THE 3rd VIETNAM SCHOOL OF BIOLOGY (VSOB-3)

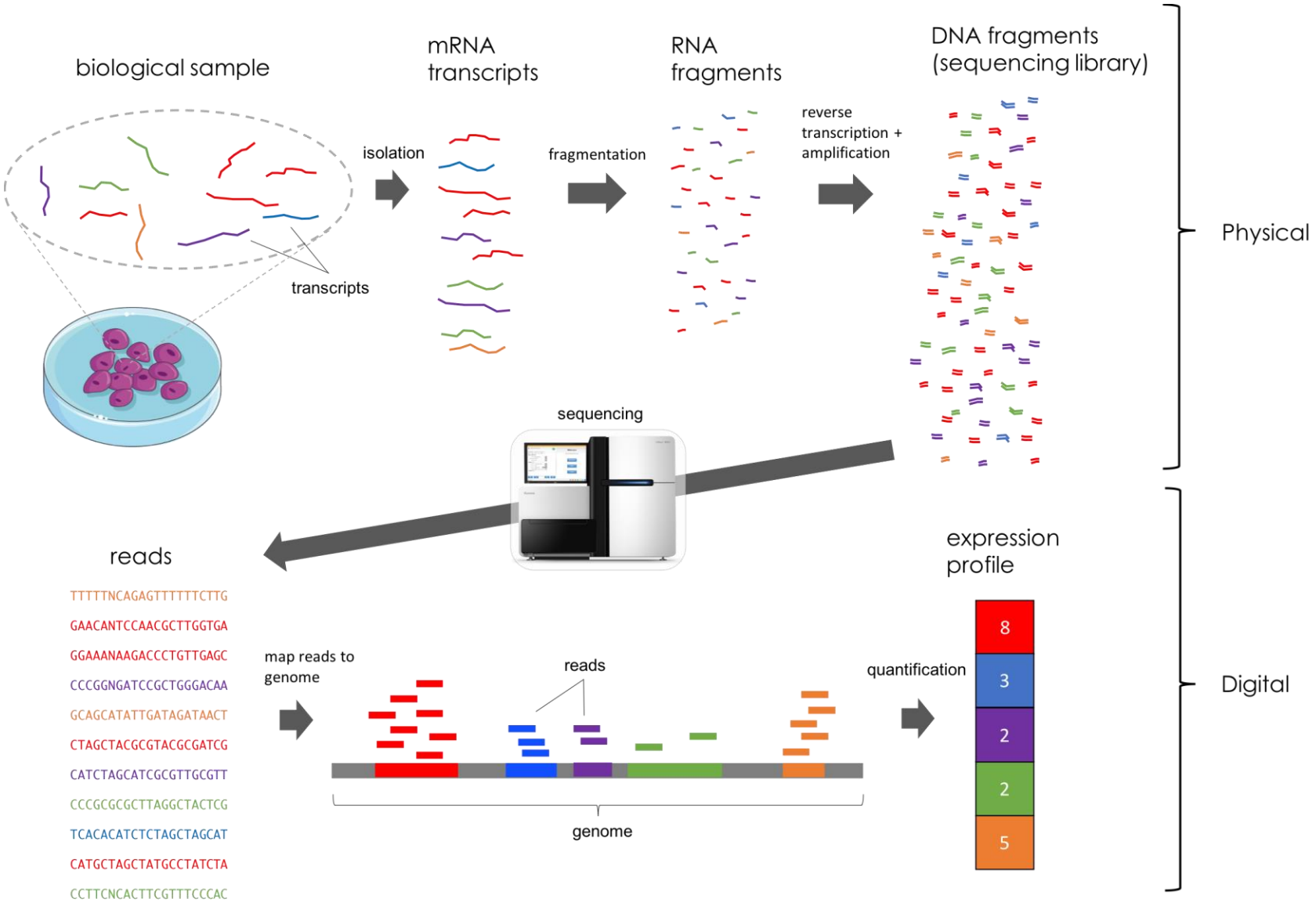
Bioinformatic Analysis For Bulk RNAseq Data

December 06th-08th, 2024, ICISE, Quy Nhon, Vietnam

A Guide for Designing and Analyzing RNAseq Data

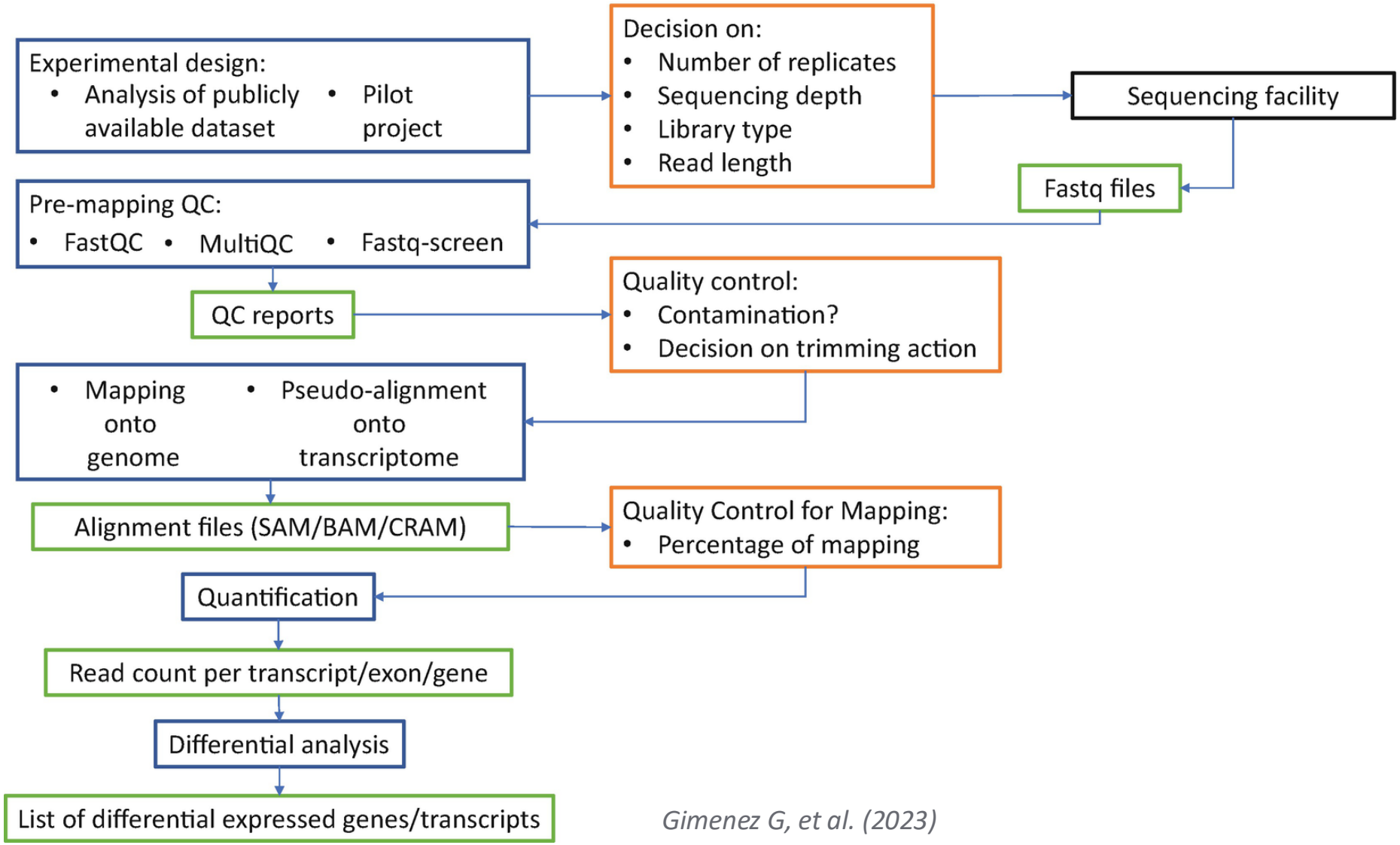
NGUYEN Thuy Vy

General steps of an RNAseq experiment



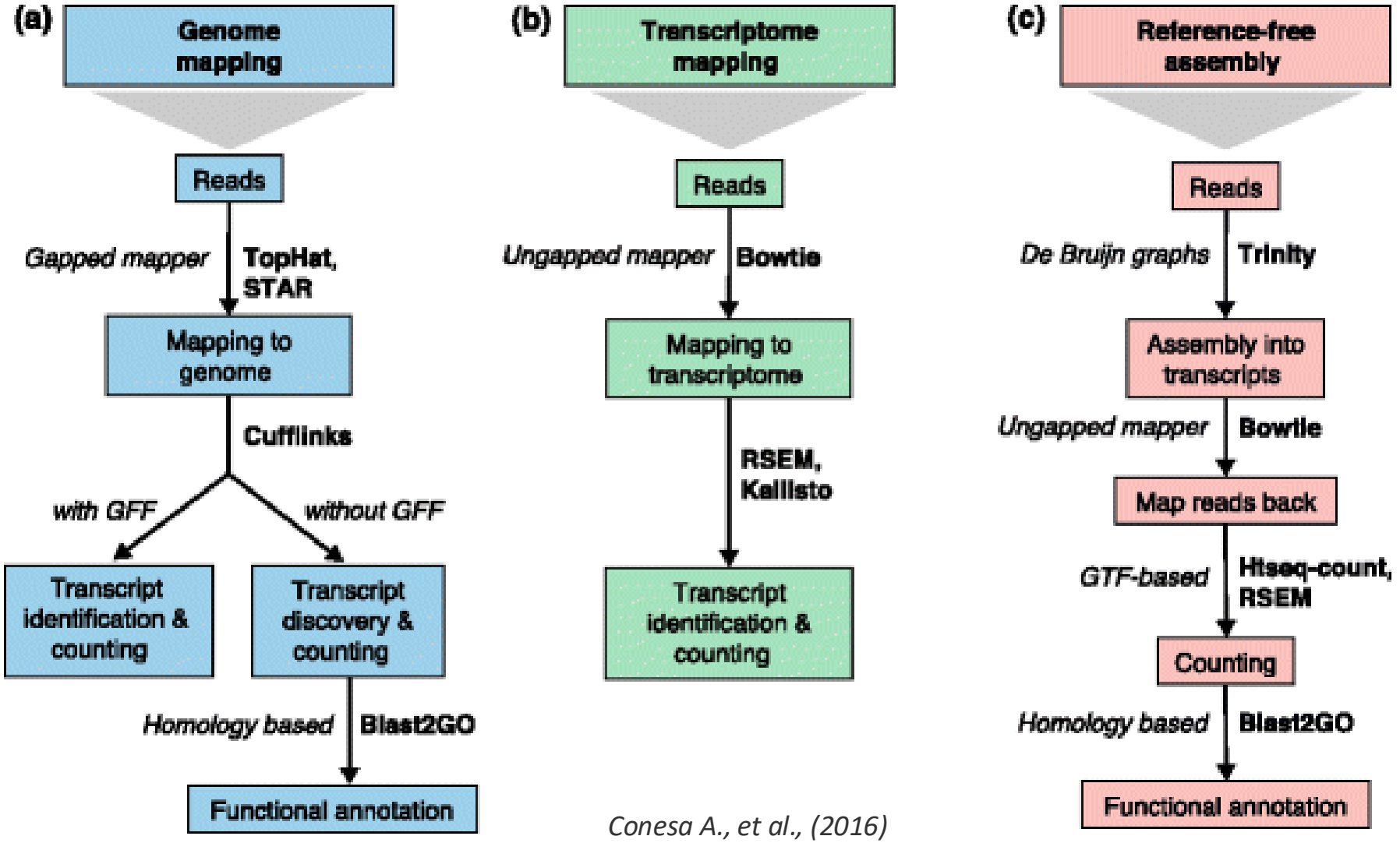
https://mbernste.github.io/posts/rna_seq_basics/

RNAseq Experimental Design and Data Analysis



Gimenez G, et al. (2023)

RNAseq Data Analysis



Conesa A., et al., (2016)

Replicates

- Compared to arrays, RNA-Seq is a very clean technical measure of expression
  - Generally, don't run **technical** replicates
  - Must run **biological** replicates

- For clean systems (eg cell lines) 3x3 or 4x4 is common
- Higher numbers required as the system gets more variable
- Always plan for at least one sample to fail
- Randomise across sample groups

How many reads

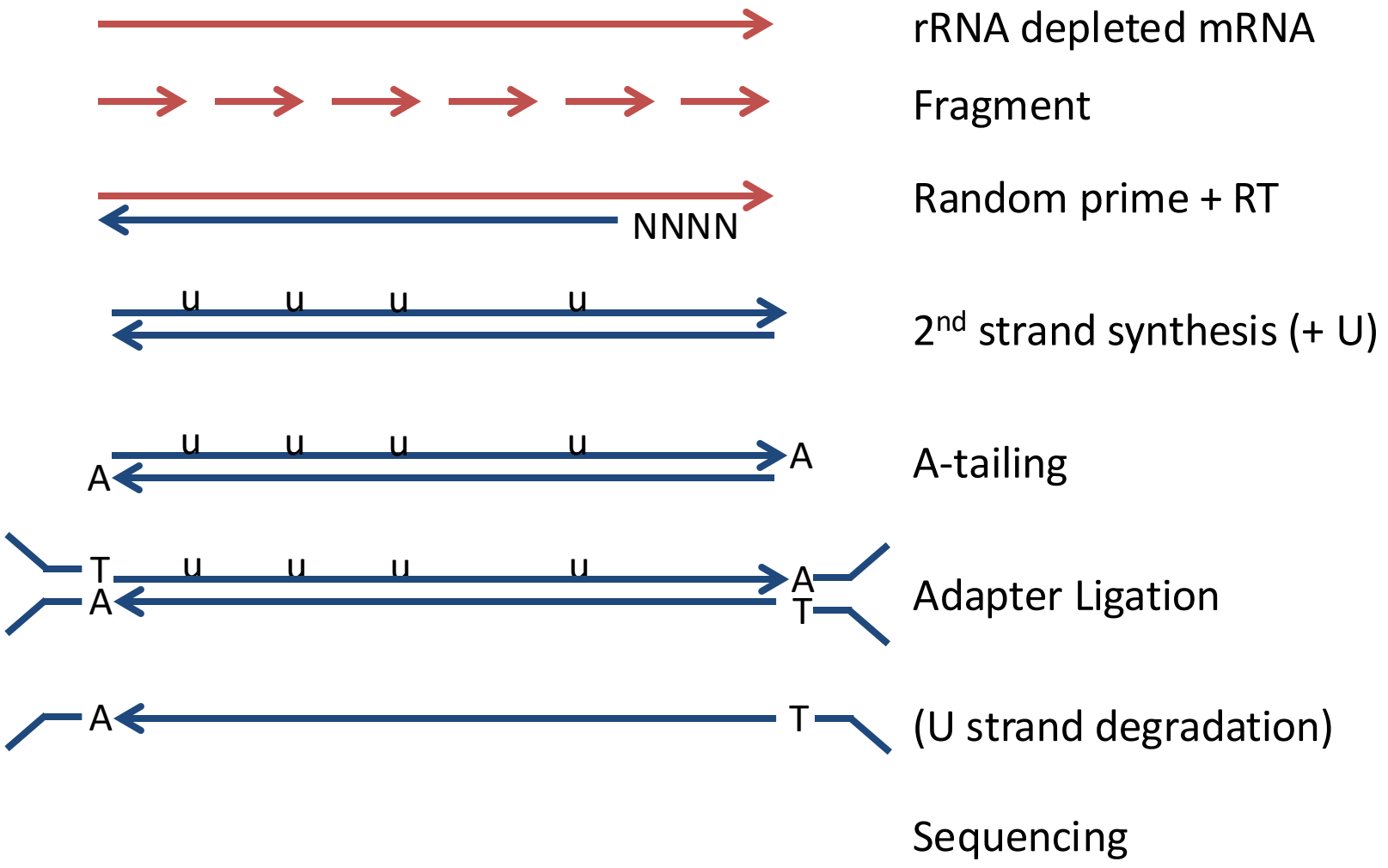
- Typically aim for 20 million reads for human/mouse sized genome
- More reads:
  - De-novo discovery
  - Low expressed transcripts
- More replicates more useful than more reads

What type of library?

- Directional libraries if possible
  - Easier to spot contamination
  - No mixed signals from antisense transcription
  - May be difficult for low input samples

- mRNA vs total vs depletion etc.
  - Down to experimental questions
  - Remember LINC RNA may not have polyA tail
  - Active transcription vs standing mRNA pool

RNAseq Libraries

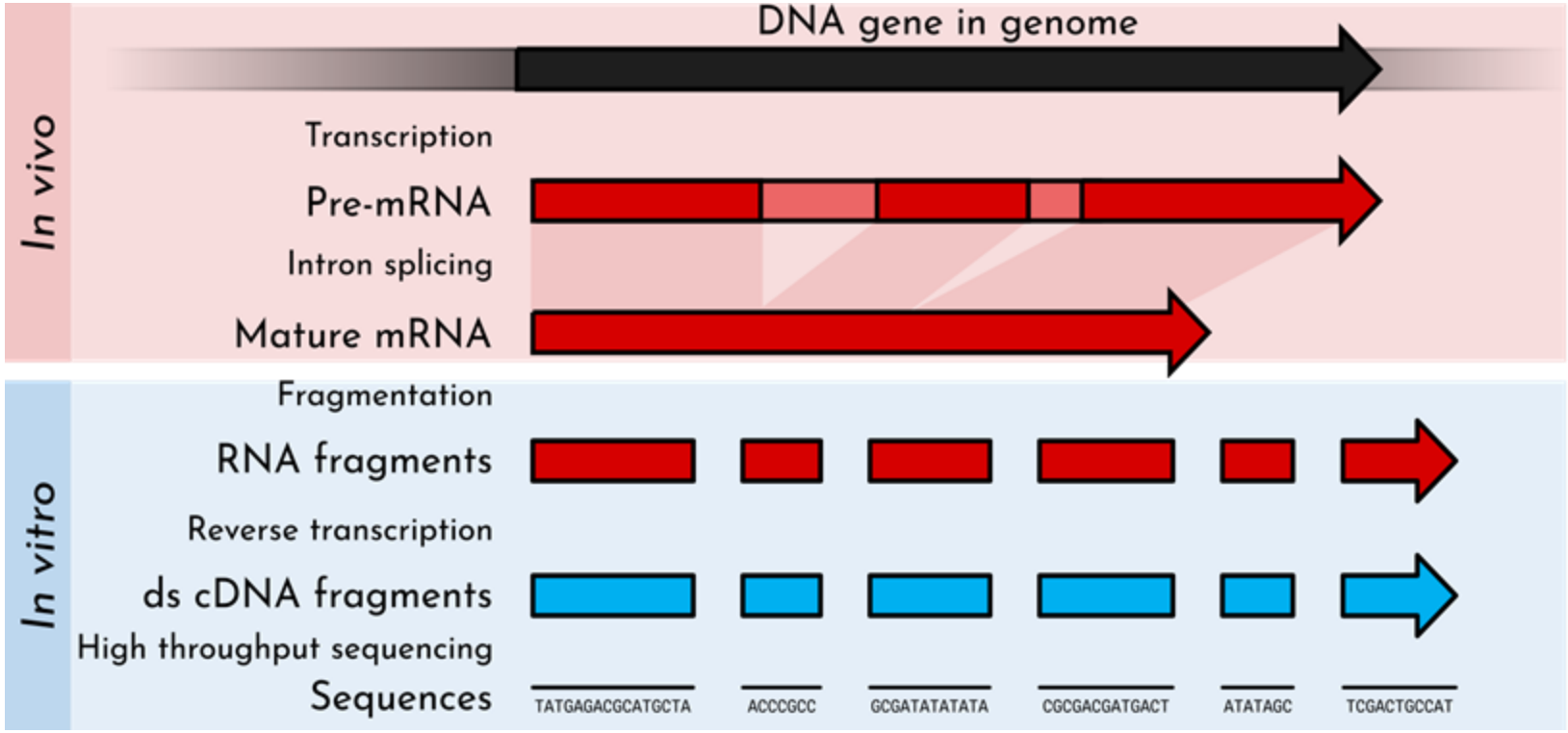


What type of sequencing

- Depends on your interest
  - Expression quantitation of known genes
    - 50 bp single end is fine
  - Expression plus splice junction usage
    - 100 bp (or longer if possible) single end
  - Novel transcript discovery or per transcript expression
    - > 100 bp paired end

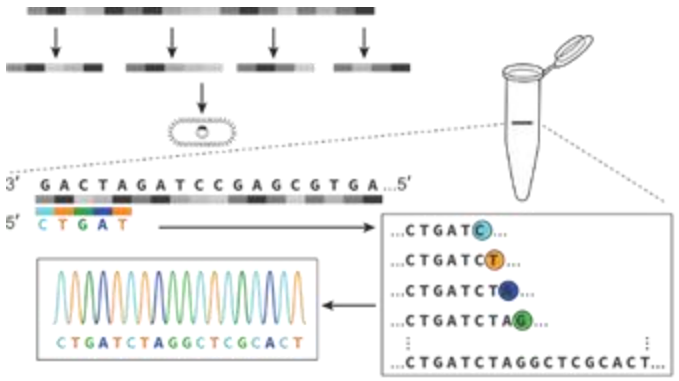
Available platforms for RNAseq

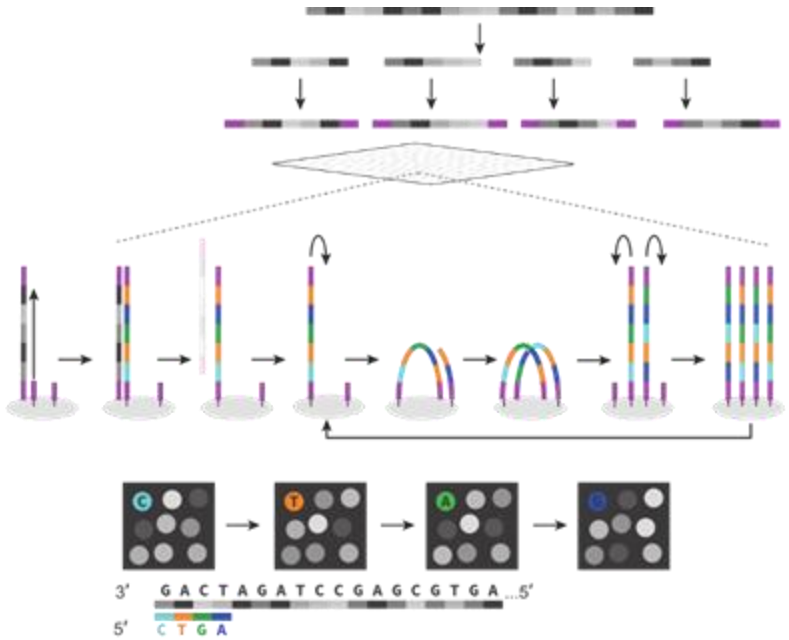
The basic principle behind RNAseq is the construction of a library, from there, we can choose different platforms to sequence our library

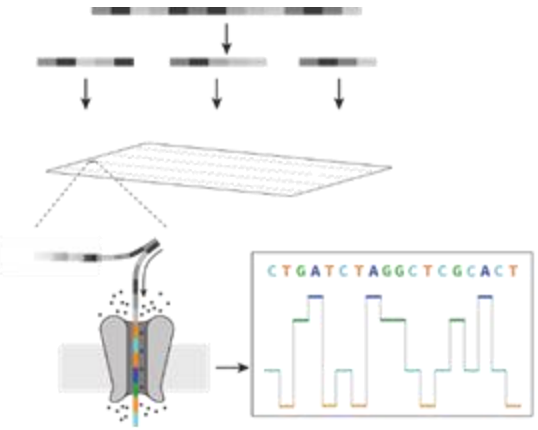


Lowe. 2017. doi: 10.1371/journal.pcbi.1005457

Available platforms for RNAseq, an overview







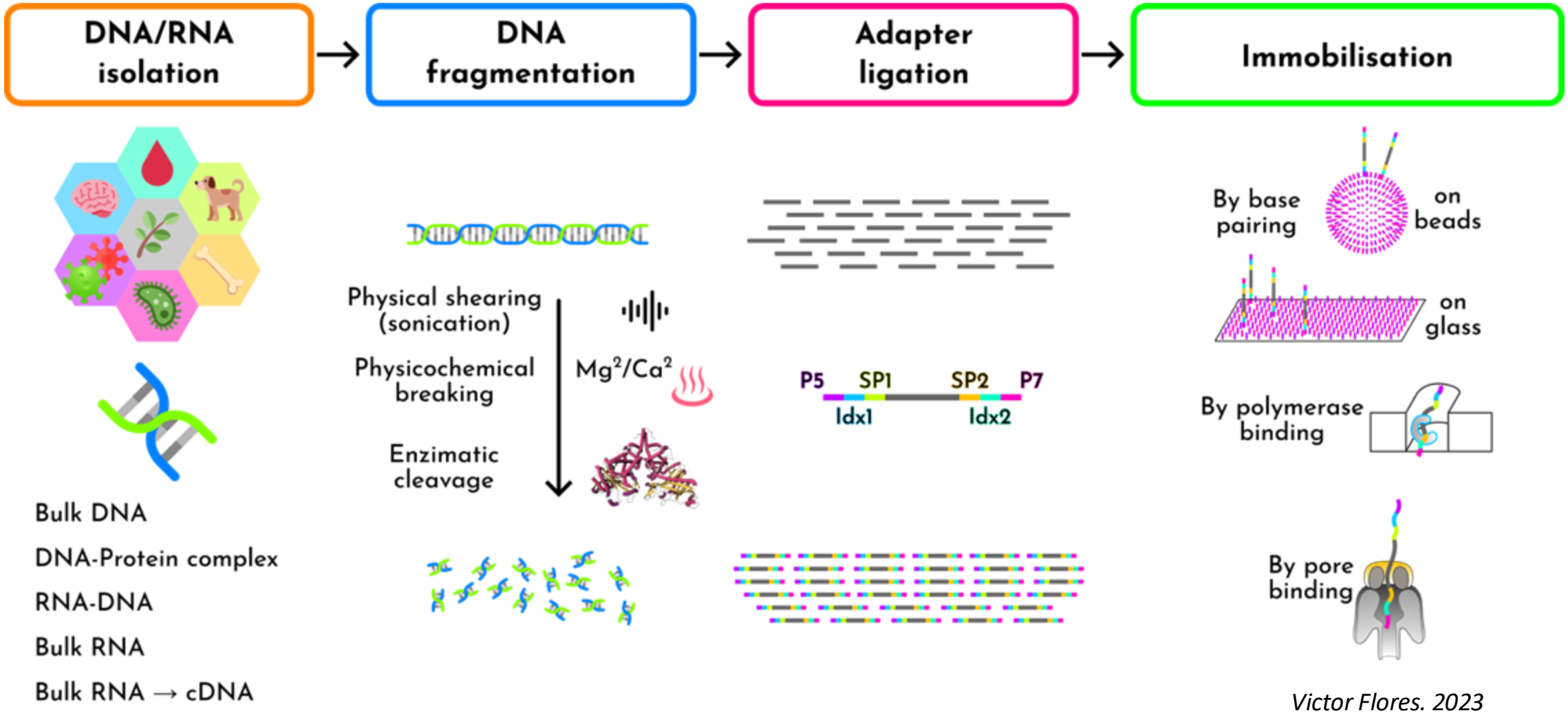
1st Generation
- Up to 1000 bp/fragment
- Up to 96 frags/run

2nd Generation
- Up to **600** bp/frag **(2 x 300)**
- Billions of frags/run

3rd Generation
- Up to **4.4 Mbp/frag**
- Millions of frags/run

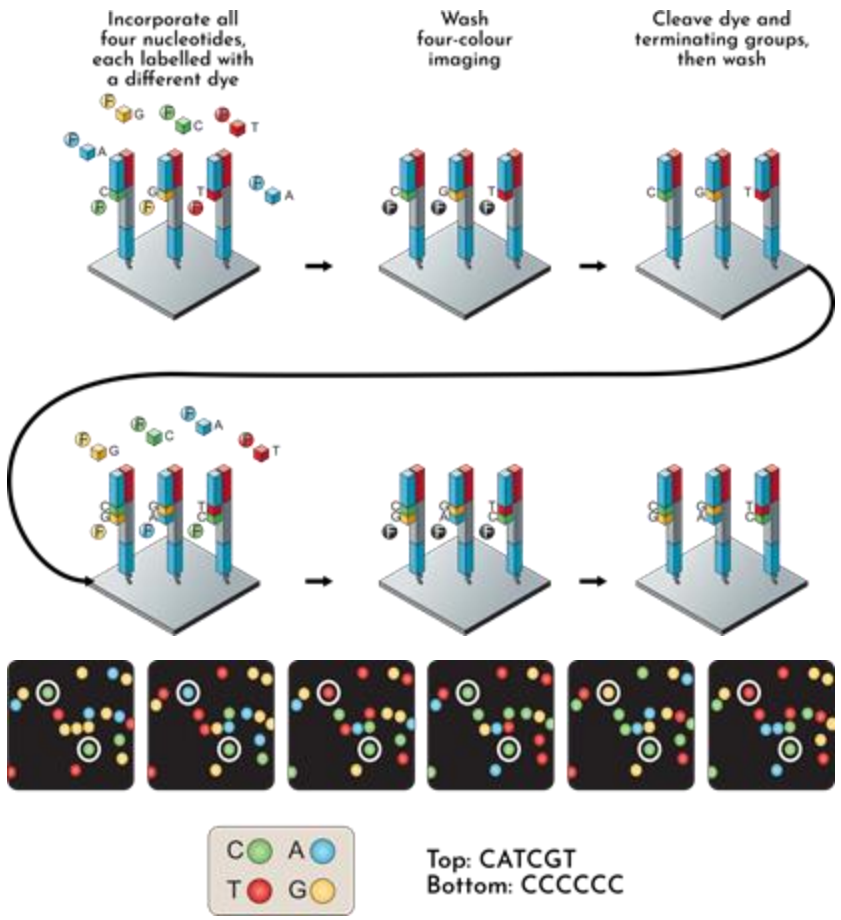
Adapted from Shendure. 2017. doi: 10.1038/nature24286. Numbers updated from 2021 data.

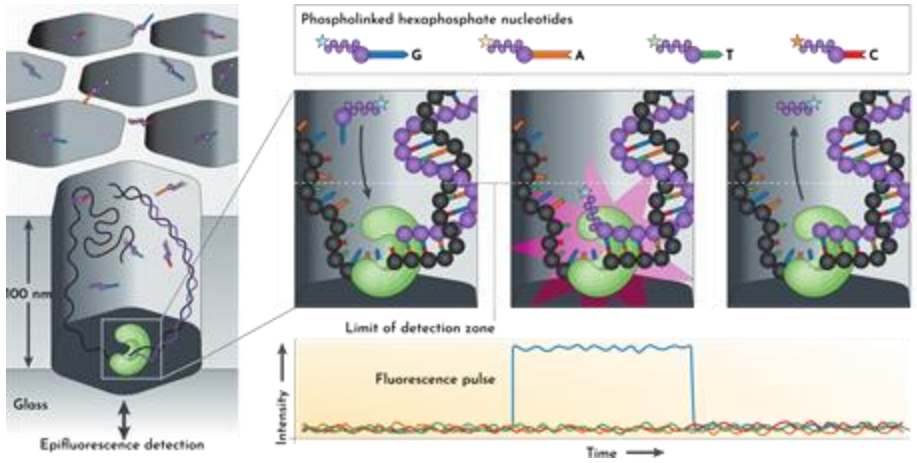
Available platforms for RNAseq: common themes

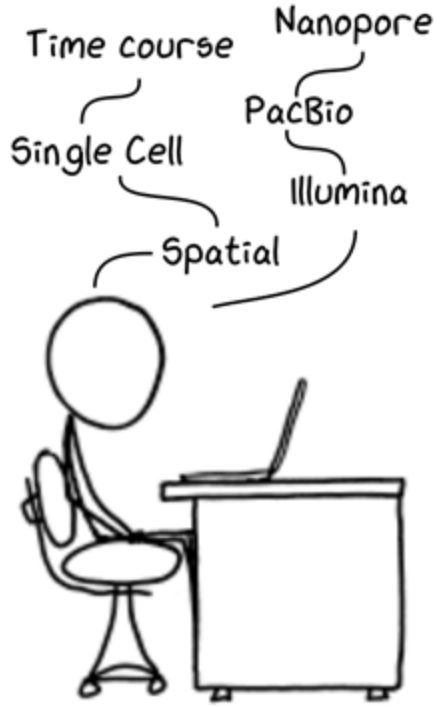


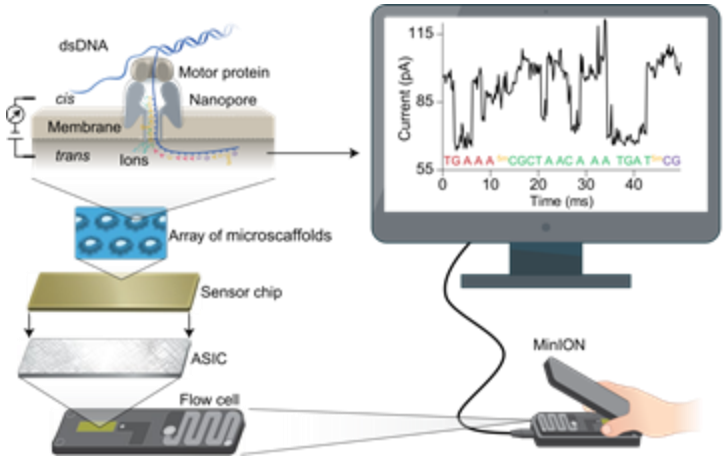
Victor Flores. 2023

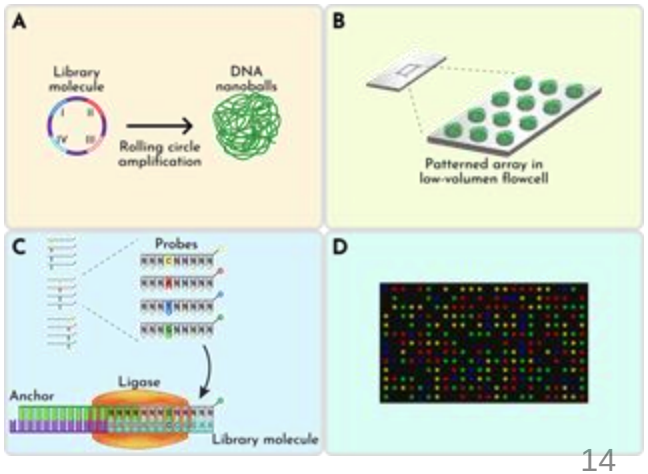
Which platform should I use?



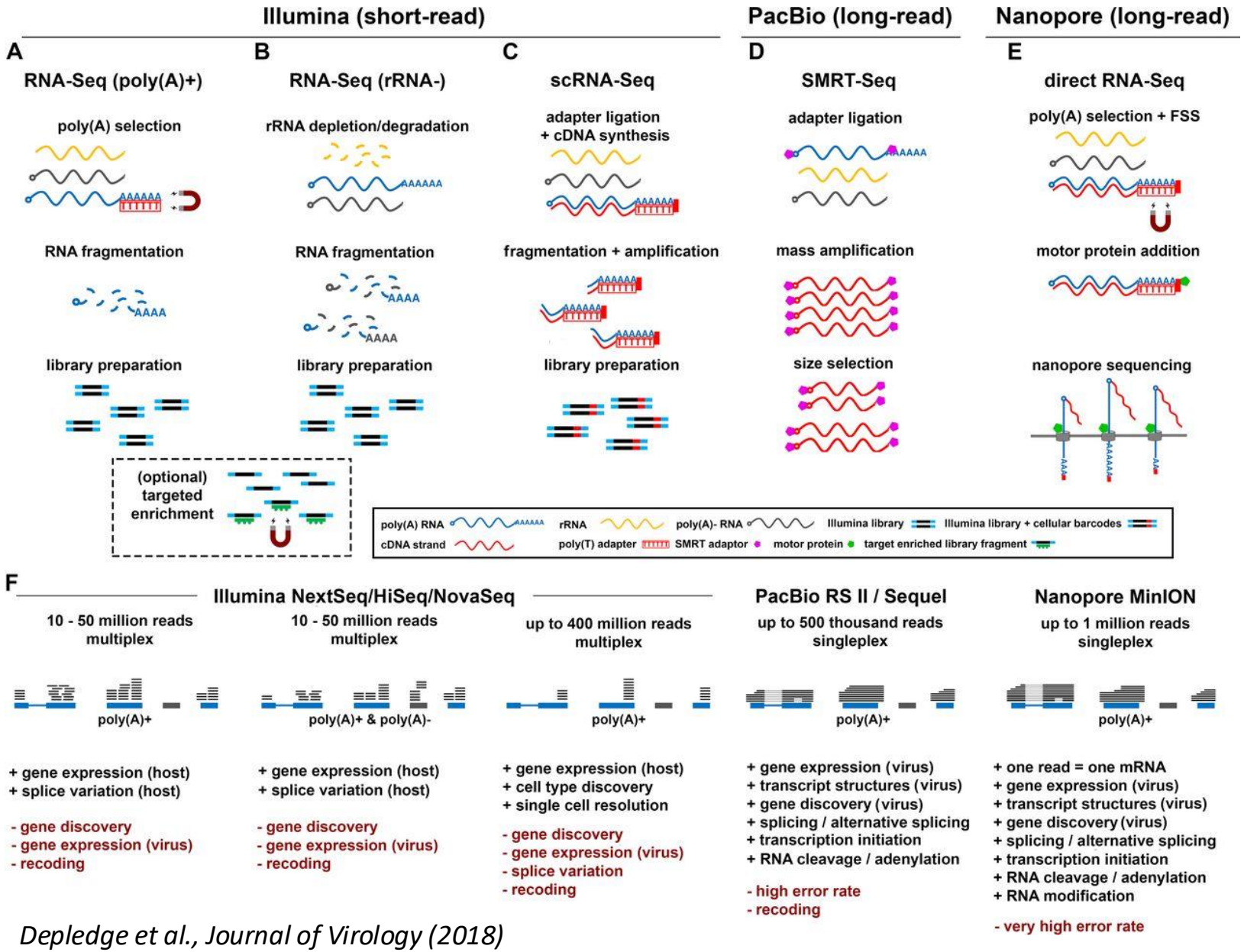






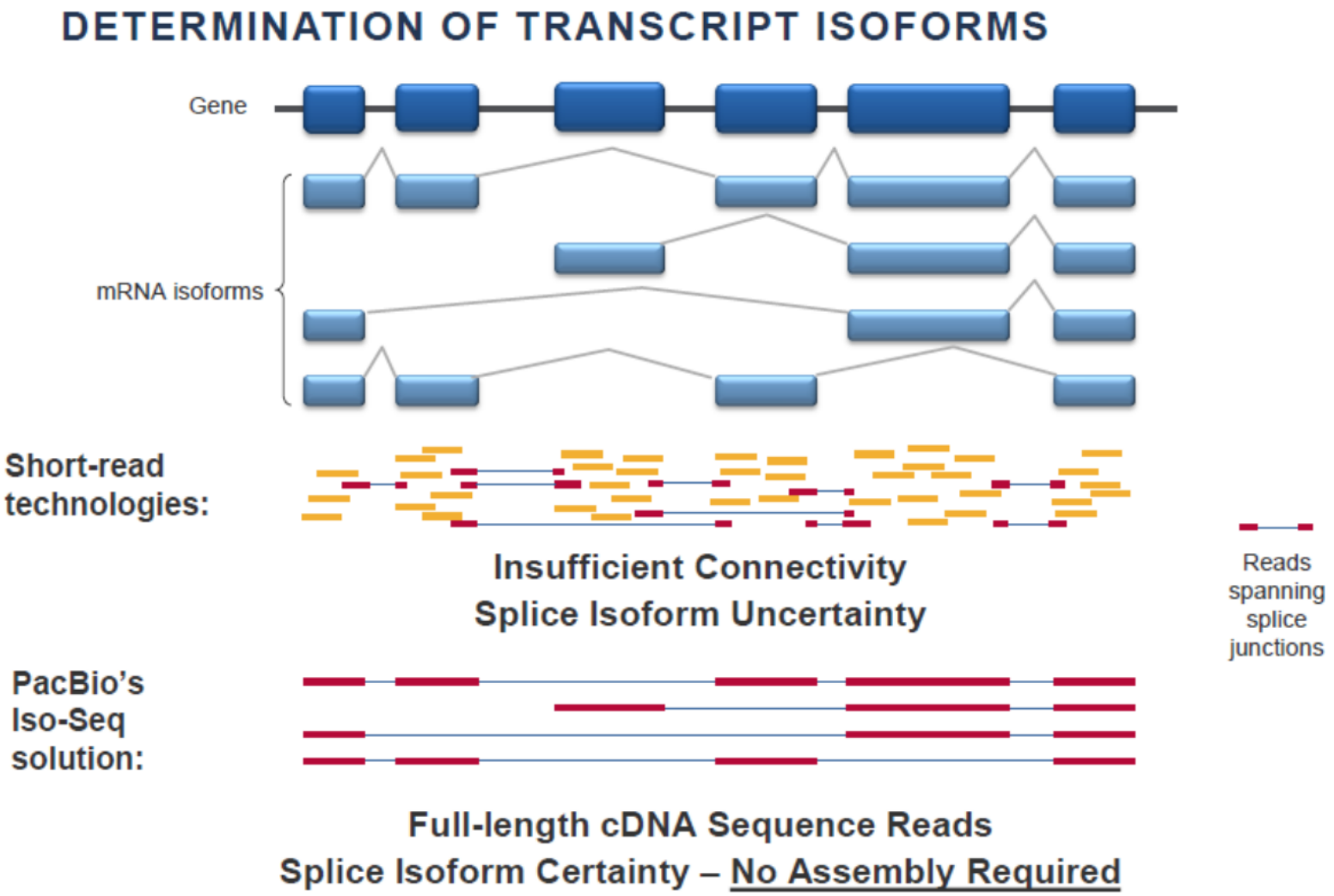


Major RNA sequencing methodologies



Depledge et al., Journal of Virology (2018)

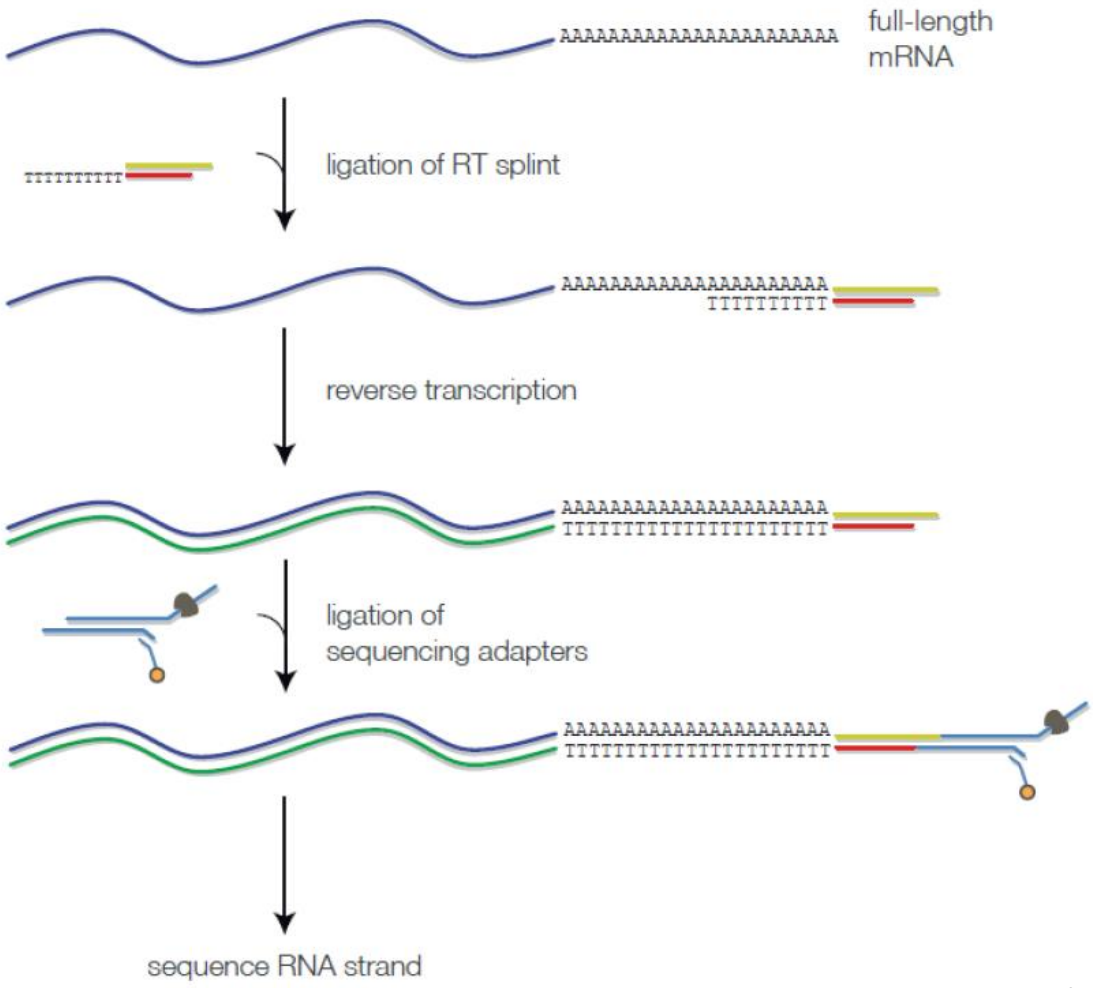
Iso-Seq by PacBio



Direct Nanopore sequencing of RNA and its modifications

- "3-prime down" reading
- RNA-specific motor protein
- RNA directly read in nanopore
- Modified bases distinguishable
- Long read lengths
- cDNA is not sequenced

- Reverse transcription bias (?)
- PCR-free
- More quantitative (?)



Garalde et al., Nature Methods (2018)

We are happy to help you!