Supplementary information S1
Segmentation of biological multivariate time-series data
Nooshin Omranian1,2, Bernd Mueller-Roeber1, Zoran Nikoloski2,*
1 Department of Molecular Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, Haus 20, 14476 Potsdam, Germany
2 Systems Biology and Mathematical Modelling Group, Max Planck Institute for Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam, Germany
* E-mail: nikoloski@mpimp-golm.mpg.de

Synthetic data
In this section, we provide additional elaboration of the results from different approaches applied on the synthetic data, described in the main text in Omranian et al. (2013). The results are succinctly summarized in Table S1, and Figure S1 showing the comparison between the obtained segmentations. The actual breakpoints are 7, 12, and 21.

Yeast’s metabolic cycle
In this section, we provide additional elaboration of the results from different approaches applied on the yeast’s metabolic cycle (YMC) data, described in the main text. The results are succinctly summarized in Table S2, showing the comparison between the obtained segmentations.

Yeast’s cell cycle
In this section, we provide additional elaboration of the results from different approaches applied on the yeast’s cell cycle (YCC) data. With the filtering step, the number of genes was reduced from 6076 to 2071. The characteristics of the resulting segmentations are summarized in Table S3 and Figure S3. Based on the work done by Spellman et al. (1998), the data should be segmented into 5 segments representing 5 cycles for which each cycle includes the following phases: M/G1, G1, S, G2, and M. Each of the M/G1, G1 and S phases lasts 2 time points while the G2 phase lasts only one time point, as described in Ramakrishnan et al. (2010). Therefore, as shown in Table S3, our method revealed the cell cycles in the YCC data.

Oxidative stress and yeast’s cell cycle
In this section, we provide additional elaboration of the results from the different approaches applied on the data capturing the effect of oxidative stress, induced by hydrogen peroxide (HP), on the yeast’s cell cycle. With the filtering step, the number of genes was reduced from 4771 to 1189.

The characteristics of the resulting segmentations are summarized in Table S4 and Figure S4. Based on the work done by Shapira et al. (2004), we could capture all phases in the system which correspond to the G1, S, G2, G2/M phases of the cell cycle.
Figure Legends

**Figure S1.** Segmentation over synthetic data from Omranian et al. (2013). The green dashed lines show the obtained breakpoints. The red dots, connected by a red line, represent the sequence \( A \) (column-averages of the absolute values of the regression coefficients in the matrix \( C \)).

**Figure S2.** Segmentation over Yeast’s metabolic cycle. The expression profile of 255 genes illustrated over 36 time points (separated by \( \sim 25\)-min intervals) over three consecutive cell cycles. The green dashed lines show the obtained breakpoints. The red dots, connected by a red line, represent the sequence \( A \) (column-averages of the absolute values of the regression coefficients in the matrix \( C \)).
Figure S3. Segmentation over Yeast’s cell cycle. The green dashed lines show the obtained breakpoints. The red dots, connected by a red line, represent the sequence A (column-averages of the absolute values of the regression coefficients in the matrix C).

Figure S4. Segmentation over Yeast’s cell cycle. The green dashed lines show the obtained breakpoints. The red dots, connected by a red line, represent the sequence A (column-averages of the absolute values of the regression coefficients in the matrix C).
Tables

Table S1. Optimal segmentation for synthetic data. The first part of the table comprises the result of the optimal segmentation for synthetic data based on the regularized regression approach implemented in Algorithm 1. The second part includes the result based on Bleakley et al. (2011) approach. The third and the forth parts show the results based on the method of Omranian et al. (2013), penalized longest path algorithm using number of segments and distribution of length of the segments to calculate the penalty of a path, respectively. The lower part contains the result based on the method of Ramakrishnan et al. (2010). The upper part of the table includes the number of segments $k$, breakpoints, and tuning parameters corresponding to fused LASSO regularization parameters, $\lambda_1$ and $\lambda_2$. In the third part of the table, the first and second columns show the name and the type of network properties used to determine the distances: G stands for global, L for local, and LG for local-global. The third column includes the number $k$ for each of the three methods and the resulting segments are given in the forth column. The fifth and sixth columns in the third and forth parts present the values of lower ($\nu_{\min}$) and upper ($\nu_{\max}$) bound of the tuning parameter $\nu$ with dynamic programming approach. The lower part also includes minimum and maximum length of the segments, i.e., $l_{\min}$ and $l_{\max}$, as parameters of the contending method.

| Algorithm 1 | $k$ | Breakpoints | $\lambda_1$ | $\lambda_2$ |
|-------------|-----|-------------|-------------|-------------|
| Bleakley et al. (2011) | 6 | 7, 12, 16, 21 and 27 | [1,50] | [1,50] |
| Omranian et al. Penalized (number) | Type | $k$ | Breakpoints | $\nu_{\min}$ | $\nu_{\max}$ |
| relative density | G | 5 | 4, 8, 12 and 21 | 0.05 | 6.00 |
| degree | L | 5 | 4, 8, 12 and 24 | 1.50 | 11.13 |
| closeness | LG | 4 | 4, 15 and 20 | 0.05 | 4.47 |
| betweenness | LG | 7 | 4, 8, 12, 17, 21 and 29 | 1.06 | 12.10 |
| Omranian et al. Penalized (length) | Type | $k$ | Breakpoints | $\nu_{\min}$ | $\nu_{\max}$ |
| relative density | G | 9 | 4, 8, 12, 16, 20, 24, 28 and 32 | 0.05 | 6.00 |
| degree | L | 8 | 5, 9, 13, 18, 23, 27 and 32 | 1.50 | 11.13 |
| closeness | LG | 4 | 4, 16 and 20 | 0.05 | 4.47 |
| betweenness | LG | 8 | 4, 9, 13, 17, 21, 25 and 30 | 1.06 | 12.10 |
| Ramakrishnan et al. (2010) | $k$ | Breakpoints | $l_{\min}$ | $l_{\max}$ |
| 6 | 5, 12, 19, 24 and 30 | 4 | 7 |
| 5 | 7, 12, 21 and 29 | 4 | 9 |
| 4 | 6, 12 and 24 | 4 | 12 |
| 3 | 7 and 22 | 4 | 15 |
| 3 | 7 and 25 | 4 | 20 |
| 2 | 25 | 4 | 28 |
| 2 | 7 | 4 | 32 |
Table S2. Optimal segmentation for yeast’s metabolic cycle (YMC) data with the same preprocessing has been applied in Ramakrishnan et al. [15]. The first part of the table comprises the result of the optimal segmentation for synthetic data based on the regularized regression approach implemented in Algorithm 1. The second part includes the result based on Bleakley et al. (2011) approach. The third and the forth parts show the results based on the method of Omranian et al. (2013), penalized longest path algorithm using number of segments and distribution of length of the segments to calculate the penalty of a path, respectively. The lower part contains the result based on the method of Ramakrishnan et al. (2010). The upper part of the table includes the number of segments \( k \), breakpoints, and tuning parameters corresponding to fused LASSO regularization parameters, \( \lambda_1 \) and \( \lambda_2 \). In the third part of the table, the first and second columns show the name and the type of network properties used to determine the distances: G stands for global, L for local, and LG for local-global. The third column includes the number \( k \) for each of the three methods and the resulting segments are given in the forth column. The fifth and sixth columns in the third and forth parts present the values of lower (\( \nu_{\min} \)) and upper (\( \nu_{\max} \)) bound of the tuning parameter \( \nu \) with dynamic programming approach. The lower part also includes minimum and maximum length of the segments, i.e., \( l_{\min} \) and \( l_{\max} \), as parameters of the contending method.

| Algorithm 1 | \( k \) | Breakpoints | \( \lambda_1 \) | \( \lambda_2 \) |
|-------------|--------|-------------|--------|--------|
| Bleakley et al. (2011) | \( k \) | Breakpoints |
| Omranian et al. Penalized (number) Type | \( k \) | Breakpoints | \( \nu_{\min} \) | \( \nu_{\max} \) |
| relative density degree | 6 | 4, 9, 13, 20 and 31 | 0.05 | 5.60 |
| closeness | L 8 | 4, 8, 12, 16, 20, 24 and 32 | 4.35 | 13.50 |
| betweenness | LG 6 | 4, 9, 17, 21 and 31 | 0.05 | 4.24 |
| Ramakrishnan et al. [15] | \( k \) | Breakpoints | \( l_{\min} \) | \( l_{\max} \) |
| 8 | 6, 10, 14, 18, 22, 26 and 31 | 4 | 7 |
Table S3. Optimal segmentation for yeast’s cell cycle (YCC) data. The first part of the table comprises the result of the optimal segmentation for synthetic data based on the regularized regression approach implemented in Algorithm 1. The second part includes the result based on Bleakley et al. (2011) approach. The third and the forth parts show the results based on the method of Omranian et al. (2013), penalized longest path algorithm using number of segments and distribution of length of the segments to calculate the penalty of a path, respectively. The lower part contains the result based on the method of Ramakrishnan et al. (2010). The upper part of the table includes the number of segments $k$, breakpoints, and tuning parameters corresponding to fused LASSO regularization parameters, $\lambda_1$ and $\lambda_2$. In the third part of the table, the first and second columns show the name and the type of network properties used to determine the distances: G stands for global, L for local, and LG for local-global. The third column includes the number $k$ for each of the three methods and the resulting segments are given in the forth column. The fifth and sixth columns in the third and forth parts present the values of lower ($\nu_{min}$) and upper ($\nu_{max}$) bound of the tuning parameter $\nu$ with dynamic programming approach. The lower part also includes minimum and maximum length of the segments, i.e., $l_{min}$ and $l_{max}$, as parameters of the contending method.

| Algorithm 1 | $k$ | Breakpoints | $\lambda_1$ | $\lambda_2$ |
|-------------|-----|-------------|-------------|-------------|
|             | 4   | 4, 9 and 14 | [1,50]      | [1,50]      |

| Bleakley et al. (2011) | $k$ | Breakpoints |
|-------------------------|-----|-------------|
|                         | 2   | 4           |

| Omranian et al. Penalized (number) Type | $k$ | Breakpoints | $\nu_{min}$ | $\nu_{max}$ |
|----------------------------------------|-----|-------------|-------------|-------------|
| relative density                       | G   | 4           | 0.05        | 4.08        |
| degree                                 | L   | 3           | 6 and 14    | 6.64        | 12.72       |
| closeness                              | LG  | 3           | 5 and 14    | 0.05        | 5.64        |
| betweenness                            | LG  | 4           | 4, 8 and 14 | 9.15        | 15.38       |

| Ramakrishnan et al. [15] | $k$ | Breakpoints | $l_{min}$ | $l_{max}$ |
|--------------------------|-----|-------------|-----------|-----------|
|                          | 6   | 3, 6, 9, 12 and 15 | 3         | 5         |
Table S4. Optimal segmentation for data from oxidative stress, induced by hydrogen peroxide (HP), on yeast’s cell cycle. The first part of the table comprises the result of the optimal segmentation for synthetic data based on the regularized regression approach implemented in Algorithm 1. The second part includes the result based on Bleakley et al. (2011) approach. The third and the forth parts show the results based on the method of Omranian et al. (2013), penalized longest path algorithm using number of segments and distribution of length of the segments to calculate the penalty of a path, respectively. The lower part contains the result based on the method of Ramakrishnan et al. (2010). The upper part of the table includes the number of segments $k$, breakpoints, and tuning parameters corresponding to fused LASSO regularization parameters, $\lambda_1$ and $\lambda_2$. In the third part of the table, the first and second columns show the name and the type of network properties used to determine the distances: G stands for global, L for local, and LG for local-global. The third column includes the number $k$ for each of the three methods and the resulting segments are given in the forth column. The fifth and sixth columns in the third and forth parts present the values of lower ($\nu_{\text{min}}$) and upper ($\nu_{\text{max}}$) bound of the tuning parameter $\nu$ with dynamic programming approach. The lower part also includes minimum and maximum length of the segments, i.e., $l_{\text{min}}$ and $l_{\text{max}}$, as parameters of the contending method.

| Algorithm 1 | $k$ Breakpoints | $\lambda_1$ | $\lambda_2$ |
|-------------|-----------------|-------------|-------------|
| 5           | 5, 10, 13 and 16| [1,50]      | [1,50]      |

| Bleakley et al. (2011) | $k$ Breakpoints |
|-------------------------|-----------------|
| 4                       | 5, 10 and 15    |

| Omranian et al. Penalized (number) Type | $k$ Breakpoints | $\nu_{\text{min}}$ | $\nu_{\text{max}}$ |
|----------------------------------------|-----------------|---------------------|---------------------|
| relative density                       | G 3 4 and 15    | 0.05                | 4.43                |
| degree                                 | L 3 5 and 15    | 5.78                | 12.60               |
| closeness                              | LG 4 4, 8 and 16| 0.05                | 5.71                |
| betweenness                            | LG 4 5, 9 and 13| 8.43                | 15.11               |

| Ramakrishnan et al. [15] | $k$ Breakpoints | $l_{\text{min}}$ | $l_{\text{max}}$ |
|--------------------------|-----------------|-----------------|-----------------|
| 4                        | 4, 11 and 14    | 3               | 7               |
null
| ProteinID | GeneSymbol | Genome | Description |
|-----------|------------|--------|-------------|
| chr2_1002284_-,chr2_1002424_-,chr2_1002564_- | hypothetical protein | | Endoplasmic reticulum targeting sequence |
| chr3 | PYK1 | 261601 | chr2_2623747_+,chr2_2623607_+,chr2_2623467_+ |
| chr17_15357_+,chr17_15217_+,chr17_15077_+ | PGAM2 | 36958 | chr6_1166008_+,chr6_1165868_+ |
| chr17_215819_-,chr17_215959_-,chr17_216099_- | ANT1 | 33312 | chr_7 |
| chr3_1694422_-,chr3_1694562_-,chr3_1694751_- | Aminotransferase, class V | 10542 | chr2_130741_+,chr2_130601_+,chr2_130144_+ |
| chr1_2589121_+,chr1_2588981_+,chr1_2588841_+ | tRNA/rRNA methyltransferase (SpoU) | 2159 | chr2_2571662_-,chr2_2571802_-,chr2_2571942_- |
| chr4_740514_+,chr4_740374_+,chr4_740234_+ | tRNA/rRNA methyltransferase (PyrK) | | |
| chr_8 | chr_20 | 22311 | chr19a_157141_-,chr19a_157406_-,chr19a_157618_- |
| chr15_611196_+,chr15_610955_+,chr15_610814_+ | tRNA/rRNA methyltransferase (PyrK) | | |
| chr3_2408537_+,chr3_2408397_+,chr3_2408257_+ | Glucose inhibited division protein | 6123 | chr7_871263_+,chr7_867123_+,chr7_863123_+ |
| chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |
| chr_20 | chr_4 | chr_10 | chr_17 | chr_19a_19 | Protein kinase | 14979 | chr15_913236_+,chr15_912696_+,chr15_912156_+ |