Biomedical Imaging and Functional Genomics

Elisa Ficarra

Politecnico di Torino, Italy – EPFL Lausanne, Switzerland – DEIS University of Bologna, Italy

Background

- Digital Image Processing
- Computer Vision
- Gene Expression analysis

Techniques for Automated Analysis of DNA molecules in high-resolution microscope images

Development of Automated Algorithms for DNA Molecules Feature Analysis and Extraction

Image Processing techniques → to automatically extract that users can see on the images

Computer Vision → Image processing + Artificial Intelligence techniques to be able to make inferences

Work in Progress (1) : Gene Expression Analysis

- **Transcription Factors Identification**

  through data mining techniques on microarray data and ChIP-Chip data, and image processing and computer vision techniques

- **Co-expressed Genes identification**

  through data mining and clustering techniques and statistical analysis

- **Modeling Gene Regulatory Networks**
  - defining gene function
  - defining biochemical pathways

  through network mathematical models design and microarray screening of RNAi knockouts

- Goals:

- Drug Development
- Therapeutic treatment

Specific Problems

- Noise in the experimental data
- Non observability of mani variables of interest for gene network modeling (data incomplete)
- Biological variability

Mathematical Models

- Boolean Networks

  Probabilistic BN
- Bayesian Networks
  - Dynamic BN
- Differential Equations

Work in Progress (2) : Clinical Bioimaging and Functional Genomics

- **Biomedical and molecular imaging**
  - techniques to extract clinical and functional biological information from tissue or molecule or live cell images (*i.e. diagnosis of a specific subtype of cancer*)
- Correlation with clinical parameters and genetic pathways
  - to enhance gene expression analysis or to increase the amount of confidence in the hypothesized gene expression paths

Non small cell lung carcinoma (NSCLC) Project

- The EGFR/erb-B family of receptors seems to play an important role for non small cell lung carcinoma (NSCLC) development
- Aim of project → to evaluate the correlation between EGFR genetic alterations, the expression profile of EGFR, of its ligands, and the activation of downstream pathways in order to better define a subgroup of NSCLC able to respond to EGFR kinase inhibitors.

In collaboration with S.Luigi Hospital, Turin, Italy

HOW?

Through

- Gene Expression Analysis
- Immunohistochemical Automated Quantification

Immunohistochemistry (IHC): characteristics and aim

- To investigate the activation of downstream EGFR/erb-B receptor family pathways
- Marked antibodies to detect anti-genes EGFR, TGF alpha, erb-B
- Correlation of staining with absolute protein levels as tool for clinical applications (i.e. diagnosis and prognosis) and therapy improvement

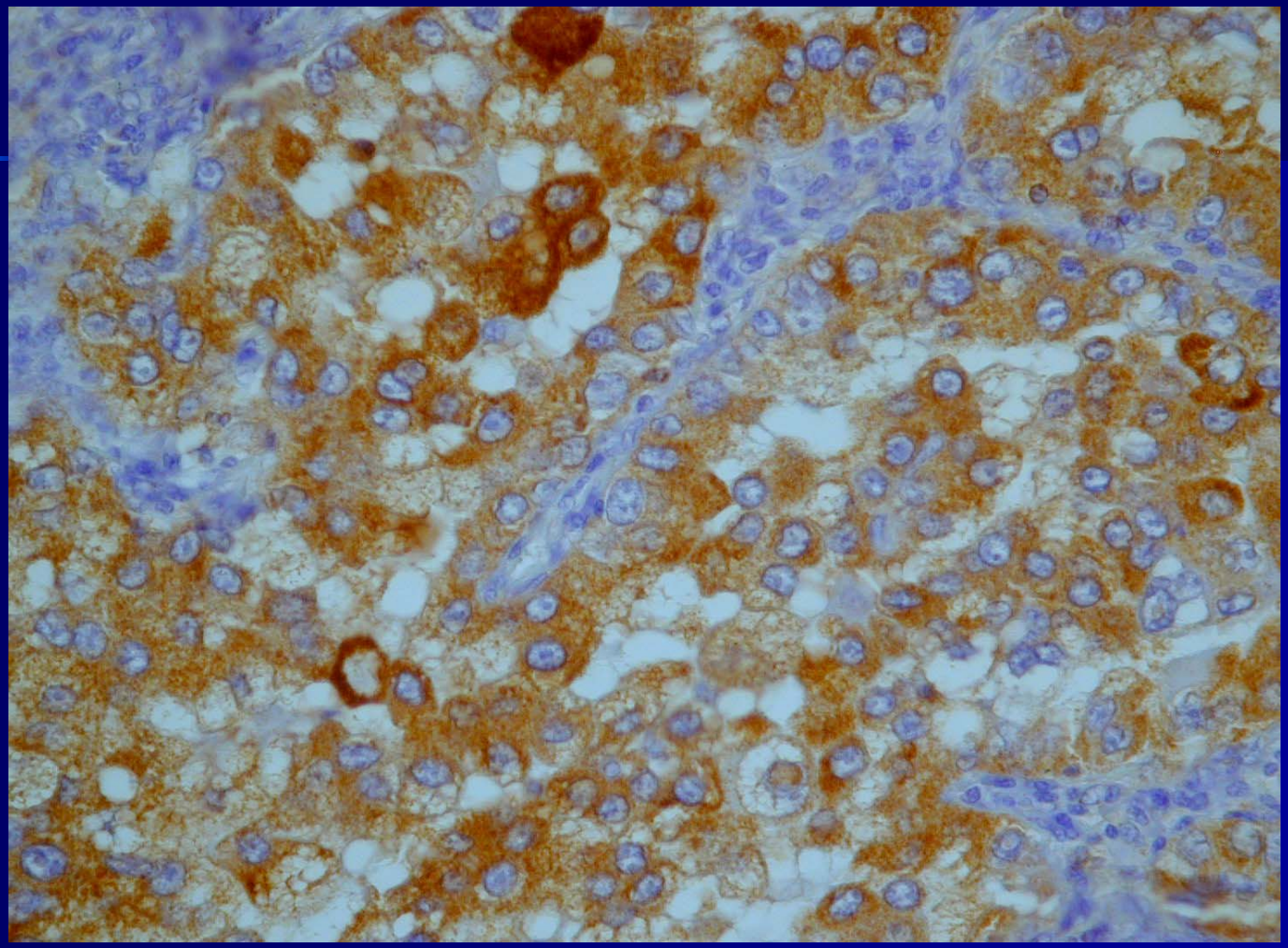
Immunohistochemical Automated Quantification

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

⇓

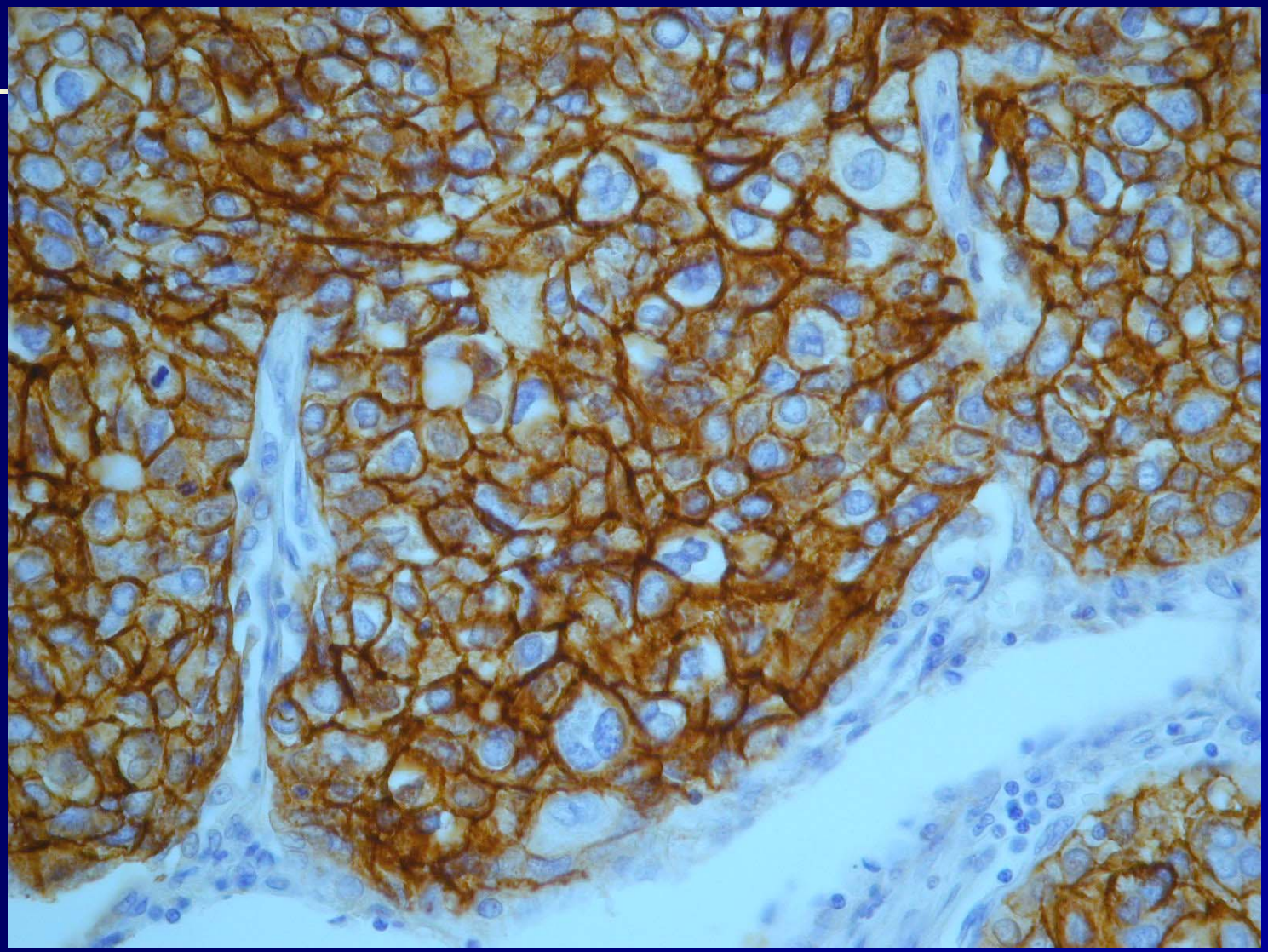
- Development of automated image processing methods to standardize IHC analysis

Example of IHC images



EGFR/erb-B receptors positivity (in carcinoma cells) as brown stain
Negative carcinoma cells and other cells in the sample as blue stain

Example of IHC images



IHC Quantification Framework

Parameters to extract:

- Localization of marker (i.e membrane, cytoplasm, nucleus)
- Reaction Intensity
- Percentage of (EGF) positivity w.r.t. all carcinoma cells in the sample

IHC Quantification Framework

Steps:

- Differentiation of carcinoma cells w.r.t. non carcinoma cells (individuation of different morphological properties)

  to quantify the percentage of positivity

- Identification and extraction of marked areas

  to quantify the reaction intensity and the percentage of positivity

- Differentiation of cell components (i.e. membrane, cytoplasm, nucleus)

  to identify the main location of the reaction

Step 2: Identification and extraction of marked areas

Color Segmentation in RGB space

- Identification of *seed* points of specified color
- Region growing from these points to points that satisfy conditions
  - Points have to belong to a specified color range (see RGB channels analysis or histogram processing on HSI color space and thresholds definition)
  - Points have to be 8-connected (for each region defined by seed points)

Step 1: Hard Problem

- Morphology Variability for same typology of cells
- Difficult to differentiate cells in some cases
- Noise in the images

Step 1 and 3: possible approaches

- Object Segmentation starting from specified points (e.g adaptive T-snakes or region-based segmentation by satisfying connectivity and some defined properties) and
- Regional Description (i.e. segmentation, representation and then description)
- Pattern Recognition: Decision-theoretical approaches (using quantitative descriptors, e.g. shape, orientation etc) or Structural approaches (using qualitative descriptors, e.g. relational)

↓

Both decision-theoretical and structural approaches are based on *learning* from sample patterns

Decision theoretical methods: Matching, Optimum Statistical Classifiers, Neural Networks
Structural methods: String Matching, Syntactic Recognition of Strings or Trees

Whole Framework

- Bioimaging as standardized IHC image analysis
- Extraction of quantitative and qualitative parameters for activation of downstream pathways analysis of EGFR/erb-B receptor family
- Validation of results
- Correlation of these parameters with other clinical parameters and gene expression analysis data