Elisa Ficarra[1], Giovanni De Micheli[1], Sungroh Yoon[2], Luca Benini[3], Enrico Macii[4]

[1] EPFL Ecole Polytechnique Fédérale
[2] Stanford University, Computer Science Dep.
[3] University of Bologna, DEIS
[4] Politecnico di Torino, DAUIN

## Bioimaging and functional genomics

For better understanding of genetic mechanisms underlying clinical observations, it is interesting to determine which genes and clinical traits are interrelated. In the last few years a consistent amount of research in genomics has been done concerning correlation of gene expression to multi-factorial genetic pathologies. Microarray data analysis, as well as real-time PCR, are useful techniques exploited so far to this purpose [1] [2]. Despite this effort, results obtained are strongly limited by the poor informative content provided by clustering techniques applied to gene expression data [3].

At the same time, in the field of biomedical and molecular imaging, new techniques have been shown to be effective in extracting clinical and functional biological information from images of molecules and tissues [4] [5]. By observing processes as they happen within the cell, these techniques add an important extra dimension to the understanding of cell behavior and functioning for early disease detection and drug response. In clinics, new applications of conventional imaging technologies are likely to play increasingly important roles, particularly in oncology.

Up to now, these two independent sources of information, namely gene expression analysis techniques and bioimaging, have never been correlated to enhance gene expression analysis or to increase the amount of confidence in the hypothesized gene expression paths. For this purpose, we developed a joint co-clustering technique able to extract clinical bioimaging parameters through a fully-automated computer-aided approach and to perform coclustering technique between clinical bioimaging parameters and gene expression data.

Our proposed method consists thus of two steps. As first step, we developed two tools. The first one is a computational method, namely Co-clustering, that can find co-clusters or groups of genes and clinical parameters that are believed to be closely related to each other based upon given empirical information. The second tool is a fully-automated tissue image processing method able to extract a set of clinical parameters that give a characterization of the pathology dynamics. The fist tool was successfully tested on Acute Myelogenous Leukemia (AML) data. Through co-clustering method statistically significant co-clusters of genes and clinical traits have been identified [6]. The clinical traits used in this test were obtained in a non-automated way by pathologists inspection of tissue samples and images. The second tool was successfully tested on non-small cell lung carcinoma (NSCLC) tissue images in order to characterize and quantify, in a standardized way, the activation of the EGFR/erb-B protein receptor family that plays an important role in non small cell lung carcinoma growing [7]. This type of analysis aims at characterizing each pathological cell, and on average the whole tissue, by performing a standardized quantitative and qualitative measurement of protein activations. This information can be treated as a clinical parameter, and can be finally correlated with the genetic expression data on same lung carcinoma tissue in order to better define a group of potential candidates to protein family-inhibiting therapy. For this purpose, we are developing the proposed fully-automated joint co-clustering approach to find correlations between genetic data and clinical and bioimaging parameters. Preliminary results show that it is a very promising approach to analyze largescale biological data and to study multi-factorial genetic pathologies through their genetic alterations. Moreover, this approach enables new opportunities for early diagnosis and provides information in future strategies for therapy.