

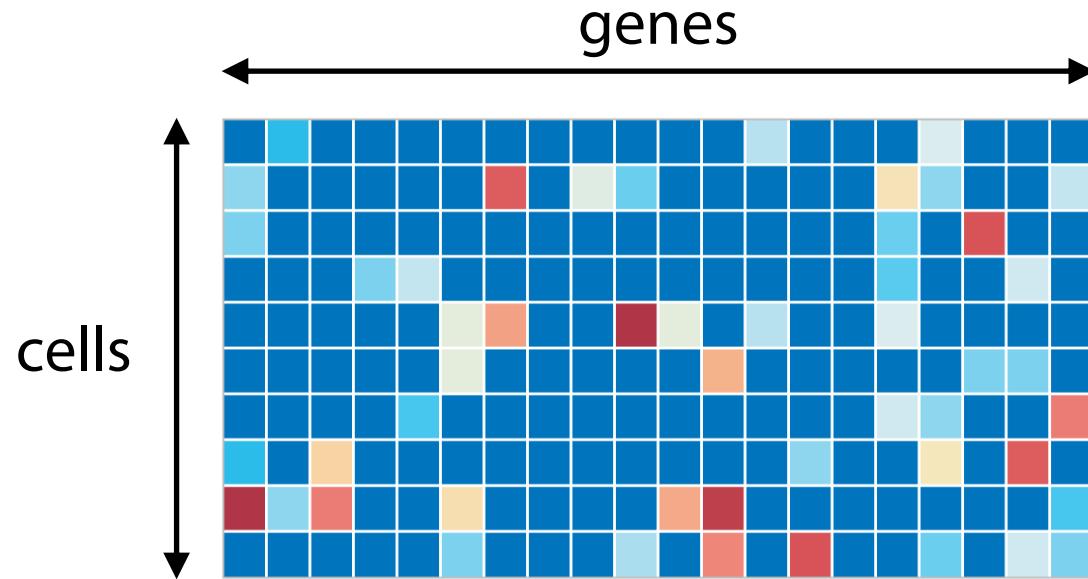
Quantifying the effect of experimental perturbations on scRNA-seq data

bioRxiv bit.ly/quantsinglecell

 github.com/KrishnaswamyLab/MELD

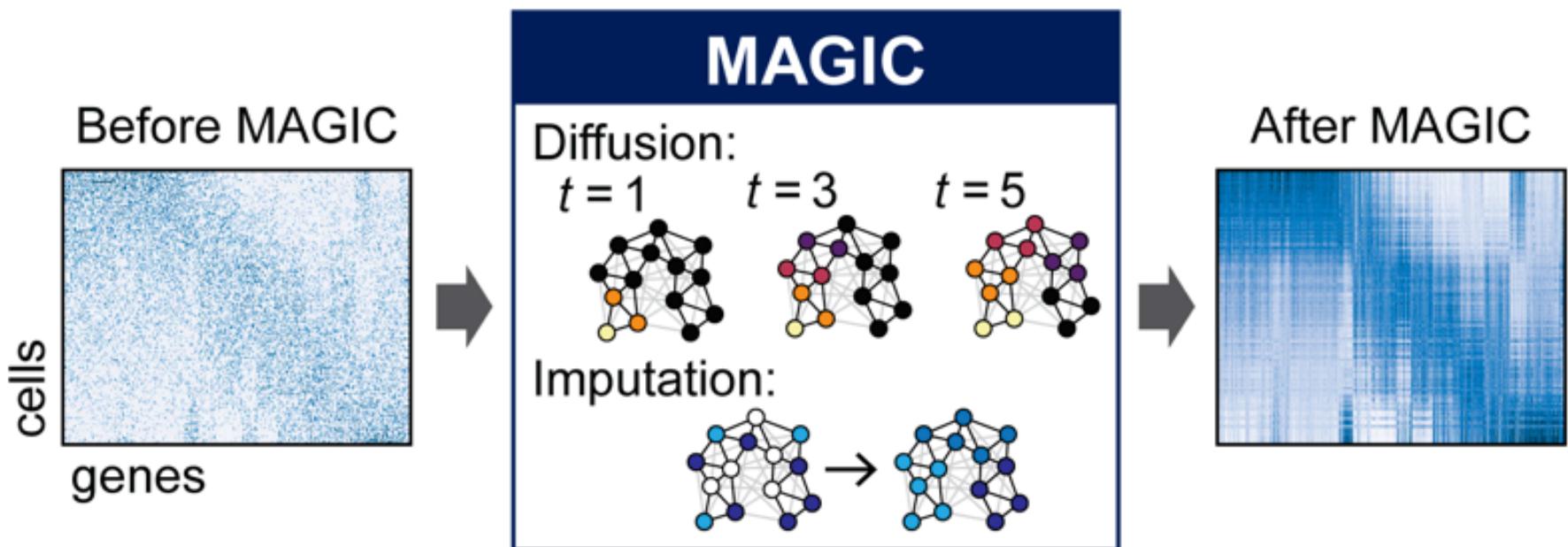
 @dbburkhardt

Daniel Burkhardt
Krishnaswamy Lab, Department of Genetics, Yale University

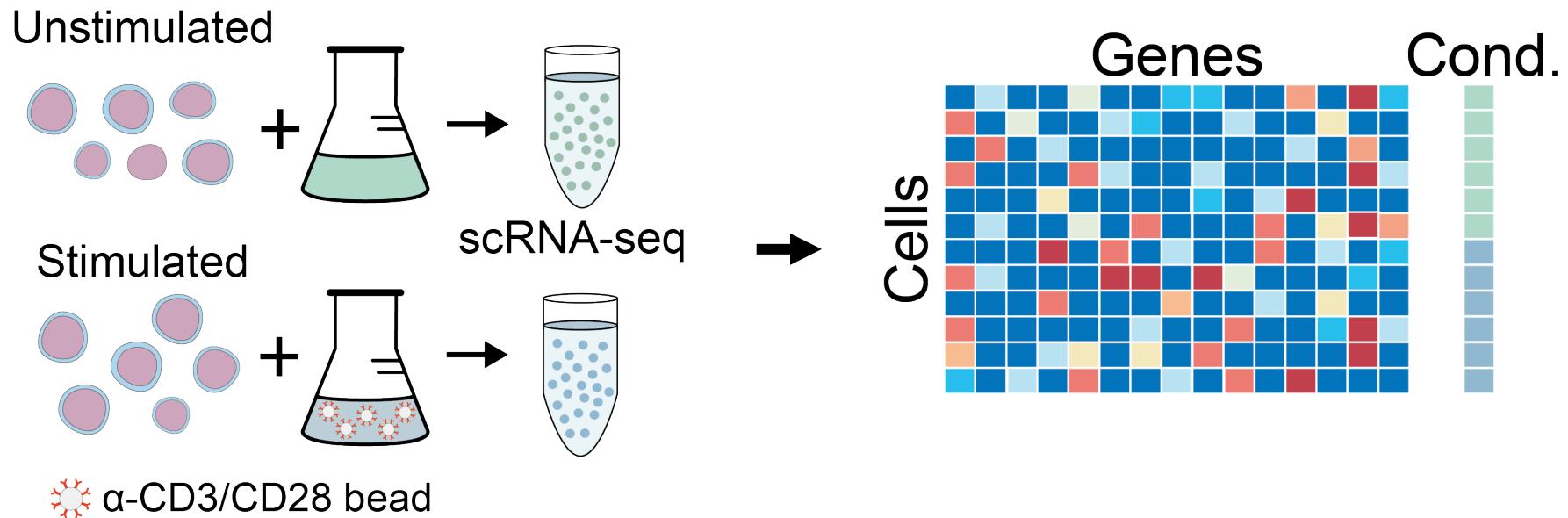


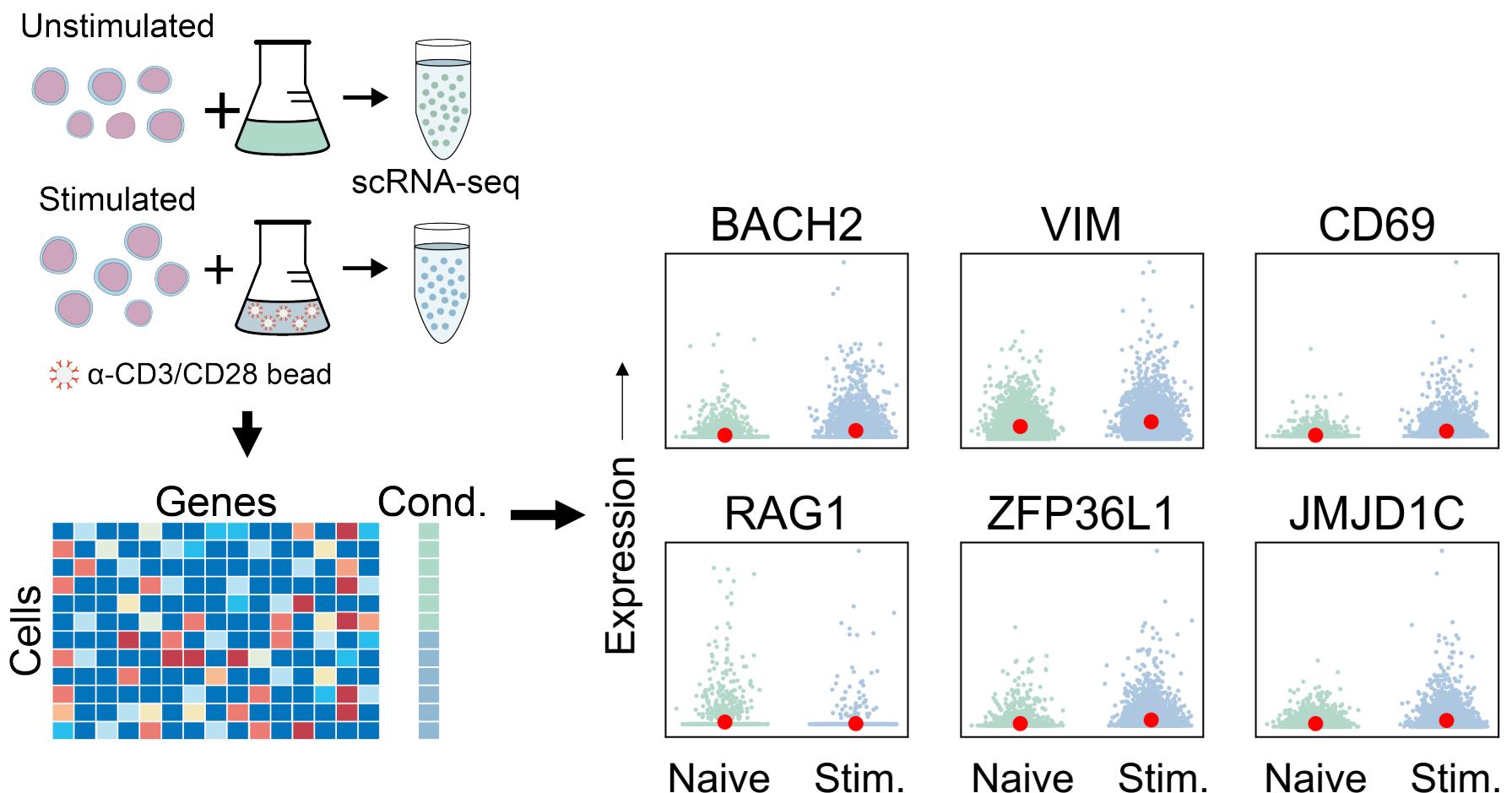
Biomedical data is noisy and sparse. How can we understand the relationships between observations (here, cells) and features (here, genes)?

Diffusing Values to Impute

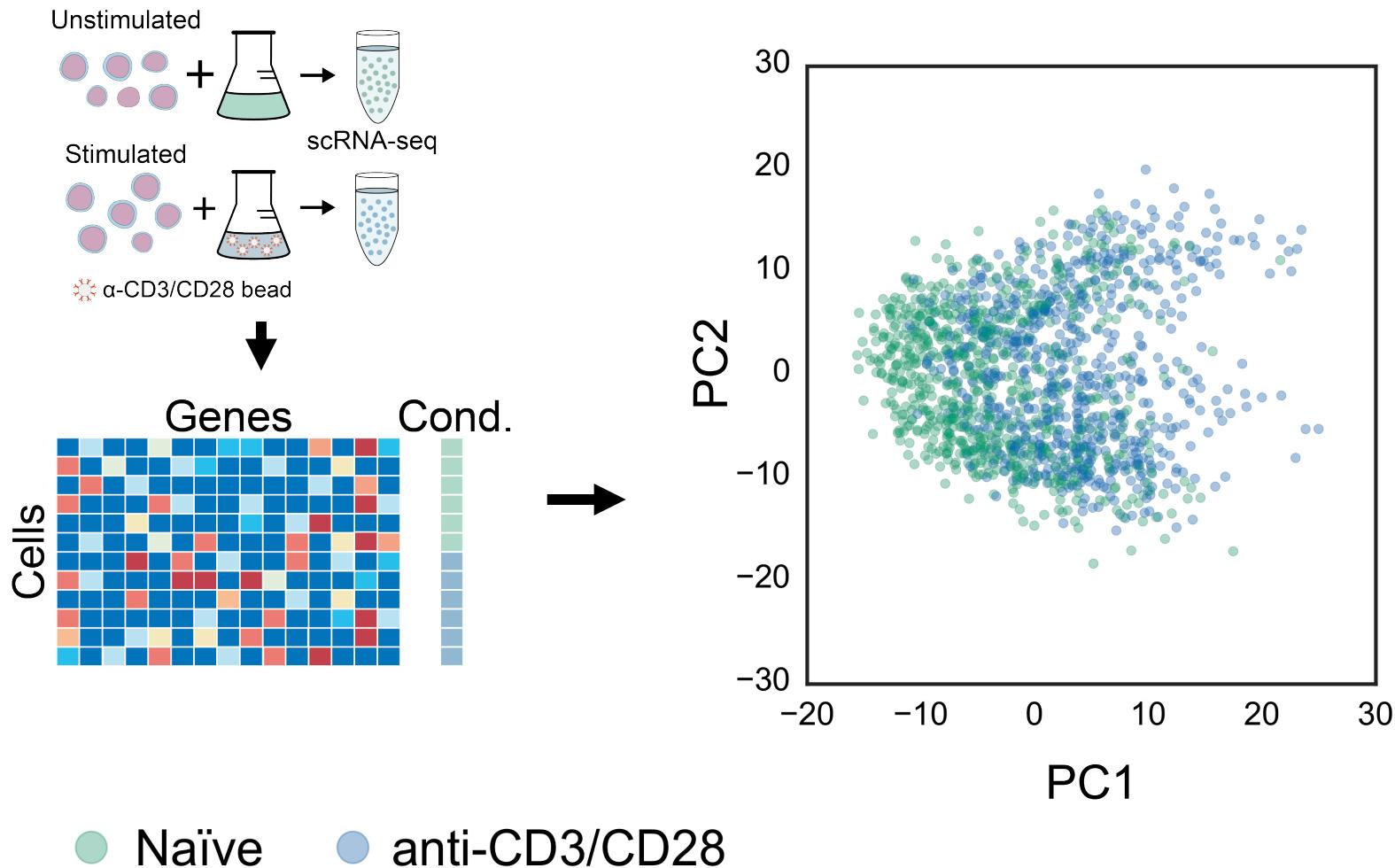


Measuring the difference between two experiments

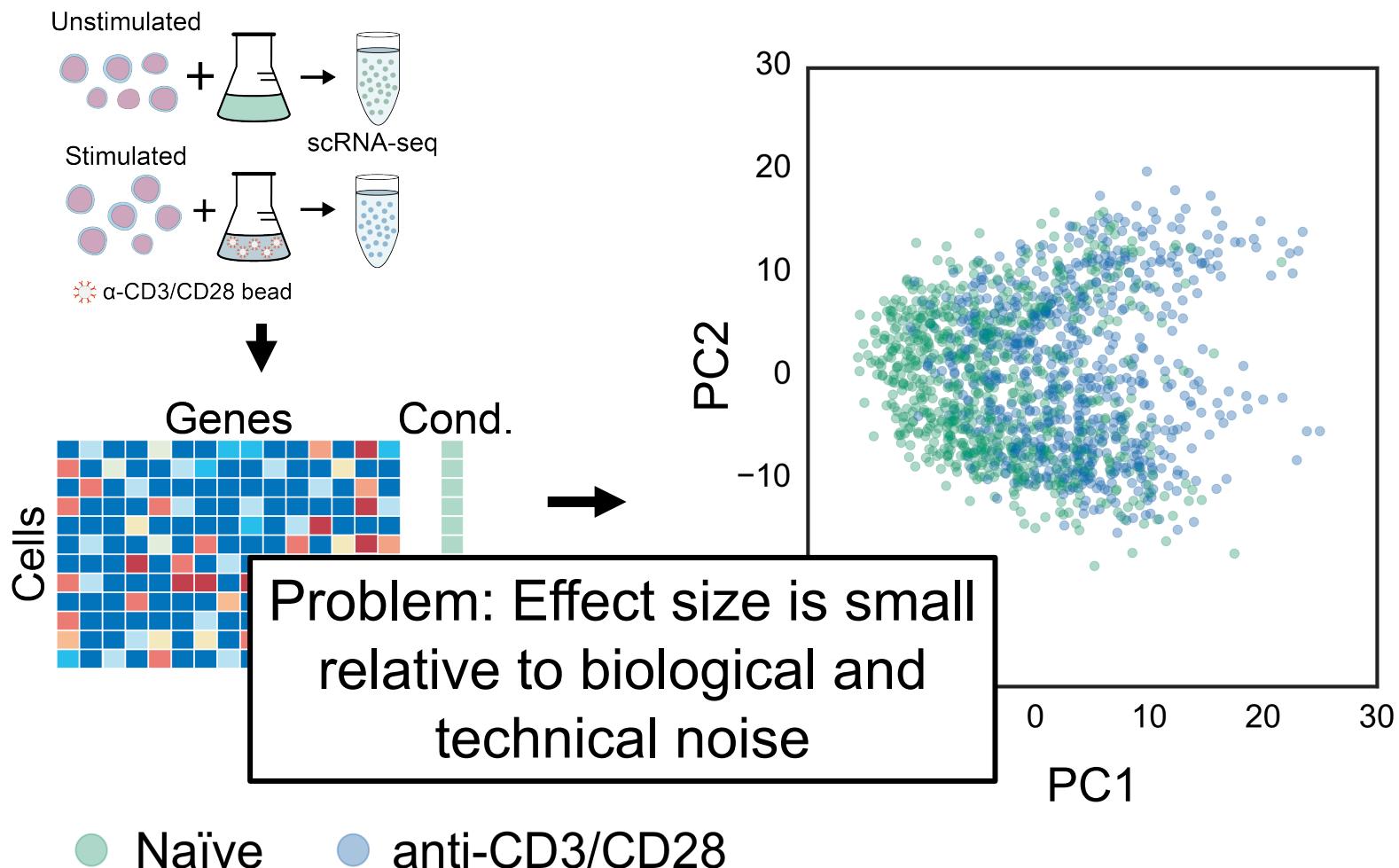




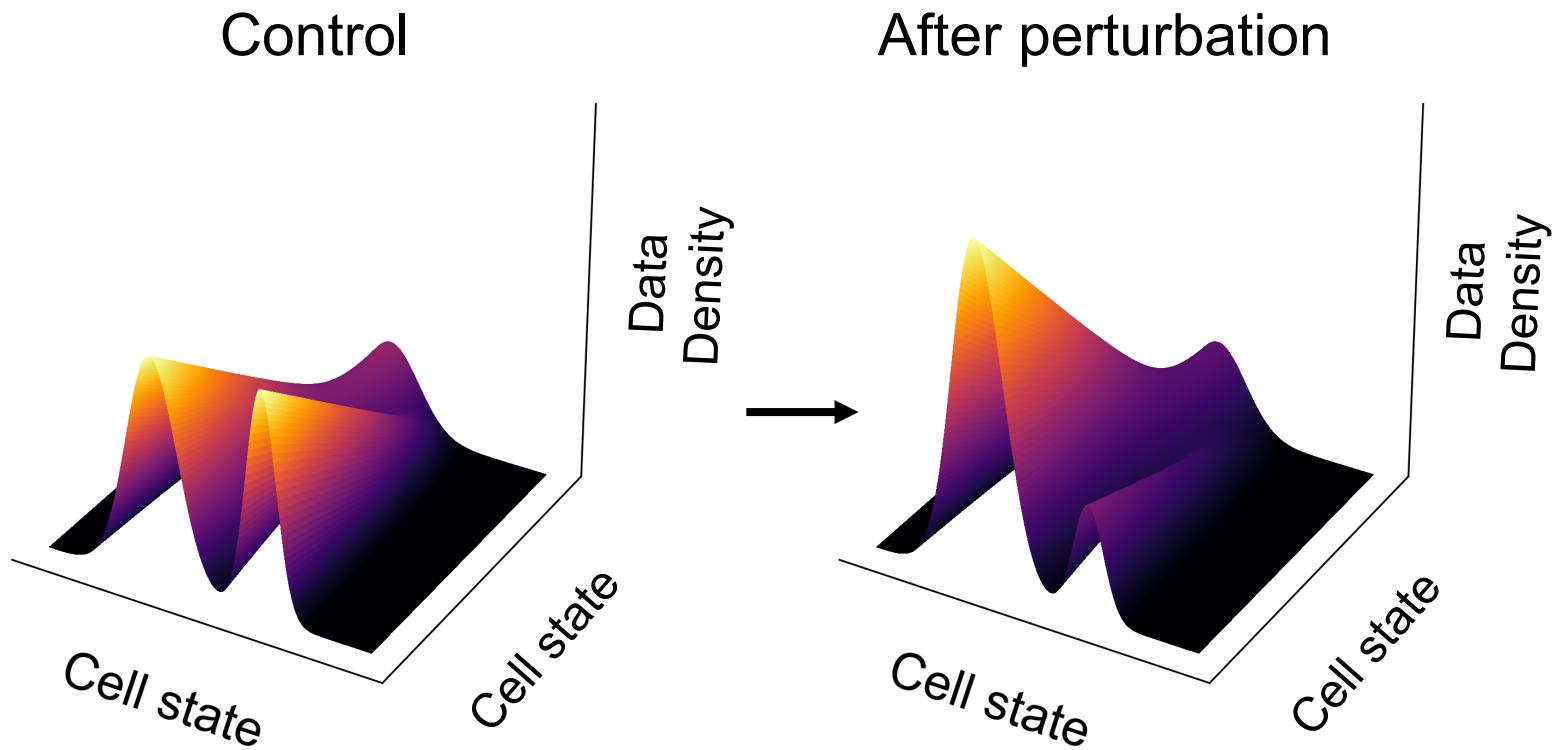
How do we measure the effect of TCR stimulation on T cell activation?



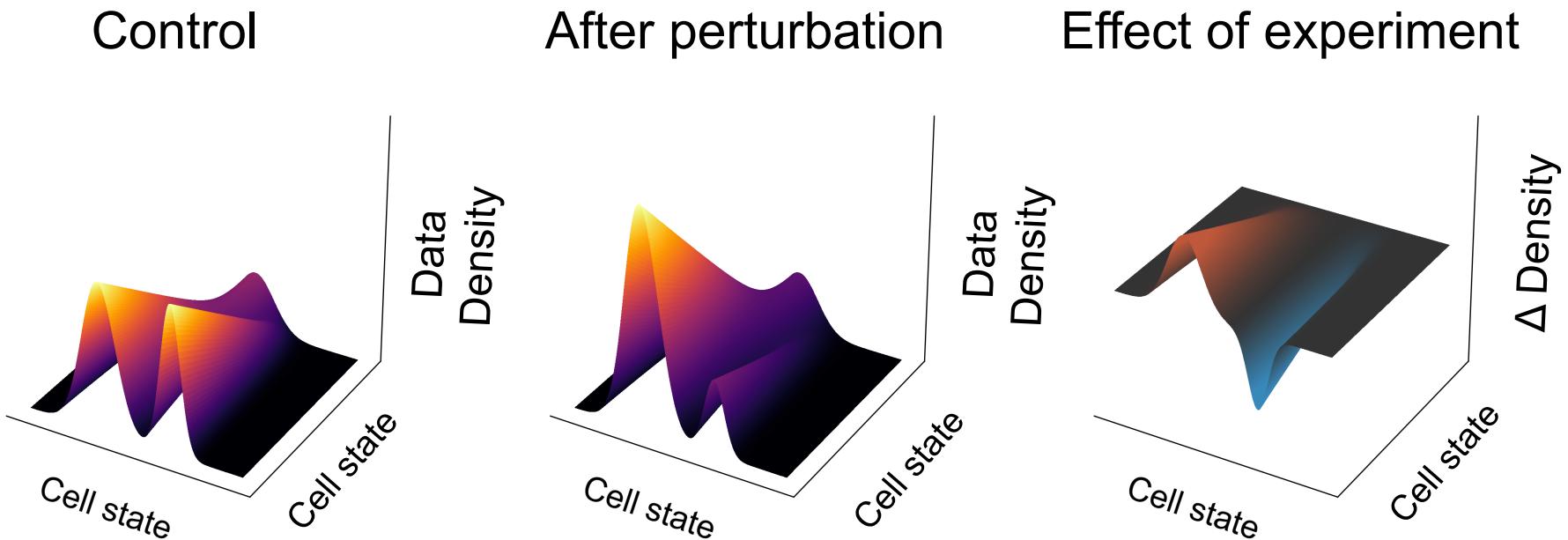
How do we measure the effect of TCR stimulation on T cell activation?



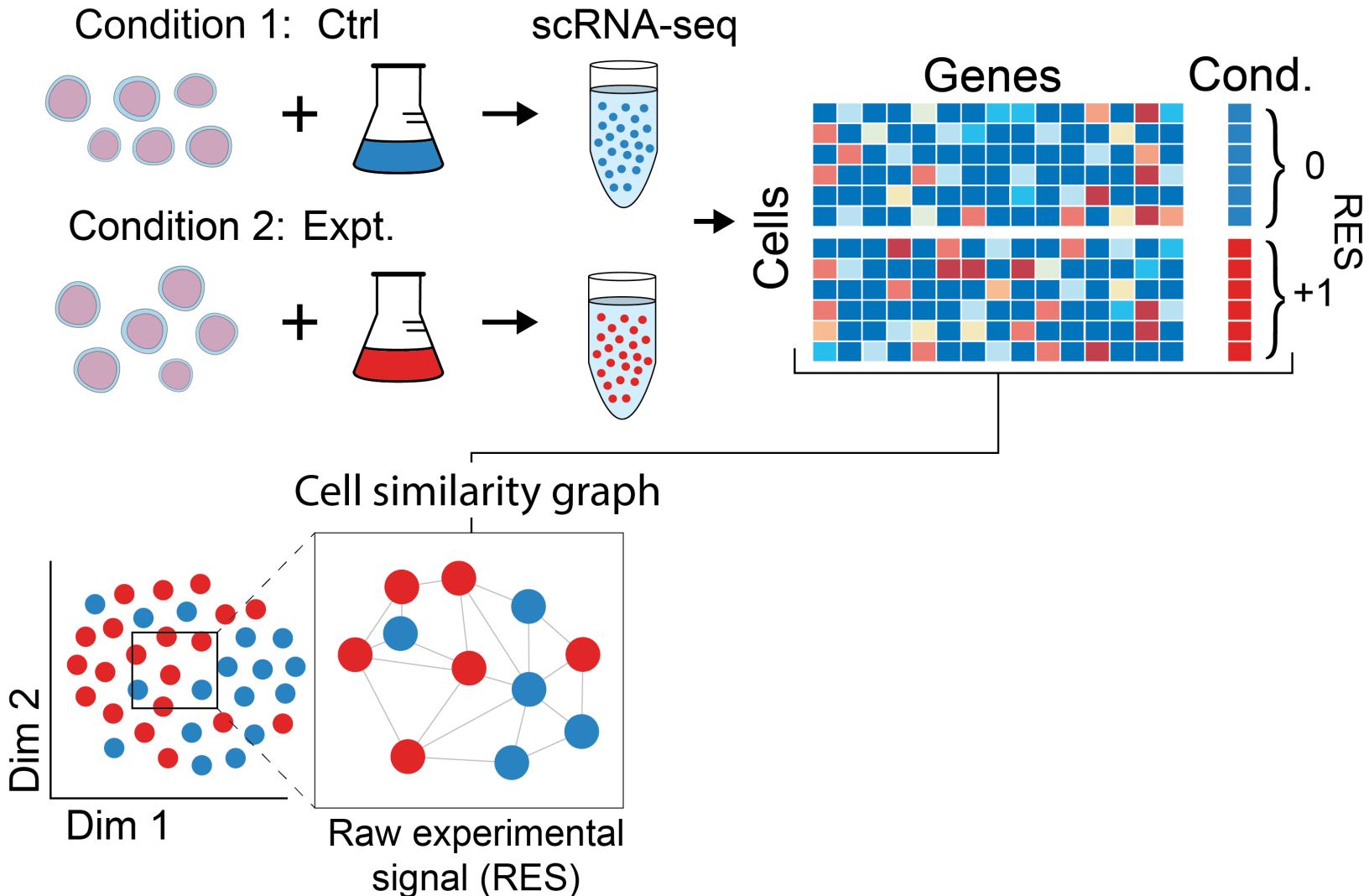
Goal: Quantify difference in data density



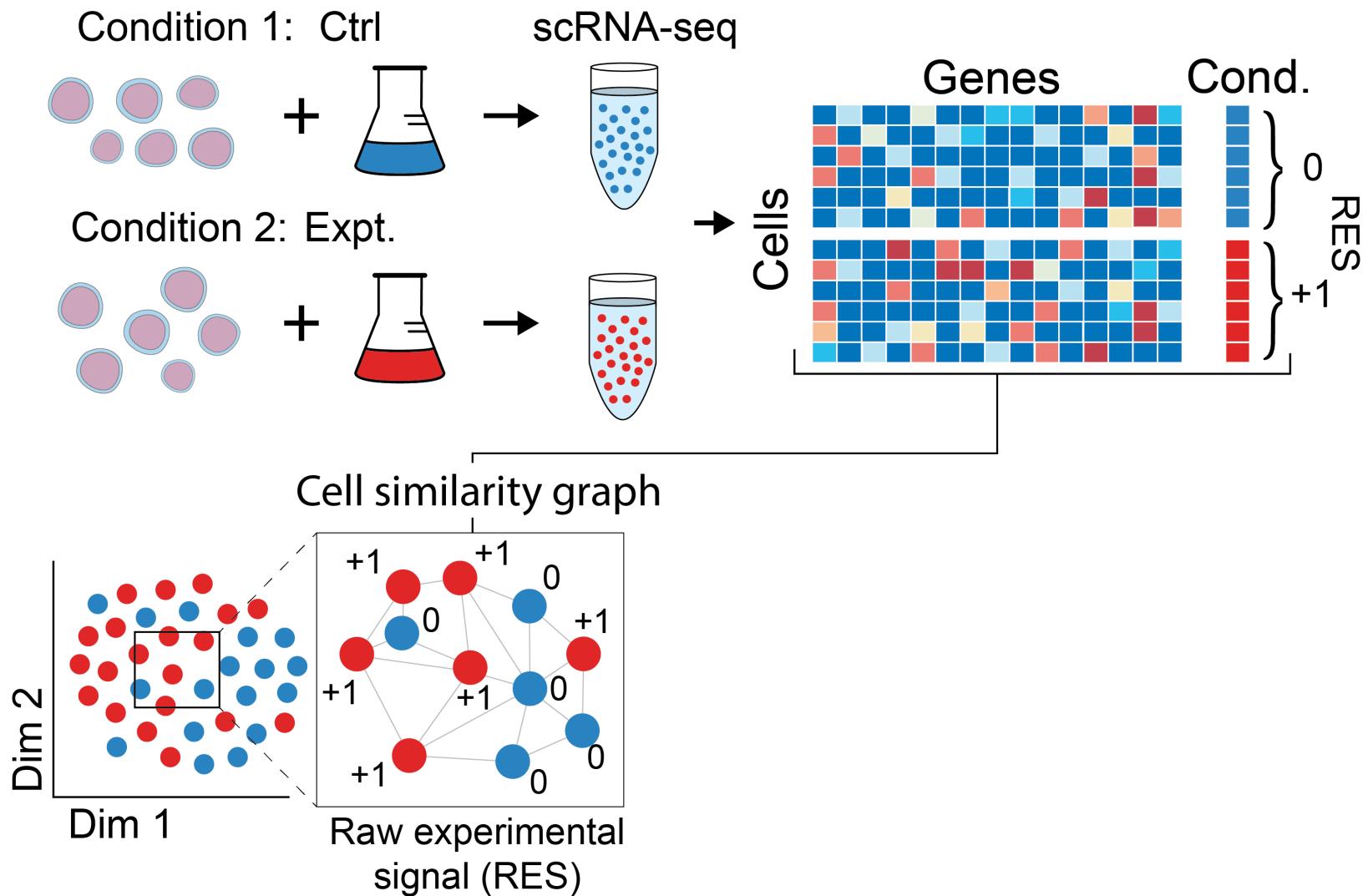
Goal: Quantify difference in data density



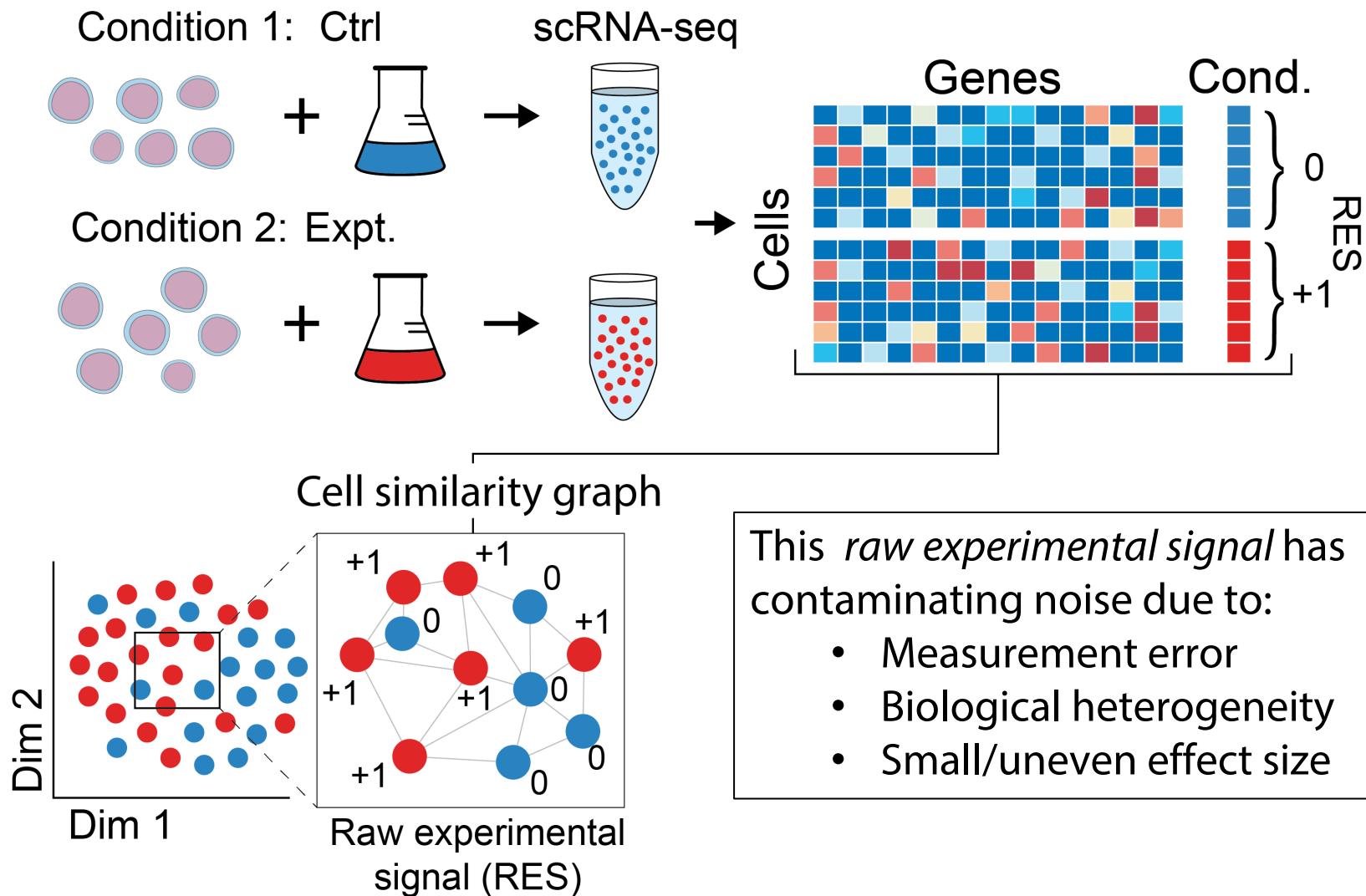
Experimental signal on cell graph



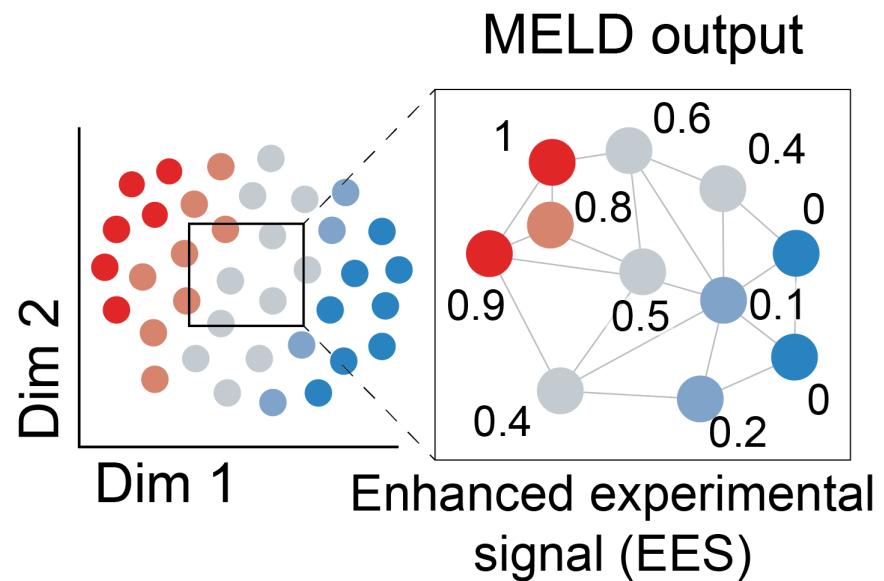
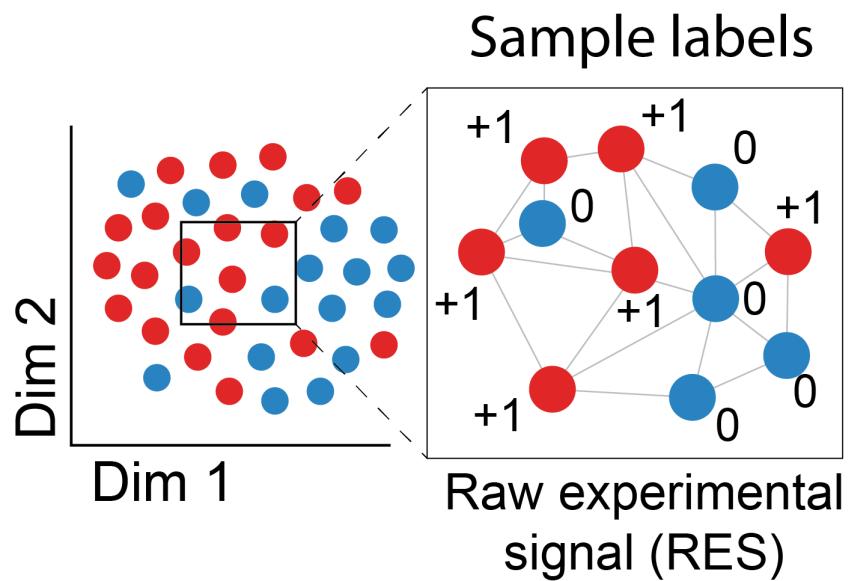
Experimental signal on cell graph



Experimental signal on cell graph



Smoothing signal removes high frequency noise



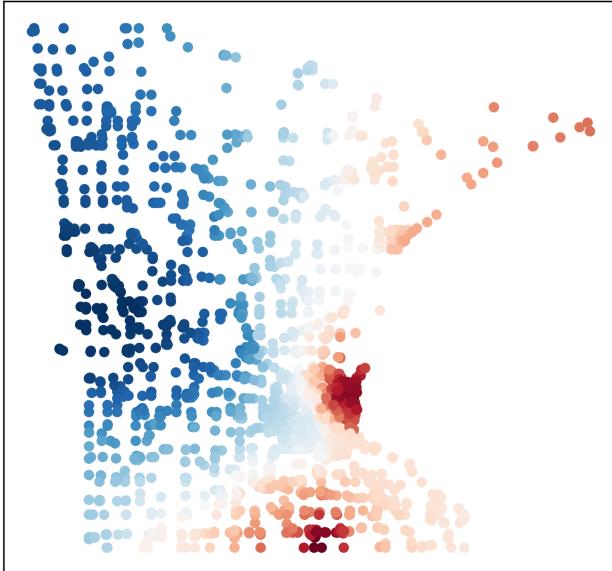
The MELD algorithm

1. Learn a graph from the data using a Gaussian kernel
2. Instantiate the Raw Experimental Signal, x , using the sample labels
3. Filter the RES via Laplacian smoothing

$$\hat{x} = \underset{\hat{x}}{\operatorname{argmin}} |x - \hat{x}|_2^2 + \beta (\hat{x}^T \mathcal{L} \hat{x})$$

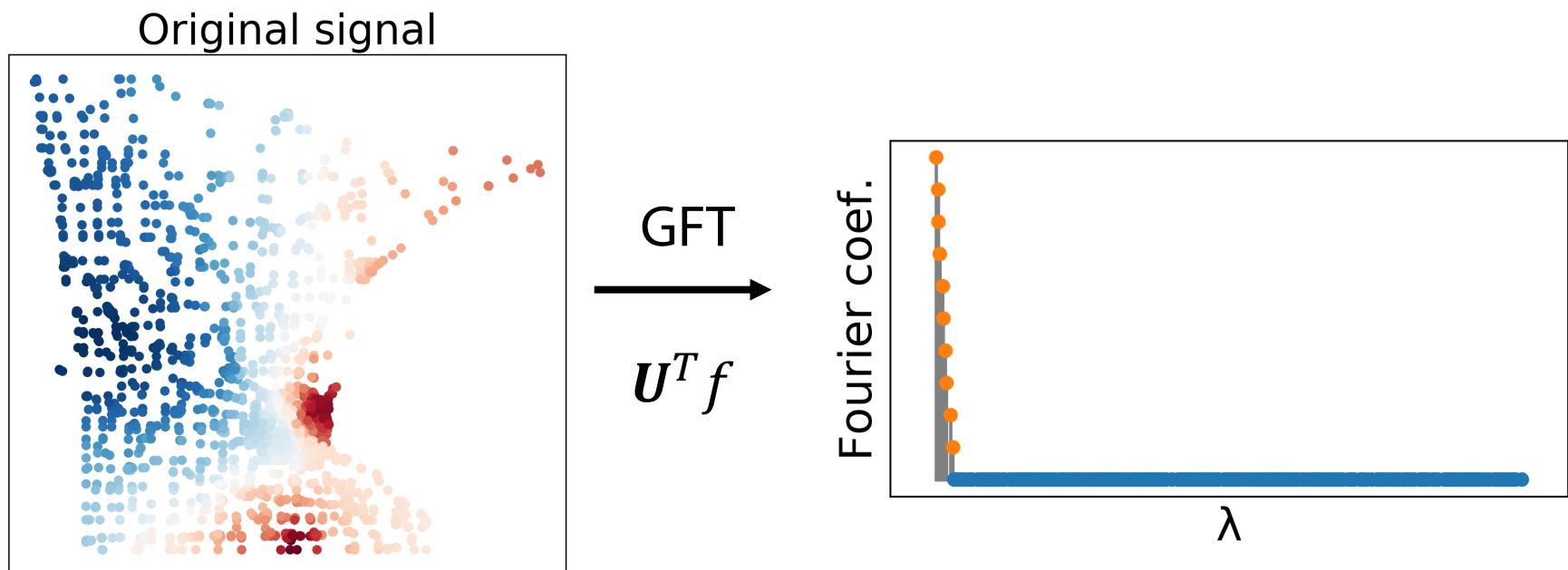
Lowpass filtering recovers low frequency signals

Original signal



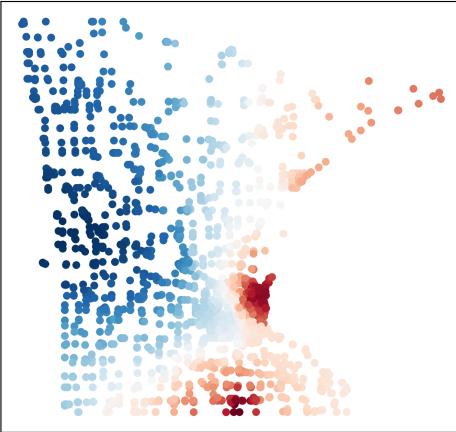
- Given the graph Laplacian $\mathcal{L} = \mathbf{U}\Lambda\mathbf{U}^T$
- GFT of f is the projection of f onto the Laplacian eigenbasis:
$$\mathcal{F}_G = \hat{f} = \mathbf{U}^T f$$

Lowpass filtering recovers low frequency signals



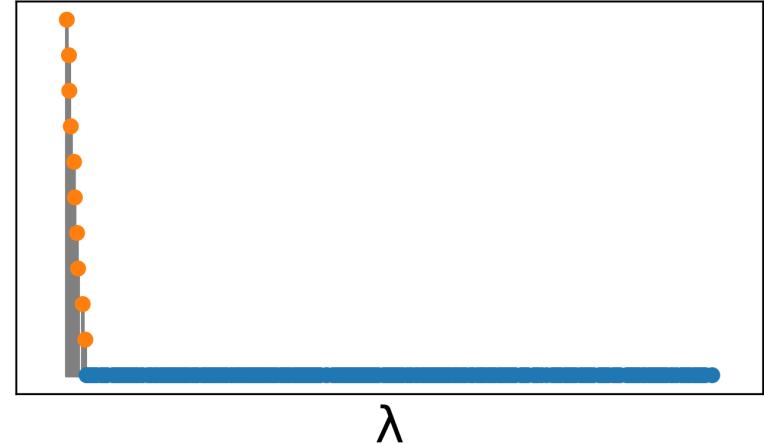
Lowpass filtering recovers low frequency signals

Original signal

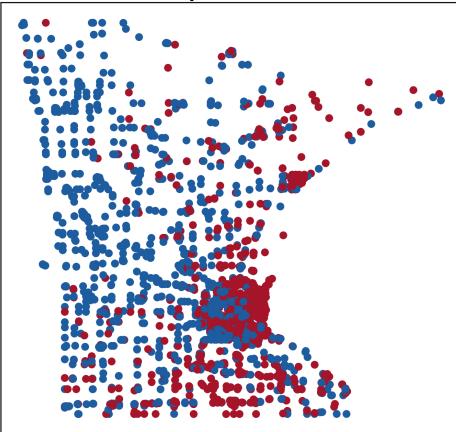


GFT
→
 $U^T f$

Fourier coef.

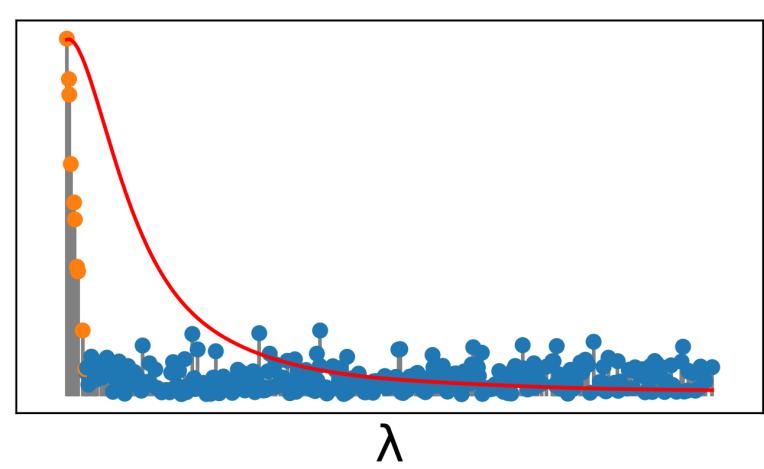


Sample labels



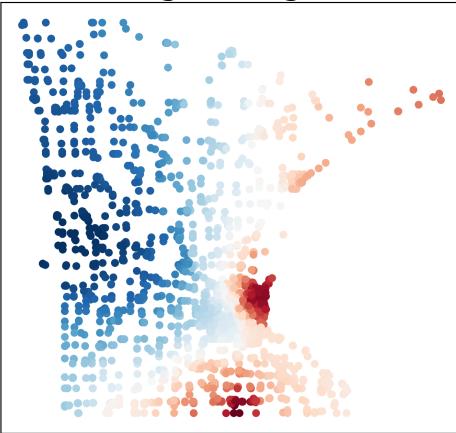
GFT
→
 $U^T f$

Fourier coef.



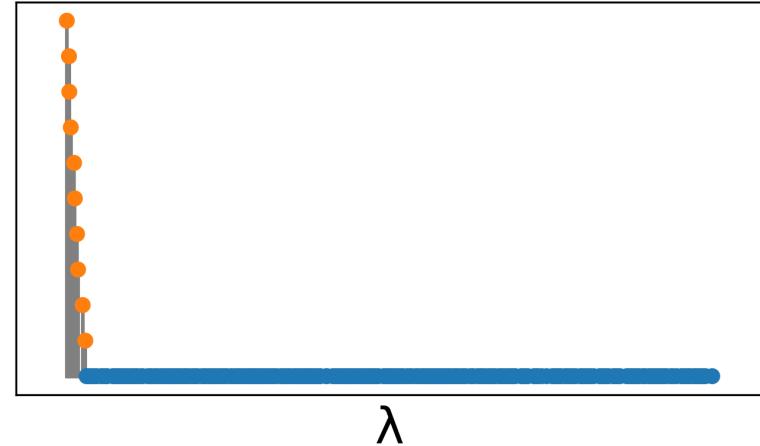
Lowpass filtering recovers low frequency signals

Original signal

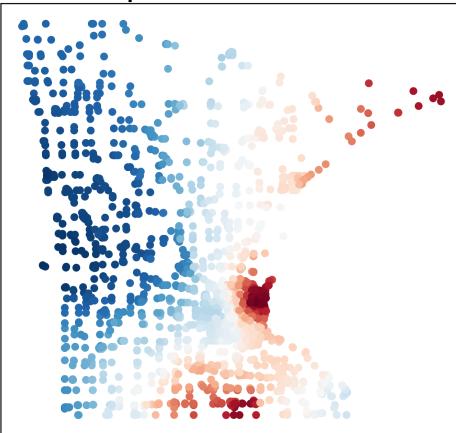


GFT
 \longrightarrow
 $U^T f$

Fourier coef.

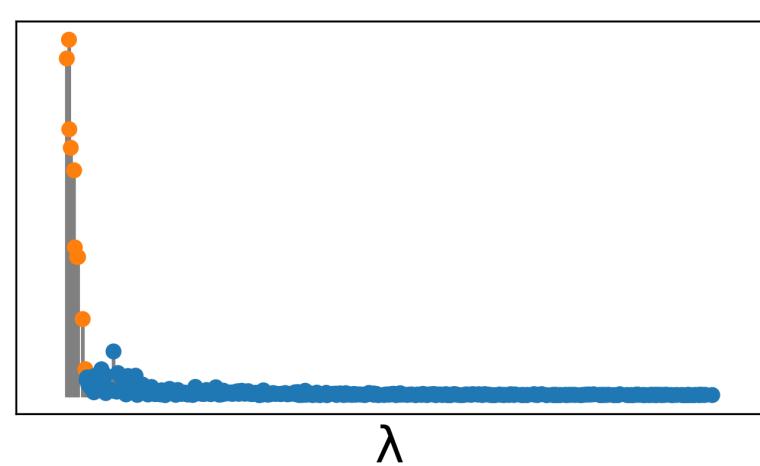


Composite recovered



GFT
 \longrightarrow
 $U^T f$

Fourier coef.

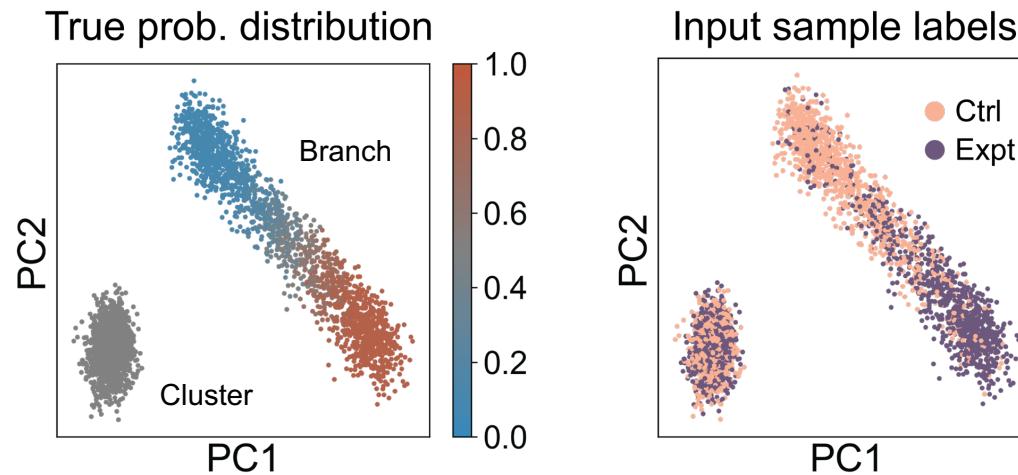


The MELD algorithm

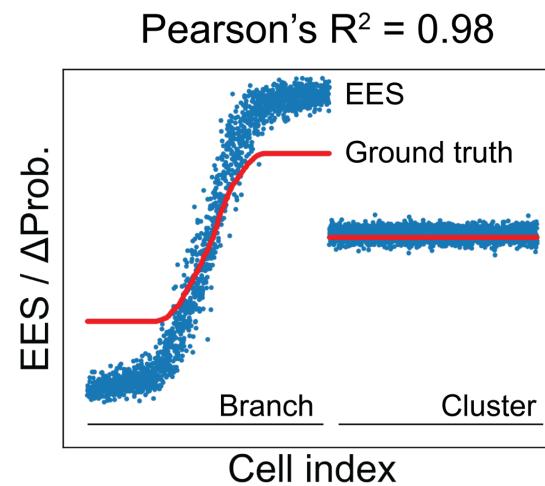
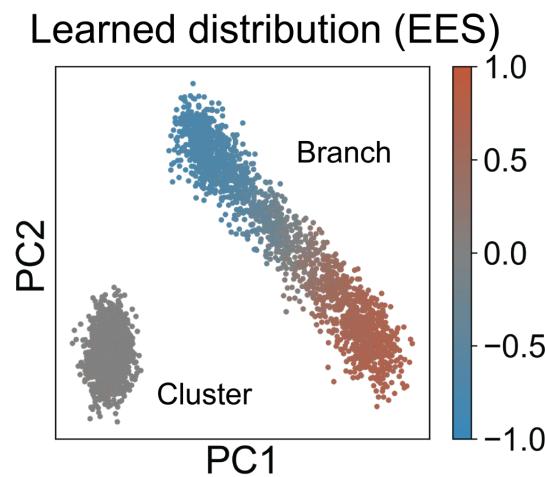
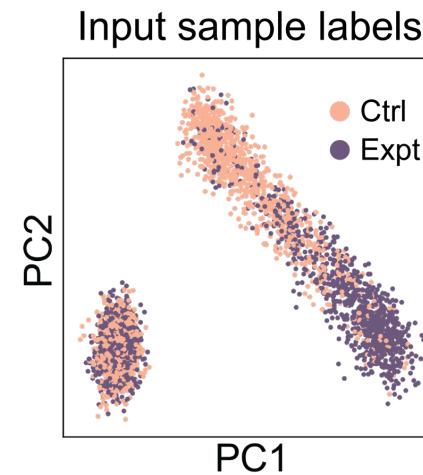
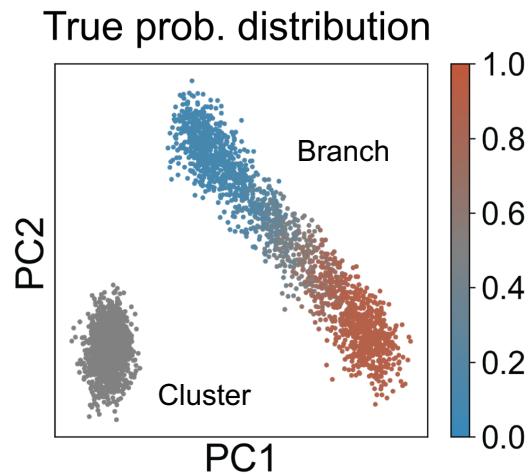
1. Learn a graph from the data using a Gaussian kernel
2. Instantiate the Raw Experimental Signal, x , using the sample labels
3. Filter the RES via Laplacian smoothing

$$\hat{x} = \underset{\hat{x}}{\operatorname{argmin}} |x - \hat{x}|_2^2 + \beta (\hat{x}^T \mathcal{L} \hat{x})$$

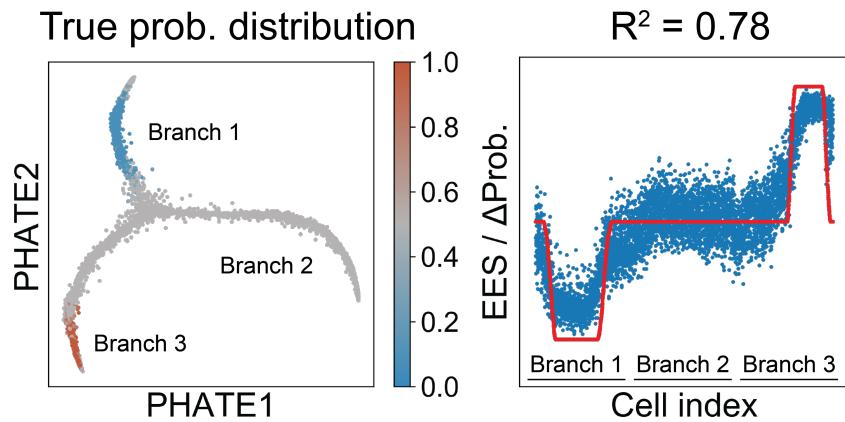
Simulated scRNA-seq data



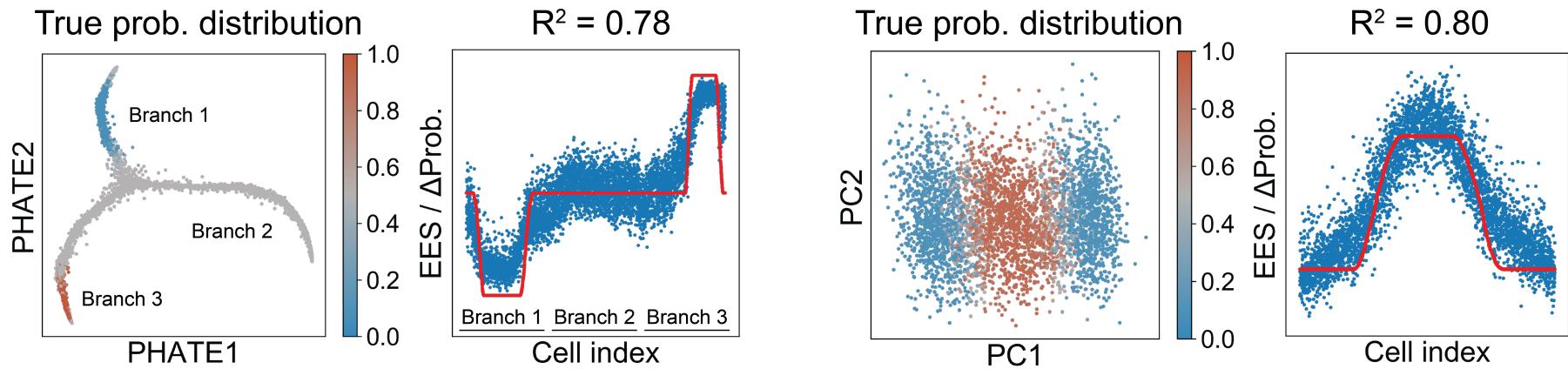
Simulated scRNA-seq data



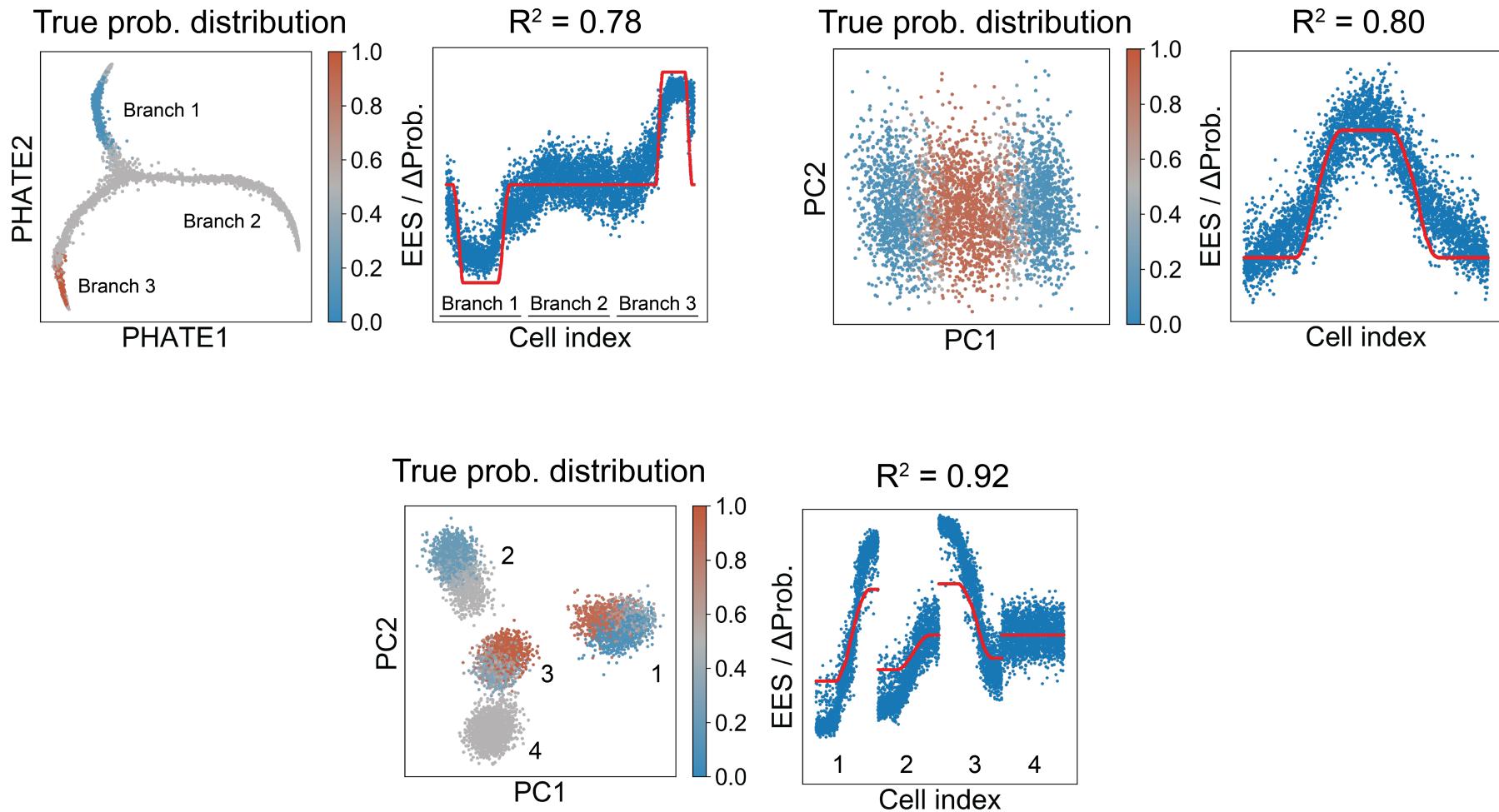
Simulated scRNA-seq data



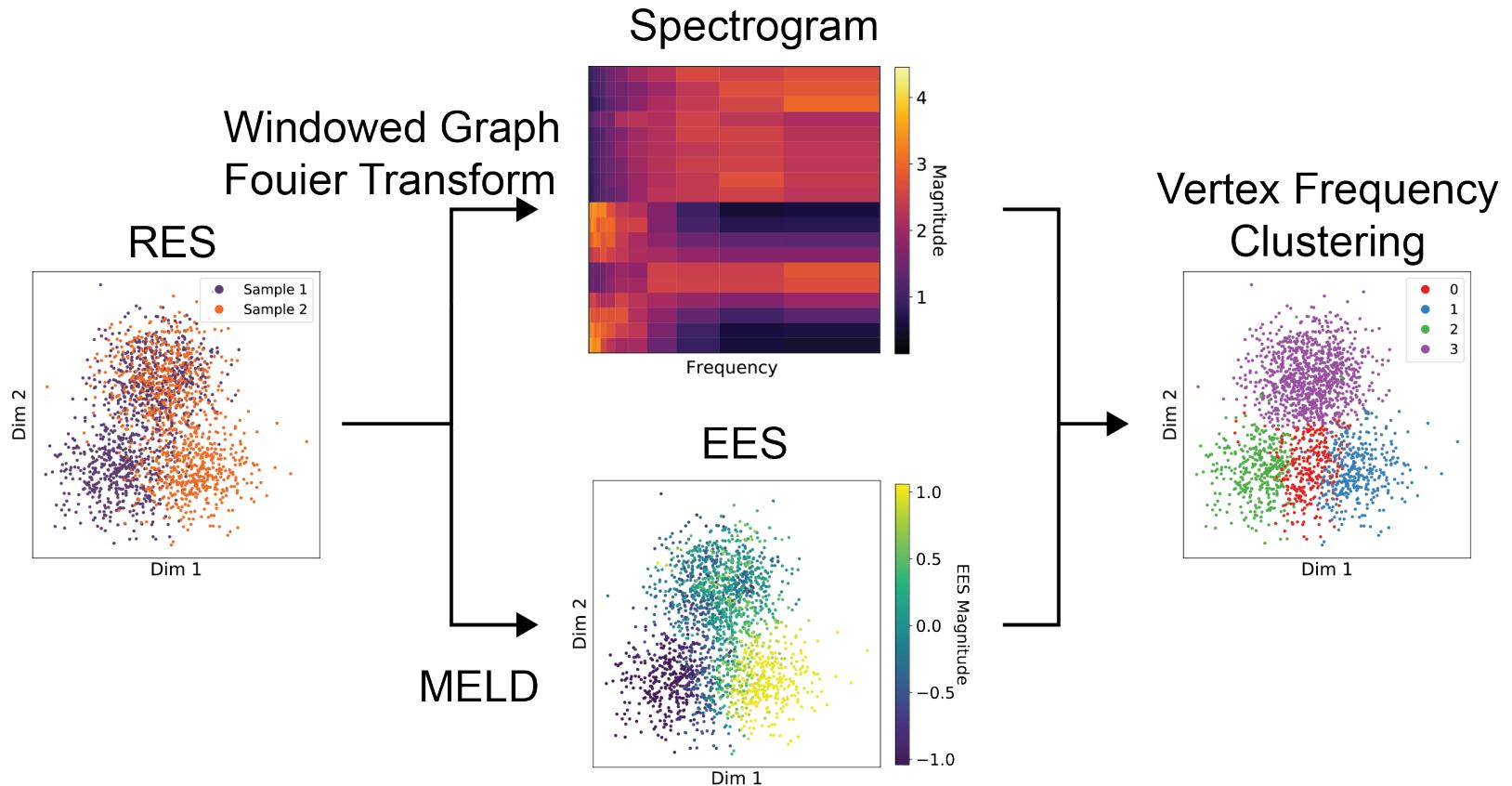
Simulated scRNA-seq data



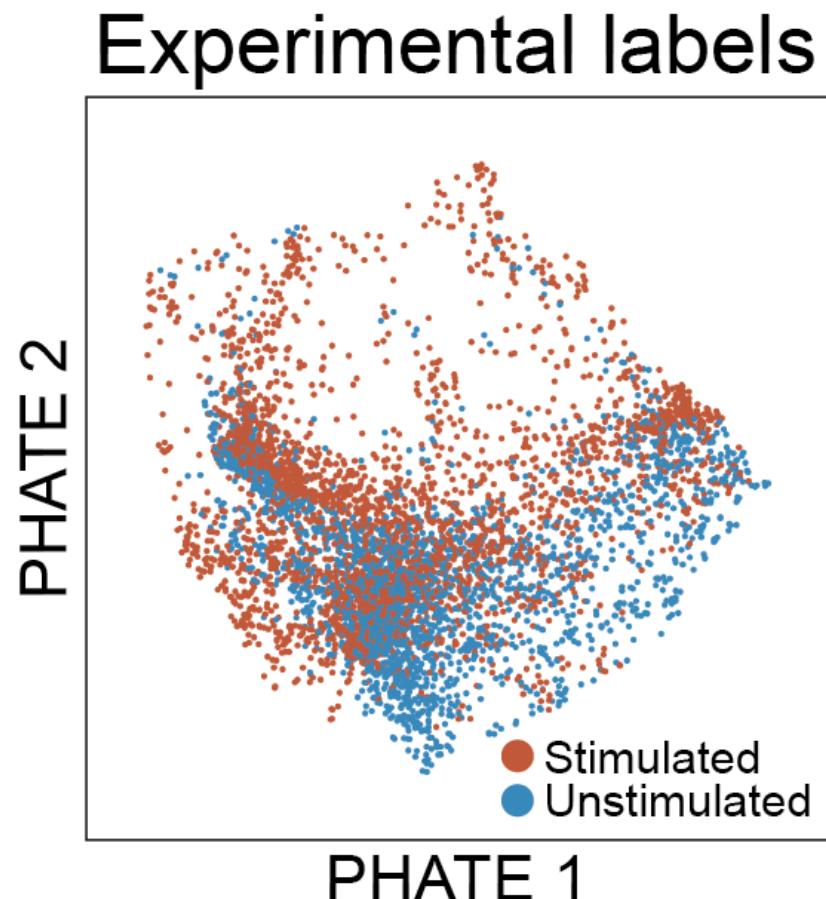
Simulated scRNA-seq data



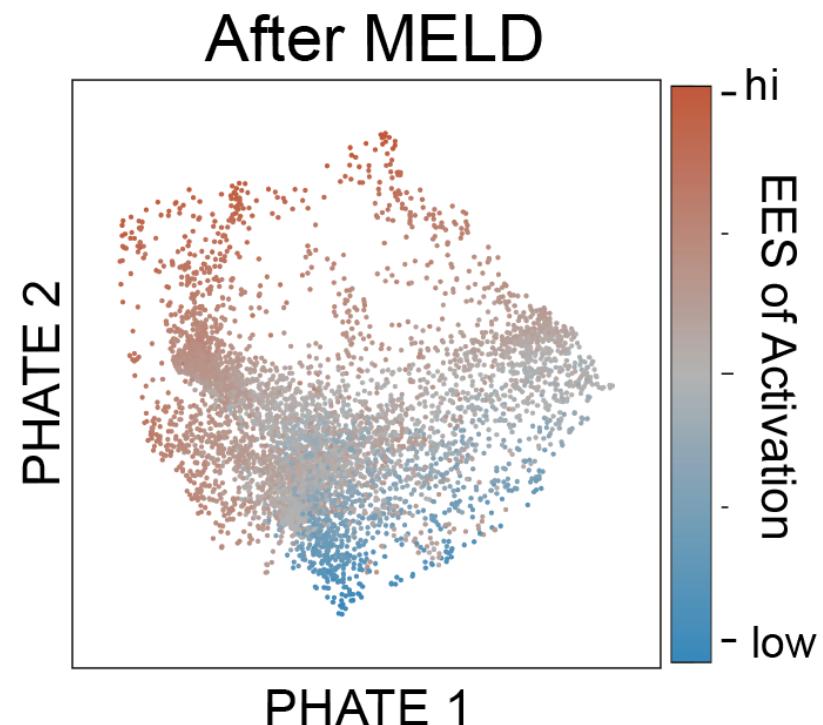
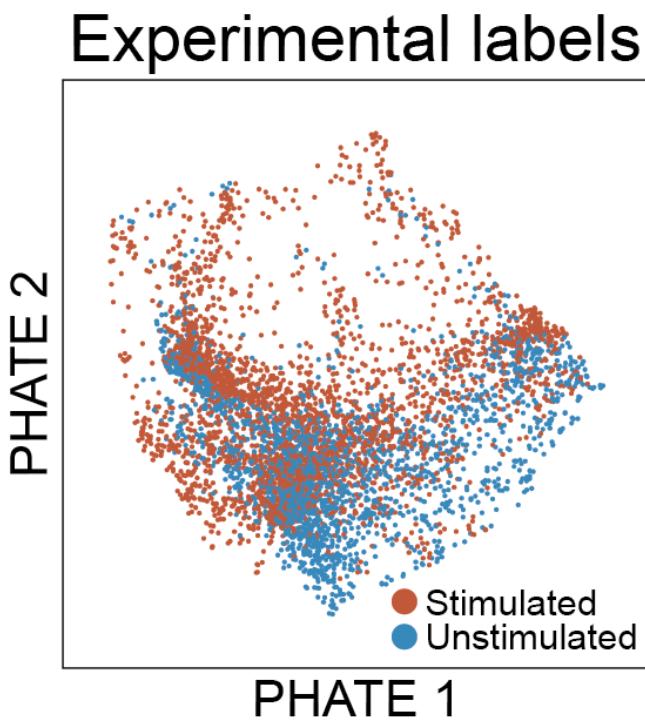
Vertex Frequency Clustering identifies clusters using RES and EES



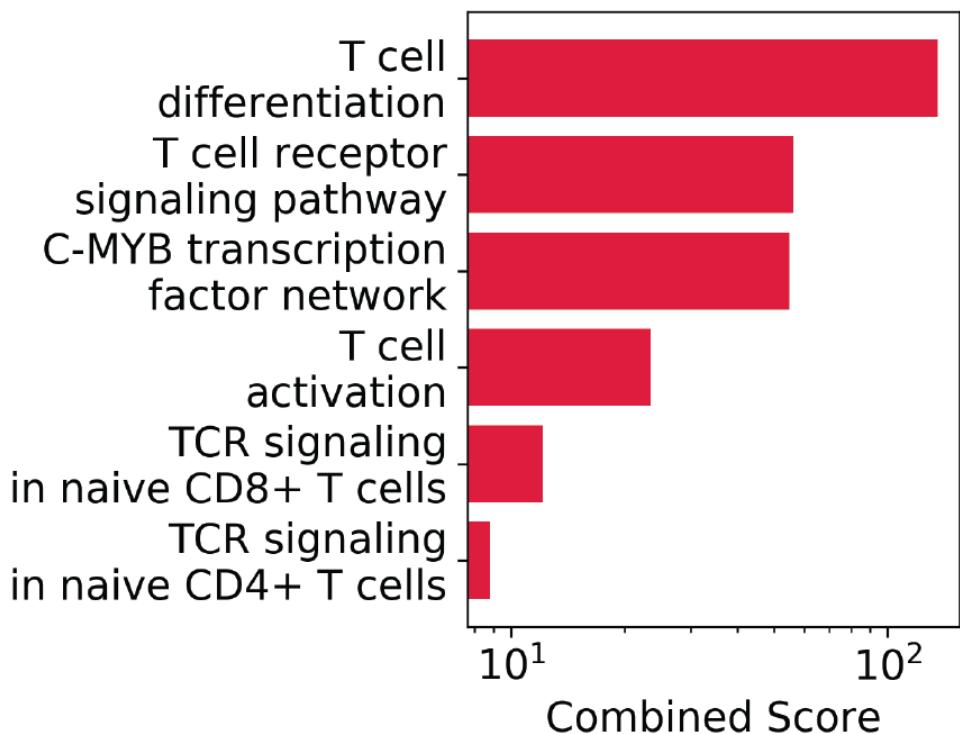
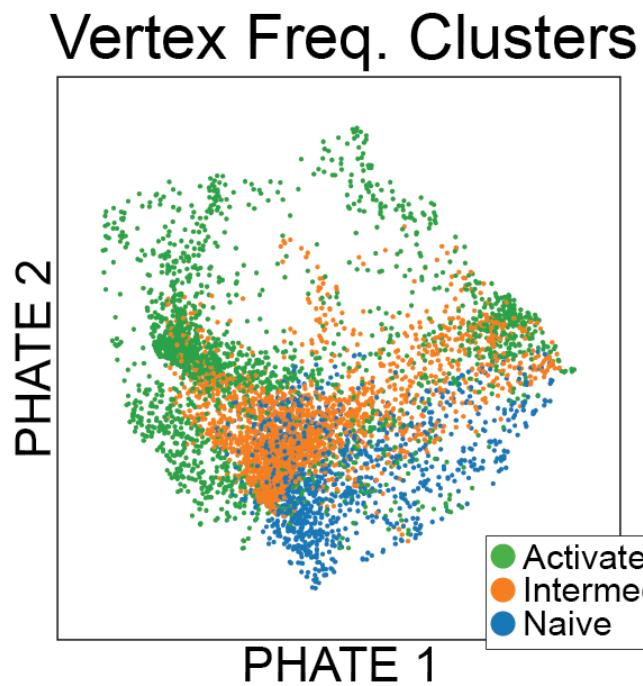
Directly analyzing experimental labels does not reveal effect of TCR induction



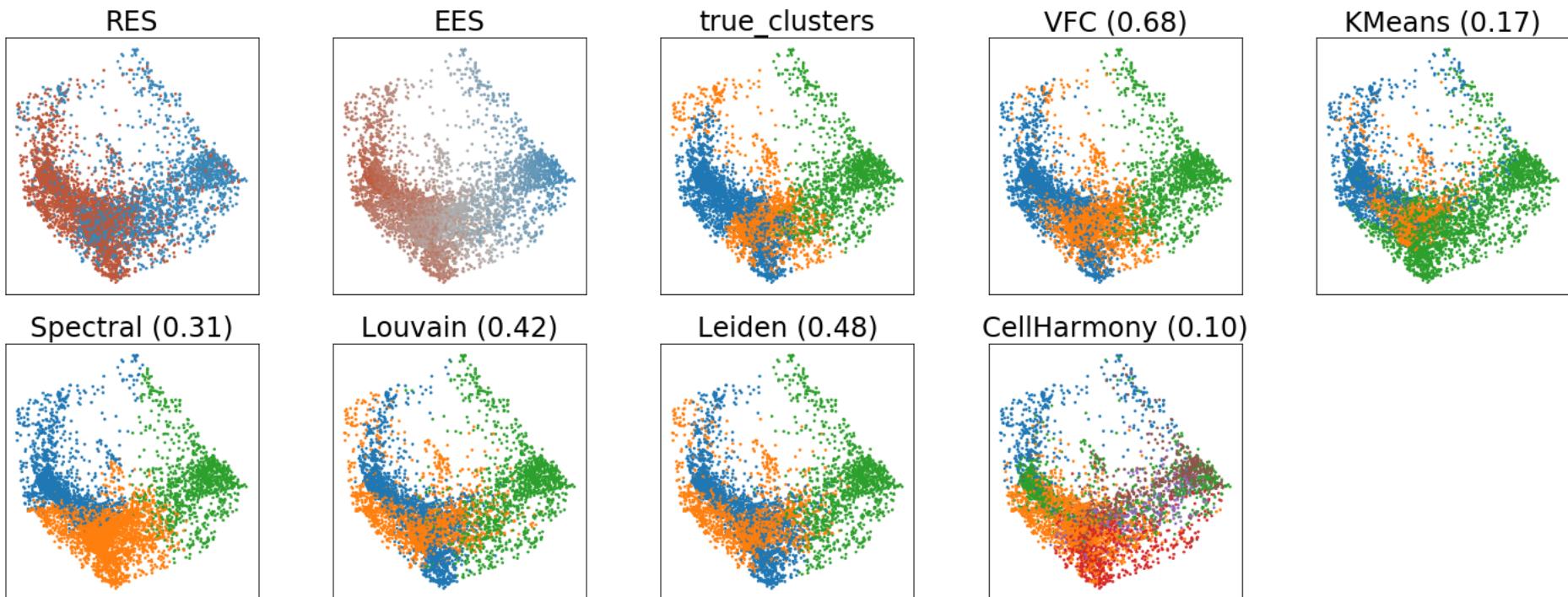
MELD reveals a continuous spectrum of TCR induction



The EES identifies the signature of TCR induction



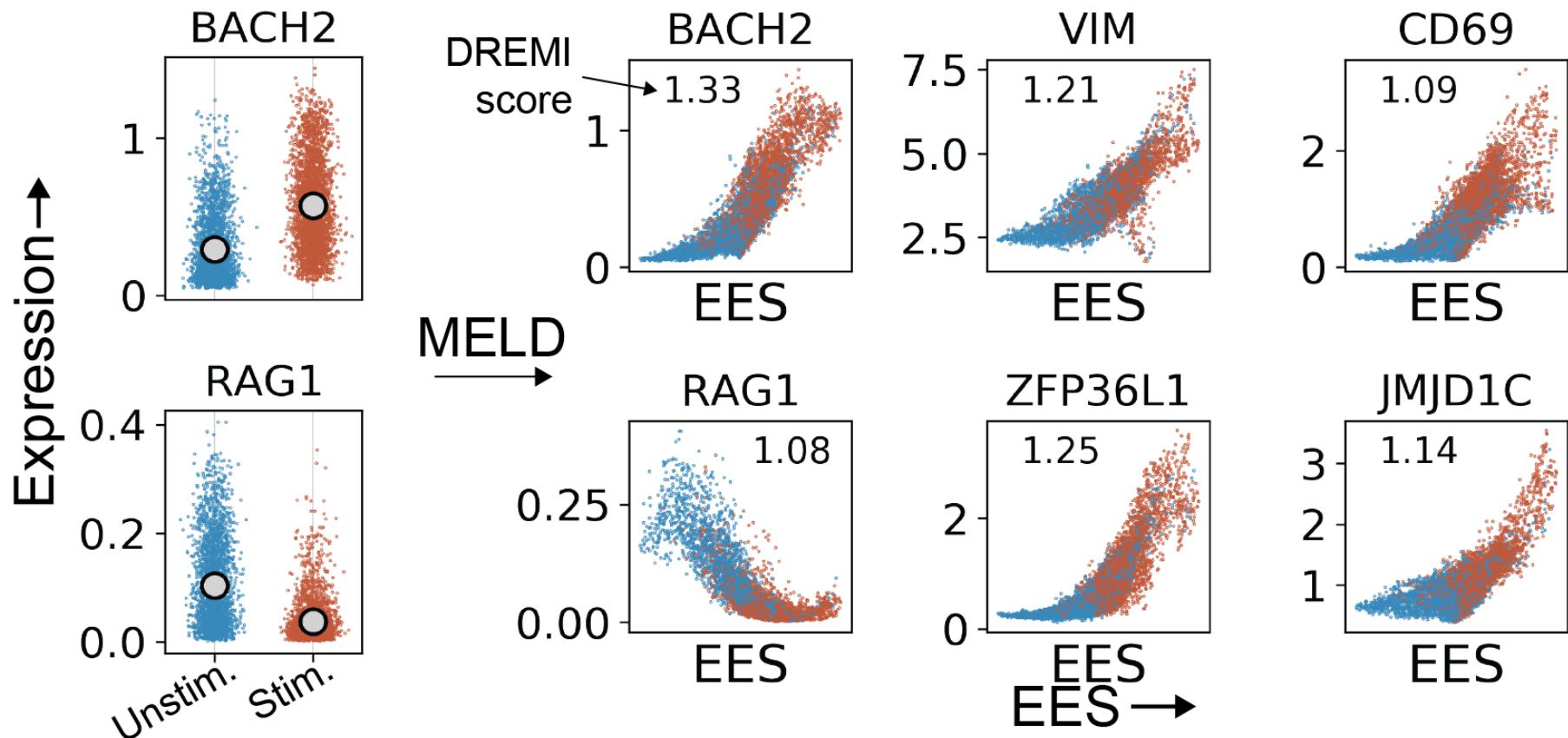
VFC identifies enriched groups of cells better than other methods

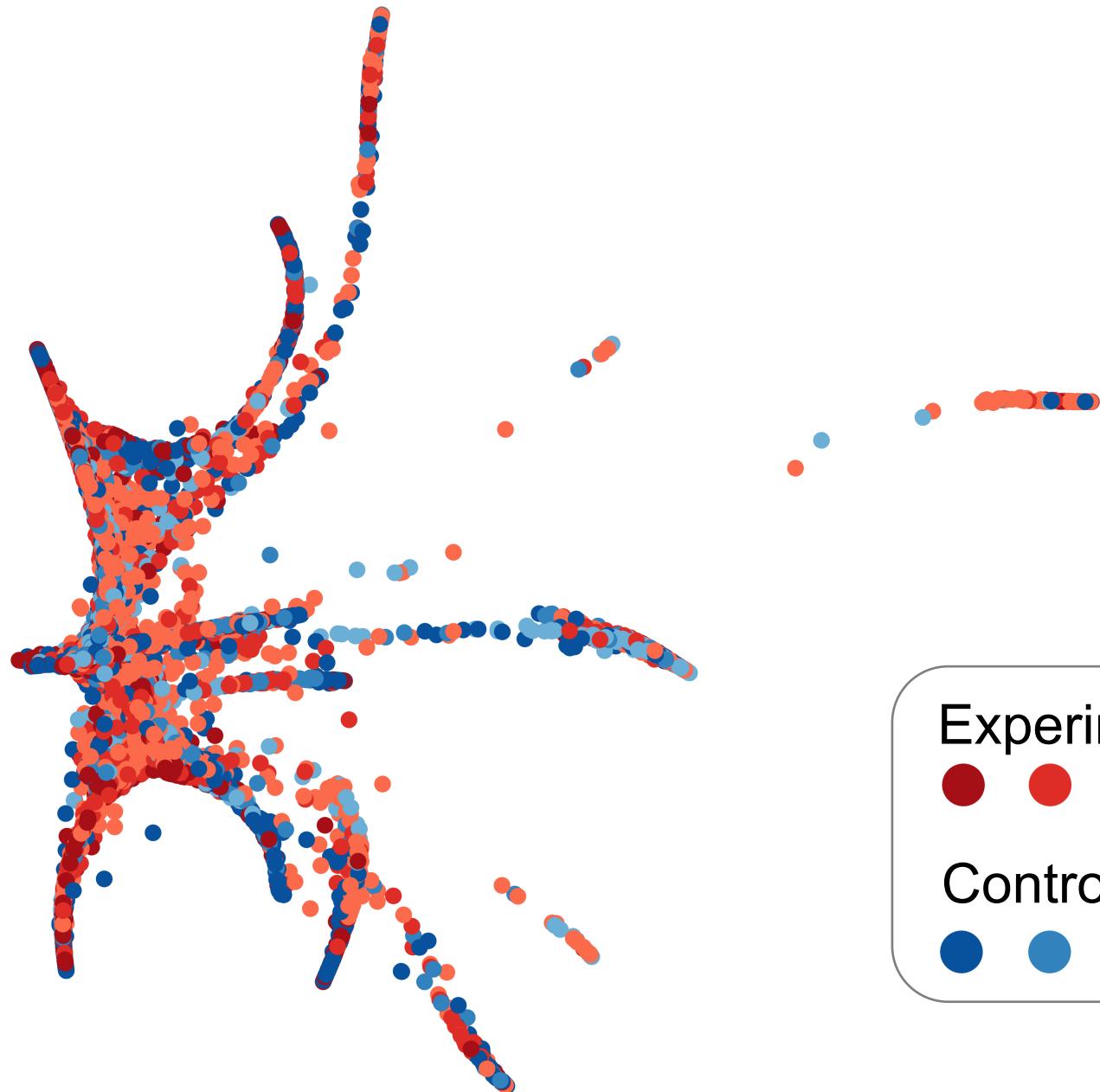


(adjusted rand score)

The EES identifies the signature of TCR induction

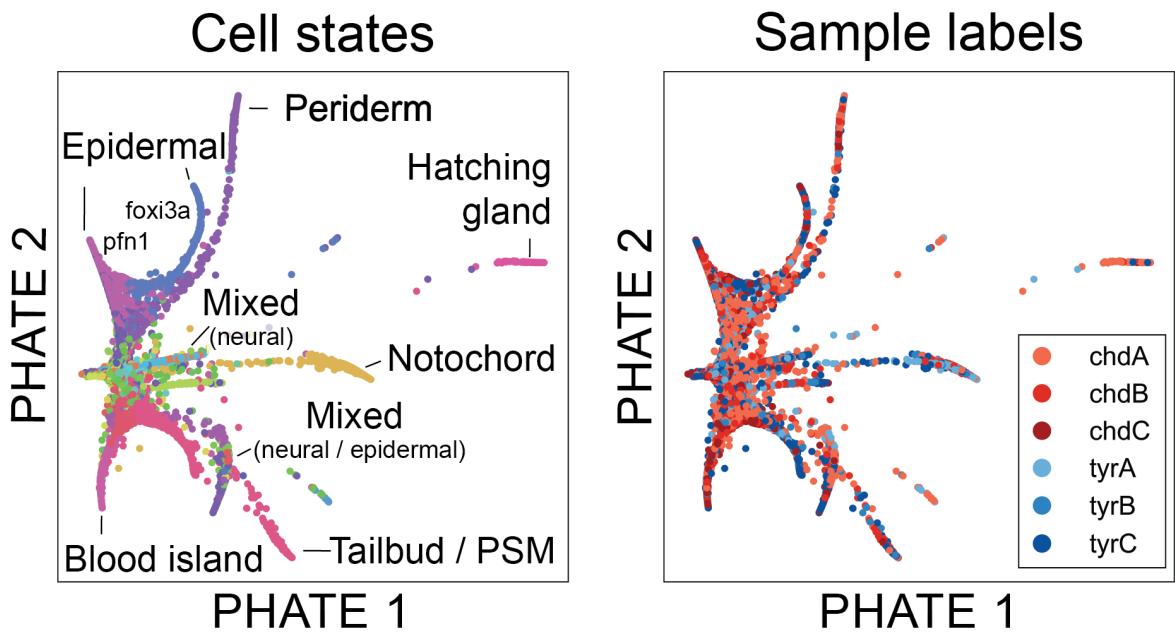
Top 1% genes associated with EES

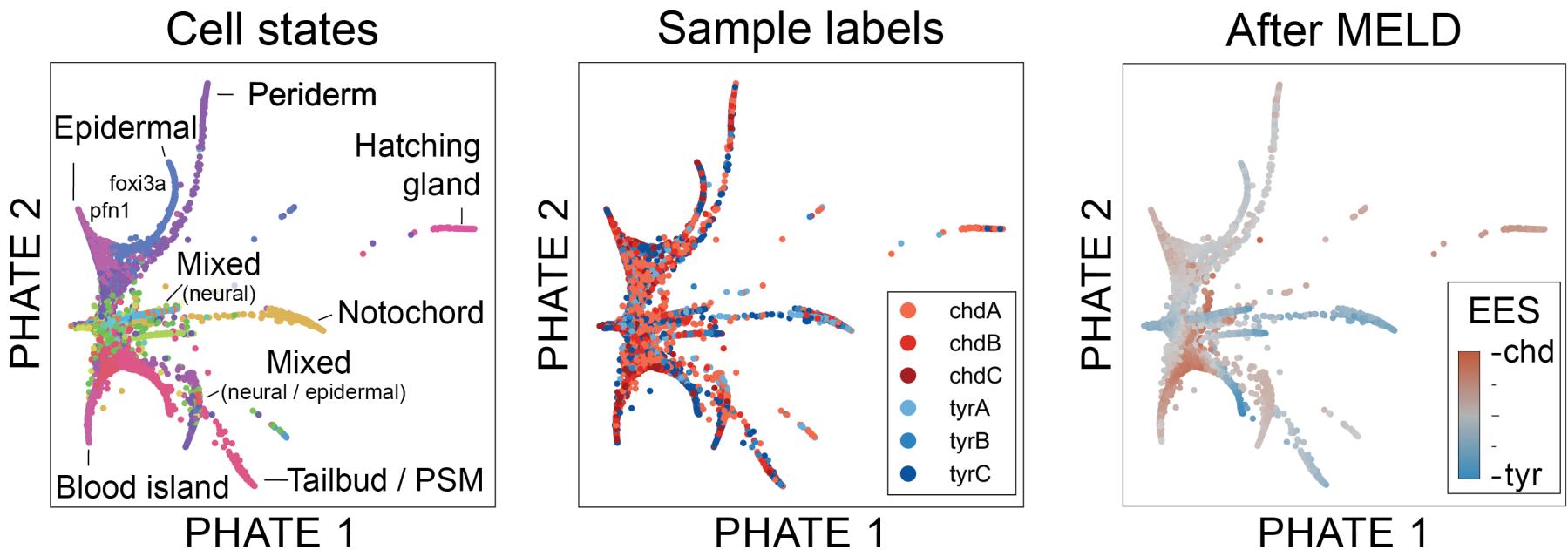


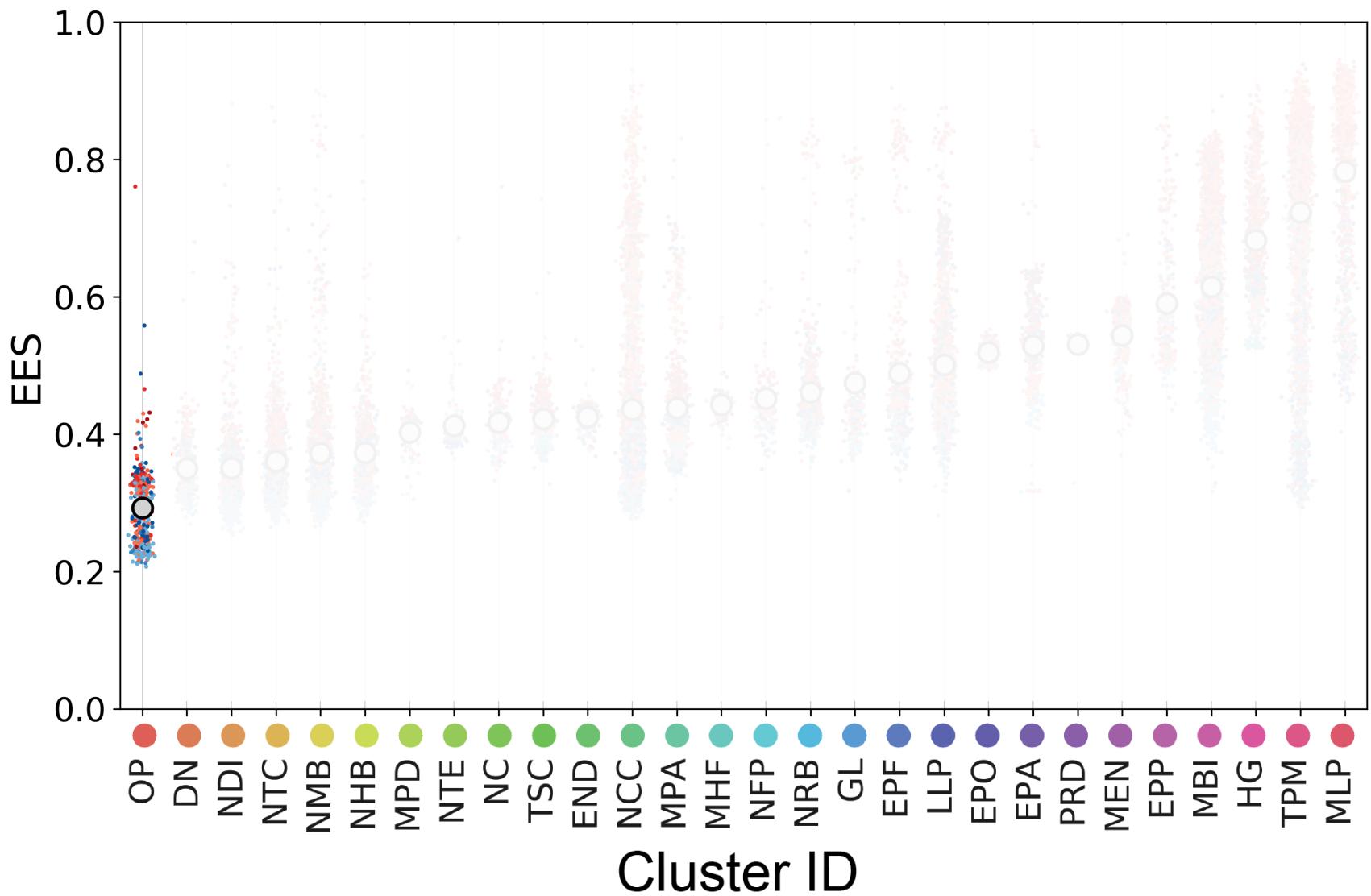


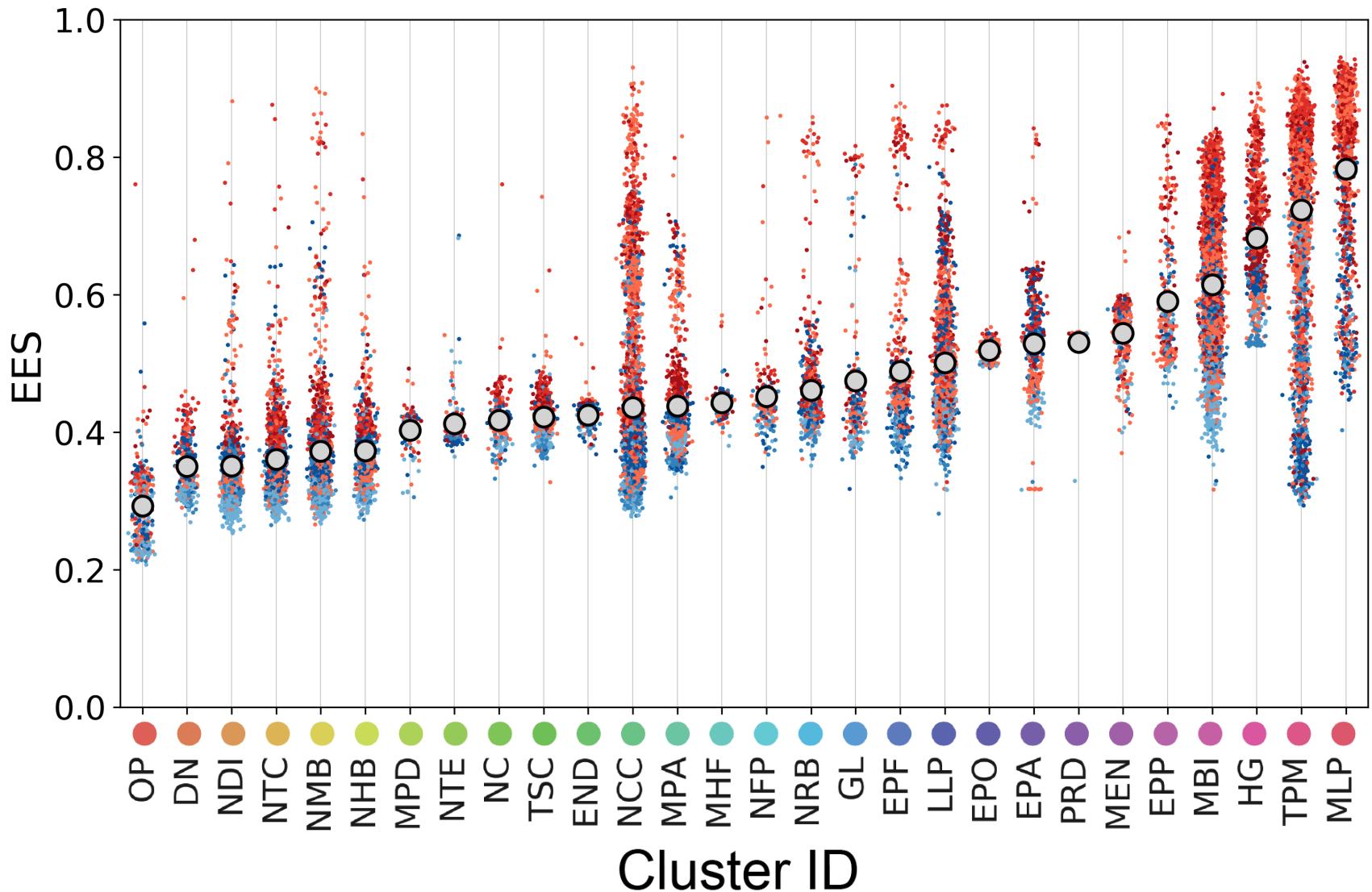
Experiment

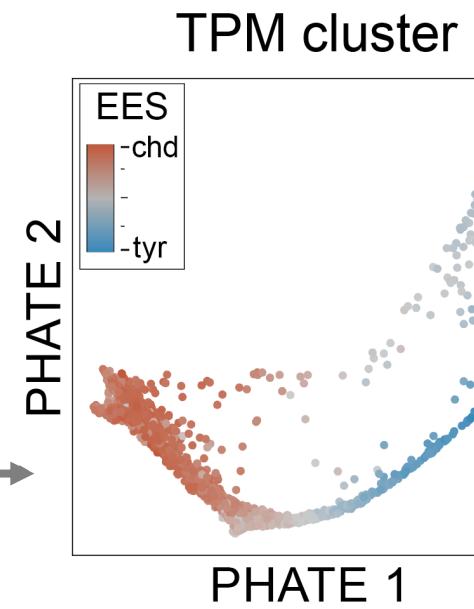
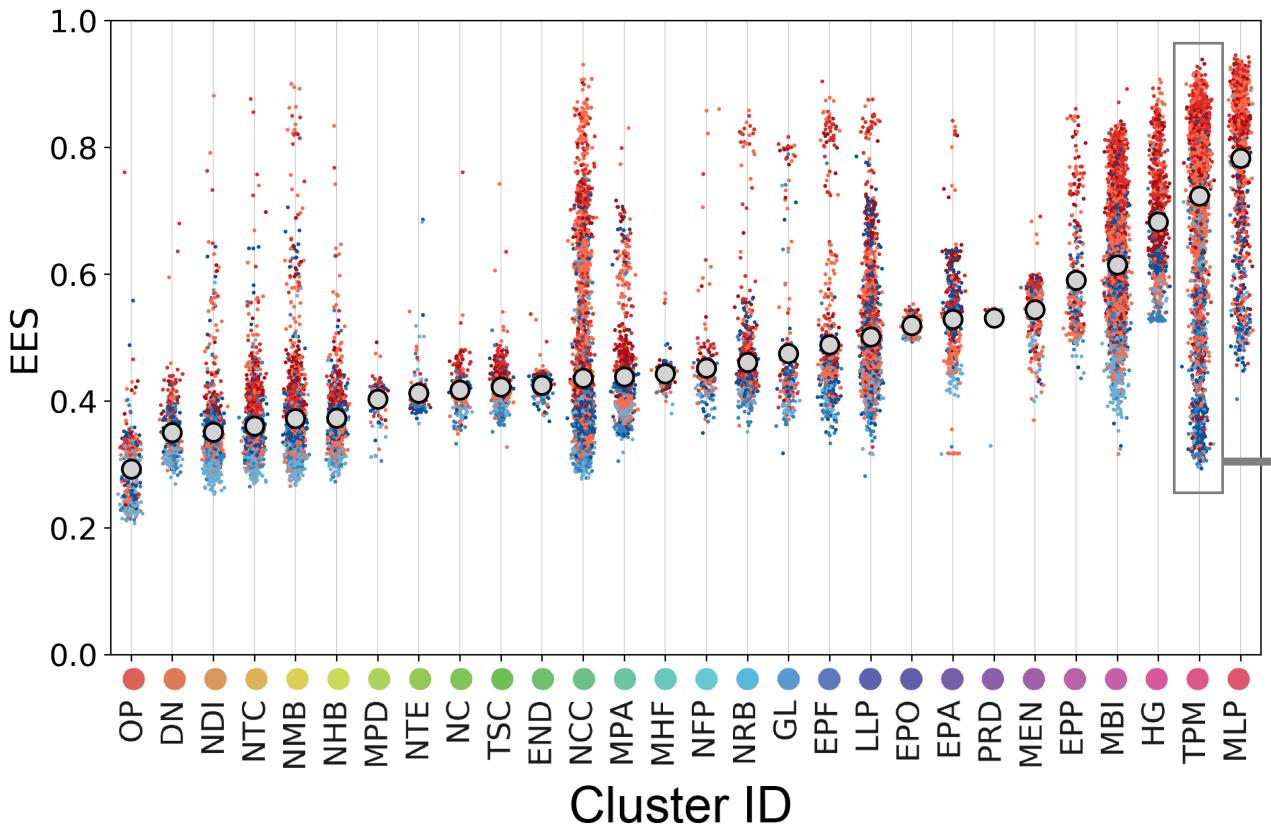
Control

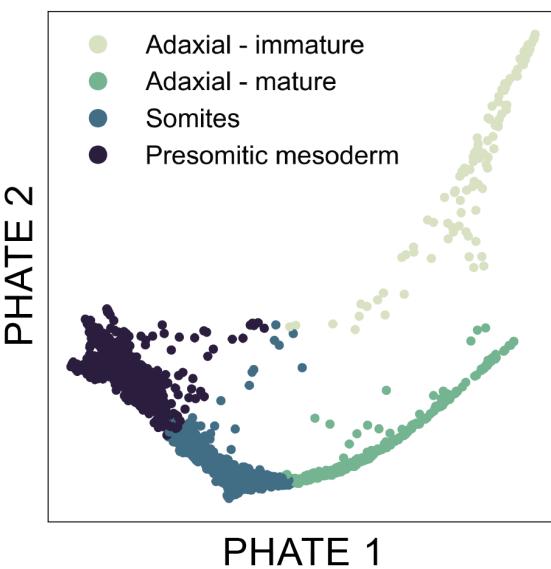
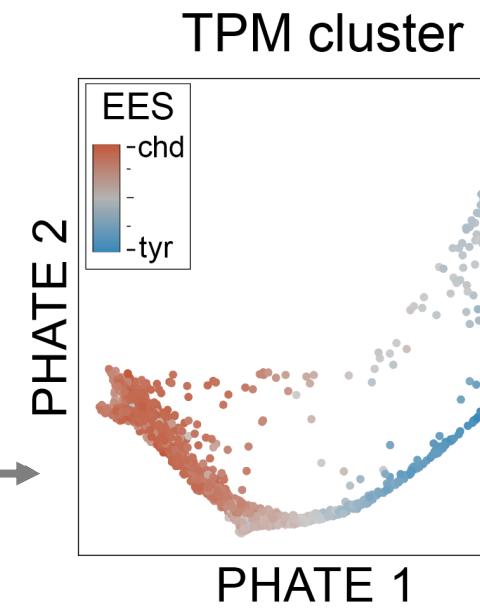
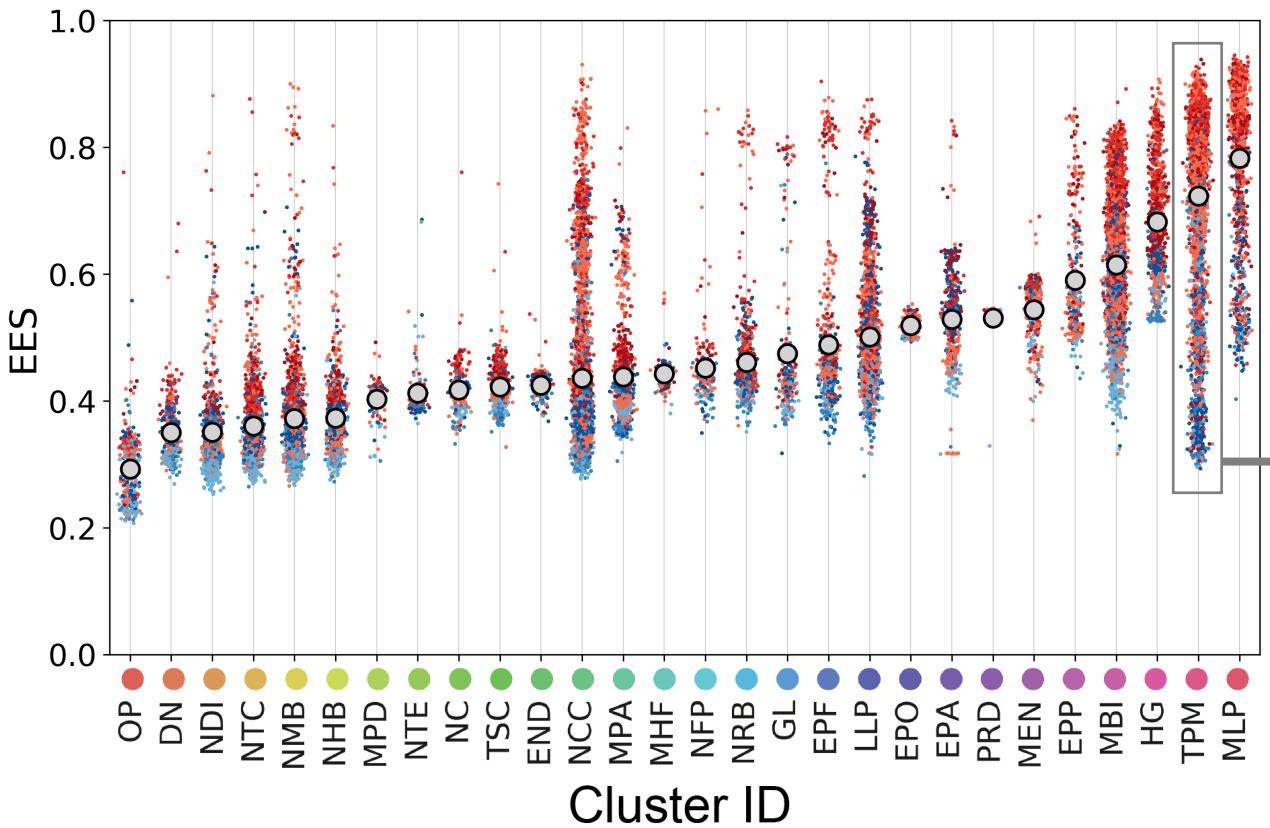












Conclusions

- MELD provides a framework to quantify the effect of an experiment on a per-cell basis
- Vertex Frequency Clustering groups cells by cell state and the effect of the perturbation
- We can apply this method to arbitrary graphs, simulated single cell data, and real experimental data

MELD quantifies the effects of experimental perturbations in single cell datasets

bioRxiv bit.ly/quantsinglecell

 github.com/KrishnaswamyLab/MELD

 @dbburkhardt