



Swiss Institute of
Bioinformatics

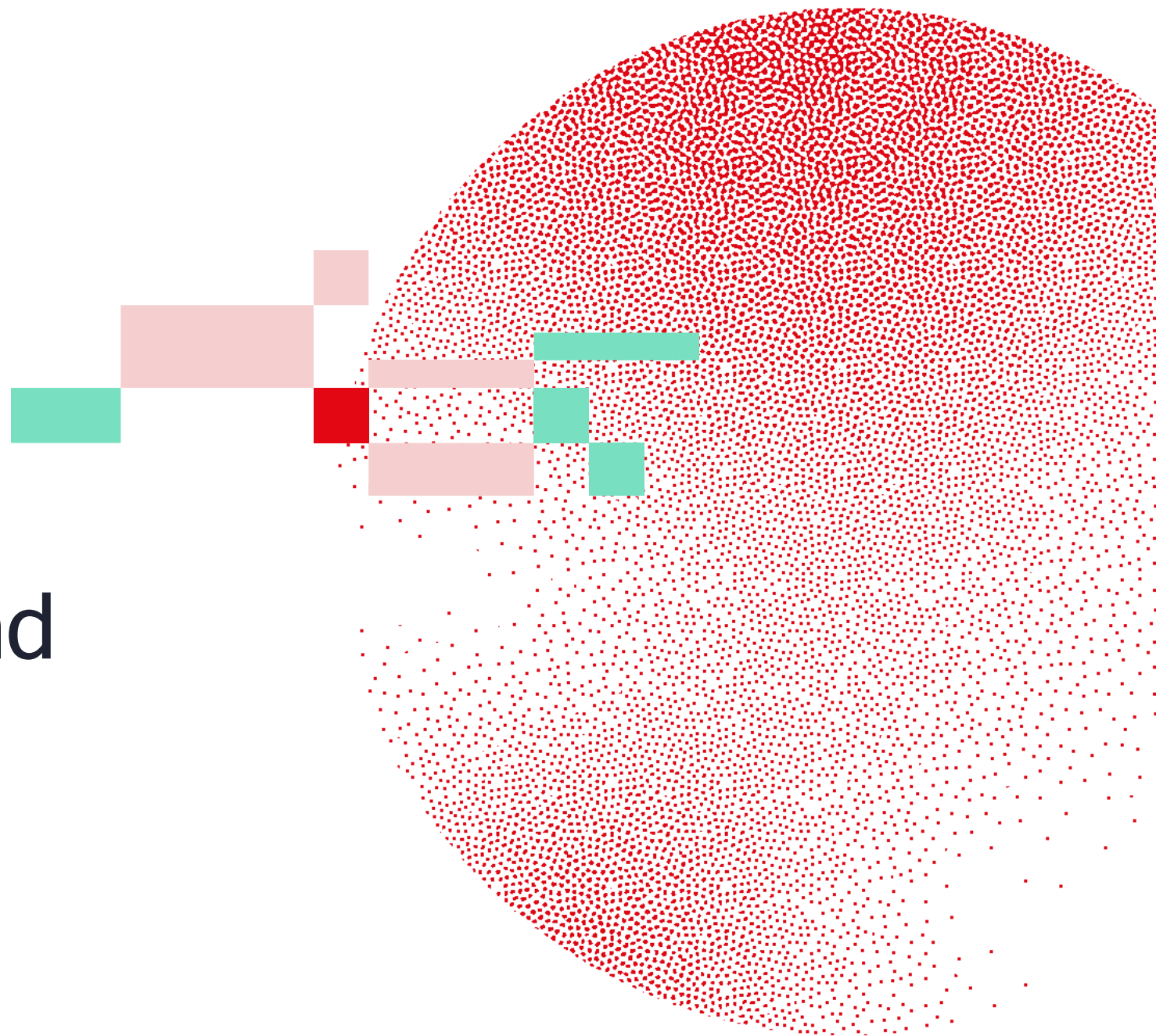
SINGLE-CELL TRANSCRIPTOMICS WITH R

Normalization, Transformation, and scaling

Deepak Tanwar

November 12-14, 2025

Adapted from previous year courses

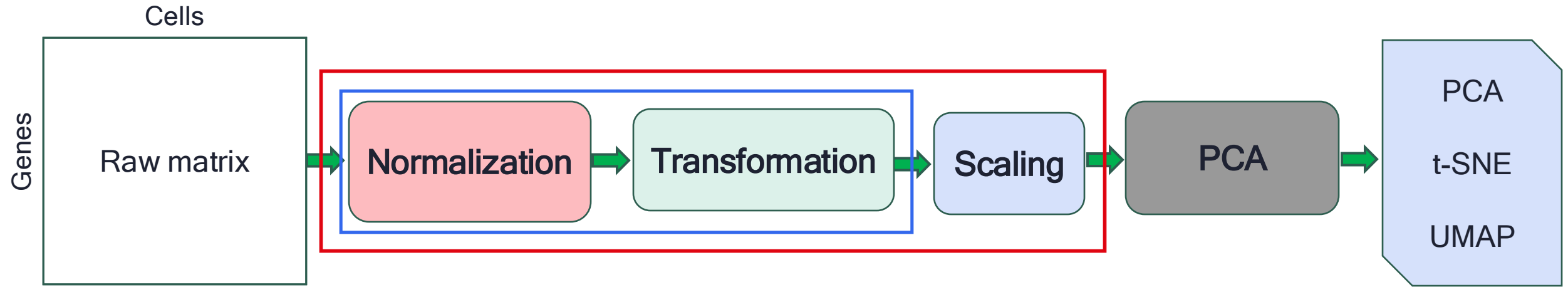


Learning objectives

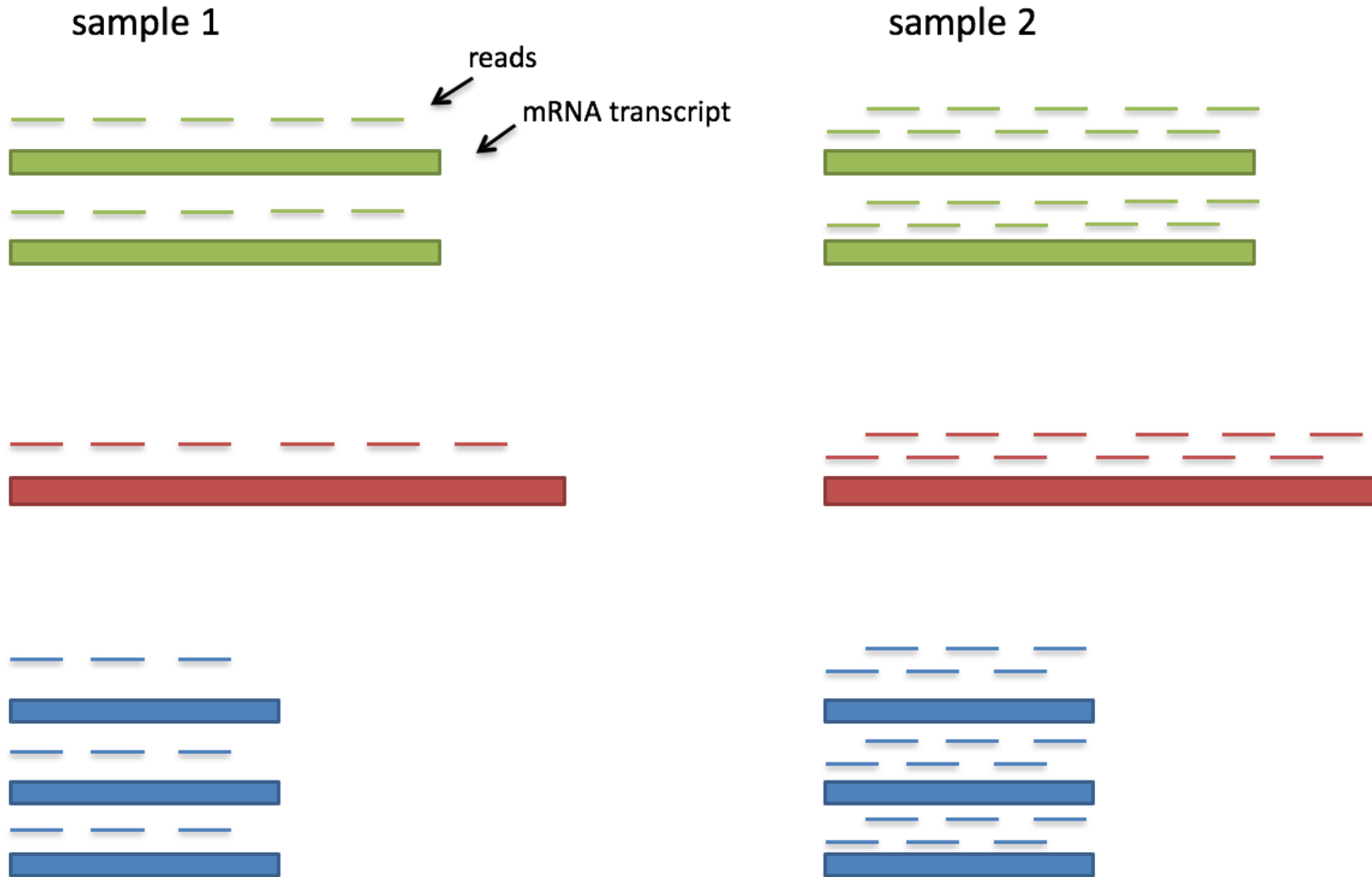
Distinguish Normalization, Transformation, and Scaling

Identify and apply Normalization techniques

Normalization and scaling



Differences from sequencing data



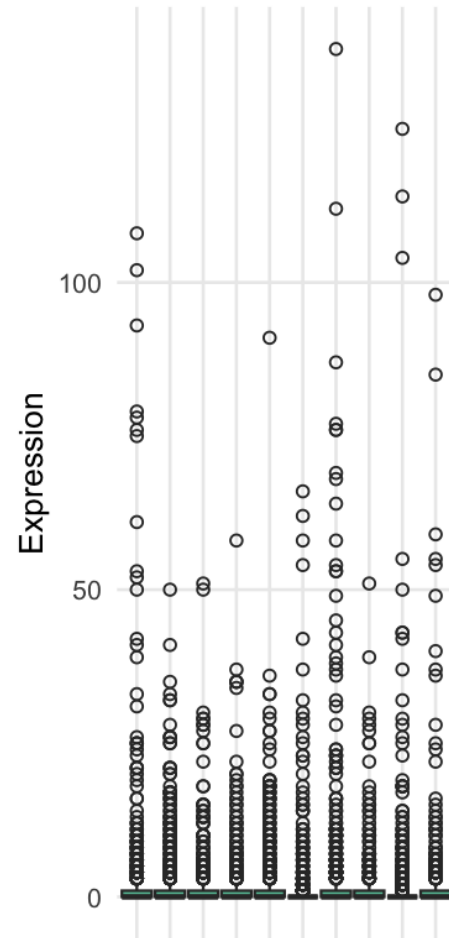
Problems with raw counts

1. Library size bias → deeper sequenced cells dominate
2. Zero-inflation → most genes = 0
3. High dynamic range

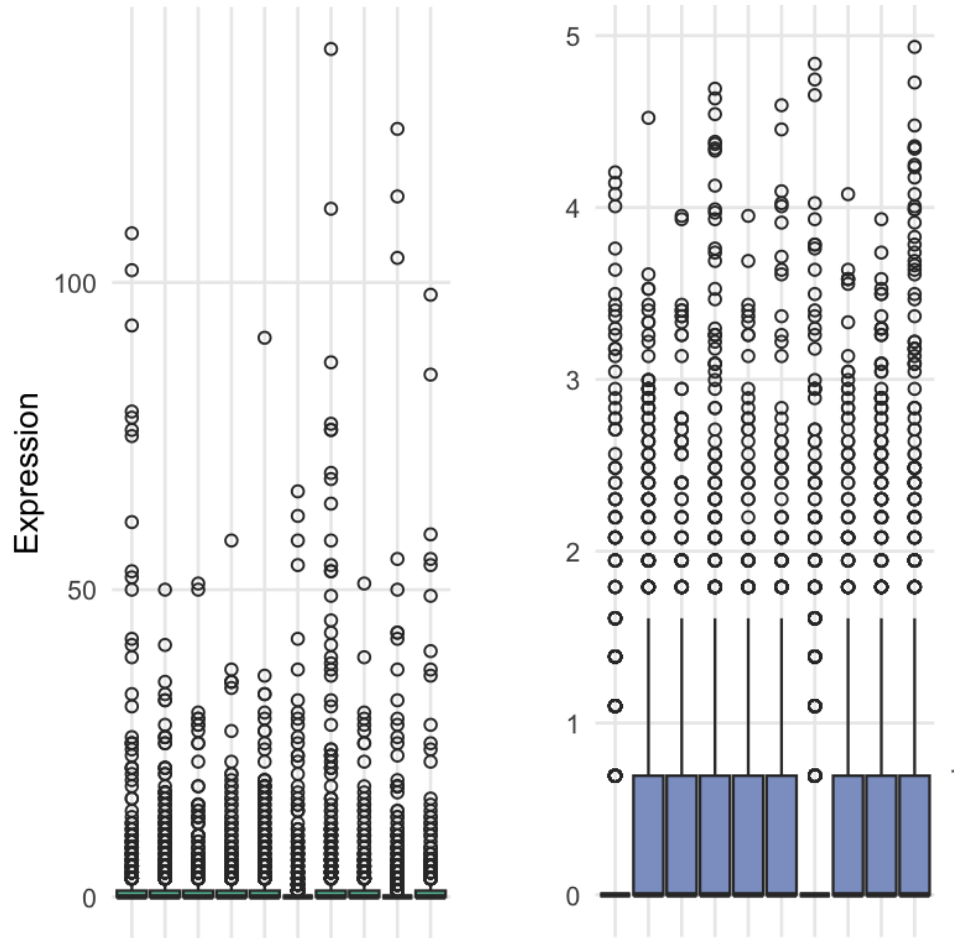
Goal of normalization and transformation

1. Remove sequencing-depth (library-size) differences so that cells are comparable.
2. Transform values with log
3. Multiply each UMI count by 10,000 to make it in a readable/manageable scale

Transformation

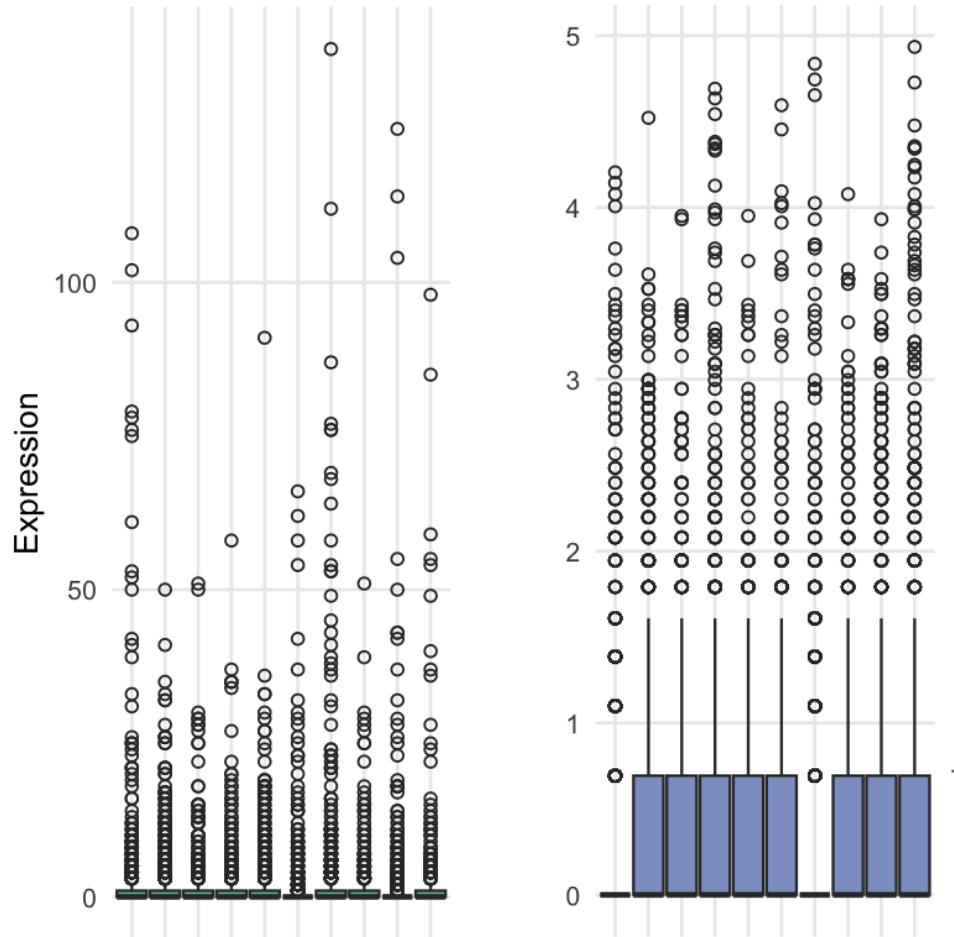


Transformation



$$\text{norm}_{g,c} = \log \left(\frac{\text{count}_{g,c}}{\sum_g \text{count}_{g,c}} \times 10,000 + 1 \right)$$

Transformation



`Seurat::NormalizeData()`

$$\text{norm}_{g,c} = \log \left(\frac{\text{count}_{g,c}}{\sum_g \text{count}_{g,c}} \times 10,000 + 1 \right)$$

What Is Scaling?

The Problem: Genes Have Different Scales

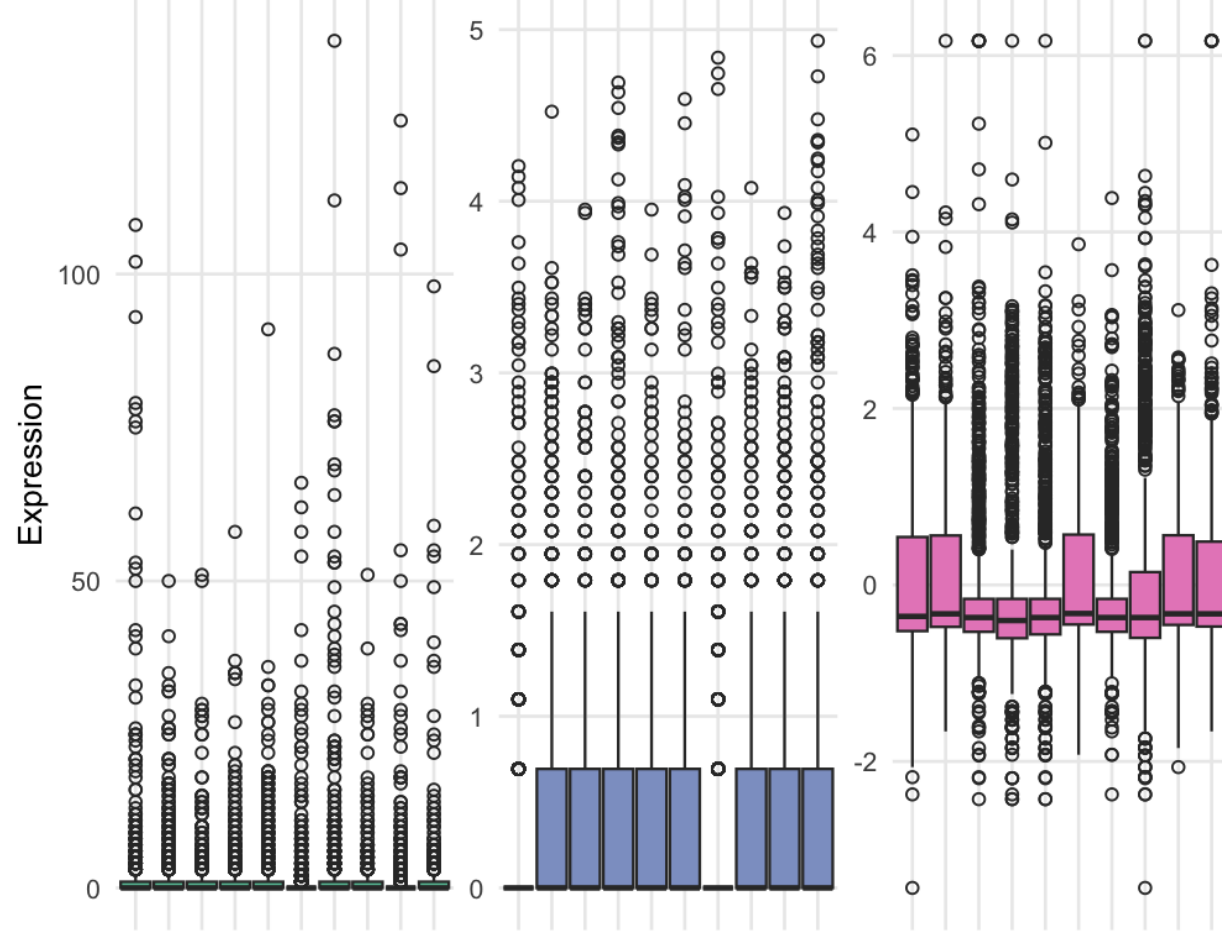
Gene	Mean	SD	Dominates PCA?
LYZ	5	3	YES
CD79A	0.1	0.2	NO

High-expression genes dominate PCA – even if low variance!

Solution: Z-Score Scaling

(Every gene → **mean = 0, SD = 1**)

Scaling



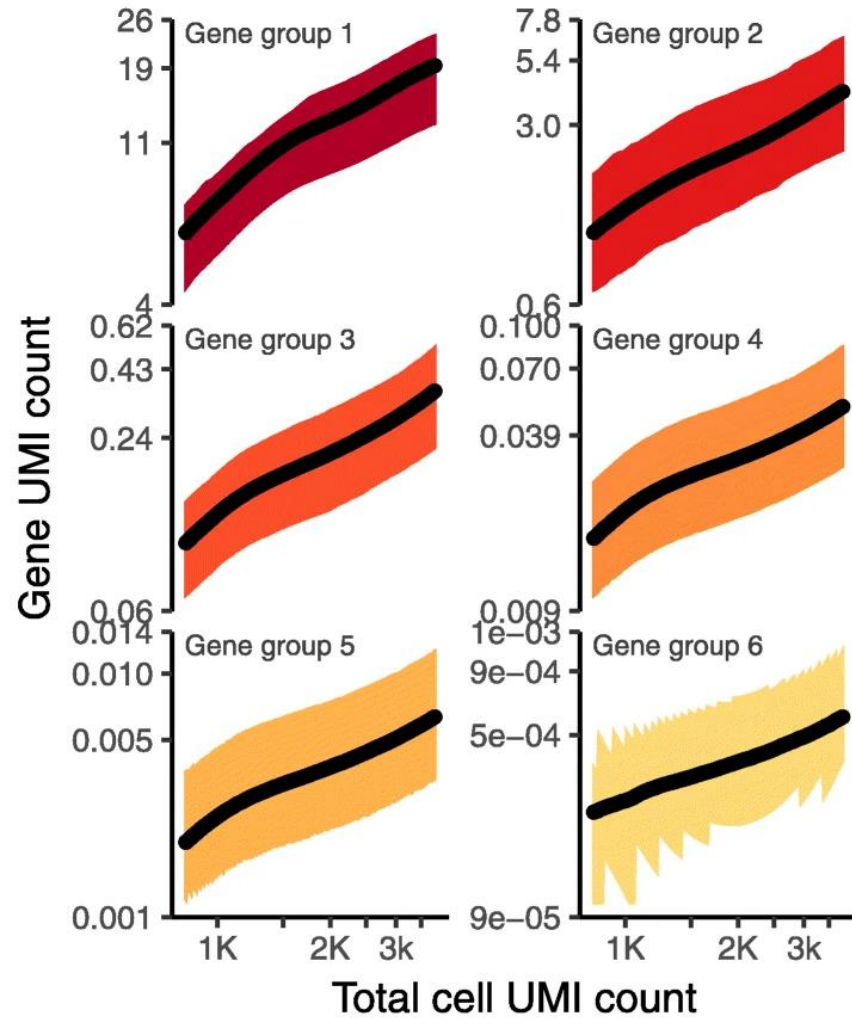
What other transformations can we use?

Does anyone know about SCTransform?

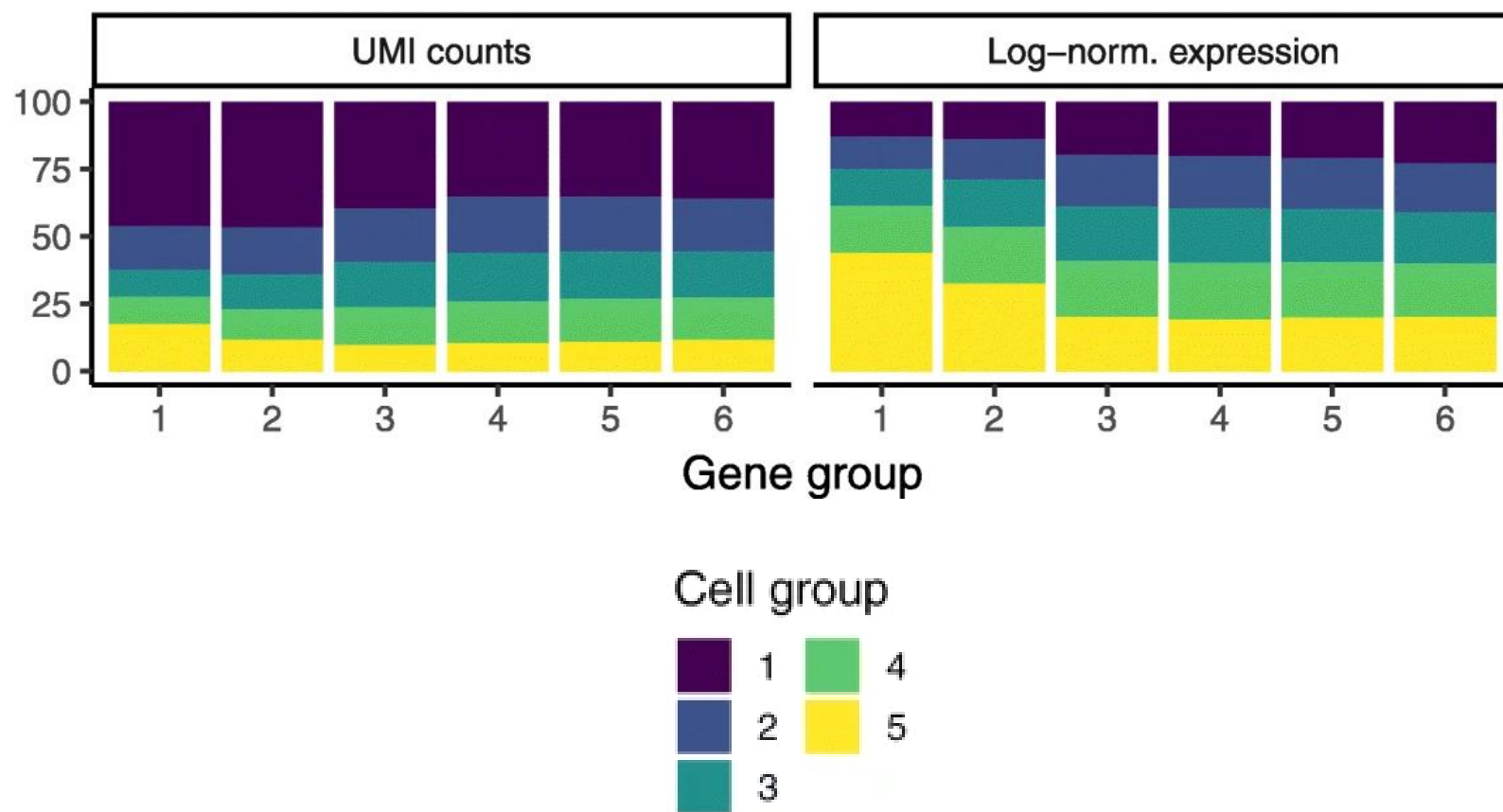
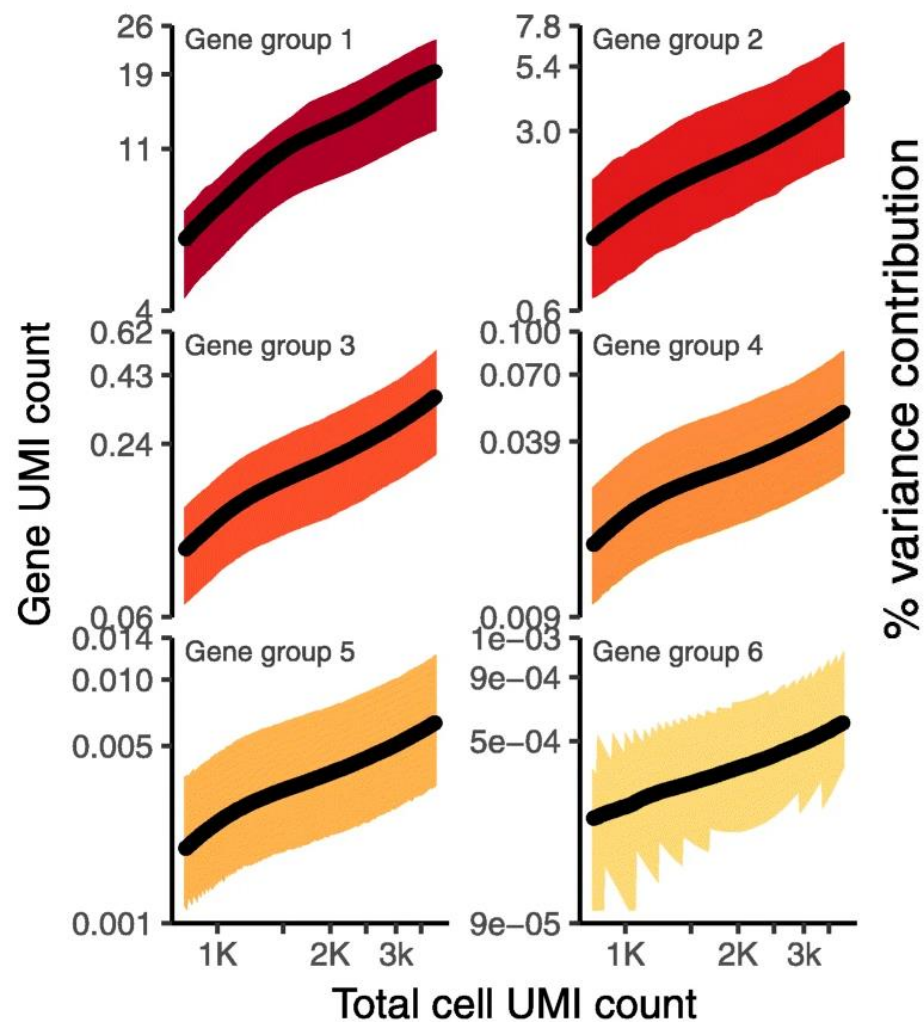
Transformation : Pearson residuals

Instead of transforming each measurements individually, Pearson residuals apply a weight to all measurements of a gene

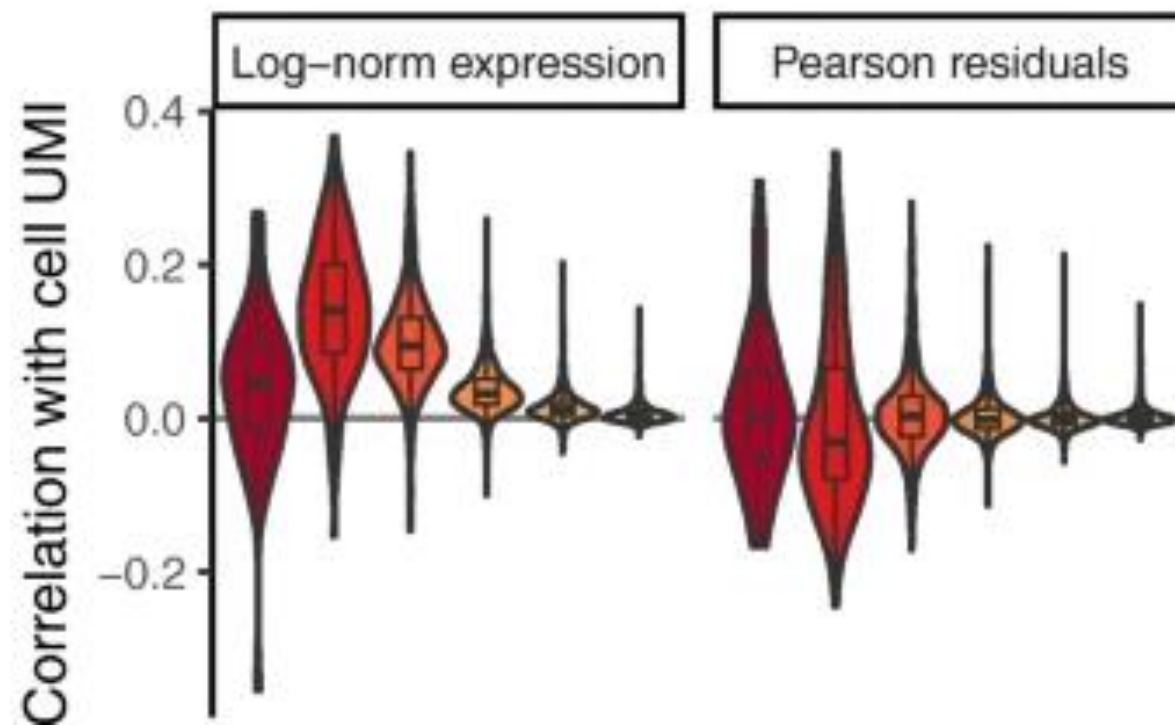
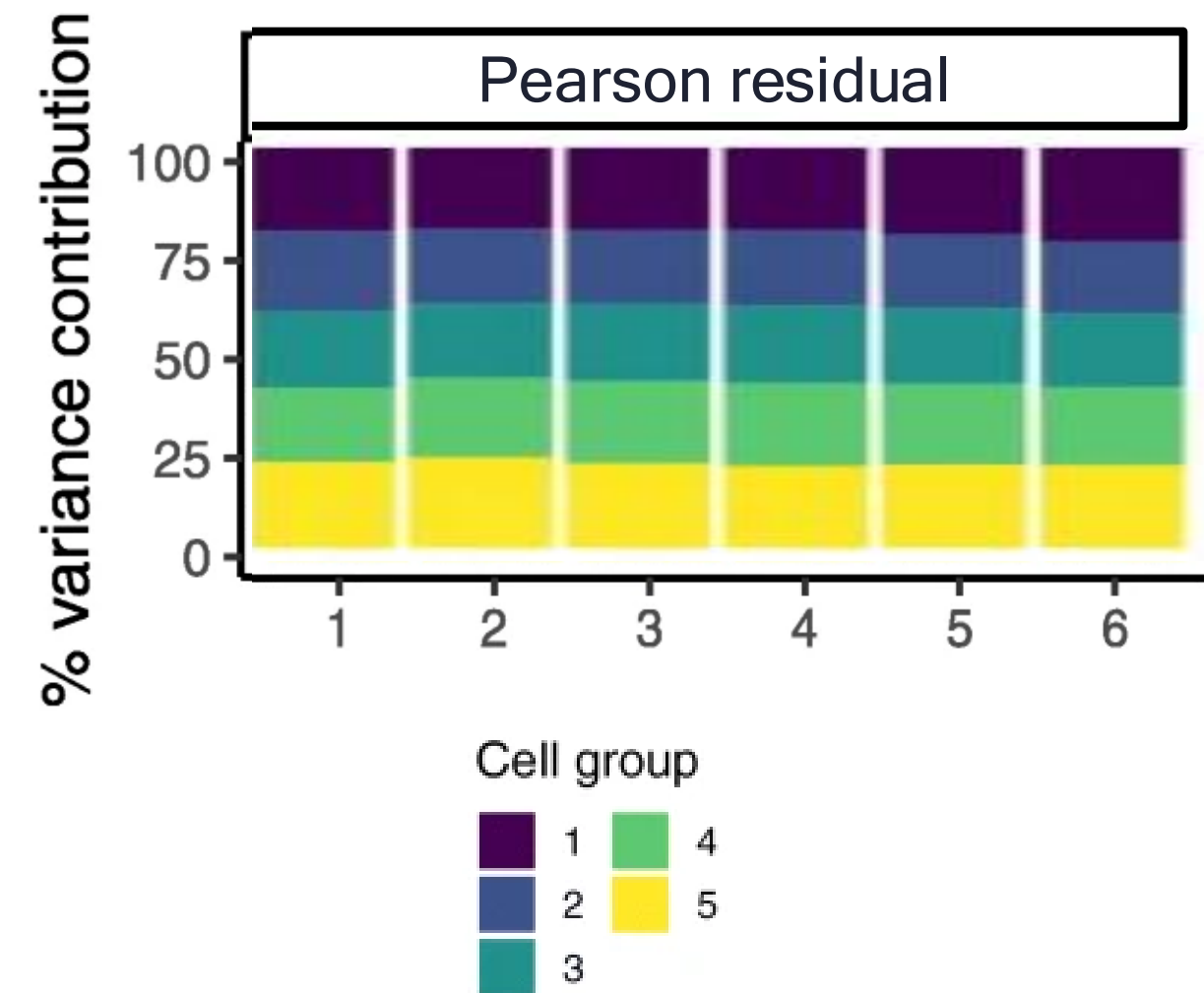
Gene counts are heavily influenced by sequencing depth



Gene counts are heavily influenced by sequencing depth



sctransform



We will not use SCTransform for the group work due to computation resources. It is advisable to evaluate it for your own single-cell RNA-seq data.

