MoSBi: Automated signature mining for molecular stratification and subtyping

Tim Daniel Rose¹, Thibault Bechtler¹, Octavia-Andreea Ciora¹, Kim Anh Lilian Le¹, Florian Molnar¹, Nikolai Koehler¹, Jan Baumbach³, Richard Röttger², Josch Konstantin Pauling¹

¹ LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Germany
² Department for Mathematics and Computer Science, University of Southern Denmark, Odense
³ Chair of Computational Systems Biology, University of Hamburg, Germany

Supplementary Information
## Individual biclustering algorithms vs. MoSBi

### Supplementary Table 1: Datasets utilized for the comparison of individual biclustering algorithms vs. MoSBi

| Publication         | Omics type   | # of groups | Data size (features x samples) |
|---------------------|--------------|-------------|--------------------------------|
| Wiśniewski et al.¹  | Proteomics   | 3           | 10927 x 32                      |
| Ku et al.²          | Proteomics   | 6           | 3209 x 90                       |
| Berger et al.³      | Transcriptom | 5           | 5000 x 1063                     |
| Curtis et al.⁴      | Transcriptom | 6           | 5000 x 1898                     |
| Tang et al.⁵        | Metabolomics | 6           | 399 x 30                        |
| Yang et al.⁶        | Metabolomics | 2           | 226 x 88                        |
Supplementary Figure 1: Unfiltered bicluster network of proteomics data from Ku et al.\textsuperscript{2}

Edges represent overlaps between biclusters/
Supplementary Figure 2: Bicluster networks of metabolomics data from Tang et al.\textsuperscript{5}

Biclusters are colored by clinical confounding factors, including cancer subtype, receptor status, Ki67 proliferation and several mutations in known cancer driver genes.

Supplementary Figure 3: Co-occurrence networks of communities from Tang et al.\textsuperscript{5}

Exemplary Bicluster communities from the networks shown in Supplementary Figure 1. Nodes represent samples, colored by subtype and metabolic features (white). They are connected, if they co-occur in multiple biclusters of a Louvain community. For further details, see the Methods section.
Synthetic evaluation scenarios

We first reproduced existing scenarios and took this as a starting point to develop further scenarios. We implemented and adapted the model described by Prelic et al.\textsuperscript{7}, focusing on the simultaneous row and column overlap. Additionally, we reproduced the dataset presented by Eren et al.\textsuperscript{8}, which used a bigger data matrix and multiple bicluster patterns. However, these benchmark datasets focused only on the classic scenarios like noise, overlap or bicluster pattern.

The workflow of our data generation approach is illustrated in Supplementary Figure 2. First, biclusters with a previously defined shape, size and pattern are generated. The biclusters are then inserted in an empty data matrix and a background distribution is added to the remaining cells. The obtained raw matrix is processed by adding noise and sparsity. After performing a random permutation of rows and columns, the dataset is finished and ready to be used for evaluation. Each data matrix has a corresponding gold standard, where the rows and columns of every hidden bicluster are listed. Before running the algorithms on the dataset, the matrix is normalized by computing the z-scores row-wise.

For the evaluation of biclustering algorithms, we generated a total of 670 synthetic datasets, grouped into 12 scenarios with 25 sub scenarios, covering many characteristics of real
biological data. An overview of all datasets and their characteristics is shown in Supplementary Table 1. Each of them will be described in this section.

Pattern

All patterns are generated using a linear model. A bicluster \( B = (I, J) \) of size \( k \times i \) is defined according to the equation:

\[
B = \begin{pmatrix}
\pi_1 \alpha_1 + \beta_1 & \pi_2 \alpha_1 + \beta_1 & \cdots & \pi_i \alpha_1 + \beta_1 \\
\pi_1 \alpha_2 + \beta_2 & \pi_2 \alpha_2 + \beta_2 & \cdots & \pi_i \alpha_2 + \beta_2 \\
\vdots & \vdots & \ddots & \vdots \\
\pi_1 \alpha_k + \beta_k & \pi_2 \alpha_k + \beta_k & \cdots & \pi_i \alpha_k + \beta_k
\end{pmatrix}
\]

(1)

Where \( \pi = \{\pi_1, \pi_2, \ldots, \pi_i\} \) is the base row vector, \( \alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_k\} \) is the scale column vector and \( \beta = \{\beta_1, \beta_2, \ldots, \beta_k\} \) is the shift column vector. From that we define six bicluster patterns: constant, constant-row, constant-column, shift, scale, shift-scale. In a constant bicluster, all elements have the same value over all rows and columns. This is illustrated in Supplementary Figure 3A. In a row-constant pattern, the elements located on the same row have equal values, but differ between distinct rows. An example of a row-constant pattern is shown in Supplementary Figure 3B. The opposite pattern is represented by the column-constant bicluster illustrated in Figure Supplementary Figure 3C. It has constant values on the column dimension and varying ones on the row dimension. The shift pattern is generated from a base vector \( \pi \) and an additive component, represented by the shift vector \( \beta \). This pattern is derived from the general linear model, with \( \alpha = 1 \) and shown in Supplementary Figure 3D. In a scale pattern, there is no additive component \( \beta = 0 \). The values inside the bicluster depend on the base row \( \pi \) and the scale vector \( \alpha \). An example is shown in Supplementary Figure 3E. The shift-scale pattern is the one corresponding to the general linear model, where the base row \( \pi \), the scale column \( \alpha \) and the shift column \( \beta \) can take any value. Supplementary Figure 3F illustrates a shift-scale pattern.

Supplementary Figure 5: Examples of bicluster patterns.
(a) Constant pattern, (b) row-constant pattern, (c) column-constant pattern, (d) shift pattern, (e) scale pattern, (f) shift-scale pattern. All patterns are particular cases of the general linear model presented in equation 1.

In a data matrix of size 500 x 200, one single bicluster is hidden. The bicluster of size 50 x 50 has one of the six patterns. The constant bicluster was generated with a constant expression value of 5, while the biclusters of other types were computed based on the previously defined linear model. The base, scale and/or shift vectors are randomly generated.
from $U(1,50)$, where $U(a,b)$ is the uniform distribution with a minimum of $a$ and a maximum of $b$. Here, we distinguish two sub scenarios having different background distributions. The background matrix contains all cells of the matrix which do not belong to any bicluster. In the first scenario, the background matrix has an expression level of 0, while in the second sub scenario, the values in the background are randomly generated from $U(1,100)$. Both sub scenarios contain 6 datasets each, one for every bicluster pattern.

**Regulation**

In this scenario we analyze the effect of different bicluster patterns when taking into account the regulation of the bicluster in relation to the background. An upregulated bicluster, which is higher expressed than the background distribution, is expected to be better detected by the algorithms than a bicluster whose expression values do not stand out from the background. We define the regulation of a bicluster in relation to the background matrix and distinguish 5 regulation levels of the bicluster: -2 (highly-downregulated), -1 (downregulated), 0 (no regulation), 1 (upregulated), 2 (highly-upregulated). First, we generate the bicluster expression levels in the same way we did for the scenario "Bicluster Pattern". Based on the bicluster values, the background distribution is created from a normal distribution with mean and standard deviation defined as it follows:

$$sd_{background} = sd_{bicluster} * U(1,3)$$  
$$mean_{background} = mean_{bicluster} - (sd_{bicluster} + sd_{background}) * 2 * reg$$

Where $sd$ are the standard deviations of the bicluster and background, respectively; $mean$ are the means of the bicluster and background; $reg \in \{-2, -1, 0, 1, 2\}$ is the regulation factor and $U(a, b)$ the uniform distribution with values between $a$ and $b$.

**Noise**

Studying the effect of noise on the performance of biclustering algorithms is one of the most frequently used scenarios when evaluating on synthetic data. The definition and implementation of noise might vary between publications. In this project, we define noise as a random number picked from a normal distribution with mean 0. The standard deviation represents the different noise levels: 0, 0.1, ..., 0.9, 1. For studying the effect of noise, we use a data matrix of size 500 x 200 and hide one highly-upregulated bicluster of size 50 x 50. The bicluster has a shift pattern which was computed using a base vector and a shift vector generated randomly from $N(0,1)$. The background distribution is generated as described in the previous section. The bicluster is hidden in the data matrix and a z-score normalization is applied row-wise on the resulting matrix. After that, noise is added to every cell of the matrix. For a more robust result, the data is generated with 5 repetitions for each of the 11 noise levels.

**Size and shape**

Size and shape are other important bicluster characteristics, which may determine differences in the performance of biclustering algorithms. In these scenarios, investigate the effect of bicluster shapes and sizes on algorithm performances. For this reason, we use a 500 x 200 data matrix and a highly-upregulated shift bicluster of variable size and shape,
generated as described before. For the "Bicluster Size" scenario, we use a square-shaped bicluster and test 11 different sizes: 5x5, 10x10, 20x20, ..., 90x90, 100x100. As a result, this scenario has 11 datasets. For the "Bicluster Shape" scenario, the lengths and the widths of the biclusters are limited to 6 different values: 30, 40, ..., 80.

Dataset size

In this scenario, we test whether the size of the dataset influences the biclustering result. Up to this point we used a data matrix with 500 rows and 200 columns. In this scenario, the dataset size is varied and takes each of the following sizes: 100x100, 200x100, 500x200, 1000x500, 5000x1000. The same shift bicluster of size 50 x 50 described previously is hidden in the data matrix.

Sparsity

Under sparsity we understand the percentage of values in a data matrix that are not available for technical or biological reasons (We set them to 0). For studying the effect of sparsity, we hide one shift bicluster of size 50 x 50 in a 500 x 200 data matrix. The shift pattern is computed using a base vector and a shift vector generated randomly from N(100,1). The background distribution is generated from a uniform distribution with the minimum value of 1 and the maximum value equal to the minimum value in the bicluster. In this way, we make sure that the bicluster is upregulated and all expression values in the matrix are greater than 0. We use 11 different sparsity levels: 0, 0.05, 0.1 ..., 0.45, 0.5. This means that 0 - 50 % of the biclustering values are randomly set to 0.

Multiple Biclusters

All previously described scenarios involved hiding one single bicluster at a time. This scenario implants multiple biclusters of the same type in a data matrix. For this purpose, we use six variable biclusters of size 40 x 40 and a bigger dataset of size 500 x 300. The reason for the small change in size is the fact that we intend to place all six biclusters in the data matrix without any overlap. The previous scenarios revealed that these small changes in dimensions have no big effect on the performance of algorithms. All six biclusters are non-overlapping and upregulated. This scenario consists of three sub scenarios, each having one changed parameter. In the first simulation, we use biclusters of fixed pattern and size. All six biclusters have a shift pattern and the same size of 40 x 40. The pattern is generated with a base vector from N(100,20) and a shift vector from N(mean,20), where the mean is picked from U(30,100). In the second sub scenario, we vary the pattern of the biclusters and keep the size fixed. We use one bicluster of each pattern, all of them having size 40 x 40. For the third sub scenario, we use six shift biclusters with varying sizes of 30 x 50, 25 x 25, 40 x 60, 30 x 30, 45 x 45, 50 x 30.

One-dimensional overlap

Here we investigate the one-dimensional overlap of biclusters in a data matrix. The overlap can be either on the row dimension or on the column dimension. We define the overlap degree between two biclusters A and B as the number of rows/columns from bicluster A that are contained in bicluster B. The overlap simulation is done on 8 different overlap degrees:
0, 0.1, ..., 0.7, each with 10 repetitions. The analysis of one-dimensional overlap is repeated with row or column-overlap and biclusters of equal or different sizes.

Two-dimensional overlap

A two-dimensional overlap, i.e. simultaneous row and column overlap, defines a set of rows and a set of columns that belong to multiple biclusters at the same time. Generating this type of overlap with our own data generation approach is a difficult task, because expression levels that belong to two biclusters at the same time, should follow the pattern and fulfill the linear model of both biclusters. For this one scenario, we used a different approach and implemented the model introduced by Prelic et al. This consists of a 100 x 100 data matrix where multiple biclusters of the same size and pattern are hidden. In their publication, the authors describe two dataset types: constant and additive. The constant dataset is a binary data matrix, where the constant biclusters have an expression level of 1 and the background is 0. The additive dataset contains biclusters with column-constant patterns and the values in the background are generated from a uniform distribution, but are strictly smaller than the bicluster expression level.

Complex Scenarios

In these complex scenarios, we incorporated all aspects studied in the previous scenarios: multiple biclusters of different patterns, shapes and sizes, noise, sparsity and row- and column-overlap. For this simulation, a dataset of size 500 x 300, a noise level of 0.3 and a sparsity level of 0.1 were used. Six upregulated biclusters were placed in the data matrix. The pattern, size, and shape of each bicluster were chosen randomly and independently for each dataset (which we repeated 10 times). It aims to reconstruct experimental omics data and is a difficult scenario, because of many inferring factors.

In addition, we generated three more scenarios following a slightly different approach. In these scenarios, we simulated the background distribution as a negative binomial distribution. In this way, the missing values in the dataset (simulated as sparsity) were already integrated in the distribution and did not have to be added synthetically. The sparsity level was varied by adjusting the parameters of the negative binomial distribution. This is meant to simulate unique molecular identifier RNA sequencing data.
Supplementary Table 2: Synthetic data scenarios for evaluation on biclustering algorithms.

| Scenario name        | Subscenario Specification | Number of Biclusters | Bicluster Pattern | Bicluster Size | Regulation | Dataset Size | Noise | Sparsity | Overlap | Number of Datasets |
|----------------------|---------------------------|-----------------------|-------------------|---------------|------------|--------------|-------|----------|---------|-------------------|
| Bicluster Pattern    |                           |                       |                   |               |            |              |       |          |         |                   |
|                      | No background             | 1                     | all               | 50x50         | -           | 500x200     | 0     | 0        | -       | 6                 |
|                      | Random background         | 1                     | all               | 50x50         | -           | 500x200     | 0     | 0        | -       | 6                 |
|                      | Highly upregulation       | 1                     | all               | 50x50         | 2           | 500x200     | 0     | 0        | -       | 6                 |
|                      | Uregulation               | 1                     | all               | 50x50         | 1           | 500x200     | 0     | 0        | -       | 6                 |
|                      | No Regulation             | 1                     | all               | 50x50         | 0           | 500x200     | 0     | 0        | -       | 6                 |
|                      | Downregulation            | 1                     | all               | 50x50         | -1          | 500x200     | 0     | 0        | -       | 6                 |
|                      | Highly downregulation     | 1                     | all               | 50x50         | -2          | 500x200     | 0     | 0        | -       | 6                 |
|                      |                           |                       |                   | 5x5          |             |              |       |          |         |                   |
|                      |                           |                       |                   | 15x10         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 20x20         |             |              |       |          |         |                   |
| Bicluster Size       | Varying sizes             | 1                     | shift             |              |             | 500x200     | 0     | 0        | -       | 11                |
|                      |                           |                       |                   | 100x100       |             |              |       |          |         |                   |
|                      |                           |                       |                   | 200x20        |             |              |       |          |         |                   |
|                      |                           |                       |                   | 300x30        |             |              |       |          |         |                   |
|                      |                           |                       |                   | 30x10         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 35x25         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 40x40         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 40x30         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 40x80         |             |              |       |          |         |                   |
| Bicluster Shape      | Varying shapes            | 1                     | shift             |              |             | 500x200     | 0     | 0        | -       | 36                |
|                      | and sizes                 |                       |                   |              |             | 100x100     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 200x100     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 500x200     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 1000x500    | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 5000x1000   | 0     | 0        | -       | 5                 |
| Dataset Size         | Varying dataset sizes     | 1                     | shift             |              |             | 500x200     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 100x100     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 200x100     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 500x200     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 1000x500    | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 5000x1000   | 0     | 0        | -       | 5                 |
| Noise                | Varying noise levels      | 1                     | shift             |              |             | 500x200     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   |              |             | 100x100     | 0     | 0.1      | -       | 5                 |
|                      |                           |                       |                   |              |             | 200x100     | 0     | 0        | -       | 5                 |

| Scenario name        | Subscenario Specification | Number of Biclusters | Bicluster Pattern | Bicluster Size | Regulation | Dataset Size | Noise | Sparsity | Overlap | Number of Datasets |
|----------------------|---------------------------|-----------------------|-------------------|---------------|------------|--------------|-------|----------|---------|-------------------|
| Sparsity             | Varying sparsity levels   | 1                     | shift             |              | 50x50      | 2            | 500x200| 0       | 0.05    | 0                 |
|                      |                           |                       |                   |              |            |              |       |          | ...     | 55                |
| Multiple Biclusters  | Same pattern              | 6                     | shift             |              | 40x40      | 1            | 500x300| 0       | 0       | -                 |
|                      | Varying patterns          | 6                     | all               | 40x40         | 1           | 500x300     | 0     | 0        | -       | 5                 |
|                      |                           |                       |                   | 30x30         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 25x25         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 40x50         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 45x45         |             |              |       |          |         |                   |
|                      |                           |                       |                   | 50x30         |             |              |       |          |         |                   |
| One-dimensional     | Column overlap            | 6                     | shift             |              | 40x40      | 1            | 500x300| 0       | 0.1     | 0                 |
| Overlap              |                           |                       |                   |              |            |              |       |          | ...     | 80                |
|                      | Row overlap               | 6                     | shift             |              | 40x40      | 1            | 500x300| 0       | 0.1     | 0                 |
|                      |                           |                       |                   |              |            |              |       |          | ...     | 80                |
|                      | Column overlap, varying   | 6                     | shift             |              | 40x40      | 1            | 500x300| 0       | 0.1     | 0                 |
|                      | bicluster sizes           |                       |                   |              |            |              |       |          | ...     | 80                |
| Scenario name | Subscenario Specification | Number of Biclusters | Bicluster Pattern | Bicluster Size | Regulation | Dataset Size | Noise | Sparsity | Overlap | Number of Datasets |
|---------------|--------------------------|----------------------|------------------|----------------|------------|--------------|-------|----------|---------|-------------------|
| Row overlap, varying bicluster sizes | 6 | shift | 30x50 25x25 40x40 30x30 45x45 50x30 | 1 | 500x300 | 0 | 0 | 0 | 80 |
| "Close to Real" Two-dimensional Overlap | Complex | 6 | random | random | 1 | 500x300 | 0.3 | 0.1 | random row and column | 10 |
| Constant model (binary matrix) | 10 | constant | ~10x10 | ~100x100 | 0 | 0 | random row and column | 45 |
| Additive model | column-constant | ~10x10 | ~100x100 | 0 | 0 | random row and column | 45 |
| Negative Binomial Background | Sparsity | 1 | shift | 50x50 | - | 500x200 | 0.3 | 0 - 0.5 | - | 11 |
| Complex without noise | 6 | random | random | 500x300 | 0 | ~ 0.1 | random | 10 |
| Complex without noise | 6 | random | random | 500x300 | 0.3 | ~ 0.1 | random | 10 |
Evaluation of biclustering algorithms on synthetic scenarios

Supplementary Figure 6: Evaluation on synthetic data.
Datasets with a single hidden bicluster per data with no background, normally distributed background, biclusters of different factors.

Supplementary Figure 6: Evaluation on synthetic data.
Performance on increasing noise levels with and without z-score normalization of the data matrix. Performance on datasets with biclusters of different shapes.
Supplementary Figure 8: Evaluation on synthetic data.

Top: Performance on varying dataset sizes. Middle: Six biclusters (of different patterns) hidden in the same matrix. Bottom: Relevance distribution six biclusters of different patterns hidden in the same matrix.
Supplementary Figure 9: Evaluation on synthetic data.
Top: Multiple biclusters of different shapes are hidden in the same matrix. Middle: Multiple biclusters with overlapping rows in the same matrix. Bottom: Multiple biclusters with overlapping columns in the same matrix.
Supplementary Figure 10: Evaluation on synthetic data.
Multiple biclusters with overlapping rows or columns in the same matrix using a constant or additive model.
Supplementary Figure 11: Synthetic evaluation with filtered algorithms.
For this figure only the biclustering algorithms Plaid, Fabia, QUBIC, Isa and UniBic were utilized. (A) Performance of biclustering algorithms and ensemble approach on a synthetic scenario including different bicluster types, sizes, sparsity and noise with a negative binomial distributed background (left) and normal distributed background (right). (B) Relevance distribution of biclustering algorithms for the scenario shown in (A, right). (C) Relevance distribution of all algorithms summed up from (B). (D) Relevance distribution of the predictions of the ensemble approach using the biclusters from (D).
Supplementary Figure 12: Parameter evaluation of the MoSBi method

(A) Recovery and relevance for different row and column thresholds. Subplots show different minimum community sizes. For the evaluation, the synthetic data from main Figure 3 was used (with Gaussian background noise. For the same data, different numbers of
randomizations were investigated in (B). (C) & (D) Same plots and in (A) & (B) but with the synthetic data with negative binomially distributed data.

Parameter optimization for biclustering algorithms

We optimized the parameters for the biclustering algorithms using naive grid optimization as performed in the clustering methods evaluation by Wiwi et al.\textsuperscript{11}. Where constraints for parameters were available by the tool’s documentation, these were used otherwise defined by us. Each tool was given 1000 runs. The performances of all runs for all tools is plotted in Supplementary Figure 8.

Supplementary Figure 13: Optimization of parameters

Optimization of parameters for biclustering algorithms on the two “complex” synthetic scenarios with a gaussian or negative binomial background distribution. The red dot showcases the performance on default parameters.
Bicluster similarity

Comparison of additive and multiplicative overlaps of biclusters. An additive score considers a one-dimensional overlap as a non-zero overlap, while a multiplicative score only returns non-zero similarities for two-dimensional overlaps (Supplementary Figure 11).

Supplementary Figure 14: Comparison of additive and multiplicative overlaps of biclusters.

Additive score as used by Hanczar and Nadif\textsuperscript{12} to evaluate bicluster similarity and multiplicative score as used in this manuscript (See methods section) for one and two-dimensional overlaps of biclusters.
References

1. Wiśniewski, J. R. et al. Absolute Proteome Analysis of Colorectal Mucosa, Adenoma, and Cancer Reveals Drastic Changes in Fatty Acid Metabolism and Plasma Membrane Transporters. *J. Proteome Res.* **14**, 4005–4018 (2015).

2. Ku, X. et al. Deciphering tissue-based proteome signatures revealed novel subtyping and prognostic markers for thymic epithelial tumors. *Mol. Oncol.* **14**, 721–741 (2020).

3. Berger, A. C. et al. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell* **33**, 690–705.e9 (2018).

4. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).

5. Tang, X. et al. A joint analysis of metabolomics and genetics of breast cancer. *Breast Cancer Res.* **16**, 415 (2014).

6. Yang, Y. et al. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics* **9**, 4101–4114 (2019).

7. Prelić, A. et al. A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics* **22**, 1122–1129 (2006).

8. Eren, K., Deveci, M., Küçüktunç, O. & Çatalyürek, Ü. V. A comparative analysis of biclustering algorithms for gene expression data. *Brief. Bioinform.* **14**, 279–292 (2013).

9. Aguilar-Ruiz, J. S. Shifting and scaling patterns from gene expression data. *Bioinformatics* **21**, 3840–3845 (2005).

10. Townes, F. W., Hicks, S. C., Aryee, M. J. & Irizarry, R. A. Author Correction: Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol.* **21**, 179 (2020).

11. Wiwie, C., Baumbach, J. & Röttger, R. Comparing the performance of biomedical clustering methods. *Nat. Methods* **12**, 1033–1038 (2015).

12. Hanczar, B. & Nadif, M. Using the bagging approach for biclustering of gene expression
data. *Neurocomputing* vol. 74 1595–1605 (2011).