Coexpression Tool

User Manual

RegulonDB database

Lucia Pannier

This presentation aims to show the user how to utilize and interprete the coexpression tool.

Summary:

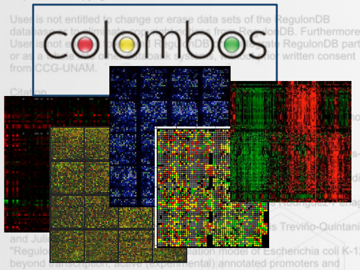
- Motivation
- Gene expression and coexpression
- How is coexpression measured?
- Interpretation of coexpression values

Motivation

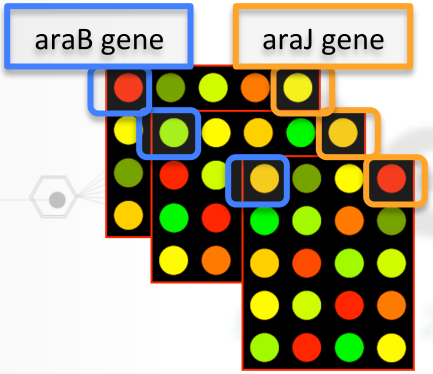
High-throughput assays enable **massive expression measurements** of thousands of genes **simultaneously**. It is possible to measure the similarity of expression between genes to see if they are coexpressed.

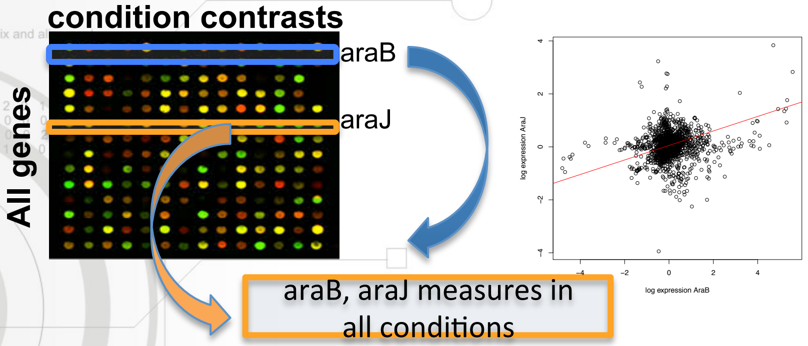
Coexpression of genes may indicate that the respective genes are controlled by the same transcriptional regulatory machinary, functionally related or they are members of the same pathway or protein complex.

Coexpression PipeLine



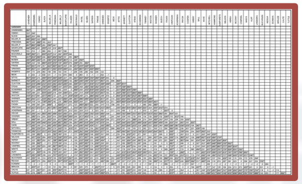
thousands of microarrays , thousands of conditions for *Escherichia coli* K-12



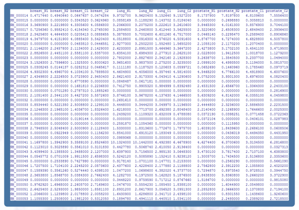


Measure of similarity of gene expression by pairs of genes, in all conditions for all genes

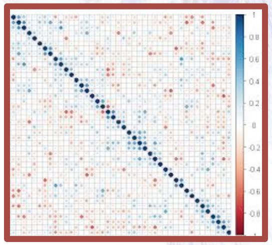
Transformations



Spearman Correlation Coefficient (SCC) matrix for all genes vs all genes



Transformed SCC into Ranked SCC matrix

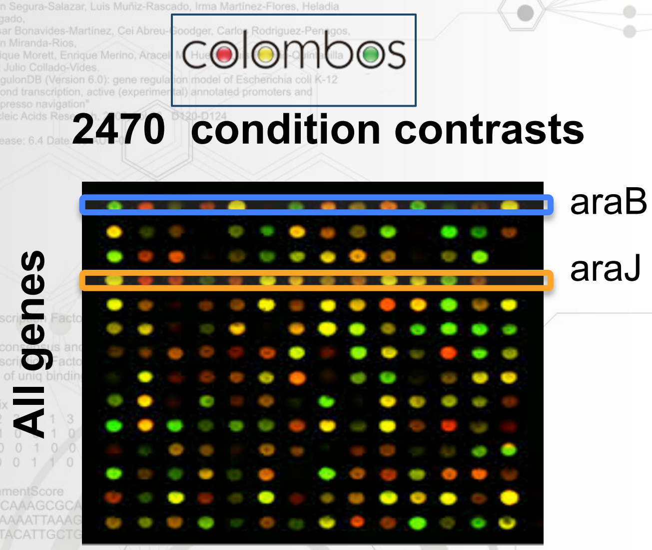


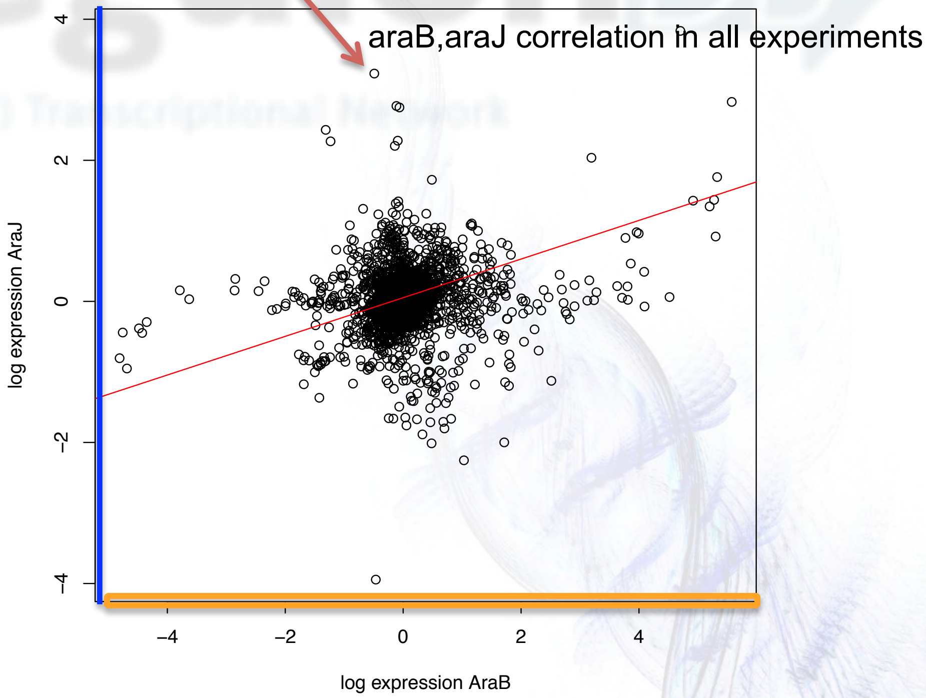
Spearman Correlation Rank (SCR) matrix

What is coexpression ?
Similarity of gene expression

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

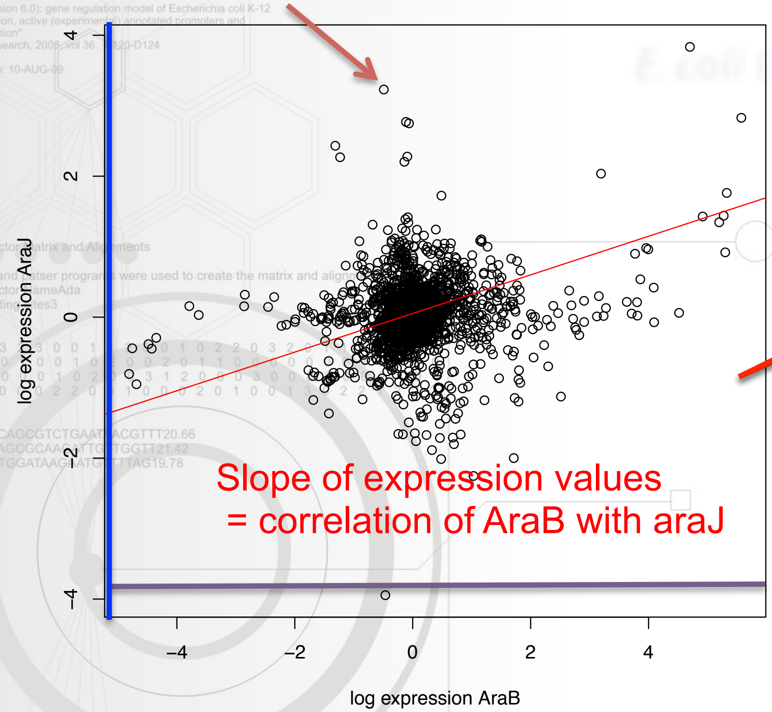
2470 condition contrasts





pairwise correlation
as a measure of similarity

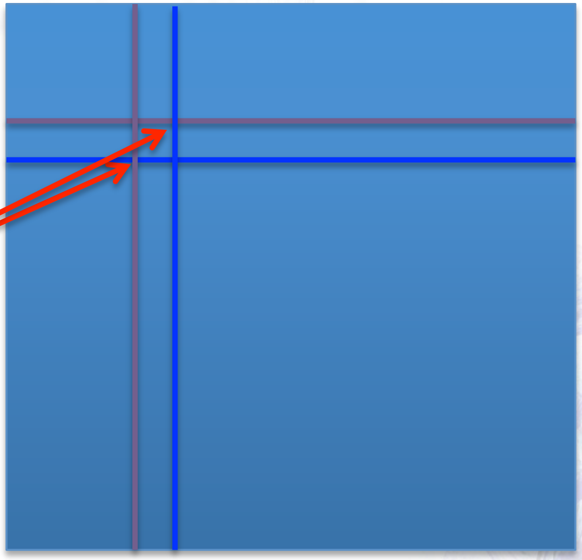
1 condition contrast



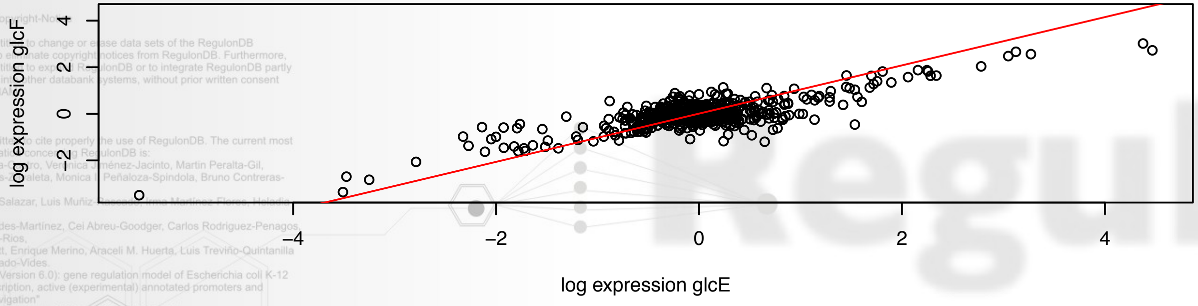
Slope of expression values
= correlation of AraB with araJ

Correlation Matrix
(all genes vs all genes)

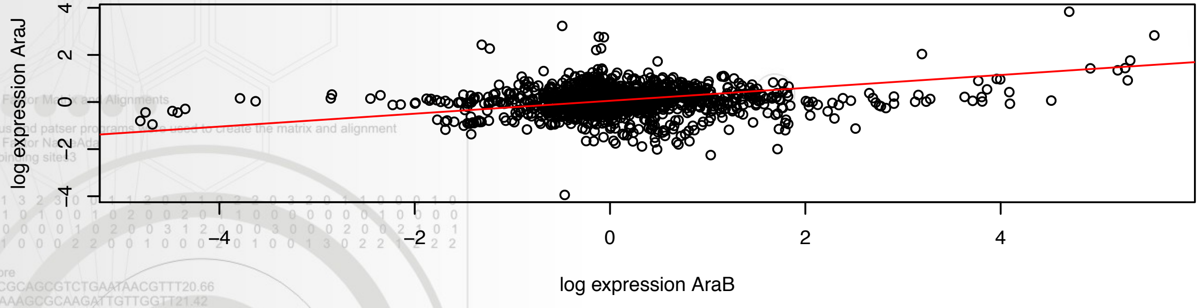
AraB AraJ



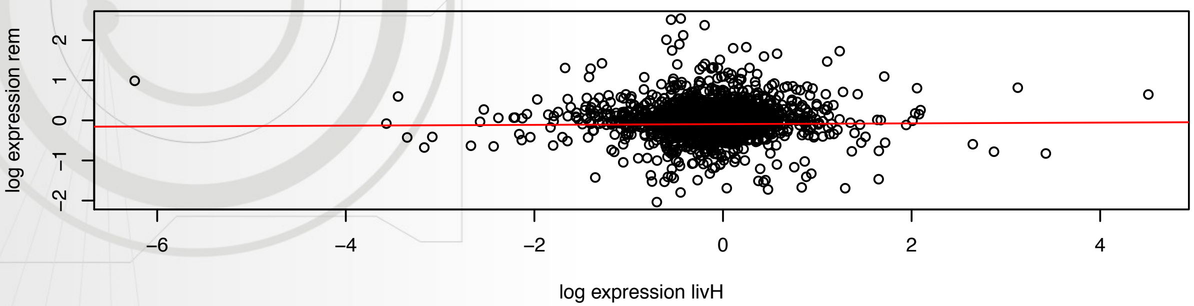
Pairwise correlation example cases



Correlation ~ 1
(high similarity)

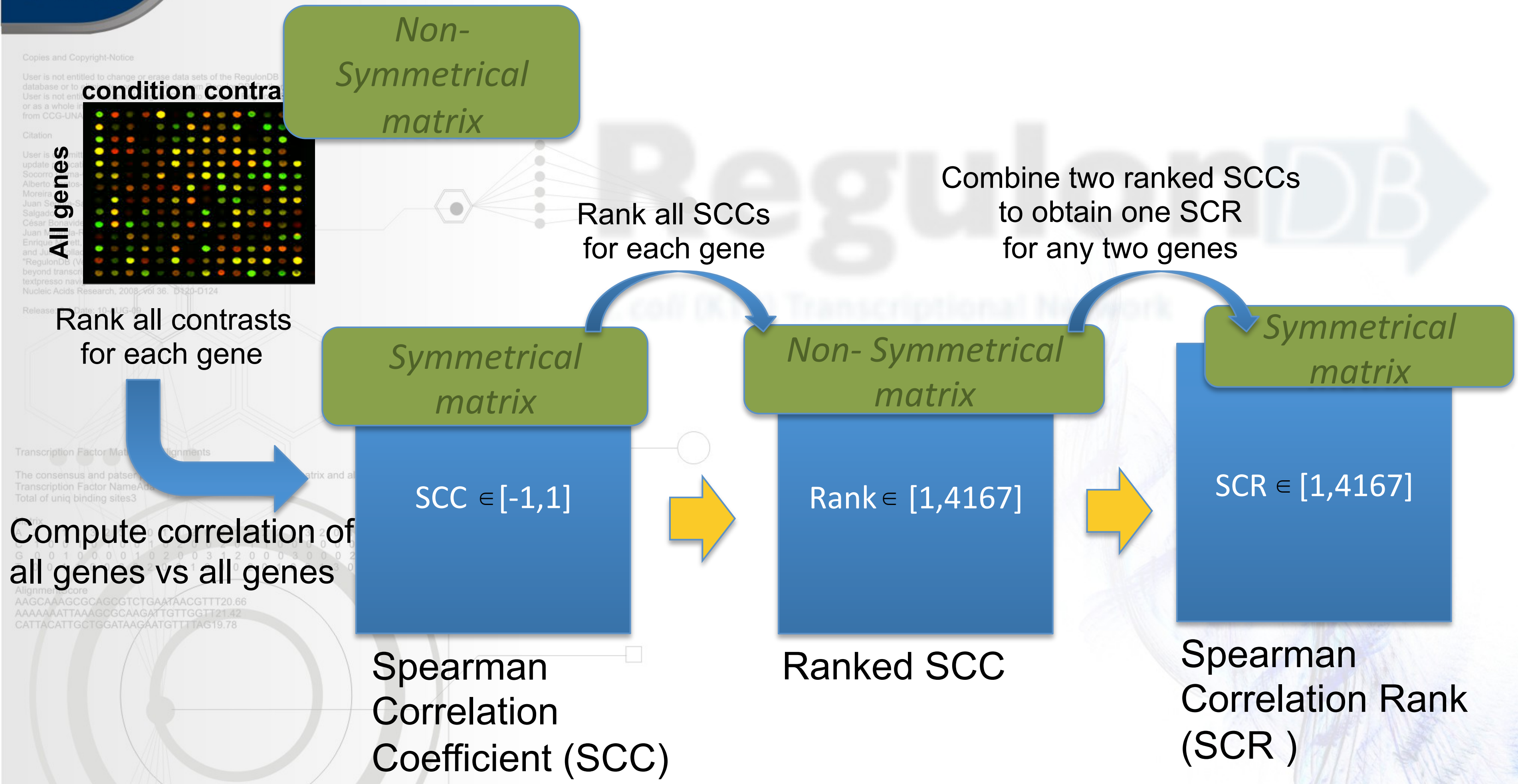


Correlation ~ 0.5
(moderate similarity)



Correlation ~ 0
(no similarity)

Transformations



condition contra

All genes

Non-Symmetrical matrix

Rank all contrasts for each gene

Compute correlation of all genes vs all genes

Symmetrical matrix

SCC ∈[-1,1]

Spearman Correlation Coefficient (SCC)

Rank all SCCs for each gene

Non- Symmetrical matrix

Rank ∈ [1,4167]

Ranked SCC

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

Symmetrical matrix

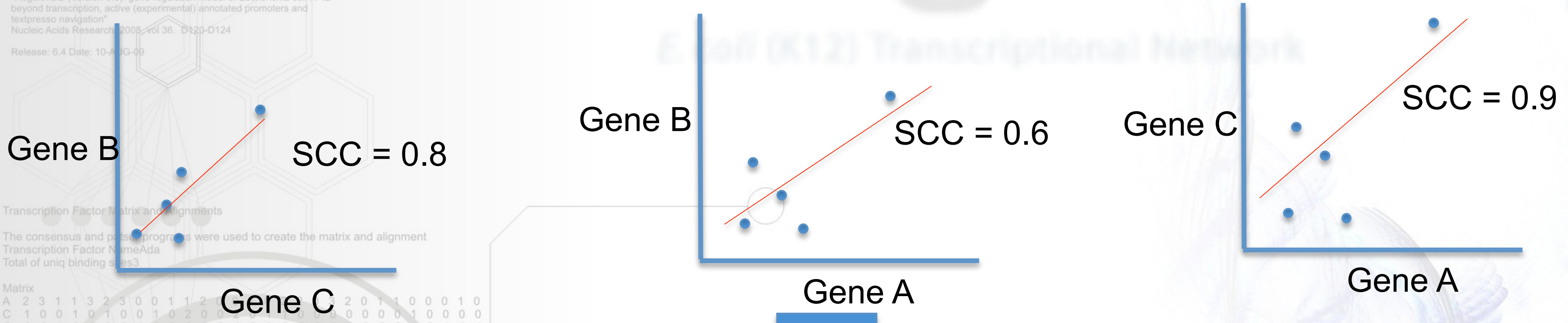
SCR ∈ [1,4167]

Spearman Correlation Rank (SCR)

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

Ranking all SCC for a gene

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



Spearman Correlation coefficient (SCC)	Gene A	Gene	
Gene A	1	0.6	0.9
Gene B	0.6	1	0.8
Gene C	0.9	0.8	1

Non- Symmetrical matrix

Obtaining SCR from two ranked SCC

Correlation coefficient	Gene A	Gene B	Gene C
Gene A	1	0.6	0.9
Gene B	0.6	1	0.8
Gene C	0.9	0.8	1

Symmetrical matrix

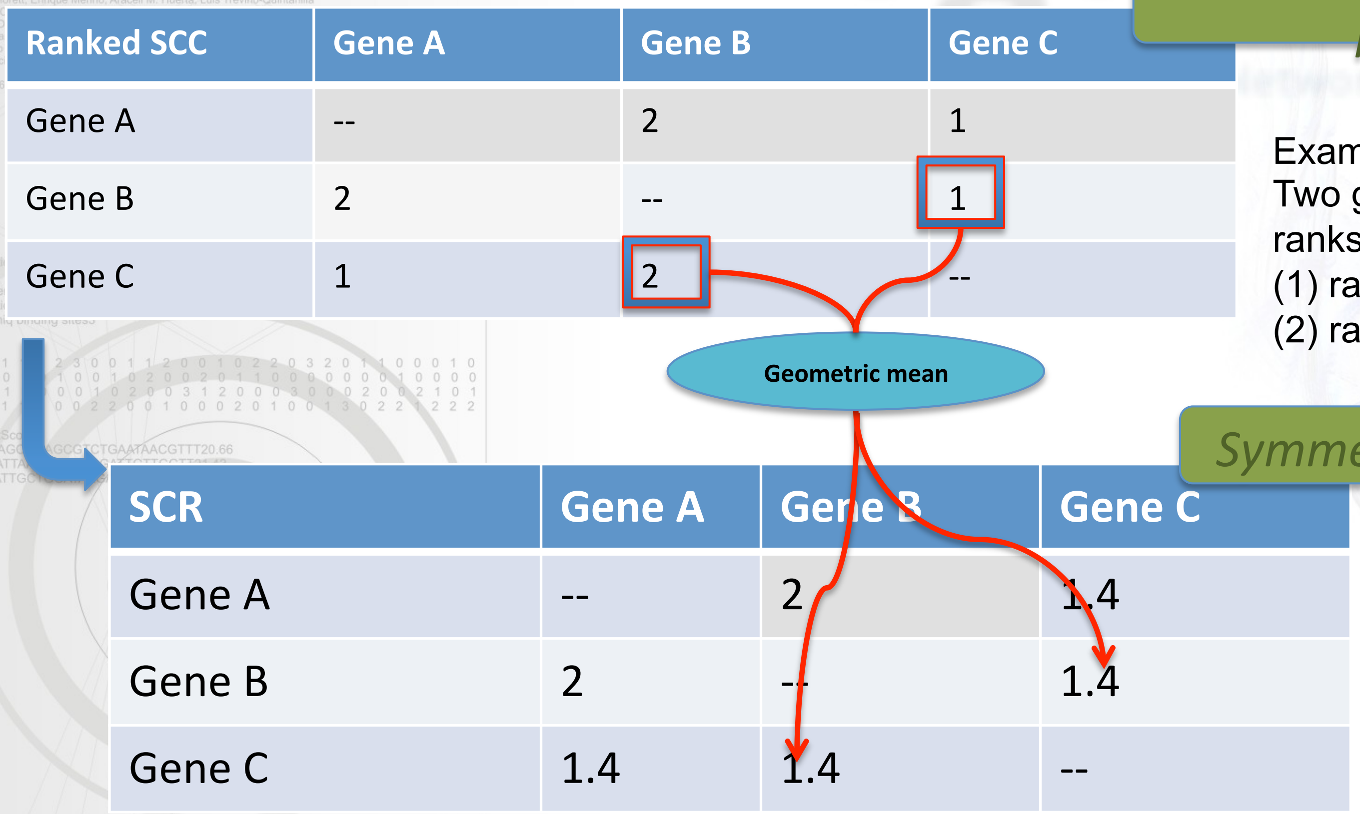
Example :
Gene B's **highest** correlation (r = 0.8)
is with gene C so **rank = 1**
Gene C's **second highes**t correlation (r=0.8)
is with gene B so **rank = 2**

Non- Symmetrical matrix

Correlation Rank	Gene A	Gene B	Gene C
Gene A	--	2	1
Gene B	2	--	1
Gene C	1	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 SC



Non- Symmetrical matrix

Ranked SCC	Gene A	Gene B	Gene C
Gene A	--	2	1
Gene B	2	--	1
Gene C	1	2	--

Example :
Two genes B,C have two ranks
(1) rank of gene B for gene C
(2) rank of gene C for gene B

Geometric mean

Symmetrical matrix

SCR	Gene A	Gene B	Gene C
Gene A	--	2	1.4
Gene B	2	--	1.4
Gene C	1.4	1.4	--

Two genes B,C have one mutual rank (1.4) = the geometrical mean of the two ranks

Obtaining SCR from two ranked SCC

Non- Symmetrical matrix

Correlation Rank	Gene A	Gene B	Gene C
Gene A	--	2	1
Gene B	2	--	1
Gene C	1	2	--

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

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)

Symmetrical matrix

Mutual Correlation Rank	Gene A	Gene B	Gene C
Gene A	--	2	1.4
Gene B	--	--	1.4
Gene C	--	--	--

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