Supporting Information for "A Bayesian Multivariate Mixture Model for High Throughput Spatial Transcriptomics" by

Carter Allen[1,4], Yuzhou Chang[1,4], Brian Neelon[2], Won Chang[3], Hang J. Kim[3], Zihai Li[4], Qin Ma[1,4], and Dongjun Chung[1,4,*]

[1] Department of Biomedical Informatics, The Ohio State University, Columbus, OH, U.S.A.
[2] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, U.S.A.
[3] Division of Statistics and Data Science, University of Cincinnati, Cincinnati, OH, U.S.A.
[4] The Pelotonia Institute for Immuno-Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH, U.S.A.

*email: chung.911@osu.edu

Web Appendix A: Proof of Proposition 1

Proposition 1 *Let π_{ik} follow the multinomial logit model defined in equation (6) of the manuscript, and let ψ_{ik} have a univariate intrinsic CAR prior as defined in equation (7) of the manuscript. Under Pólya–Gamma data augmentation, the full conditional distribution of ψ_{ik} is $N(m_{ik}, V_{ik})$, where*

$$m_{ik} = \frac{\frac{1}{m_i}\sum_{l\in\delta_i}\psi_{lk} + U^*_{ik}}{\frac{m_i^2}{\nu_k^2} + \frac{1}{\omega_{ik}^2}}, \text{ and } V_{ik} = \frac{1}{\frac{m_i^2}{\nu_k^2} + \frac{1}{\omega_{ik}^2}}, \tag{1}$$

where $U^*_{ik} = \frac{U_{ik}-1/2}{\omega_{ik}} + c_{ik} - \mathbf{w}_i^T\boldsymbol{\rho}_k$, U_{ik} is an indicator equal to 1 if $z_i = k$ and 0 otherwise, $c_{ik} = log\{\sum_{h\neq k}^K exp(\mathbf{w}_i^T\boldsymbol{\rho}_h + \psi_{ih})\}$, and $\omega_{ik} \sim PG(1,0)$.

Proof. *The full conditional distribution of ψ_{ik}, denoted $p(\psi_{ik}|\cdots)$, may be expressed as*

$$\begin{aligned} p(\psi_{ik}|\cdots) &= p(\psi_{ik}|\psi_{-ik}, \mathbf{z}, ...) = \frac{p(\psi_{ik}, \psi_{-ik}, \mathbf{z}, ...)}{p(\psi_{-ik}, \mathbf{z}, ...)} \\ &\propto \underbrace{p(\psi_{ik}|\psi_{-ik})}_{CAR\ prior}\underbrace{p(\mathbf{z}|\psi_{ik}, \psi_{-ik}, ...)}_{likelihood} \\ &\propto \left[N\left(\frac{1}{m_i}\sum_{l\in\delta_i}\psi_{lk}, \frac{\nu_k^2}{m_i}\right)\right]\prod_{i=1}^n\prod_{k=1}^K \pi_{ik}^{U_{ik}}, \end{aligned}$$

where U_{ik} is an indicator equal to 1 if $z_i = k$ and 0 otherwise. Given $\mathbf{U}_k = (U_{1k}, ..., U_{nk})^T$, we may re-parameterize the model for π_{ik} as

$$\begin{aligned} \pi_{ik} &= P(U_{ik} = 1) \\ &= \frac{exp(\mathbf{w}_i^T\boldsymbol{\rho}_k + \psi_{ik})}{\sum_{h=1}^K exp(\mathbf{w}_i^T\boldsymbol{\rho}_h + \psi_{ih})} \\ &= \frac{exp(\mathbf{w}_i^T\boldsymbol{\rho}_k + \psi_{ik})}{\sum_{h\neq k}^K exp(\mathbf{w}_i^T\boldsymbol{\rho}_h + \psi_{ih}) + exp(\mathbf{w}_i^T\boldsymbol{\rho}_k + \psi_{ik})}. \end{aligned}$$

Dividing through by $\sum_{h\neq k}^K exp(\mathbf{w}_i^T\boldsymbol{\rho}_h + \psi_{ih})$ gives

$$\pi_{ik} = \frac{exp(\mathbf{w}_i^T\boldsymbol{\rho}_k + \psi_{ik} - c_{ik})}{1 + exp(\mathbf{w}_i^T\boldsymbol{\rho}_k + \psi_{ik} - c_{ik})} = \frac{exp(\gamma_{ik})}{1 + exp(\gamma_{ik})},$$

where $c_{ik} = log\{\sum_{h \neq k}^{K} exp(\mathbf{w}_i^T \boldsymbol{\rho}_h + \psi_{ih})\}$ *and* $\gamma_{ik} = \mathbf{w}_i^T \boldsymbol{\rho}_k + \psi_{ik} - c_{ik}$. *Now, we notice that with respect to* ψ_{ik}, *the likelihood contribution may be simplified as follows.*

$$\begin{aligned}
p(\psi_{ik}|\cdots) &\propto p(\psi_{ik}|\psi_{-ik}) \left(\frac{e^{\gamma_{ik}}}{1+e^{\gamma_{ik}}}\right)^{U_{ik}} \left(\frac{1}{1+e^{\gamma_{ik}}}\right)^{1-U_{ik}} \\
&= p(\psi_{ik}|\psi_{-ik}) \left\{ \frac{\left(\frac{e^{\gamma_{ik}}}{1+e^{\gamma_{ik}}}\right)^{U_{ik}} \left(\frac{1}{1+e^{\gamma_{ik}}}\right)}{\left(\frac{1}{1+e^{\gamma_{ik}}}\right)^{U_{ik}}} \right\} \\
&= p(\psi_{ik}|\psi_{-ik}) \left\{ \left(\frac{\frac{e^{\gamma_{ik}}}{1+e^{\gamma_{ik}}}}{\frac{1}{1+e^{\gamma_{ik}}}}\right)^{U_{ik}} \left(\frac{1}{1+e^{\gamma_{ik}}}\right) \right\} \\
&= p(\psi_{ik}|\psi_{-ik}) \frac{\left(e^{\gamma_{ik}}\right)^{U_{ik}}}{1+e^{\gamma_{ik}}}.
\end{aligned}$$

The likelihood is now in the logistic form, which Polson et al. (2013) showed can be written as a scale mixture of normals with Pólya–Gamma precision terms $\omega_{ik} \sim PG(1,0)$. *Thus, we have*

$$\begin{aligned}
p(\psi_{ik}|\cdots) &\propto p(\psi_{ik}|\psi_{-ik}) \left\{ e^{(U_{ik}-1/2)\gamma_{ik}} \int_0^\infty e^{-\frac{\omega_{ik}\gamma_{ik}^2}{2}} p(\omega_{ik}) d\omega_{ik} \right\} \\
&= p(\psi_{ik}|\psi_{-ik}) exp\left\{(U_{ik}-1/2)\gamma_{ik} - \omega_{ik}\gamma_{ik}^2/2\right\} \\
&\propto p(\psi_{ik}|\psi_{-ik}) exp\left\{ -\frac{1}{2}\left(\frac{(U_{ik}^* - \psi_{ik})^2}{\omega_{ik}^2}\right)\right\},
\end{aligned}$$

where $U_{ik}^* = \frac{U_{ik}-1/2}{\omega_{ik}} + c_{ik} - \mathbf{w}_i^T \boldsymbol{\rho}_k$. *Following standard results from Bayesian normal models (Hoff, 2009), we find the full conditional of* ψ_{ik} *is* $N(m_{ik}, V_{ik})$, *as defined in equation (1).*

Web Appendix B: MCMC Algorithm

1. *Update multivariate skew-normal outcome model parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\xi}_k$, and $\boldsymbol{\Sigma}_k$.*

   For $k = 1, ..., K$:

   (a) Update $\boldsymbol{\mu}_k$:
   - i. Define $n_k = \sum_{i=1}^{n} I_{z_i=k}$ as the number of spots in mixture component k.
   - ii. Define $\mathcal{Z}_k$ as the set of all spot indices assigned to mixture component k.
   - iii. Define $\mathbf{Y}_k$ as the $n_k \times g$ matrix of gene expression values for mixture component k. Similarly define $\mathbf{t}_k$ as the length n_k vector of truncated normal random effects for mixture component k and define $\boldsymbol{\Phi}_k$ as the $n_k \times g$ matrix with rows $\boldsymbol{\phi}_1^T, ..., \boldsymbol{\phi}_{n_k}^T$.
   - iv. Define $\mathbf{E}_k = (\mathbf{Y}_k - \boldsymbol{\Phi}_k - \mathbf{t}_k^T \boldsymbol{\xi}_k)$ and $\bar{\mathbf{e}}_k$ as the column means of $\mathbf{E}_k$.
   - v. Update $\boldsymbol{\mu}_k$ from its $\mathrm{N}_g(\boldsymbol{\mu}_{nk}, \mathbf{V}_{nk})$ full conditional, where $\boldsymbol{\mu}_{nk} = \mathbf{V}_{nk}(\mathbf{V}_{0k}^{-1}\boldsymbol{\mu}_{0k} + n_k \boldsymbol{\Sigma}_k^{-1}\bar{\mathbf{e}}_k)$ and $\mathbf{V}_{nk} = (\mathbf{V}_{0k}^{-1} + n_k \boldsymbol{\Sigma}_k^{-1})^{-1}$.

   (b) Update $\boldsymbol{\xi}_k$:
   - i. Define $\mathbf{M}_k$ as the $n_k \times g$ matrix with each row equal to $\boldsymbol{\mu}_k^T$.
   - ii. Re-define $\mathbf{E}_k = \mathbf{t}_k \circ (\mathbf{Y}_k - \mathbf{M}_k - \boldsymbol{\Phi}_k)$, where "$\circ$" denotes the Hadamard product, and let $\bar{\mathbf{e}}_k$ be the $g \times 1$ vector of columns means of $\mathbf{E}_k$.
   - iii. Update $\boldsymbol{\xi}_k$ from its $\mathrm{N}_g(\boldsymbol{\xi}_{nk}, \mathbf{X}_{nk})$ full conditional, where $\boldsymbol{\xi}_{nk} = \mathbf{X}_{nk}(\mathbf{X}_{0k}^{-1}\boldsymbol{\xi}_{0k} + n_k \boldsymbol{\Sigma}_k^{-1}\bar{\mathbf{e}}_k)$ and $\mathbf{X}_{nk} = \left\{\mathbf{X}_{0k}^{-1} + (\sum_{i \in \mathcal{Z}_k} t_i^2)\boldsymbol{\Sigma}_k^{-1}\right\}^{-1}$.

   (c) Update $\boldsymbol{\Sigma}_k$:
   - i. Re-define $\mathbf{E}_k = (\mathbf{Y}_k - \mathbf{M}_k - \boldsymbol{\Phi}_k - \mathbf{t}_k^T \boldsymbol{\xi}_k)$.
   - ii. Update $\boldsymbol{\Sigma}_k$ from its $\mathrm{IW}(\nu_{nk}, \mathbf{S}_{nk})$ full conditional, where $\nu_{nk} = \nu_{0k} + n_k$ and $\mathbf{S}_{nk} = \mathbf{S}_{0k} + \mathbf{E}_k^T \mathbf{E}_k$.

   (d) (optional) Back-transform to original MSN parameterization:

$$\begin{aligned} \boldsymbol{\eta}_{ik} &= \boldsymbol{\mu}_k + \boldsymbol{\psi}_i, \\ \boldsymbol{\Omega}_k &= \boldsymbol{\Sigma}_k + \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T, \\ \boldsymbol{\alpha}_k &= \frac{1}{\sqrt{1 - \boldsymbol{\xi}_k^T \boldsymbol{\Omega}_k^{-1} \boldsymbol{\xi}_k}} \mathrm{Diag}(\boldsymbol{\Omega}_k)^{1/2} \boldsymbol{\Omega}_k^{-1} \boldsymbol{\xi}_k. \end{aligned}$$

2. *Update multivariate skew-normal conditional representation random effects t_i.*

   For $i = 1, ..., n$:

   (a) Given $z_i = k$, define $A_k = (1 + \boldsymbol{\xi}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\xi}_k)^{-1}$ and define $a_{ik} = A_k(\boldsymbol{\xi}_k^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k - \boldsymbol{\phi}_i))$.

   (b) Update t_i from $\mathrm{N}_{[0,\infty)}(a_{ik}, \sqrt{A_k})$.

3. *Update outcome model multivariate CAR random effects $\boldsymbol{\phi}_i$.*

   For $i = 1, ..., n$:

   (a) Given $z_i = k$, update $\boldsymbol{\phi}_i$ from its $g-$dimensional normal full conditional, where $\mathrm{E}(\boldsymbol{\phi}_i|...) = (\boldsymbol{\Sigma}_k^{-1} + m_i \boldsymbol{\Lambda})^{-1}(\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k - t_i \boldsymbol{\xi}_k) + \boldsymbol{\Lambda}^{-1} \sum_{l \in \delta_i} \boldsymbol{\phi}_l)$ and $\mathrm{Cov}(\boldsymbol{\phi}_i|...) = (\boldsymbol{\Sigma}_k^{-1} + m_i \boldsymbol{\Lambda})^{-1}$.

4. *Update multinomial regression CAR random effects* (See Proposition 1).

   For $i = 1, ..., n$ and $k = 2, ..., K$, the full conditional distribution of ψ_{ik} is $\mathrm{N}(m_{ik}, V_{ik})$, where

$$m_{ik} = \frac{\frac{1}{m_i} \sum_{l \in \delta_i} \psi_{lk} + U_{ik}^*}{\frac{m_i^2}{\nu_k^2} + \frac{1}{\omega_{ik}^2}}, \text{ and } V_{ik} = \frac{1}{\frac{m_i^2}{\nu_k^2} + \frac{1}{\omega_{ik}^2}}, \tag{2}$$

   where $U_{ik}^* = \frac{U_{ik} - 1/2}{\omega_{ik}} + c_{ik} - \mathbf{w}_i^T \boldsymbol{\rho}_k$, U_{ik} is an indicator equal to 1 if $z_i = k$ and 0 otherwise, $c_{ik} = \log(\sum_{h \neq k}^{K} \exp(\mathbf{w}_i^T \boldsymbol{\rho}_h + \psi_{ih}))$, and $\omega_{ik} \sim \mathrm{PG}(1, 0)$.

5. *Update multinomial regression mixing weight parameters $\boldsymbol{\rho}_k$ and latent variables ω_{ik}.*

   For $k = 2, ..., K$:

   (a) For $i = 1, ..., n$, update PG latent variables ω_{ik} from $\text{PG}(1, \eta_{ik})$, where $\eta_{ik} = \mathbf{w}_i^T \boldsymbol{\rho}_k + \psi_{ik} - c_{ik}$.

   (b) Compute $\mathbf{R}_{nk} = (\mathbf{R}_{0k}^{-1} + \mathbf{W}^T \mathbf{O}_k \mathbf{W})$, where $\mathbf{O}_k$ is the diagonal matrix with entries $(\omega_{1k}, ..., \omega_{nk})$ and $\mathbf{W}$ is the $n \times p$ matrix of covariates with rows $\mathbf{w}_1^T, ..., \mathbf{w}_n^T$.

   (c) Compute $\boldsymbol{\rho}_{nk} = \mathbf{R}_{nk}(\mathbf{R}_{0k}^{-1}\boldsymbol{\rho}_{0k} + \mathbf{W}^T \mathbf{O}_k \mathbf{U}_k^*)$, where $\mathbf{U}_k^* = \left(\frac{U_{1k} - 1/2}{\omega_{1k}} + c_{1k}, ..., \frac{U_{nk} - 1/2}{\omega_{nk}} + c_{nk}\right)$.

   (d) Update $\boldsymbol{\rho}_k$ from $\text{N}_p(\boldsymbol{\rho}_{nk}, \mathbf{R}_{nk})$.

6. *Update mixture component labels $z_1, ..., z_n$.*

   For $i = 1, ..., n$:

   (a) Compute the probability of spot i belonging to cluster k under current values of model parameters. For $k = 1, ..., K$, compute $P_{ik} = \text{dnorm}(\mathbf{y}_i; \boldsymbol{\mu}_k + \boldsymbol{\phi}_k + t_i \boldsymbol{\xi}_k, \boldsymbol{\Sigma}_k)$.

   (b) Compute $\pi_{ik} = \frac{\exp(\mathbf{w}_i^T \boldsymbol{\rho}_k + \psi_{ik})}{\sum_{h=1}^K \exp(\mathbf{w}_i^T \boldsymbol{\rho}_h + \psi_{ih})}$.

   (c) Compute $\text{P}(z_i = k|...) = \frac{P_{ik}\pi_{ik}}{\sum_{h=1}^K P_{ih}\pi_{ih}}$.

   (d) Update z_i from $\text{Categorical}\{\text{P}(z_i = 1|...), ..., \text{P}(z_i = K|...)\}$.

7. *Re-map mixture component labels to protect against label switching.*

   (a) Define $\text{ord}(\mathbf{z})$ as the function to return a length K vector containing the order in which each unique mixture component label appears in $\mathbf{z}$. E.g., in `R` this is the `unique()` function.

   (b) Initialize $\mathbf{z}^*$ as an empty length n vector for the re-mapped mixture component labels.

   (c) For $k = 1, ..., K$, let $\mathbf{z}^*[\mathbf{z} = \text{ord}(\mathbf{z})[k]] = k$.

8. Update $\boldsymbol{\Lambda}$ from its $\text{IW}(\lambda_n, \mathbf{D}_n)$ full conditional, were $\lambda_n = \lambda_0 + n$ and $\mathbf{D}_n = \mathbf{D}_0 + \boldsymbol{\Phi}^T(\mathbf{M} - \mathbf{A})\boldsymbol{\Phi}$, where $\boldsymbol{\Phi}$ as the $n \times g$ matrix with rows $\boldsymbol{\phi}_1^T, ..., \boldsymbol{\phi}_n^T$, $\mathbf{M}$ is an $n \times n$ matrix with diagonal elements $m_1, ..., m_n$ and all other elements 0, and $\mathbf{A}$ is the $n \times n$ adjacenty matrix with elements $a_{ij} = 1$ if cell spots i and j are neighbors and 0 otherwise.

Web Appendix C: Supplementary Figures

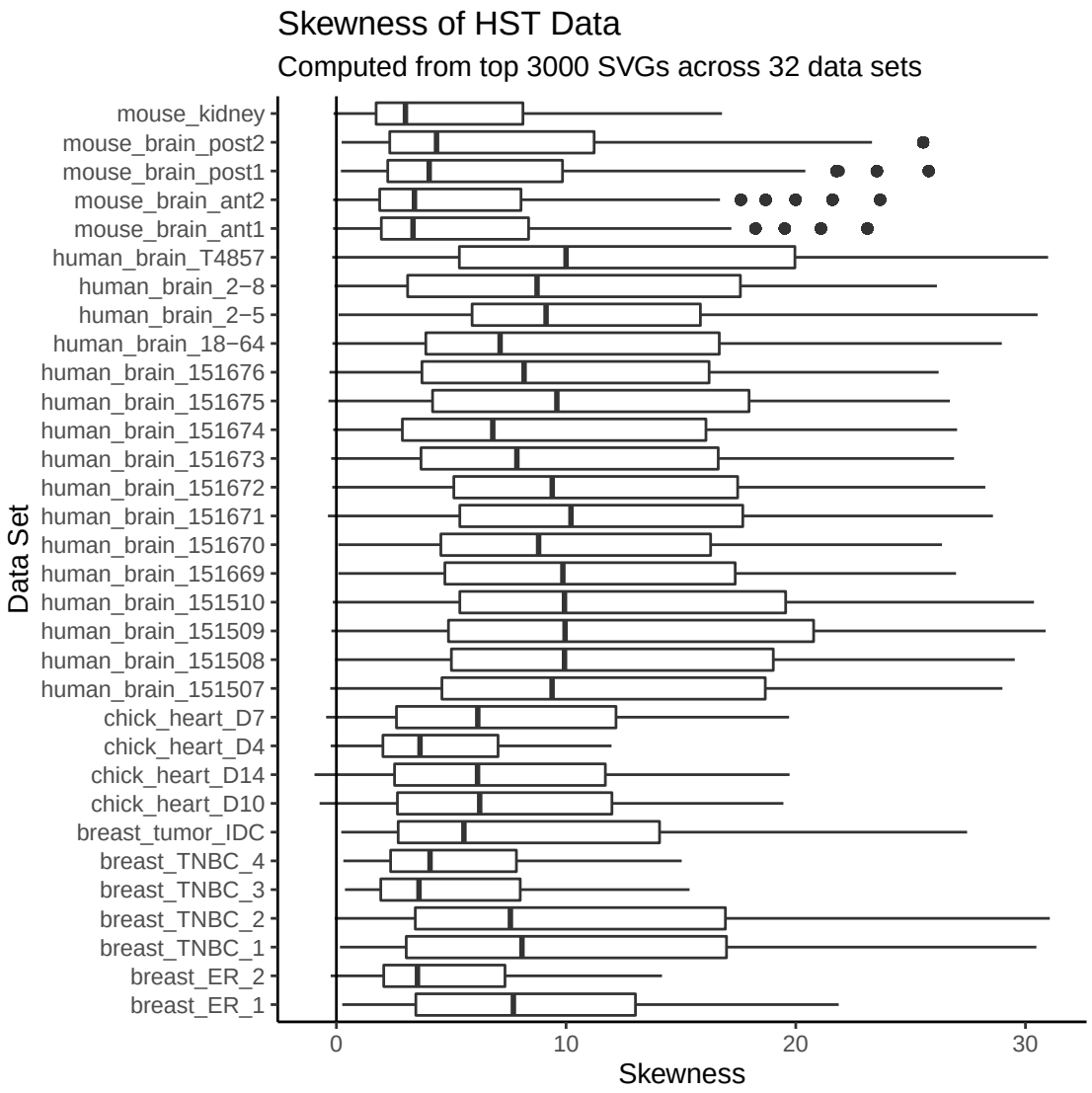

Figure S1: Skewness of top 3000 spatially variable genes across 32 publicly available HST data sets. The skewness coefficient $s(X)$ is defined as $s(X) = m_3/m_2^{3/2}$, where m_2 and m_3 are the 2nd and 3rd sample moments of X, respectively.

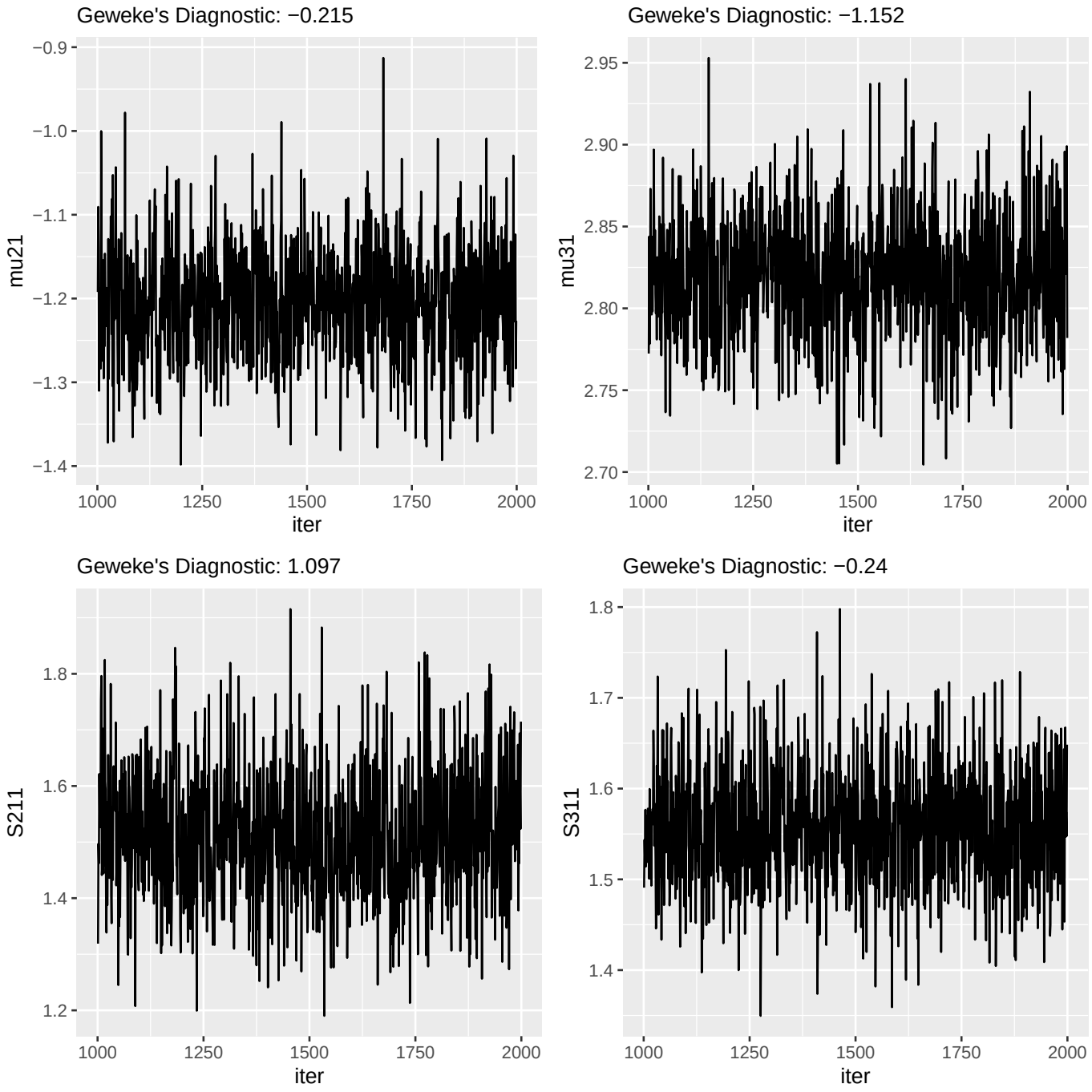

Figure S2: Trace plots for a selection of mean and variance parameters for the simulated mouse brain data analyzed in Section 4 of the main manuscript. Geweke's diagnostic z-score statistic is shown to assess model convergence.

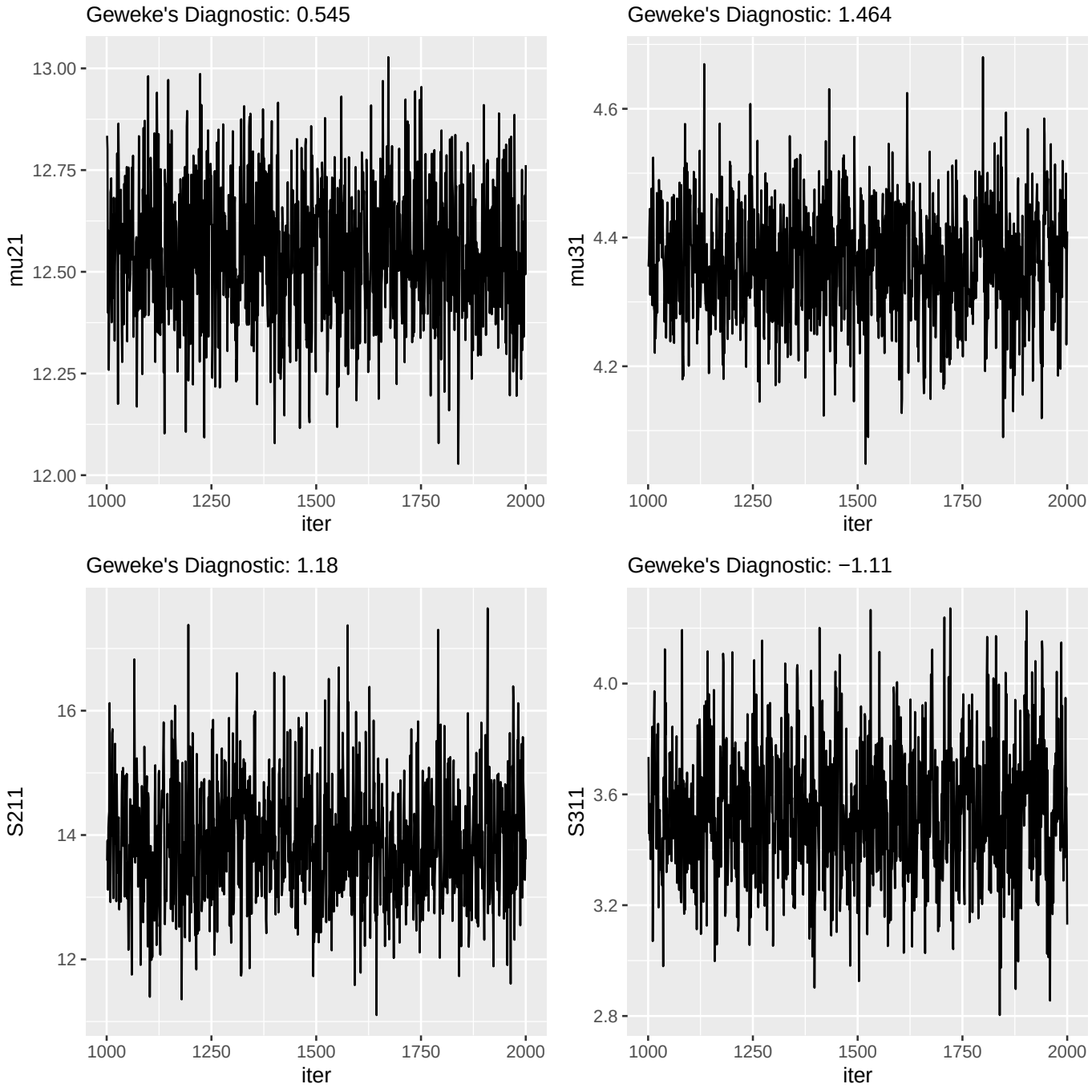

Figure S3: Trace plots for a selection of mean and variance parameters for the real breast cancer data analyzed in Section 5.2 of the main manuscript. Geweke's diagnostic z-score statistic is shown to assess model convergence.

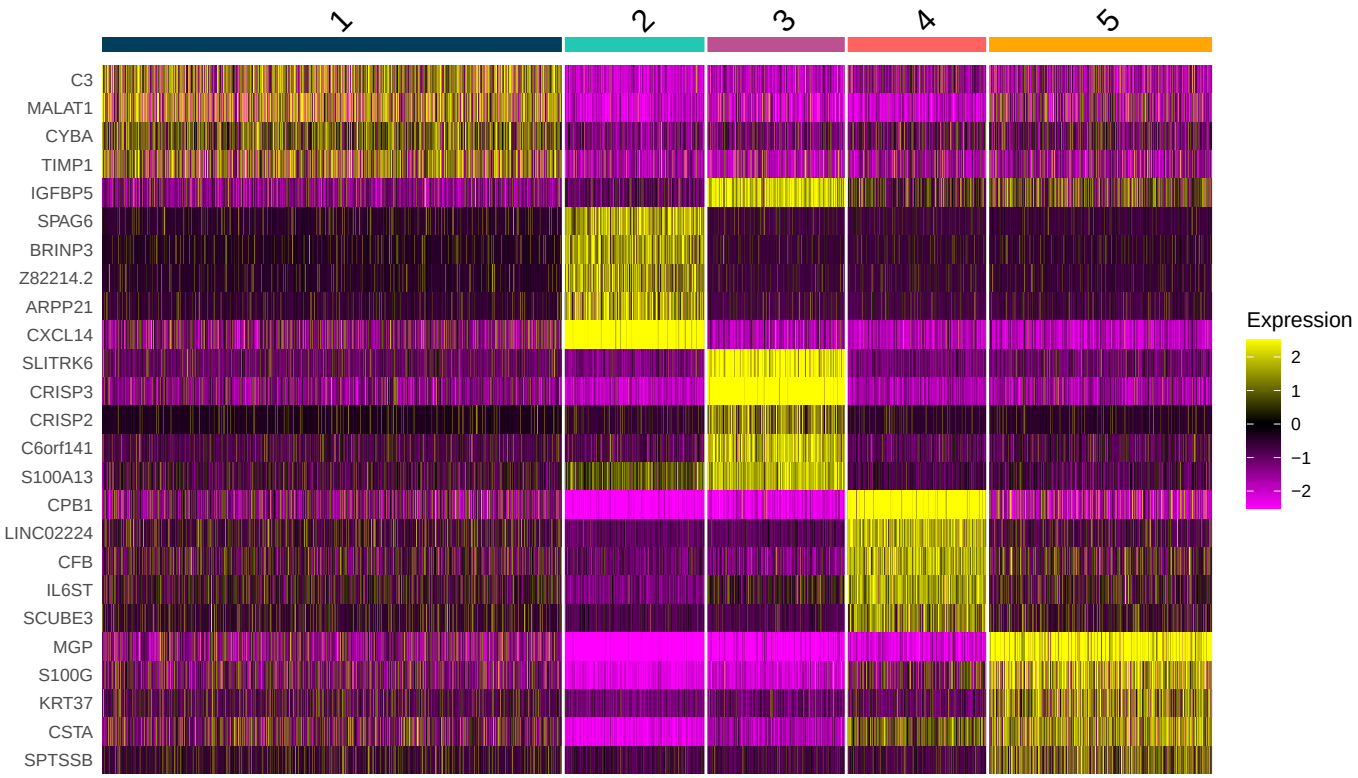

Figure S4: Top 5 most differentially expressed marker genes in each of the five sub-populations inferred from the invasive ductal carcinoma analysis in Section 5.2. Each color band in the colored heatmap represents one cell spot, and normalized gene expression intensities in each cell spot are shown with color, with bright yellow indicating higher expression and dark purple indicating lower expression. Marker genes were identified using adjusted p-values obtained from the Wilcoxon Rank Sum test of each gene in each sub-population vs. the same gene in all other sub-populations.

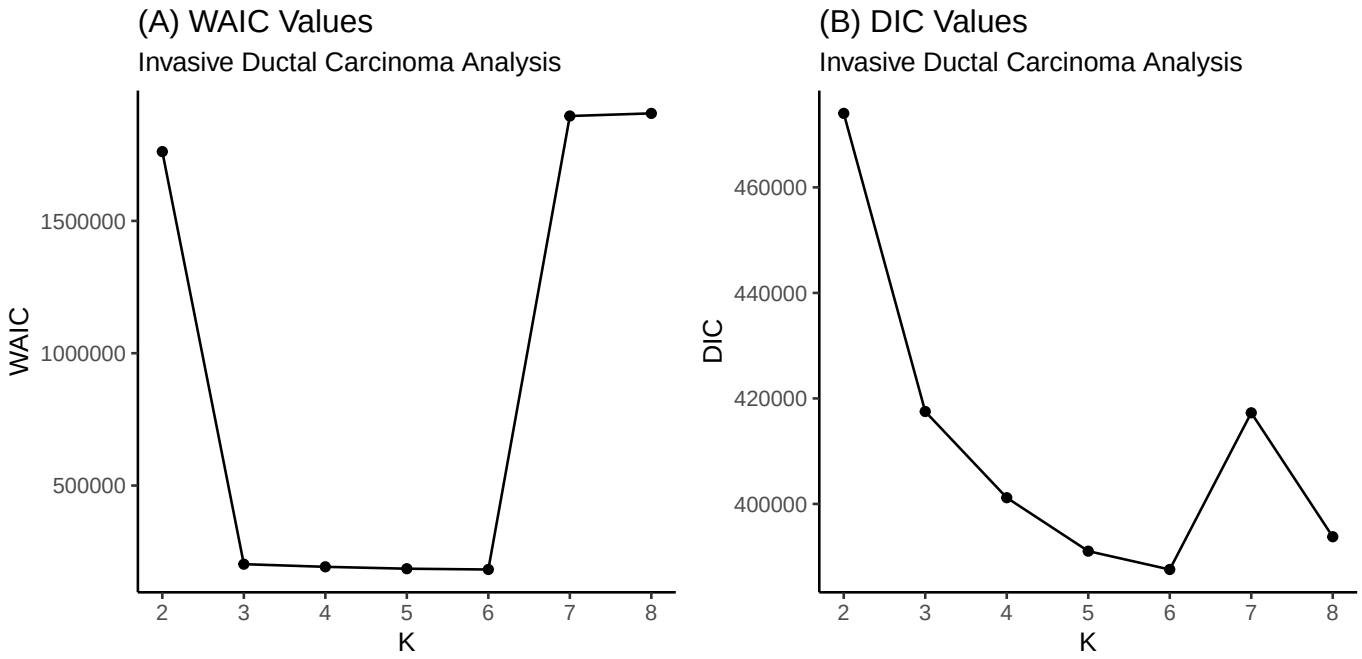

Figure S5: (A) WAIC model fit criterion across a range of K for invasive ductal carcinoma analysis. (B) DIC_3 model fit criterion.

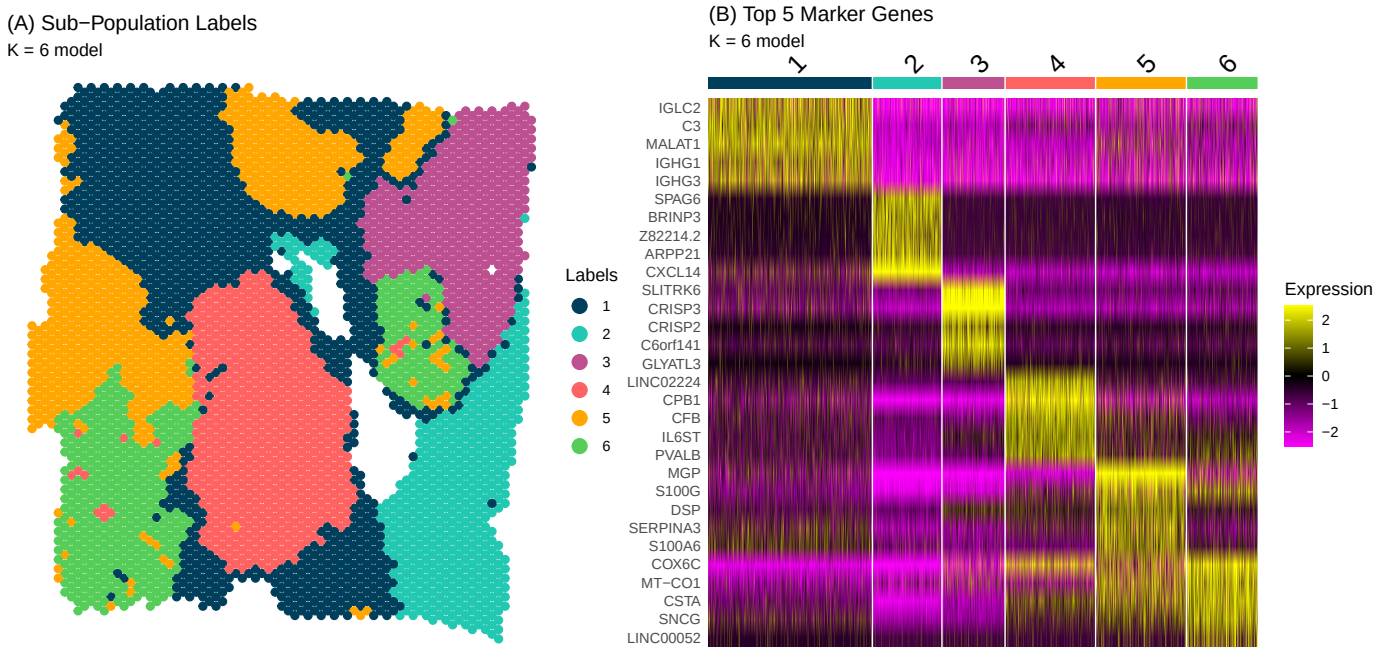

Figure S6: $K = 6$ results for invasive ductal carcinoma analysis. (A) Sub-population labels from $K = 6$ model. (B) Top 5 markers for each sub-population from $K = 6$ model.

Web Appendix D: Supplementary Tables

Table S1: Simulated parameter values and estimates obtained from three model variants: (i) MVN: multivariate normal clustering without spatial random effects; (ii) MVN Spatial: multivariate normal clustering with CAR spatial random effects; and (iii) MSN Spatial: multivariate skew-normal clustering with CAR spatial random effects. Parameter estimates are shown as posterior means with associated 95 % credible intervals.

Parameter	True	MVN	MSN	MSN Spatial
μ_{11}	-2.00	-2.72 (-2.86, 1.77)	-2.14 (-2.74, 1.62)	-2.01 (-2.86, -1.55)
μ_{12}	-1.00	-2.34 (-2.45, 1.01)	-0.83 (-1.27, -0.79)	-0.87 (-1.04, -0.63)
μ_{13}	1.00	-0.97 (-1.07, 0.64)	1.32 (0.85, 1.53)	1.03 (0.82, 1.24)
μ_{14}	2.00	1.79 (1.66, 2.89)	2.12 (1.95,2.36)	2.03 (1.84, 2.22)
ξ_{11}	-1.50	-	-0.95 (-1.59, -0.28)	-1.64 (-1.82, -1.46)
ξ_{12}	-0.75	-	-0.41 (-0.79, 0.09)	-0.78 (-1.14, -0.51)
ξ_{13}	0.75	-	0.63 (0.36, 0.89)	0.75 (0.43, 0.95)
ξ_{14}	1.50	-	1.27 (0.80, 1.52)	1.43 (0.8, 1.74)
Σ_{111}	1.50	2.01 (1.78, 2.68)	1.45 (0.92, 2.21)	1.57 (1.42, 1.74)
Σ_{112}	1.00	1.37 (1.17, 2.13)	1.11 (0.93, 1.75)	1.13 (0.89, 1.25)
Σ_{113}	0.75	0.60 (0.42, 1.77)	0.75 (0.50, 1.45)	0.78 (0.64, 0.95)
Σ_{114}	0.50	0.32 (-0.24, 1.22)	0.42 (0.25, 1.11)	0.49 (0.39, 0.60)
Σ_{122}	1.50	1.37 (1.17, 2.13)	1.59 (0.39, 0.90)	1.61 (1.41, 1.72)
Σ_{123}	1.00	0.79 (0.63, 1.74)	1.05 (0.68, 1.39)	0.88 (0.64, 1.05)
Σ_{124}	0.75	0.32 (0.11, 1.26)	0.82 (0.49, 0.99)	0.71 (0.54, 0.99)
Σ_{133}	1.50	1.71 (1.54, 2.13)	1.41 (1.01, 1.65)	1.52 (1.24, 1.83)
Σ_{134}	1.00	0.32 (0.11, 1.26)	1.03 (0.76, 1.48)	1.11 (0.54, 2.13)
Σ_{144}	1.50	1.48 (1.15, 1.70)	1.87 (1.01, 1.91)	1.44 (1.30, 1.96)

Table S2: Run time in seconds per 100 MCMC iterations for sagittal anterior mouse brain data. All models were fit on an M1 iMac with 8gb RAM.

Dimension (g)	Run Time (K = 6)	Run Time (K = 8)	Run Time (K = 10)	Run Time (K = 12)
6	54	68	87	106
7	62	77	93	114
8	62	83	97	117
9	63	81	99	122
10	64	89	109	129
11	65	92	118	128
12	72	88	102	131
13	69	90	111	130
14	73	93	114	136
15	77	101	119	146
16	80	97	121	138
17	73	98	121	144
18	83	106	132	151

Web Appendix E: Inverse-Wishart Sensitivity Analysis

We sought to investigate the sensitivity of posterior inference of $z_1, ..., z_n$ to the choice of inverse-Wishart prior scale matrix $\mathbf{S}_{0k}$ and degrees of freedom ν_{0k}. We considered the simulated mouse brain data discussed in Section 4 of the manuscript, which featured $n = 2696$ cell spots divided into $K = 4$ sub-populations. We simulated multivariate outcomes $\mathbf{y}_1, ...\mathbf{y}_n$ with $g = 4$ gene expression features. The outcomes $\mathbf{y}_1, ...\mathbf{y}_n$ were generated according to model (3) of the manuscript, where mixture component-specific parameters $\boldsymbol{\mu}_k$ were used to achieve separation between mixture components as:

$$\boldsymbol{\mu}_1 = (-3, 3, -3, 3),\ \boldsymbol{\mu}_2 = (1, -1, 1, -1),$$
$$\boldsymbol{\mu}_3 = (-1, 1, -1, 1),\ \boldsymbol{\mu}_4 = (3, -3, 3, -3),$$

and all mixture components shared a common variance parameter $\boldsymbol{\Sigma}$, i.e.,

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_4 = \begin{bmatrix} 1.50 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1.50 & 0.25 & 0.25 \\ 0.25 & 0.25 & 1.50 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1.50 \end{bmatrix}.$$

We then fit the SPRUCE model using each combination of prior degrees of freedom $\nu_{0k} \in \{2, 4, 6, 10, 100\}$ for $k = 1, ..., 4$, and four prior scale parameters $\mathbf{S}_{0k}^{(1)}, ..., \mathbf{S}_{0k}^{(4)}$ defined as:

$$\mathbf{S}_{0k}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},\ \mathbf{S}_{0k}^{(2)} = \begin{bmatrix} 2 & -1 & -1 & -1 \\ -1 & 2 & -1 & -1 \\ -1 & -1 & 2 & -1 \\ -1 & -1 & -1 & 2 \end{bmatrix},\ \mathbf{S}_{0k}^{(3)} = \begin{bmatrix} 10 & 1 & 1 & 1 \\ 1 & 10 & 1 & 1 \\ 1 & 1 & 10 & 1 \\ 1 & 1 & 1 & 10 \end{bmatrix},\ \mathbf{S}_{0k}^{(4)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

In Table S3 below, we show the count (%) of misclassified cell spots from each model fit using the the inverse-Wishart parameter combinations shown above. We found that the model was able to achieve a low misclassification across all prior parameter combinations. Since the mixture component labeling parameters $z_1, ..., z_n$ are our main object of inference, we conclude that most analyses will be robust to specification of inverse-Wishart priors for $\boldsymbol{\Sigma}_k$.

Table S3: Count (%) of misclassified cell spots out of a total of 2696 from model fit across a range of inverse-Wishart parameter combinations. Posterior estimates of cell spot labels were computed from 1000 post burn-in MCMC samples using a burn-in period of 1000 iterations.

	$\nu_{0k} = 2$	$\nu_{0k} = 4$	$\nu_{0k} = 6$	$\nu_{0k} = 10$	$\nu_{0k} = 100$
$\mathbf{S}_{0k} = \mathbf{S}_{0k}^{(1)}$	2 (0.07%)	2 (0.07%)	0 (0.00%)	2 (0.07%)	2 (0.07%)
$\mathbf{S}_{0k} = \mathbf{S}_{0k}^{(2)}$	2 (0.07%)	2 (0.07%)	1 (0.04%)	2 (0.07%)	2 (0.07%)
$\mathbf{S}_{0k} = \mathbf{S}_{0k}^{(3)}$	2 (0.07%)	2 (0.07%)	1 (0.04%)	2 (0.07%)	2 (0.07%)
$\mathbf{S}_{0k} = \mathbf{S}_{0k}^{(4)}$	2 (0.07%)	3 (0.11%)	1 (0.04%)	2 (0.07%)	2 (0.07%)

Web Appendix F: Comparison of Model Fit Criteria

We conducted a simulation study to assess the performance of the six model fit criteria discussed in Section 3.3.2 of the manuscript across various levels of clustering signal in the data. Using the ground truth 4−component mouse brain labels and coordinates studied in Section 4, we simulated three data sets: setting 1, setting 2, and setting 3, where the separation of simulated mixture components is high in setting 1, moderate in setting 2, and low in setting 3. As a result, setting 1 presents an "easy" scenario for the model, while settings 2 and 3 represent "medium" and "hard" scenarios, respectively. We control levels of separation in mixture components using the mean $\boldsymbol{\mu}_k$ and variance $\boldsymbol{\Sigma}_k$ parameters of model (3) of the manuscript. UMAP plots summarizing separation of mixture components under each setting are provided in Figure S7. For each simulated data set, we fit the SPRUCE model for a range of $K = 1, ..., 8$ and computed each of the six model fit criteria discussed in Section 3.3.2. The results are displayed below in Figure S8.

We find that entropy increased markedly for $K \geq 5$, reflecting the lack of separation between artificially separated sub-populations. The negative log-likelihood criteria identified the correct K under settings 1 and 2, but overestimates K in setting 3. Similar patterns were observed with AIC and BIC. DIC identified the correct model dimension in settings 1 and 2, but was unable to differentiate between models using $K \in \{2, 3, 4\}$ in setting 3. Finally, WAIC identified the correct model dimension across all settings, and generally preferred more parsimonious models. Generally speaking, AIC and BIC more clearly identified the correct K in settings 1 and 2 when compared to WAIC and DIC. Thus, while we utilize WAIC due to its reliable performance and preference for parsimonious models, we conclude that viable alternatives exist for this setting.

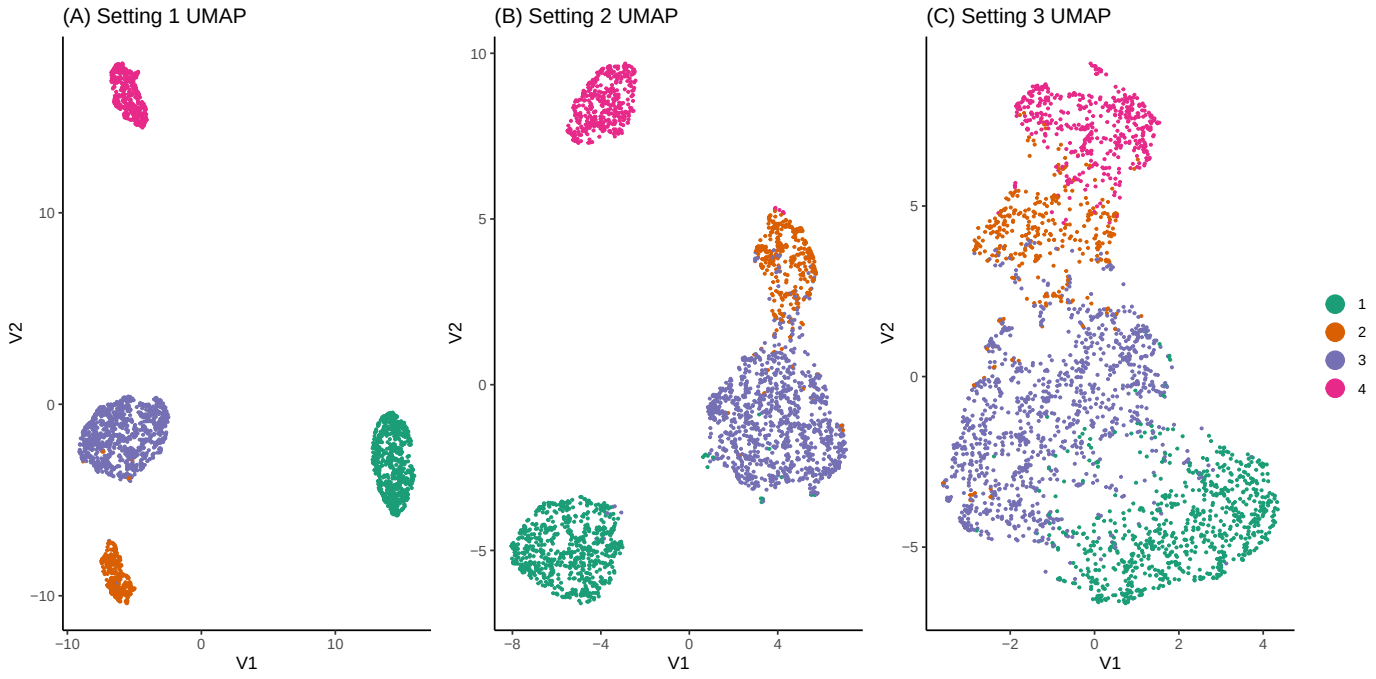

Figure S7: UMAP plots visualizing the relative separation of mixture components across three simulation settings considered in Section 4.2. (A) Setting 1 with high separation, (B) Setting 2 with moderate separation, (C) Setting 3 with low separation. Dimensions V1 and V2 represent UMAP coordinates.

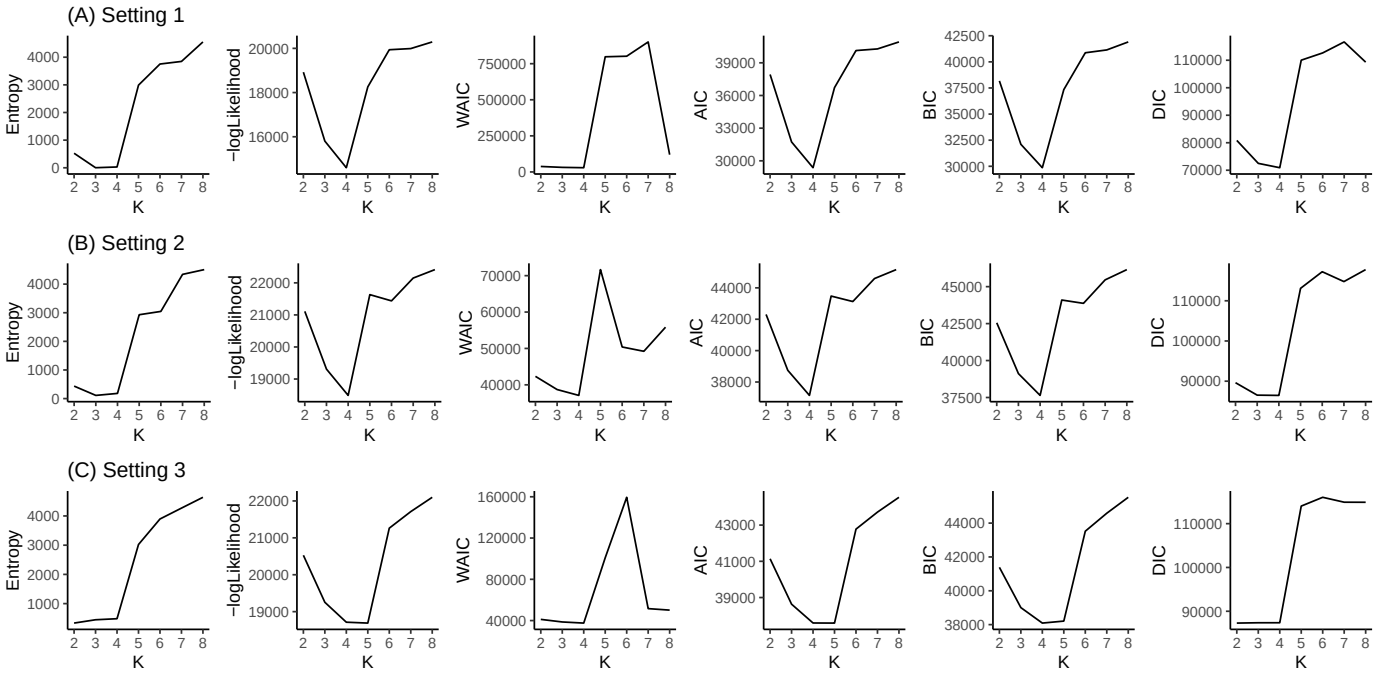

Figure S8: Comparison of six model fit criteria across three simulation settings: (A) setting 1: "easy" data with clear separation of mixture components, (B) setting 2: "medium" data with moderate separation of mixture components, and (C) setting 3: "hard" data with low separation of mixture components.

References

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association* **108,** 1339–1349.