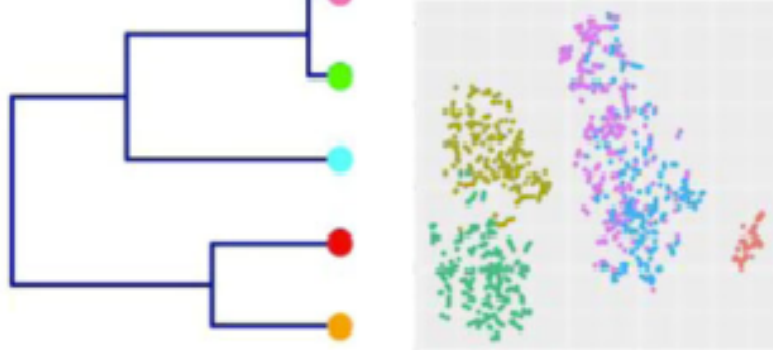
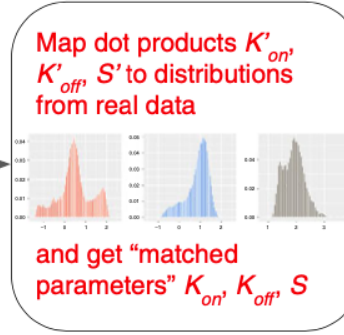
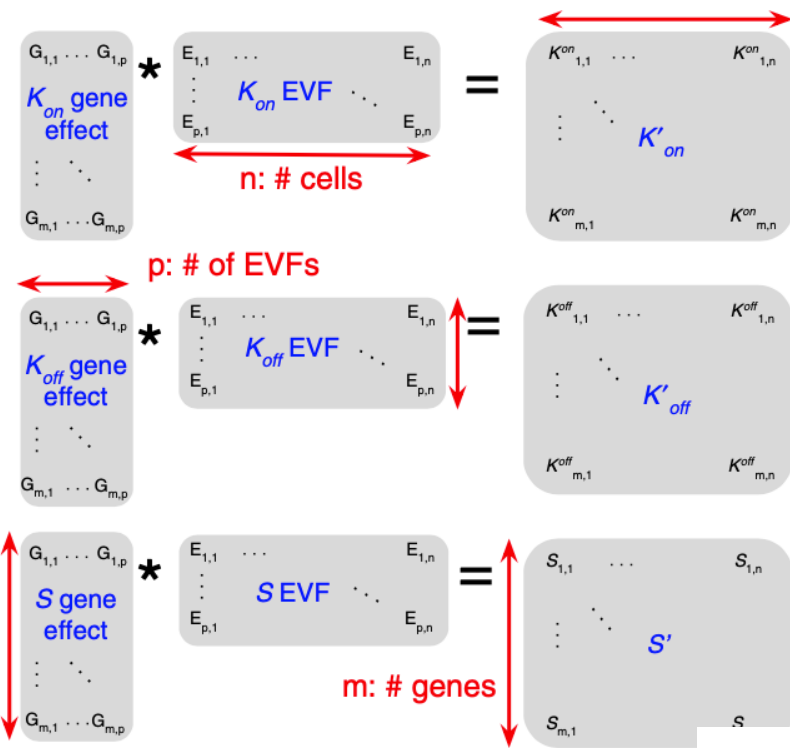
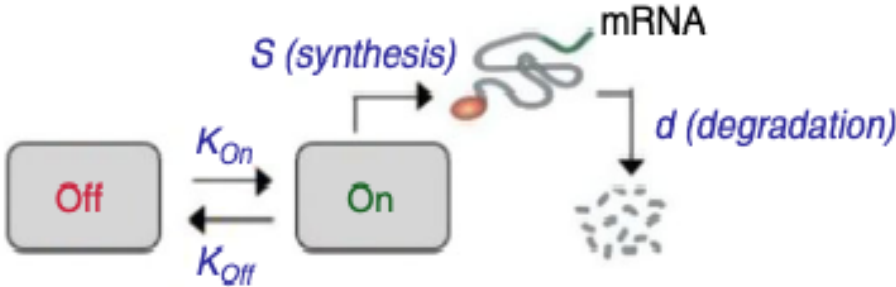


Generating True Counts based on two-state kinetic model

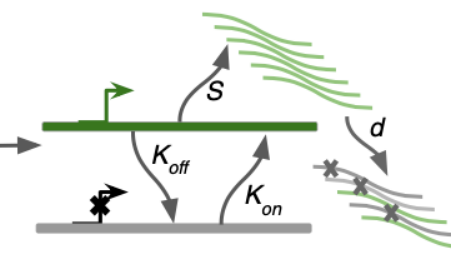
Cells: p-dim EVF vectors



Genes: p-dim gene specific vectors



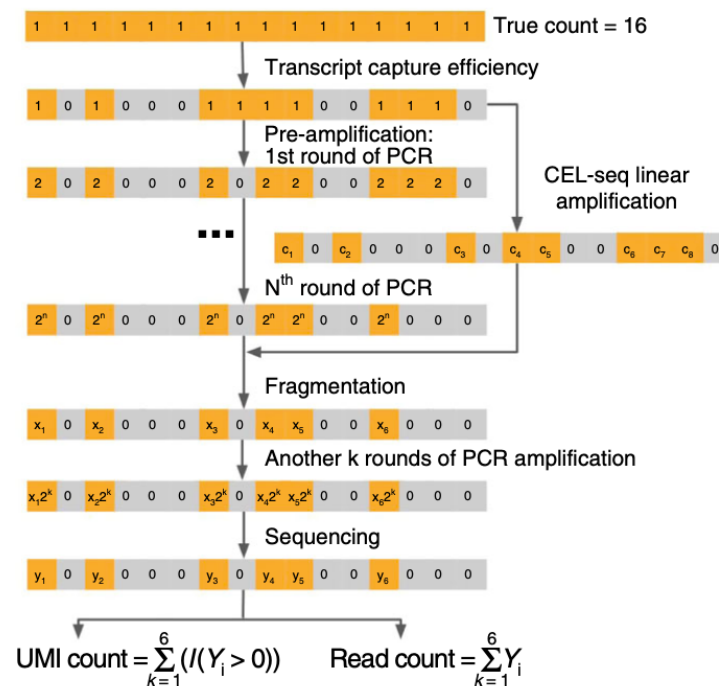
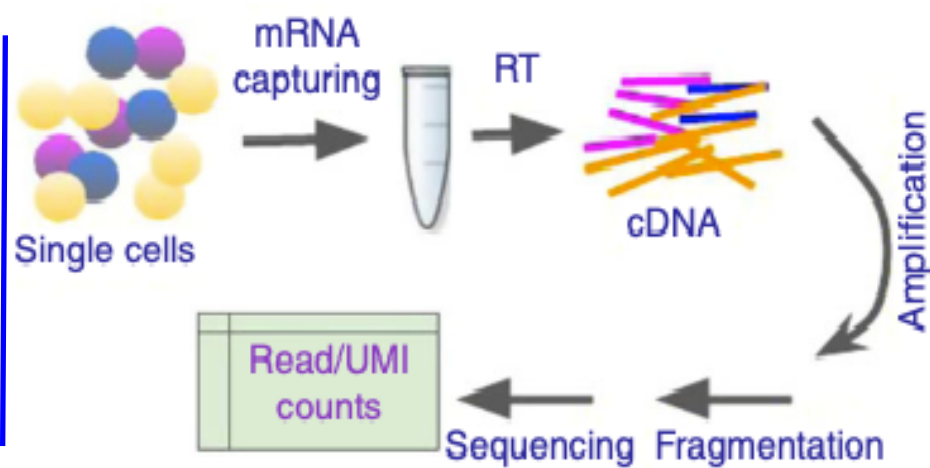
Sample transcript counts



Definition of differentially expressed (DE) genes.

- ▶ The number of diff-evfs that are associated with a non-zero gene effect value based on the kinetic parameters (should larger than zero)
- ▶ Theoretical log2 fold change analysis is based on the kinetic parameters (should larger than 0.6 ~ 1).

Simulating PCR and Sequencing platform



Batch Effect Added on Counts

- ▶ Each cell c is assigned to the batch b randomly as b_c
- ▶ Given the user-defined batch effect size (be , default as 1), SymSim will firstly samples $m_{gb} \sim \mathcal{N}(0, 1)$; then update it by sampling $\hat{m}_{gb} \sim \mathcal{U}(m_{gb} - be, m_{gb} + be)$.
- ▶ Then for cell c and gene g , let $m_{gc} \sim \mathcal{N}(m_{gb_c}, 0.01)$
- ▶ Lastly, the observed count $\hat{o}_{gc} = o_{gc} * 2^{m_{gc}}$ (i.e., added in the \log_2 level).