
Supplementary information

Integrative spatial analysis of cell morphologies and transcriptional states with MUSE

In the format provided by the
authors and unedited

Supplementary Information

Supplementary Table 1 | Parameter settings in simulation experiments.

Supplementary Table 2 | Overexpressed genes in pancreatic tissues or cancer regions for PDAC dataset.

Supplementary Table 3 | Top 50 overexpressed genes (ranked by P value) identified from different brain regions in Alzheimer's disease dataset, related to **Figure 6**.

Supplementary Table 4 | Genes consistently identified from top 50 overexpressed genes across four brain regions and their significance ranks based on p-values in Alzheimer's disease dataset, related to **Figure 6**.

Supplementary Table 5 | Genes uniquely identified from one of four analyzed regions based top 50 overexpressed genes, related to **Figure 6**.

Supplementary Table 6 | Default parameters for MUSE.

Supplementary Table 1 | Parameter settings used in simulation experiments.

Experiment	Purpose	Ground truth cluster	Cluster merge probability	sample size	latent code dimension	Feature dimension		Noise σ		Dropout coefficient	
						Trans.	Morph.	Trans.	Morph.	Trans.	Morph.
Figures 1c, Extended Data Figure 1d	Performance when the ability to discriminate subpopulations in each modality decreases	10	Each single modality had 10, 8, or 6 clusters	1,000	30	500	500	0.1	0.1	0.5	0.1
Figures 1d, e	Performance as data quality in one modality degrades by dropouts	10	0.7	1,000	30	500	500	0.1	0.1	0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5	0.1
Extended Data Figure 1e	Performance as the number of ground-truth subpopulations increases	10, 20, 30, 50	0.7	a) 1,000 b) 3,000	30	500	500	0.1	0.1	0.5	0.1
Extended Data Figure 1f	Performance with Gaussian random noise with increasing variance	10	0.7	1,000	30	500	500	0.1, 0.5, 1, 2	0.1, 0.5, 1, 2	0.5	0.1
Extended Data Figure 1l	Effects from input dimensions	10	0.7	1,000	30	500	100, 300, 500, 1,000	0.1	0.1	0.5	0.1
Extended Data Figures 1g, i-k, and m-u	Effects from downstream analysis or parameters	10	0.7	1,000	30	500	500	0.1	0.1	0.5	0.1

Supplementary Table 2 | Overexpressed genes in pancreatic tissues (**a**) or cancer regions (**b**). P-values (one-sided ranksum test) and fold changes were calculated by comparing two cancer regions vs two pancreatic tissue regions (**Methods**). Genes in red color were overexpressed in PDAC patients with worse prognosis.

a

Gene	P value	Log2FC
<i>Cel</i>	3.37E-22	5.31144436
<i>Ctrb1</i>	1.17E-21	5.07476495
<i>Cela3a</i>	7.67E-21	5.01636881
<i>Clps</i>	1.16E-19	4.77410973
<i>Cela3b</i>	3.59E-19	4.85858436
<i>Ctrc</i>	1.50E-18	4.91412558
<i>Cpa2</i>	2.67E-18	5.10829479
<i>Ctrb2</i>	1.00E-17	4.35512874
<i>Gp2</i>	7.81E-14	4.44267959
<i>Cpa1</i>	8.58E-14	4.84660111

b

Gene	P value	Log2FC
<i>Krt17</i>	2.28E-24	3.06647052
<i>Krt19</i>	6.48E-17	1.7686434
<i>Lamc2</i>	1.99E-16	2.60552734
<i>Apol1</i>	8.06E-14	2.42874519
<i>Ifi27</i>	1.70E-11	2.87460762
<i>Itgb4</i>	1.28E-09	1.91554134
<i>Fxyd3</i>	2.07E-09	1.9109301
<i>Lmna</i>	5.63E-08	1.36949752
<i>Hmga1</i>	1.07E-06	1.14514764
<i>Aebp1</i>	1.37E-06	1.29541453

Supplementary Table 3 | Top 50 overexpressed genes (ranked by P value) identified from different brain regions in the Alzheimer's disease dataset, related to **Figure 6**. TH: thalamus; HY: hypothalamus; HPd: dendritic hippocampus; CTX: isocortex.

Rank	TH	CTX	HY	HPd
0	<i>Ctsd</i>	<i>Ctsd</i>	<i>Ctsd</i>	<i>Ctsd</i>
1	<i>Cst3</i>	<i>Gfap</i>	<i>C4b</i>	<i>C4b</i>
2	<i>Gfap</i>	<i>C4b</i>	<i>Cst3</i>	<i>B2m</i>
3	<i>Tyrobp</i>	<i>Cst3</i>	<i>Gfap</i>	<i>Tyrobp</i>
4	<i>B2m</i>	<i>Cst7</i>	<i>Rn18s-rs5</i>	<i>Gfap</i>
5	<i>C4b</i>	<i>B2m</i>	<i>Trem2</i>	<i>Cst3</i>
6	<i>Cst7</i>	<i>Ctss</i>	<i>Ttr</i>	<i>C1qb</i>
7	<i>Trem2</i>	<i>C1qb</i>	<i>Supt7l</i>	<i>C4a</i>
8	<i>Serpina3n</i>	<i>Tyrobp</i>	<i>B2m</i>	<i>Cst7</i>
9	<i>Apoe</i>	<i>C4a</i>	<i>Tyrobp</i>	<i>C1qa</i>
10	<i>Ctsz</i>	<i>C1qa</i>	<i>Cst7</i>	<i>Trem2</i>
11	<i>C1qa</i>	<i>Lyz2</i>	<i>C4a</i>	<i>Lyz2</i>
12	<i>Ctss</i>	<i>Ly86</i>	<i>Lyz2</i>	<i>Ly86</i>
13	<i>Rn18s-rs5</i>	<i>Trem2</i>	<i>C1qb</i>	<i>C1qc</i>
14	<i>C4a</i>	<i>C1qc</i>	<i>Ly86</i>	<i>Apoe</i>
15	<i>C1qc</i>	<i>Hexb</i>	<i>Apoe</i>	<i>Ctss</i>
16	<i>Hexb</i>	<i>Cd74</i>	<i>C1qa</i>	<i>Serpina3n</i>
17	<i>Ly86</i>	<i>Apoe</i>	<i>Lars2</i>	<i>Ctsz</i>
18	<i>Sez6</i>	<i>Gm13340</i>	<i>Serpina3n</i>	<i>Ctsb</i>
19	<i>Gm20721</i>	<i>Ctsb</i>	<i>C1qc</i>	<i>Cd74</i>
20	<i>C1qb</i>	<i>Serpina3n</i>	<i>Gm42418</i>	<i>RP23-269H21.1</i>
21	<i>Lyz2</i>	<i>RP23-269H21.1</i>	<i>Ctsz</i>	<i>Rn18s-rs5</i>
22	<i>RP23-269H21.1</i>	<i>Rn18s-rs5</i>	<i>Ctss</i>	<i>Hexb</i>
23	<i>Gm13340</i>	<i>Ctsz</i>	<i>Enpp2</i>	<i>Ccl6</i>
24	<i>Cd74</i>	<i>Gm20721</i>	<i>RP23-269H21.1</i>	<i>Hexa</i>
25	<i>Ctsb</i>	<i>Ccl6</i>	<i>Gm10059</i>	<i>H2-D1</i>
26	<i>Clu</i>	<i>Psap</i>	<i>Mt3</i>	<i>Fbxl16</i>
27	<i>Gm12155</i>	<i>AC149090.1</i>	<i>Dnm1</i>	<i>Clu</i>
28	<i>Nme7</i>	<i>Hexa</i>	<i>Gm20721</i>	<i>Trf</i>
29	<i>AC149090.1</i>	<i>Mpeg1</i>	<i>Camk2n1</i>	<i>Gm20721</i>
30	<i>Eef1a2</i>	<i>Sez6</i>	<i>Ctsb</i>	<i>Cox4i1</i>
31	<i>Dnm1</i>	<i>Lgals3bp</i>	<i>Gm9396</i>	<i>Gm11410</i>
32	<i>Supt7l</i>	<i>Gm8326</i>	<i>Eef1a2</i>	<i>Gm7618</i>

33	<i>Scn1b</i>	<i>Cox4i1</i>	<i>Ccl6</i>	<i>AC149090.1</i>
34	<i>Vgf</i>	<i>Gm7618</i>	<i>Trf</i>	<i>Mt3</i>
35	<i>Gm10925</i>	<i>Gm11410</i>	<i>Vgf</i>	<i>Gm8326</i>
36	<i>Ids</i>	<i>Clu</i>	<i>Gm5575</i>	<i>Mpeg1</i>
37	<i>Calr</i>	<i>H2-D1</i>	<i>Hexb</i>	<i>Purb</i>
38	<i>Camk2n1</i>	<i>Gm12155</i>	<i>1500015O10Rik</i>	<i>Gm42418</i>
39	<i>Psap</i>	<i>Mt3</i>	<i>Prkcd</i>	<i>Laptm5</i>
40	<i>Penk</i>	<i>Malat1</i>	<i>Podxl2</i>	<i>Psap</i>
41	<i>Gm7618</i>	<i>Fth1</i>	<i>Cd74</i>	<i>Sez6</i>
42	<i>Gm42418</i>	<i>Laptm5</i>	<i>Zfp365</i>	<i>Cd9</i>
43	<i>Spp1</i>	<i>Dnm1</i>	<i>Fam212a</i>	<i>Lars2</i>
44	<i>Gm29216</i>	<i>Fcer1g</i>	<i>H2-D1</i>	<i>Dnm1</i>
45	<i>Serinc1</i>	<i>Gm26723</i>	<i>Rasgrp1</i>	<i>Pkm</i>
46	<i>Dlgap3</i>	<i>Cd9</i>	<i>Gm10925</i>	<i>Fth1</i>
47	<i>Pvalb</i>	<i>Pkm</i>	<i>Serinc1</i>	<i>Lgals3bp</i>
48	<i>Pmch</i>	<i>Clec7a</i>	<i>Pcdh10</i>	<i>Camk2n1</i>
49	<i>Grina</i>	<i>Klf9</i>	<i>Mir6236</i>	<i>Eef1a2</i>

Supplementary Table 4 | Genes consistently identified from top 50 overexpressed genes across four brain regions and their significance ranks based on p-values in the Alzheimer's disease dataset, related to **Figure**

6. TH: thalamus; HY: hypothalamus; HPd: dendritic hippocampus; CTX: isocortex.

Gene name	TH	CTX	HY	HPd
<i>Ctsd</i>	0	0	0	0
<i>C4b</i>	5	2	1	1
<i>Gfap</i>	2	1	3	4
<i>Cst3</i>	1	3	2	5
<i>B2m</i>	4	5	8	2
<i>Tyrobp</i>	3	8	9	3
<i>Cst7</i>	6	4	10	8
<i>Trem2</i>	7	13	5	10
<i>C4a</i>	14	9	11	7
<i>C1qb</i>	20	7	13	6
<i>C1qa</i>	11	10	16	9
<i>ApoE</i>	9	17	15	14
<i>Ctss</i>	12	6	22	15
<i>Ly86</i>	17	12	14	12
<i>Lyz2</i>	21	11	12	11
<i>Rn18s-rs5</i>	13	22	4	21
<i>C1qc</i>	15	14	19	13
<i>Serpina3n</i>	8	20	18	16
<i>Ctsz</i>	10	23	21	17
<i>RP23-269H21.1</i>	22	21	24	20
<i>Hexb</i>	16	15	37	22
<i>Ctsb</i>	25	19	30	18
<i>Gm20721</i>	19	24	28	29
<i>Cd74</i>	24	16	41	19
<i>Dnm1</i>	31	43	27	44

Supplementary Table 5 | Genes uniquely identified from one of four analyzed regions based top 50 overexpressed genes. TH: thalamus; HY: hypothalamus; HPd: dendritic hippocampus; CTX: isocortex.

Brain region	Unique genes in one region from top 50 DE genes
TH	<i>Nme7, Scn1b, Ids, Calr, Penk, Spp1, Gm29216, Dlgap3, Pvalb, Pmch, Grina</i>
CTX	<i>Malat1, Fcer1g, Gm26723, Clec7a, Klf9</i>
HY	<i>Ttr, Enpp2, Gm10059, Gm9396, Gm5575, 1500015O10Rik, Prkcd, Podxl2, Zfp365, Fam212a, Rasgrp1, Pcdh10, Mir6236</i>
HPd	<i>Fbxl16, Purb</i>

Supplementary Table 6 | Default parameters used in MUSE.

Hyperparameter	Value
Latent dimension	100
Learning rate	1.00E-04
Training epoch	500
Initialization epoch	200
Cluster update interval	200
$\lambda_{\text{supervise}}$	5
$\lambda_{\text{regularization}}$	5