

## New metaheuristics approaches for biclustering of gene expression data

F. Musacchia<sup>1</sup> ✉, A. Marabotti<sup>2</sup>, A. Facchiano<sup>3</sup>, L. Milanese<sup>2</sup>, P. Festa<sup>1</sup>

<sup>1</sup>Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università Federico II, Napoli, Italy

<sup>2</sup>Istituto di Tecnologie Biomediche - CNR, Milano, Italy

<sup>3</sup>Istituto di Scienze dell' Alimentazione - CNR, Avellino, Italy

### Motivations

Biclustering or simultaneous clustering of both genes and conditions have generated considerable interest over the past few decades, particularly related to the analysis of high-dimensional gene expression data in information retrieval, knowledge discovery, and data mining [1]. Given a gene expression data matrix, a bicluster is a submatrix of genes and conditions that exhibits a high correlation of expression activity across both rows and columns. The problem of locating the most significant bicluster has been shown to be NP-complete. Therefore, given the inner "intractability" of the problem from a computational point of view, to efficiently find good solutions in a reasonable running times we have designed and implemented several different metaheuristics based on a GRASP framework.

### Methods

All the procedures are based on a GRASP (Greedy Randomized Adaptive Search Procedure) framework [2]. A GRASP is a randomized multistart iterative metaheuristic consisting of two phases: a construction phase and a local search phase. The construction phase builds iteratively a feasible solution in a greedy randomized fashion. Once a feasible solution is obtained, a local search procedure attempts to improve it by producing a locally optimal solution with respect to suitable defined neighborhood structure. The construction and the local search phases are repeatedly applied. The best local optimum solution found is returned as an approximation of the optimal one. We designed and implemented several different GRASP-based metaheuristics that differ in both construction and local search phase. In the construction phase, one implements a k-means and a second one a greedy randomized procedure inspired by a minimum spanning tree of a suitable weighted graph. Then, two types of lo-

cal searches have been implemented: one has been already proposed [3] and a second one is an Iterated Local Search [4]. All the designed algorithms have been tested and compared using the lymphoma dataset [5].

### Results

Both the solution construction procedure and the local search procedure use a gain function that combines the mean squared residue [1], the row variance, and the size of the bicluster. Experimental results on Lymphoma microarray data set [5] are promising indicating that the methods are able to find significant biclusters, also from a biological point of view.

### Acknowledgements

A.M. and L.M. are supported by MIUR FIRB ITALBIONET (RBPR05ZK2Z and RBIN064YAT\_003). The work has been made in the frame of the Flagship Project InterOmics.

### References

1. Y.Cheng and G. Church. (2000) Biclustering of Expression Data, Proc. Int. Conf. Intell. Syst. Mod. Biol., 93-103.
2. T.A. Feo and M. G.C. Resende (1995) Greedy Randomized Adaptive Search Procedures, J. Global Optim., 6, 109-134.
3. F. Musacchia, A. Marabotti, A. Facchiano, L. Milanese, and P. Festa (2011) Biclustering of gene expression data based on GRASP-like algorithms. BITS2011, ISBN 978-884673069-5, pp. 100-101.
4. W.Ayadi, M. Elloumi and J.-K. Hao (2010) Iterated Local Search for Biclustering of Microarray Data. Lect. Notes Comput. Sci., 6282, 219-229
5. A.A. Alizadeh, M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu, J.I. Powell, L. Yang, G.E. Marti, T. Moore, J. Jr Hudson, L. Lu, D.B. Lewis, R. Tibshirani, G. Sherlock, W.C. Chan, T.C. Greiner, D.D. Weisenburger, J.O. Armitage, R. Warnke, R. Levy, W. Wilson, M.R. Grever, J.C. Byrd, D. Botstein, P.O. Brown, L.M. Staudt (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 403, 503-511