Coexpression Tool
User Manual
RegulonDB database
Lucia Pannier
This presentation aims to show the user how to utilize and interpret the coexpression tool.

Summary:

- **Motivation**
- **Gene expression and coexpression**
- **How is coexpression measured?**
- **Interpretation of coexpression values**
Motivation

High-throughput assays enable massive expression measurements of thousands of genes simultaneously. It is possible to measure the similarity of expression between genes to see if they are coexpressed. Coexpression of genes may indicate that the respective genes are controlled by the same transcriptional regulatory machinery, functionally related or they are members of the same pathway or protein complex.
thousands of microarrays, thousands of conditions for *Escherichia coli* K-12

**Coexpression PipeLine**

Measure of similarity of gene expression by pairs of genes, in all conditions for all genes

**Transformations**

- Spearman Correlation Coefficient (SCC) matrix for all genes vs all genes
- Transformed SCC into Rank (SCR) matrix

**Condition contrasts**

- araB
- araJ

**araB, araJ measures in all conditions**
What is coexpression?
Similarity of gene expression

2470 condition contrasts

All genes

araB
araJ

One circle represents expression of two genes in one condition **contrast**
= **test versus control** of one experiment

araB, araJ correlation in all experiments
pairwise correlation as a measure of similarity

Correlation Matrix (all genes vs all genes)

AraB AraJ

Slope of expression values = correlation of AraB with araJ
Pairwise correlation example cases

Correlation ~ 1 (high similarity)

Correlation ~ 0.5 (moderate similarity)

Correlation ~ 0 (no similarity)
Compute correlation of all genes vs all genes

Rank all contrasts for each gene

All genes

Rank all SCCs for each gene

Combine two ranked SCCs to obtain one SCR for any two genes

Spearman Correlation Coefficient (SCC)

Symmetrical matrix

SCC ∈ [-1,1]

Non-Symmetrical matrix

Rank ∈ [1,4167]

Combine two ranked SCCs to obtain one SCR for any two genes

Spearman Correlation Rank (SCR)

\[ SCR(A, B) = \frac{\sqrt{\text{rankedSCC}_{A(B)} \cdot \text{rankedSCC}_{B(A)}}}{\text{rankedSCC}_{A(B)} + \text{rankedSCC}_{B(A)}} \]
Ranking all SCC for a gene

Gene expression measurements of genes A, B and C in 5 experiments (conditions)

<table>
<thead>
<tr>
<th>Spearman Correlation coefficient (SCC)</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Gene C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>1</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Gene B</td>
<td>0.6</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Gene C</td>
<td>0.9</td>
<td>0.8</td>
<td>1</td>
</tr>
</tbody>
</table>

Non-Symmetrical matrix
Obtaining SCR from two ranked SCCs

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Gene C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>1</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Gene B</td>
<td>0.6</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Gene C</td>
<td>0.9</td>
<td>0.8</td>
<td>1</td>
</tr>
</tbody>
</table>

**Example:**
- Gene B’s **highest** correlation \( r = 0.8 \) is with gene C so **rank = 1**
- Gene C’s **second highest** correlation \( r = 0.8 \) is with gene B so **rank = 2**

Two genes B, C have two ranks (1) rank of gene B for gene C (2) rank of gene C for gene B
### Obtaining SCR from two ranked SCC

#### Ranked SCC

<table>
<thead>
<tr>
<th>Ranked SCC</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Gene C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>--</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Gene B</td>
<td>2</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td>Gene C</td>
<td>1</td>
<td>2</td>
<td>--</td>
</tr>
</tbody>
</table>

#### Geometric mean

Example:
Two genes B, C have two ranks
1. Rank of gene B for gene C
2. Rank of gene C for gene B

<table>
<thead>
<tr>
<th>SCR</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Gene C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>--</td>
<td>2</td>
<td>1.4</td>
</tr>
<tr>
<td>Gene B</td>
<td>2</td>
<td>--</td>
<td>1.4</td>
</tr>
<tr>
<td>Gene C</td>
<td>1.4</td>
<td>1.4</td>
<td>--</td>
</tr>
</tbody>
</table>

Two genes B, C have one mutual rank (1.4) = the geometrical mean of the two ranks
Obtaining SCR from two ranked SCC

Example (cont.) Calculating SCR

<table>
<thead>
<tr>
<th>Correlation Rank</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Gene C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene A</td>
<td>--</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Gene B</td>
<td>2</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td>Gene C</td>
<td>1</td>
<td>2</td>
<td>--</td>
</tr>
</tbody>
</table>

**SCR(A,B) = \sqrt{rankedSCC_{A(B)} \cdot rankedSCC_{B(A)}}**

Why geometric mean?
The geometric mean is closer to the smallest argument than to the largest argument. → the mutual rank tends to be closer to the “best” rank than to the “worst” rank (see Pannier L et al, *in prep.* for more detailed explanation)
EVALUATION OF THE SCR

RANGE : 1 until 4167 (number of genes -1)

INTERPRETATION : If two genes have SCR=1, it means that these two genes have the highest similarity with each other compared to the similarities they have with all other genes

If two genes have SCR=4167, it means these two genes have the lowest similarity with each other compared to the similarities with all other genes.