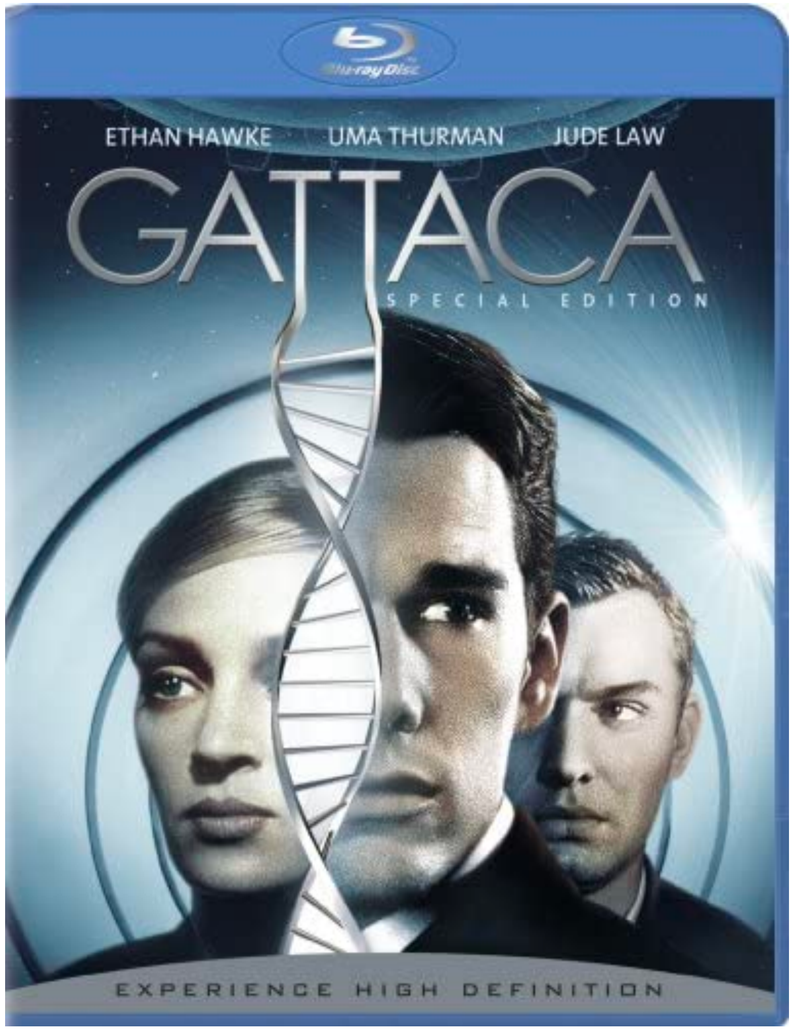
Bioinformatics analysis for next generation sequencing

Hirak Sarkar
Princeton University
www.hiraksarkar.com

Tools for measuring the “sight” and “sound” of biology (Genome and Genomics)

1993

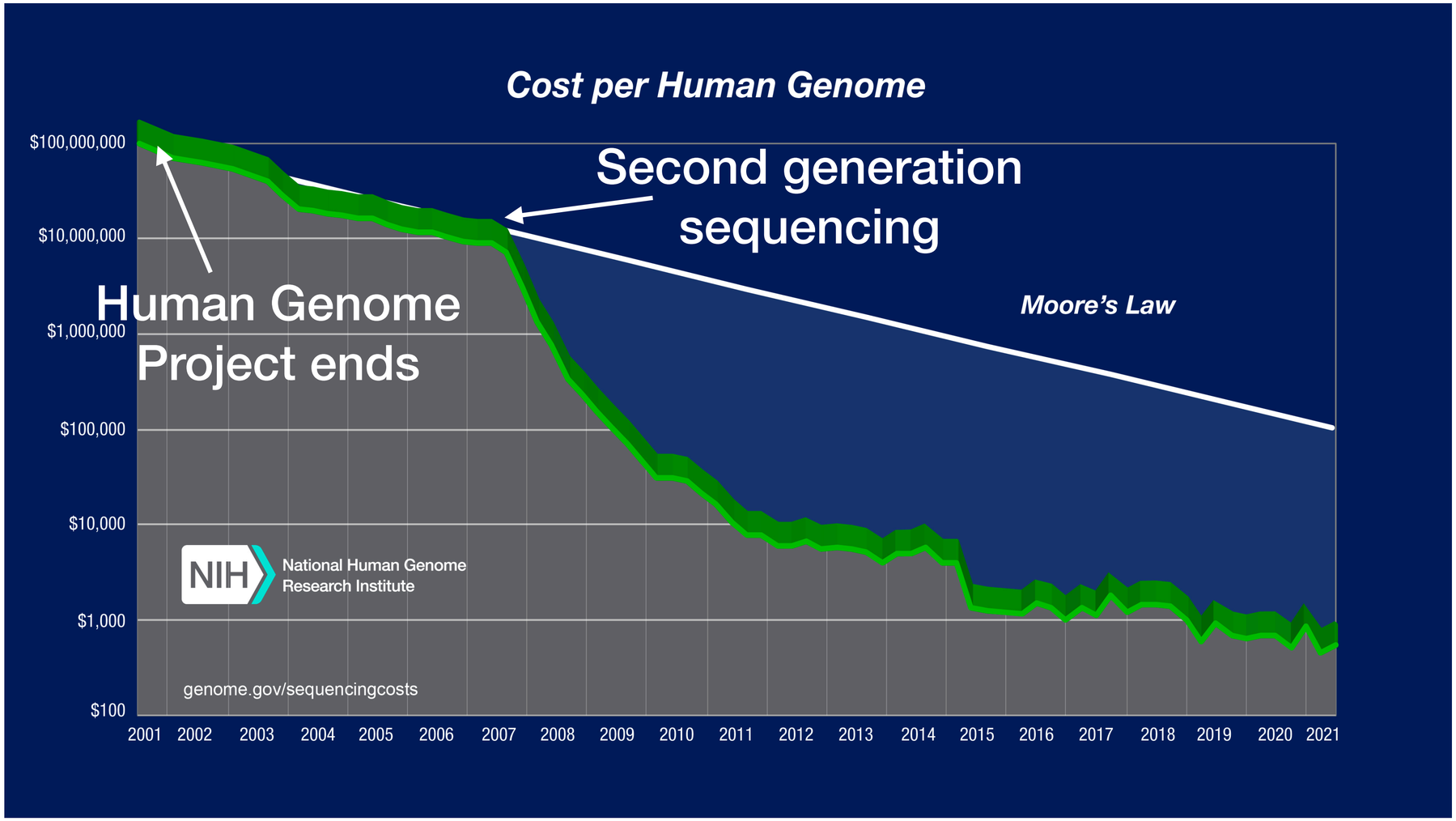
1997



Inspiration - Ben Langmead’s lecture notes

Tools for measuring the “sight” and “sound” of biology (Genome and Genomics)

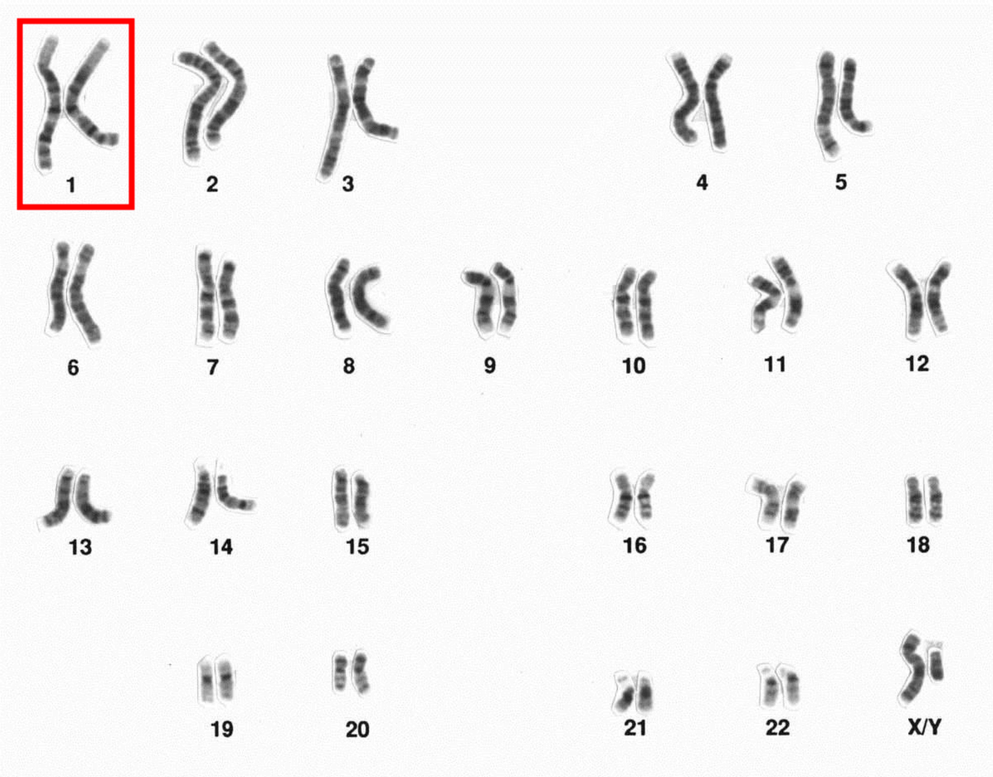
2001

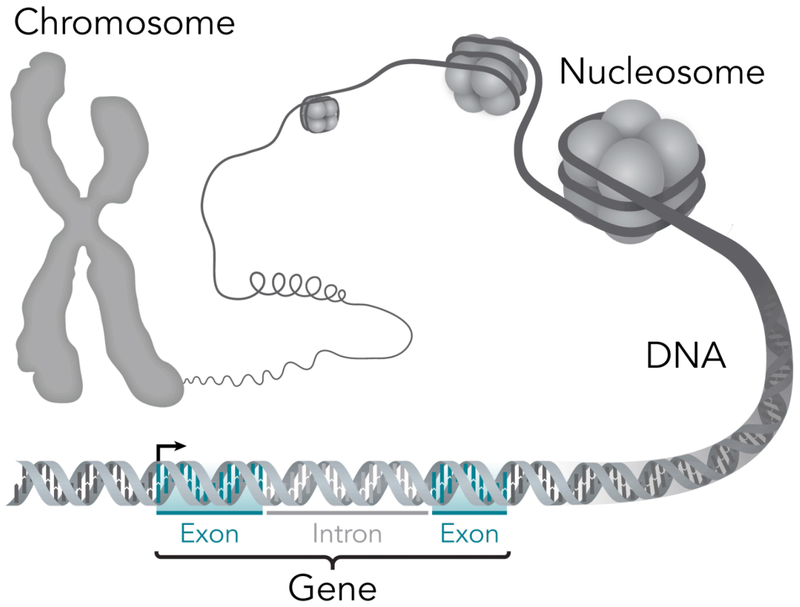


NIH - Genome Project

Inspiration - Ben Langmead’s lecture notes

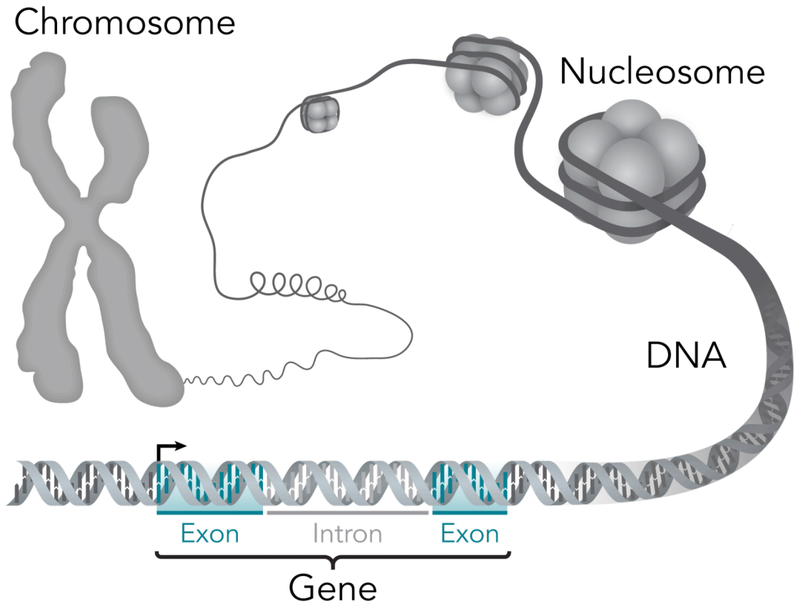
Tools for measuring the “sight” and “sound” of biology (Genome and Genomics)





https://medicalxpress.com/news/2017-02-effects-genes-network.html

Tools for measuring the “sight” and “sound” of biology (Genome and Genomics)



NovaSeq X series specifications

Output Range	~165 Gb - 16 Tb
Single reads per run	1.6 billion - 52 billion
Read length	2 × 150 bp
Run time	~13 hr - 48 hr

View All NovaSeq X Specs View AR

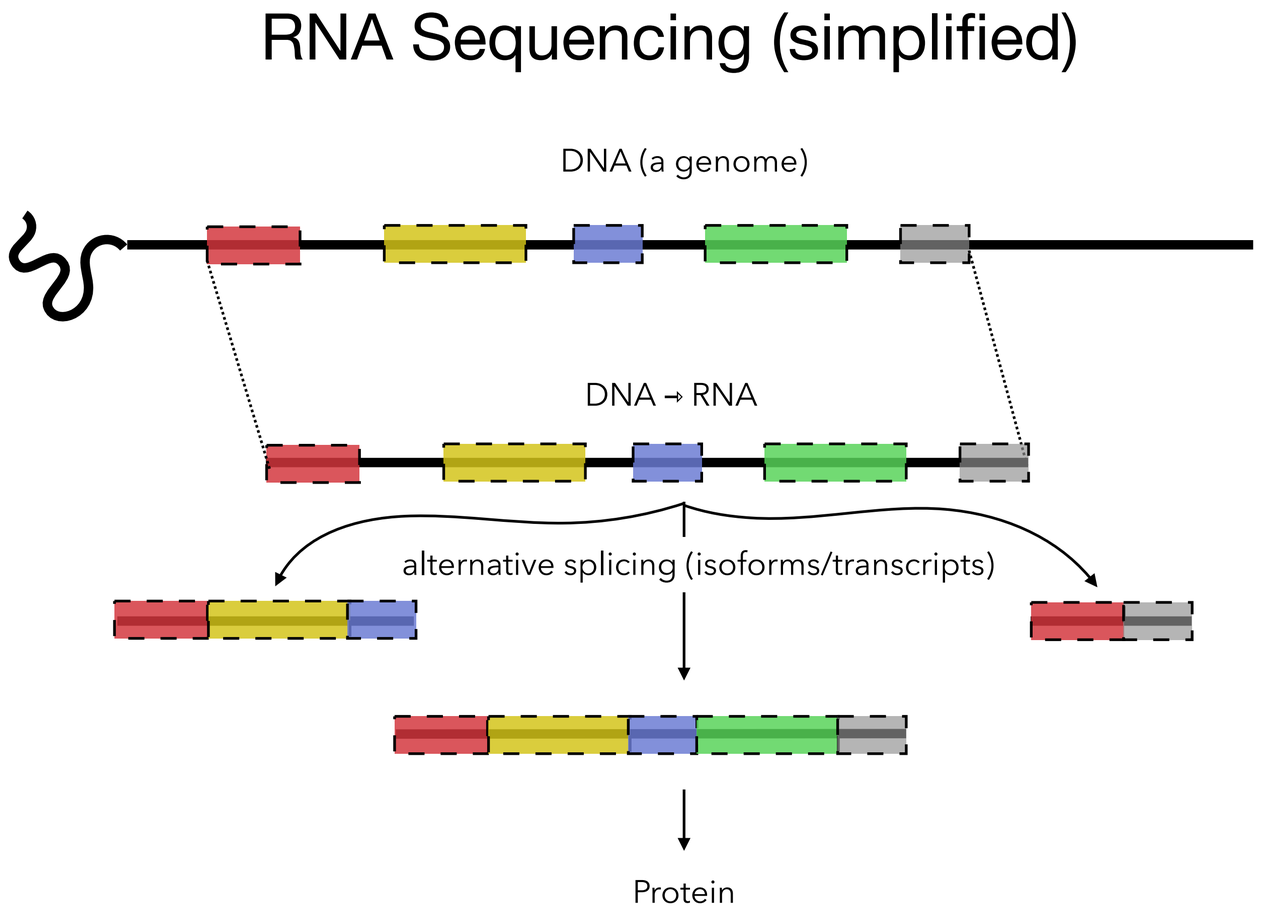
ATAGCAAGCTTTT

https://medicalxpress.com/news/2017-02-effects-genes-network.html

https://www.illumina.com/systems/sequencing-platforms/novaseq-x-plus.html

How to measure gene expression

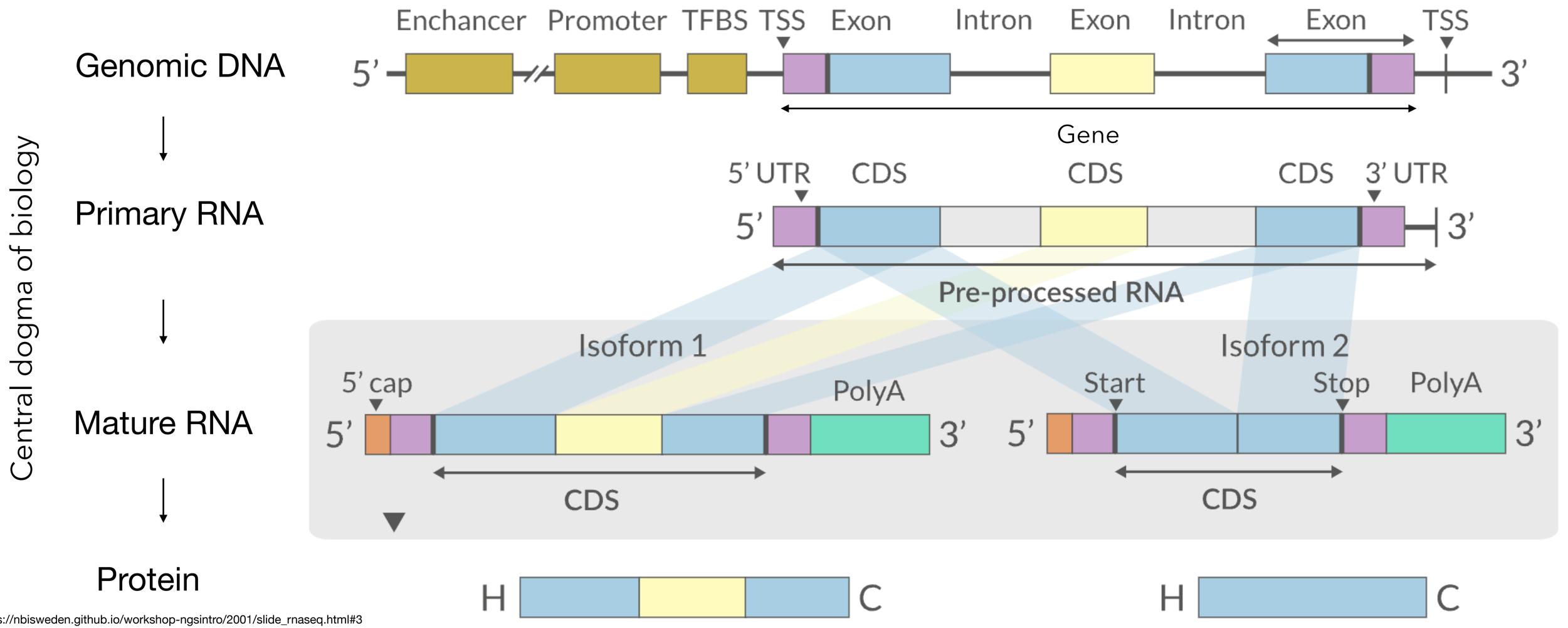
Central dogma of biology



https://medicalxpress.com/news/2017-02-effects-genes-network.html

How to measure gene expression

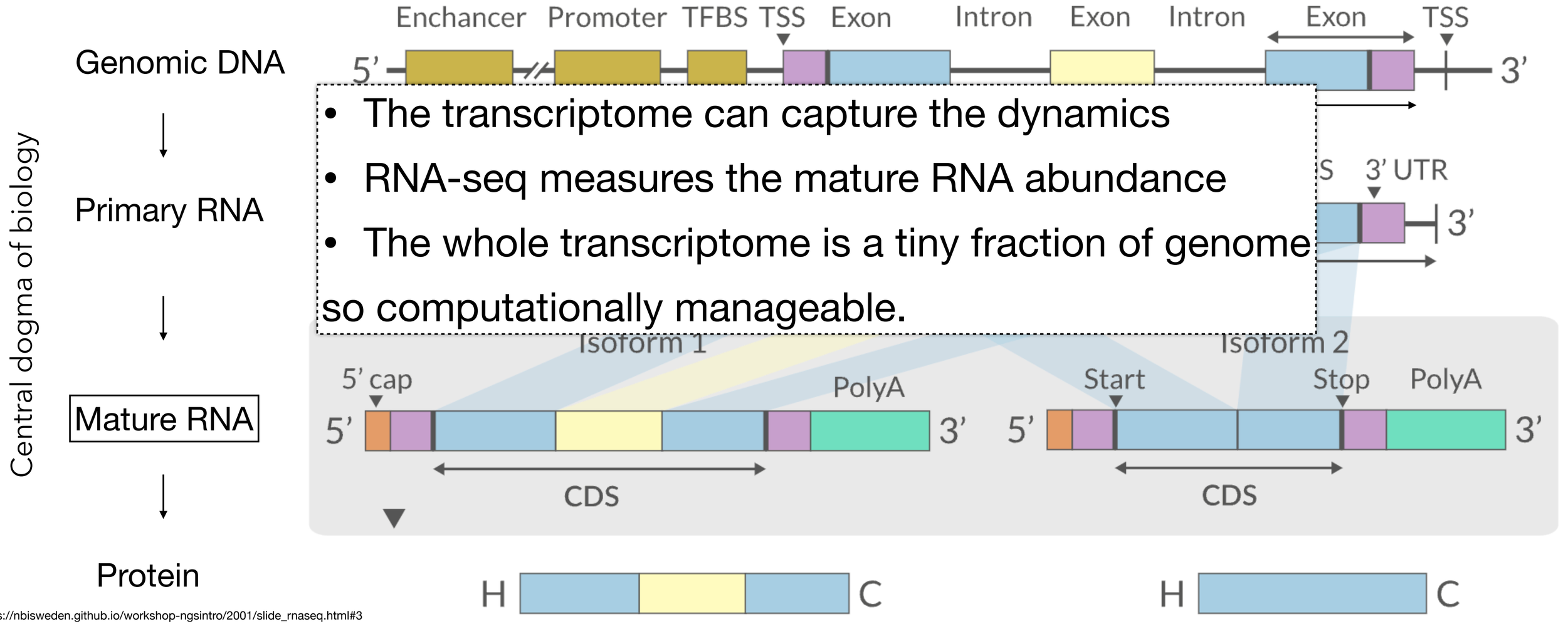
RNA Sequencing (*not* simplified)



https://nbisweden.github.io/workshop-ngsintro/2001/slide_rnaseq.html#3

How to measure gene expression

RNA Sequencing (*not* simplified)



- The transcriptome can capture the dynamics
- RNA-seq measures the mature RNA abundance
- The whole transcriptome is a tiny fraction of genome

so computationally manageable.

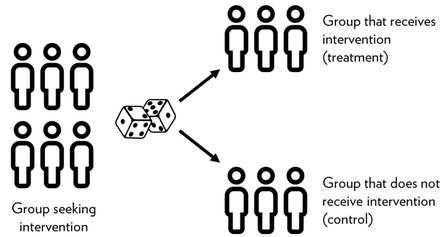
https://nbisweden.github.io/workshop-ngsintro/2001/slide_rnaseq.html#3

What can we do with RNA-seq

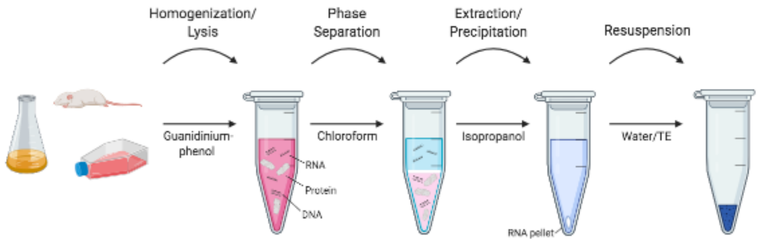
- Measure the expression of genes across different conditions.
- Understanding function of gene.
- Differential gene expression analysis.
  - Measure the degree of change in different conditions (e.g. a sample with special disease).
  - Discover pathways, gene networks, co-expression.
  - Fusion detection from genes
  - Much much more …

How can we do such exciting stuff

Experimental design



Experimental design

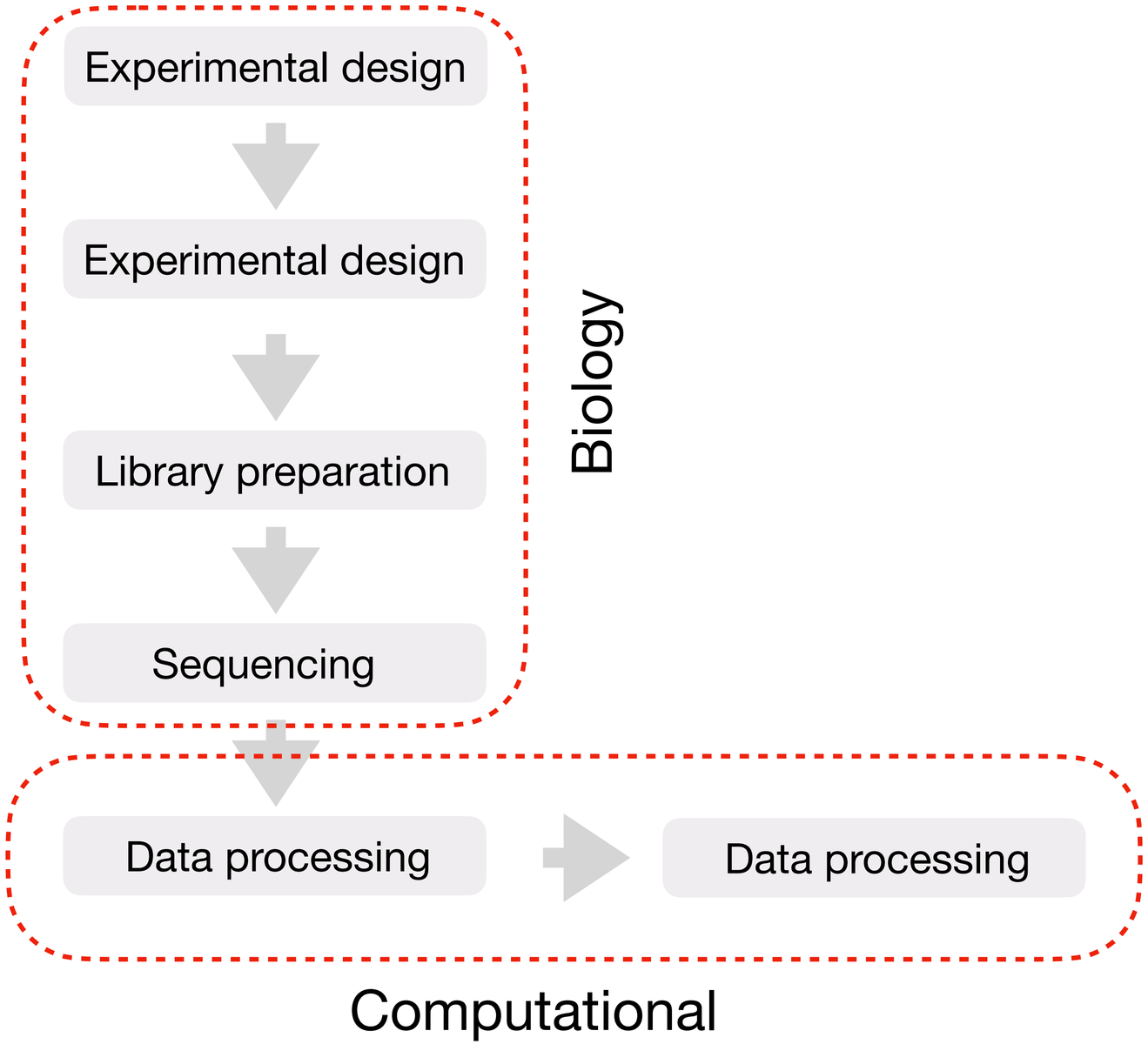


Library preparation

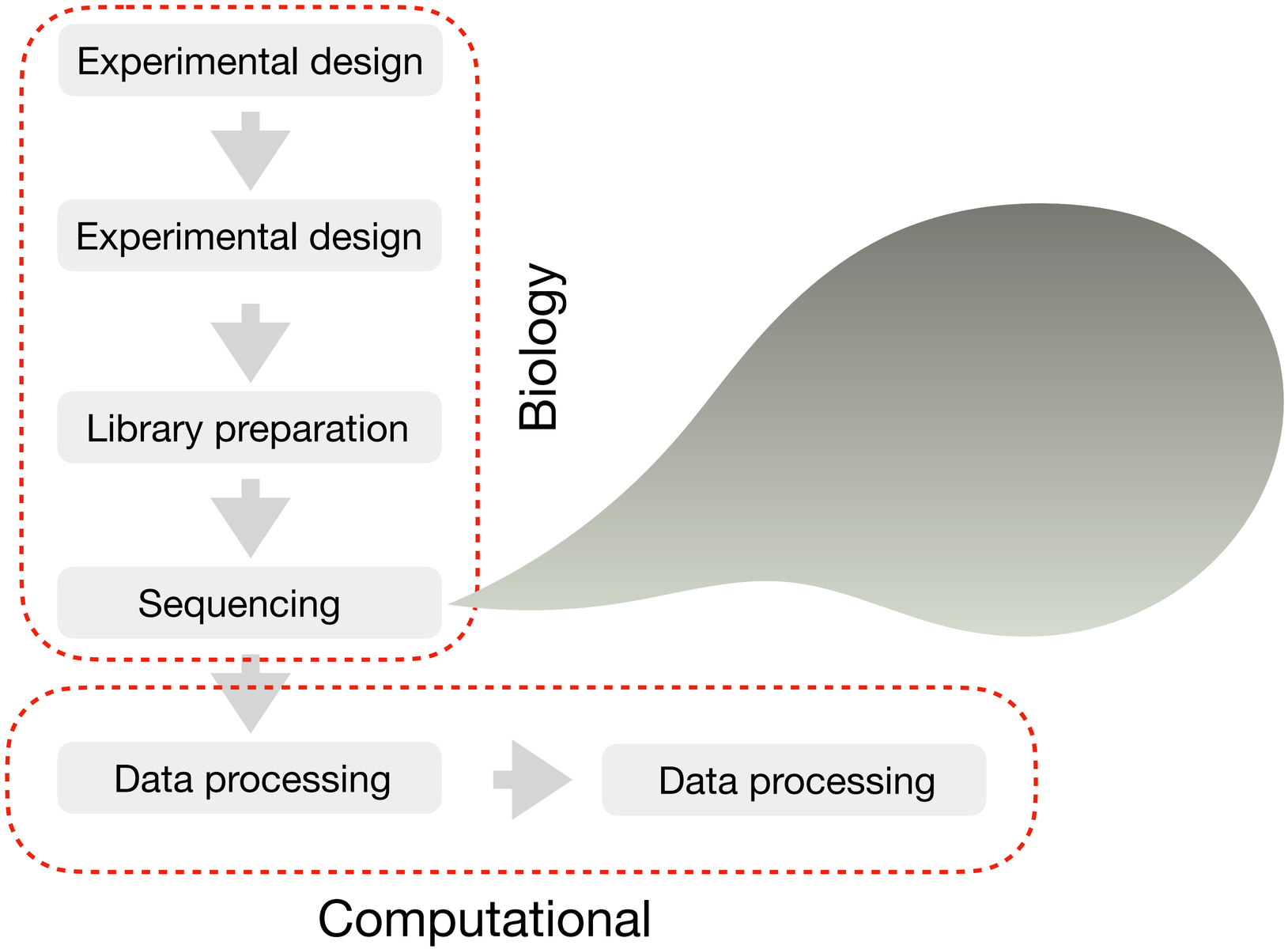
Sequencing

Data processing → Data processing

How can we do such exciting stuff



How can we do such exciting stuff

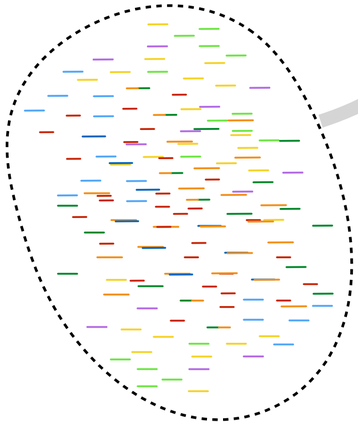


Basic workflow for RNA-seq (or any sequence based)

Reads

10s-100s of millions of
short (35-300 character) “fragments”

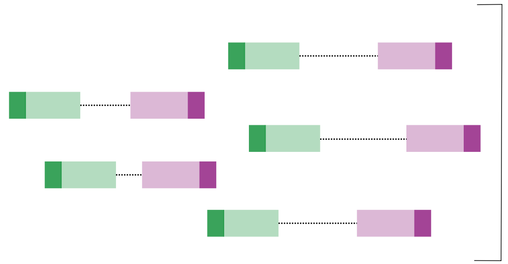
Measurement
(RNA-seq)



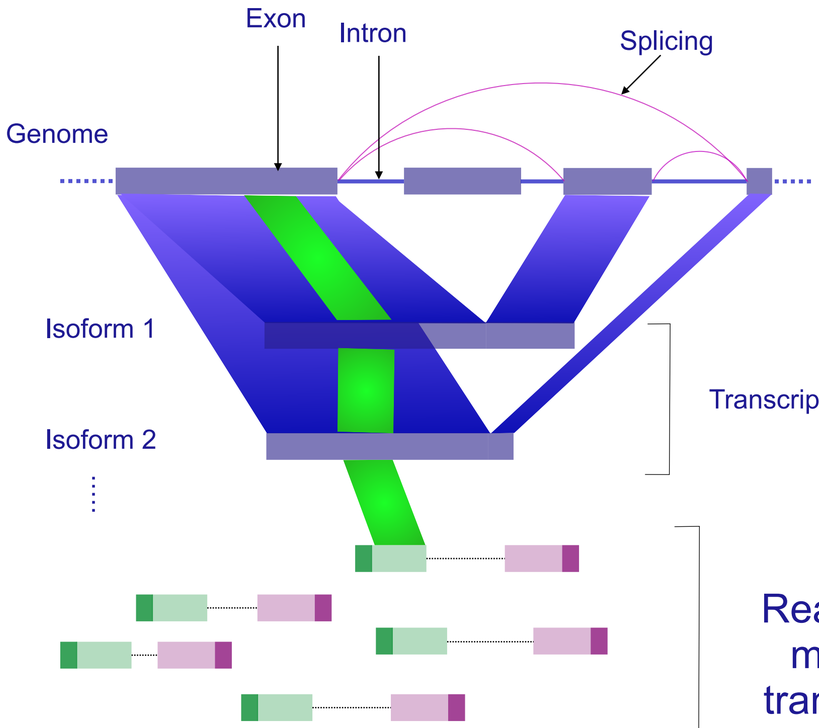
QC

Mapping

Quantification



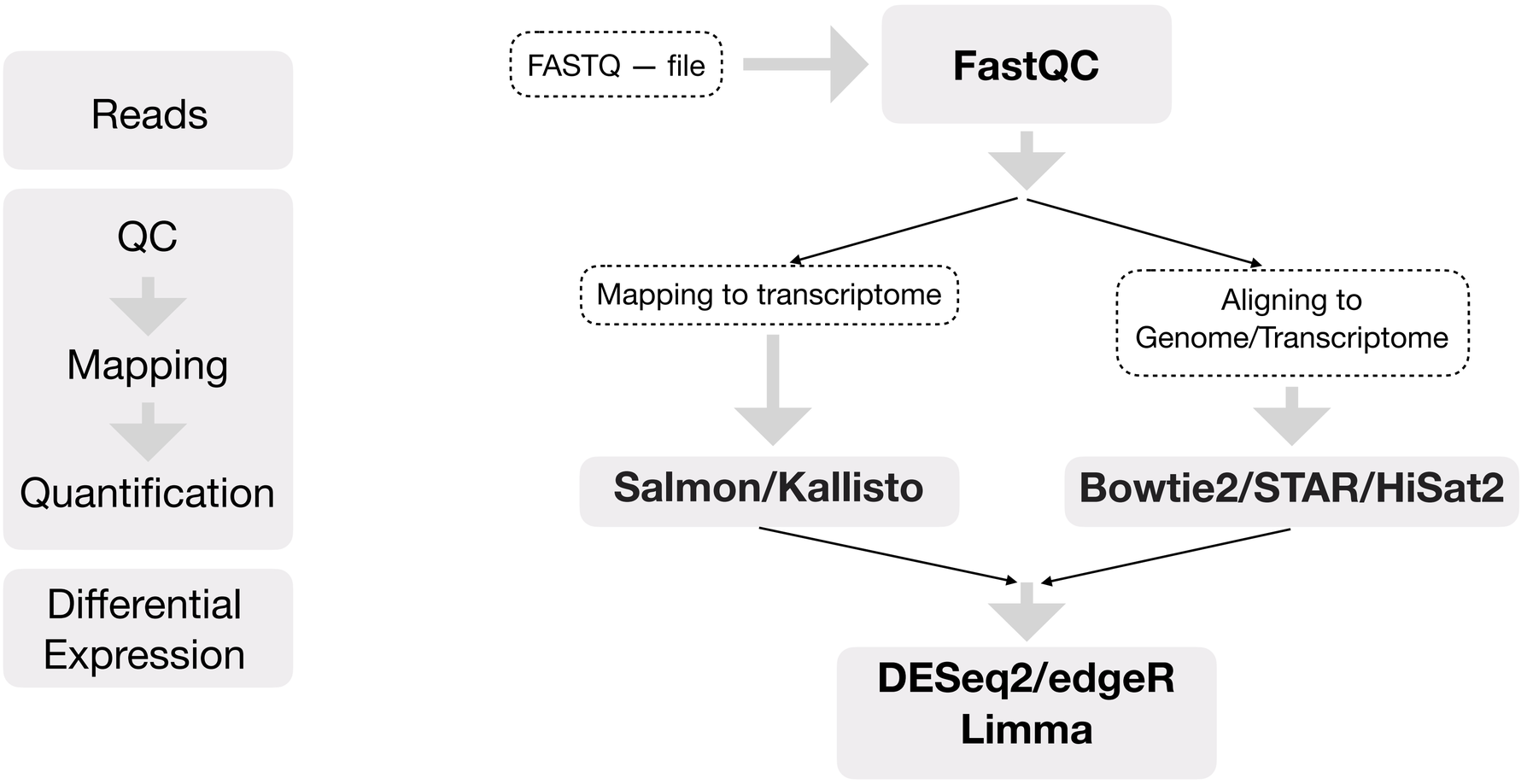
Reads can be mapped to genome



Reads can be mapped to transcriptome

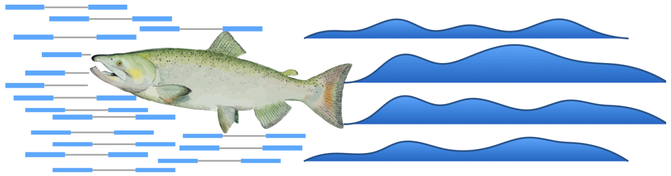
https://nbisweden.github.io/workshop-ngsintro/2001/slide_rnaseq.html#5

Basic workflow for RNA-seq (or any sequence based)



Tools we will use/learn today

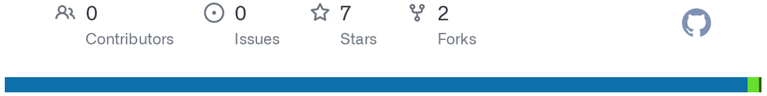
Salmon



ostrokach/**gseapy**

Gene Set Enrichment Analysis in Python

0 Contributors | 0 Issues | 7 Stars | 2 Forks

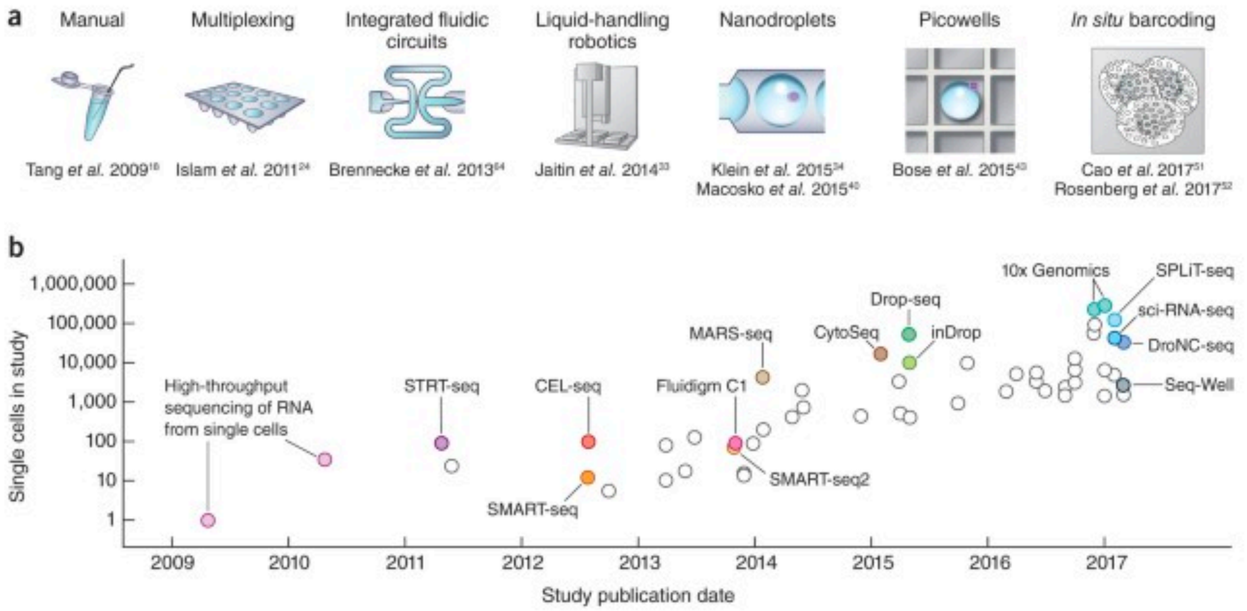


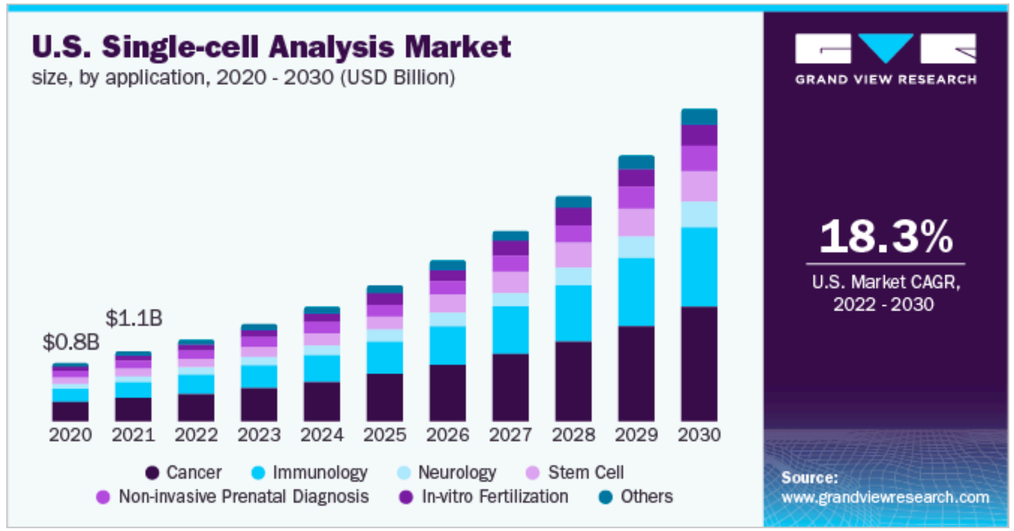
Acknowledgments

- https://rpubs.com/qshenfeng/DESeq2_analysis_tutorial
- https://rpubs.com/qshenfeng/bulk_RNAseq
- https://biocorecrg.github.io/PHINDaccess_RNAseq_2020/salmon.html
- https://www.hadriengourle.com/tutorials/rna/
- https://github.com/mousepixels/sanbomics_scripts/blob/main/PyDeseq2_DE_tutorial.ipynb
- https://www.hadriengourle.com/tutorials/rna/
- Documentation pages from softwares

Real big data — (Song of Cells)*

https://www.grandviewresearch.com/industry-analysis/single-cell-analysis-market





*The Song of the Cell: An Exploration of Medicine and the New Human Book by Siddhartha Mukherjee