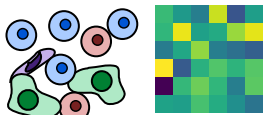


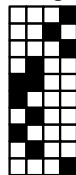
## Expression count matrix for heterogeneous population



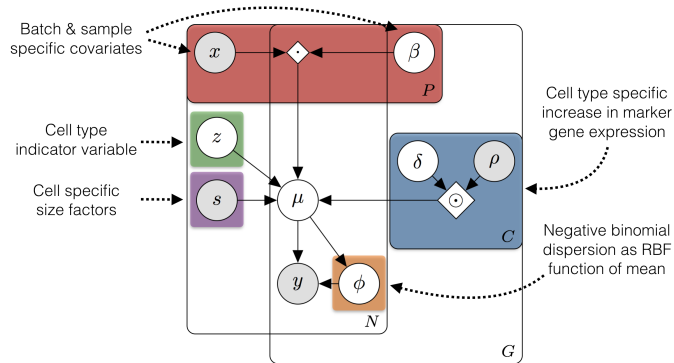
## Prior knowledge of marker genes

- **T cell:** CD2, CD3, CD45
- **B cell:** CD19, CD20, CD79A, CD45
- **Tumor cell:** EPCAM
- **Fibroblast:** COL1A1, ACTA2, THY1

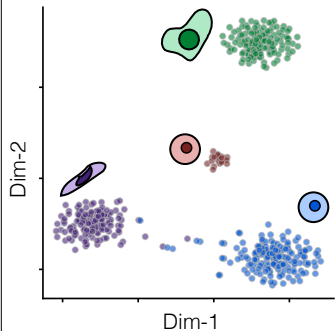
Genes



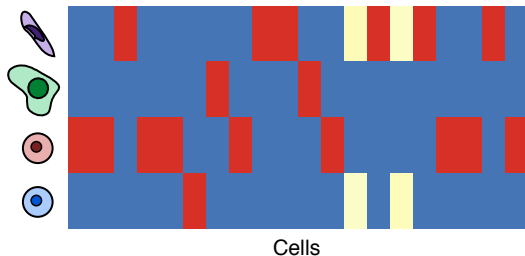
**CellAssign:** automated probabilistic assignment of cells to known cell types



## Cell type assignments



## Probabilities



Variable	Distribution	Description
$y_{ng}$	Negative Binomial	Single-cell counts
$s_n$	None	Cell size factor
$z_n$	Categorical	Cell type indicator
$\mu_{ngc}$	Deterministic $f^n$	Modelled average expression
$\phi_{ngc}$	Deterministic $f^n$	Negative binomial dispersion
$\delta_{gc}$	Log-Normal	Marker over-expression
$\rho_{gc}$	None	Marker-cell type matrix
$x_{np}$	None	Covariates (batch or sample)
$\beta_{pg}$	Gaussian	Covariate coefficients
$a, b$	None	Dispersion basis coefficients