Bioimaging and Functional Genomics

Elisa Ficarra, *EPF Lausanne*
Giovanni De Micheli, *EPF Lausanne*
Sungroh Yoon, *Stanford University*
Luca Benini, *University of Bologna*
Enrico Macii, *Politecnico di Torino*

Outline

- Context and aim
- Linking genes with clinical traits
- Joint co-clustering
  - Co-clustering
  - Automated clinical image processing
- Conclusion

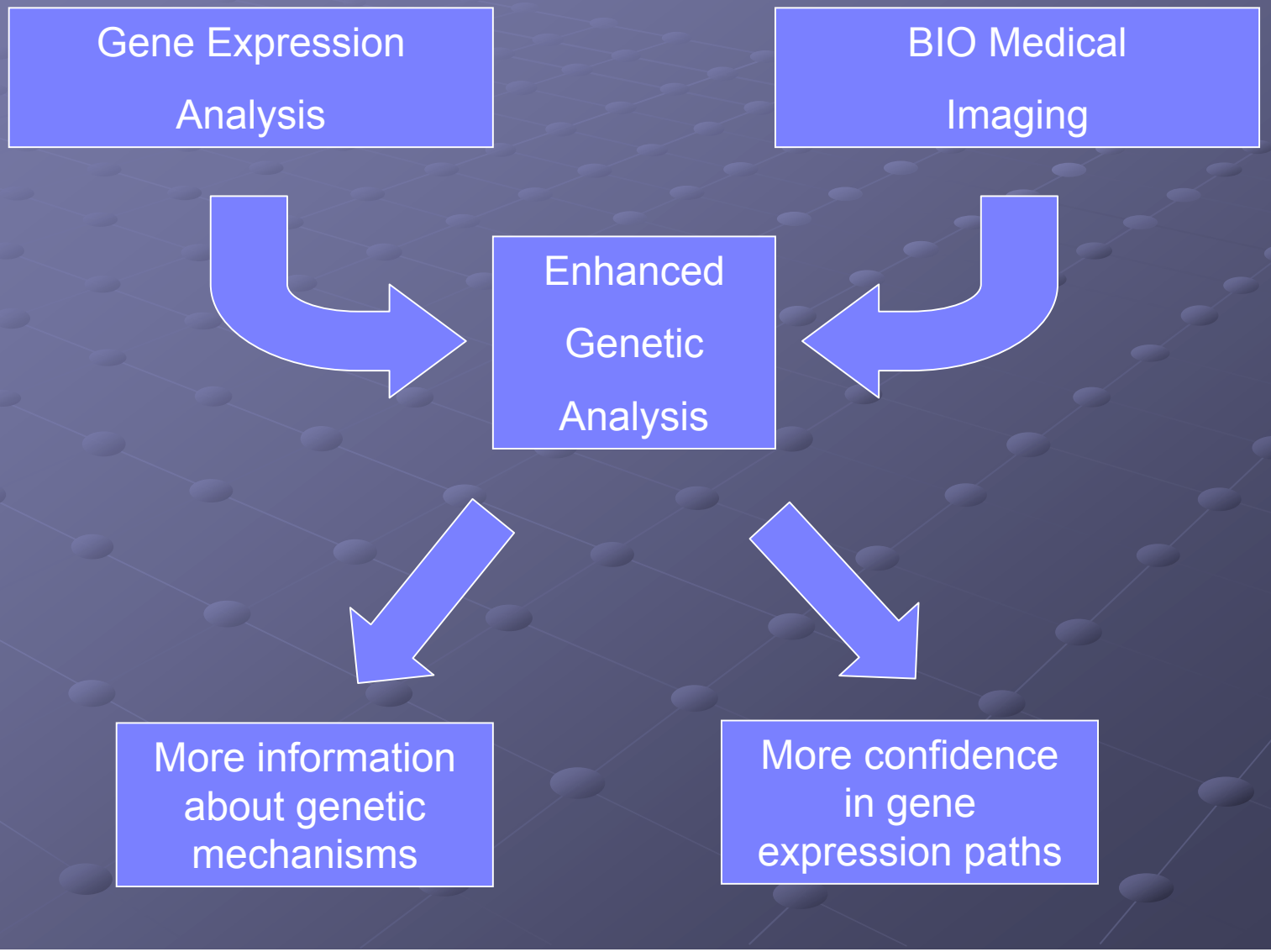
Context and Aim

- Study of multi-factorial genetic pathologies
- Improve clinical diagnosis and develop specific new therapy
  - Understanding of genetic mechanisms underlying clinical observations
  - Increase the amount of confidence in the hypothesized gene expression paths

Approach

- Linking genes with clinical traits
- How?
  - Correlation of gene expression with clinical trait measurement
  - Clinical traits can be obtained through several techniques
- Example:
  - Biopsy genetic analysis and tissue imaging data

Solution


Gene Expression Analysis
BIO Medical Imaging
Enhanced Genetic Analysis
More information about genetic mechanisms
More confidence in gene expression paths

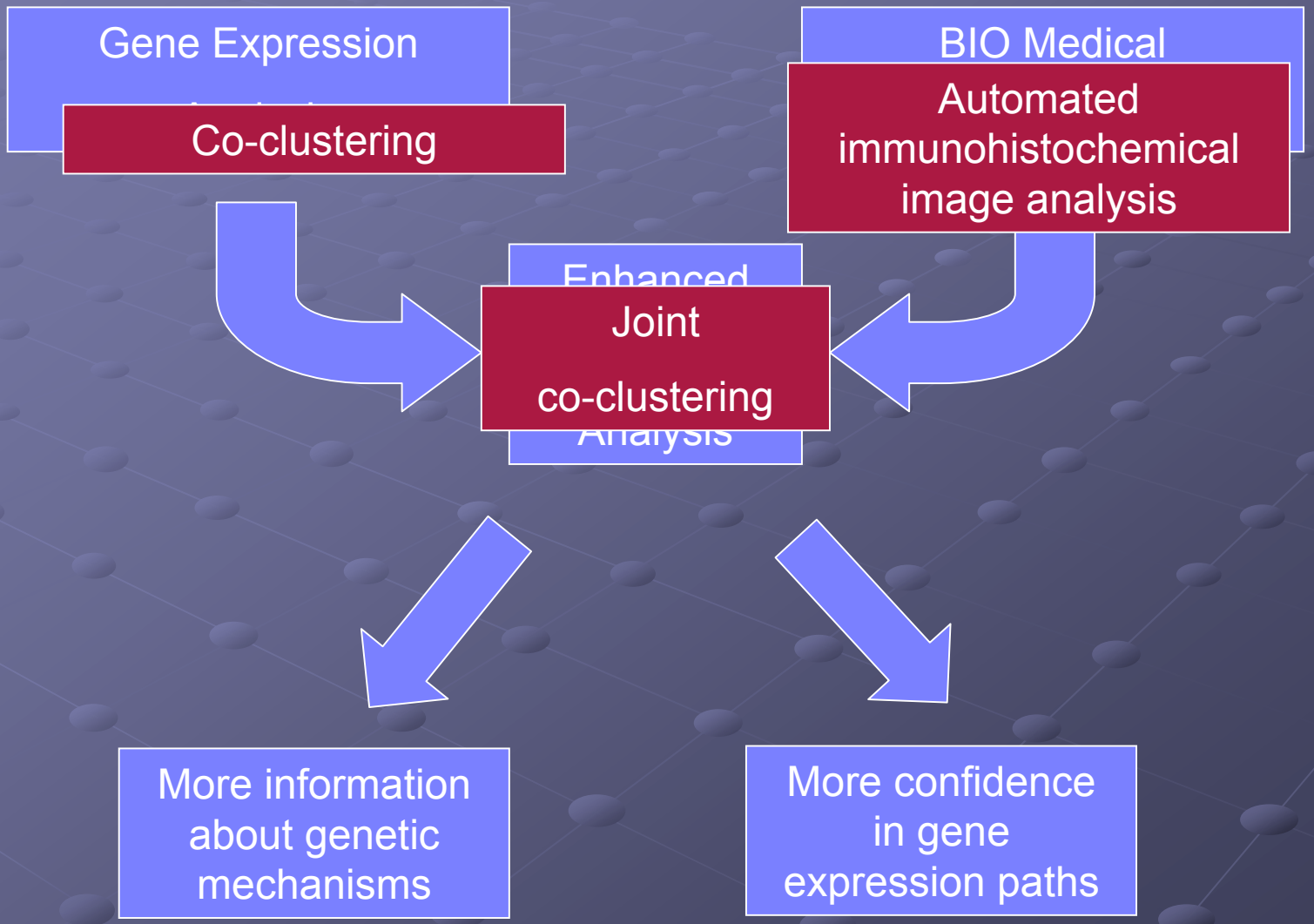
Gene Expression Analysis

- Objective
  - Correlation of gene expression with multi-factorial genetic pathologies
- Techniques exploited so far
  - Real-time PCR (qPCR)
  - Microarray data analysis
- Drawback
  - Information provided by genetic clustering techniques is not sufficient to draw pathological conclusions

Bio and Molecular Imaging

- Extraction of clinical and functional biological information from images of molecules and tissues
- Important roles in clinics
  - E.g, study protein activations in carcinoma tissues
- Objectives
  - Early disease detection
  - Drug response
  - New Therapies

Our Solution



Joint co-clustering

- Correlation between
  - Gene expression analysis
  - Bioimaging
- *Joint co-clustering*:
  - *Fully-automated computer-aided image analysis approach* to extract clinical bioimaging parameters
  - *Co-clustering* applied to gene expression data and clinical bioimaging parameters

Outline

- Context and aim
- Linking genes with clinical traits
- Joint co-clustering
  - Co-clustering
  - Automated clinical image processing
- Conclusion

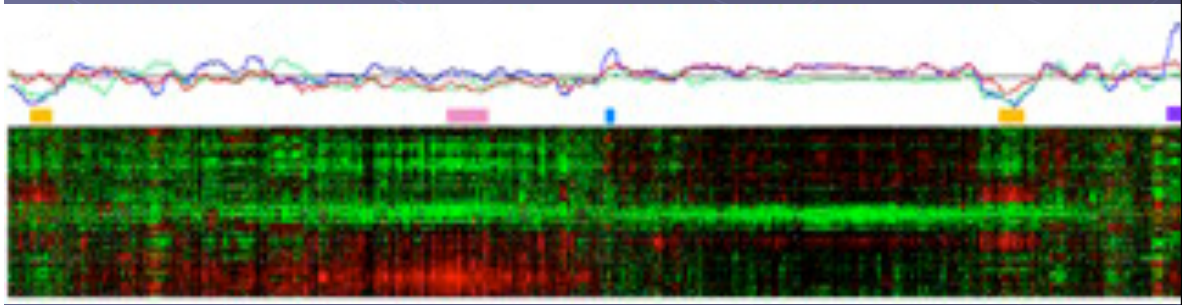
Co-clustering

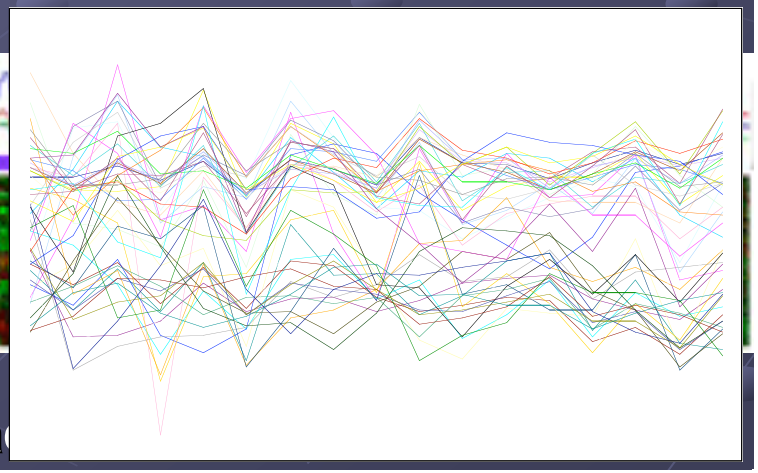
- Unsupervised cluster search method in high dimensional spaces
- Linking genes and clinical traits
  - Provides qualitative information about pathology

S. Yoon, L. Benini, and G. De Micheli, Finding Co-Clusters of Genes and Clinical Parameters, *IEEE EMBC 2005* no. 11.3.1.4, 2005

Challenges

- As the number of clinical traits increases, the inspection method breaks down
  - We need a systematic approach
- The values of clinical traits are not necessarily continuous or numeric
  - We need a more general statistics than the Pearson correlation coefficient





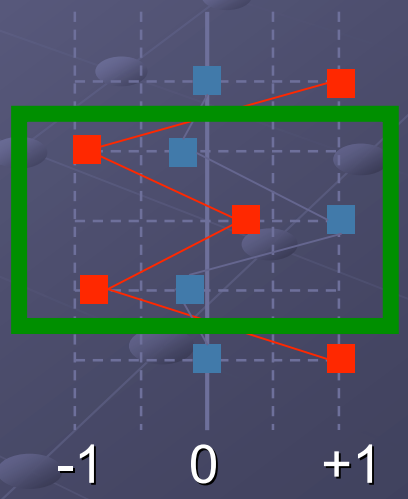
Lapointe, Jacques et al. (2004) Proc. Natl. Aca

Definition

- *Co-cluster* of genes and clinical traits
  - A submatrix of the correlation matrix
  - Elements are statistically significant
  - For any pair of column vectors, the inter-column distance is less than a threshold

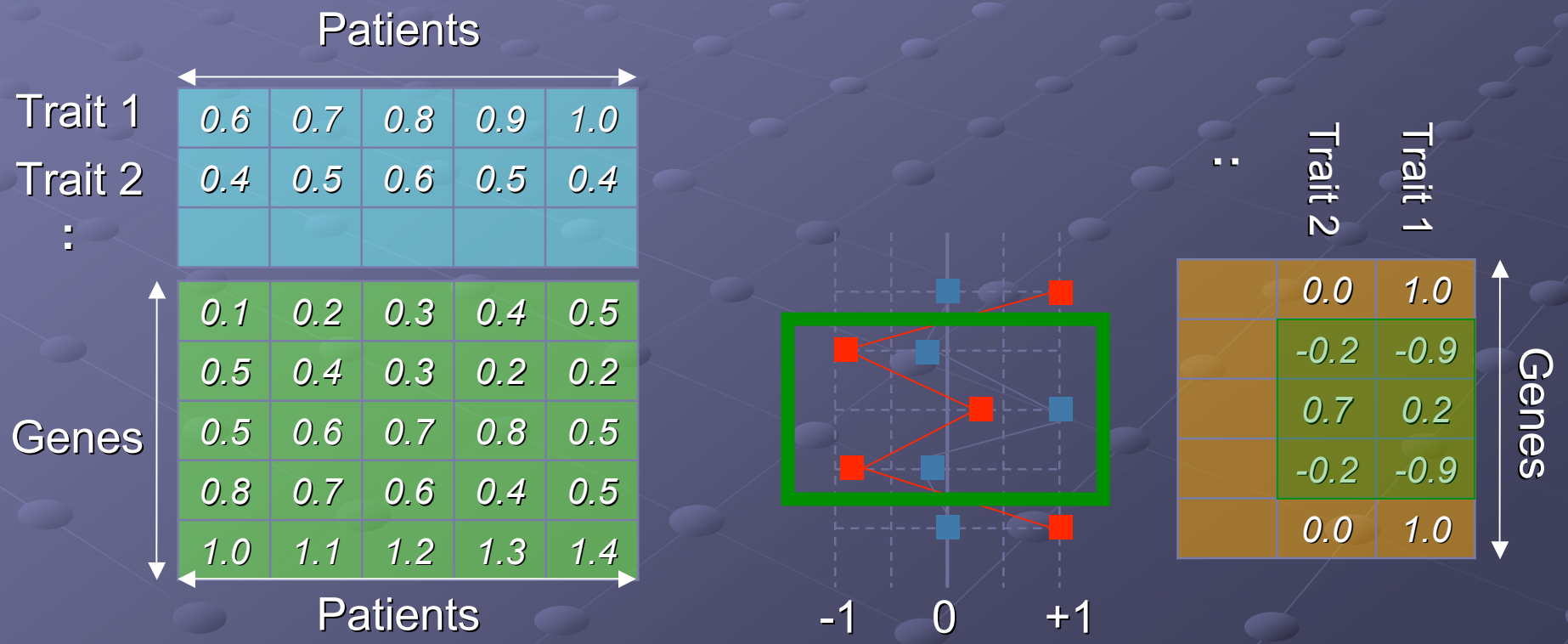
..	Trait 2	Trait 1	
	0.0	1.0	
	-0.2	-0.9	Genes
	0.7	0.2	
	-0.2	-0.9	
	0.0	1.0	

Correlation matrix



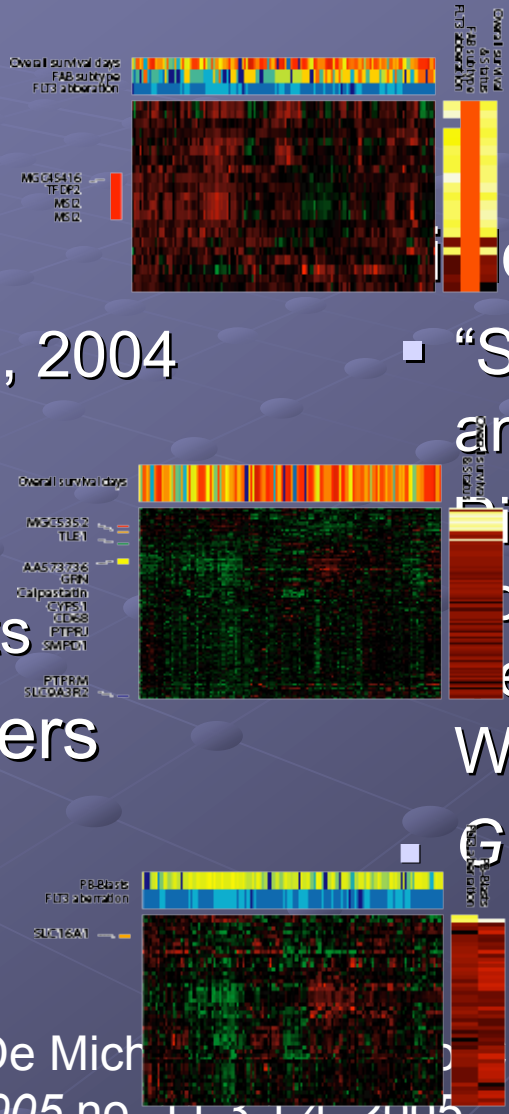
Our Approach

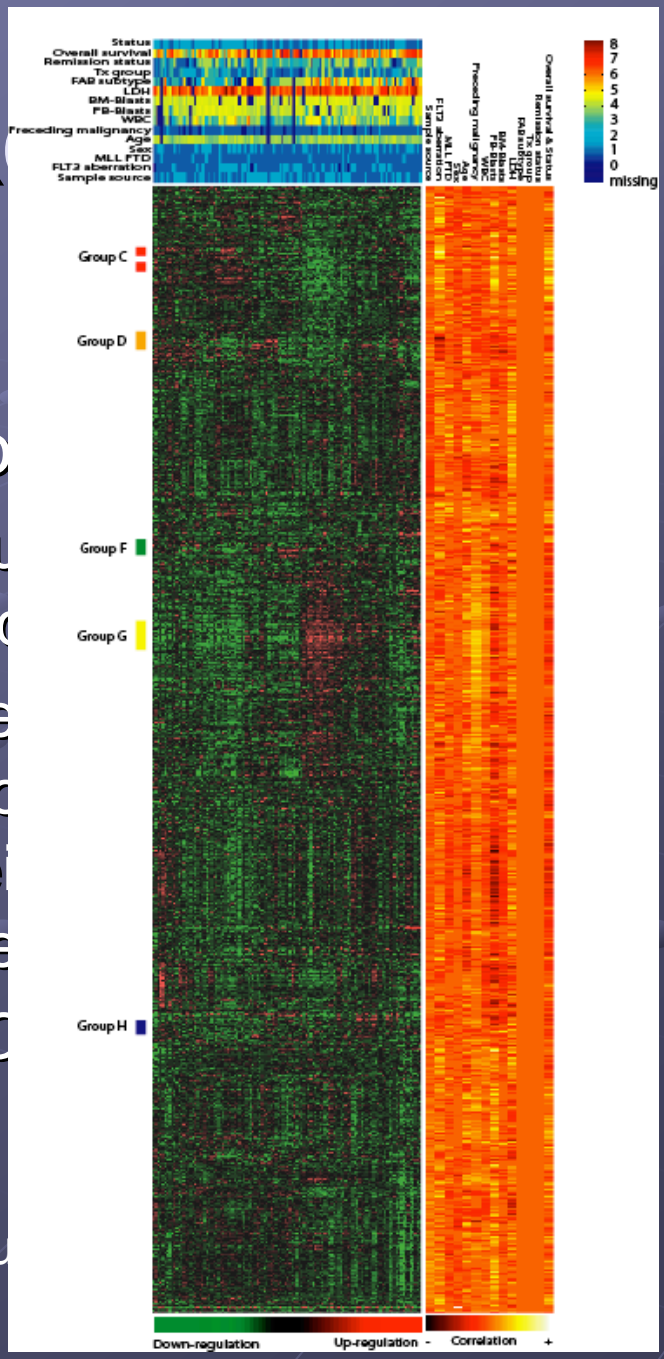
- Construct a “correlation matrix”
  - Use the *SAM* statistic (Tusher *et al., “Significance analysis of microarrays applied to the ionizing radiation response”*, 2001)
- Find co-clusters of genes and traits appearing on the correlation matrix



Experimental R

- AML data set
  - Bullinger *et al.*, 2004
  - 6283 genes
  - 119 patients
  - 15 clinical traits
- 43 GT co-clusters





- "Su
- anc
- ie
- co
- ei
- We
- GO

S. Yoon, L. Benini, and G. De Mich lu
Parameters, *IEEE EMBC 2005* no. 11.3.1.4, 2005

Outstanding Issue

- Extraction clinical parameter from image
- Can this be automated?

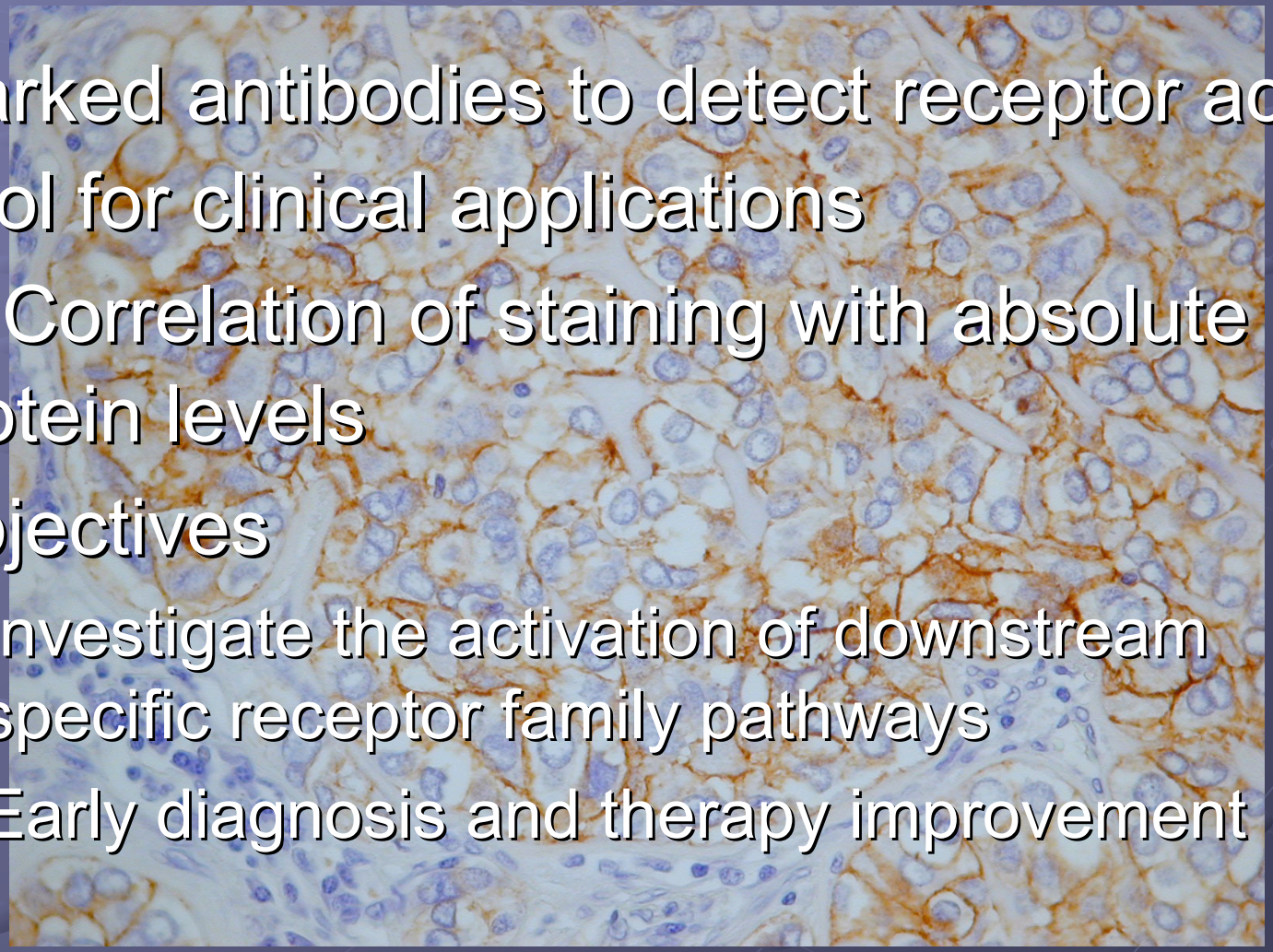
Immunohistochemical (IHC) automated quantification

- Development of bioimaging techniques for acquiring quantitative and qualitative information from immunostains

⇓

- Development of an automated image processing method to standardize IHC analysis

Immunohistochemistry (IHC): characteristics and aim



- Marked antibodies to detect receptor activity
- Tool for clinical applications

  → Correlation of staining with absolute protein levels
- Objectives
  - Investigate the activation of downstream specific receptor family pathways
  - Early diagnosis and therapy improvement

Non small cell lung carcinoma (NSCLC) Project

- The EGFR/erb-B family of receptors
  - Important role for non small cell lung carcinoma (NSCLC) development
- Project objectives
  - Evaluation of the correlation between EGFR genetic alterations
  - Evaluation of the activation of downstream pathways
  - Definition of a subgroup of NSCLC able to respond to EGFR inhibiting therapy

IHC quantification framework: parameters

- Localization of marker (i.e membrane, cytoplasm, nucleus)
- Reaction intensity
- Percentage of (EGFR) positivity w.r.t. negative protein reaction in the carcinoma cells computed at the same cellular area

Cell segmentation

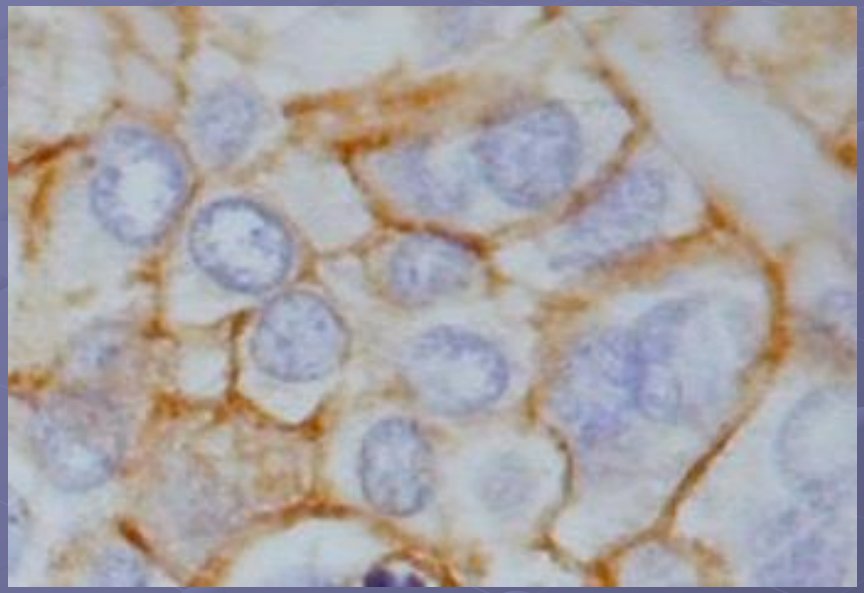
- Challenging issue
  - Morphology variability
    - ⇒ difficult to design geometrical models
  - Noise in the images
  - Cancer cell membranes with negative protein reaction are not visible

- Standard methods based on geometrical and/or gradient variation analysis are not effective

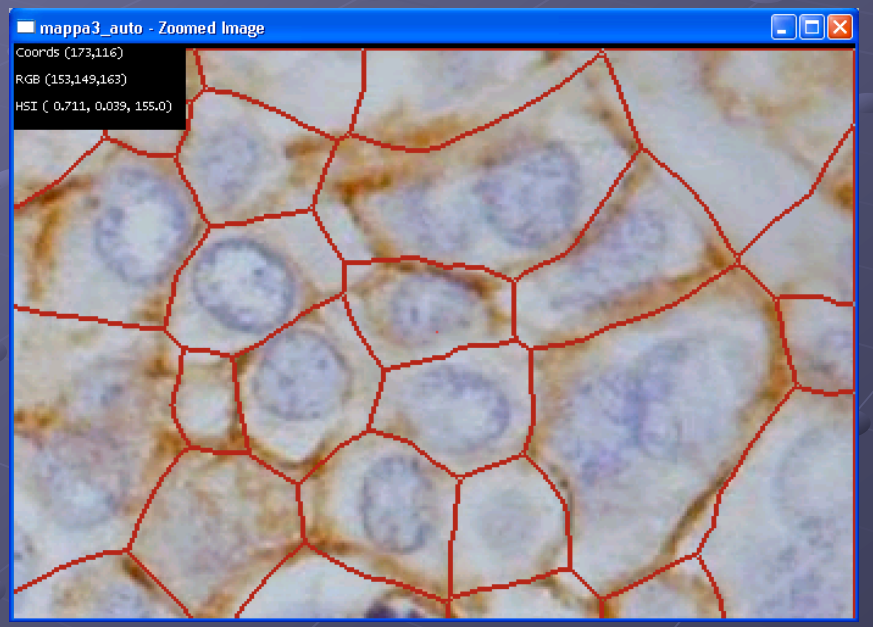
⇒ Customized automated procedure for cellular segmentation in IHC tissue images

Bioimaging tool(1)

Original image (a detail)



Virtual cell membrane segmentation



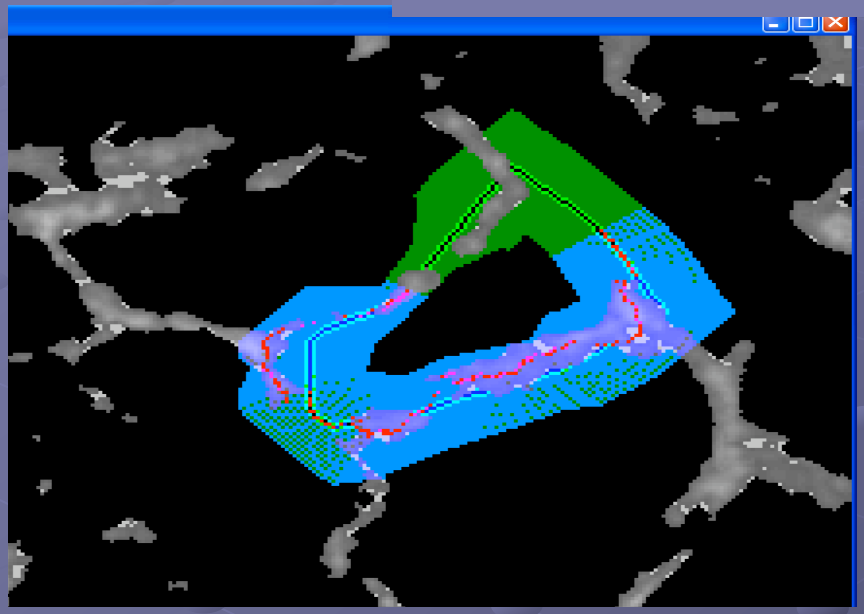
Several steps:

- Color filtering
- Nuclear membrane segmentation
- Virtual cell membrane definition

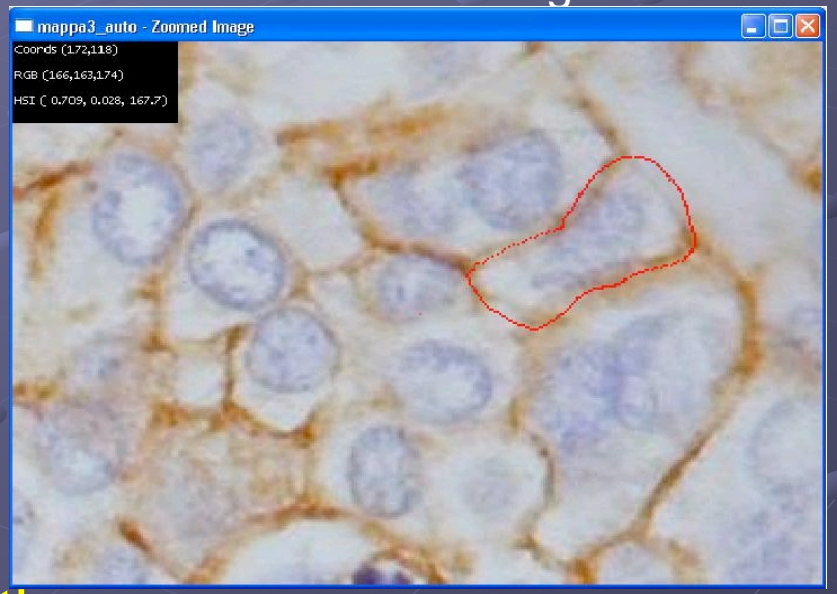
Original automated customized procedure

• Detection positive reaction (brown) areas when they exist
• Connection positive cellular membran es with cell virtual ones where positive reaction doesn't exist

Scanning Proced



Cellular membrane segmentation



Several steps:

4. Cellular membrane segmentation:
   - Scanning procedure
   - Fitting membrane points
5. Clinical parameters computation

Original automated customized procedures

Clinical Parameter Computation

- Protein activity computation
  - Percentage of brown pixels in the cellular membranes w.r.t. total number of membrane pixels
  - Validation: comparison with manual measurements

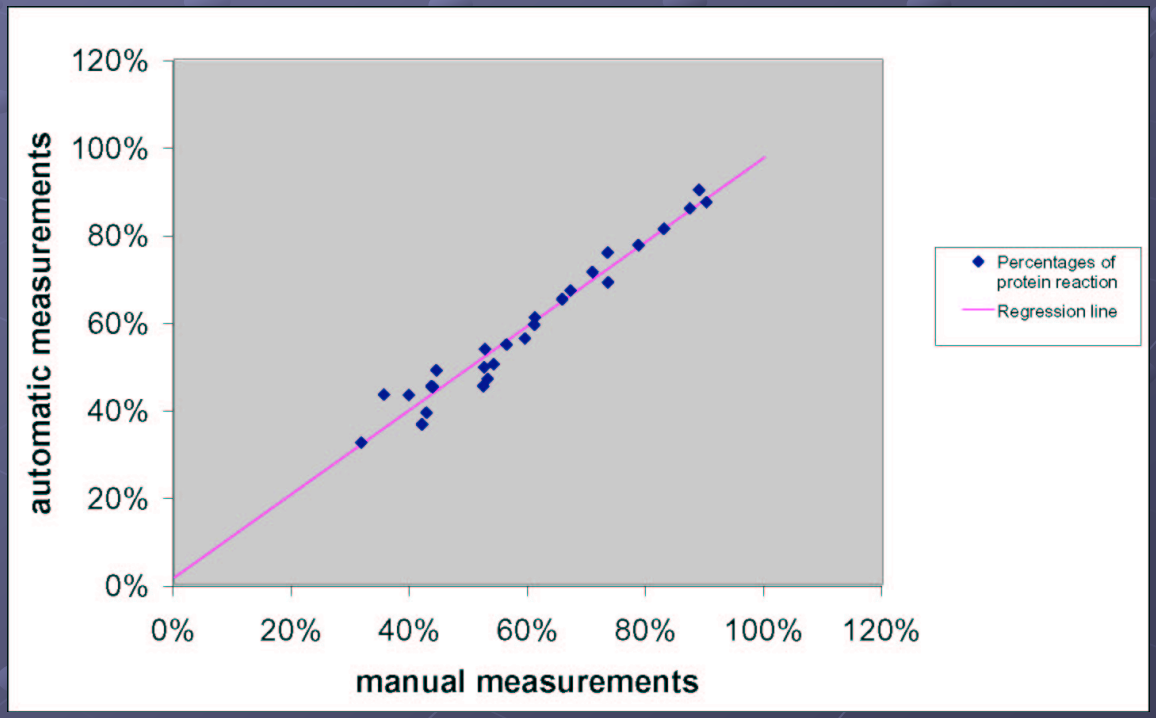
Results:

Coeff. Correl 0.98

Perc Reac = 59%

Aver Err = -0.77%

RMSE = 3.3%



Conclusion

- Bioimaging as standardized IHC image analysis
  - Extraction of quantitative and qualitative parameters for activation of downstream pathways analysis of receptor families
- Correlation of these parameters with data coming from gene expression analysis → ***joint co-clustering* tool**
- Preliminary results (85% correlation) show *joint co-clustering* is promising approach:
  - To analyze largescale biological data
  - To study multi-factorial genetic pathologies through their genetic alterations
  - To enable new opportunities for early diagnosis
  - To provide information in future strategies for therapy