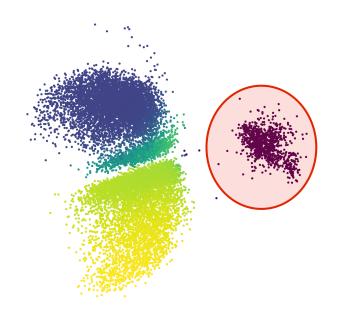
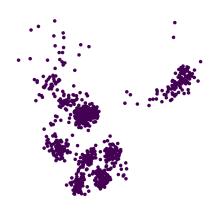
Iterative Latent Spaces for Interpretable High-Dimensional Modeling

Jackson Loper, David Blei, and Liam Paninski Columbia University

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	
S_1	0.00	65.00	701.00	30.00	0.00	1.00	20.00	
S_2	0.00	224.00	628.00	143.00	34.00	0.00	71.00	
S_3	0.00	52.00	326.00	212.00	24.00	0.00	22.00	
S_4	0.00	135.00	428.00	122.00	0.00	0.00	97.00	
S_5	0.00	132.00	555.00	19.00	0.00	0.00	30.00	
S_6	0.00	164.00	557.00	164.00	17.00	0.00	59.00	
S_7	0.00	140.00	520.00	28.00	11.00	0.00	39.00	
S_8	0.00	193.00	661.00	89.00	204.00	0.00	45.00	
S_9	0.00	89.00	691.00	93.00	0.00	0.00	3.00	
S_{10}	0.00	196.00	804.00	189.00	12.00	0.00	33.00	
S_{11}	0.00	92.00	480.00	105.00	8.00	0.00	17.00	
S_{12}	0.00	214.00	371.00	19.00	0.00	0.00	189.00	
S_{13}	0.00	334.00	866.00	69.00	187.00	0.00	103.00	
S_{14}	0.00	41.00	844.00	38.00	0.00	0.00	25.00	
S_{15}	0.00	270.00	306.00	2.00	20.00	0.00	102.00	
		:	:	:	:		:	٠.,







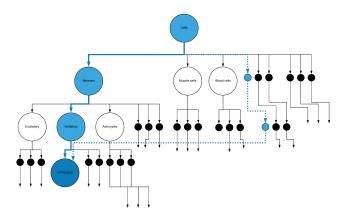
No two-dimensional space can smoothly represent the full diversity of cells.

Latent-variable generative modelling is a natural idea

Especially natural for multi-modality data.

$$\mathbb{P}(X_1, X_2, X_3, \cdots) = \int \left(\left(\prod_i \mathbb{P}(X_i | Z) \right) \mu(dZ) \right)$$

Put them together



Empirical results

Tasic, Bosiljka, Zizhen Yao, Kimberly A. Smith, Lucas Graybuck, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy et al. "Shared and distinct transcriptomic cell types across neocortical areas." bioRxiv (2017): 229542.

Empirical results

- Amortized variational inference
- Zero-inflated negative binomial distribution for gene expression

Empirical results

- Amortized variational inference
- Zero-inflated negative binomial distribution for gene expression

Method	Train	Test
Factor analysis $k = 16$	-4.2	-4.2
Factor analysis $k = 64$	-4.1	-4.3
HNBPF $k=2$	-5.7	-5.7
HNBPF $k = 16$	-3.8	-3.9
HNBPF $k = 64$	-3.8	-5.7
ZINB-NN $k = 2$	-3.6	-3.6
ILS $k = 2 + 2$	-3.5	-3.5
ILS $k = 2 + 2$	-3.5	-3.5

More flexible, but still interpretable

