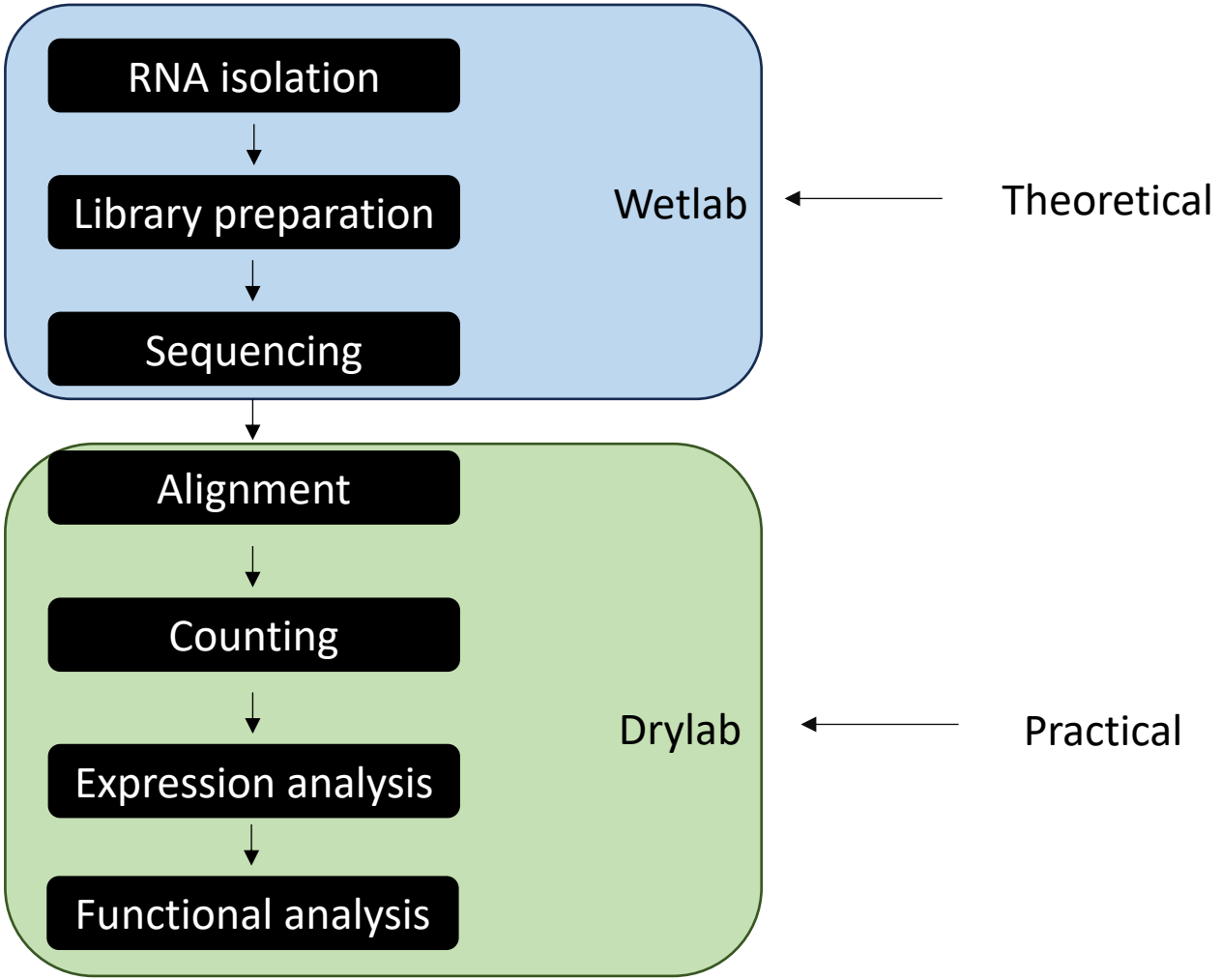
2. Gene expression profiling

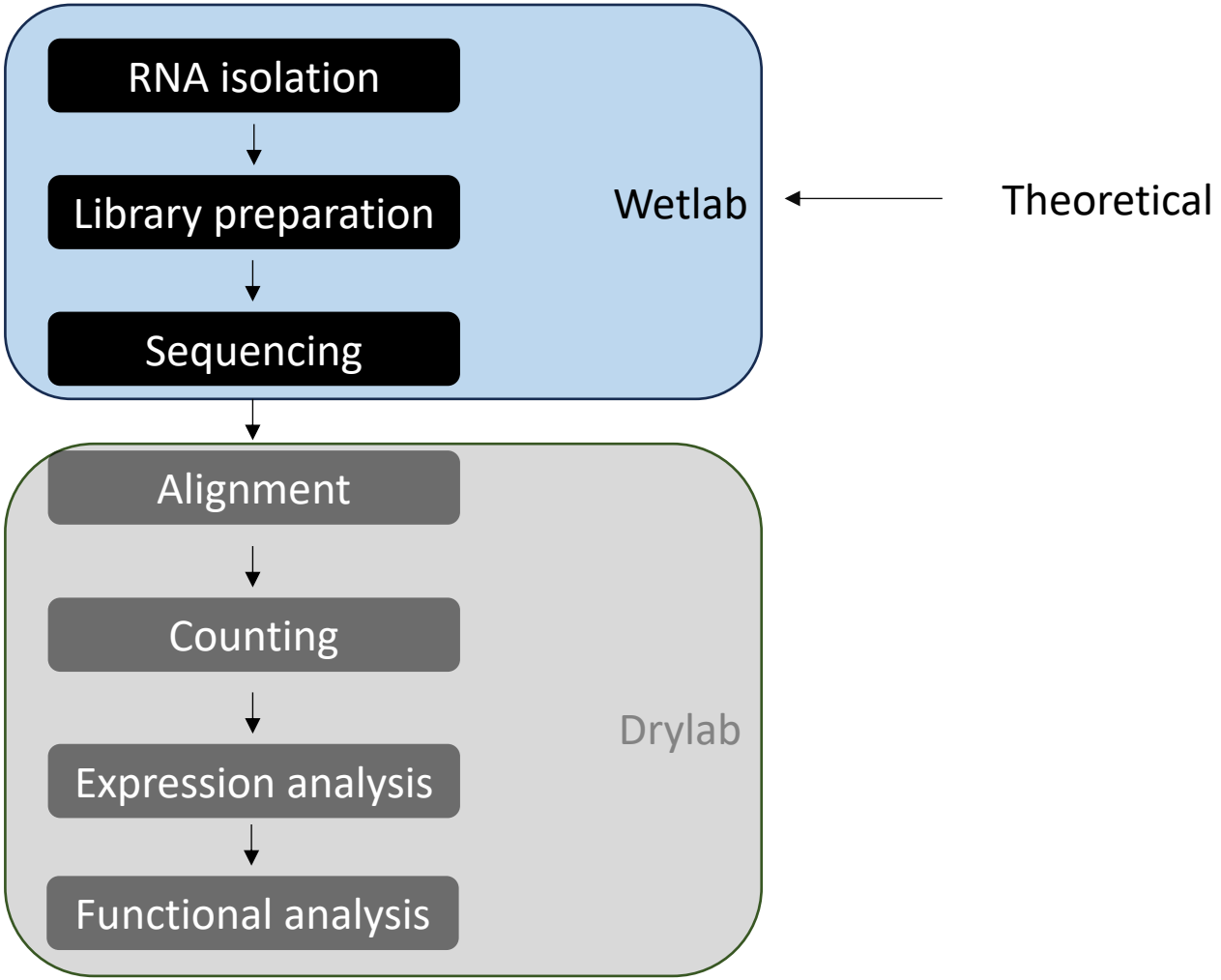
I. Gene expression profiling at the bulk level.
II. Gene expression profiling at the single-cell level.
III. Gene expression profiling at the spatial level.

2.1 Gene expression profiling at the bulk level



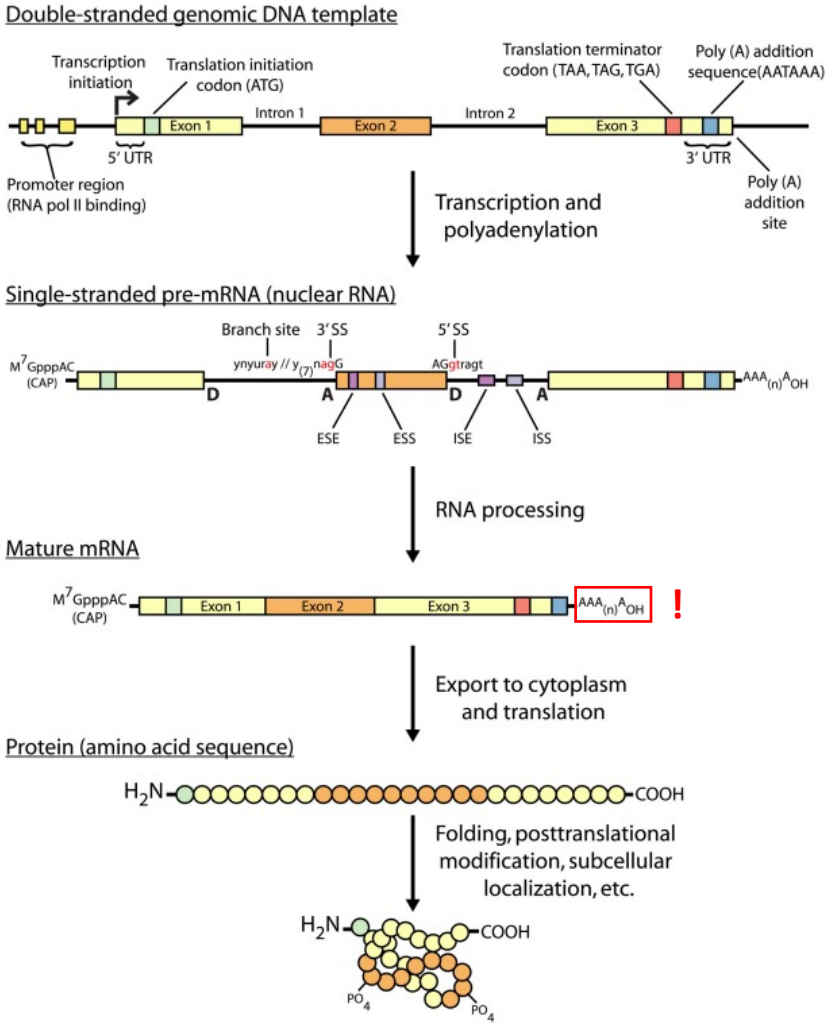
Ref: Griffith *et al.* 2015

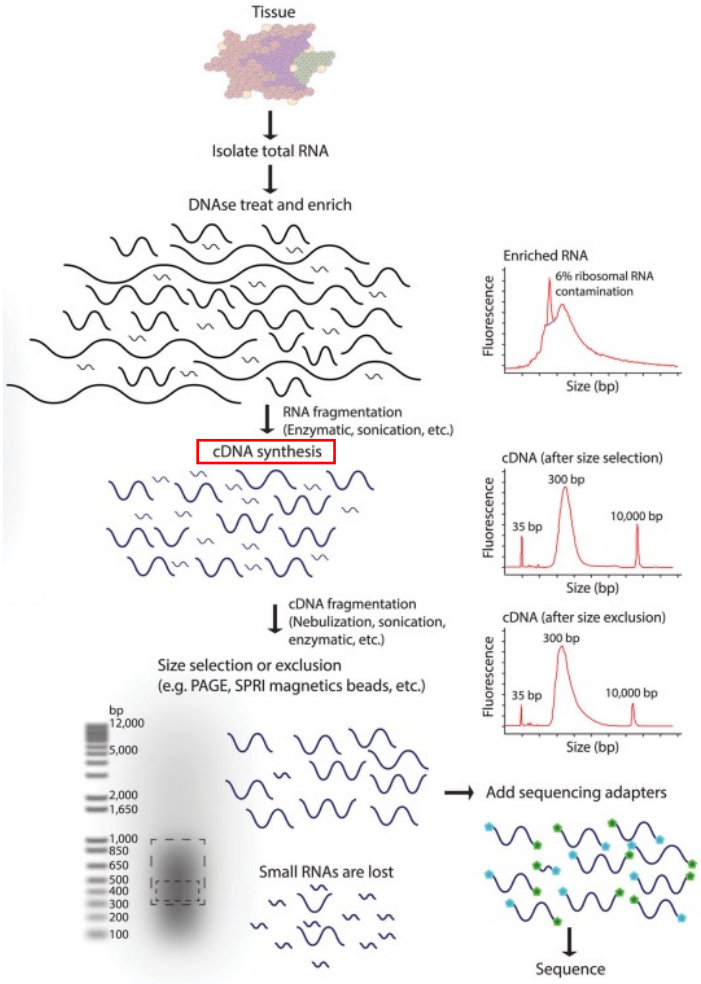
2.1 Gene expression profiling at the bulk level



Ref: Griffith *et al.* 2015

2.1 Gene expression profiling at the bulk level

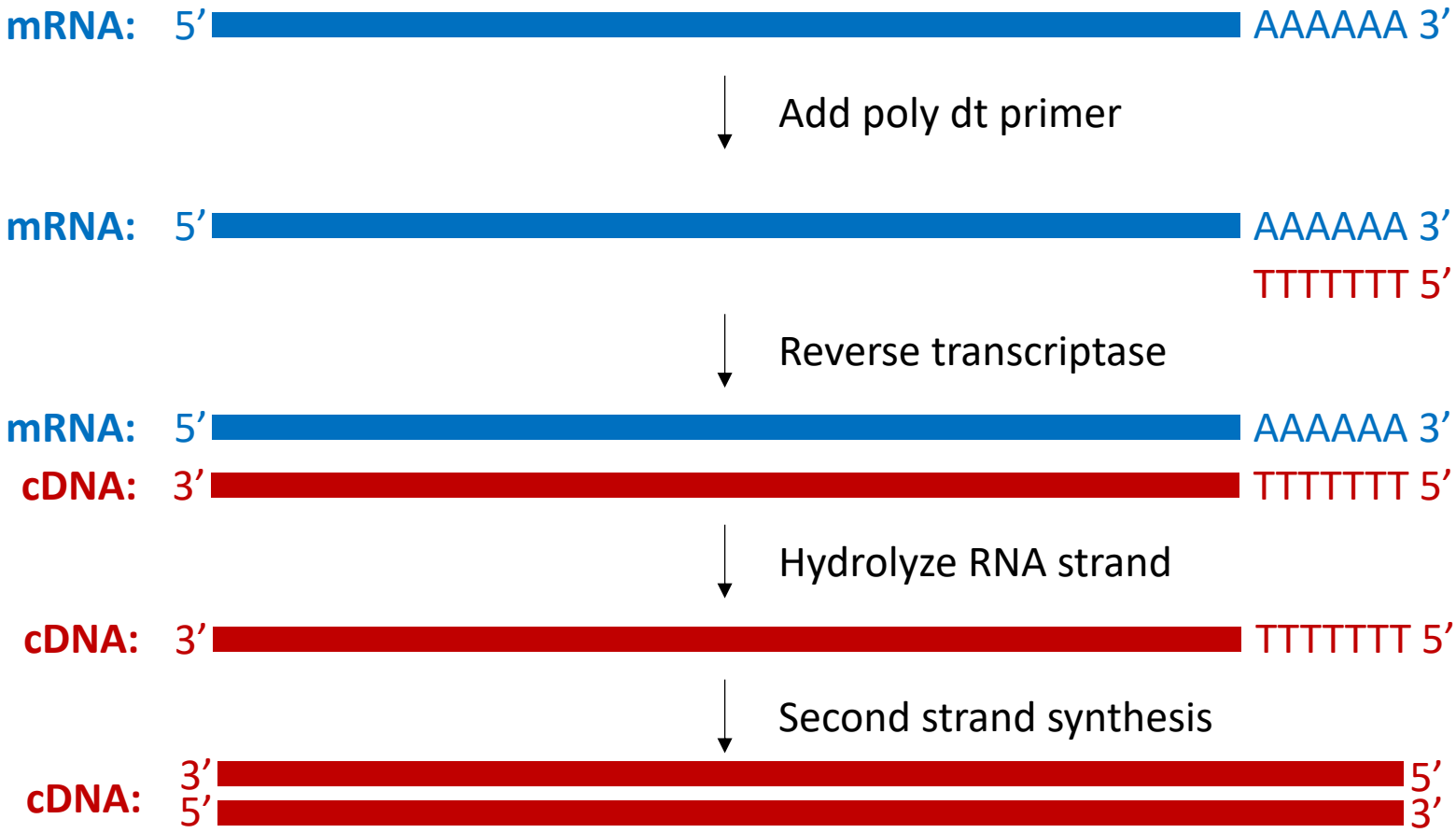




Ref: Griffith *et al.* 2015

2.1 Gene expression profiling at the bulk level

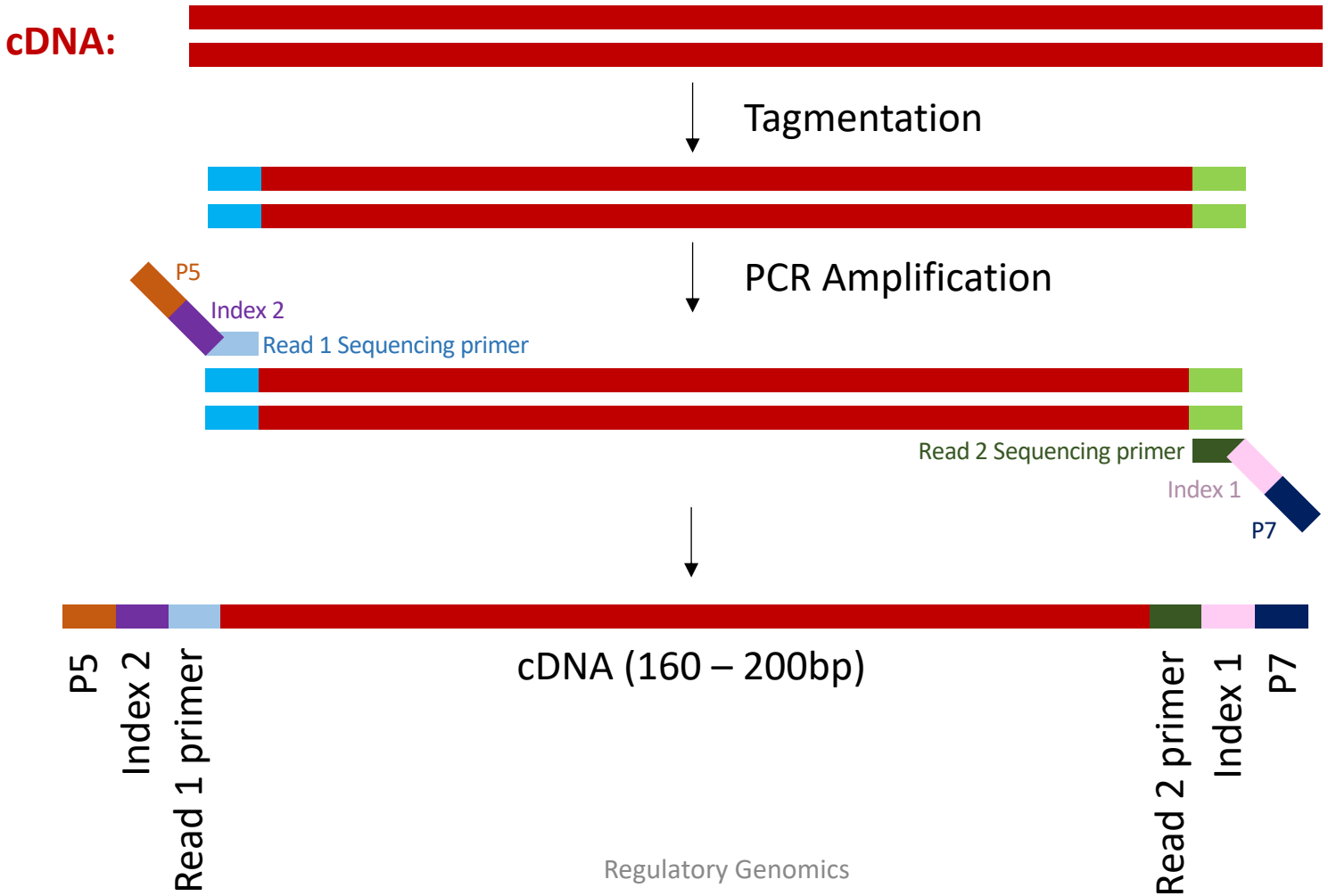
cDNA synthesis



Why cDNA synthesis? DNA is <u>a lot</u> more stable compared to RNA!

2.1 Gene expression profiling at the bulk level

Library prep



2.1 Gene expression profiling at the bulk level

What is the index?

The index is a short (6-8 nt) unique barcode which is used to identify each sample (unique identifier for each sample). These barcodes should be tolerant of 1-2 sequencing errors.

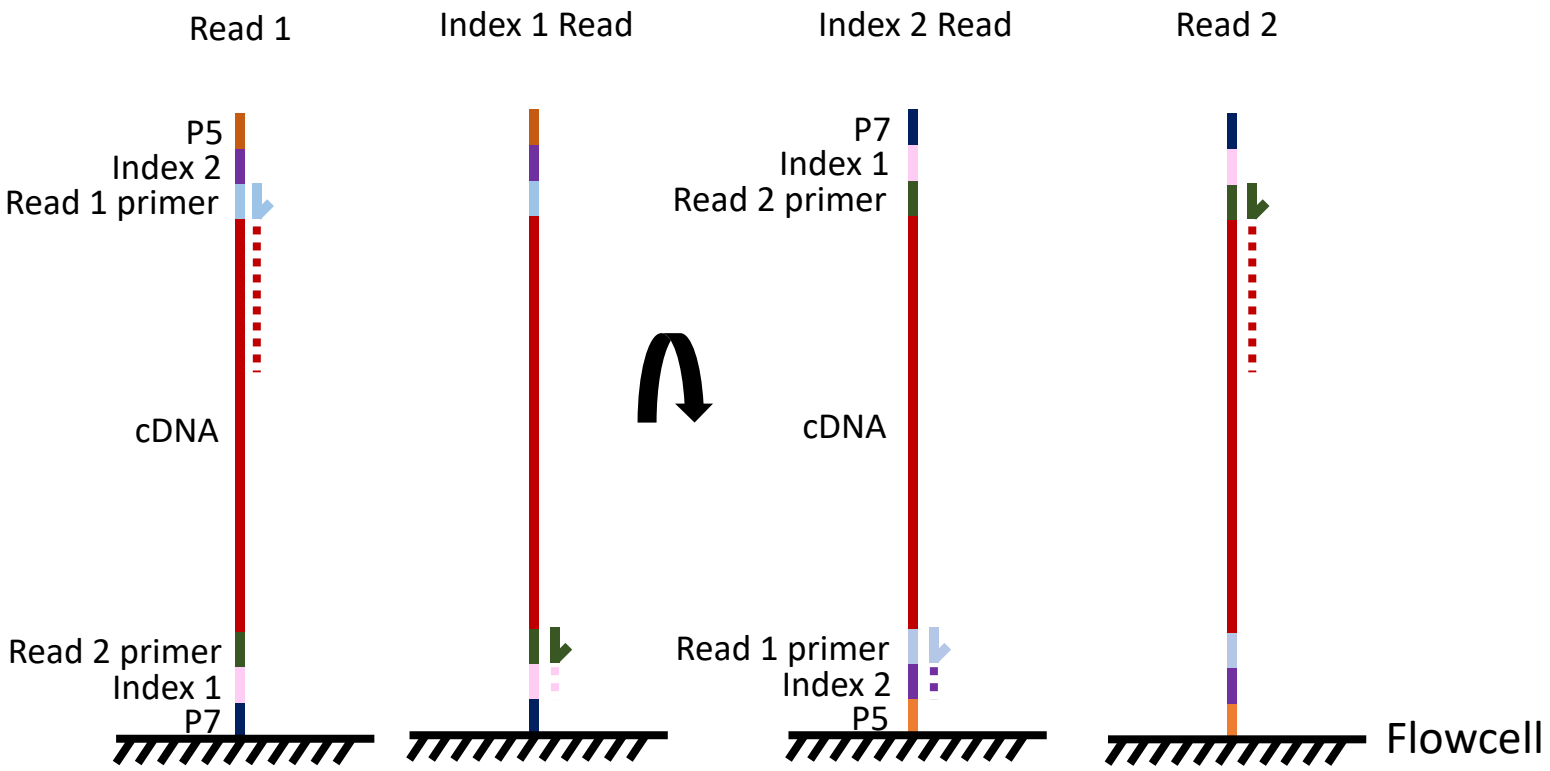
Why use an index?

The use of these unique barcodes allows pooling of multiple samples on a single sequencing lane. The barcodes can be used to computationally demultiplex the samples.

This furthermore mitigates lane effects and allows sequencing capacity to be used efficiently.

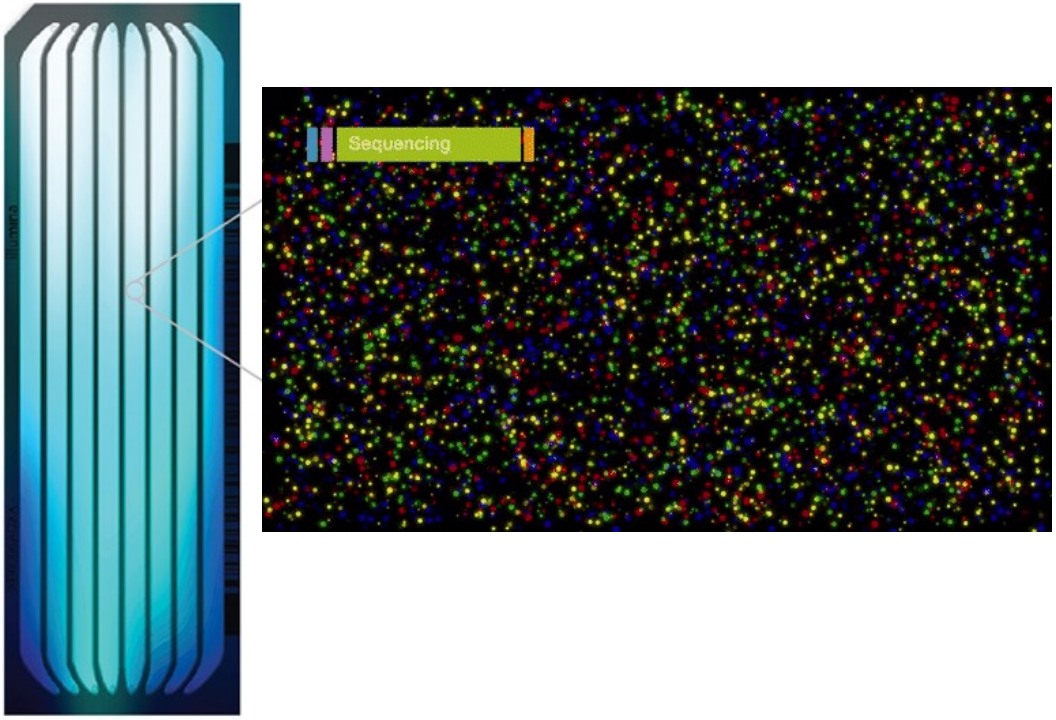
2.1 Gene expression profiling at the bulk level

Next generation sequencing



2.1 Gene expression profiling at the bulk level

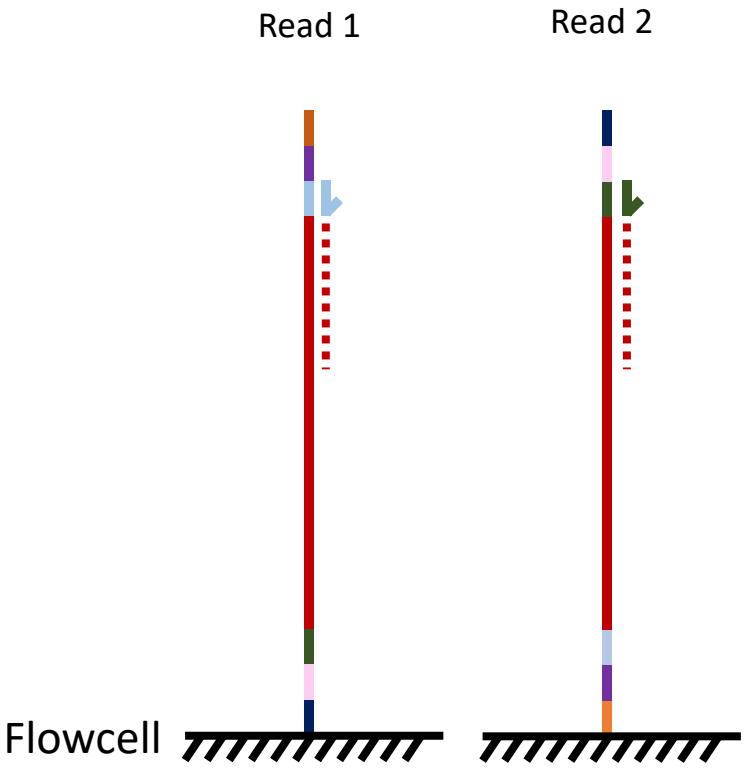
Next generation sequencing



Flowcell

2.1 Gene expression profiling at the bulk level

Next generation sequencing



Why paired-end sequencing?

Paired-end sequencing enables the sequencing of both ends of the fragment. The distance between each paired read is known and this information can be used by alignment algorithms to more accuratly map reads, especially across repetitive regions of the genome

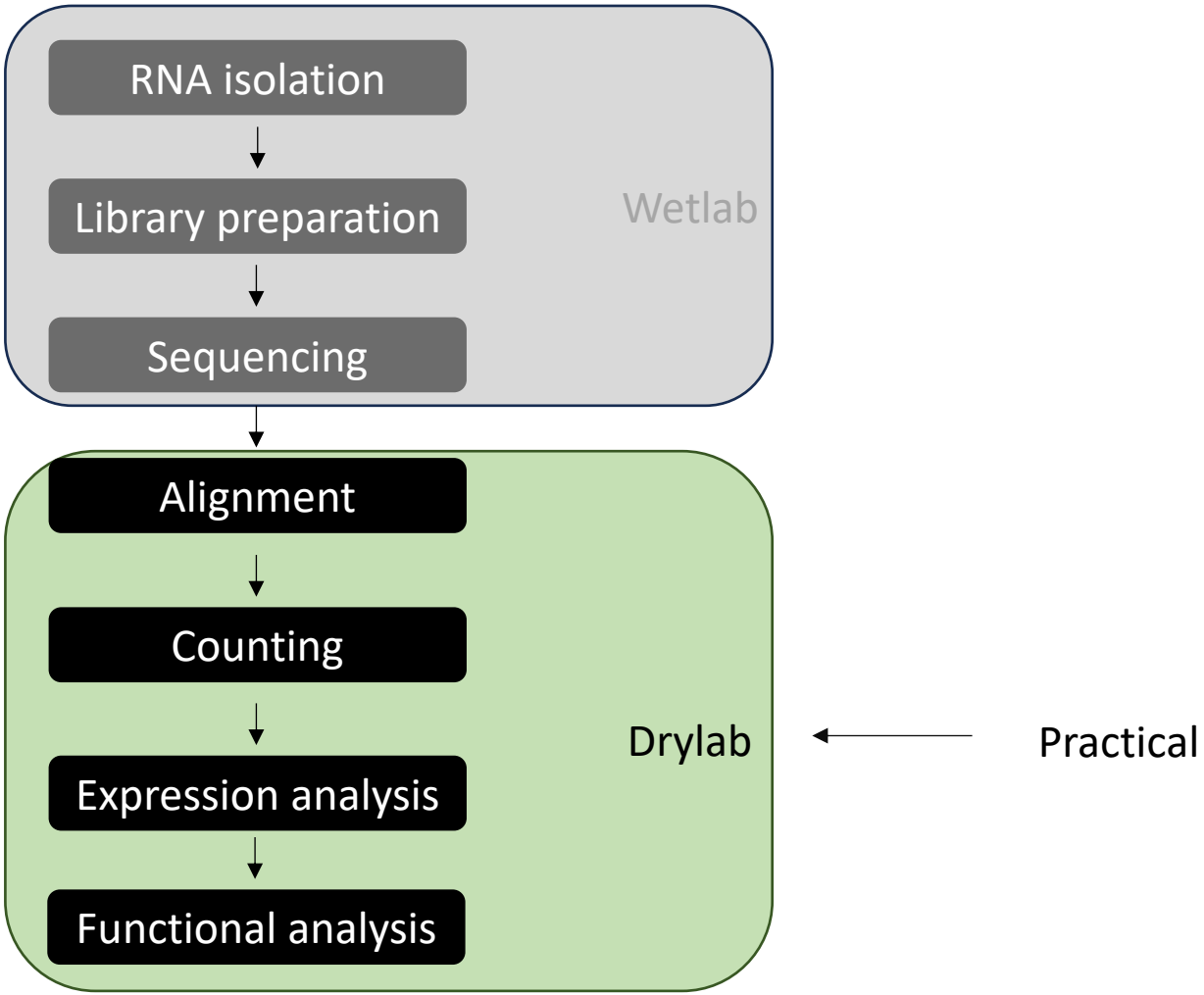
2.1 Gene expression profiling at the bulk level

gene expression omnibus (GEO) & sequencing read archive (SRA)

Functional genomics data (like transcriptomics) can be stored in publicly available repositories along with metadata. GEO is on of the biggest repositories.

In the case of GEO, raw sequencing reads are stored in the SRA database.

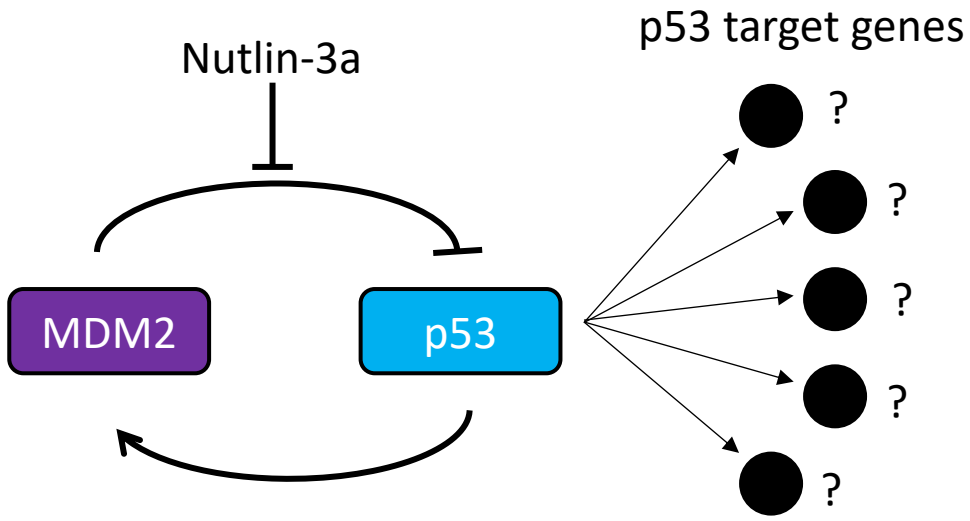
2.1 Gene expression profiling at the bulk level



Ref: Griffith *et al.* 2015

2.1 Gene expression profiling at the bulk level

Case study: Differential gene expression between cells with and without activation of the tumor suppressor p53



MCF-7 (breast cancer cell line) cells with and without Nutlin-3a stimilation.

Research question: Which genes are downstream of p53 (i.e. are upregulated by p53)?

2.1 Gene expression profiling at the bulk level

Case study: Differential gene expression between cells with and without activation of the tumor suppressor p53

See notebook: bulk_RNA_seq_part_1.ipynb

2.1 Gene expression profiling at the bulk level

Differential gene expression analysis using DESeq2

Why specialised algorithms are needed?

- non-normality and a dependence of the variance on the mean
- A small number of replicates (in our case study only two per condition)

Both of these require special consideraiton for statistical testing.

Goal of this algorithm is to test for which genes the expression level is significantly different between two (or more) conditions.

2.1 Gene expression profiling at the bulk level

Case study: Differential gene expression between cells with and without activation of the tumor suppressor p53

See notebook: bulk_RNA_seq_part_2.ipynb

2.1 Gene expression profiling at the bulk level
Further reading

- Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. Malachi Griffith ,Jason R. Walker, Nicholas C. Spies, Benjamin J. Ainscough, Obi L. Griffith. 2015. PLOS comp. biol.
- Mapping RNA-seq Reads with STAR. Alexander Dobin and Thomas R. Gingeras. 2015. Curr. Protoc. Bioinformatics.
- Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Michael I love, Wolfgang Huber & Simon Anders. 2014. Genome Biol.