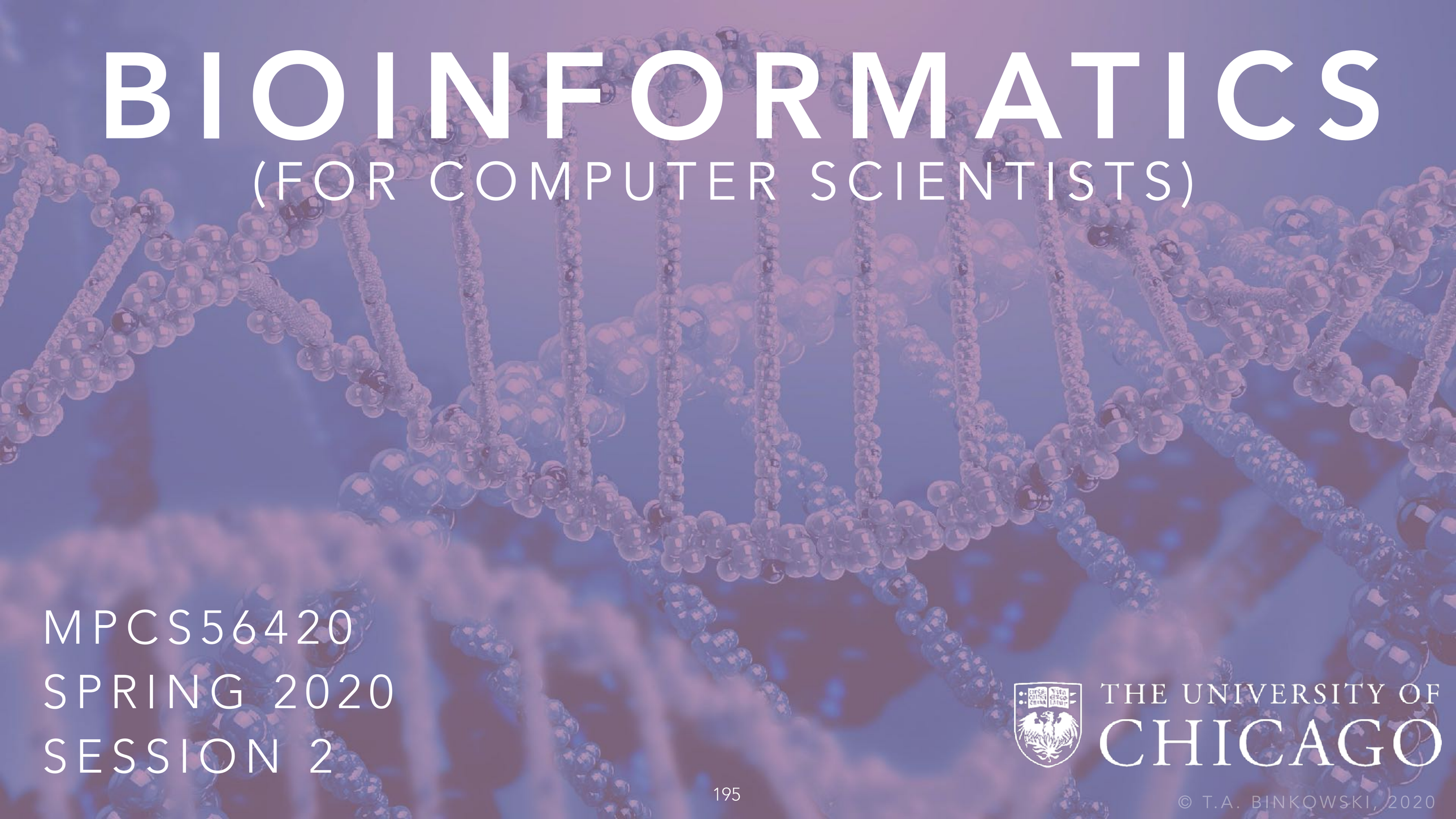


BIOINFORMATICS

(FOR COMPUTER SCIENTISTS)

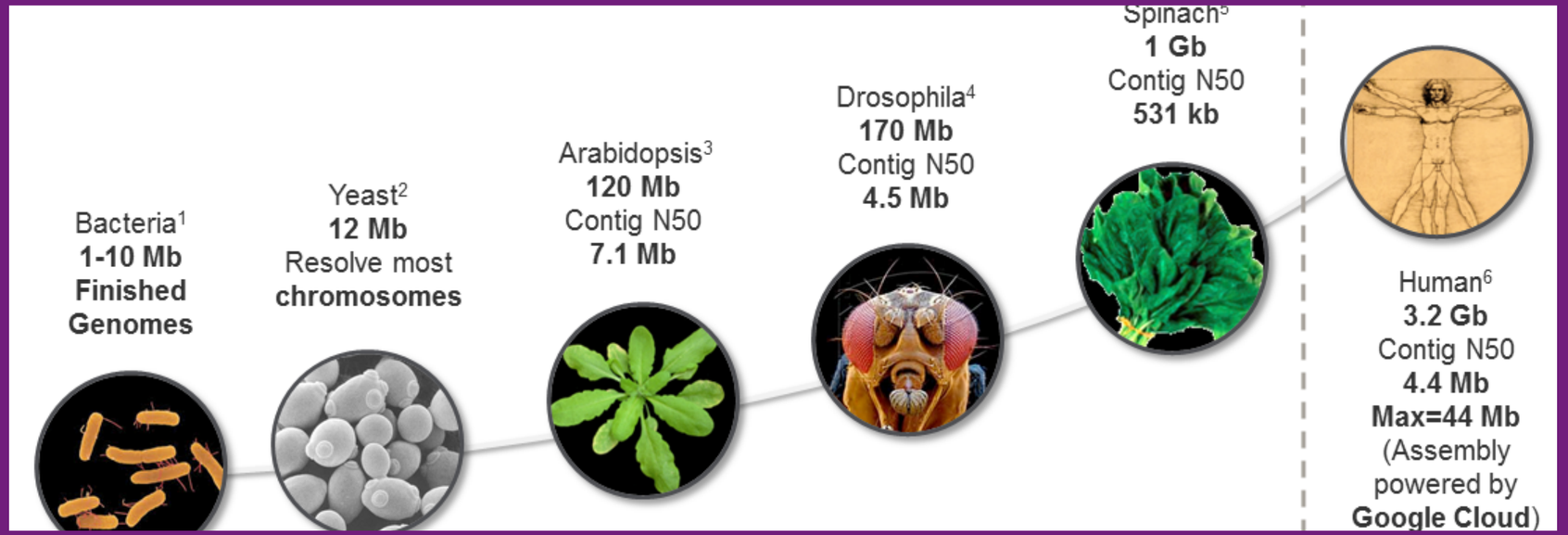
MPCS56420
SPRING 2020
SESSION 2

THE UNIVERSITY OF CHICAGO

GENES AND GENOMES

GENOMES

GENES AND GENOMES



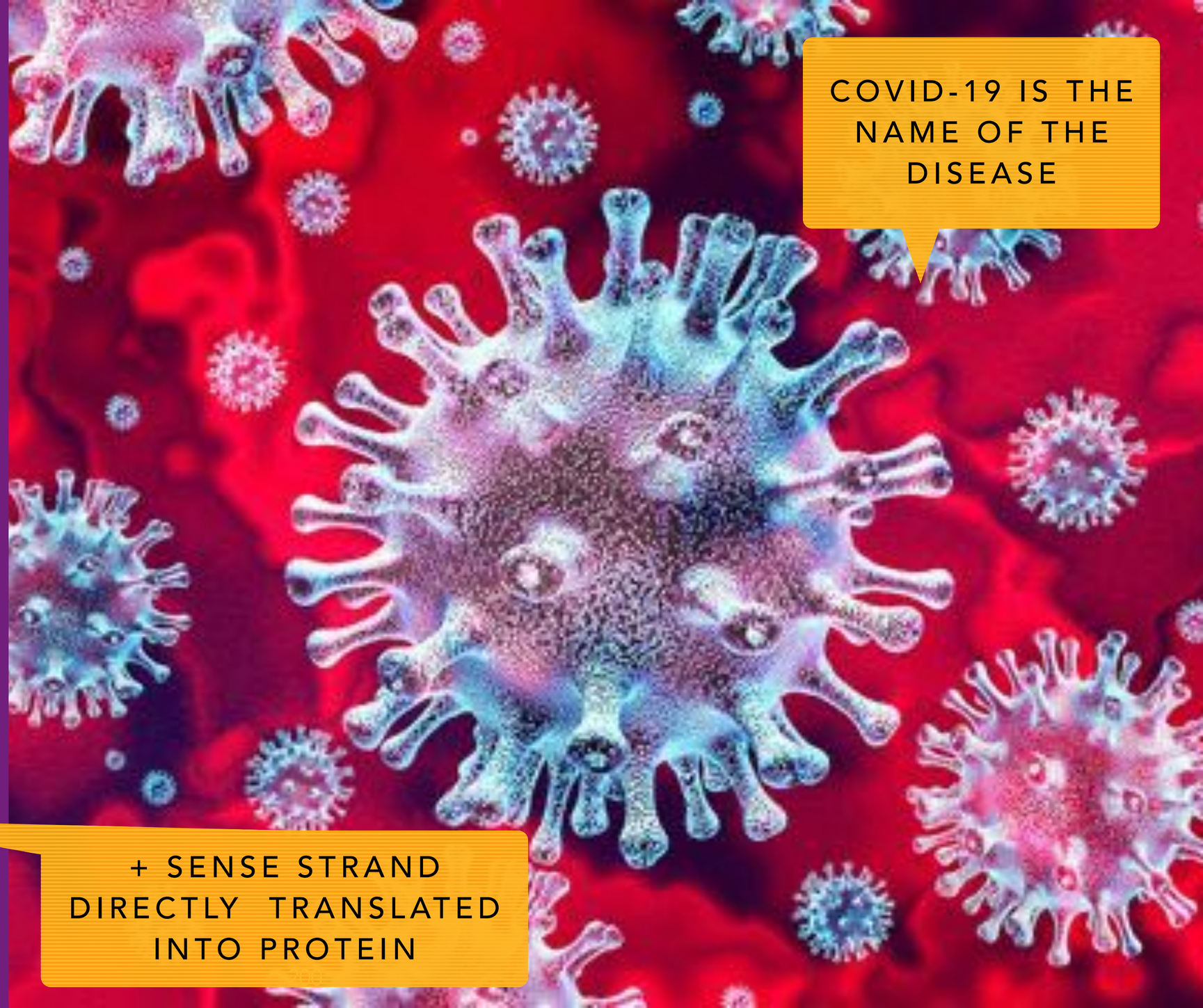
- The genome is an organism's complete set of DNA
  - Bacteria contains about 600,000 DNA base pairs
  - Human and mouse genomes have some 3 billion

GENES AND GENOMES

- Ebola
  - Genome is only 19 kb

GENES AND GENOMES

- Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 or coronavirus)
  - Genome is only 26-32 kb
  - One of the largest positive RNA viruses



COVID-19 IS THE NAME OF THE DISEASE

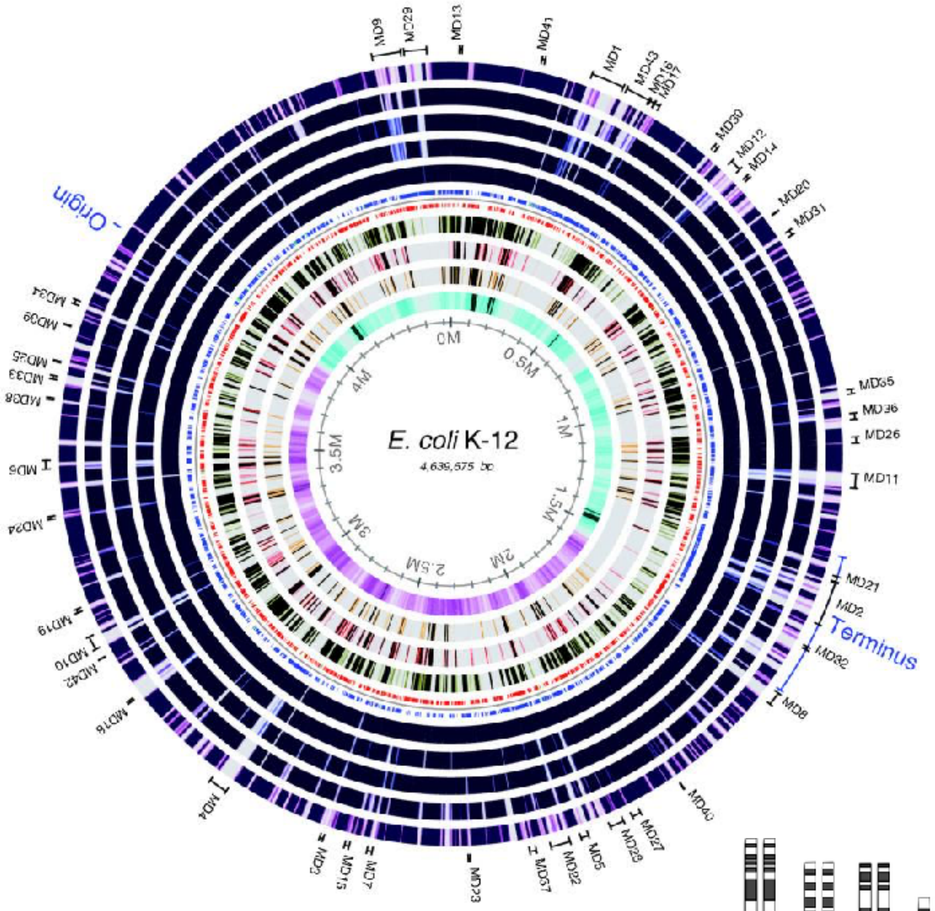
+ SENSE STRAND DIRECTLY TRANSLATED INTO PROTEIN

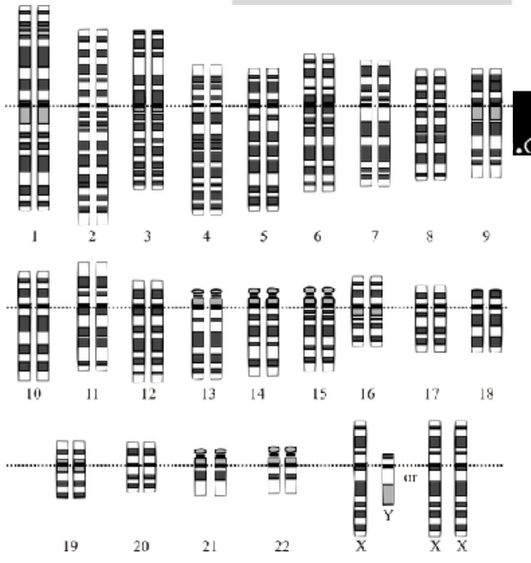
GENES AND GENOMES

- Japanese flower named Paris japonica 149 billion base pairs
  - 50 times the size of human genome

GENES AND GENOMES

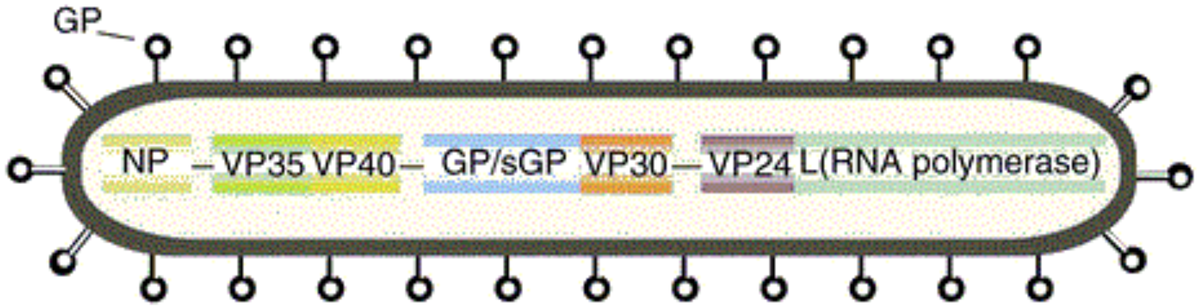
- Eukaryote genome
  - Organized on chromosomes Polycistronic
    - More than one coding region on an mRNA transcript
- Bacterial genome
  - Typically circular genome of double stranded DNA



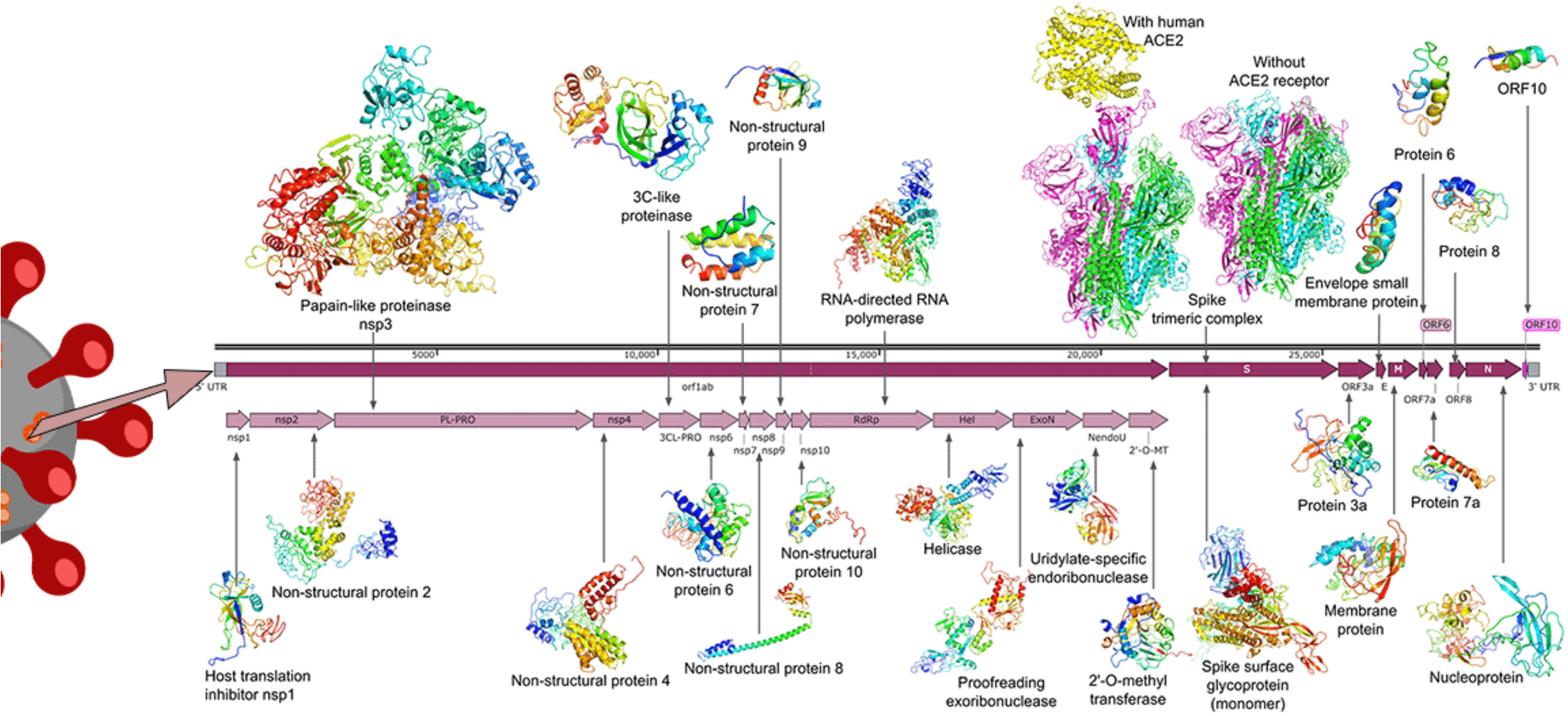


VIRAL GENOMES

- Can be circular or linear
- Three viral genome types
  - RNA viruses
    - Proteins translated from RNA
    - SARS, polio, West Nile, Coronavirus
  - DNA viruses
    - Transcription to mRNA -> protein
  - Retroviruses (negative RNA)
    - Reverse transcriptase convers RNA to DNA
    - Ebola, measles, influenza



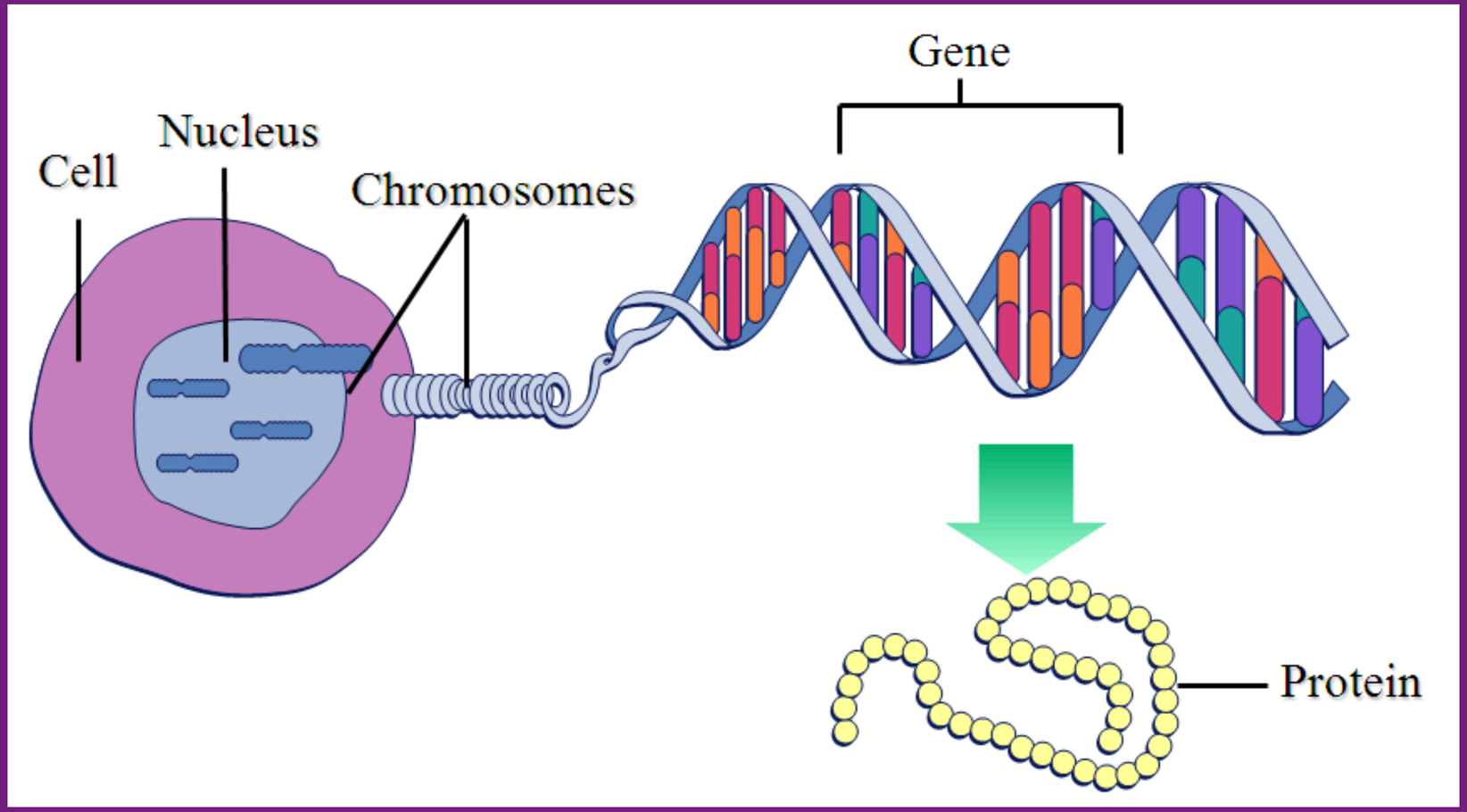
VIRAL GENOMES



Genome-wide structure and function modeling of SARS-COV-2

GENES

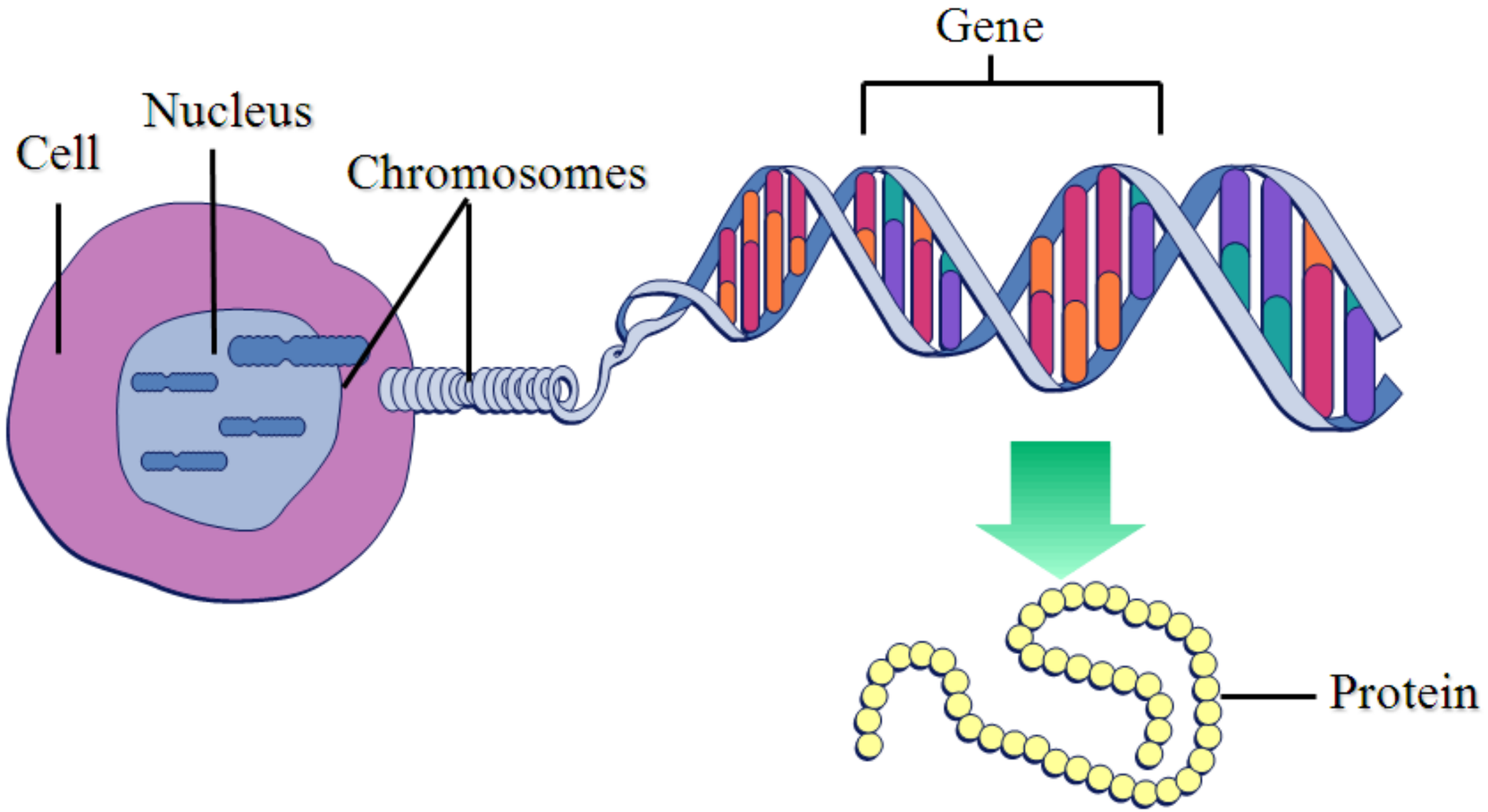
GENES AND GENOMES



- Gene: Specific sequences of DNA bases that encode instructions on how to make proteins

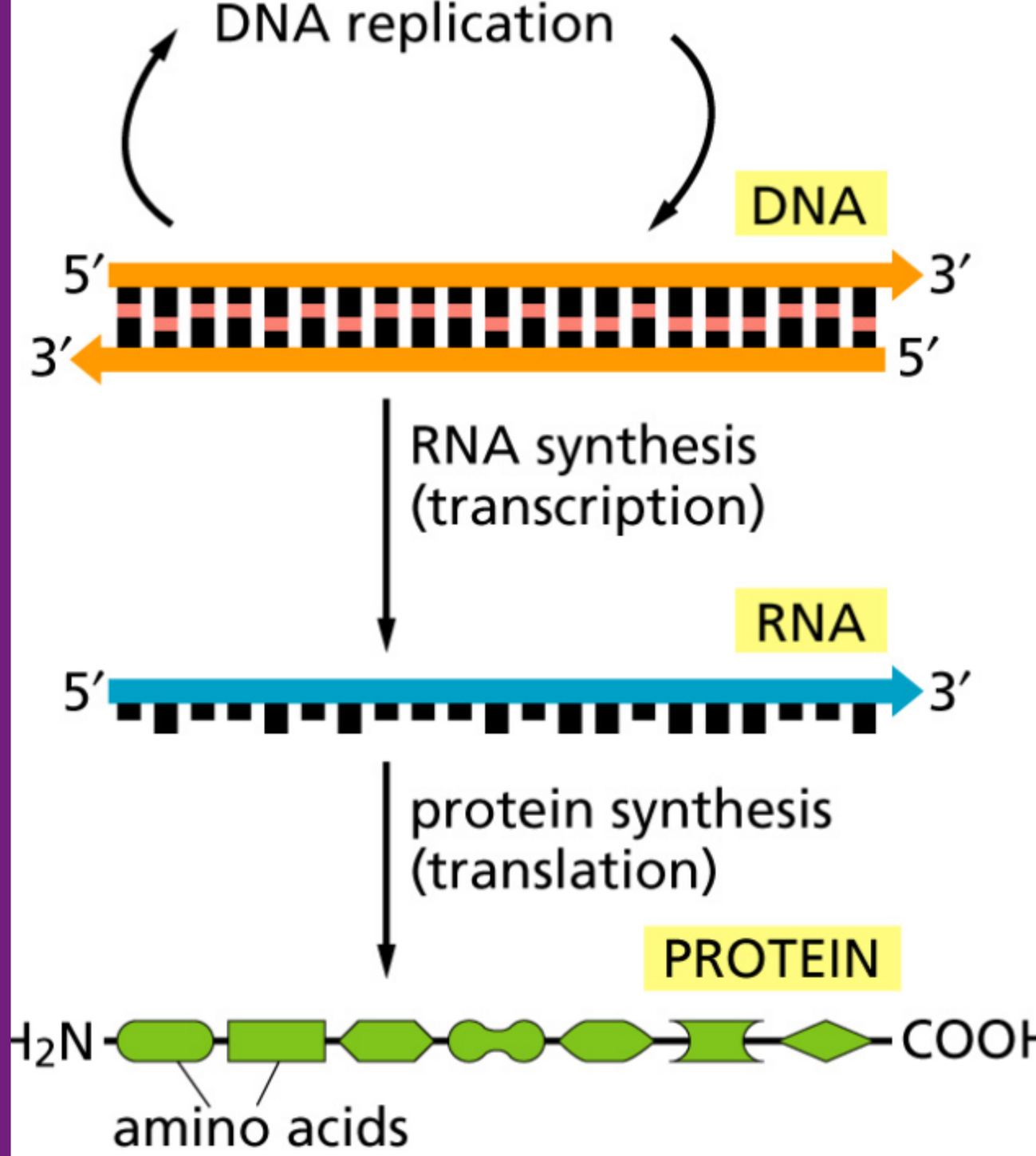
GENES AND GENOMES

- Proteins
  - Make up the cellular structure
  - Large, complex molecules made up of smaller subunits called amino acids



GENES AND GENOMES

- Not all genetic information in DNA encodes proteins
- Non-coding “junk” DNA
  - 98% is non-coding in human genome
  - Play other roles



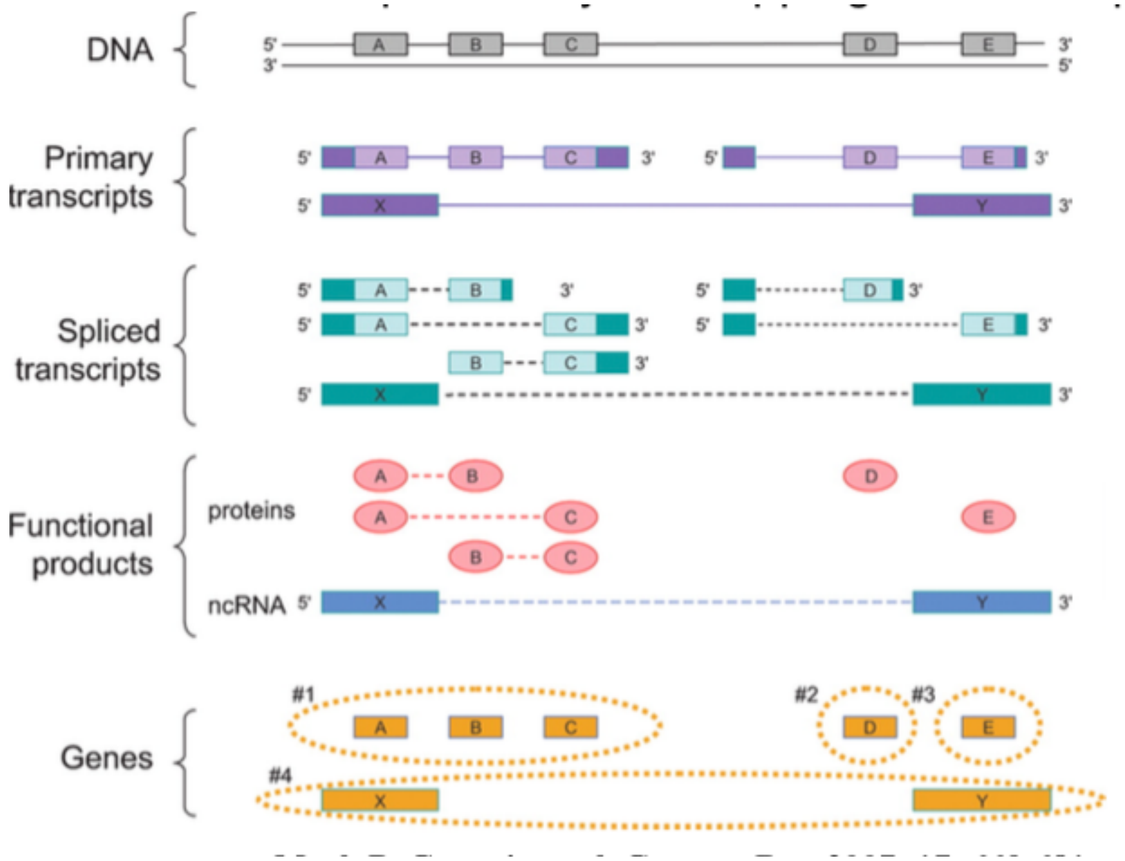
GENE FINDING

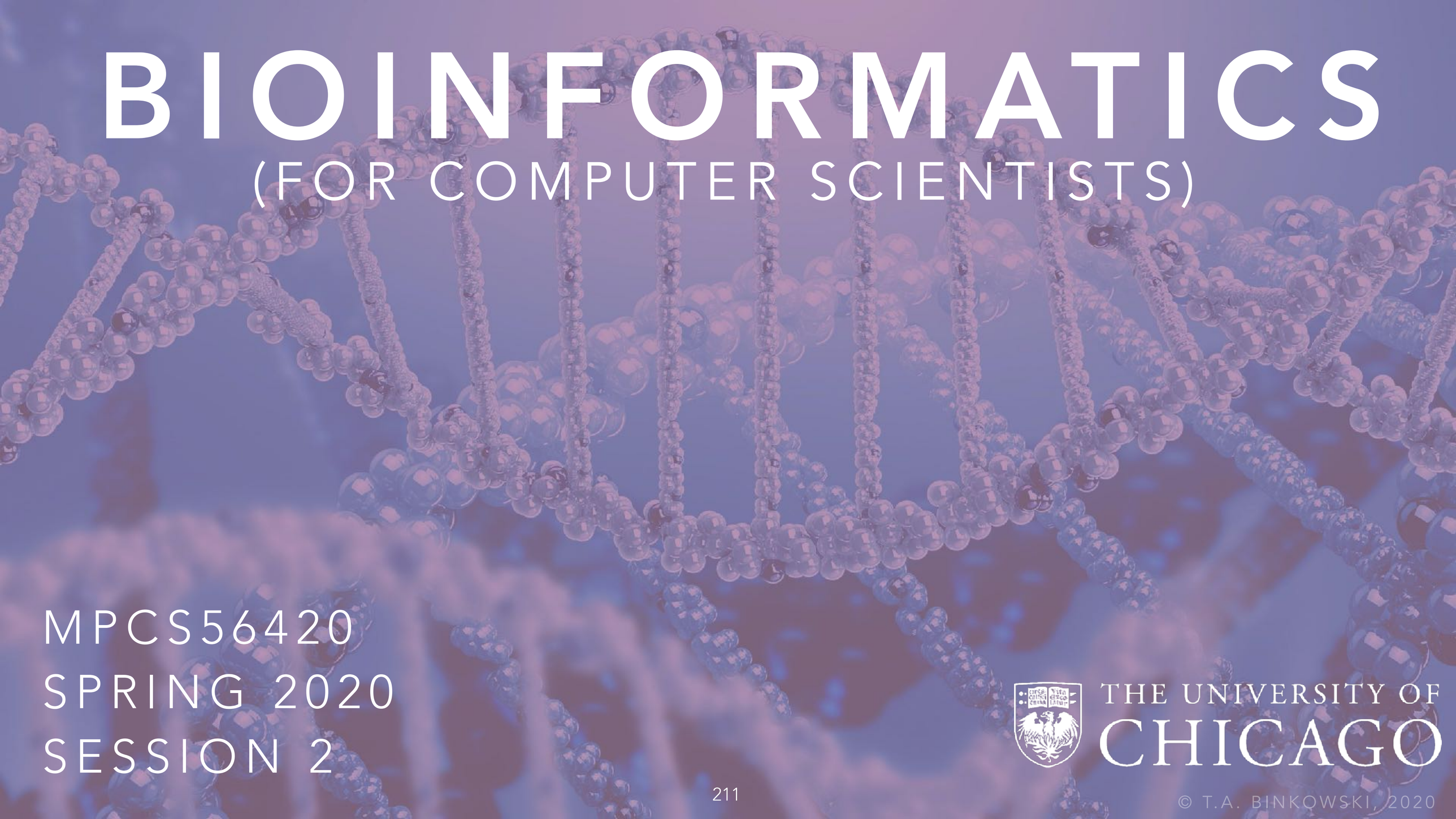
- What is a gene?
  - 1970s–1980s: Gene as open reading frame (ORF) sequence pattern
  - 1990s–2000s: Annotated genomic entity, enumerated in the databanks
  - ENCODE: The gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products

- ENCODE, the Encyclopedia Of DNA Elements
  - NHGRI effort in September 2003, to carry out a project to identify all functional elements in the human genome sequence.

GENE FINDING

- Post ENCODE definition





BIOINFORMATICS

(FOR COMPUTER SCIENTISTS)

MPCS56420
SPRING 2020
SESSION 2

THE UNIVERSITY OF CHICAGO