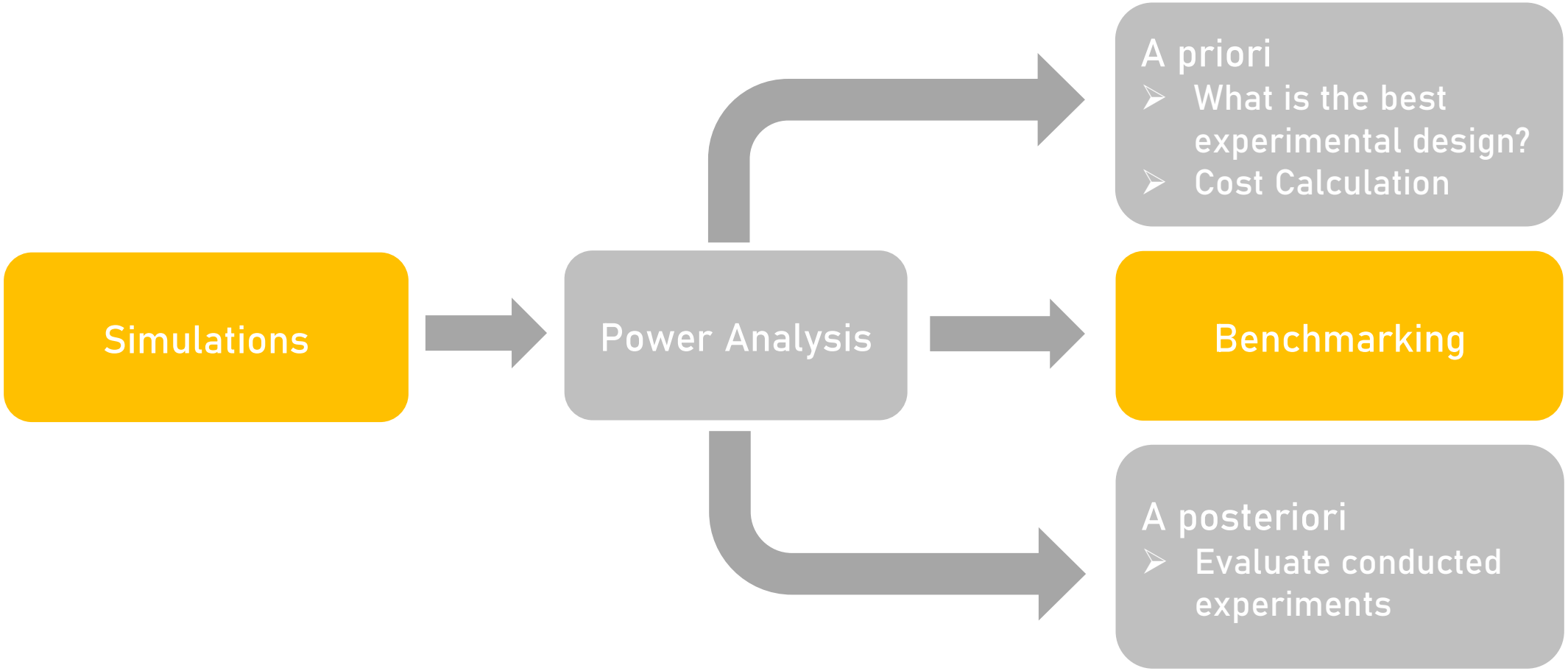
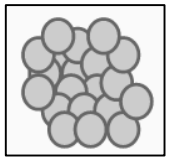
Power simulations for cell type identification from single cell RNA-seq data



Beate Vieth, Philipp Janssen, Wolfgang Enard, Ines Hellmann
Department of Biology II, Ludwig-Maximilians University, Munich, Germany

Observed Expression = Unwanted Technical Factors + Unwanted Biological Factors + Biological Factors

X_{ng} = W_{nr} η_{rg} + Z_{nk} γ_{kg} + Y_{ni} β_{ig}



X_{gn}	Cell 1	Cell 2	Cell 3	Cell n
Gene 1	0	80	0	10
Gene 2	10	0	5	0
Gene 3	25	65	39	0
Gene g	0	0	54	24

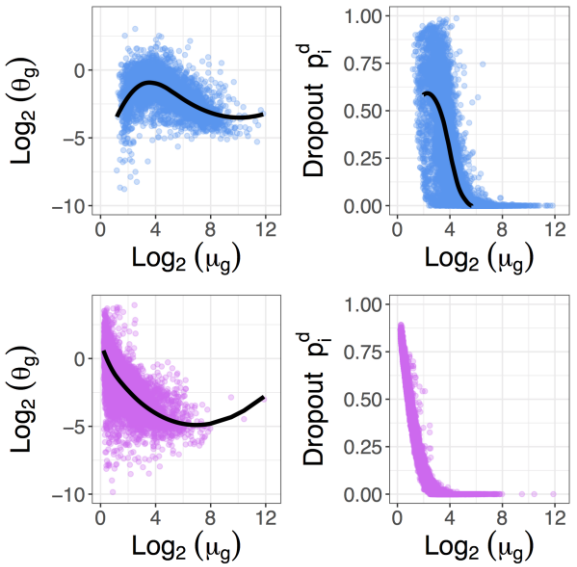
non-UMI Protocols

$$X_{ng} \sim ZINB(\mu_{g+}, \theta_{g+}, \pi_g)$$

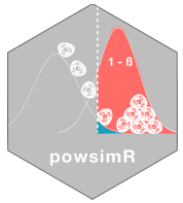
UMI Protocols

$$X_{ng} \sim NB(\mu_g, \theta_g)$$

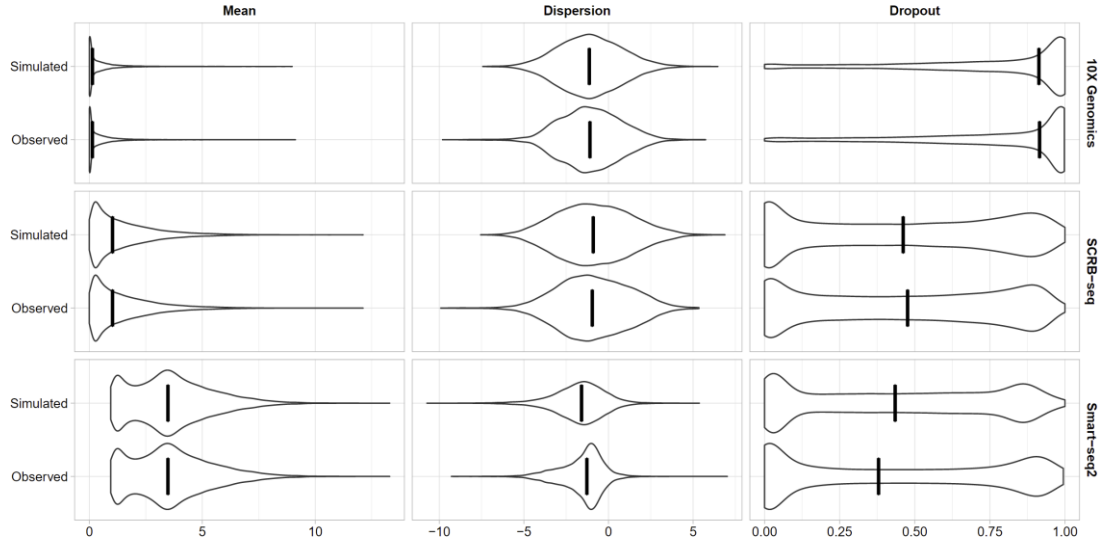
Fitting mean-variance and mean-dropout relationship:



Simulate expression



Vieth et al. 2017 Bioinformatics

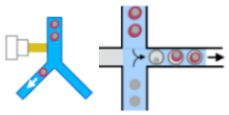


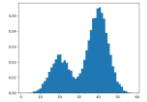
Vieth et al. 2019 Nat. Comm.

Observed Expression = Unwanted Technical Factors + Unwanted Biological Factors + Biological Factors

X_{ng} = W_{nr} η_{rg} + Z_{nk} γ_{kg} + Y_{ni} β_{ig}

Lib Prep | **Doublets** | **Background RNA** | **Seq Depth**



DE-Testing | scDbl Finder | cellbender | scran

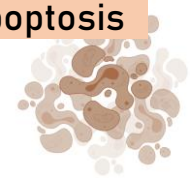
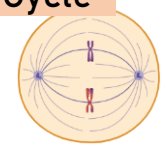
$\eta_{rg} \sim lN(\mu, \sigma^2)$
$\mu = E[\eta_r]$
$\sigma^2 = Var[\eta_r]$

$d \in [d_1,...,d_r]$
$\rho_n \sim Beta(\rho_\alpha, \rho_\beta)$
Barcode: $\bar{\chi}_g$
Ambient: χ_g^a

$z_r \sim MN(\pi)$
$sf_r \sim lN(\mu_z, \sigma_z^2)$

Feature | Estimate | Simulate

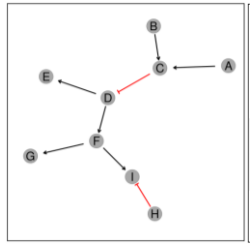
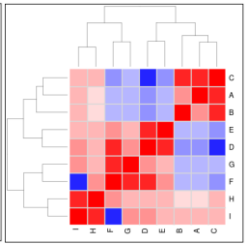
Apoptosis | **Cell Cycle** | **Hypoxia**


slalom, cNMF

$q \in [q_1,...,q_k]$
$\gamma_{kg} \sim MVN(\mu, \Sigma)$
$\mu = \Sigma q_k E[\gamma_k]$
$\Sigma = \sigma^2 [\gamma]_{k*k}$

Pathway | **Covariance**

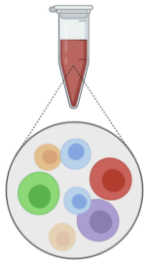



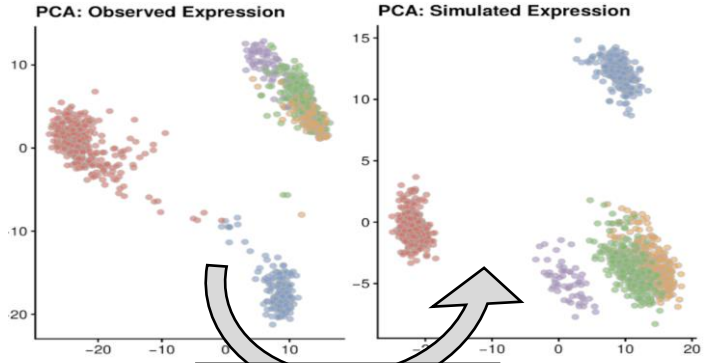
graphsim

Observed Expression = Unwanted Technical Factors + Unwanted Biological Factors + Biological Factors

X_{ng} W_{nr} η_{rg} Z_{nk} γ_{kg} Y_{ni} β_{ig}

Blood





$\beta_{ig} \sim MVN(\mu, \Sigma)$

Ornstein Uhlenbeck Process:

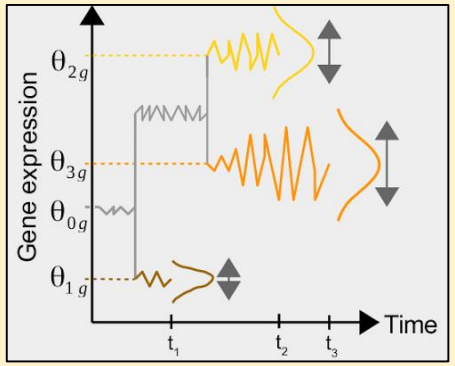
$$dX_{ig}(t) = \alpha_g[\theta_{ig} - X_{ig}(t)]dt + \sigma_g dB(t)$$

$$\beta_{ig} \sim F(f_{ig}, \theta_{0g}, \alpha_g, \sigma_g, \phi)$$

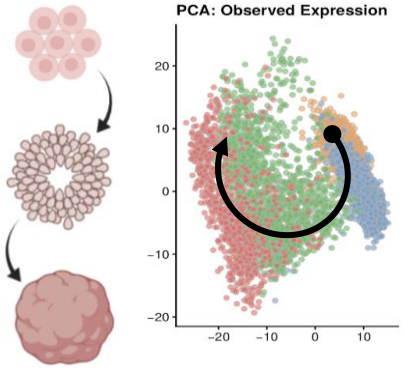
$\phi\ Tree$ $\quad \alpha_g \sim \Gamma(\alpha, \beta)$ $\quad \sigma_g \sim N(\mu, \sigma^2)$

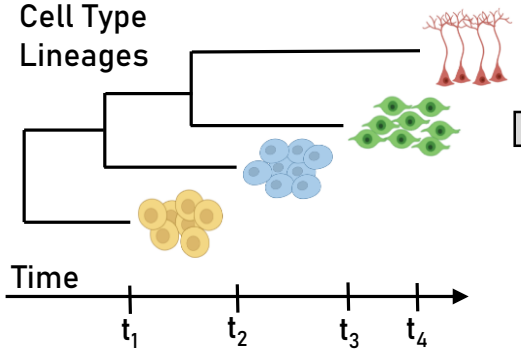
$f_{ig} \sim \Gamma(\alpha, \beta)$

$\theta_{0g} \sim N(Par(\alpha, \beta), \frac{\sigma_g^2}{2\alpha_g})$



Differentiation





Evaluate classification results