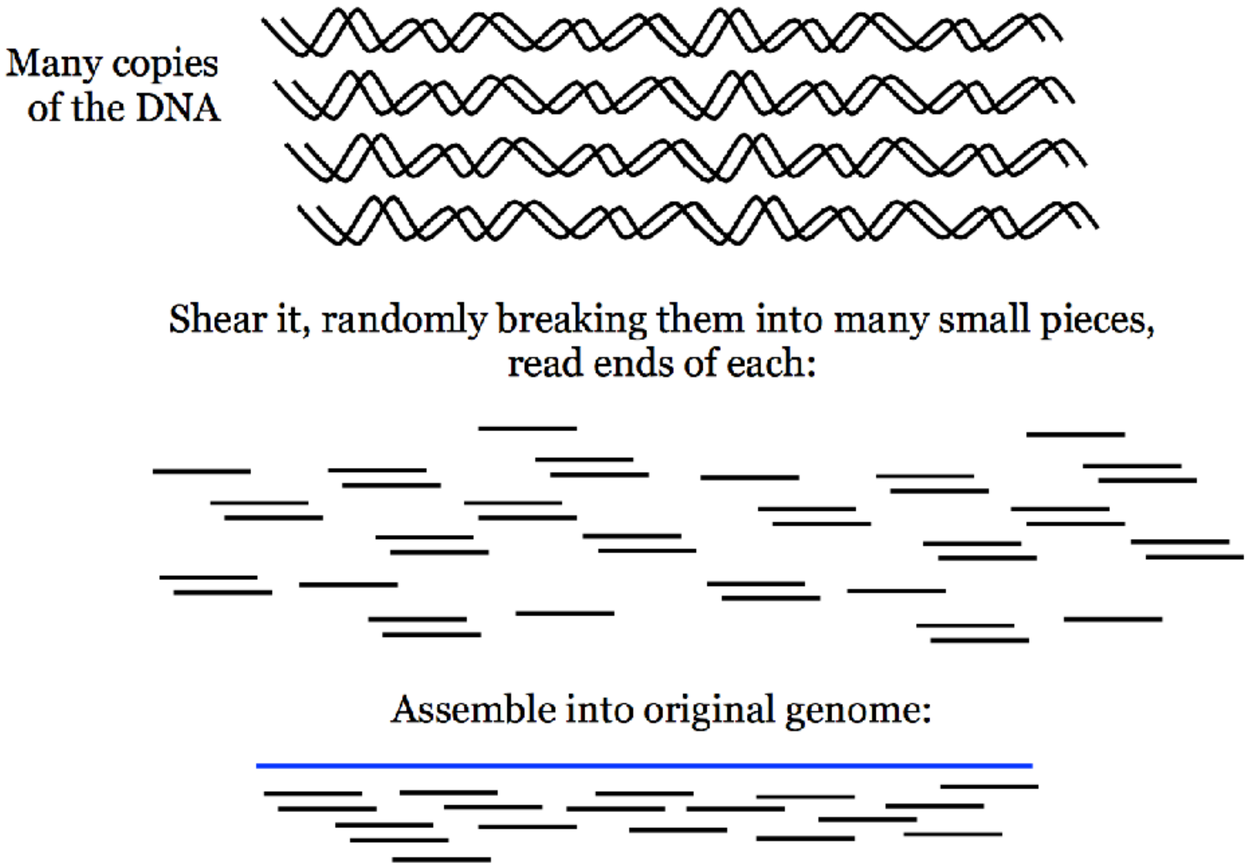
- Human Genome project
  - Public shotgun sequencing approach

SHOTGUN SEQUENCING

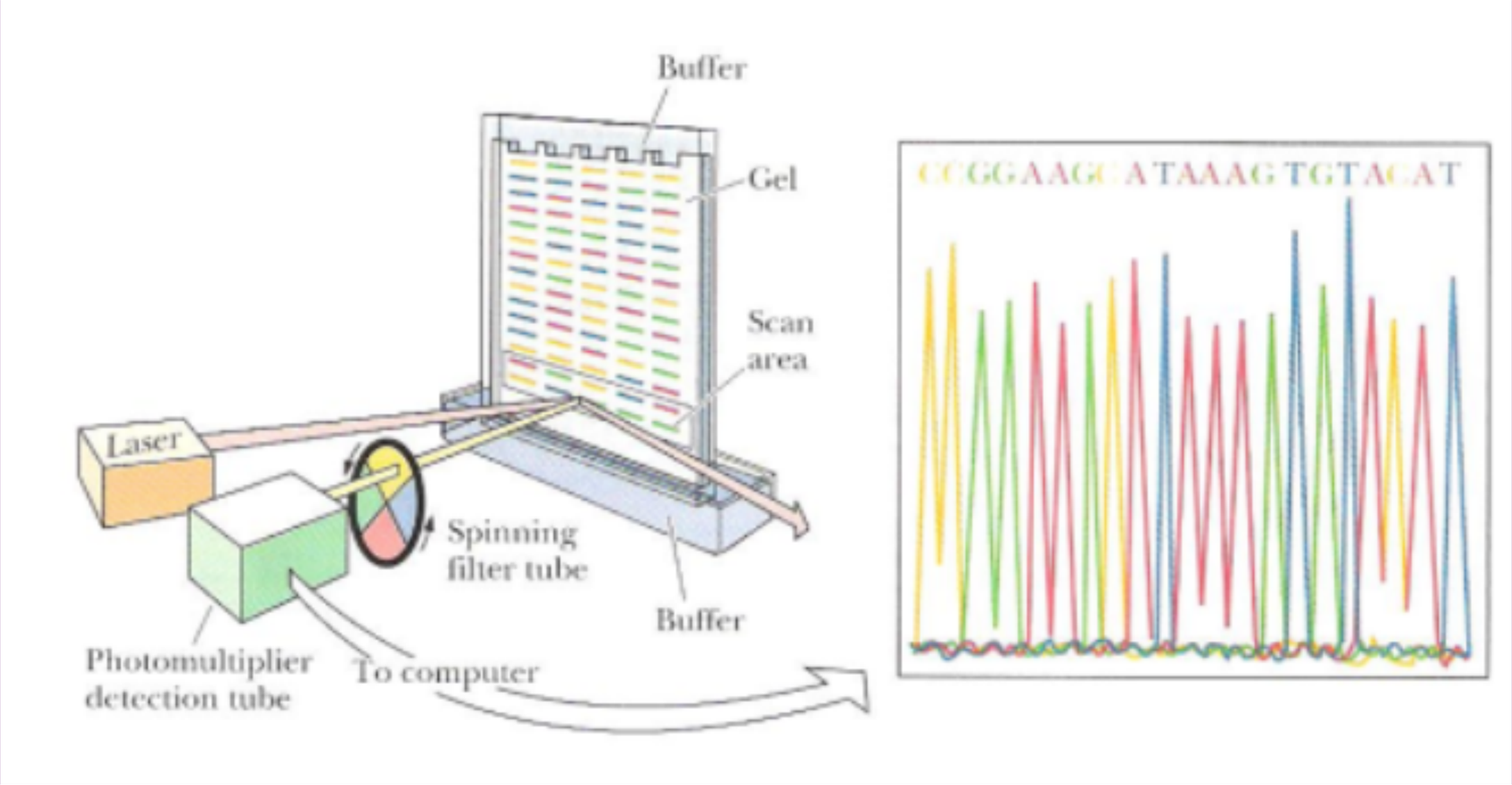
- Human Genome project
  - Celera shotgun sequencing

SHOTGUN SEQUENCING

- Used to sequence whole genomes
- Steps:
  - DNA is broken up randomly into smaller fragments
  - Dideoxy method produces reads
  - Look for overlap of reads
- Technique used to accelerate human genome project



SHOTGUN SEQUENCING



- Automated version of dideoxy method

SHOTGUN SEQUENCING

STRAND	SEQUENCE
FIRST SHOTGUN SEQUENCE	AGCATGCTGCAGTCATGCT------- ----------TAGGCTA
SECOND SHOTGUN SEQUENCE	AGCATG---------- ------CTGCAGTCATGCTTAGGCTA
RECONSTRUCTION	AGCATGCTGCAGTCATGCTTAGGCTA

Next Generation Sequencing

NEXT GENERATION SEQUENCING

- Next-generation sequencing (NGS) is the catch-all term used to describe a number of different modern sequencing technologies
- Cheaper and faster but based on similar mechanism
  - Probes on a chip
  - Tagging bases

- **Illumina GAIIx, HiSeq2000 (Sequencing by Synthesis)**
- **Roche 454 Genome Sequencer FLX, GS Junior (Pyrosequencing)**
- **Life Technologies SOLiD (Sequencing by Ligation)**
- Life Technologies Ion Torrent (PostLight™ semi-conductor sequencing)
- Pacific Biosciences (Real Time Single Molecule Sequencing)
- Helicos (True Single Molecule Sequencing)
- Oxford nanopore (Single Molecule Sequencing)
- IBM nanopore sequencing (Single Molecule Sequencing)
- Nabsys (nanopore + electronic sequencing)
- ZS Genetics
- Visigen
- Halcyon Molecular
- Complete Genomics (only human)
- Mobious (Nexus I)
- Polonator (MPS by ligation)
- Cracker (SMRT on a chip)
- Kavli Institute of Nanoscience; Graphene Nanopores

NEXT GENERATION SEQUENCING

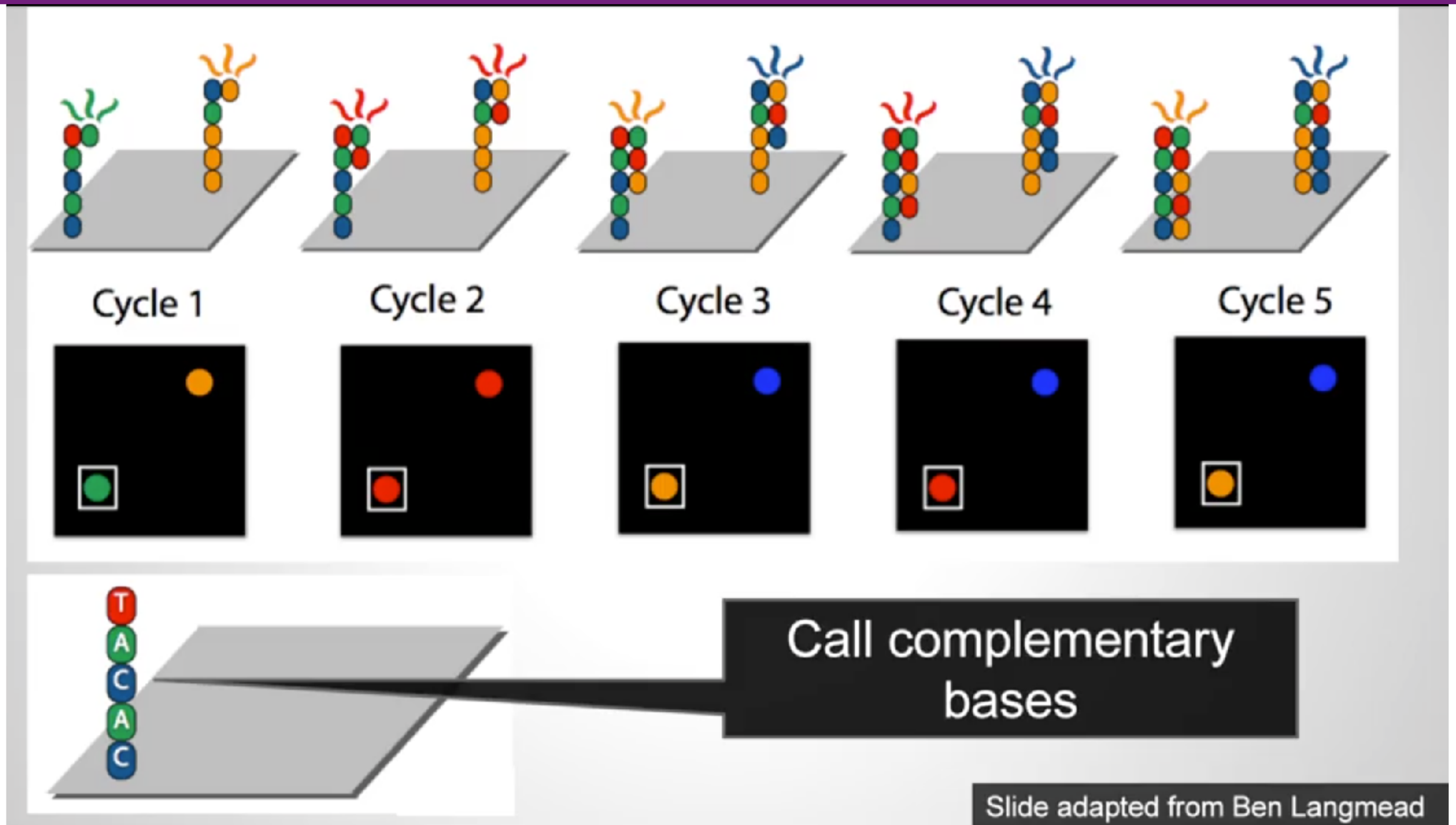
- **Illumina GAIIx, HiSeq2000 (Sequencing by Synthesis)**
- **Roche 454 Genome Sequencer FLX, GS Junior (Pyrosequencing)**
- **Life Technologies SOLiD (Sequencing by Ligation)**
- Life Technologies Ion Torrent (PostLight™ semi-conductor sequencing)
- Pacific Biosciences (Real Time Single Molecule Sequencing)
- Helicos (True Single Molecule Sequencing)
- Oxford nanopore (Single Molecule Sequencing)
- IBM nanopore sequencing (Single Molecule Sequencing)
- Nabsys (nanopore + electronic sequencing)
- ZS Genetics
- Visigen
- Halcyon Molecular
- Complete Genomics (only human)
- Mobious (Nexus I)
- Polonator (MPS by ligation)
- Cracker (SMRT on a chip)
- Kavli Institute of Nanoscience; Graphene Nanopores

NEXT GENERATION SEQUENCING

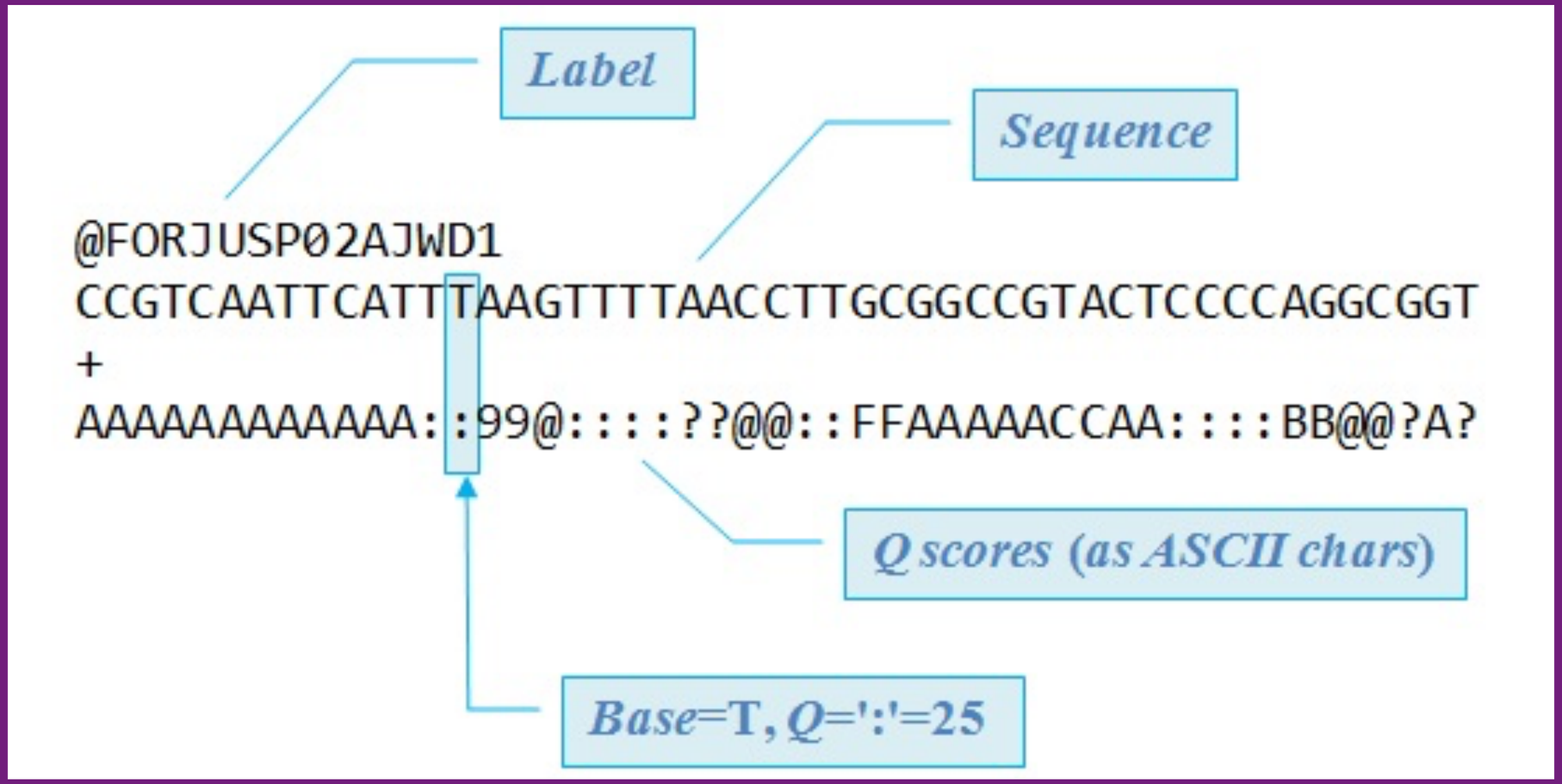
ION TORRENT

NEXT GENERATION SEQUENCING

NEXT GENERATION SEQUENCING



NEXT GENERATION SEQUENCING

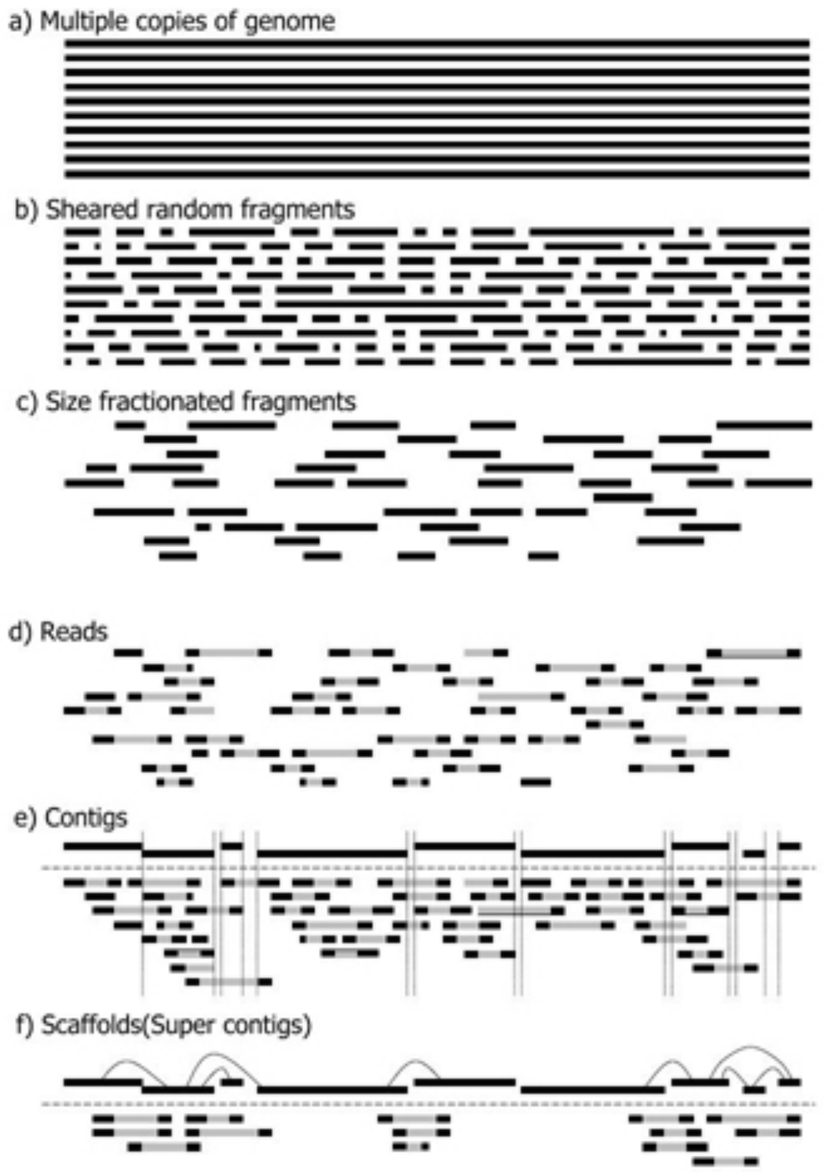


- Scale of experiments introduces errors; need to evaluate the confidence of every base

NEXT GENERATION SEQUENCING

- Genome Assembly Problem
  - How to piece together small fragments
  - Repeats
  - Reference/de novo assembly

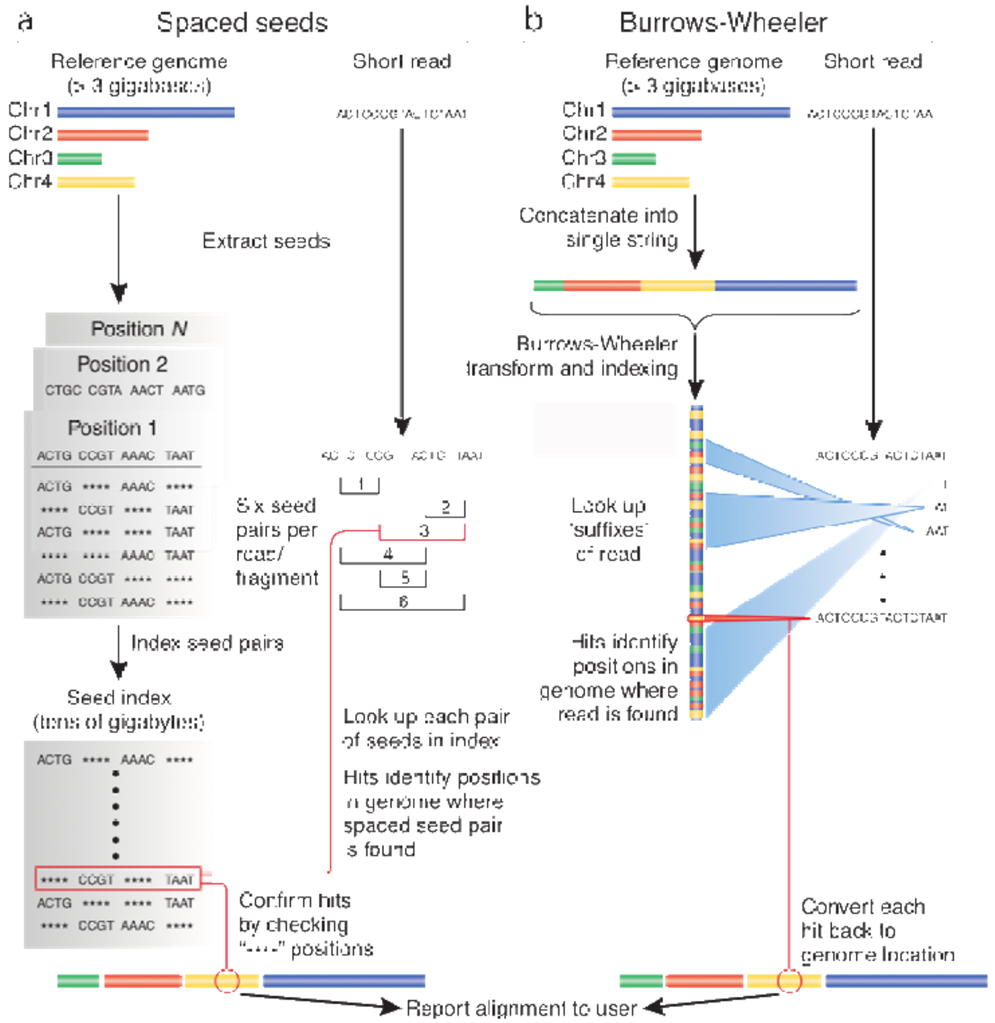
- We'll get back to this…



NEXT GENERATION SEQUENCING

- Human microbiome
  - 50-1000 genomes
  - Each 500,000-10,000,000 characters long
  - Alphabet of only 4 letters
  - Abundance varies
    - Billions copies
    - Tens of copies
  - Each genome is randomly cut up into “reads” ranging in length from 75-500 characters long

NEXT GENERATION SEQUENCING

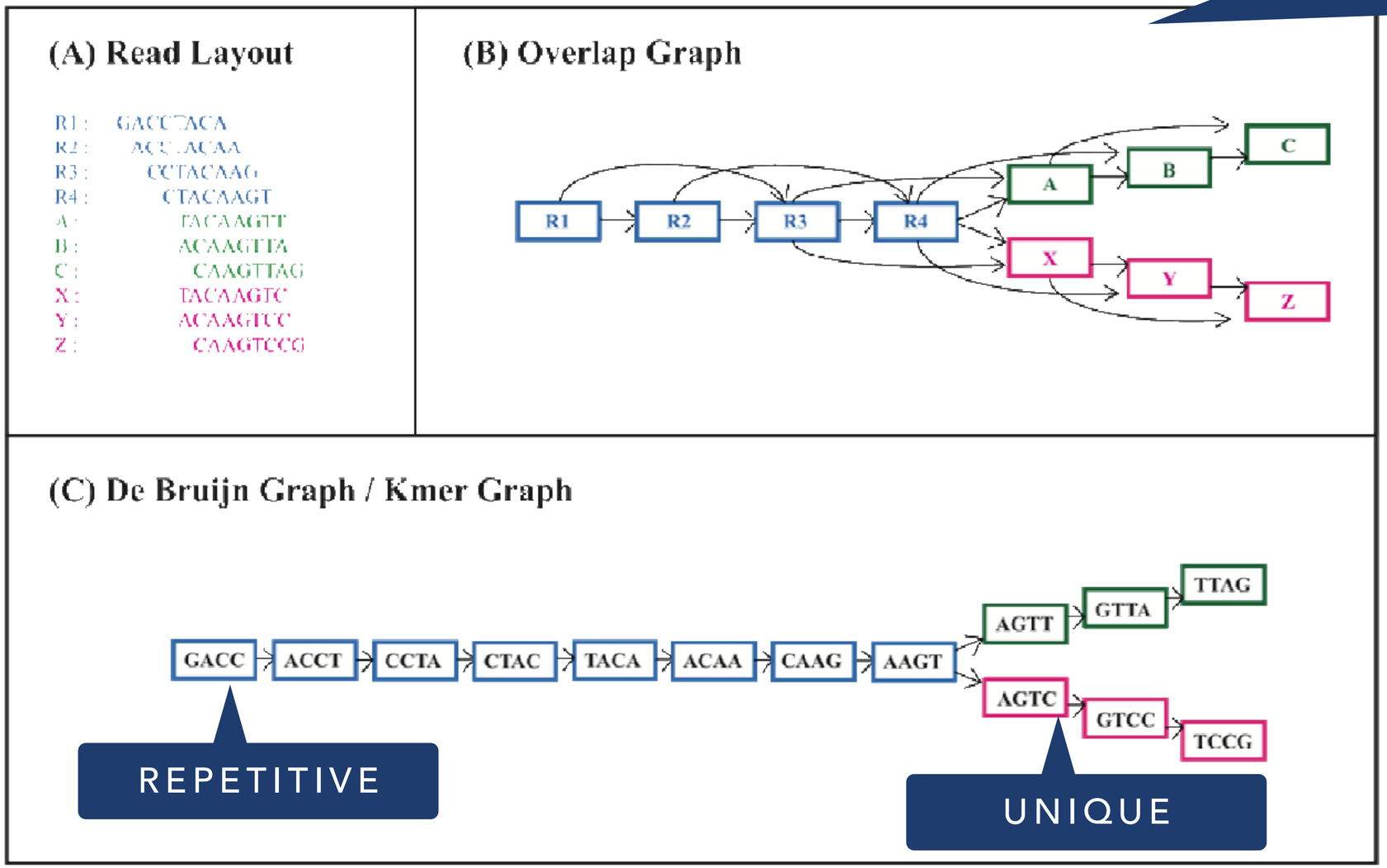


MAP TO REFERENCE GENOMES

Image from Trapnell & Salzberg (2009) Nature Biotech. 27:455-457.

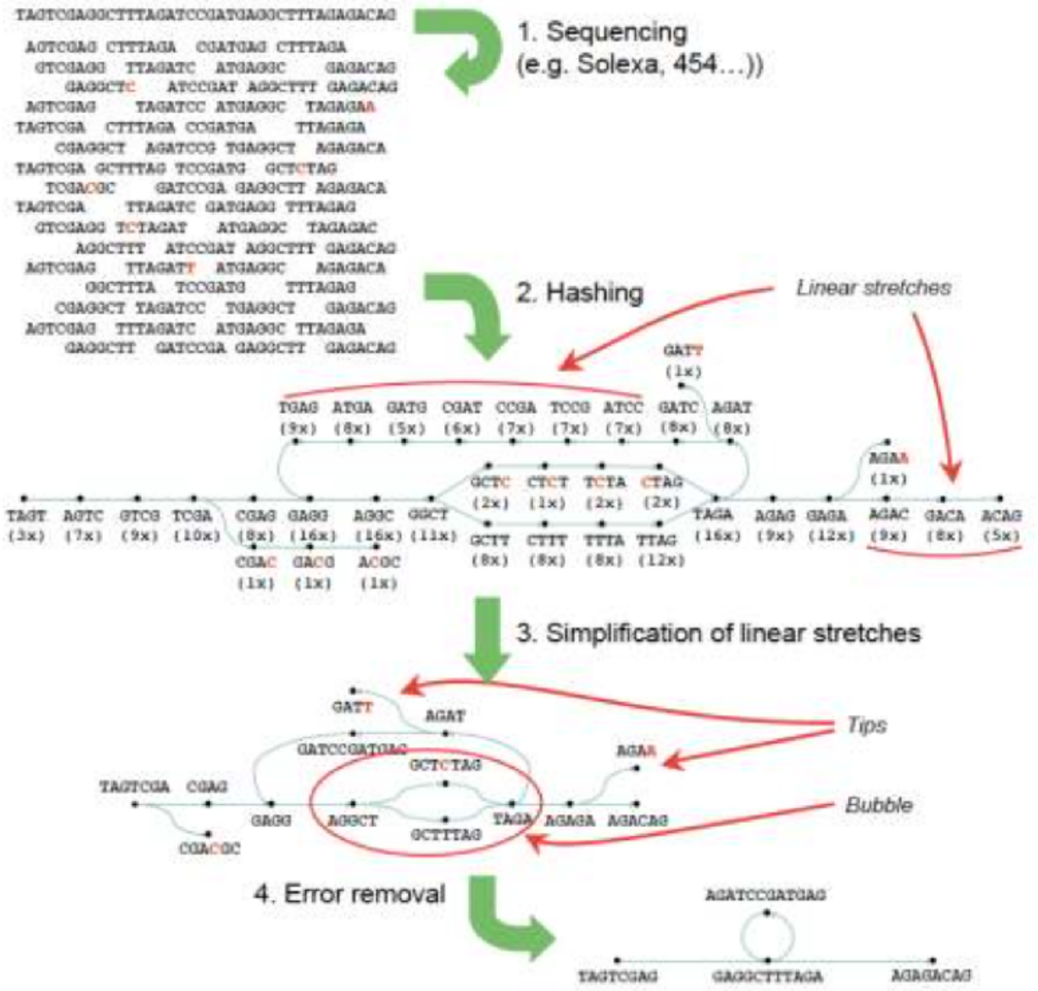
NEXT GENERATION SEQUENCING

DE NOVO ASSEMBLY

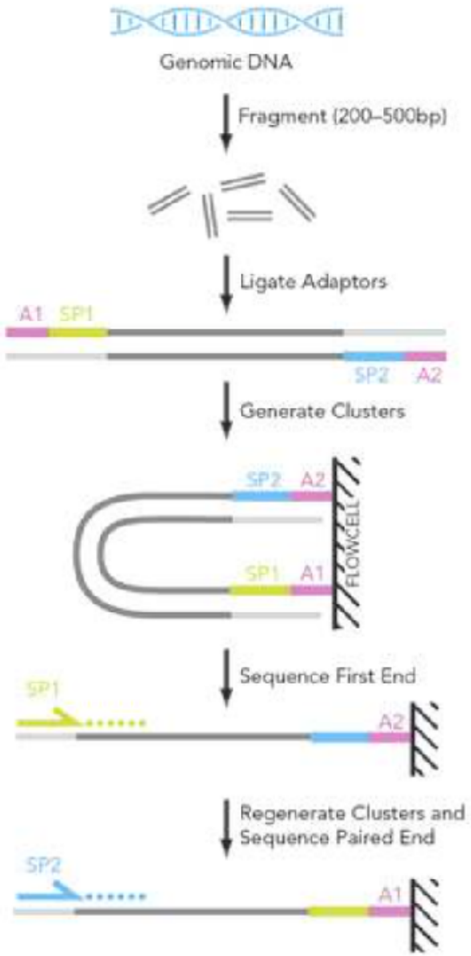


NEXT GENERATION SEQUENCING

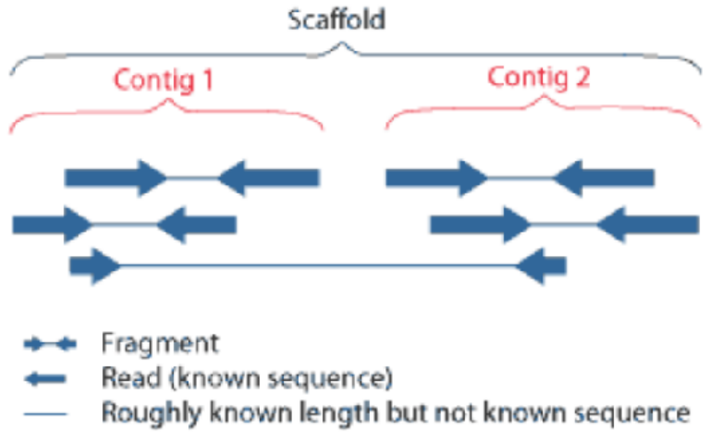
DE NOVO ASSEMBLY



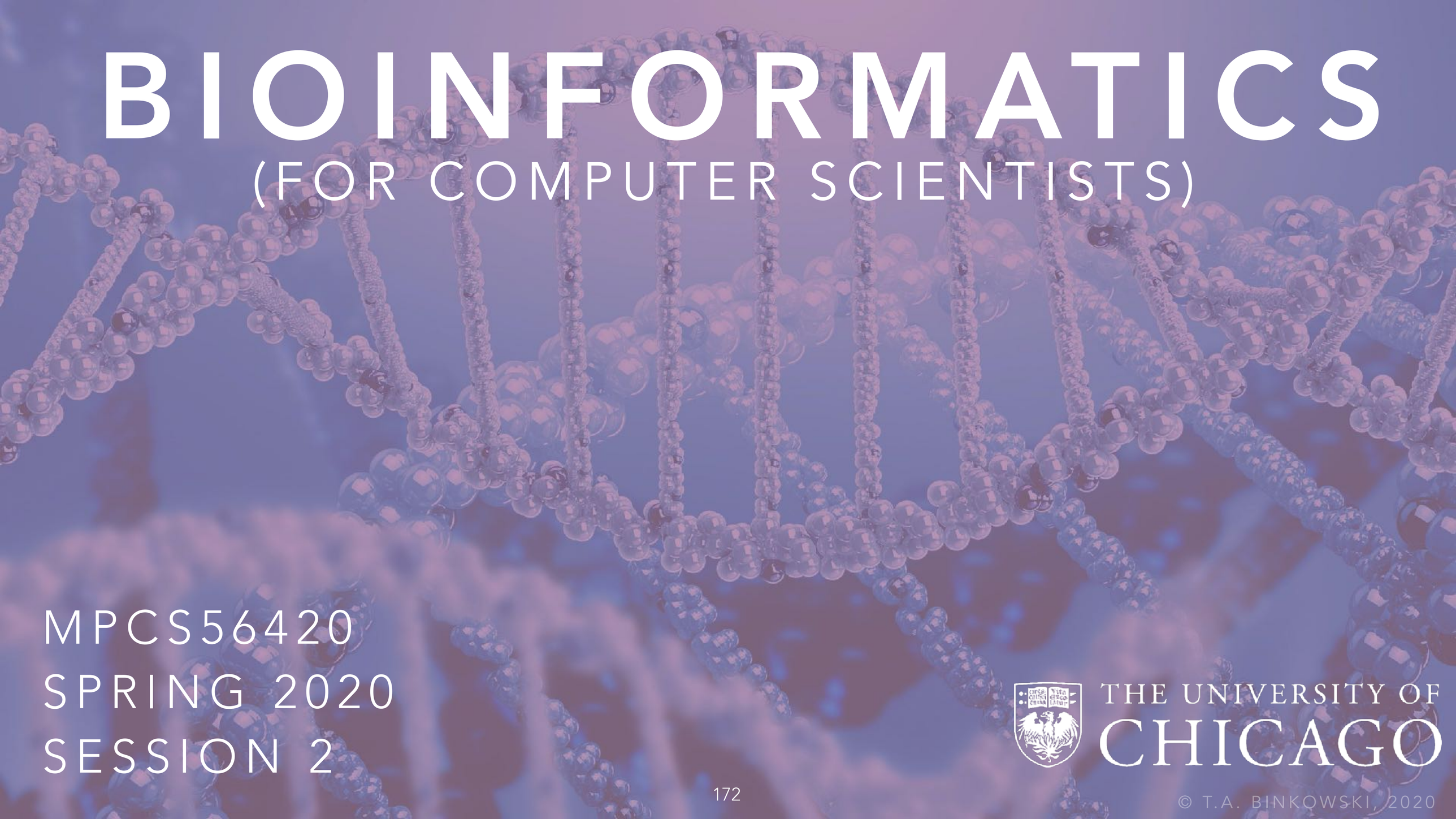
NEXT GENERATION SEQUENCING



PAIRED-END READS (ILLUMINA) CAN SIMPLIFY ASSEMBLY



From Illumina.com website and http://genome.jgi.doe.gov/help/scaffolds.html



BIOINFORMATICS

(FOR COMPUTER SCIENTISTS)

MPCS56420
SPRING 2020
SESSION 2

THE UNIVERSITY OF CHICAGO