Web-based Supplementary Materials for "Statistical Significance for Hierarchical Clustering"

Patrick K. Kimes, Yufeng Liu, D. Neil Hayes, and J. S. Marron

Web Appendix A: Proofs

Proof of Theorem 1

Let $d_W(\cdot, \cdot)$ denote the Ward's linkage function defined over sets of observation indices. Additionally, let $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ denote n and m samples drawn from two Gaussian components with distributions $N(\boldsymbol{\mu}_1, \sigma_1^2 \mathbf{I}_p)$ and $N(\boldsymbol{\mu}_2, \sigma_2^2 \mathbf{I}_p)$, with corresponding observation index sets, $\mathbb{C}^{(1)}$ and $\mathbb{C}^{(2)}$. For $k = 1, 2$, let $C_0^{(k)}$, $C_1^{(k)}$ and $C_2^{(k)}$ denote subsets of $\mathbb{C}^{(k)}$, where $C_1^{(k)}$ and $C_2^{(k)}$ are necessarily disjoint. Let $n_0 = |C_0^{(1)}|$, $n_1 = |C_1^{(1)}|$, $n_2 = |C_2^{(1)}|$, $m_0 = |C_0^{(2)}|$, $m_1 = |C_1^{(2)}|$ and $m_2 = |C_2^{(2)}|$ denote the size of each subset. Finally, let $\boldsymbol{X}_0^{(k)}$, $\boldsymbol{X}_1^{(k)}$, and $\boldsymbol{X}_2^{(k)}$, denote the corresponding subsets of $\mathbb{X}^{(k)}$, with corresponding sample means, $\overline{\boldsymbol{X}}_0^{(k)}$, $\overline{\boldsymbol{X}}_1^{(k)}$, and $\overline{\boldsymbol{X}}_2^{(k)}$.

Consider the two events: $A = \{\max d_W(C_1^{(1)}, C_2^{(1)}) < \min d_W(C^{(1)}, C^{(2)})\}$, and $B = \{\max d_W(C_1^{(2)}, C_2^{(2)}) < \min d_W(C^{(1)}, C^{(2)})\}$, where maxima and minima are taken with respect to the possible values of $C_1^{(k)}$, $C_2^{(k)}$, and $C^{(k)}$. Note that the joint occurrence of A and B is sufficient for correctly separating observations from the two components at the root node. It therefore suffices to show that $P(A \cap B) \to 1$ as $p \to \infty$.

For some $0 < a_1 \leq a_2$, define the following events: $E_1 = \{\max d_W(C_1^{(1)}, C_2^{(1)}) < a_1 \cdot p\}$, $E_2 = \{\max d_W(C_1^{(2)}, C_2^{(2)}) < a_1 \cdot p\}$, $E_3 = \{\min d_W(C_0^{(1)}, C_0^{(2)}) > a_2 \cdot p\}$. Note that the

probability of the joint event $(A \cap B)$ can be bounded below by:

$$\begin{aligned} P(A \cap B) &= 1 - P(A^C \cup B^C) \\ &\geq 1 - \big(P(E_1^C) + P(E_2^C) + P(E_3^C)\big). \end{aligned}$$

We complete the proof by showing that $P(E_1^C)$, $P(E_2^C)$, $P(E_3^C)$ all tend to 0 as $p \to \infty$.

By Lemmas 3 and 4 of Borysov et al. (2014) for $a_1 > 2\sigma_1^2$, we have

$$\begin{aligned} P(E_1^C) &= P(\max d_W(C_1^{(1)}, C_2^{(1)}) > a_1 \cdot p) \\ &= P\left(\max \frac{2n_1 n_2}{n_1 + n_2} \left\| \overline{\boldsymbol{X}}_1^{(1)} - \overline{\boldsymbol{X}}_2^{(1)} \right\|^2 > a_1 \cdot p\right) \\ &\leq 3^n P\left(\left\| \left(\frac{2n_1 n_2}{n_1 + n_2}\right)^{1/2} \left(\overline{\boldsymbol{X}}_1^{(1)} - \overline{\boldsymbol{X}}_2^{(1)}\right) \right\|^2 > a_1 \cdot p\right) \\ &\leq e^{-c_1 p + n \log 3}, \end{aligned}$$

where $c_1 = a_1/\sigma_1^2 - (1 + \log(a_1/\sigma_1^2))$. Note that since $c_1 > 0$ and $n = o(p) + p^\alpha$ for some $\alpha \in (0, 1)$, $P(E_1^C) \to 0$ as $p \to \infty$. Similarly, for $a_2 > 2\sigma_2^2$, we have

$$P(E_2^C) \leq e^{-c_2 p + m \log 3},$$

where $c_2 = a_2/\sigma_2^2 - (1 + \log(a_2/\sigma_2^2))$, such that $P(E_2^C) \to 0$ as $p \to \infty$. Finally, to bound $P(E_3^C)$, we make use of Lemmas 2 and 4 of Borysov et al. (2014):

$$\begin{aligned} P(E_3^C) &= P(\min d_W(C_0^{(1)}, C_0^{(2)}) < a_2 \cdot p) \\ &\leq \sum_{i=1}^{n} \sum_{j=1}^{m} P\left(\frac{2ij}{i+j} \left\| \overline{\boldsymbol{X}}_0^{(1)} - \overline{\boldsymbol{X}}_0^{(2)} \right\|^2 < a_2 \cdot p\right) \\ &\leq 2^{n+m} \max_{i \leq n,\ j \leq m} P\left(\frac{2ij}{i+j} \left\| \overline{\boldsymbol{X}}_0^{(1)} - \overline{\boldsymbol{X}}_0^{(2)} \right\|^2 < a_2 \cdot p\right). \end{aligned}$$

Suppose that i and j are fixed, and let $\mu^2 = p^{-1}(\frac{2ij}{i+j})\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$, $\mu_k = (\frac{2ij}{i+j})^{1/2}(\boldsymbol{\mu}_{1,k} - \boldsymbol{\mu}_{2,k})$, $\sigma^2 = (\frac{2ij}{i+j})(\frac{i\sigma_2^2+j\sigma_1^2}{ij})$. Then, using the result of Lemma 2 of Borysov et al. (2014), for $0 < a_2 < \sigma^2 + \mu^2$, we have

$$P\left(\frac{2ij}{i+j}\left\|\overline{\boldsymbol{X}}_0^{(1)} - \overline{\boldsymbol{X}}_0^{(2)}\right\|^2 < a_2 \cdot p\right) \leq e^{-c_3 p},$$

where $c_3 = (a_2 - \sigma^2 - \mu^2)^2/(6\sigma^4 + 12\sigma^2\mu^2 + 2p^{-1}\sum_{k=1}^{p}\mu_k^4)$.

Using the fourth moment bound of (2), and the fact that $n = o(p) + p^\alpha$, $m = o(p) + p^\beta$, we have that $P(E_3^C) \to 0$ as $p \to \infty$. Thus, for $a_1 > 2\sigma_1^2$, $P(E_1^C) \to 0$, for $a_1 > 2\sigma_2^2$, $P(E_2^C) \to 0$, and for $a_2 < (\frac{2nm}{n+m})(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p})$, $P(E_3^C) \to 0$. Combining the necessary inequalities on a_1, a_2, we obtain the stated condition:

$$\begin{aligned} 2\sigma_2^2 < a_1 \leq a_2 & \\ &< \frac{2nm}{n+m}\left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p}\right) \\ \frac{1}{n}(\sigma_2^2 - \sigma_1^2) &< \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p}. \end{aligned}$$

Proof of Theorem 2

Let $d_W(\cdot, \cdot)$ denote the Ward's linkage function. Further, let $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(2)}$ denote the n and m observations from the first and second Gaussian components with distributions $N(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I}_p)$ and $N(\boldsymbol{\mu}_2, \sigma_2^2\mathbf{I}_p)$. Assume that $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, σ_1^2 and σ_2^2 are known. Then, the theoretical best fit Gaussian to the mixture distribution is equivalent (up to a mean shift and rotation) to

$N(\mathbf{0}, \widehat{\mathbf{\Sigma}})$, where

$$\begin{aligned}
\widehat{\mathbf{\Sigma}} &= \text{diag}\{\widehat{\lambda}_k\}_{k=1}^{p} \\
\widehat{\lambda}_1 &= \frac{nm}{n+m}\left((n+m)^{-1}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 + \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right) \\
\widehat{\lambda}_k &= \frac{nm}{n+m}\left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right), \quad \text{for } k \geq 2,
\end{aligned}$$

where the $\widehat{\lambda}_k$ are derived by the formula for the variance of a univariate mixture of Gaussians. Let $\mathbb{X}^{(3)}$ denote a sample of $n+m$ observations drawn from $N(\mathbf{0}, \widehat{\mathbf{\Sigma}})$. Let $\mathbb{C}^{(k)}$ denote the corresponding observation indices for $k = 1, 2, 3$. Additionally, let $C_1^{(3)}$ and $C_2^{(3)}$ denote disjoint subsets of $\mathbb{C}^{(3)}$, and let $r_1 = |C_1^{(3)}|$ and $r_2 = |C_2^{(2)}|$ denote the size of the subsets. Finally, let $\boldsymbol{X}_1^{(3)}$ and $\boldsymbol{X}_2^{(3)}$ denote the corresponding subsets of $\mathbb{X}^{(3)}$ with means $\overline{\boldsymbol{X}}_1^{(3)}$ and $\overline{\boldsymbol{X}}_2^{(3)}$.

Consider the event: $D = \{\max d_W(C_1^{(3)}, C_2^{(3)}) < d_W(\mathbb{C}^{(1)}, \mathbb{C}^{(2)})\}$, where the maximum is taken with respect the possible values of $C_1^{(3)}$ and $C_2^{(3)}$. By Theorem 1 we have that, asymptotically, Ward's linkage clustering achieves the correct partition of $\mathbb{C}^{(1)}$ and $\mathbb{C}^{(2)}$. Therefore, D is precisely the event that a linkage value simulated from the null distribution is less than the observed linkage value. The proof is completed by showing $P(D) \to 1$ as $p \to \infty$. That is, we wish to show that the empirical p-value tends to 0 as $p \to \infty$.

For some $a > 0$, define the following events: $E_4 = \{\max d_W(C_1^{(3)}, C_2^{(3)}) < a \cdot p\}$, and $E_5 = \{d_W(\mathbb{C}^{(1)}, \mathbb{C}^{(2)}) > a \cdot p\}$. Note that $P(D)$ can be bounded below by:

$$\begin{aligned}
P(D) &= 1 - P(D^C) \\
&\geq 1 - \big(P(E_4^C) + P(E_5^C)\big).
\end{aligned}$$

Thus, it suffices to show the probabilities of E_4^C, E_5^C,both tend to 0 as $p \to \infty$.

First, we state a generalization of Lemma 3 from Borysov et al. (2014) for Gaussian distributions with diagonal covariance.

Lemma 1 *Suppose n independent observations, $\mathbb{X}$, are drawn from the p-dimensional Gaussian distribution, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix with diagonal entries $\{\lambda_k\}_{k=1}^p$. Define scalars $\mu^2 = p^{-1}\|\boldsymbol{\mu}\|^2$, $\overline{\lambda} = p^{-1}\sum_{k=1}^p \lambda_k$, and let $a > \overline{\lambda} + \mu^2$. Then, for any $0 < i \leq n$,*

$$P(\|X_i\|^2 > a \cdot p) \leq e^{-cp},$$

where $c = \left[a + \mu^2 - \sqrt{\overline{\lambda}^2 + 4\mu^2 a} + \overline{\lambda}\left(\frac{\overline{\lambda} + \sqrt{\overline{\lambda}^2 + 4\mu^2 a}}{2a}\right)\right] / \overline{\lambda}.$

The proof of Lemma 1 is omitted as it follows exactly as that of Lemma 3 from Borysov et al. (2014). By Lemma 1 given above and Lemma 4 of Borysov et al. (2014) , for $a > 2\overline{\widehat{\lambda}}$, where $\overline{\widehat{\lambda}} = p^{-1}\sum_{k=1}^p \widehat{\lambda}_k$, we have

$$\begin{aligned}
P(E_4^C) &= P(\max d_W(C_1^{(3)}, C_2^{(3)}) > a \cdot p) \\
&= P\left(\max \frac{2r_1 r_2}{r_1 + r_2} \left\|\overline{\boldsymbol{X}}_1^{(3)} - \overline{\boldsymbol{X}}_2^{(3)}\right\|^2 > a \cdot p\right) \\
&\leq 3^{n+m} P\left(\left\|\left(\frac{2r_1 r_2}{r_1 + r_2}\right)^{1/2}\left(\overline{\boldsymbol{X}}_1^{(3)} - \overline{\boldsymbol{X}}_2^{(3)}\right)\right\|^2 > a \cdot p\right) \\
&\leq e^{-c_4 p + (n+m)\log 3},
\end{aligned}$$

where $c_4 = a/\overline{\widehat{\lambda}} - (1 + \log(a/\overline{\widehat{\lambda}}))$. As for $P(E_1^C)$ from the proof of Theorem 2, we have that as $p \to \infty$, $P(E_4^C) \to 0$ as $p \to 0$. Next, using an argument similar to the one presented above for $P(E_3^C) \to 0$, we show that $P(E_5^C) \to 0$. Let $\mu^2 = p^{-1}(\frac{2nm}{n+m})\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$, $\mu_k = (\frac{2nm}{n+m})^{1/2}(\boldsymbol{\mu}_{1,k} - \boldsymbol{\mu}_{2,k})$, $\sigma^2 = (\frac{2nm}{n+m})(\frac{i\sigma_2^2 + j\sigma_1^2}{nm})$. Then, by Lemmas 2 and 4 of Borysov et al. (2014), for $0 < a < \sigma^2 + \mu^2$, we have

$$\begin{aligned}
P(E_4^C) &= P(d_W(\mathbb{C}^{(1)}, \mathbb{C}^{(2)}) < a \cdot p) \\
&= P\left(\frac{2nm}{n+m}\left\|\overline{\boldsymbol{X}}^{(1)} - \overline{\boldsymbol{X}}^{(2)}\right\|^2 < a \cdot p\right) \\
&\leq e^{-c_5 p},
\end{aligned}$$

where $c_5 = (a - \sigma^2 - \mu^2)^2/(6\sigma^4 + 12\sigma^2\mu^2 + 2p^{-1}\sum_{k=1}^{p}\mu_k^4)$. As for $P(E_3^C)$ from the proof of Theorem 2, we have $P(E_5^C) \to 0$ as $p \to \infty$. Thus, for $a > 2\overline{\widehat{\lambda}}$, $P(E_4^C) \to 0$, and for $a < (\frac{2nm}{n+m})(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p})$, $P(E_5^C) \to 0$. Combining the two inequalities on a, we obtain the stated condition:

$$
\begin{aligned}
2p^{-1}\sum_{k=1}^{p}\widehat{\lambda}_k &< a \\
&< \left(\frac{2nm}{n+m}\right)\left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p}\right) \\
\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} + (n+m)^{-1} \cdot \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p} &< \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} + \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p} \\
\frac{(m^2 - n^2)(\sigma_2^2 - \sigma_1^2)}{nm(n+m-1)} &< \frac{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{p}.
\end{aligned}
$$

Web Appendix B: Additional Simulation Discussion and Results

In Maitra et al. (2012), the BootClust approach was only tested in settings with dimension up to 80, suggesting that the method may not be suitable for inference in settings of substantially higher dimension, e.g. $p = 1000$. Therefore, to keep the total computational time of the simulations manageable, for each setting a single replication of the BootClust approach was first run with a maximum runtime of two hours. If more than two hours was needed for the approach to finish, the method was assumed to be incapable of performing the analysis. Most simulations with $p = 1000$ required longer than the pre-specified two hours to complete. In all other $p = 1000$ settings, BootClust consistently returned the maximum number of clusters considered, 10, regardless of the underlying dataset. Representative results for applying the BootClust approach to high-dimensional settings with $K = 1, 2, 3, 4$ are shown in Table S1. Since these results illustrate poor behavior of the method when $p = 1000$, performance using BootClust is not reported for these settings.

We next provide a brief review of the fundamental differences between `pvclust` and our proposed SHC method. Similar to SHC, `pvclust` also tests at nodes along the dendrogram. However, no test is performed at the root node, and the corresponding hypotheses tested at each node is given by:

$$H_0 : \text{ the cluster does not exist}$$

$$H_1 : \text{ the cluster exists.}$$

The difference between the two approaches can be understood by examining the dendrogram presented in Figure 2B of the main text. Using SHC, significant evidence of the three clusters is obtained if the null hypothesis is rejected at the top two nodes of the dendrogram. In contrast, to identify the three clusters using `pvclust`, the null hypothesis must be rejected at the three nodes directly above each cluster, denoted by their respective cluster symbol.

Table S1: $K = 1, 2, 3, 4$ simulation results for the BootClust (BC) method in select high-dimensional settings ($p = 1000$). Each setting was replicated 5 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = K|$), the mean number of significant clusters (mean $\hat{K}$), and median computing time are reported. All simulations were allowed to run until completion. (K: number of clusters, N: sample size, p: dimension, δ: separating signal between clusters, shape: orientation of cluster centroids, F: cluster distribution.)

parameters						$\lvert\hat{K} = K\rvert$ (mean $\hat{K}$)	median time (sec.)
K	N	p	δ	shape	F	BC	BC
1	50	1000	0	–	Gaus.	0 (10)	1482.73
1	100	1000	0	–	Gaus.	0 (10)	1803.63
1	200	1000	0	–	Gaus.	0 (10)	3878.31
1	50	1000	0	–	t_3	0 (10)	11432.72
1	100	1000	0	–	t_3	0 (10)	11241.9
1	200	1000	0	–	t_3	0 (10)	13108.73
2	100	1000	6	Line	Gaus.	0 (10)	2752.52
2	200	1000	6	Line	Gaus.	0 (10)	3728.86
2	100	1000	6	Line	t_3	0 (10)	15115.67
2	200	1000	6	Line	t_3	0 (10)	16875.53
3	150	1000	12	Triangle	Gaus.	0 (10)	3369.85
3	300	1000	12	Triangle	Gaus.	0 (10)	4818.35
3	150	1000	12	Triangle	t_3	0 (10)	3535.31
3	300	1000	12	Triangle	t_3	0 (10)	14395.84
4	200	1000	12	Tetrahedron	Gaus.	0 (10)	2192.96
4	400	1000	12	Tetrahedron	Gaus.	0 (10)	5897.81
4	200	1000	12	Tetrahedron	t_3	0 (10)	12890.73
4	400	1000	12	Tetrahedron	t_3	0 (10)	4585.91

Table S2: $K = 1$ simulation results for low-dimensional settings ($p = 10$). Each setting was replicated 100 times. For each setting, the number of false positives ($|p\text{-value} < 0.05|$) and median computing time are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, p: dimension, v: scaling factor of first dimension to mimic low-dimensional noise, F: cluster distribution.)

parameters			$\lvert p\text{-value} < 0.05\rvert$				median time (sec.)			
N	v	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC
50	1	Gaus.	1	0	0	0	9.81	0.09	0.15	0.43
50	10	Gaus.	8	0	1	0	10.33	0.08	0.17	0.4
50	50	Gaus.	10	0	1	0	8.45	0.06	0.12	0.29
50	100	Gaus.	19	0	3	0	8.52	0.06	0.12	0.29
50	1000	Gaus.	49	0	1	0	8.34	0.06	0.12	0.28
50	1	t_6	1	0	0	0	8.6	0.07	0.12	0.29
50	10	t_6	15	0	0	0	8.57	0.06	0.12	0.29
50	50	t_6	22	0	2	0	8.62	0.06	0.12	0.29
50	100	t_6	12	0	0	0	8.65	0.06	0.12	0.3
50	1000	t_6	51	0	0	0	8.49	0.06	0.12	0.28
50	1	t_3	16	3	10	12	8.59	0.06	0.12	0.3
50	10	t_3	24	0	1	4	8.72	0.06	0.12	0.29
50	50	t_3	32	2	4	4	8.55	0.06	0.12	0.29
50	100	t_3	40	0	2	0	8.54	0.06	0.12	0.29
50	1000	t_3	53	0	1	0	8.53	0.06	0.12	0.28
100	1	Gaus.	0	0	0	0	22.38	0.16	0.23	0.78
100	10	Gaus.	3	0	1	0	23.83	0.22	0.31	1.04
100	50	Gaus.	5	0	1	0	21.95	0.17	0.26	0.72
100	100	Gaus.	10	0	4	0	25.29	0.22	0.31	1.04
100	1000	Gaus.	21	0	3	0	22.6	0.16	0.25	0.7
100	1	t_6	1	0	0	0	22.44	0.16	0.24	0.79
100	10	t_6	0	0	0	0	22.83	0.16	0.24	0.78
100	50	t_6	10	0	0	0	22.38	0.16	0.23	0.77
100	100	t_6	12	0	0	0	22.48	0.16	0.23	0.76
100	1000	t_6	20	0	0	0	22.55	0.15	0.23	0.74
100	1	t_3	18	16	20	28	22.02	0.22	0.31	1.09
100	10	t_3	20	2	4	4	22.6	0.17	0.31	0.86
100	50	t_3	23	1	2	3	21.71	0.21	0.29	1.01
100	100	t_3	22	0	0	0	22.96	0.17	0.26	0.72
100	1000	t_3	32	1	3	4	22.57	0.16	0.23	0.75
200	1	Gaus.	0	0	0	0	68.81	0.51	0.63	2.79
200	10	Gaus.	0	0	1	0	69.11	0.51	0.62	2.77
200	50	Gaus.	5	0	1	0	68.18	0.5	0.61	2.73
200	100	Gaus.	6	0	3	0	68.6	0.5	0.61	2.7
200	1000	Gaus.	8	0	4	0	68.38	0.49	0.6	2.64
200	1	t_6	0	0	0	1	81.41	0.79	0.92	4.54
200	10	t_6	2	0	0	0	62.89	0.49	0.62	2.43
200	50	t_6	3	0	0	0	61.84	0.61	0.74	3.25
200	100	t_6	2	0	0	0	62.84	0.48	0.62	2.38
200	1000	t_6	5	0	0	0	61.53	0.61	0.71	3.19
200	1	t_3	7	9	11	32	68.72	0.51	0.62	2.84
200	10	t_3	14	2	4	8	68.94	0.51	0.62	2.8
200	50	t_3	12	2	2	3	68.3	0.5	0.61	2.76
200	100	t_3	14	2	2	2	68.72	0.5	0.62	2.74
200	1000	t_3	23	0	0	0	62.52	0.62	0.85	3.17

Table S3: $K = 1$ simulation results for moderate-dimensional settings ($p = 100$). Each setting was replicated 100 times. For each setting, the number of false positives ($|p\text{-value} < 0.05|$) and median computing time are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, p: dimension, v: scaling factor of first dimension to mimic low-dimensional noise, F: cluster distribution.)

parameters			$\lvert p\text{-value} < 0.05\rvert$				median time (sec.)			
N	v	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC
50	1	Gaus.	0	0	0	100	10.06	0.39	0.49	4.89
50	10	Gaus.	2	0	0	100	9.99	0.37	0.48	4.92
50	50	Gaus.	5	0	0	100	14.47	0.38	0.49	6.07
50	100	Gaus.	3	0	0	100	10.25	0.38	0.48	5.02
50	1000	Gaus.	22	0	2	100	10.4	0.32	0.48	5.41
50	1	t_6	2	0	0	100	11.7	0.38	0.48	5.62
50	10	t_6	5	0	0	100	14.43	0.34	0.44	6.88
50	50	t_6	8	0	0	100	10.35	0.36	0.48	5.63
50	100	t_6	5	0	0	100	10.48	0.27	0.36	5.62
50	1000	t_6	24	0	1	100	10.44	0.34	0.45	5.48
50	1	t_3	16	15	20	100	10.18	0.38	0.49	7.79
50	10	t_3	15	13	22	100	12.45	0.34	0.44	15.24
50	50	t_3	25	4	7	100	18.01	0.38	0.48	16.42
50	100	t_3	19	1	3	100	10.5	0.38	0.49	6.81
50	1000	t_3	42	0	1	100	10.31	0.38	0.5	5.97
100	1	Gaus.	0	0	0	46	28.59	1.2	1.38	7.35
100	10	Gaus.	0	0	0	51	47.24	1.2	1.39	9.42
100	50	Gaus.	1	0	0	47	30.28	1.21	1.38	7.02
100	100	Gaus.	3	0	2	44	34.15	1.06	1.24	8.3
100	1000	Gaus.	10	0	2	41	27.41	1.2	1.39	5.07
100	1	t_6	1	0	0	57	34.01	1.21	1.4	8.06
100	10	t_6	1	0	0	50	32.59	1.2	1.4	7.95
100	50	t_6	1	0	0	48	36.85	1.2	1.39	8.98
100	100	t_6	3	0	0	48	34.52	1.2	1.39	8.25
100	1000	t_6	5	0	0	51	31.59	1.19	1.38	7.83
100	1	t_3	14	40	45	72	36.6	1.21	1.45	18.9
100	10	t_3	10	20	21	59	32.45	1.2	1.39	14.09
100	50	t_3	19	5	9	61	34.59	1.2	1.4	12.18
100	100	t_3	16	0	2	51	36.05	1.2	1.39	10.15
100	1000	t_3	18	1	1	45	33.66	1.2	1.38	8.25
200	1	Gaus.	0	0	0	0	97.76	4.41	4.78	16.28
200	10	Gaus.	0	6	17	1	101.68	4.48	4.88	16.82
200	50	Gaus.	0	0	2	0	110.09	4.45	4.82	18.03
200	100	Gaus.	2	0	1	0	94.17	3.59	4.31	13.5
200	1000	Gaus.	5	0	2	0	115.4	4.45	4.84	20.57
200	1	t_6	0	0	0	0	107.66	4.57	4.97	19.4
200	10	t_6	0	0	0	0	112.94	4.15	4.51	19.8
200	50	t_6	0	0	0	0	115.66	4.55	4.94	21.21
200	100	t_6	1	0	0	0	95.23	4.51	4.87	18.28
200	1000	t_6	5	0	0	0	93.76	4.02	4.41	18.72
200	1	t_3	23	65	65	74	104.59	9	9.74	29.73
200	10	t_3	9	27	27	56	120.94	4.59	4.98	25.83
200	50	t_3	6	8	8	13	116.16	4.61	5.02	20.46
200	100	t_3	13	4	4	2	118.74	4.53	4.9	20.42
200	1000	t_3	16	1	2	1	113.63	4.44	4.81	20.01

Table S4: $K = 1$ simulation results for high-dimensional settings ($p = 1000$). Each setting was replicated 100 times. For each setting, the number of false positives ($|p\text{-value} < 0.05|$) and median computing time are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting (−). (N: sample size, p: dimension, v: scaling factor of first dimension to mimic low-dimensional noise, F: cluster distribution.)

parameters			$\lvert p\text{-value} < 0.05\rvert$				median time (sec.)			
N	v	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC
50	1	Gaus.	0	0	0	–	34.98	3.41	4.26	–
50	10	Gaus.	0	0	0	–	31.79	3.38	4.09	–
50	50	Gaus.	0	0	1	–	36	3.43	4.32	–
50	100	Gaus.	1	0	0	–	30.66	3.42	4.3	–
50	1000	Gaus.	12	0	1	–	35.8	3.06	3.75	–
50	1	t_6	3	0	0	–	24.63	3.43	4.27	–
50	10	t_6	0	1	1	–	34.53	2.41	3.07	–
50	50	t_6	5	0	0	–	31.82	3.32	4.14	–
50	100	t_6	9	0	0	–	35.42	3.32	4.14	–
50	1000	t_6	12	0	0	–	28.03	3.47	4.33	–
50	1	t_3	20	9	14	–	32.13	3.39	4.28	–
50	10	t_3	23	14	17	–	27.83	3.41	4.29	–
50	50	t_3	14	8	9	–	39.5	3.35	4.15	–
50	100	t_3	14	7	9	–	32.19	3.03	3.9	–
50	1000	t_3	28	1	3	–	33.44	3.43	4.27	–
100	1	Gaus.	0	0	0	–	96.79	12.03	13.65	–
100	10	Gaus.	0	0	0	–	130.63	11.96	13.76	–
100	50	Gaus.	0	0	0	–	81.52	11.74	13.58	–
100	100	Gaus.	0	0	0	–	94.58	12.16	13.74	–
100	1000	Gaus.	1	0	0	–	109.47	11.46	13	–
100	1	t_6	1	0	0	–	115.5	11.09	13.02	–
100	10	t_6	0	0	0	–	78.09	12.27	13.95	–
100	50	t_6	3	0	0	–	129.13	11.1	13.04	–
100	100	t_6	1	0	0	–	99.47	12.09	13.78	–
100	1000	t_6	5	0	0	–	97.05	10.7	12.52	–
100	1	t_3	22	45	48	–	101.62	12.43	14.58	–
100	10	t_3	22	48	50	–	102.03	12.6	18.78	–
100	50	t_3	4	18	19	–	128.57	11.94	13.52	–
100	100	t_3	11	9	9	–	117.9	12.28	14.23	–
100	1000	t_3	18	3	3	–	130.98	11.62	13.27	–
200	1	Gaus.	0	0	0	–	324.1	44.96	48.2	–
200	10	Gaus.	0	0	1	–	457.49	43.65	46.58	–
200	50	Gaus.	0	0	0	–	410.82	43.86	46.62	–
200	100	Gaus.	0	0	0	–	467.12	45.87	49.1	–
200	1000	Gaus.	1	0	0	–	386.62	46.16	49.71	–
200	1	t_6	3	0	0	–	286.82	46.01	49.27	–
200	10	t_6	1	0	0	–	286.01	45.79	49.16	–
200	50	t_6	0	0	0	–	287	42.55	45.6	–
200	100	t_6	0	0	0	–	289.29	46.07	49.22	–
200	1000	t_6	3	0	0	–	285.66	45.7	48.76	–
200	1	t_3	28	88	88	–	459.98	91.14	97.77	–
200	10	t_3	18	76	76	–	298.76	87.14	93.34	–
200	50	t_3	14	20	19	–	288.86	46.27	50.1	–
200	100	t_3	16	10	10	–	287.57	41.91	46.11	–
200	1000	t_3	9	1	1	–	288.84	39.81	42.68	–

Table S5: $K = 2$ simulation results for low-dimensional settings ($p = 10$) with equal cluster sizes ($\pi_1, \pi_2 = 1/2, 1/2$). Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 2|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 2\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
100	2	Gaus.	0 (19.02)	0 (1)	0 (1)	0 (1)	22.37	0.16	0.23	0.78	0.01	0	0	0
100	4	Gaus.	3 (18.92)	73 (1.73)	96 (1.96)	84 (1.84)	22.44	0.28	0.42	0.77	0.05	0.62	0.79	0.71
100	6	Gaus.	32 (13.56)	100 (2)	100 (2)	100 (2)	22.23	0.33	0.46	0.96	0.34	0.98	0.98	0.98
100	8	Gaus.	78 (5.66)	100 (2)	100 (2)	100 (2)	22.14	0.29	0.46	0.75	0.79	1	1	1
100	10	Gaus.	92 (3.41)	100 (2)	100 (2)	100 (2)	26.54	0.35	0.48	1.04	0.92	1	1	1
100	2	t_6	0 (18.54)	0 (1)	0 (1)	0 (1)	25.27	0.22	0.31	1.09	0.01	0	0	0
100	4	t_6	0 (19)	31 (1.31)	65 (1.65)	55 (1.55)	22.77	0.23	0.42	1.04	0.02	0.24	0.48	0.41
100	6	t_6	4 (18.22)	100 (2)	100 (2)	100 (2)	22.21	0.29	0.46	0.72	0.08	0.93	0.93	0.93
100	8	t_6	44 (11.33)	100 (2)	100 (2)	100 (2)	24.59	0.34	0.52	1	0.45	0.98	0.98	0.98
100	10	t_6	71 (6.7)	100 (2)	100 (2)	100 (2)	24.61	0.38	0.53	1.09	0.72	0.99	0.99	0.99
100	2	t_3	0 (16.19)	8 (1.1)	10 (1.14)	0 (3.07)	22.62	0.16	0.23	0.79	0.01	0	0	0.01
100	4	t_3	0 (18.18)	6 (1.06)	11 (1.13)	7 (4.49)	28.09	0.22	0.31	1.19	0.02	0	0.03	0.14
100	6	t_3	4 (16.77)	65 (1.77)	74 (2)	73 (3.44)	29.46	0.41	0.64	1.32	0.1	0.6	0.72	0.67
100	8	t_3	11 (15.12)	92 (2.06)	82 (2.17)	83 (3.27)	27.15	0.39	0.54	1.15	0.2	0.91	0.91	0.83
100	10	t_3	24 (12.87)	94 (2.04)	83 (2.18)	81 (3.52)	21.39	0.32	0.47	1.01	0.33	0.95	0.96	0.86
200	2	Gaus.	0 (43.44)	0 (1)	2 (1.02)	1 (1.01)	64.16	0.65	0.79	3.55	0.01	0	0.01	0
200	4	Gaus.	0 (43.51)	98 (1.98)	99 (1.99)	98 (1.98)	65.9	1.17	1.39	3.66	0.01	0.79	0.8	0.79
200	6	Gaus.	27 (32.1)	100 (2)	100 (2)	100 (2)	103.41	1.42	1.69	4.8	0.28	0.98	0.98	0.98
200	8	Gaus.	78 (11.07)	100 (2)	100 (2)	100 (2)	85.1	1.21	1.42	4.08	0.78	1	1	1
200	10	Gaus.	79 (10.15)	100 (2)	100 (2)	100 (2)	70.01	1.2	1.4	3.81	0.79	1	1	1
200	2	t_6	0 (41.3)	0 (1)	0 (1)	0 (1)	100.51	0.9	1.06	4.84	0	0	0	0
200	4	t_6	1 (40.82)	86 (1.86)	97 (1.97)	93 (1.96)	69.1	1.1	1.39	3.76	0.02	0.61	0.68	0.66
200	6	t_6	5 (40.09)	100 (2)	100 (2)	100 (2)	64.19	1.1	1.24	3.53	0.06	0.92	0.92	0.92
200	8	t_6	15 (35.69)	100 (2)	100 (2)	100 (2)	62.36	0.98	1.19	3.25	0.16	0.98	0.98	0.98
200	10	t_6	61 (17.18)	100 (2)	100 (2)	99 (2.08)	61.06	1.02	1.18	3.23	0.62	0.99	0.99	0.98
200	2	t_3	2 (34.7)	12 (1.12)	16 (1.16)	0 (3.97)	62.93	0.65	0.8	3.4	0	0	0	0.02
200	4	t_3	0 (39.2)	22 (1.28)	31 (1.47)	30 (4.63)	62.57	0.51	0.95	2.5	0.01	0.1	0.18	0.27
200	6	t_3	1 (41.23)	82 (2.02)	83 (2.09)	73 (4.16)	85.49	1.22	1.44	4.27	0.03	0.75	0.78	0.7
200	8	t_3	3 (38.3)	86 (2.15)	77 (2.25)	67 (4.64)	85.58	1.21	1.43	4.16	0.09	0.91	0.91	0.74
200	10	t_3	11 (33.38)	81 (2.16)	78 (2.21)	67 (4.64)	86.21	1.21	1.43	4.02	0.21	0.93	0.93	0.79

Table S6: $K = 2$ simulation results for moderate-dimensional settings ($p = 100$) with equal cluster sizes ($\pi_1, \pi_2 = 1/2, 1/2$). Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 2|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 2\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
100	2	Gaus.	24 (3.04)	0 (1)	0 (1)	0 (5.68)	29.07	0.83	0.98	10.51	0	0	0	0.01
100	4	Gaus.	15 (3.98)	1 (1.01)	1 (1.01)	0 (4.96)	40.9	0.85	1.15	8.32	0	0.01	0.01	0.06
100	6	Gaus.	11 (4.32)	99 (1.99)	100 (2)	0 (9.82)	28.94	1.65	2	11.93	0	0.9	0.91	0.21
100	8	Gaus.	7 (4.8)	100 (2)	100 (2)	0 (10)	28.41	1.54	1.83	11.87	0.01	0.99	0.99	0.22
100	10	Gaus.	65 (2.98)	100 (2)	100 (2)	0 (10)	28.35	1.64	1.99	11.89	0.62	1	1	0.22
100	2	t_6	24 (3.44)	2 (1.02)	2 (1.02)	0 (5.5)	28.89	0.86	1.13	9.91	0	0	0	0.01
100	4	t_6	13 (3.25)	1 (1.01)	1 (1.01)	0 (4.78)	29.35	0.98	1.12	7.22	0	0	0	0.04
100	6	t_6	18 (4)	44 (1.44)	55 (1.55)	0 (7.21)	28.46	1.21	1.57	11.95	0	0.38	0.46	0.16
100	8	t_6	10 (4.73)	100 (2)	100 (2)	0 (10)	28.6	1.96	2.33	12.15	0.01	0.94	0.94	0.26
100	10	t_6	6 (4.95)	100 (2)	100 (2)	0 (10)	40.57	1.4	1.82	15.03	0.03	0.99	0.99	0.27
100	2	t_3	19 (2.73)	26 (1.5)	31 (1.51)	0 (8.11)	29.04	1.21	1.49	14.45	0	0	0	0
100	4	t_3	7 (3.13)	28 (1.38)	30 (1.42)	0 (8.11)	35.65	1.2	1.38	21.77	0	0	0	0.04
100	6	t_3	12 (3.37)	27 (1.4)	27 (1.42)	0 (8.2)	32.68	1.2	1.39	19.44	0	0	0	0.25
100	8	t_3	15 (4.08)	35 (1.55)	44 (1.7)	0 (9.19)	47.92	1.23	1.68	24.9	0.01	0.23	0.34	0.47
100	10	t_3	10 (4.29)	83 (2.08)	78 (2.2)	0 (10)	29.54	1.36	1.78	17.66	0.04	0.84	0.87	0.62
200	2	Gaus.	4 (6.53)	0 (1)	0 (1)	0 (1)	95.54	4.35	4.72	18.73	0	0	0	0
200	4	Gaus.	2 (8.35)	78 (1.78)	100 (2)	1 (6.32)	93.77	4.66	5.18	33.83	0	0.45	0.56	0.1
200	6	Gaus.	0 (9.24)	100 (2)	100 (2)	4 (9.68)	94.34	4.7	5.21	34.65	0	0.92	0.92	0.24
200	8	Gaus.	0 (10.81)	100 (2)	100 (2)	4 (9.68)	114.48	6.92	7.69	24.56	0	0.99	0.99	0.26
200	10	Gaus.	47 (6.08)	100 (2)	100 (2)	2 (9.84)	139.92	6.59	7.28	26.26	0.47	1	1	0.24
200	2	t_6	3 (7.22)	1 (1.01)	1 (1.01)	0 (1)	95.15	3.04	3.28	18.52	0	0	0	0
200	4	t_6	5 (7.78)	15 (1.15)	16 (1.16)	0 (1.18)	95.23	3.17	3.89	18.8	0	0.07	0.07	0
200	6	t_6	0 (9.13)	100 (2)	100 (2)	3 (9.76)	92.66	4.59	5.22	34.7	0	0.78	0.78	0.24
200	8	t_6	3 (9.35)	99 (2.01)	99 (2.01)	3 (9.76)	93.91	4.65	5.29	34.86	0.02	0.95	0.95	0.28
200	10	t_6	2 (10.76)	99 (2.01)	99 (2.01)	5 (9.6)	138.88	4.79	5.33	26.98	0.02	0.99	0.99	0.3
200	2	t_3	2 (6.29)	44 (1.94)	44 (1.93)	0 (7.3)	112.59	5.85	6.36	30.88	0	0	0	0
200	4	t_3	3 (6.09)	31 (1.76)	30 (1.76)	0 (7.84)	94.75	5.72	6.28	25.97	0	0	0	0.07
200	6	t_3	0 (7.79)	33 (1.85)	32 (1.92)	0 (8.92)	95.24	5.84	6.49	32.06	0	0.09	0.12	0.35
200	8	t_3	0 (9.34)	45 (2.76)	46 (2.87)	1 (9.92)	102.98	7.05	7.78	34.84	0	0.72	0.74	0.61
200	10	t_3	1 (10.23)	48 (2.82)	42 (2.97)	0 (10)	132.49	7.1	7.96	41.41	0.01	0.87	0.87	0.71
400	2	Gaus.	0 (15.93)	0 (1)	1 (1.01)	0 (1)	327.85	15.52	16.06	46.11	0	0	0	0
400	4	Gaus.	0 (18.67)	99 (1.99)	100 (2)	96 (1.96)	305.35	18.8	19.64	41.46	0	0.58	0.59	0.58
400	6	Gaus.	0 (21.48)	100 (2)	100 (2)	100 (2)	319.15	19.81	20.66	41.59	0	0.93	0.93	0.93
400	8	Gaus.	1 (22.37)	100 (2)	100 (2)	100 (2)	303.51	18.69	19.49	41.03	0.01	0.99	0.99	0.99
400	10	Gaus.	43 (13.65)	100 (2)	100 (2)	100 (2)	341.81	21.63	22.83	53.22	0.43	1	1	1
400	2	t_6	0 (16.76)	1 (1.01)	1 (1.01)	0 (1)	357.37	13.02	13.48	49.08	0	0	0	0
400	4	t_6	0 (16.89)	60 (1.62)	60 (1.62)	40 (1.4)	406.8	17.06	18.2	63.97	0	0.25	0.26	0.17
400	6	t_6	0 (20.36)	99 (2.01)	100 (2)	99 (2.08)	315.49	19.17	20.02	41.78	0	0.8	0.8	0.79
400	8	t_6	1 (20.79)	99 (2.01)	99 (2.01)	99 (2.08)	317.96	25.68	26.79	41.37	0.01	0.95	0.95	0.95
400	10	t_6	1 (21.67)	100 (2)	100 (2)	100 (2)	323.59	24.01	24.79	43.86	0.01	0.99	0.99	0.99
400	2	t_3	3 (13.63)	38 (2.48)	35 (2.51)	0 (9.55)	325	31.91	34.18	57.36	0	0	0	0
400	4	t_3	4 (13.47)	40 (2.46)	38 (2.48)	0 (9.46)	320.67	33.31	34.9	57.15	0	0	0	0.11
400	6	t_3	0 (18.63)	27 (3.03)	23 (3.27)	3 (9.76)	361.27	27.37	29.8	60.19	0	0.32	0.35	0.44
400	8	t_3	0 (20.35)	22 (3.6)	22 (3.63)	8 (9.36)	338.65	28.48	29.66	54.85	0	0.73	0.73	0.68
400	10	t_3	0 (22.29)	26 (3.51)	21 (3.64)	9 (9.28)	330.3	27.52	29.02	55.73	0	0.87	0.87	0.81

Table S7: $K = 2$ simulation results for high-dimensional settings ($p = 1000$) with equal cluster sizes ($\pi_1, \pi_2 = 1/2, 1/2$). Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 2|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting ($-$). (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 2\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
100	2	Gaus.	30 (2.15)	0 (1)	0 (1)	–	77.68	8.86	10.14	–	0	0	0	–
100	4	Gaus.	32 (2.08)	0 (1)	0 (1)	–	77.86	8.38	9.6	–	0	0	0	–
100	6	Gaus.	26 (2.24)	0 (1)	0 (1)	–	77.56	7.7	9.03	–	0	0	0	–
100	8	Gaus.	27 (2.71)	29 (1.29)	94 (1.94)	–	77.13	11.75	15.29	–	0	0.25	0.75	–
100	10	Gaus.	24 (3.08)	100 (2)	100 (2)	–	78.26	14.82	17.37	–	0.01	0.95	0.95	–
100	2	t_6	27 (2.22)	0 (1)	0 (1)	–	78.53	8.35	9.38	–	0	0	0	–
100	4	t_6	22 (2.22)	0 (1)	1 (1.01)	–	100.38	8.33	9.57	–	0	0	0	–
100	6	t_6	29 (2.14)	0 (1)	0 (1)	–	99.47	8.47	9.73	–	0	0	0	–
100	8	t_6	28 (2.38)	0 (1)	0 (1)	–	102.59	10.76	12.52	–	0	0	0	–
100	10	t_6	30 (2.29)	72 (1.72)	73 (1.73)	–	130.74	12.67	14.65	–	0	0.61	0.62	–
100	2	t_3	17 (1.53)	43 (1.66)	43 (1.7)	–	99.14	14.89	17.52	–	0	0	0	–
100	4	t_3	12 (1.42)	40 (1.6)	38 (1.6)	–	80.52	13.42	14.58	–	0	0	0	–
100	6	t_3	23 (1.45)	34 (1.68)	34 (1.68)	–	91.08	14.51	16.51	–	0	0	0	–
100	8	t_3	9 (1.58)	39 (1.63)	40 (1.62)	–	96.26	13.54	15.66	–	0	0	0	–
100	10	t_3	27 (1.7)	38 (1.65)	40 (1.66)	–	96.1	14.76	17.45	–	0	0	0	–
200	2	Gaus.	11 (4.6)	0 (1)	0 (1)	–	410.95	35.6	38.56	–	0	0	0	–
200	4	Gaus.	11 (4.56)	0 (1)	0 (1)	–	459.59	40.31	43.2	–	0	0	0	–
200	6	Gaus.	9 (4.73)	20 (1.2)	56 (1.56)	–	438.72	31.74	44.62	–	0	0.12	0.3	–
200	8	Gaus.	6 (5.43)	100 (2)	100 (2)	–	421.5	44.84	49.34	–	0	0.82	0.82	–
200	10	Gaus.	5 (6.14)	100 (2)	100 (2)	–	438.66	47.8	51.53	–	0.02	0.96	0.96	–
200	2	t_6	14 (4.12)	0 (1)	0 (1)	–	504.84	31.69	33.87	–	0	0	0	–
200	4	t_6	15 (4.23)	0 (1)	0 (1)	–	329.8	28.83	31.37	–	0	0	0	–
200	6	t_6	12 (4.24)	1 (1.01)	1 (1.01)	–	466.4	30.09	32.53	–	0	0	0	–
200	8	t_6	11 (4.9)	68 (1.68)	71 (1.71)	–	408.21	45.61	49.8	–	0	0.43	0.44	–
200	10	t_6	5 (5.18)	100 (2)	100 (2)	–	467.84	45.93	49.63	–	0	0.83	0.83	–
200	2	t_3	12 (2.9)	39 (2.5)	36 (2.54)	–	369.67	89.11	96.99	–	0	0	0	–
200	4	t_3	5 (2.85)	28 (2.81)	27 (2.83)	–	349.23	84.88	91.4	–	0	0	0	–
200	6	t_3	13 (2.72)	37 (2.37)	37 (2.43)	–	290.8	80.9	89.34	–	0	0	0	–
200	8	t_3	8 (3.34)	33 (2.6)	30 (2.62)	–	292.14	79.43	86.33	–	0	0	0	–
200	10	t_3	16 (2.98)	40 (2.28)	39 (2.3)	–	373.17	62.86	67.59	–	0	0	0	–

Table S8: $K = 2$ simulation results with unequal cluster sizes ($\pi_1, \pi_2 = 1/3, 2/3$). Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 2|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting ($-$). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 2\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	p	δ	F	pvAU	SHC_L	SHC_2	BC	`pv`	SHC_L	SHC_2	BC	`pvAU`	SHC_L	SHC_2	BC
100	10	2	Gaus.	0 (18.64)	0 (1)	0 (1)	0 (1)	35.5	0.25	0.36	1.24	0.01	0	0	0
100	10	4	Gaus.	3 (18.94)	58 (1.58)	84 (1.84)	69 (1.69)	27.32	0.29	0.55	1.21	0.05	0.5	0.7	0.58
100	10	6	Gaus.	27 (14.41)	100 (2)	100 (2)	100 (2)	25.86	0.35	0.57	1.03	0.28	0.98	0.98	0.98
100	10	8	Gaus.	78 (5.58)	100 (2)	100 (2)	100 (2)	26.35	0.41	0.64	1.12	0.79	1	1	1
100	10	10	Gaus.	81 (5.24)	100 (2)	100 (2)	100 (2)	27.87	0.43	0.64	1.21	0.81	1	1	1
200	10	2	Gaus.	0 (43.03)	2 (1.02)	1 (1.01)	2 (1.02)	87.12	0.78	0.98	3.81	0	0.01	0.01	0.01
200	10	4	Gaus.	0 (43.31)	96 (1.96)	98 (1.98)	96 (1.96)	89.91	1.34	1.62	4.57	0.01	0.78	0.8	0.78
200	10	6	Gaus.	29 (31.86)	100 (2)	100 (2)	100 (2)	79.69	1.17	1.47	3.77	0.29	0.98	0.98	0.98
200	10	8	Gaus.	73 (12.8)	100 (2)	100 (2)	100 (2)	87.32	1.21	1.58	4.14	0.73	1	1	1
200	10	10	Gaus.	78 (10.95)	100 (2)	100 (2)	100 (2)	85.65	1.29	1.56	4.24	0.78	1	1	1
100	100	2	Gaus.	19 (3.32)	0 (1)	0 (1)	0 (5.5)	37.51	1.03	1.21	8.56	0	0	0	0.01
100	100	4	Gaus.	16 (3.87)	0 (1)	0 (1)	0 (5.68)	32.22	0.97	1.13	8.88	0	0	0	0.06
100	100	6	Gaus.	19 (4.09)	91 (1.91)	97 (1.97)	0 (9.73)	32.57	1.98	2.37	11.93	0	0.84	0.88	0.17
100	100	8	Gaus.	7 (4.77)	100 (2)	100 (2)	0 (10)	34.66	1.45	1.79	12.19	0.01	0.99	0.99	0.18
100	100	10	Gaus.	69 (2.9)	100 (2)	100 (2)	0 (10)	36.81	1.5	1.83	11.62	0.63	1	1	0.18
200	100	2	Gaus.	2 (6.87)	0 (1)	0 (1)	0 (1)	113.37	4.56	4.98	20.08	0	0	0	0
200	100	4	Gaus.	1 (8.35)	56 (1.56)	95 (1.95)	0 (3.97)	110.2	4.83	7.71	21.22	0	0.35	0.54	0.04
200	100	6	Gaus.	2 (8.8)	100 (2)	100 (2)	0 (10)	113.56	6.4	7.09	28.89	0.01	0.92	0.92	0.17
200	100	8	Gaus.	4 (9.52)	100 (2)	100 (2)	0 (10)	113.01	4.86	5.43	29.26	0.04	0.99	0.99	0.19
200	100	10	Gaus.	48 (6.11)	100 (2)	100 (2)	0 (10)	112.62	4.89	5.44	35.44	0.48	1	1	0.18
100	1000	2	Gaus.	25 (2.34)	0 (1)	0 (1)	–	79.56	7.58	8.87	–	0	0	0	–
100	1000	4	Gaus.	29 (2.39)	0 (1)	0 (1)	–	96.62	8.03	9.41	–	0	0	0	–
100	1000	6	Gaus.	25 (2.17)	0 (1)	1 (1.01)	–	108.93	10.14	11.82	–	0	0	0.01	–
100	1000	8	Gaus.	21 (2.51)	12 (1.12)	87 (1.87)	–	115.93	8.37	15.35	–	0	0.11	0.71	–
100	1000	10	Gaus.	26 (3.13)	100 (2)	100 (2)	–	115.47	14.98	17.2	–	0.01	0.95	0.95	–
200	1000	2	Gaus.	9 (4.76)	0 (1)	0 (1)	–	386.58	29.23	32.1	–	0	0	0	–
200	1000	4	Gaus.	13 (3.93)	0 (1)	0 (1)	–	379.9	27.61	30.15	–	0	0	0	–
200	1000	6	Gaus.	7 (5.21)	5 (1.05)	42 (1.42)	–	339.59	32.37	48.99	–	0	0.03	0.24	–
200	1000	8	Gaus.	10 (5.19)	100 (2)	100 (2)	–	404.11	49.67	53.81	–	0	0.83	0.83	–
200	1000	10	Gaus.	4 (6.09)	100 (2)	100 (2)	–	297.35	47.2	51.3	–	0.01	0.96	0.96	–

Table S9: $K = 2$ simulation results with unequal cluster sizes ($\pi_1, \pi_2 = 1/4, 3/4$). Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 2|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting ($-$). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 2\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	p	δ	F	pvAU	SHC_L	SHC_2	BC	`pv`	SHC_L	SHC_2	BC	`pvAU`	SHC_L	SHC_2	BC
100	10	2	Gaus.	0 (19)	0 (1)	0 (1)	0 (1)	26.9	0.24	0.35	1.2	0.01	0	0	0
100	10	4	Gaus.	2 (18.43)	27 (1.27)	61 (1.61)	42 (1.42)	29.3	0.25	0.46	1.21	0.03	0.23	0.52	0.36
100	10	6	Gaus.	36 (12.91)	100 (2)	100 (2)	100 (2)	30.19	0.45	0.65	1.2	0.37	0.98	0.98	0.98
100	10	8	Gaus.	82 (4.98)	100 (2)	100 (2)	100 (2)	30.04	0.44	0.66	1.19	0.82	1	1	1
100	10	10	Gaus.	82 (4.82)	100 (2)	100 (2)	100 (2)	25.42	0.33	0.59	1.13	0.82	1	1	1
200	10	2	Gaus.	0 (41.65)	0 (1)	1 (1.01)	0 (1)	91.71	0.81	1	4.54	0	0	0	0
200	10	4	Gaus.	2 (42.43)	87 (1.87)	94 (1.94)	94 (1.94)	89.53	1.32	1.67	4.45	0.03	0.73	0.78	0.78
200	10	6	Gaus.	28 (32.01)	100 (2)	100 (2)	100 (2)	92.3	1.38	1.7	4.49	0.28	0.99	0.99	0.99
200	10	8	Gaus.	76 (11.37)	100 (2)	100 (2)	100 (2)	95.31	1.35	1.68	4.47	0.76	1	1	1
200	10	10	Gaus.	81 (9.49)	100 (2)	100 (2)	100 (2)	80.25	1.14	1.5	3.55	0.81	1	1	1
100	100	2	Gaus.	23 (2.99)	0 (1)	0 (1)	0 (5.5)	33.36	1.2	1.38	8.84	0	0	0	0.01
100	100	4	Gaus.	19 (3.53)	0 (1)	0 (1)	0 (5.41)	35.61	1.05	1.23	8.82	0	0	0	0.04
100	100	6	Gaus.	5 (4.47)	83 (1.83)	85 (1.85)	0 (9.19)	33.89	1.43	1.75	11.79	0	0.76	0.78	0.12
100	100	8	Gaus.	17 (4.09)	100 (2)	100 (2)	0 (10)	35.25	1.8	2.15	14.17	0.06	0.99	0.99	0.14
100	100	10	Gaus.	58 (3.2)	100 (2)	100 (2)	0 (10)	35.32	1.7	2.06	10.91	0.53	1	1	0.14
200	100	2	Gaus.	1 (7.08)	0 (1)	0 (1)	0 (1)	111.17	2.99	3.35	20.12	0	0	0	0
200	100	4	Gaus.	0 (7.97)	33 (1.33)	86 (1.86)	0 (2.53)	108.1	4.29	6.42	19.1	0	0.22	0.52	0.02
200	100	6	Gaus.	1 (8.65)	100 (2)	100 (2)	0 (10)	114.39	5.01	5.59	40.72	0	0.92	0.92	0.13
200	100	8	Gaus.	0 (9.72)	100 (2)	100 (2)	0 (10)	113.36	6.28	6.97	32.28	0	0.99	0.99	0.14
200	100	10	Gaus.	52 (5.24)	100 (2)	100 (2)	0 (10)	111.21	5.02	5.58	30.44	0.52	1	1	0.15
100	1000	2	Gaus.	28 (2.23)	0 (1)	0 (1)	–	100.63	7.84	9.18	–	0	0	0	–
100	1000	4	Gaus.	35 (2.12)	0 (1)	0 (1)	–	101.05	8.34	9.72	–	0	0	0	–
100	1000	6	Gaus.	27 (2.14)	0 (1)	0 (1)	–	89.12	8.5	9.73	–	0	0	0	–
100	1000	8	Gaus.	35 (2.4)	2 (1.02)	74 (1.74)	–	107.48	8.22	15.27	–	0	0.02	0.61	–
100	1000	10	Gaus.	17 (3.22)	94 (1.94)	100 (2)	–	100.41	13.61	16.15	–	0.01	0.89	0.94	–
200	1000	2	Gaus.	17 (4.38)	0 (1)	0 (1)	–	377.7	34.79	37.73	–	0	0	0	–
200	1000	4	Gaus.	8 (4.74)	0 (1)	0 (1)	–	390.02	43.52	45.97	–	0	0	0	–
200	1000	6	Gaus.	11 (4.61)	0 (1)	19 (1.19)	–	386.08	40.65	44.54	–	0	0	0.12	–
200	1000	8	Gaus.	4 (5.31)	100 (2)	100 (2)	–	382.49	52.99	57.31	–	0	0.83	0.83	–
200	1000	10	Gaus.	4 (5.88)	100 (2)	100 (2)	–	392.09	56.6	61.28	–	0.02	0.96	0.96	–

Table S10: $K = 3$ simulation results for low-dimensional settings ($p = 10$) with equal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/3, 1/3, 1/3$) and clusters arranged along a line. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	2	Gaus.	0 (31.52)	0 (1.05)	0 (1.19)	0 (1.08)	50.11	0.45	0.57	2.43	0.02	0.02	0.07	0.03
150	4	Gaus.	0 (30.34)	4 (1.2)	59 (2.32)	12 (1.27)	38.69	0.41	0.85	1.95	0.07	0.1	0.57	0.12
150	6	Gaus.	0 (20.69)	20 (1.4)	96 (2.92)	66 (2.32)	38.07	0.39	0.95	2.01	0.26	0.2	0.94	0.65
150	8	Gaus.	1 (7.37)	47 (1.94)	99 (2.98)	93 (2.86)	37.86	0.47	0.97	2.03	0.5	0.47	0.99	0.93
150	10	Gaus.	1 (7.35)	43 (1.86)	100 (3)	98 (2.96)	39.07	0.48	0.89	1.37	0.49	0.43	1	0.98
150	2	t_6	0 (29.59)	0 (1.03)	0 (1.05)	0 (1.14)	51.39	0.45	0.56	2.27	0.01	0.01	0.02	0.02
150	4	t_6	0 (28.86)	2 (1.15)	20 (1.74)	14 (1.53)	51.17	0.45	0.72	2.33	0.05	0.07	0.32	0.15
150	6	t_6	0 (26.87)	14 (1.29)	93 (2.89)	65 (2.31)	42.67	0.31	0.88	1.6	0.13	0.14	0.87	0.61
150	8	t_6	0 (18.25)	39 (1.78)	99 (2.98)	97 (2.94)	41.85	0.31	0.85	1.57	0.3	0.38	0.97	0.95
150	10	t_6	0 (11.21)	40 (1.8)	100 (3)	100 (3)	42.37	0.31	0.85	1.56	0.43	0.4	0.99	0.99
150	2	t_3	0 (26.48)	1 (1.14)	0 (1.16)	0 (4.69)	46.46	0.46	0.56	2.31	0.01	0	0	0.05
150	4	t_3	0 (28.47)	3 (1.18)	7 (1.38)	2 (4.82)	39.27	0.41	0.55	2.14	0.02	0.05	0.11	0.26
150	6	t_3	0 (28.03)	6 (1.24)	48 (2.34)	59 (4.07)	39.69	0.43	0.88	2	0.08	0.09	0.52	0.62
150	8	t_3	2 (25.31)	27 (1.61)	78 (2.94)	76 (4.14)	49.05	0.46	1.16	2.43	0.19	0.28	0.83	0.82
150	10	t_3	1 (19.71)	41 (1.91)	71 (3.16)	67 (5.22)	62.72	0.55	1.38	2.74	0.32	0.42	0.91	0.86
300	2	Gaus.	0 (68.22)	0 (1.29)	0 (1.43)	3 (1.24)	134.97	1.17	1.61	6.19	0.01	0.11	0.16	0.08
300	4	Gaus.	0 (69.98)	64 (2.34)	92 (2.86)	96 (2.97)	134.48	1.99	2.38	6.11	0.02	0.56	0.76	0.8
300	6	Gaus.	0 (50.43)	89 (2.78)	100 (3)	100 (3)	137.22	2.13	2.45	5.98	0.18	0.88	0.98	0.98
300	8	Gaus.	0 (13.7)	97 (2.94)	100 (3)	100 (3)	134.46	2.04	2.35	5.97	0.49	0.97	1	1
300	10	Gaus.	2 (12.18)	95 (2.9)	100 (3)	100 (3)	134.12	2.08	2.37	5.96	0.51	0.95	1	1
300	2	t_6	0 (66.21)	0 (1.1)	0 (1.18)	0 (1.16)	134.69	1.08	1.34	6.24	0.01	0.03	0.06	0.05
300	4	t_6	0 (66.82)	29 (1.77)	66 (2.5)	83 (2.94)	135.17	1.25	2.32	6.16	0.01	0.3	0.55	0.64
300	6	t_6	0 (62.29)	80 (2.6)	100 (3)	100 (3)	133.77	2.03	2.41	6.06	0.08	0.74	0.92	0.92
300	8	t_6	0 (52.95)	94 (2.88)	100 (3)	100 (3)	134.57	2.1	2.39	6.01	0.17	0.92	0.98	0.98
300	10	t_6	0 (21.61)	94 (2.88)	100 (3)	100 (3)	135.46	2.06	2.39	5.99	0.44	0.93	0.99	0.99
300	2	t_3	0 (61.51)	0 (1.06)	1 (1.09)	0 (5.42)	139.72	1.09	1.37	6.32	0.01	0	0	0.08
300	4	t_3	0 (62.93)	5 (1.36)	11 (1.62)	19 (6.35)	136.35	1.09	1.42	6.36	0.01	0.11	0.18	0.39
300	6	t_3	0 (65.05)	45 (2.28)	60 (2.93)	63 (5.5)	135.98	1.97	2.4	6.19	0.03	0.48	0.68	0.74
300	8	t_3	2 (60.44)	73 (2.87)	74 (3.22)	57 (6.01)	135.24	2.08	2.42	6.14	0.07	0.78	0.89	0.81
300	10	t_3	0 (54.43)	82 (2.95)	79 (3.25)	67 (5.31)	134.33	2.09	2.4	6.1	0.15	0.88	0.96	0.89

Table S11: $K = 3$ simulation results for moderate-dimensional settings ($p = 100$) with equal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/3, 1/3, 1/3$) and clusters arranged along a line. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	2	Gaus.	15 (5.26)	0 (1)	0 (1)	0 (1)	64.72	1.74	2.01	11.8	0	0	0	0
150	4	Gaus.	4 (6.9)	0 (1.39)	0 (1.59)	0 (2.35)	66.76	1.92	3.22	13.52	0	0.18	0.27	0.04
150	6	Gaus.	2 (7.83)	29 (1.68)	71 (2.48)	0 (4.78)	64.24	2.57	3.95	13.88	0.02	0.33	0.68	0.16
150	8	Gaus.	2 (8.99)	69 (2.38)	96 (2.92)	0 (8.47)	66.66	3.43	4.27	15.52	0.03	0.68	0.95	0.34
150	10	Gaus.	5 (4.98)	58 (2.16)	97 (2.94)	0 (9.46)	68.9	3.21	4.05	17.53	0.39	0.58	0.97	0.39
150	2	t_6	14 (5.07)	0 (1.01)	0 (1.01)	0 (1.27)	65.18	2.22	2.47	11.7	0	0	0	0
150	4	t_6	6 (6.29)	0 (1.26)	0 (1.4)	0 (1.72)	61.71	1.85	2.62	11.07	0	0.12	0.17	0.02
150	6	t_6	4 (7.38)	5 (1.51)	5 (1.64)	0 (2.71)	60.95	2.57	3.21	11.45	0.01	0.27	0.33	0.08
150	8	t_6	3 (8.12)	53 (2.07)	90 (2.8)	0 (7.48)	68.01	3.06	4.01	15.55	0.03	0.51	0.85	0.34
150	10	t_6	4 (8.59)	63 (2.26)	95 (2.9)	0 (8.74)	55.48	3.25	4.05	13.61	0.05	0.62	0.94	0.41
150	2	t_3	7 (4.18)	16 (1.82)	15 (1.83)	0 (7.93)	69.83	3.37	3.81	24.69	0	0	0	0.01
150	4	t_3	8 (6.06)	9 (1.64)	8 (1.67)	0 (6.4)	70.22	2.6	3.55	22.44	0	0	0	0.16
150	6	t_3	4 (7.12)	5 (1.39)	9 (1.51)	0 (7.66)	67.31	2.53	2.94	23.68	0	0.1	0.12	0.33
150	8	t_3	4 (7.27)	9 (1.48)	17 (1.88)	0 (8.47)	72.23	2.22	3.31	26.22	0.01	0.18	0.32	0.56
150	10	t_3	2 (7.96)	19 (1.78)	41 (2.72)	0 (9.19)	67.73	2.66	4.12	24.31	0.05	0.3	0.62	0.72
300	2	Gaus.	1 (11.35)	0 (1.13)	0 (1.66)	0 (1.01)	202.26	6.49	10.75	31.06	0	0.04	0.15	0
300	4	Gaus.	0 (14.8)	34 (2.2)	85 (2.81)	0 (3.84)	187.38	10.3	12.55	29.28	0	0.43	0.55	0.3
300	6	Gaus.	0 (17.18)	99 (2.98)	100 (3)	0 (10)	185.83	13.9	15.07	35.72	0	0.92	0.92	0.39
300	8	Gaus.	0 (18.76)	100 (3)	100 (3)	1 (9.93)	187.76	12.02	13.13	35.79	0.02	0.99	0.99	0.41
300	10	Gaus.	0 (8.4)	100 (3)	100 (3)	0 (10)	206.64	12.11	13.13	35.84	0.36	1	1	0.41
300	2	t_6	0 (12.34)	0 (1.02)	0 (1.03)	0 (1)	201.26	6.53	6.97	36.88	0	0	0	0
300	4	t_6	0 (13.83)	0 (1.62)	0 (1.68)	0 (1.94)	197.22	10.05	10.79	35.75	0	0.25	0.27	0.25
300	6	t_6	0 (17.13)	85 (2.75)	91 (2.84)	1 (9.77)	200.42	11.63	12.68	66.57	0	0.69	0.73	0.39
300	8	t_6	0 (17.5)	99 (2.98)	100 (3)	0 (10)	200.28	12.98	14.75	66.8	0.01	0.94	0.95	0.46
300	10	t_6	0 (17.83)	99 (2.98)	100 (3)	1 (9.93)	177.04	14.32	16.94	35.01	0.02	0.98	0.99	0.46
300	2	t_3	0 (9.88)	27 (2.32)	27 (2.36)	0 (8.92)	191.15	18.76	20.05	41.04	0	0	0	0.02
300	4	t_3	0 (12.84)	8 (1.91)	8 (1.96)	0 (9.37)	193.62	11.92	12.06	42.83	0	0.06	0.06	0.25
300	6	t_3	0 (14.72)	21 (2.17)	29 (2.3)	0 (9.08)	190.42	10.86	12.19	42.58	0	0.24	0.28	0.48
300	8	t_3	0 (17.11)	27 (3.25)	27 (3.64)	0 (9.91)	217.02	12.81	14.82	46.09	0	0.55	0.63	0.72
300	10	t_3	0 (18.59)	31 (3.87)	30 (4.05)	0 (10)	199.92	17.57	19.12	63.58	0.01	0.85	0.88	0.84

Table S12: $K = 3$ simulation results for high-dimensional settings ($p = 1000$) with equal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/3, 1/3, 1/3$) and clusters arranged along a line. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting (−). (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	4	Gaus.	21 (3.36)	0 (1.01)	0 (1.18)	–	253.74	18.2	20.73	–	0	0.01	0.07	–
150	8	Gaus.	10 (4.69)	7 (1.71)	62 (2.37)	–	207.16	27.13	35.41	–	0	0.36	0.58	–
150	12	Gaus.	7 (6.25)	88 (2.76)	97 (2.94)	–	253.92	31.48	36.01	–	0.02	0.87	0.96	–
150	16	Gaus.	12 (3.93)	74 (2.48)	99 (2.98)	–	213.87	31.87	36.63	–	0.35	0.74	0.99	–
150	20	Gaus.	4 (2.58)	63 (2.26)	94 (2.88)	–	273.94	29.1	35.01	–	0.5	0.63	0.94	–
150	4	t_6	26 (3.09)	0 (1.01)	0 (1.01)	–	166.96	18.1	20	–	0	0	0	–
150	8	t_6	22 (3.8)	0 (1.23)	0 (1.3)	–	165.05	23.27	28.27	–	0	0.12	0.15	–
150	12	t_6	9 (5.21)	63 (2.3)	76 (2.57)	–	252.08	30.89	36.51	–	0.01	0.62	0.74	–
150	16	t_6	9 (6.09)	70 (2.4)	84 (2.68)	–	168.76	32.07	36.2	–	0.01	0.7	0.84	–
150	20	t_6	6 (3.48)	65 (2.3)	90 (2.83)	–	249.68	29.62	36.8	–	0.4	0.65	0.91	–
150	4	t_3	14 (2.04)	22 (2.24)	19 (2.25)	–	245.08	41.82	39.35	–	0	0	0	–
150	8	t_3	19 (2.87)	14 (1.81)	15 (1.85)	–	237.56	31.02	35.1	–	0	0.01	0.01	–
150	12	t_3	22 (3.47)	5 (1.44)	7 (1.54)	–	228.1	25.06	28.55	–	0	0.09	0.13	–
150	16	t_3	14 (4.12)	20 (2)	26 (2.28)	–	217.58	27.96	34.61	–	0.03	0.37	0.46	–
150	20	t_3	11 (4.84)	53 (2.53)	63 (2.97)	–	225.43	31.25	37.44	–	0.11	0.67	0.83	–
300	4	Gaus.	3 (8)	0 (1.63)	0 (1.69)	–	871.51	91.24	97.95	–	0	0.21	0.23	–
300	8	Gaus.	0 (10.13)	96 (2.93)	96 (2.93)	–	710.39	115.55	124.68	–	0	0.79	0.79	–
300	12	Gaus.	1 (12.61)	100 (3)	100 (3)	–	796.69	118.11	125.24	–	0.02	0.99	0.99	–
300	16	Gaus.	0 (8)	100 (3)	100 (3)	–	765.97	161.54	171.5	–	0.27	1	1	–
300	20	Gaus.	5 (3.32)	100 (3)	100 (3)	–	769.39	145.15	161.57	–	0.51	1	1	–
300	4	t_6	1 (7.65)	0 (1.03)	0 (1.01)	–	837.69	71.6	75.6	–	0	0.01	0	–
300	8	t_6	0 (8.84)	22 (2.1)	21 (2.11)	–	878.79	106.56	114.31	–	0	0.45	0.46	–
300	12	t_6	1 (11.4)	100 (3)	100 (3)	–	762.92	113.31	122.7	–	0.01	0.95	0.95	–
300	16	t_6	0 (12.03)	99 (3.01)	99 (3.01)	–	603.21	134.31	139.97	–	0.01	1	1	–
300	20	t_6	1 (7.47)	99 (3.01)	99 (3.01)	–	658.54	117.83	126.54	–	0.31	1	1	–
300	4	t_3	6 (4.69)	17 (3.8)	16 (3.8)	–	662.47	248.46	261.9	–	0	0	0	–
300	8	t_3	10 (5.77)	26 (3.09)	29 (3.19)	–	594.85	138.15	162.77	–	0	0.21	0.22	–
300	12	t_3	3 (7.03)	28 (3.39)	32 (3.49)	–	752.2	140.12	153.14	–	0	0.47	0.5	–
300	16	t_3	1 (7.67)	23 (4.36)	22 (4.45)	–	808.84	180.89	192.34	–	0	0.9	0.9	–
300	20	t_3	2 (8.97)	21 (4.41)	19 (4.5)	–	719.21	176.14	188.4	–	0.1	0.96	0.96	–

Table S13: $K = 3$ simulation results for low-dimensional settings ($p = 10$) with equal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/3, 1/3, 1/3$) and clusters arranged as a triangle. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	2	Gaus.	0 (30.5)	0 (1.07)	0 (1.13)	5 (1.15)	41.85	0.31	0.4	1.62	0.01	0.03	0.05	0.04
150	4	Gaus.	1 (29.17)	21 (1.89)	36 (2.14)	41 (2.19)	42.2	0.52	0.7	1.59	0.08	0.43	0.52	0.53
150	6	Gaus.	11 (16.19)	71 (2.65)	81 (2.78)	79 (2.75)	42.19	0.64	0.87	1.57	0.41	0.81	0.85	0.84
150	8	Gaus.	25 (10.17)	89 (2.87)	94 (2.94)	94 (2.94)	42.04	0.64	0.86	1.56	0.64	0.93	0.96	0.96
150	10	Gaus.	30 (6.65)	89 (2.89)	91 (2.91)	90 (2.92)	42.6	0.65	0.87	1.56	0.68	0.95	0.95	0.95
150	2	t_6	0 (29.67)	0 (1.02)	0 (1.02)	0 (1.18)	42.89	0.31	0.41	1.63	0.01	0	0	0
150	4	t_6	0 (30.07)	12 (1.75)	26 (1.97)	34 (2.12)	42.34	0.51	0.68	1.61	0.06	0.36	0.43	0.47
150	6	t_6	0 (23.87)	57 (2.48)	68 (2.65)	69 (2.72)	42.31	0.63	0.84	1.59	0.19	0.71	0.76	0.76
150	8	t_6	11 (15.13)	82 (2.8)	86 (2.86)	86 (2.94)	42.06	0.64	0.86	1.58	0.43	0.88	0.9	0.9
150	10	t_6	19 (10.19)	90 (2.89)	93 (2.95)	94 (2.94)	42.24	0.64	0.86	1.56	0.57	0.93	0.95	0.95
150	2	t_3	0 (26.75)	1 (1.13)	1 (1.13)	0 (3.25)	42.45	0.31	0.4	1.65	0.01	0	0	0.02
150	4	t_3	0 (28.7)	3 (1.29)	7 (1.45)	8 (5.36)	42.62	0.31	0.42	1.65	0.02	0.09	0.14	0.3
150	6	t_3	1 (24.88)	22 (1.95)	28 (2.25)	26 (4.97)	43.06	0.52	0.69	1.65	0.11	0.41	0.47	0.53
150	8	t_3	2 (22.69)	48 (2.37)	55 (2.75)	55 (4.61)	42.24	0.62	0.85	1.6	0.2	0.63	0.73	0.73
150	10	t_3	7 (17.64)	77 (2.79)	66 (3.04)	66 (4.97)	42.44	0.64	0.86	1.58	0.38	0.83	0.85	0.82
300	2	Gaus.	0 (68.56)	0 (1.16)	1 (1.2)	9 (1.38)	134.2	1.09	1.38	6.2	0.01	0.05	0.06	0.08
300	4	Gaus.	0 (64.76)	54 (2.45)	61 (2.52)	62 (2.57)	134.92	2	2.32	6.1	0.05	0.62	0.63	0.65
300	6	Gaus.	10 (31.91)	90 (2.9)	91 (2.91)	91 (2.91)	137.18	2.09	2.42	6.04	0.42	0.91	0.92	0.92
300	8	Gaus.	21 (21.41)	95 (2.95)	98 (2.98)	96 (2.96)	134.84	2.14	2.44	5.98	0.59	0.96	0.97	0.96
300	10	Gaus.	34 (16.93)	97 (2.97)	97 (2.97)	97 (2.97)	134.06	2.15	2.44	5.96	0.66	0.98	0.98	0.98
300	2	t_6	0 (66.42)	0 (1)	0 (1.02)	6 (1.25)	168.45	1.67	1.9	9.44	0	0	0	0.03
300	4	t_6	0 (64.97)	40 (2.19)	43 (2.27)	53 (2.47)	134.69	2.46	2.84	8.05	0.02	0.48	0.5	0.55
300	6	t_6	1 (54.11)	72 (2.71)	78 (2.78)	81 (2.89)	125.36	2.46	2.81	7.81	0.16	0.79	0.8	0.81
300	8	t_6	7 (38.92)	88 (2.87)	89 (2.89)	90 (2.9)	124.24	2.41	2.69	7.1	0.32	0.9	0.9	0.9
300	10	t_6	26 (24.49)	97 (2.97)	97 (2.97)	97 (2.97)	119.19	2.36	2.62	6.99	0.53	0.95	0.95	0.95
300	2	t_3	0 (57.68)	3 (1.2)	3 (1.21)	0 (4.7)	136.38	1.09	1.4	6.33	0	0	0	0.04
300	4	t_3	0 (62.81)	8 (1.48)	10 (1.63)	16 (5.87)	119.57	1.42	1.85	7.11	0.01	0.17	0.2	0.34
300	6	t_3	0 (62.16)	46 (2.55)	50 (2.73)	44 (5.88)	120.84	1.94	2.31	5.14	0.03	0.59	0.62	0.65
300	8	t_3	0 (57.29)	60 (2.83)	57 (2.89)	52 (5.72)	152.04	2.78	3.33	9.12	0.1	0.76	0.75	0.74
300	10	t_3	5 (49.3)	66 (3.03)	63 (3.22)	49 (6.09)	135.73	2.09	2.45	6.13	0.22	0.84	0.85	0.78

Table S14: $K = 3$ simulation results for moderate-dimensional settings ($p = 100$) with equal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/3, 1/3, 1/3$) and clusters arranged as a triangle. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	2	Gaus.	20 (4.75)	0 (1.01)	0 (1.01)	0 (1)	89.11	1.74	2.07	13.34	0	0	0	0
150	4	Gaus.	11 (6.02)	5 (1.58)	3 (1.61)	0 (5.23)	56.64	2.73	3.2	13.07	0	0.26	0.27	0.13
150	6	Gaus.	3 (6.76)	44 (2.39)	46 (2.42)	0 (9.19)	56.4	3.29	3.85	21.77	0.07	0.67	0.68	0.32
150	8	Gaus.	6 (6.04)	59 (2.59)	62 (2.62)	0 (9.91)	55.97	4.08	4.75	21.42	0.23	0.8	0.81	0.38
150	10	Gaus.	18 (5.39)	81 (2.81)	84 (2.84)	0 (9.91)	56.22	3.36	3.98	21.56	0.42	0.91	0.92	0.39
150	2	t_6	11 (5.75)	0 (1)	0 (1)	0 (1.09)	57.02	2.3	2.65	8.99	0	0	0	0
150	4	t_6	9 (5.61)	0 (1.29)	0 (1.26)	0 (2.53)	56.18	2.52	2.77	12.03	0	0.11	0.1	0.05
150	6	t_6	5 (6.58)	17 (2.06)	21 (2.09)	0 (7.75)	55.07	2.84	3.39	21.37	0.01	0.5	0.51	0.26
150	8	t_6	4 (6.79)	53 (2.47)	53 (2.5)	0 (9.46)	56.58	3.3	3.85	21.85	0.08	0.71	0.72	0.4
150	10	t_6	13 (6.7)	74 (2.74)	73 (2.73)	0 (9.91)	56	3.52	3.97	21.97	0.21	0.86	0.85	0.46
150	2	t_3	5 (4.55)	13 (1.79)	17 (1.83)	0 (6.13)	57.12	3.38	3.88	16.03	0	0	0	0.01
150	4	t_3	3 (4.79)	16 (1.89)	17 (1.91)	0 (6.4)	57.74	3.38	3.83	17.65	0	0	0	0.05
150	6	t_3	4 (6.3)	20 (1.83)	19 (1.9)	0 (8.38)	68.6	3.13	3.89	24.8	0	0.15	0.15	0.36
150	8	t_3	2 (6.98)	31 (2.11)	36 (2.36)	0 (9.46)	64.76	2.96	3.89	22.27	0.03	0.38	0.43	0.57
150	10	t_3	2 (7.18)	38 (2.54)	38 (2.74)	0 (9.73)	88.06	3.37	4.04	29.48	0.05	0.57	0.59	0.65
300	2	Gaus.	1 (11.3)	0 (1)	0 (1.07)	0 (1)	201.71	6.52	6.95	35.77	0	0	0.01	0
300	4	Gaus.	0 (12.68)	28 (2.06)	57 (2.48)	0 (5.1)	195.56	11.92	16.7	39.62	0.01	0.39	0.45	0.26
300	6	Gaus.	0 (15.52)	76 (2.76)	87 (2.87)	0 (8.22)	203.48	12.03	13.11	35.67	0.01	0.8	0.81	0.38
300	8	Gaus.	7 (12.67)	86 (2.86)	90 (2.9)	0 (9.27)	252.56	13.92	14.99	45.22	0.24	0.9	0.9	0.4
300	10	Gaus.	12 (10.49)	92 (2.92)	96 (2.96)	0 (9.83)	194.89	11.95	13	67.29	0.35	0.94	0.95	0.4
300	2	t_6	0 (12.03)	0 (1)	0 (1)	0 (1.18)	277.79	6.46	6.81	38.38	0	0	0	0
300	4	t_6	0 (13.04)	11 (1.71)	12 (1.72)	0 (3.89)	209.07	10.41	11.13	35.94	0	0.24	0.24	0.16
300	6	t_6	0 (14.83)	56 (2.54)	60 (2.59)	0 (7.17)	226.37	11.61	12.76	45.31	0	0.63	0.64	0.36
300	8	t_6	0 (15.21)	78 (2.77)	77 (2.76)	0 (8.92)	223.16	12	12.95	41.45	0.04	0.81	0.8	0.4
300	10	t_6	3 (13.92)	83 (2.85)	84 (2.86)	0 (9.1)	214.09	12.17	13.23	43.04	0.13	0.88	0.88	0.44
300	2	t_3	0 (9.34)	25 (2.42)	25 (2.42)	0 (9.01)	199.65	18.81	20.04	42.67	0	0	0	0.01
300	4	t_3	0 (11.05)	30 (2.28)	32 (2.31)	0 (9.11)	205.86	18.84	20.09	46.54	0	0.03	0.03	0.13
300	6	t_3	0 (14.19)	31 (2.86)	28 (3.05)	0 (9.22)	196.78	14.81	16.55	51.73	0	0.28	0.29	0.39
300	8	t_3	0 (15.12)	32 (3.2)	31 (3.32)	1 (9.45)	201.85	14.16	15.53	57.04	0.01	0.53	0.54	0.6
300	10	t_3	0 (16.51)	34 (3.49)	35 (3.67)	1 (9.93)	199.2	13.91	15.28	61.51	0.02	0.73	0.73	0.75

Table S15: $K = 3$ simulation results for high-dimensional settings ($p = 1000$) with equal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/3, 1/3, 1/3$) and clusters arranged as a triangle. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting (−). (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	4	Gaus.	19 (3.57)	0 (1.01)	0 (1.02)	–	221.07	20.65	23.07	–	0	0	0.01	–
150	8	Gaus.	12 (4.32)	24 (2.09)	45 (2.35)	–	180.93	34.32	42.07	–	0.01	0.53	0.61	–
150	12	Gaus.	16 (4.72)	58 (2.56)	73 (2.72)	–	196.5	32.84	36.96	–	0.14	0.79	0.84	–
150	16	Gaus.	23 (4.4)	82 (2.81)	90 (2.89)	–	194.7	34.69	38.96	–	0.38	0.91	0.93	–
150	20	Gaus.	26 (3.96)	96 (2.96)	97 (2.97)	–	221.87	31.86	36.35	–	0.48	0.98	0.98	–
150	4	t_6	28 (3.16)	0 (1)	0 (1)	–	167.34	23.89	26.58	–	0	0	0	–
150	8	t_6	24 (3.87)	4 (1.55)	3 (1.54)	–	165.14	25.98	29	–	0	0.26	0.26	–
150	12	t_6	19 (4.41)	47 (2.39)	49 (2.4)	–	163.93	37.51	42.22	–	0.01	0.69	0.69	–
150	16	t_6	16 (4.56)	81 (2.8)	81 (2.8)	–	165.66	33.15	36.96	–	0.17	0.89	0.89	–
150	20	t_6	25 (3.86)	84 (2.84)	85 (2.85)	–	164.58	38.5	43.38	–	0.36	0.92	0.92	–
150	4	t_3	20 (2.06)	26 (2.19)	27 (2.22)	–	183.82	35.72	39.5	–	0	0	0	–
150	8	t_3	16 (2.34)	22 (2.09)	25 (2.11)	–	171.07	35.54	43.33	–	0	0.01	0.01	–
150	12	t_3	18 (3.3)	24 (2.33)	23 (2.42)	–	176.81	40.64	47.66	–	0	0.2	0.21	–
150	16	t_3	19 (3.76)	28 (2.44)	32 (2.51)	–	187.08	40.22	45.36	–	0.03	0.5	0.51	–
150	20	t_3	22 (3.8)	54 (2.89)	50 (3.02)	–	188.17	45.05	51.34	–	0.18	0.75	0.75	–
300	4	Gaus.	5 (7.21)	0 (1.09)	0 (1.13)	–	659.73	77.43	89.06	–	0	0.03	0.04	–
300	8	Gaus.	2 (9.7)	49 (2.42)	56 (2.48)	–	657.83	110.35	121.24	–	0.01	0.63	0.64	–
300	12	Gaus.	2 (9.36)	89 (2.89)	89 (2.89)	–	658.55	137.89	147.3	–	0.11	0.91	0.91	–
300	16	Gaus.	8 (7.17)	94 (2.94)	94 (2.94)	–	658.07	124.99	134.47	–	0.35	0.95	0.95	–
300	20	Gaus.	13 (6.56)	92 (2.92)	94 (2.94)	–	696.86	113.87	123.51	–	0.43	0.96	0.97	–
300	4	t_6	6 (6.95)	0 (1)	0 (1)	–	587.49	90.89	95.64	–	0	0	0	–
300	8	t_6	2 (8.13)	28 (2.09)	28 (2.09)	–	757.31	99.74	108.53	–	0	0.43	0.43	–
300	12	t_6	1 (9.22)	74 (2.73)	76 (2.75)	–	802.25	155.95	169.55	–	0.02	0.8	0.8	–
300	16	t_6	1 (9.5)	88 (2.88)	88 (2.88)	–	755.59	117.67	126.03	–	0.09	0.91	0.91	–
300	20	t_6	15 (6.78)	99 (2.99)	99 (2.99)	–	795.03	148.68	159.76	–	0.37	0.97	0.97	–
300	4	t_3	4 (3.61)	33 (3.25)	32 (3.24)	–	790.72	272.77	283.99	–	0	0	0	–
300	8	t_3	7 (4.79)	29 (3.41)	27 (3.45)	–	703.06	213.98	246.24	–	0	0.06	0.06	–
300	12	t_3	6 (5.73)	27 (3.75)	25 (3.85)	–	793.85	191.86	207.88	–	0	0.37	0.37	–
300	16	t_3	1 (7.32)	30 (4.03)	29 (4.24)	–	823.61	169.18	181.83	–	0.02	0.69	0.7	–
300	20	t_3	2 (7.57)	29 (4.21)	21 (4.35)	–	754.8	144.55	164.02	–	0.08	0.82	0.82	–

Table S16: $K = 3$ simulation results with unequal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/5, 2/5, 2/5$) and clusters arranged as a triangle. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting ($-$). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	p	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	10	2	Gaus.	0 (30.91)	0 (1.06)	0 (1.08)	2 (1.09)	50.12	0.44	0.59	2.35	0.01	0.03	0.04	0.04
150	10	4	Gaus.	1 (28.93)	12 (1.78)	26 (2.05)	34 (2.12)	50.43	0.66	0.95	2.43	0.09	0.42	0.52	0.52
150	10	6	Gaus.	12 (15.78)	60 (2.54)	73 (2.7)	74 (2.73)	48.37	0.83	1.13	2.38	0.42	0.79	0.85	0.86
150	10	8	Gaus.	23 (9.98)	86 (2.84)	90 (2.9)	89 (2.88)	47.25	0.84	1.1	2.23	0.62	0.92	0.95	0.94
150	10	10	Gaus.	22 (7.41)	90 (2.89)	94 (2.93)	96 (2.96)	44.67	0.84	1.1	2.17	0.69	0.96	0.97	0.98
300	10	2	Gaus.	0 (67.61)	1 (1.19)	2 (1.25)	8 (1.34)	165.51	1.63	1.89	9.05	0.01	0.08	0.1	0.12
300	10	4	Gaus.	0 (67.46)	41 (2.28)	52 (2.42)	57 (2.56)	166.74	2.52	3	8.92	0.02	0.6	0.63	0.67
300	10	6	Gaus.	9 (38.07)	78 (2.78)	82 (2.82)	82 (2.84)	166.8	2.73	3.25	8.97	0.38	0.88	0.89	0.89
300	10	8	Gaus.	21 (21.76)	86 (2.85)	89 (2.89)	90 (2.9)	182.04	3	3.57	9.65	0.57	0.93	0.94	0.94
300	10	10	Gaus.	26 (10.9)	99 (2.99)	99 (2.99)	98 (2.98)	171.46	3	3.5	9.11	0.7	0.99	0.99	0.98
150	100	2	Gaus.	16 (5.09)	0 (1.01)	0 (1.01)	0 (1)	61.89	1.85	2.07	10.3	0	0	0	0
150	100	4	Gaus.	10 (5.8)	7 (1.65)	5 (1.68)	0 (4.69)	71.57	2.74	3.18	15.04	0	0.32	0.33	0.1
150	100	6	Gaus.	2 (6.75)	26 (2.17)	27 (2.2)	0 (8.65)	68.36	2.88	3.35	21.66	0.04	0.62	0.63	0.26
150	100	8	Gaus.	5 (6.53)	64 (2.62)	70 (2.68)	0 (9.37)	71.4	3.26	3.8	18.42	0.2	0.82	0.84	0.32
150	100	10	Gaus.	9 (5.6)	79 (2.78)	83 (2.82)	0 (9.91)	75.16	3.43	4.08	17.89	0.33	0.9	0.91	0.35
300	100	2	Gaus.	0 (12.56)	0 (1.04)	0 (1.11)	0 (1)	244.17	6.59	7.27	39.89	0	0.01	0.03	0
300	100	4	Gaus.	0 (13.71)	23 (2.04)	41 (2.35)	0 (3.92)	242.96	10.56	12.58	41.9	0	0.43	0.48	0.27
300	100	6	Gaus.	0 (14.76)	78 (2.78)	91 (2.91)	0 (8.28)	242.52	11.77	13.02	49.08	0.03	0.82	0.84	0.36
300	100	8	Gaus.	2 (14.02)	78 (2.78)	89 (2.89)	0 (8.7)	231.3	11.81	13.11	47.58	0.12	0.87	0.89	0.39
300	100	10	Gaus.	12 (10.03)	95 (2.95)	97 (2.97)	0 (9.52)	238.64	13.25	14.31	55.36	0.4	0.96	0.97	0.39
150	1000	4	Gaus.	24 (3.36)	0 (1.01)	0 (1.04)	–	233.03	17.2	19.38	–	0	0	0.02	–
150	1000	8	Gaus.	20 (3.94)	14 (1.83)	25 (2.07)	–	228.16	27.73	34.31	–	0.02	0.44	0.53	–
150	1000	12	Gaus.	18 (4.34)	63 (2.62)	75 (2.75)	–	212.6	31.62	36.15	–	0.06	0.84	0.87	–
150	1000	16	Gaus.	17 (4.42)	83 (2.82)	90 (2.9)	–	214.99	30.08	35.23	–	0.24	0.91	0.94	–
150	1000	20	Gaus.	21 (3.75)	90 (2.9)	95 (2.95)	–	213.61	37.09	41.4	–	0.45	0.97	0.98	–
300	1000	4	Gaus.	4 (7.61)	0 (1.09)	0 (1.1)	–	669.75	71.95	75.21	–	0	0.03	0.04	–
300	1000	8	Gaus.	1 (9.18)	39 (2.33)	45 (2.4)	–	661.61	101.79	110.8	–	0	0.63	0.64	–
300	1000	12	Gaus.	3 (9.18)	78 (2.77)	83 (2.82)	–	865.78	127.34	133.58	–	0.13	0.87	0.88	–
300	1000	16	Gaus.	6 (8.05)	95 (2.95)	96 (2.96)	–	785.7	120.32	127.35	–	0.32	0.97	0.97	–
300	1000	20	Gaus.	20 (5.85)	97 (2.97)	97 (2.97)	–	843.61	112.5	122.43	–	0.5	0.98	0.98	–

Table S17: $K = 3$ simulation results with unequal cluster sizes ($\pi_1, \pi_2, \pi_3 = 1/5, 1/5, 3/5$) and clusters arranged as a triangle. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 3|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting ($-$). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 3\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	p	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
150	10	2	Gaus.	0 (31.14)	0 (1.02)	0 (1.03)	0 (1.02)	51.71	0.45	0.57	2.46	0.01	0.01	0.02	0.01
150	10	4	Gaus.	0 (28.24)	4 (1.69)	17 (1.92)	18 (1.93)	48.6	0.63	0.85	2.28	0.09	0.44	0.52	0.52
150	10	6	Gaus.	14 (14.9)	42 (2.38)	71 (2.69)	65 (2.64)	47.48	0.76	1.08	2.29	0.47	0.8	0.87	0.86
150	10	8	Gaus.	23 (8.48)	68 (2.65)	83 (2.8)	85 (2.85)	53.27	0.84	1.25	2.4	0.65	0.89	0.91	0.93
150	10	10	Gaus.	23 (8.36)	75 (2.69)	92 (2.91)	89 (2.88)	55.54	0.93	1.29	2.5	0.65	0.89	0.95	0.95
300	10	2	Gaus.	0 (69.01)	0 (1.09)	0 (1.11)	5 (1.19)	164.95	1.74	1.95	10.01	0	0.04	0.05	0.06
300	10	4	Gaus.	0 (64.1)	43 (2.27)	53 (2.4)	61 (2.59)	161.21	2.61	2.92	8.22	0.06	0.64	0.68	0.72
300	10	6	Gaus.	5 (38.85)	78 (2.77)	84 (2.83)	80 (2.81)	163.35	2.85	3.27	8.44	0.36	0.87	0.88	0.87
300	10	8	Gaus.	18 (22.37)	90 (2.9)	93 (2.93)	91 (2.91)	168.16	3.04	3.64	9.28	0.59	0.96	0.96	0.96
300	10	10	Gaus.	27 (9.54)	97 (2.96)	99 (2.99)	99 (2.99)	175.66	2.88	3.4	9.07	0.77	0.98	0.99	0.99
150	100	2	Gaus.	7 (5.67)	0 (1)	0 (1)	0 (1)	69.65	2.35	2.61	11.41	0	0	0	0
150	100	4	Gaus.	14 (5.41)	1 (1.61)	1 (1.67)	0 (4.15)	68.92	2.76	3.29	14.82	0	0.34	0.36	0.07
150	100	6	Gaus.	7 (6.49)	33 (2.27)	41 (2.37)	0 (8.74)	69.44	3.72	4.47	18.33	0.08	0.71	0.73	0.21
150	100	8	Gaus.	4 (5.86)	59 (2.57)	69 (2.67)	0 (9.64)	66.55	4.1	4.92	22.91	0.21	0.86	0.87	0.25
150	100	10	Gaus.	20 (5.08)	71 (2.7)	83 (2.83)	0 (9.73)	71.27	3.12	4.18	18.58	0.36	0.9	0.92	0.26
300	100	2	Gaus.	0 (11.88)	0 (1.02)	0 (1.09)	0 (1.09)	237.46	6.65	8	41.46	0	0.01	0.02	0
300	100	4	Gaus.	0 (13.66)	16 (1.95)	28 (2.2)	0 (5.45)	234.95	10.69	12.03	43.22	0	0.45	0.5	0.23
300	100	6	Gaus.	0 (14.41)	61 (2.61)	66 (2.66)	0 (8.38)	242.7	13.73	15	52.71	0.03	0.8	0.81	0.32
300	100	8	Gaus.	4 (12.27)	80 (2.79)	83 (2.83)	0 (9.11)	238.55	14.62	16.08	47.28	0.2	0.89	0.9	0.32
300	100	10	Gaus.	11 (9.82)	89 (2.89)	92 (2.92)	0 (9.68)	227.12	16.12	17.1	53.54	0.35	0.93	0.94	0.3
150	1000	4	Gaus.	19 (3.51)	0 (1.01)	0 (1.02)	–	166.44	18.42	20.17	–	0	0.01	0.01	–
150	1000	8	Gaus.	6 (4.21)	9 (1.77)	18 (2)	–	165.09	26.6	31.24	–	0	0.47	0.56	–
150	1000	12	Gaus.	20 (4.62)	58 (2.53)	68 (2.66)	–	165.73	35.37	40.05	–	0.12	0.83	0.86	–
150	1000	16	Gaus.	20 (3.87)	77 (2.75)	91 (2.9)	–	166.22	33.06	37.86	–	0.27	0.93	0.96	–
150	1000	20	Gaus.	18 (3.6)	86 (2.86)	89 (2.89)	–	165.16	36.72	40.76	–	0.49	0.95	0.96	–
300	1000	4	Gaus.	3 (7.1)	0 (1.13)	0 (1.15)	–	673.4	94.01	98.73	–	0	0.06	0.07	–
300	1000	8	Gaus.	4 (8.89)	31 (2.23)	34 (2.27)	–	891.79	145.25	154.96	–	0	0.62	0.62	–
300	1000	12	Gaus.	3 (9.78)	80 (2.8)	81 (2.81)	–	663.69	136.83	145.92	–	0.07	0.88	0.89	–
300	1000	16	Gaus.	9 (6.93)	92 (2.91)	93 (2.92)	–	660.42	126.95	136.8	–	0.33	0.96	0.96	–
300	1000	20	Gaus.	13 (6.22)	94 (2.94)	95 (2.95)	–	666.59	136.72	145.78	–	0.45	0.97	0.97	–

Table S18: $K = 4$ simulation results for low-dimensional settings ($p = 10$) with equal sized clusters arranged as a tetrahedron. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 4|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 4\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
200	2	Gaus.	0 (42.78)	0 (1.04)	0 (1.04)	3 (1.32)	62.79	0.59	0.7	3.26	0.01	0.01	0.01	0.05
200	4	Gaus.	0 (37.3)	8 (2.19)	20 (2.63)	34 (3.09)	67.86	0.87	1.24	2.73	0.11	0.39	0.5	0.6
200	6	Gaus.	7 (16.73)	55 (3.39)	72 (3.63)	72 (3.69)	67.75	1.08	1.48	2.68	0.45	0.79	0.85	0.87
200	8	Gaus.	14 (8.18)	75 (3.7)	88 (3.88)	84 (3.84)	68.11	1.17	1.5	2.67	0.6	0.9	0.95	0.94
200	10	Gaus.	21 (5.39)	92 (3.88)	94 (3.94)	93 (3.93)	67.66	1.17	1.5	2.67	0.67	0.96	0.98	0.97
200	2	t_6	0 (40.91)	0 (1)	0 (1)	0 (1.13)	71.47	0.52	0.65	2.81	0.01	0	0	0.01
200	4	t_6	0 (39.75)	1 (1.82)	10 (2.13)	19 (2.82)	67.78	0.85	1.05	2.76	0.05	0.26	0.33	0.44
200	6	t_6	1 (32.03)	30 (2.95)	51 (3.3)	57 (3.51)	68.06	1.05	1.35	2.72	0.22	0.64	0.72	0.74
200	8	t_6	3 (14.86)	63 (3.47)	79 (3.74)	77 (3.75)	64.43	1.32	1.68	3.42	0.45	0.81	0.88	0.89
200	10	t_6	13 (9.85)	79 (3.73)	89 (3.9)	92 (3.99)	71.37	1.21	1.66	2.7	0.59	0.9	0.95	0.95
200	2	t_3	0 (36.63)	0 (1.14)	0 (1.16)	0 (3.52)	68.54	0.51	0.62	2.84	0.01	0	0	0.02
200	4	t_3	0 (38.41)	0 (1.34)	0 (1.47)	2 (5.76)	68.33	0.67	0.78	2.82	0.02	0.06	0.1	0.29
200	6	t_3	0 (37.14)	5 (2.2)	19 (2.6)	16 (6.29)	68.91	0.93	1.2	2.79	0.1	0.35	0.43	0.61
200	8	t_3	0 (28.4)	31 (2.98)	42 (3.43)	46 (6.04)	68.5	1.07	1.46	2.74	0.24	0.61	0.69	0.77
200	10	t_3	3 (21.12)	60 (3.54)	70 (4.01)	55 (6.14)	68.88	1.19	1.58	2.74	0.38	0.79	0.88	0.86
400	2	Gaus.	0 (95.23)	0 (1.17)	0 (1.21)	3 (1.9)	224.71	2.79	3.15	13	0.01	0.04	0.04	0.11
400	4	Gaus.	0 (90.11)	38 (3.06)	45 (3.18)	45 (3.29)	224.38	3.6	3.98	10.88	0.05	0.6	0.62	0.64
400	6	Gaus.	5 (44.72)	80 (3.79)	82 (3.81)	82 (3.81)	227.64	3.99	4.44	10.71	0.34	0.88	0.89	0.89
400	8	Gaus.	9 (15.96)	91 (3.87)	92 (3.9)	92 (3.92)	233.6	4.07	4.5	10.61	0.55	0.94	0.95	0.96
400	10	Gaus.	26 (10.05)	95 (3.95)	98 (3.98)	97 (3.97)	202.74	4.52	4.94	12.53	0.7	0.98	0.99	0.98
400	2	t_6	0 (91.25)	0 (1.03)	0 (1.04)	4 (1.45)	226.26	2.05	2.23	11.18	0.01	0	0.01	0.04
400	4	t_6	0 (91.94)	18 (2.53)	28 (2.79)	34 (3.22)	226.15	3.31	3.84	11.03	0.01	0.4	0.45	0.51
400	6	t_6	0 (70.74)	60 (3.47)	67 (3.66)	72 (3.84)	233.42	3.84	4.35	10.82	0.15	0.75	0.79	0.8
400	8	t_6	4 (39.28)	88 (3.84)	89 (3.89)	93 (3.99)	203.3	4.47	4.91	12.59	0.38	0.91	0.92	0.93
400	10	t_6	10 (24.48)	93 (3.9)	98 (3.98)	98 (3.98)	229.91	3.98	4.48	10.7	0.53	0.95	0.97	0.97
400	2	t_3	0 (84.43)	0 (1.12)	0 (1.12)	0 (5.81)	227.01	2.05	2.25	11.35	0	0	0	0.04
400	4	t_3	0 (89.68)	2 (1.68)	3 (1.85)	7 (6.67)	226.48	2.53	3.3	11.22	0.01	0.15	0.17	0.35
400	6	t_3	0 (89.74)	32 (3.03)	31 (3.3)	25 (6.94)	203.75	4.28	4.81	12.96	0.02	0.5	0.53	0.61
400	8	t_3	0 (79.62)	57 (3.72)	56 (3.91)	39 (7.36)	230.28	3.88	4.32	10.97	0.1	0.77	0.78	0.79
400	10	t_3	2 (63.1)	66 (3.93)	65 (4.13)	43 (6.93)	229.28	3.93	4.41	10.85	0.24	0.86	0.87	0.85

Table S19: $K = 4$ simulation results for moderate-dimensional settings ($p = 100$) with equal sized clusters arranged as a tetrahedron. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 4|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 4\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
200	2	Gaus.	5 (7.45)	0 (1)	0 (1.03)	0 (1)	115.86	2.96	3.25	21.03	0	0	0.01	0
200	4	Gaus.	7 (8.54)	1 (1.83)	2 (2.16)	0 (6.85)	111.83	4.75	5.74	24.01	0	0.24	0.31	0.2
200	6	Gaus.	4 (9.65)	20 (2.94)	27 (3.05)	0 (9.82)	103.58	5.93	6.84	22.45	0.04	0.63	0.65	0.46
200	8	Gaus.	4 (9.25)	62 (3.58)	67 (3.65)	0 (10)	113.61	6.51	7.24	25.64	0.2	0.85	0.87	0.52
200	10	Gaus.	11 (6.7)	80 (3.78)	87 (3.87)	0 (10)	109.51	6.43	7.25	23.91	0.46	0.92	0.94	0.54
200	2	t_6	9 (7.56)	0 (1.02)	0 (1.02)	0 (1)	92.64	3.05	3.37	18.33	0	0	0	0
200	4	t_6	7 (7.92)	0 (1.24)	0 (1.24)	0 (4.15)	93.56	3.09	3.38	19.17	0	0.06	0.06	0.1
200	6	t_6	1 (9.75)	7 (2.47)	10 (2.62)	0 (9.37)	95.12	5.7	6.55	34.29	0.01	0.44	0.47	0.43
200	8	t_6	0 (9.79)	38 (3.22)	44 (3.29)	0 (9.91)	94.41	6.06	6.96	34.74	0.06	0.73	0.75	0.53
200	10	t_6	7 (8.42)	61 (3.56)	68 (3.66)	0 (10)	91.66	6.31	7.15	34.67	0.22	0.84	0.87	0.6
200	2	t_3	6 (5.76)	7 (2.11)	7 (2.13)	0 (7.3)	113.49	9.02	9.69	30.1	0	0	0	0.01
200	4	t_3	4 (6.78)	4 (1.85)	3 (1.83)	0 (7.75)	94.87	5.79	6.36	26.96	0	0	0	0.09
200	6	t_3	2 (8.9)	11 (2.04)	10 (2.08)	0 (8.29)	95.09	5.16	6.02	31.04	0	0.12	0.12	0.3
200	8	t_3	1 (9.88)	9 (2.57)	16 (2.75)	0 (9.65)	97.34	6.49	7.23	34.78	0.01	0.36	0.37	0.58
200	10	t_3	2 (10.13)	31 (3.18)	33 (3.36)	0 (10)	96.33	7.73	8.42	35.71	0.05	0.61	0.63	0.73
400	2	Gaus.	0 (16.52)	0 (1.01)	0 (1.15)	0 (1.01)	331.97	12.93	13.51	57.21	0	0	0.01	0
400	4	Gaus.	0 (18.5)	16 (2.62)	31 (3.07)	0 (6.36)	336.06	22.06	24	100.5	0	0.4	0.44	0.32
400	6	Gaus.	0 (21.31)	55 (3.51)	80 (3.79)	0 (9.19)	469.93	24.08	26.06	132.4	0.05	0.75	0.78	0.5
400	8	Gaus.	2 (20.17)	85 (3.85)	94 (3.94)	0 (9.35)	306.94	22.31	24.54	56.85	0.12	0.91	0.92	0.54
400	10	Gaus.	5 (12)	94 (3.94)	99 (3.99)	0 (9.86)	460.19	25.91	27.32	130.96	0.41	0.96	0.96	0.56
400	2	t_6	0 (17.11)	0 (1.02)	0 (1.02)	0 (1)	472.9	12.86	13.28	73.84	0	0	0	0
400	4	t_6	0 (18.54)	1 (1.81)	3 (1.93)	0 (4.76)	467.89	19.05	20.88	74.05	0	0.18	0.2	0.19
400	6	t_6	0 (20.79)	43 (3.27)	40 (3.25)	0 (7.72)	318.59	22.96	24.48	58.12	0	0.6	0.6	0.44
400	8	t_6	0 (20.9)	63 (3.59)	64 (3.62)	1 (9.07)	322.19	23.33	24.88	59.43	0.05	0.79	0.79	0.57
400	10	t_6	0 (18.01)	83 (3.8)	84 (3.86)	0 (9.56)	321.42	23.74	25.39	59.03	0.18	0.89	0.9	0.59
400	2	t_3	0 (14.7)	17 (2.65)	18 (2.67)	0 (9.37)	359.93	35.55	37.04	67.36	0	0	0	0
400	4	t_3	0 (14.27)	6 (2.19)	7 (2.25)	0 (9.48)	314.18	25.93	27.17	56.33	0	0.02	0.02	0.12
400	6	t_3	0 (19.87)	26 (3.03)	26 (3.26)	0 (9.85)	309.81	28.64	31.07	59.14	0	0.22	0.24	0.36
400	8	t_3	0 (22.08)	34 (3.82)	29 (4.01)	0 (9.77)	313.05	24.8	26.48	60.91	0.01	0.52	0.52	0.59
400	10	t_3	0 (22.8)	39 (4.57)	34 (4.85)	0 (9.93)	315.53	29.11	30.68	62	0.01	0.71	0.72	0.76

Table S20: $K = 4$ simulation results for high-dimensional settings ($p = 1000$) with equal sized clusters arranged as a tetrahedron. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 4|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), SHC_L, SHC_2, and BootClust (BC). Results for `pvclust` BP p-values are omitted since no significant clusters were identified across any settings. BootClust results are not reported for the high-dimensional setting (−). (N: sample size, δ: separating signal between clusters, F: cluster distribution.)

parameters			$\lvert\hat{K} = 4\rvert$ (mean $\hat{K}$)				median time (sec.)				mean ARI			
N	δ	F	pvAU	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	SHC_L	SHC_2	BC
200	4	Gaus.	24 (4.64)	0 (1.03)	0 (1.08)	–	287.59	31.64	34.49	–	0	0.01	0.02	–
200	8	Gaus.	17 (5.88)	8 (2.5)	24 (2.82)	–	287.68	60.91	68.5	–	0	0.49	0.55	–
200	12	Gaus.	5 (7.59)	58 (3.51)	72 (3.7)	–	301.51	57.63	64.69	–	0.11	0.83	0.87	–
200	16	Gaus.	12 (5.72)	85 (3.84)	90 (3.9)	–	288.94	66.91	74.58	–	0.35	0.94	0.95	–
200	20	Gaus.	20 (4.66)	88 (3.85)	94 (3.93)	–	287.86	61.05	67.74	–	0.56	0.95	0.97	–
200	4	t_6	15 (4.71)	0 (1.02)	0 (1.02)	–	295.92	31.52	33.61	–	0	0	0	–
200	8	t_6	15 (5.17)	1 (1.71)	1 (1.75)	–	292.58	50.82	63.25	–	0	0.2	0.22	–
200	12	t_6	12 (6.39)	27 (3.06)	30 (3.15)	–	285.55	55.22	63.07	–	0.02	0.68	0.7	–
200	16	t_6	8 (6.4)	76 (3.74)	76 (3.74)	–	288.51	61.74	68.94	–	0.19	0.9	0.9	–
200	20	t_6	13 (5.57)	86 (3.84)	87 (3.85)	–	301.24	66.21	73.03	–	0.37	0.95	0.95	–
200	4	t_3	12 (2.63)	15 (2.75)	14 (2.71)	–	290.49	83.68	88.93	–	0	0	0	–
200	8	t_3	19 (3.04)	9 (2.43)	11 (2.44)	–	290.43	75.93	79.17	–	0	0.02	0.01	–
200	12	t_3	15 (4.12)	14 (2.83)	13 (2.89)	–	290.84	79.15	84.67	–	0	0.26	0.26	–
200	16	t_3	23 (4.98)	29 (3.13)	27 (3.18)	–	291.32	71.36	78.44	–	0.05	0.5	0.51	–
200	20	t_3	16 (5.29)	37 (3.58)	35 (3.77)	–	293.21	67.61	74	–	0.17	0.66	0.69	–
400	4	Gaus.	1 (9.86)	0 (1.12)	0 (1.14)	–	1301.21	118.97	126.88	–	0	0.02	0.02	–
400	8	Gaus.	0 (13.31)	36 (3.23)	44 (3.34)	–	1307.24	221.41	239.46	–	0.01	0.62	0.64	–
400	12	Gaus.	1 (13.45)	79 (3.79)	81 (3.81)	–	1304.33	298.74	311.4	–	0.1	0.89	0.9	–
400	16	Gaus.	4 (10.48)	93 (3.91)	97 (3.97)	–	1307.39	230.55	246.18	–	0.35	0.95	0.97	–
400	20	Gaus.	12 (7.37)	98 (3.96)	98 (3.96)	–	1301.7	237.29	251.45	–	0.5	0.98	0.98	–
400	4	t_6	1 (9.84)	0 (1.01)	0 (1.02)	–	1300.13	117.12	122.58	–	0	0	0	–
400	8	t_6	1 (10.88)	19 (2.78)	21 (2.79)	–	1161.19	258.65	271	–	0	0.45	0.45	–
400	12	t_6	1 (12.38)	47 (3.38)	48 (3.39)	–	1156.08	232.06	245.57	–	0.02	0.73	0.74	–
400	16	t_6	0 (11.61)	84 (3.84)	83 (3.83)	–	1170.46	232.33	244.81	–	0.15	0.91	0.91	–
400	20	t_6	3 (10.3)	88 (3.9)	88 (3.9)	–	1291.06	231.1	245.73	–	0.31	0.95	0.95	–
400	4	t_3	4 (5.86)	21 (4.35)	19 (4.36)	–	1318.48	553.35	571.23	–	0	0	0	–
400	8	t_3	3 (6.66)	19 (3.94)	21 (4)	–	1318.62	443.06	464.14	–	0	0.05	0.06	–
400	12	t_3	2 (9.21)	27 (4.32)	24 (4.44)	–	1297.52	332.13	354.86	–	0	0.43	0.44	–
400	16	t_3	1 (10.07)	23 (5.28)	19 (5.41)	–	1304.03	294.21	307.06	–	0.02	0.71	0.71	–
400	20	t_3	0 (11.16)	19 (5.5)	18 (5.7)	–	1211.91	249.4	269.83	–	0.09	0.85	0.85	–

Table S21: $K = 5$ simulation results with equal sized clusters. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 5|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), BP p-values (`pvBP`), SHC_L, SHC_2, and BootClust (BC). BootClust results are not reported for the high-dimensional setting (−). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 5\rvert$ (mean $\hat{K}$)					median time (sec.)				mean ARI				
N	p	δ	F	pvAU	pvBP	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	pvBP	SHC_L	SHC_2	BC
250	10	4	Gaus.	0 (52.42)	0 (1)	0 (2.24)	2 (2.62)	13 (3.52)	119.44	1.58	2.17	5.61	0.06	0	0.31	0.38	0.51
250	10	8	Gaus.	0 (8.13)	0 (1)	44 (4.28)	66 (4.65)	66 (4.65)	118.38	2.07	2.88	5.58	0.5	0	0.83	0.9	0.9
250	10	12	Gaus.	10 (5.59)	0 (1)	80 (4.7)	89 (4.87)	92 (4.92)	117.61	2.12	2.74	5.37	0.62	0	0.93	0.97	0.97
250	10	16	Gaus.	19 (3.85)	0 (1.01)	93 (4.83)	99 (4.99)	99 (4.99)	120.89	2.49	3.02	6.63	0.67	0	0.96	1	1
250	10	20	Gaus.	24 (3.97)	0 (1.01)	94 (4.84)	99 (4.99)	99 (4.99)	117.49	2.09	2.97	5.46	0.68	0.03	0.96	1	1
250	10	4	t_6	0 (54.12)	0 (1)	0 (1.82)	0 (2.08)	11 (3.12)	123.53	1.44	1.95	5.86	0.03	0	0.21	0.26	0.38
250	10	8	t_6	0 (17.56)	0 (1)	34 (3.82)	57 (4.4)	58 (4.69)	120.43	2.21	2.85	6.03	0.38	0	0.72	0.83	0.84
250	10	12	t_6	4 (7.43)	0 (1)	77 (4.69)	88 (4.84)	90 (4.89)	123.55	2.64	3.24	6.79	0.53	0	0.92	0.95	0.96
250	10	16	t_6	7 (4.54)	0 (1)	81 (4.74)	93 (4.93)	92 (4.92)	118.55	2.09	2.95	5.32	0.6	0	0.94	0.98	0.98
250	10	20	t_6	13 (3.89)	0 (1)	92 (4.89)	98 (4.98)	98 (4.98)	123.63	2.44	3.11	6	0.67	0.01	0.97	0.99	0.99
250	10	4	t_3	0 (51.07)	0 (1)	0 (1.32)	0 (1.35)	0 (6.3)	121.06	1.22	1.4	5.88	0.02	0	0.06	0.06	0.29
250	10	8	t_3	0 (42.22)	0 (1)	4 (2.73)	20 (3.43)	22 (7)	108.62	1.68	2.18	4.45	0.2	0	0.43	0.53	0.72
250	10	12	t_3	0 (18.34)	0 (1)	35 (3.98)	48 (4.6)	43 (6.93)	123.86	2.09	3	5.79	0.43	0	0.72	0.8	0.86
250	10	16	t_3	4 (10.53)	0 (1)	68 (4.63)	66 (5.11)	57 (6.79)	122.74	2.3	3.1	5.9	0.55	0	0.89	0.94	0.93
250	10	20	t_3	5 (7.89)	0 (1)	82 (4.81)	71 (5.13)	62 (6.84)	122.67	2.72	3.36	6.81	0.6	0	0.93	0.95	0.96
250	100	4	Gaus.	2 (11.12)	0 (1)	0 (2.07)	0 (2.46)	0 (5.75)	171.68	9.36	11.27	33.04	0	0	0.23	0.29	0.23
250	100	8	Gaus.	1 (10.73)	0 (1)	20 (3.92)	28 (4.08)	0 (9.92)	161.24	10.52	11.72	33.63	0.17	0	0.74	0.76	0.6
250	100	12	Gaus.	7 (5.9)	0 (1)	64 (4.63)	71 (4.7)	0 (10)	162.91	10.29	11.62	35.46	0.51	0	0.91	0.92	0.67
250	100	16	Gaus.	14 (4.28)	0 (1)	85 (4.78)	91 (4.91)	0 (10)	169.22	9.48	10.57	36.43	0.64	0	0.95	0.98	0.68
250	100	20	Gaus.	30 (4.17)	0 (1)	95 (4.92)	97 (4.97)	0 (10)	161.5	13.31	14.39	33.58	0.72	0	0.98	0.99	0.69
250	100	4	t_6	1 (9.96)	0 (1)	0 (1.4)	0 (1.41)	0 (3.42)	165.22	5.07	5.73	28.2	0	0	0.08	0.08	0.08
250	100	8	t_6	2 (13.11)	0 (1)	5 (3.48)	10 (3.59)	0 (9.76)	167.99	9.08	10.09	35.76	0.04	0	0.62	0.64	0.58
250	100	12	t_6	5 (8.69)	0 (1)	35 (4.24)	43 (4.4)	0 (10)	164	10.96	12.3	36.17	0.31	0	0.82	0.85	0.71
250	100	16	t_6	7 (4.7)	0 (1)	80 (4.73)	82 (4.8)	0 (10)	147.08	9.83	11.15	32.37	0.52	0	0.93	0.95	0.73
250	100	20	t_6	13 (5.07)	0 (1)	83 (4.71)	89 (4.87)	0 (10)	169.32	11.78	13.48	38.56	0.62	0	0.93	0.97	0.76
250	100	4	t_3	1 (8.28)	0 (1)	3 (2.2)	3 (2.18)	0 (7.94)	173.71	9.06	9.74	42.47	0	0	0	0	0.08
250	100	8	t_3	0 (12.22)	0 (1)	5 (2.77)	5 (2.94)	0 (9.67)	172.02	9.92	10.57	44.76	0.01	0	0.3	0.33	0.51
250	100	12	t_3	0 (13.32)	0 (1)	22 (3.91)	25 (4.21)	0 (10)	174	10.59	11.86	43.66	0.09	0	0.64	0.66	0.76
250	100	16	t_3	2 (11.09)	0 (1)	46 (4.55)	48 (4.94)	0 (10)	130.43	13.15	13.93	30.11	0.29	0	0.8	0.84	0.88
250	100	20	t_3	4 (8.64)	0 (1)	57 (4.66)	46 (5.19)	0 (10)	136.92	9.88	11.21	30.94	0.44	0	0.84	0.9	0.92
250	1000	5	Gaus.	18 (6.03)	0 (1)	0 (1.3)	0 (1.36)	−	571.44	66.82	72.59	−	0	0	0.06	0.07	−
250	1000	10	Gaus.	10 (8.56)	0 (1)	4 (3.3)	13 (3.69)	−	605.9	92.05	102.7	−	0.02	0	0.58	0.64	−
250	1000	15	Gaus.	3 (7.67)	0 (1)	53 (4.39)	69 (4.67)	−	623.62	111.32	125.83	−	0.22	0	0.86	0.9	−
250	1000	20	Gaus.	13 (5.44)	0 (1)	79 (4.75)	86 (4.84)	−	611.89	94.44	102.01	−	0.51	0	0.94	0.95	−
250	1000	25	Gaus.	16 (4.23)	0 (1)	83 (4.74)	91 (4.88)	−	661.95	94.67	104.36	−	0.58	0	0.94	0.96	−
250	1000	5	t_6	15 (5.49)	0 (1)	0 (1.03)	0 (1.03)	−	622.27	46.42	48.78	−	0	0	0.01	0.01	−
250	1000	10	t_6	10 (6.83)	0 (1)	1 (2.62)	2 (2.71)	−	587.17	84.05	90.13	−	0	0	0.41	0.42	−
250	1000	15	t_6	8 (7.93)	0 (1)	25 (3.92)	31 (4)	−	599.97	92.61	100.2	−	0.08	0	0.75	0.76	−
250	1000	20	t_6	9 (6.41)	0 (1)	58 (4.5)	60 (4.52)	−	729.29	86.43	94.67	−	0.31	0	0.88	0.88	−
250	1000	25	t_6	13 (5.39)	0 (1)	81 (4.77)	84 (4.83)	−	607.59	94.81	103.58	−	0.51	0	0.94	0.96	−
250	1000	5	t_3	9 (3.49)	0 (1)	13 (3.21)	13 (3.26)	−	678.42	148.44	159.03	−	0	0	0	0	−
250	1000	10	t_3	14 (4.33)	0 (1)	6 (2.84)	6 (2.87)	−	699.39	140.13	149.98	−	0	0	0.1	0.11	−
250	1000	15	t_3	20 (5.73)	0 (1)	13 (3.62)	10 (3.68)	−	687.27	115.76	124.45	−	0.02	0	0.4	0.4	−
250	1000	20	t_3	11 (6.6)	0 (1)	29 (4.18)	29 (4.32)	−	680.23	105.84	116.26	−	0.11	0	0.62	0.63	−
250	1000	25	t_3	17 (6.26)	0 (1)	29 (4.57)	34 (4.66)	−	622.88	112.8	123.2	−	0.27	0	0.74	0.75	−

Table S22: $K = 6$ simulation results with equal sized clusters. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 6|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), BP p-values (`pvBP`), SHC_L, SHC_2, and BootClust (BC). BootClust results are not reported for the high-dimensional setting (−). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 6\rvert$ (mean $\hat{K}$)					median time (sec.)				mean ARI				
N	p	δ	F	pvAU	pvBP	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	pvBP	SHC_L	SHC_2	BC
300	10	4	Gaus.	0 (64.17)	0 (1)	0 (2.46)	1 (2.95)	13 (4.22)	140.26	1.98	2.48	6.1	0.05	0	0.31	0.36	0.52
300	10	8	Gaus.	2 (6.73)	0 (1)	55 (5.26)	78 (5.72)	80 (5.78)	145.95	2.68	3.42	6.75	0.47	0	0.86	0.93	0.94
300	10	12	Gaus.	10 (4.25)	0 (1)	91 (5.79)	98 (5.94)	99 (5.99)	166.73	3.21	4.17	8.32	0.65	0	0.96	0.98	0.99
300	10	16	Gaus.	19 (4.5)	0 (1)	92 (5.82)	97 (5.97)	98 (5.98)	170.82	3.26	4.32	8.62	0.71	0.02	0.97	1	1
300	10	20	Gaus.	24 (4.67)	0 (1.17)	96 (5.88)	99 (5.99)	99 (5.99)	172.88	3.29	4.19	8.4	0.74	0.14	0.98	1	1
300	10	4	t_6	0 (66.27)	0 (1)	0 (1.98)	0 (2.33)	12 (4.1)	163.89	1.88	2.51	9.92	0.03	0	0.19	0.24	0.42
300	10	8	t_6	0 (25.18)	0 (1)	20 (4.43)	50 (5.16)	64 (5.69)	180.95	3.74	4.47	9.84	0.4	0	0.7	0.8	0.87
300	10	12	t_6	2 (5.17)	0 (1)	73 (5.55)	93 (5.89)	95 (5.95)	170.98	3.96	4.58	9.81	0.49	0	0.91	0.96	0.98
300	10	16	t_6	4 (4.11)	0 (1)	89 (5.81)	96 (5.94)	98 (6.03)	174.05	4.09	4.74	9.83	0.55	0	0.96	0.98	0.99
300	10	20	t_6	14 (4.06)	0 (1)	91 (5.8)	96 (5.98)	96 (6.01)	175.32	4.04	4.68	9.88	0.65	0.01	0.96	0.99	0.99
300	10	4	t_3	0 (62.07)	0 (1)	0 (1.39)	0 (1.53)	0 (7.37)	176.01	1.75	2.2	9.17	0.02	0	0.06	0.07	0.3
300	10	8	t_3	0 (47.88)	0 (1)	9 (3.74)	26 (4.49)	20 (7.54)	178.72	2.91	3.87	9.36	0.2	0	0.54	0.61	0.77
300	10	12	t_3	2 (22.8)	0 (1)	42 (5.13)	62 (5.9)	55 (7.55)	142	2.99	3.52	6.04	0.48	0	0.8	0.89	0.92
300	10	16	t_3	2 (9.71)	0 (1)	71 (5.64)	70 (6.04)	66 (7.26)	169.76	3.62	4.25	9.69	0.54	0	0.9	0.94	0.96
300	10	20	t_3	8 (7.11)	0 (1)	84 (5.61)	66 (6.27)	56 (7.76)	185.79	3.89	4.66	9.87	0.61	0	0.92	0.97	0.98
300	100	4	Gaus.	1 (13.28)	0 (1)	0 (2.24)	0 (2.85)	0 (6.78)	216.47	11.25	12.99	40.72	0	0	0.21	0.28	0.27
300	100	8	Gaus.	2 (12.3)	0 (1)	24 (4.96)	32 (5.09)	0 (10)	226.32	13.92	15.38	46.67	0.22	0	0.8	0.81	0.73
300	100	12	Gaus.	10 (5.79)	0 (1)	78 (5.76)	83 (5.83)	0 (10)	224.3	13.51	14.99	48.97	0.57	0	0.96	0.97	0.78
300	100	16	Gaus.	14 (4.67)	0 (1)	96 (5.89)	96 (5.92)	0 (10)	221.38	14.66	16.19	47.3	0.66	0	0.98	0.98	0.78
300	100	20	Gaus.	24 (4.43)	0 (1)	94 (5.9)	98 (5.98)	0 (10)	182.94	13.78	15.34	34.91	0.72	0	0.98	1	0.79
300	100	4	t_6	3 (12.57)	0 (1)	0 (1.46)	0 (1.49)	0 (3.27)	241.71	9.15	10.19	43.89	0	0	0.07	0.07	0.09
300	100	8	t_6	0 (17.31)	0 (1)	7 (4.13)	14 (4.4)	0 (9.84)	237.17	16.45	17.93	78.29	0.03	0	0.63	0.65	0.68
300	100	12	t_6	4 (9.41)	0 (1)	49 (5.3)	58 (5.44)	0 (10)	242.79	14.68	16.44	80.12	0.4	0	0.87	0.89	0.8
300	100	16	t_6	5 (5.74)	0 (1)	84 (5.79)	88 (5.86)	0 (10)	253.98	16.91	18.52	78.95	0.57	0	0.96	0.97	0.83
300	100	20	t_6	8 (4.84)	0 (1)	95 (5.82)	98 (5.92)	0 (10)	201.5	14.06	15.39	39.42	0.63	0	0.96	0.98	0.83
300	100	4	t_3	2 (10.65)	0 (1)	0 (2.08)	0 (2.07)	0 (9.1)	245.4	12.78	13.8	51.94	0	0	0	0	0.08
300	100	8	t_3	0 (16.16)	0 (1)	3 (2.98)	2 (3.18)	0 (9.84)	250.7	13.66	14.85	73.51	0.01	0	0.29	0.3	0.52
300	100	12	t_3	0 (16.24)	0 (1)	16 (4.64)	19 (5.03)	0 (10)	242.7	14.5	15.97	54.93	0.07	0	0.64	0.67	0.8
300	100	16	t_3	1 (13.21)	0 (1)	39 (5.52)	43 (6.11)	0 (10)	199.2	16.67	17.98	41.78	0.32	0	0.83	0.87	0.92
300	100	20	t_3	6 (9)	0 (1)	62 (5.76)	46 (6.37)	0 (10)	200.8	14.3	16	41.71	0.51	0	0.9	0.93	0.95
300	1000	5	Gaus.	10 (7.89)	0 (1)	0 (1.37)	0 (1.41)	–	933.71	96.03	102.6	–	0	0	0.06	0.06	–
300	1000	10	Gaus.	5 (10.79)	0 (1)	7 (4.26)	23 (4.78)	–	950.41	146.17	157.97	–	0.02	0	0.65	0.71	–
300	1000	15	Gaus.	5 (8.1)	0 (1)	63 (5.58)	76 (5.76)	–	1014.4	136.28	146.92	–	0.33	0	0.91	0.93	–
300	1000	20	Gaus.	12 (6.01)	0 (1)	85 (5.85)	91 (5.91)	–	866.43	142.45	155.67	–	0.59	0	0.97	0.98	–
300	1000	25	Gaus.	15 (4.85)	0 (1)	93 (5.93)	96 (5.96)	–	675.07	138.29	149.48	–	0.67	0	0.99	0.99	–
300	1000	5	t_6	14 (7.39)	0 (1)	0 (1.02)	0 (1.02)	–	921.02	71.71	75.68	–	0	0	0	0	–
300	1000	10	t_6	7 (8.9)	0 (1)	0 (3.2)	0 (3.21)	–	696.63	114.13	123.69	–	0	0	0.42	0.42	–
300	1000	15	t_6	5 (10.45)	0 (1)	25 (4.99)	28 (5)	–	748.48	129.67	140.22	–	0.13	0	0.8	0.8	–
300	1000	20	t_6	8 (7.5)	0 (1)	57 (5.42)	59 (5.5)	–	868.52	131.78	143.09	–	0.4	0	0.89	0.9	–
300	1000	25	t_6	12 (5.1)	0 (1)	80 (5.73)	84 (5.81)	–	942.12	130.85	142.88	–	0.56	0	0.95	0.96	–
300	1000	5	t_3	11 (3.99)	0 (1)	3 (3.26)	3 (3.26)	–	951.81	194.75	211.01	–	0	0	0	0	–
300	1000	10	t_3	18 (5.2)	0 (1)	7 (3.04)	4 (3.03)	–	949.23	193.98	200.28	–	0	0	0.09	0.09	–
300	1000	15	t_3	15 (8.44)	0 (1)	4 (3.88)	8 (3.97)	–	921.06	150.42	161.42	–	0.01	0	0.39	0.4	–
300	1000	20	t_3	9 (8.76)	0 (1)	14 (4.82)	22 (5.12)	–	837.73	163.08	182.24	–	0.1	0	0.64	0.67	–
300	1000	25	t_3	12 (8.21)	0 (1)	34 (5.31)	38 (5.43)	–	917.31	149.73	162.98	–	0.28	0	0.77	0.79	–

Table S23: $K = 7$ simulation results with equal sized clusters. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 7|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), BP p-values (`pvBP`), SHC_L, SHC_2, and BootClust (BC). BootClust results are not reported for the high-dimensional setting (−). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 7\rvert$ (mean $\hat{K}$)					median time (sec.)				mean ARI				
N	p	δ	F	pvAU	pvBP	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	pvBP	SHC_L	SHC_2	BC
350	10	4	Gaus.	0 (79.26)	0 (1)	0 (2.99)	2 (3.75)	25 (5.56)	238.6	4.15	5	13.28	0.05	0	0.34	0.42	0.59
350	10	8	Gaus.	0 (7.02)	0 (1)	56 (6.23)	84 (6.79)	90 (6.9)	238.92	4.96	6.08	12.22	0.45	0	0.87	0.95	0.96
350	10	12	Gaus.	7 (4.68)	0 (1)	90 (6.7)	99 (6.96)	100 (7)	248.55	5.19	6.16	13.27	0.63	0	0.95	0.99	1
350	10	16	Gaus.	13 (4.98)	0 (1.12)	97 (6.9)	100 (7)	100 (7)	243.66	5.11	6.08	12.81	0.7	0.08	0.99	1	1
350	10	20	Gaus.	33 (5.54)	0 (1.71)	95 (6.86)	100 (7)	100 (7)	227.67	4.93	5.79	11.15	0.78	0.27	0.98	1	1
350	10	4	t_6	0 (78.96)	0 (1)	0 (2.01)	0 (2.42)	11 (4.76)	248.56	3.83	4.53	13.39	0.03	0	0.17	0.21	0.41
350	10	8	t_6	2 (20.95)	0 (1)	25 (5.44)	61 (6.41)	72 (6.76)	233.66	4.53	5.71	12.87	0.39	0	0.75	0.87	0.91
350	10	12	t_6	1 (4.85)	0 (1)	83 (6.77)	96 (6.97)	99 (6.99)	247.27	5.36	6.23	13.18	0.5	0	0.96	0.98	0.99
350	10	16	t_6	1 (5.65)	0 (1)	91 (6.68)	96 (6.89)	100 (7)	244.72	5.45	6.25	13.13	0.59	0	0.95	0.98	1
350	10	20	t_6	13 (4.75)	0 (1.01)	96 (6.83)	100 (7)	100 (7)	236.56	5.25	6.07	12.97	0.64	0.03	0.97	1	1
350	10	4	t_3	0 (75.42)	0 (1)	0 (1.35)	0 (1.46)	1 (7.21)	242.54	2.42	2.77	12.42	0.02	0	0.04	0.05	0.3
350	10	8	t_3	0 (63.18)	0 (1)	7 (4.3)	23 (5.33)	20 (8.4)	242.14	4.54	5.43	13.09	0.2	0	0.54	0.64	0.79
350	10	12	t_3	0 (25)	0 (1)	29 (5.62)	55 (6.74)	52 (8.4)	234.65	5	5.84	13.14	0.47	0	0.76	0.87	0.93
350	10	16	t_3	0 (9.12)	0 (1)	72 (6.51)	67 (7.3)	55 (8.35)	246.48	5.04	6.05	12.79	0.57	0	0.91	0.97	0.97
350	10	20	t_3	1 (6.63)	0 (1)	79 (6.63)	65 (7.27)	53 (8.37)	240.94	5.07	6.14	13.02	0.56	0	0.93	0.96	0.98
350	100	4	Gaus.	0 (16.54)	0 (1)	0 (2.46)	0 (3.27)	0 (7.7)	324.38	17.23	21.59	76.32	0	0	0.21	0.28	0.29
350	100	8	Gaus.	1 (14.02)	0 (1)	26 (5.91)	37 (6.17)	0 (10)	319.99	20.6	22.76	98.47	0.24	0	0.82	0.85	0.82
350	100	12	Gaus.	4 (5.39)	0 (1)	90 (6.73)	93 (6.92)	0 (10)	328.11	26.49	28.89	67.84	0.55	0	0.96	0.98	0.86
350	100	16	Gaus.	14 (5.37)	0 (1)	94 (6.94)	98 (6.98)	0 (10)	306.18	19.2	21.07	62.88	0.71	0	0.99	1	0.86
350	100	20	Gaus.	20 (5.5)	0 (1)	98 (6.91)	99 (6.99)	0 (10)	320.68	22.76	24.66	102.17	0.74	0.01	0.99	1	0.86
350	100	4	t_6	0 (15.4)	0 (1)	0 (1.43)	0 (1.44)	0 (3.55)	302.96	14.11	14.74	56.09	0	0	0.05	0.05	0.1
350	100	8	t_6	0 (19.08)	0 (1)	5 (4.9)	7 (5.13)	0 (10)	305.46	22.67	24.61	66.65	0.07	0	0.64	0.66	0.75
350	100	12	t_6	3 (10.28)	0 (1)	67 (6.49)	76 (6.69)	0 (10)	321.59	19.93	21.76	100.38	0.4	0	0.91	0.94	0.88
350	100	16	t_6	8 (5.31)	0 (1)	86 (6.72)	92 (6.91)	0 (10)	326.31	24.02	26.72	102.9	0.56	0	0.95	0.98	0.9
350	100	20	t_6	10 (4.74)	0 (1)	95 (6.9)	98 (6.98)	0 (10)	315.78	20.75	22.66	67.85	0.63	0	0.99	1	0.9
350	100	4	t_3	1 (13.26)	0 (1)	0 (2.2)	0 (2.23)	0 (9.01)	323.13	25.73	27.29	64.57	0	0	0	0	0.06
350	100	8	t_3	0 (19.45)	0 (1)	1 (3.26)	3 (3.51)	0 (10)	312.85	19.23	20.83	72.54	0.01	0	0.28	0.3	0.51
350	100	12	t_3	0 (20.49)	0 (1)	9 (5.31)	22 (5.92)	0 (10)	321.97	26.29	28.68	77.96	0.08	0	0.63	0.68	0.82
350	100	16	t_3	3 (14.84)	0 (1)	36 (6.19)	43 (6.97)	0 (10)	321.12	22.74	24.58	70.48	0.32	0	0.82	0.88	0.92
350	100	20	t_3	5 (9.09)	0 (1)	60 (6.74)	43 (7.39)	0 (10)	328.22	19.8	21.85	75.87	0.49	0	0.92	0.95	0.97
350	1000	5	Gaus.	15 (9.62)	0 (1)	0 (1.28)	0 (1.32)	–	1294.04	119.57	126.47	–	0	0	0.04	0.04	–
350	1000	10	Gaus.	1 (13.19)	0 (1)	2 (5)	22 (5.76)	–	1211.38	222.52	240.64	–	0.01	0	0.67	0.74	–
350	1000	15	Gaus.	7 (8.57)	0 (1)	64 (6.54)	80 (6.7)	–	1294.04	184.71	198.92	–	0.37	0	0.92	0.94	–
350	1000	20	Gaus.	8 (6.02)	0 (1)	87 (6.81)	95 (6.95)	–	1248.75	240.04	263.66	–	0.62	0	0.97	0.99	–
350	1000	25	Gaus.	12 (5.08)	0 (1)	94 (6.84)	98 (6.95)	–	1315.66	184.7	196.18	–	0.7	0	0.97	0.99	–
350	1000	5	t_6	12 (8.26)	0 (1)	0 (1.02)	0 (1.03)	–	998.98	87.48	92.81	–	0	0	0	0	–
350	1000	10	t_6	7 (11)	0 (1)	0 (3.71)	0 (3.86)	–	1220.82	170.81	184.09	–	0	0	0.44	0.45	–
350	1000	15	t_6	5 (11.26)	0 (1)	33 (5.94)	34 (6.09)	–	1134.6	176.73	189.43	–	0.18	0	0.82	0.84	–
350	1000	20	t_6	8 (7.17)	0 (1)	78 (6.77)	78 (6.77)	–	1325.48	237.55	256.75	–	0.47	0	0.96	0.96	–
350	1000	25	t_6	7 (5.95)	0 (1)	93 (6.9)	96 (6.96)	–	1294.97	182.46	199.81	–	0.61	0	0.98	0.99	–
350	1000	5	t_3	8 (5.09)	0 (1)	3 (3.91)	3 (3.9)	–	1318.84	324.41	341.81	–	0	0	0	0	–
350	1000	10	t_3	11 (6.23)	0 (1)	3 (3.62)	3 (3.72)	–	1403.09	257.03	276.48	–	0	0	0.08	0.09	–
350	1000	15	t_3	7 (9.01)	0 (1)	10 (4.64)	12 (4.72)	–	1306.18	256.93	263.95	–	0	0	0.41	0.42	–
350	1000	20	t_3	5 (9.7)	0 (1)	17 (5.87)	17 (6.04)	–	934.4	195.34	209.9	–	0.11	0	0.68	0.69	–
350	1000	25	t_3	11 (8.64)	0 (1)	27 (6.16)	32 (6.67)	–	1111.95	244.85	268.45	–	0.37	0	0.79	0.82	–

Table S24: $K = 8$ simulation results with equal sized clusters. Each setting was replicated 100 times. For each setting, the number of replications detecting the correct number of clusters ($|\hat{K} = 8|$), the mean number of significant clusters (mean $\hat{K}$), median computing time, and mean adjusted Rand Index (ARI) are reported for `pvclust` AU p-values (`pvAU`), BP p-values (`pvBP`), SHC_L, SHC_2, and BootClust (BC). BootClust results are not reported for the high-dimensional setting (−). (N: sample size, p: dimension, δ: separating signal between clusters, F: cluster distribution.)

parameters				$\lvert\hat{K} = 8\rvert$ (mean $\hat{K}$)					median time (sec.)				mean ARI				
N	p	δ	F	pvAU	pvBP	SHC_L	SHC_2	BC	pv	SHC_L	SHC_2	BC	pvAU	pvBP	SHC_L	SHC_2	BC
400	10	4	Gaus.	0 (91.67)	0 (1)	0 (3.18)	1 (4.06)	10 (6.26)	285.66	4.92	5.85	15.36	0.05	0	0.33	0.41	0.59
400	10	8	Gaus.	0 (5.06)	0 (1)	55 (6.99)	85 (7.79)	91 (7.91)	257.84	4.78	6.14	12.32	0.39	0	0.87	0.96	0.97
400	10	12	Gaus.	3 (4.85)	0 (1)	98 (7.92)	99 (7.96)	100 (8)	289.68	6.24	7.06	15.54	0.58	0	0.99	0.99	1
400	10	16	Gaus.	21 (6.04)	0 (1.29)	98 (7.95)	100 (8)	100 (8)	293.42	6.66	7.61	16.96	0.75	0.18	0.99	1	1
400	10	20	Gaus.	33 (6.47)	3 (3.16)	97 (7.84)	100 (8)	100 (8)	290.11	6.52	7.36	16.89	0.8	0.49	0.98	1	1
400	10	4	t_6	0 (91.95)	0 (1)	0 (2.19)	0 (2.51)	9 (5.6)	292.15	5.21	5.75	17.66	0.03	0	0.17	0.19	0.43
400	10	8	t_6	0 (24.48)	0 (1)	18 (6.32)	63 (7.5)	74 (7.82)	296.2	6.54	7.58	17.29	0.41	0	0.77	0.89	0.92
400	10	12	t_6	0 (4.46)	0 (1)	83 (7.47)	96 (7.87)	98 (7.98)	240.5	5.74	6.42	14.25	0.43	0	0.93	0.98	0.99
400	10	16	t_6	1 (4.74)	0 (1)	91 (7.84)	98 (7.98)	99 (7.99)	302.73	6.28	7.46	16.97	0.58	0	0.98	1	1
400	10	20	t_6	9 (5.68)	0 (1.03)	99 (7.95)	100 (8)	100 (8)	247.22	5.84	6.7	14.22	0.68	0.05	0.99	1	1
400	10	4	t_3	0 (88.99)	0 (1)	0 (1.28)	0 (1.34)	0 (8.05)	295.22	3.26	3.56	17.84	0.02	0	0.02	0.03	0.29
400	10	8	t_3	0 (74.68)	0 (1)	2 (4.26)	13 (5.54)	23 (8.96)	289.16	6.24	7.38	17.53	0.17	0	0.46	0.58	0.8
400	10	12	t_3	2 (29.68)	0 (1)	32 (6.3)	53 (7.89)	53 (8.76)	294.11	6.82	7.93	17.38	0.5	0	0.76	0.9	0.94
400	10	16	t_3	2 (8.56)	0 (1)	75 (7.52)	67 (8.33)	58 (8.81)	292.4	6.62	7.58	17.2	0.49	0	0.92	0.98	0.98
400	10	20	t_3	5 (6.17)	0 (1)	87 (7.65)	75 (8.24)	61 (8.78)	291.11	6.63	7.46	17.13	0.53	0	0.95	0.98	0.99
400	100	4	Gaus.	0 (19.87)	0 (1)	0 (2.56)	0 (3.58)	0 (7.66)	374.96	20.42	23.98	81.52	0	0	0.18	0.25	0.31
400	100	8	Gaus.	0 (16.48)	0 (1)	31 (6.92)	43 (7.23)	0 (10)	363.33	33.34	36.47	101.84	0.27	0	0.85	0.87	0.87
400	100	12	Gaus.	4 (5.42)	0 (1)	87 (7.87)	91 (7.91)	0 (10)	427.08	25.59	27.7	127.09	0.55	0	0.98	0.99	0.92
400	100	16	Gaus.	10 (5.68)	0 (1)	93 (7.84)	97 (7.92)	0 (10)	434.55	31.94	33.93	127.23	0.7	0	0.98	0.99	0.92
400	100	20	Gaus.	18 (6.26)	0 (1)	97 (7.89)	100 (8)	0 (10)	435.14	30.58	33.22	127.79	0.78	0.01	0.99	1	0.92
400	100	4	t_6	0 (19.03)	0 (1)	0 (1.55)	0 (1.55)	0 (3.56)	363.62	18.47	19.36	59.61	0	0	0.06	0.06	0.09
400	100	8	t_6	0 (23.39)	0 (1)	3 (5.75)	6 (6.23)	0 (10)	372.79	27.14	29.05	103.36	0.05	0	0.68	0.72	0.82
400	100	12	t_6	6 (9.75)	0 (1)	60 (7.46)	69 (7.62)	0 (10)	389.26	36.14	38.3	104.53	0.44	0	0.92	0.94	0.93
400	100	16	t_6	2 (5.07)	0 (1)	89 (7.79)	94 (7.84)	0 (10)	426.85	26.67	28.35	85.9	0.56	0	0.97	0.98	0.95
400	100	20	t_6	9 (5.56)	0 (1)	96 (7.93)	99 (7.99)	0 (10)	381.52	25.26	27.32	95.78	0.68	0	0.99	1	0.94
400	100	4	t_3	1 (14.59)	0 (1)	0 (2.36)	0 (2.36)	0 (9.01)	414.51	33.28	34.64	76.35	0	0	0	0	0.05
400	100	8	t_3	0 (23.16)	0 (1)	0 (3.79)	2 (4.3)	0 (10)	445.4	24.27	26.42	106.95	0.01	0	0.3	0.33	0.5
400	100	12	t_3	0 (24.31)	0 (1)	12 (6.16)	20 (6.95)	0 (10)	380.76	33.72	36.64	91.18	0.08	0	0.67	0.72	0.83
400	100	16	t_3	3 (17.12)	0 (1)	40 (7.09)	41 (8.14)	0 (10)	443.57	27.74	30	122.97	0.39	0	0.83	0.9	0.94
400	100	20	t_3	4 (9.9)	0 (1)	59 (7.73)	42 (8.66)	0 (10)	365.52	26.29	28.51	102.98	0.52	0	0.91	0.96	0.97
400	1000	5	Gaus.	11 (10.43)	0 (1)	0 (1.28)	0 (1.32)	–	1691.74	158.63	167.07	–	0	0	0.03	0.03	–
400	1000	10	Gaus.	2 (15.05)	0 (1)	6 (5.88)	25 (6.71)	–	1874.58	239.22	264.73	–	0.02	0	0.7	0.76	–
400	1000	15	Gaus.	4 (8.98)	0 (1)	77 (7.66)	92 (7.88)	–	1620.74	262.55	279.85	–	0.41	0	0.95	0.97	–
400	1000	20	Gaus.	11 (6.24)	0 (1)	93 (7.93)	98 (7.98)	–	1455.25	239.3	255.36	–	0.69	0	0.99	0.99	–
400	1000	25	Gaus.	12 (5.74)	0 (1)	97 (7.91)	100 (8)	–	1629.73	264.83	273.85	–	0.7	0	0.99	1	–
400	1000	5	t_6	16 (9.81)	0 (1)	0 (1.02)	0 (1.03)	–	1660.07	157.77	164.52	–	0	0	0	0	–
400	1000	10	t_6	6 (12.99)	0 (1)	0 (4.4)	1 (4.54)	–	1648.44	223.97	241.77	–	0	0	0.46	0.47	–
400	1000	15	t_6	2 (13.92)	0 (1)	35 (7.17)	43 (7.28)	–	1570.98	255.36	279.31	–	0.17	0	0.87	0.89	–
400	1000	20	t_6	9 (7.76)	0 (1)	84 (7.76)	85 (7.8)	–	1541.65	247.08	267.1	–	0.51	0	0.97	0.97	–
400	1000	25	t_6	6 (5.63)	0 (1)	96 (7.86)	96 (7.9)	–	1697.77	284.26	300.82	–	0.64	0	0.98	0.99	–
400	1000	5	t_3	9 (5.1)	0 (1)	3 (4.09)	3 (4.08)	–	1316.81	544.11	567.2	–	0	0	0	0	–
400	1000	10	t_3	7 (6.38)	0 (1)	0 (3.91)	0 (3.9)	–	1351.71	423.56	442.09	–	0	0	0.07	0.07	–
400	1000	15	t_3	3 (10.07)	0 (1)	6 (5.53)	8 (5.7)	–	1387.78	340.85	360.39	–	0.01	0	0.44	0.45	–
400	1000	20	t_3	6 (11.95)	0 (1)	8 (6.41)	14 (6.79)	–	1342.06	248.16	269.51	–	0.13	0	0.67	0.7	–
400	1000	25	t_3	8 (8.93)	0 (1)	32 (7.4)	41 (7.86)	–	1618.96	273.9	302.78	–	0.32	0	0.82	0.85	–

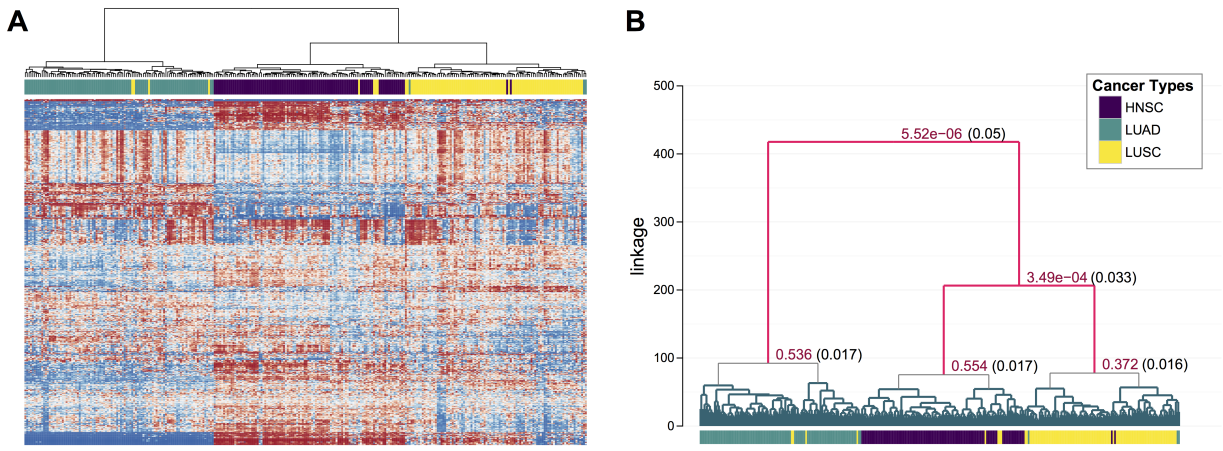

Figure S1: Analysis of gene expression for 300 LUAD, LUSC, and HNSC samples. (A) Heatmap of log-transformed gene expression for the 300 samples (columns), clustered by Ward's linkage. (B) Dendrogram with corresponding SHC p-values (red) and α^* cutoffs (black) given only at nodes tested according to the FWER controlling procedure at $\alpha = 0.05$.

Web Appendix C: Multi-Cancer Gene Expression Dataset

A dataset of 300 samples was constructed by combining 100 samples from each of head and neck squamous cell carcinoma (HNSC), lung squamous cell carcinoma (LUSC), and lung adenocarcinoma (LUAD), all obtained from The Cancer Genome Atlas (TCGA) project (The Cancer Genome Atlas Research Network, 2012, 2014). Gene expression was estimated for 20,531 genes from RNA-seq data using RSEM (Li and Dewey, 2011), as described in the TCGA RNA-seq v2 pipeline (`https://wiki.nci.nih.gov/display/TCGA/RNASeq+Version+2`). To adjust for technical effects of the data collection process, expression values were first normalized using the upper-quartile procedure of Bullard et al. (2010). Then, all expression values of zero were replaced by the smallest non-zero expression value across all genes and samples. A subset of 500 most variably expressed genes were selected according to the median absolute deviation about the median (MAD) expression across all samples. Finally, SHC was applied to the log-transformed expression levels at the 500 most variable loci. Similar results were also obtained with the 100, 1000, and 2000 most variable genes.

In Figure S1A, the log-transformed expression values are visualized using a heatmap,

with rows corresponding to genes, and columns corresponding to samples. Lower and higher expression values are shown in blue and red, respectively. For easier visual interpretation, rows and columns of the heatmap were independently clustered using Ward's linkage clustering. The corresponding dendrogram and cancer type labels are shown above the heatmap. The dendrogram and labels in Panel A of Figure S1 are reproduced in Panel B, along with the SHC p-values (red) and modified significance cutoffs (black) at nodes tested according to the FWER controlling procedure. Diagnostic plots investigating the null Gaussian and factor analysis model assumptions at the first two nodes are shown in Figure S2. Moderately heavy tails are observed in the density plots. However, as in the BRCA dataset, these may be partially attributed to the factor analysis model illustrated in the eigenvalue plots. Ward's linkage clustering correctly separates the three cancer types, with the exception of 8 LUSC samples, 2 HNSC samples, and 3 LUAD samples. The resulting ARI is 0.87, highlighting the strong separation between the three cancer types. Interestingly, the LUSC and HNSC samples cluster prior to joining with the LUAD samples, suggesting the greater molecular similarity between squamous cell tumors of different sites, than different cancers of the lung. This agrees with the recently identified genomic similarity of the two tumors reported in Hoadley et al. (2014). Furthermore, we note that no HNSC and LUAD samples are jointly clustered, highlighting the clear difference between tumors of both distinct histology and site. As shown in Figure S1B, statistically significant evidence of clustering was determined at the top two nodes, with respective Gaussian-fit p-values $5.52e-6$ and $3.49e-4$ at the modified significance cutoffs, $\alpha_1^* = 0.05$ and $\alpha_2^* = 0.033$. Additionally, the three candidate nodes corresponding to splitting each of the cancer types all give insignificant results, suggesting no further clustering in the cohort. We note that when analyzed using `pvclust`, no statistically significant evidence of clustering was found, with AU p-values of 0.14, 0.54, and 0.48 obtained at the three nodes corresponding to primarily LUAD, HNSC and LUSC samples. Finally, when the BootClust approach was applied as described in Section 5, the maximum of 10 clusters was predicted.

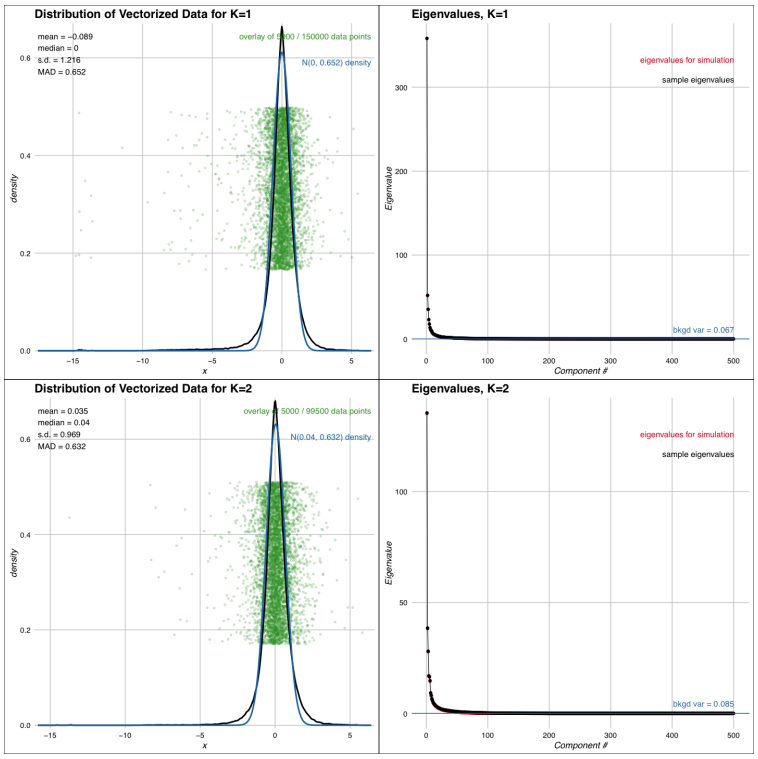

Figure S2: SHC diagnostic plots for multi-cancer gene expression dataset. Diagnostic plots are shown for the hypothesis test performed at the top two nodes: $j = 1$ (top row) and $j = 2$ (bottom row). A kernel density estimate for the raw data (black) and the best-fit Gaussian (blue) are shown in the left panels. Sample eigenvalues (black) and eigenvalues used for simulation (red) are shown in the right panels.

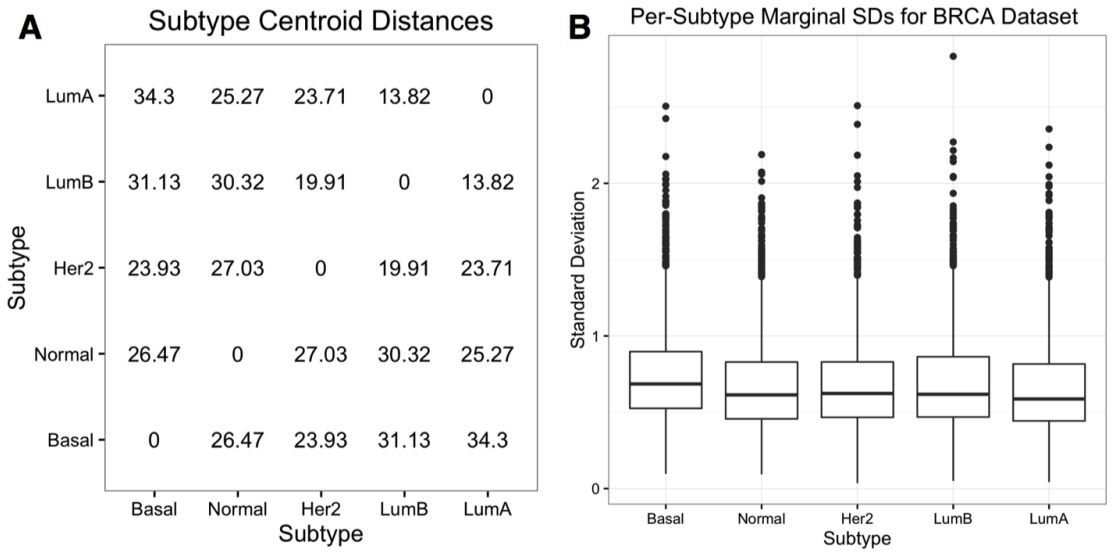

Figure S3: Distribution statistics for BRCA gene expression dataset for comparison with simulation setting parameters. (A) Subtype centroid distances to be compared with δ parameter used in simulations. (B) Distribution of marginal standard deviations (SDs) calculated for each subtype. Marginal SDs of $\sigma = 1$ were used in all non-null simulation settings. The values suggest the δ and σ values used in simulations fall within realistic ranges.

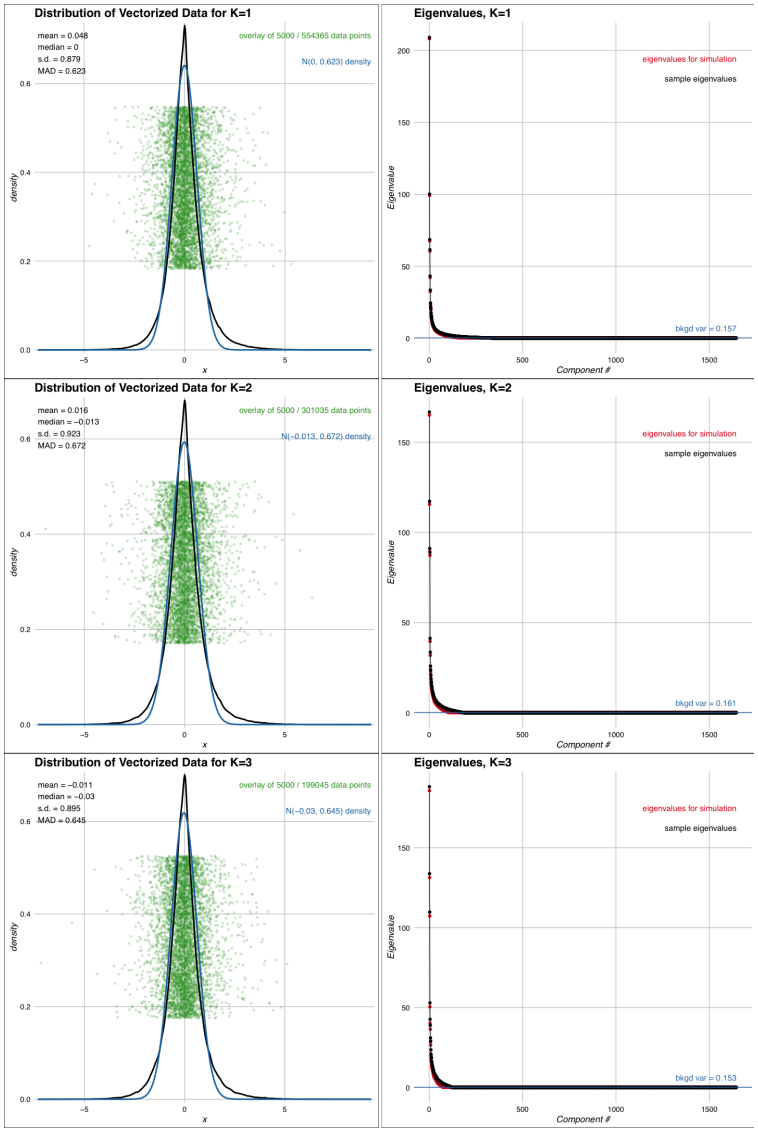

Figure S4: SHC diagnostic plots for BRCA gene expression dataset. Diagnostic plots are shown for the hypothesis test performed at the top three nodes: $j = 1$ (top row), $j = 2$ (middle row), and $j = 3$ (bottom row). A kernel density estimate for the raw data (black) and the best-fit Gaussian (blue) are shown in the left panels. Sample eigenvalues (black) and eigenvalues used for simulation (red) are shown in the right panels.

Web Appendix D: Null Covariance Estimation

As mentioned in Section 2.2, several approaches have been proposed for estimating the low-level background noise, σ_b^2, and the w non-zero entries of Λ_0, including the hard-threshold, soft-threshold, and sample-based approaches (Liu et al., 2008; Huang et al., 2015). Briefly, given the eigenvalues of the sample covariance matrix, $\hat{\lambda}_j$, and an estimate of the background noise, $\hat{\sigma}_b^2$, the hard and soft approaches estimate the diagonal entries of Λ to be $\max\{\hat{\lambda}_j, \hat{\sigma}_b^2\}$ and $\max\{\hat{\lambda}_j - \tau, \hat{\sigma}_b^2\}$, respectively, with tuning parameter $\tau \geq 0$. The sample-based approach simply estimates the diagonal entries of Λ by the $\hat{\lambda}_j$. In the original SigClust paper, Liu et al. (2008) proposed estimating the background noise by:

$$\hat{\sigma}_{\text{RAW}} = \frac{MAD_{p \cdot N \text{ data}}}{MAD_{N(0,1)}},$$

where $MAD_{p \cdot N \text{ data}}$ is used to denote the median absolute deviation about the median (MAD) computed from the $p \cdot N$ total entries of the original data matrix, and $MAD_{N(0,1)}$ is used to denote the MAD of a standard Gaussian distribution.

The $\hat{\sigma}_{\text{RAW}}$ estimator relies on the assumption that a majority of the p variables are pure noise, as in microarray studies where expression is measured for thousands of genes, most of which are of no interest. However, when this assumption does not hold, and the few directions of true signal lie across a large number of variables, $\hat{\sigma}_{\text{RAW}}$ can vastly overestimate the noise level and produce poor estimates of Λ. To address this problem, when $N \ll p$ we propose estimating σ_b from the $N(N-1)$ non-zero principal component (PC) scores of the $p \times N$ data matrix. Intuitively, since the directions of true signal approximately lie within the first few PC directions, a robust estimate of spread based on the PC scores should accurately target the background noise. In this way, the estimator is a natural modification of $\hat{\sigma}_{\text{RAW}}$.

Specifically, we propose the new estimator:

$$\hat{\sigma}_{\text{PC}} = \frac{MAD_{N(N-1) \text{ scaled PC scores}}}{MAD_{N(0,1)}}.$$

where $MAD_{N(N-1) \text{ scaled PC scores}}$ is used to denote the MAD of the $N(N-1)$ non-zero PC scores scaled by $(\frac{p}{N-1})^{1/2}$. The scaling factor is obtained by noting that the variance of the PC scores for N samples drawn from a p-dimensional spherical distribution with covariance $\sigma_b^2 \mathbf{I}_p$ is precisely $(\frac{p}{N-1})\sigma_b^2$.

Since $\hat{\sigma}_{\text{PC}}$ relies on a PC decomposition of the data, when the true signal is relatively weak, the approach may over-estimate σ_b, similar to $\hat{\sigma}_{\text{RAW}}$. Thus, we propose using:

$$\hat{\sigma}_{\min} = \min\{\hat{\sigma}_{\text{RAW}}, \hat{\sigma}_{\text{PC}}\},$$

as a more conservative estimator of σ_b. For all moderate and high dimensional simulations in Section 5, and real data analyses in Section 6, we use $\hat{\sigma}_{\min}$ along with the soft thresholding approach of Huang et al. (2015).

References

Borysov, P., Hannig, J., and Marron, J. S. (2014). Asymptotics of hierarchical clustering for growing dimension. *Journal of Multivariate Analysis* **124,** 465–479.

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94.

Hoadley, K. A., ..., and Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158,** 929–944.

Huang, H., Liu, Y., Yuan, M., and Marron, J. S. (2015). Statistical significance of clustering using soft thresholding. *Journal of Computational and Graphical Statistics* **24,** 975–993.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323.

Liu, Y., Hayes, D. N., Nobel, A. B., and Marron, J. S. (2008). Statistical significance of clustering for high-dimension, low-sample size data. *Journal of the American Statistical Association* **103,** 1281–1293.

Maitra, R., Melnykov, V., and Lahiri, S. N. (2012). Bootstrapping for significance of compact clusters in multidimensional datasets. *Journal of the American Statistical Association* **107,** 378–392.

The Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525.

The Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550.