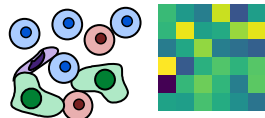


Expression count matrix for heterogeneous population



Prior knowledge of marker genes

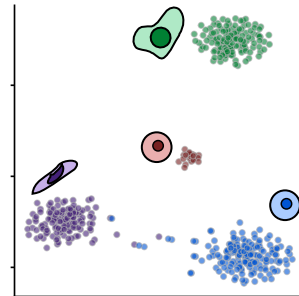
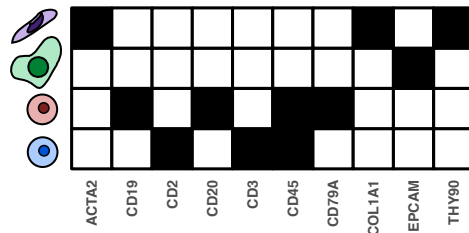
- T cell: CD2, CD3, CD45
- B cell: CD19, CD20, CD79A, CD45
- Tumor cell (epithelial): EPCAM
- Fibroblast: COL1A1, ACTA2, THY90



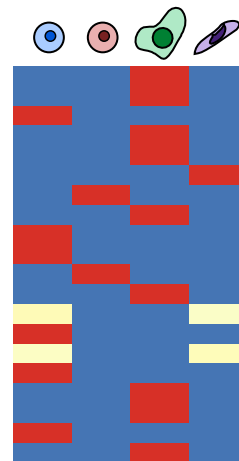
CellAssign: automated probabilistic assignment of cells to known cell types

$$\overbrace{\log \mu_{ngc}}^{\text{Mean expression}} = \underbrace{\log s_n}_{\text{Size factor}} + \underbrace{\delta_{gc} \rho_{gc}}_{\text{Cell type}} + \underbrace{\beta_{g0}}_{\text{Base expression}} + \underbrace{\sum_{p=1}^P \beta_{gp} x_{pn}}_{\text{Other covariates (batch)}}$$


Binarization



Cell type assignments



Probabilities