

## Supplementary material

Computational insights on the molecular  
mechanisms across breast cancer progression  
combining gene differential expression and  
co-expression \*

Emmanouil K. Ikonomakis  
*Computational Intelligence Laboratory*  
*Department of Mathematics*  
*University of Patras*  
Patras, Greece  
eki@math.upatras.gr

Marilena M. Bourdakou  
*Laboratory of Pharmacology*  
*Faculty of Medicine*  
*Democritus University of Thrace*  
Alexandroupolis, Greece  
kmarlen1988@gmail.com

George Kolios  
*Laboratory of Pharmacology*  
*Faculty of Medicine*  
*Democritus University of Thrace*  
Alexandroupolis, Greece  
gkolios@med.duth.g

Michael N. Vrahatis  
*Computational Intelligence Laboratory*  
*Department of Mathematics*  
*University of Patras*  
Patras, Greece  
vrahatis@math.upatras.gr

George M. Spyrou  
*The Cyprus Institute of Neurology and Genetics*  
*The Cyprus School of Molecular Medicine*  
Ayios Dometios, Nicosia, Cyprus  
georges@cing.ac.cy

## Abstract

As a complex disease with mechanisms that are not fully understood, breast cancer pathology and progression is expected to be determined not only by individual genes, but also by their coordinated effects through a more systemic framework. In this contribution, we combine gene differential analysis (single gene view) with gene co-expression analysis (systemic view) to provide insights on the implicated molecular mechanisms across breast cancer progression. Important gene-gene links in a gene co-expression network are identified and clustered with DeDiCo algorithm, which has been recently published by members of our groups. This algorithm presents very good performance regarding its accuracy on both synthetic and real-life datasets and ability to determine arbitrarily shaped clusters without any assumptions on the shape and the number of the clusters. This work provides a pipeline, exploiting a novel clustering algorithm, in order to analyse gene-gene links based on their co-expression relationships and gene differential analysis across breast cancer progression stages. Existing bibliography verifies the validity of the resulted clusters while exclusive pathways per breast cancer stage as well as a common pathway signature across all breast cancer stages are reported and proper discussion takes place.

---

\*George M. Spyrou holds the Bioinformatics ERA Chair, The Cyprus Institute of Neurology and Genetics, 6 International Airport Avenue, 2370 Ayios Dometios, Nicosia, Cyprus, funded by the European Commission Research Executive Agency (REA) Grant BIORISE (Num. 669026), under the Spreading Excellence, Widening Participation, Science with and for Society Framework. Marilena M. Bourdakou's postdoctoral research is being funded by the IKY (State Scholarships Foundation) scholarship, co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the project "Reinforcement of Postdoctoral Researchers - 2nd Cycle" (MIS5033021), implemented by the State Scholarships Foundation (IKY). This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility – ARIS – under project ID PA070211-DeDiCoBC.

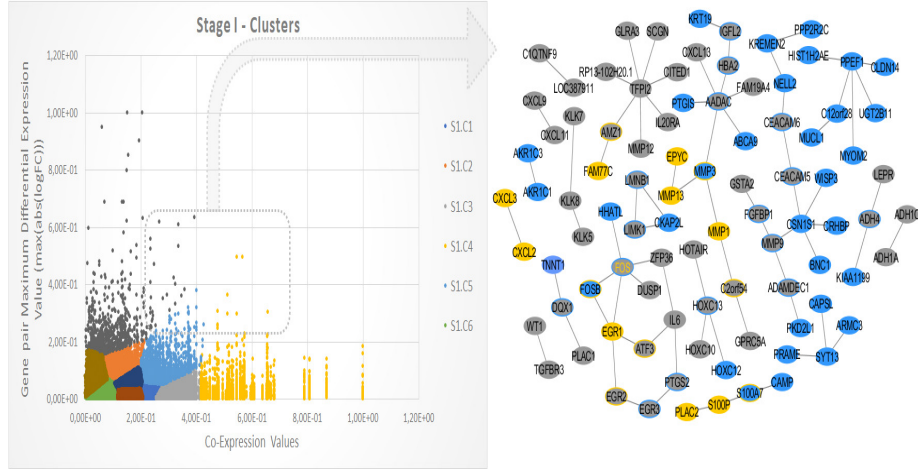


Figure 1: Stage I: Scatterplot regarding the gene-pair maximum differential expression value vs gene-pair co-expression value. A rectangular region in this scatterplot with down-left corner at (Dif.Expr. Threshold, Coexpr.Threshold) = (0.25, 0.25) includes the best gene-gene links of simultaneous high co-expression and dif. expression values. The small network pieces that correspond to these links are embedded in a picture-in-picture mode.

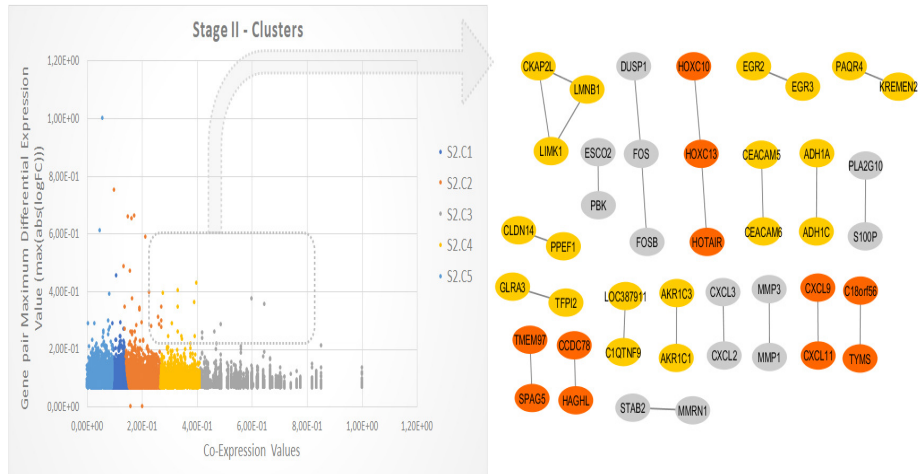


Figure 2: Stage II: Scatterplot regarding the gene-pair maximum differential expression value vs gene-pair co-expression value. A rectangular region in this scatterplot with down-left corner at (Dif.Expr. Threshold, Coexpr.Threshold) = (0.25, 0.25) includes the best gene-gene links of simultaneous high co-expression and dif. expression values. The small network pieces that correspond to these links are embedded in a picture-in-picture mode.

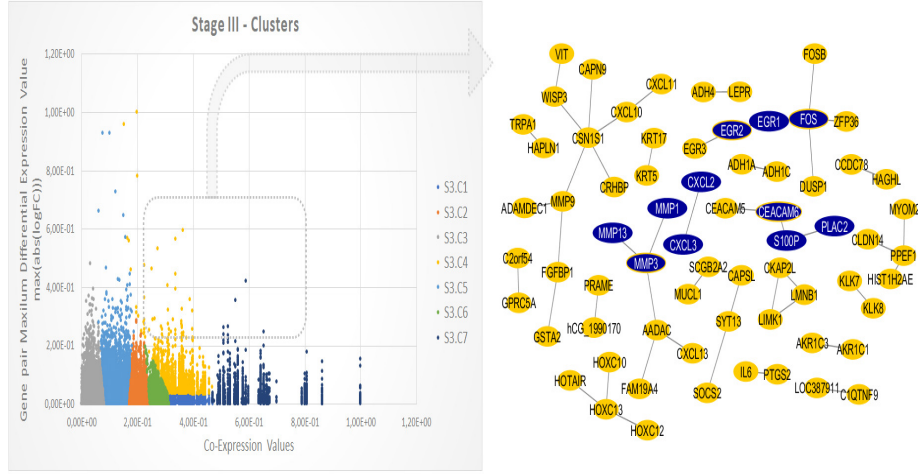


Figure 3: Stage III: Scatterplot regarding the gene-pair maximum differential expression value vs gene-pair co-expression value. A rectangular region in this scatterplot with down-left corner at (Dif.Expr. Threshold, Coexpr.Threshold) = (0.25, 0.25) includes the best gene-gene links of simultaneous high co-expression and dif. expression values. The small network pieces that correspond to these links are embedded in a picture-in-picture mode.

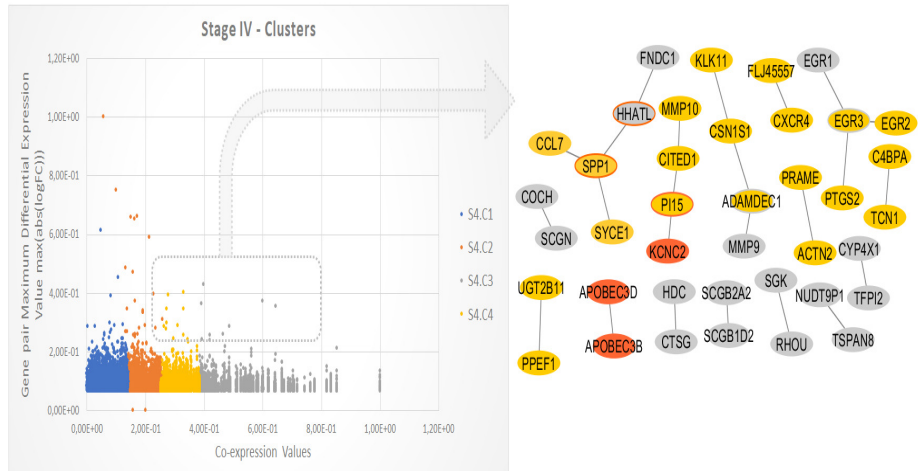


Figure 4: Stage IV: Scatterplot regarding the gene-pair maximum differential expression value vs gene-pair co-expression value. A rectangular region in this scatterplot with down-left corner at (Dif.Expr. Threshold, Coexpr.Threshold) = (0.25, 0.25) includes the best gene-gene links of simultaneous high co-expression and dif. expression values. The small network pieces that correspond to these links are embedded in a picture-in-picture mode.