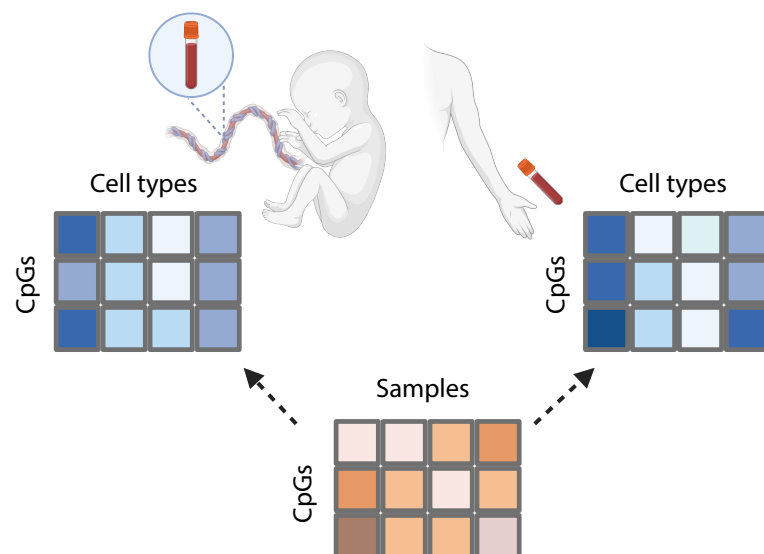


1

Pick an appropriate
reference DNAm
data set

getRef()

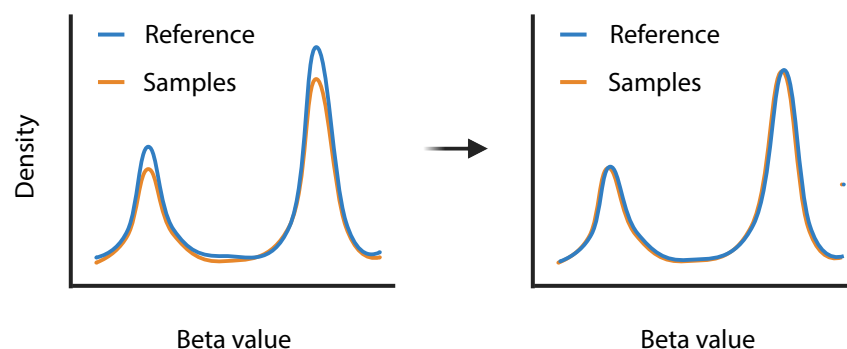


- **IDOL reference (2018)**
- IDOL extended reference (2022)
- Cord reference (2019)
- UniBlood (adult + cord) reference - 7 cell types, 13 cell types, and 19 cell types

2

Normalize
reference and
sample data sets

combData()

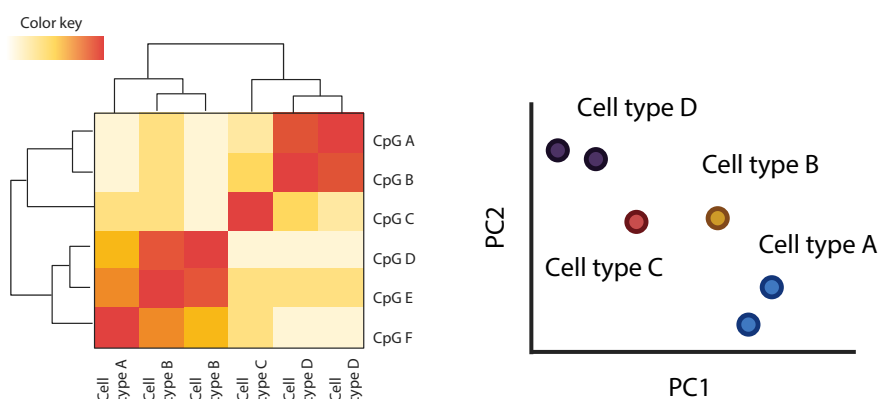


- **quantile normalize (RGset)**
- noob normalize (RGset)
- functional normalize (RGset)
- quantile normalize (beta matrix)
- no normalization

3

Select DNAm
features that best
distinguish cell types

pickProbes()

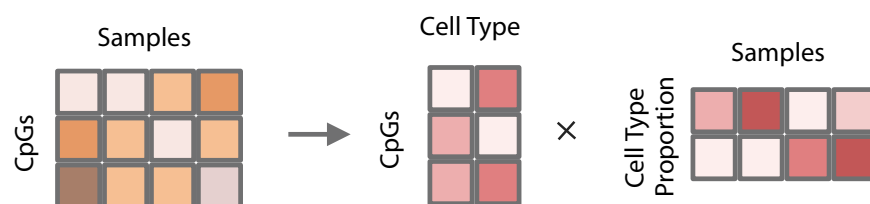


- **t-test (top 100 probes)**
- preselected probes: IDOL
- lasso
- elastic net (EL)
- random forest (RF)
- boosted logistic regression (BLR)
- decision tree (CART)
- gradient boosting machine (GBM)

4

Estimate cell type
proportion with
regression

predictCT()



Estimate prediction
accuracy with CETYGO

- **constraint projection (CP)**
- robust partial correlation (RPC)
- support vector regression (SVR)

Calculate absolute error
(predicted - actual)

Compare predicted
value to clinical
reference values

Compare variance
explained (R²) by
prediction outcomes