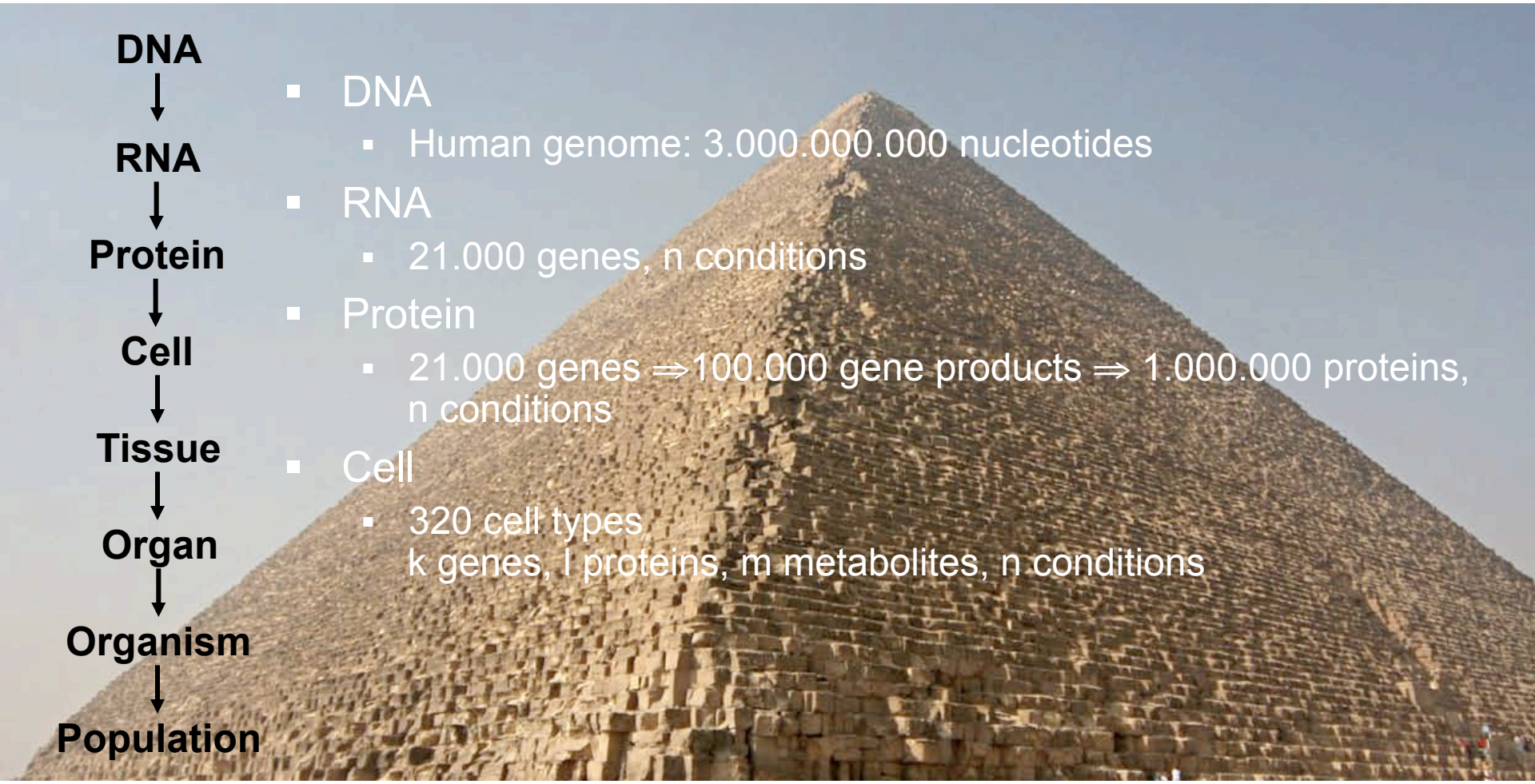
Integrating Biomolecular and Clinical Data for Cancer Research: Challenges and Concepts

Zlatko Trajanoski
Biocenter, Division for Bioinformatics
Innsbruck Medical University
Innrain 80, 6020 Innsbruck, Austria
Email: zlatko.trajanoski@i-med.ac.at
http://icbi.at

Cancer Research and Challenges

- Clinical research is driven by access to patient samples
- New technologies (sequencing, microarrays, proteomics) are driving discovery
- Clinical and research data are in different domains with no links between them
- Interpretation of the data requires integration of information across domains

Omics Technologies



DNA
↓
RNA
↓
Protein
↓
Cell
↓
Tissue
↓
Organ
↓
Organism
↓
Population

- DNA
  - Human genome: 3.000.000.000 nucleotides
- RNA
  - 21.000 genes, n conditions
- Protein
  - 21.000 genes ⇒100.000 gene products ⇒ 1.000.000 proteins, n conditions
- Cell
  - 320 cell types
    k genes, l proteins, m metabolites, n conditions

Omics Technologies

New high-throughput technologies

- DNA: deep sequencing
- RNA: high-density arrays
- Protein: MALDI-TOF, LC-MS/MS
- Tissue: tissue microarrays

Complementary technologies, real value in integrating diverse datasets

Data management and analyses?

Omics Technologies

Drowning in data, starving for information?

- Microarray data (n=1):

  Affymetrix HG U133A2 chip
  - Raw data: 80 MB per sample (incl. TIFF)
  - MAGE-ML (public repositories): 30 MB
  - Normalized data: 5-10 MB (Excel table or text file)

Omics Technologies

Drowning in data, starving for information?

- Proteomics data (n=1)

  Kisslinger et al, Cell 2006, 125:173-186

  one organ (heart), one organelle (cytosol)
  - Raw data: 1.55 GB (mzXML format)
  - Sequest search folders: 235 MB
  - Results in PRIDE format: 320 MB
  - Results incl. protein sequences: 374 KB

Medical Systems Biology

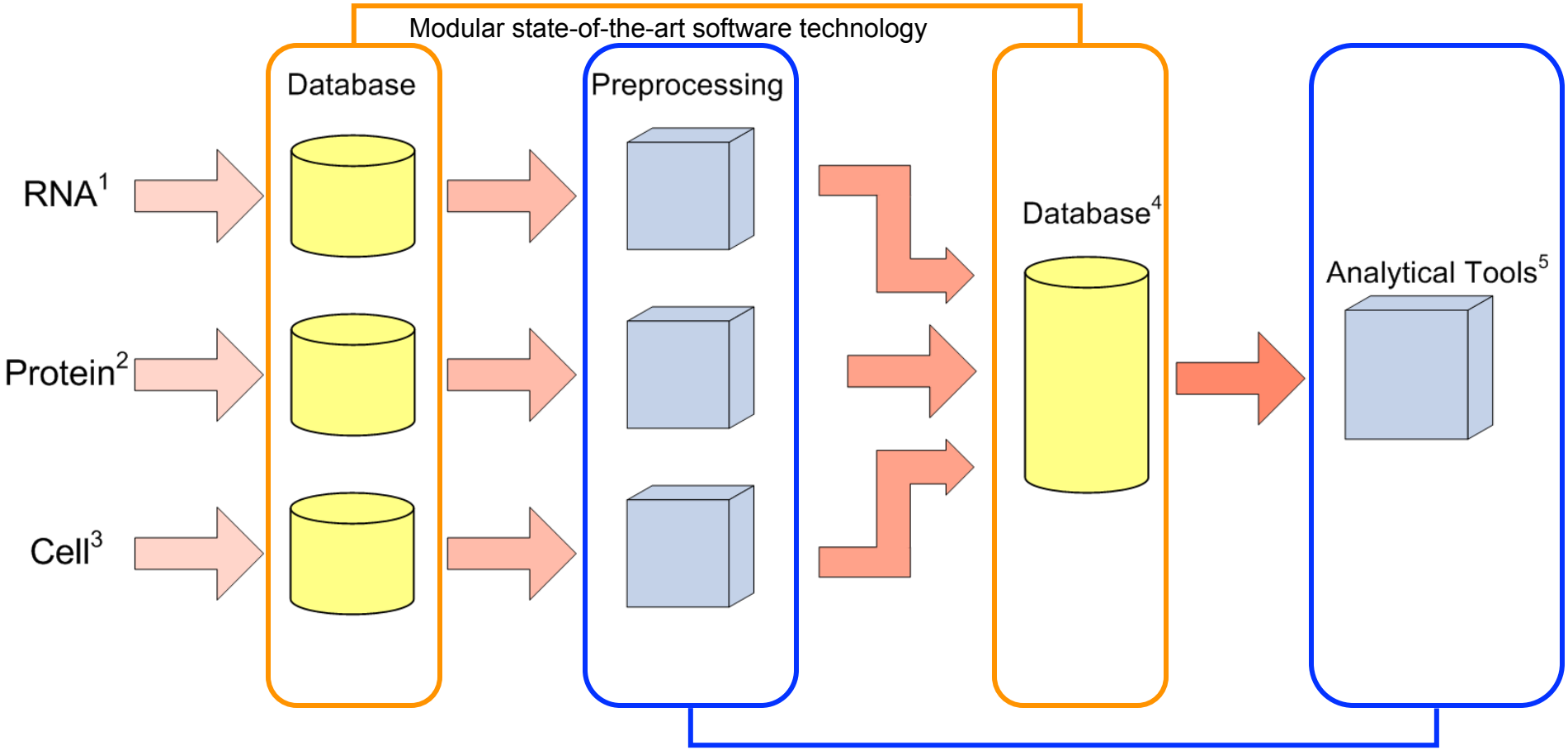
- n>100
- Phenotype data (clinical parameters)
- Genomics data (SNPs)
- Gene expression data (microarrays)
- Proteomics data (LC-MS/MS)
- Pharmacology data (pharmacokinetics/dynamics)
- Medical Images (CT, MR, PET, Ultrasound)
- Literature data (PubMed, Cochrane)
- Computational biology data (Ensembl, HPRD,...)

Data warehouse („Google“ for biomedical data)?

Data Integration

- Herculian task
- Few standards
- System incompatibilities
- Organizational issues
- Specific requirements in specific institutions

Data integration



[1]RNomics
Sturn et al., *Bioinformatics,* 2002
Pieler et al., *Bioinformatics*, 2002
Maurer et al., *BMC Bioinformatics*, 2005
Rainer et al., *Nucleic Acids Res*, 2006
Pabinger et al., *BMC Bioinformatics*, 2009

[2]Proteomics
Yu et al., *Bioinformatics*, 2005
Hartler et al., *BMC Bioinformatics*, 2007

[3]Cellomics
Habeler et al., *Nucleic Acids Res,* 2002
Thallinger et al., *BMC Bioinformatics*, 2007

[4]Databases
Hackl et al., *BMC Genomics*, 2004
Hackl et al., *Genome Biology,* 2005

[5]Analytical tools
Mlecnik et al., *Nucleic Acids Res*, 2005
Vogl et al., *Bioinformatics,* 2005
Bindea et al., *Bioinformatics*, 2009

Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house.

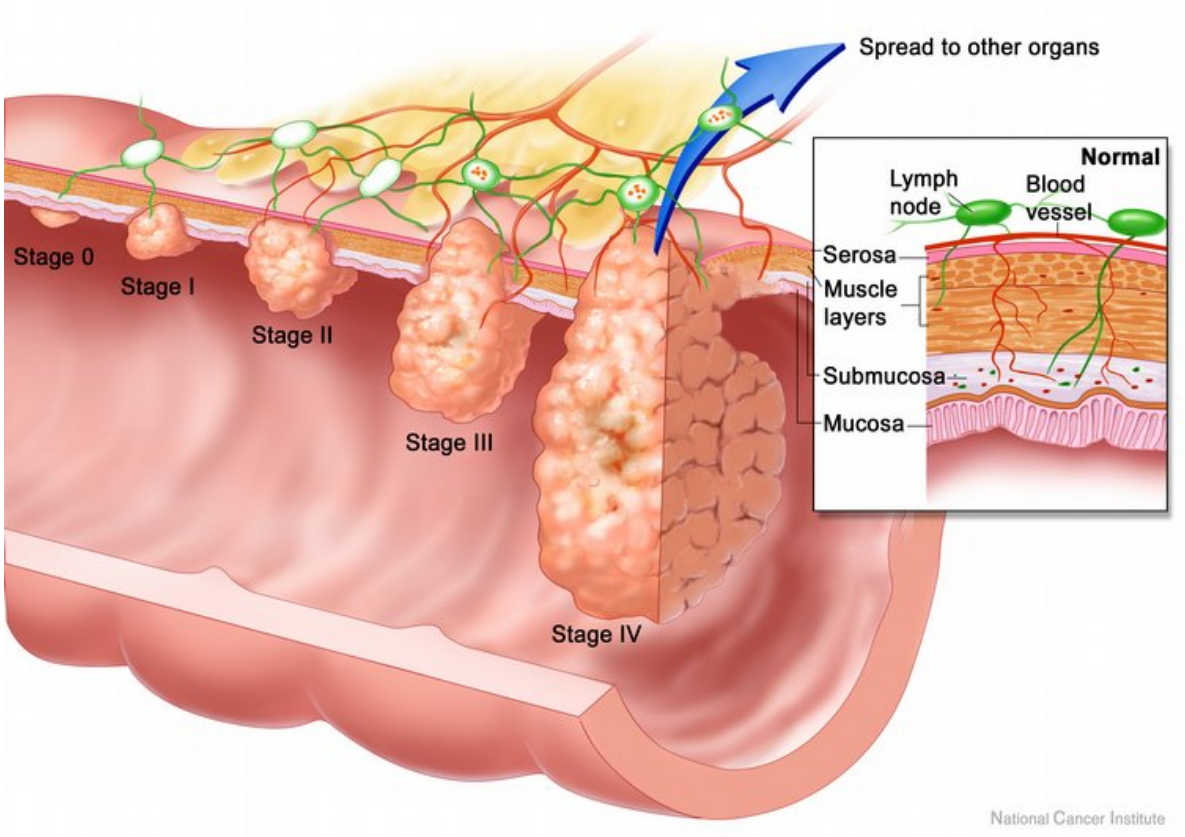
- Jules Henri Poincaré

From Data Collection to Discovery

- Case study: Colorectal cancer
  - Second leading cause of death among cancer patients
  - 1932: Dukes classification for postoperative outcome*
  - Today: Classification accuracy unchanged
  - Predictive molecular markers and rationale for adjuvant therapy?

*Dukes C, *J Pathol Bacteriol*, 35:232, 1932

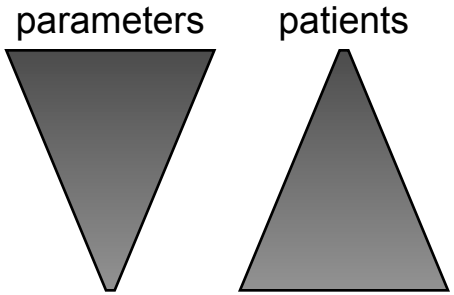
Cancer Immunology



Role of the immune system in colorectal cancer?

Data Generation and Integration

- Retrospective cohort (1986-2005)
- Clinical data and follow-up (n>1000)
- Patient material: paraffin-embedded tumors (>1000) and frozen tissue (>100)
- Assays: double-funnel approach
  - FACS analysis of 410 parameters (n=50)
  - qPCR of 50 mRNAs (n>100)
  - Tissue microarrays (n>500)



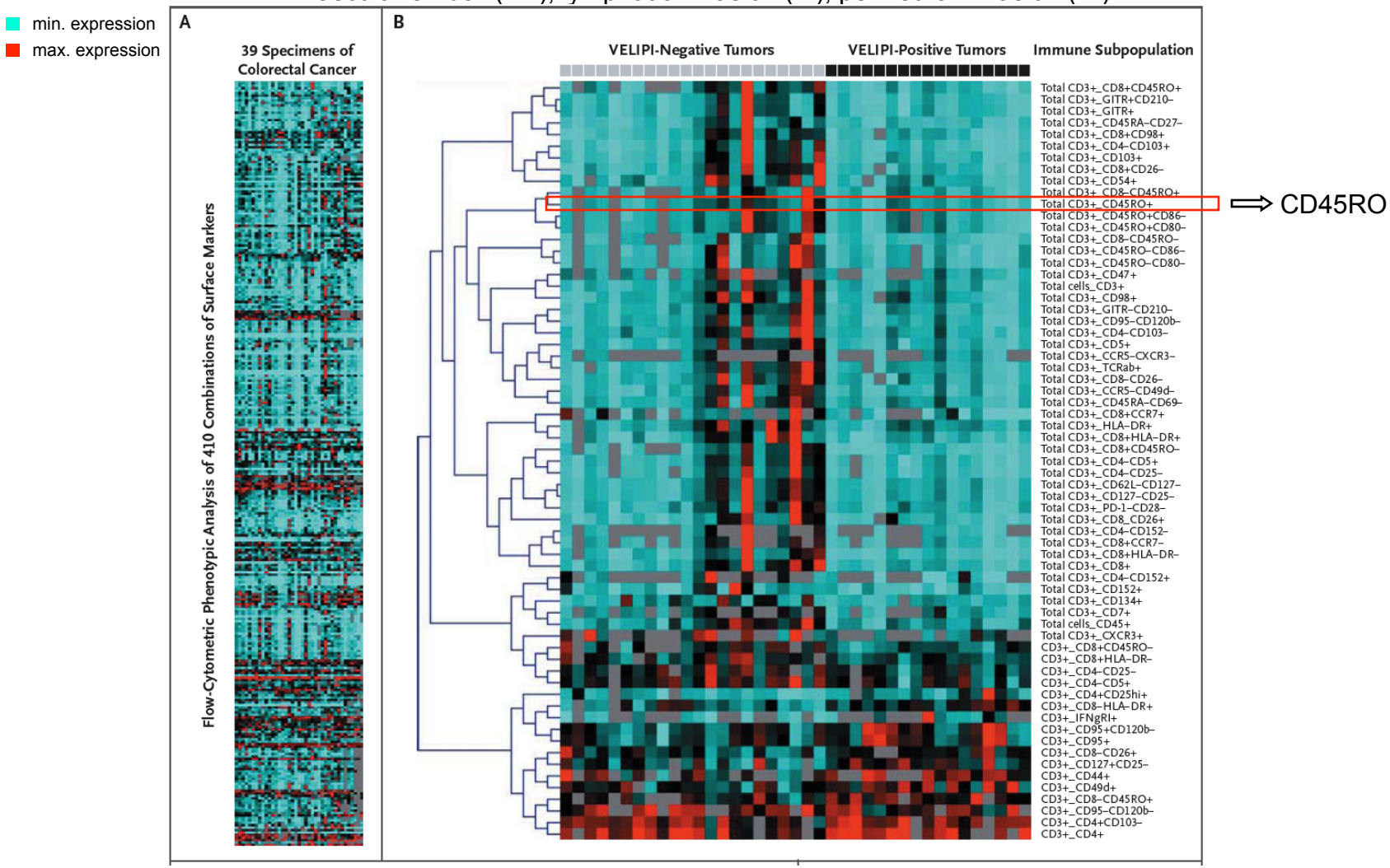
- Dedicated database for biomolecular and clinical data (http://tme.tugraz.at)*

*Mlecnik et al. *BMC Genomics*, 2010

Phenotypes of tumor-infiltrating immune cells

Significantly different markers between invasion positive (VELIPI+) and negative (VELIPI-) patients

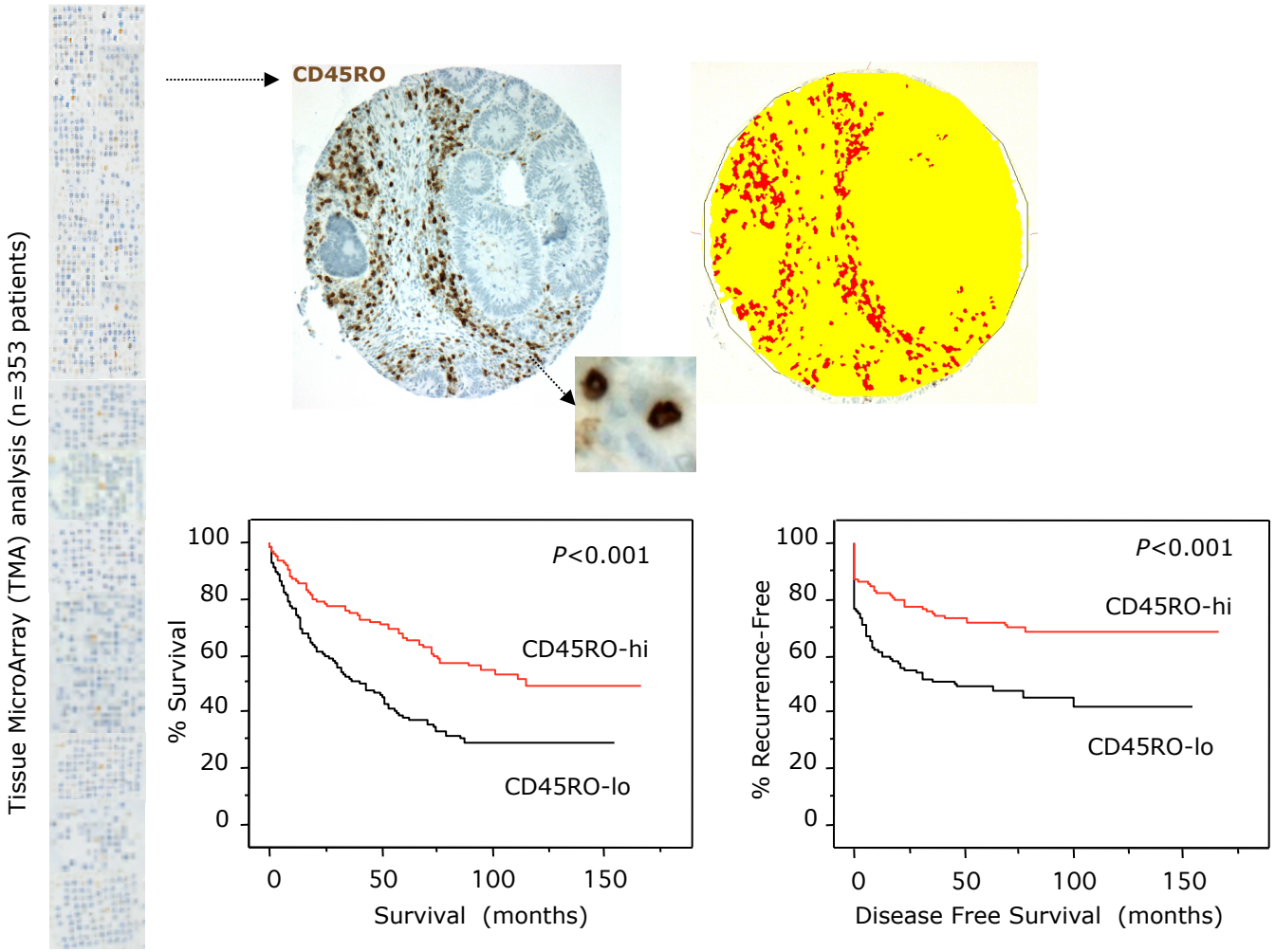
VELIPI: vascular emboli (VE), lymphatic invasion (LI), perineural invasion (PI)



Pagès et al. *N Engl J Med*, 353:2654-2666, 2005

Effector Memory T-cells and Survival

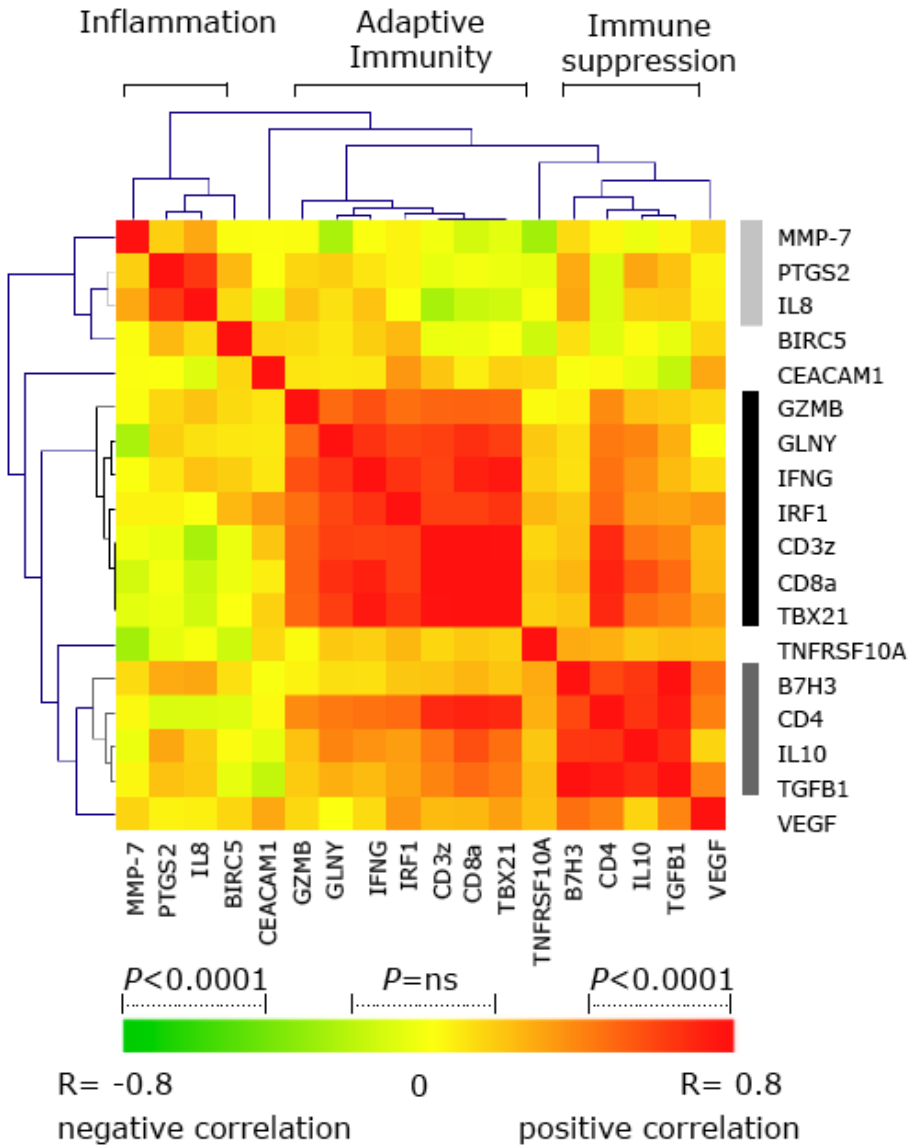
Disease-free and overall survival of CD45RO^{hi} patients

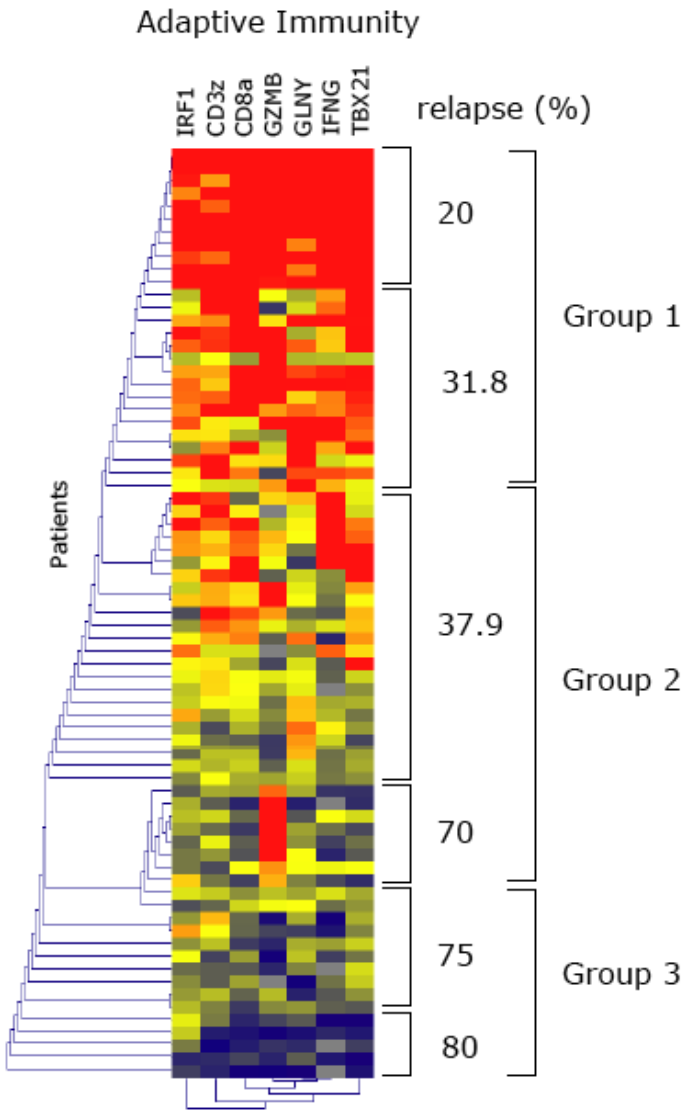


Pagès et al. *N Engl J Med*, 353:2654-2666, 2005

What is the Relationship between the Type, Density, and Location of Immune Cells within Tumors and the Clinical Outcome?

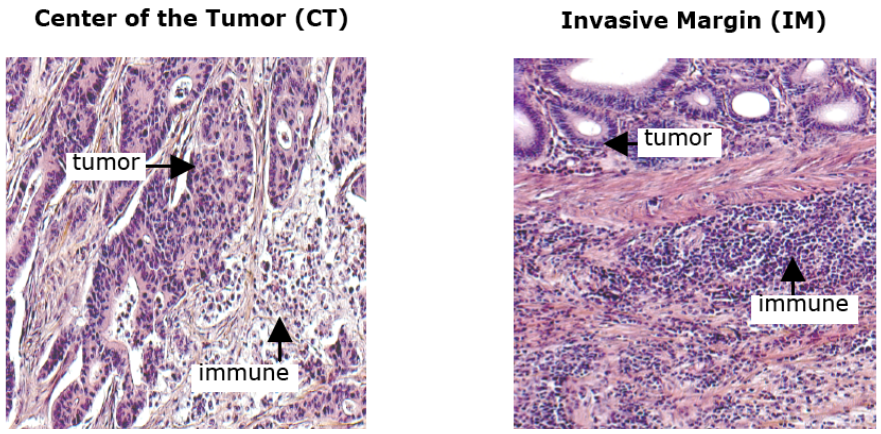
Adaptive Immunity has a Beneficial Effect on Clinical Outcome

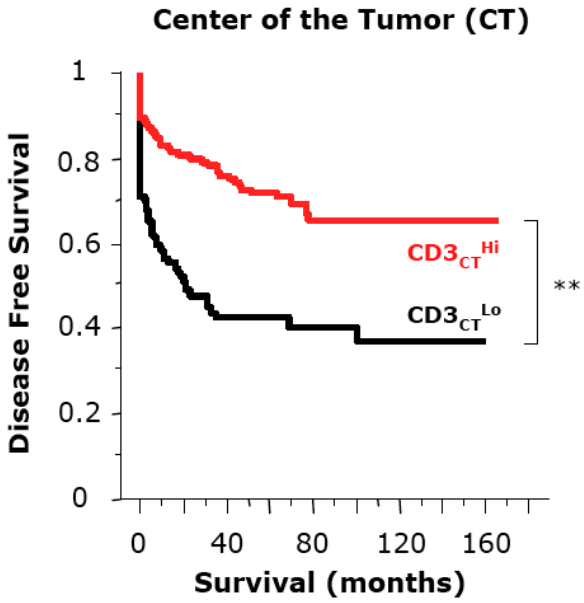


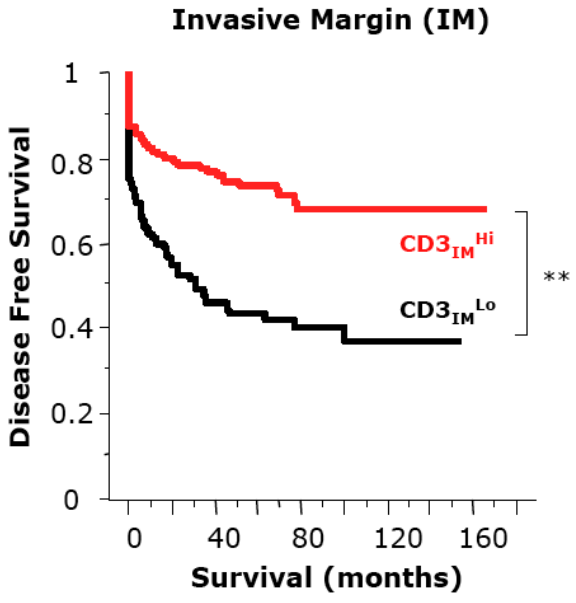


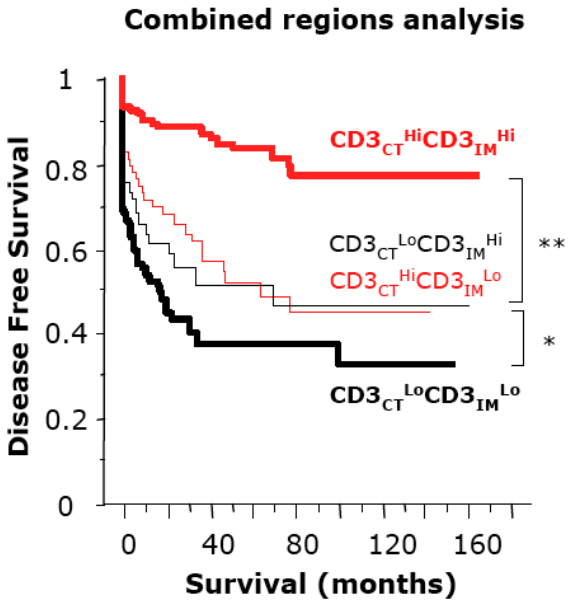
Galon et al. *Science*, 313:1960-1964, 2006

Combined Analysis of Tumor Regions Improves Prediction of Patient Survival







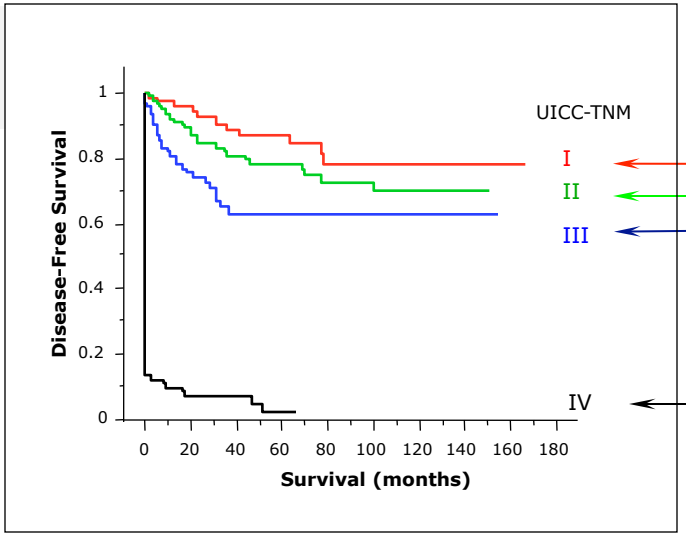


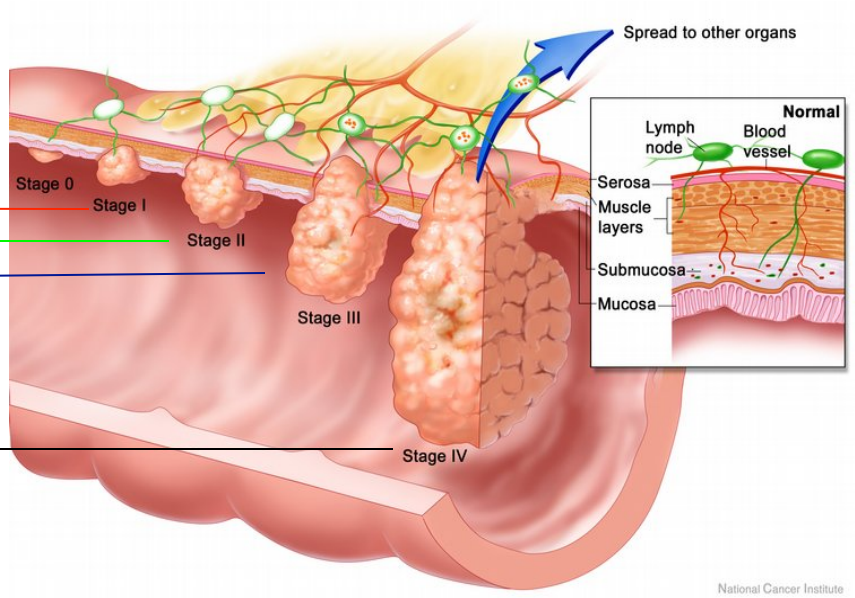
Galon et al. *Science*, 313:1960-1964, 2006

Patient Stratification

Tumor histopathology

UICC-TNM
Staging system

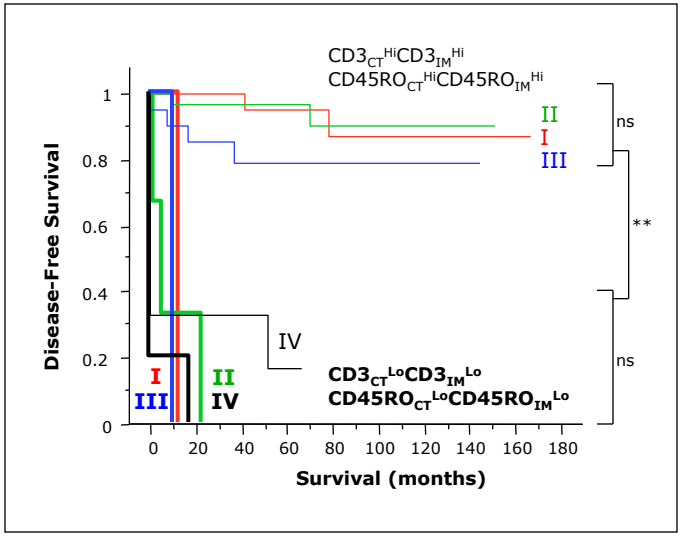




$CD3_{CT}CD3_{IM}$
evaluation

plus

$CD45RO_{CT}CD45RO_{IM}$
evaluation



Galon et al. *Science*, 313:1960-1964, 2006

Immunological criterion for predicting tumor recurrence

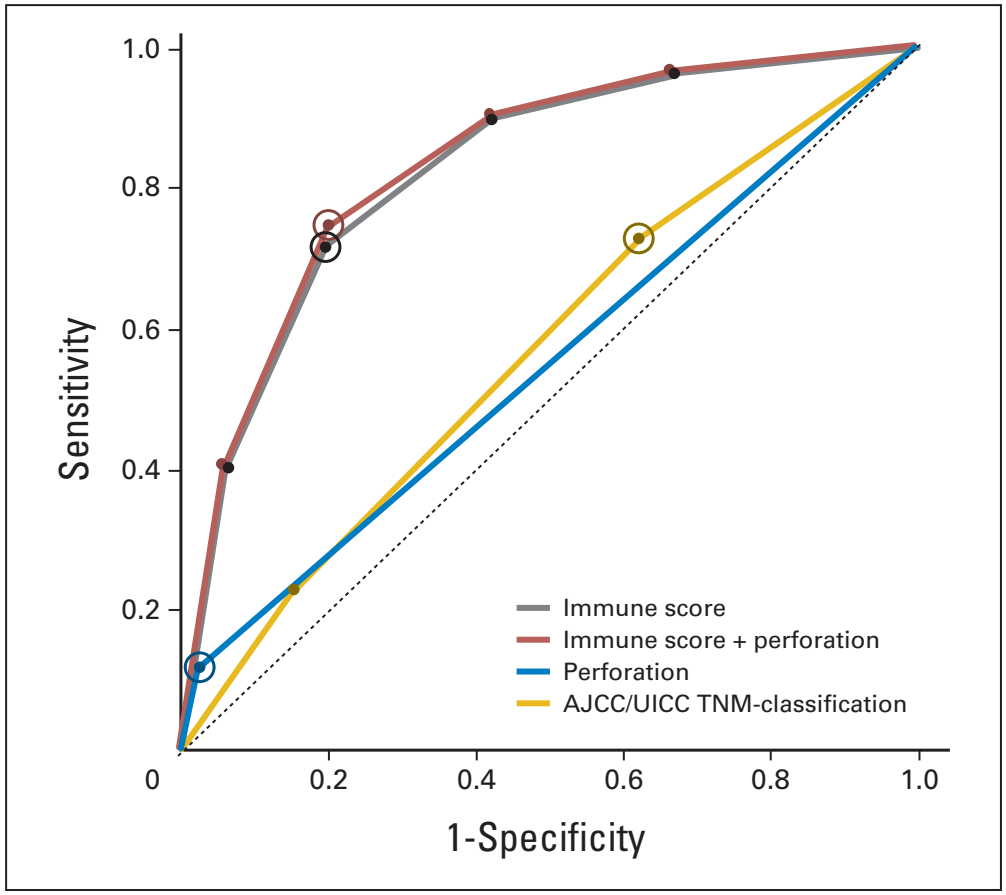
Once human colorectal cancers become clinically detectable, the adaptive immune system plays a role in preventing tumor recurrence

Type, density, and location of immune cells within colorectal tumors influence the clinical outcome of the patients

An immune score based on the combined evaluation of memory and cytotoxicity markers identifies patients with early-stage (stage I and II) tumor at high-risk of tumor recurrence and death*

*Pagès et al. *J Clin Oncol*, 2009

Immunological criterion for predicting tumor recurrence



Immune score:
(2 markers, 2 regions)

Im0: 0 - hi, 4 - lo (lolololo)
Im1: 1 - hi, 3 - lo
Im2: 2 - hi, 2 - lo
Im3: 3 - hi, 1 - lo
Im4: 4 - hi, 0 - lo (hihihihi)

Mlecnik et al. *J Clin Oncol*, 2011

Immunological criterion for predicting tumor recurrence

VOLUME 29 · NUMBER 6 · FEBRUARY 20 2011

JOURNAL OF CLINICAL ONCOLOGY

E D I T O R I A L S

TNM Staging in Colorectal Cancer: T Is for T Cell and M Is for Memory

Elizabeth K. Broussard and Mary L. Disis, *Tumor Vaccine Group, Center for Translational Medicine in Women's Health, University of Washington, Seattle, WA*

metastatic phenotype.[12] A focused investigation of patients with early-stage colorectal cancer suggested that a multimarker panel of CD45RO-positive and CD8-positive T cells and cytotoxicity-related genes could predict prognosis even in these patients with minimal disease.[13] Taken together, these studies laid the foundation for the immune score presented by Mlecnik et al[9] as a clinical prognostic marker at any stage of colorectal cancer.

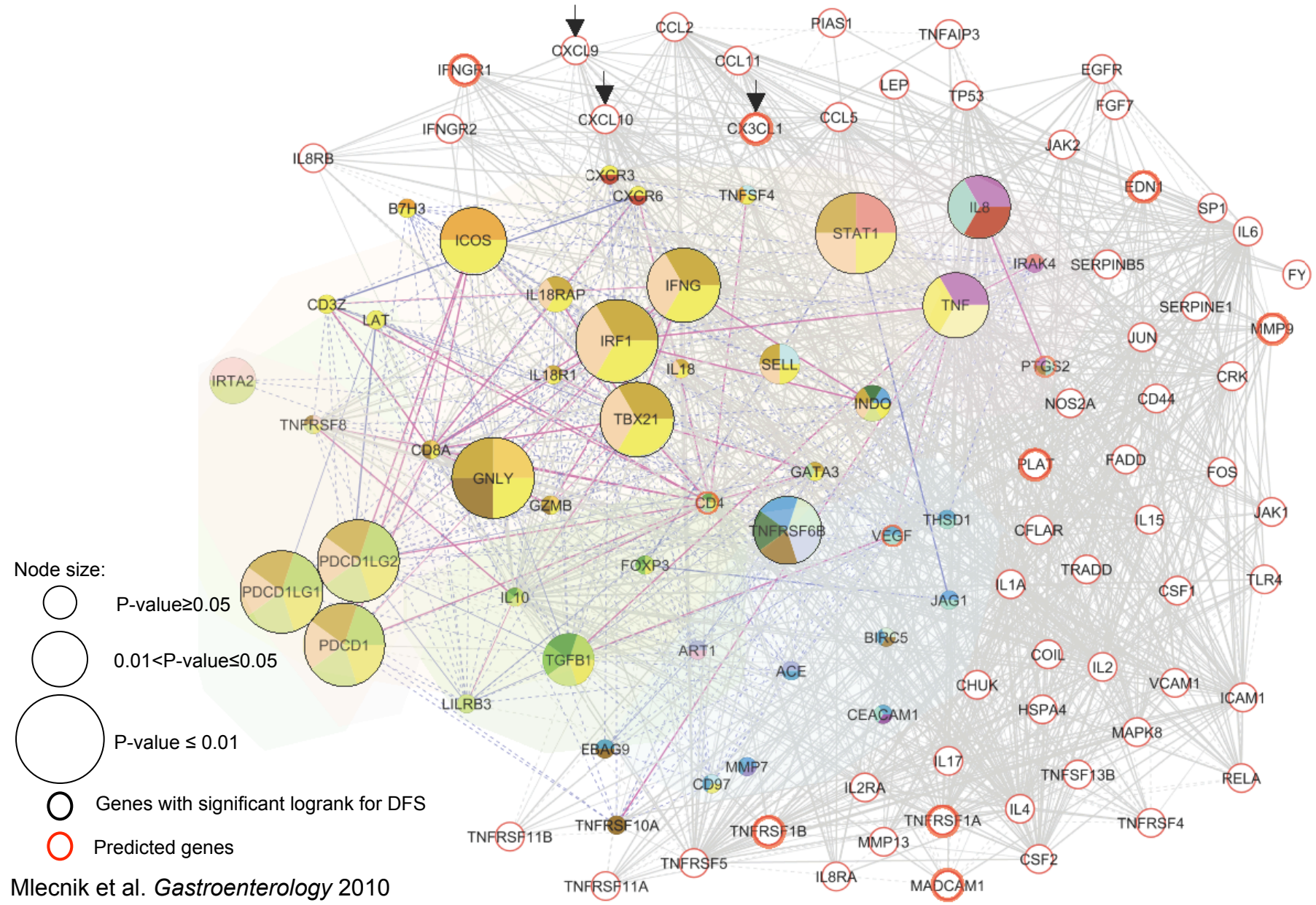
Results presented in the current report by Mlecnik et al[9] have both biologic and clinical importance. The presence of high-density memory T cells that are cytotoxic and display markers that suggest the cell is activated and capable of killing tumor cells (granzyme B) suggests that the adaptive immune system does, indeed, play a major role in tumor eradication and disease outcome. The fact that the density of these infiltrating T cells can be quantitated and that they are an independent predictor of prognosis suggests that the evaluation of CD8-positive and CD45RO-positive T-cell density should become part of the standard practice of evaluating colorectal cancer, or even other tumors, at the time of diagnosis. In addition, with the advent of several successful immune-based cancer therapies that result in a statistically significant survival benefit in randomized clinical trials,[30-32] perhaps an immune score would identify a population of patients who would derive substantial benefit from further stimulating their adaptive immune response.

Biomolecular networks

- Sequential analyses of datasets: powerful but limited
- Mechanistic insights?
- Clues for developing (immuno)therapy?

$\Rightarrow$ Integrative data analyses using biomolecular networks

Reconstruction of immune network in colorectal cancer



Node size:

P-value≥0.05

0.01<P-value≤0.05

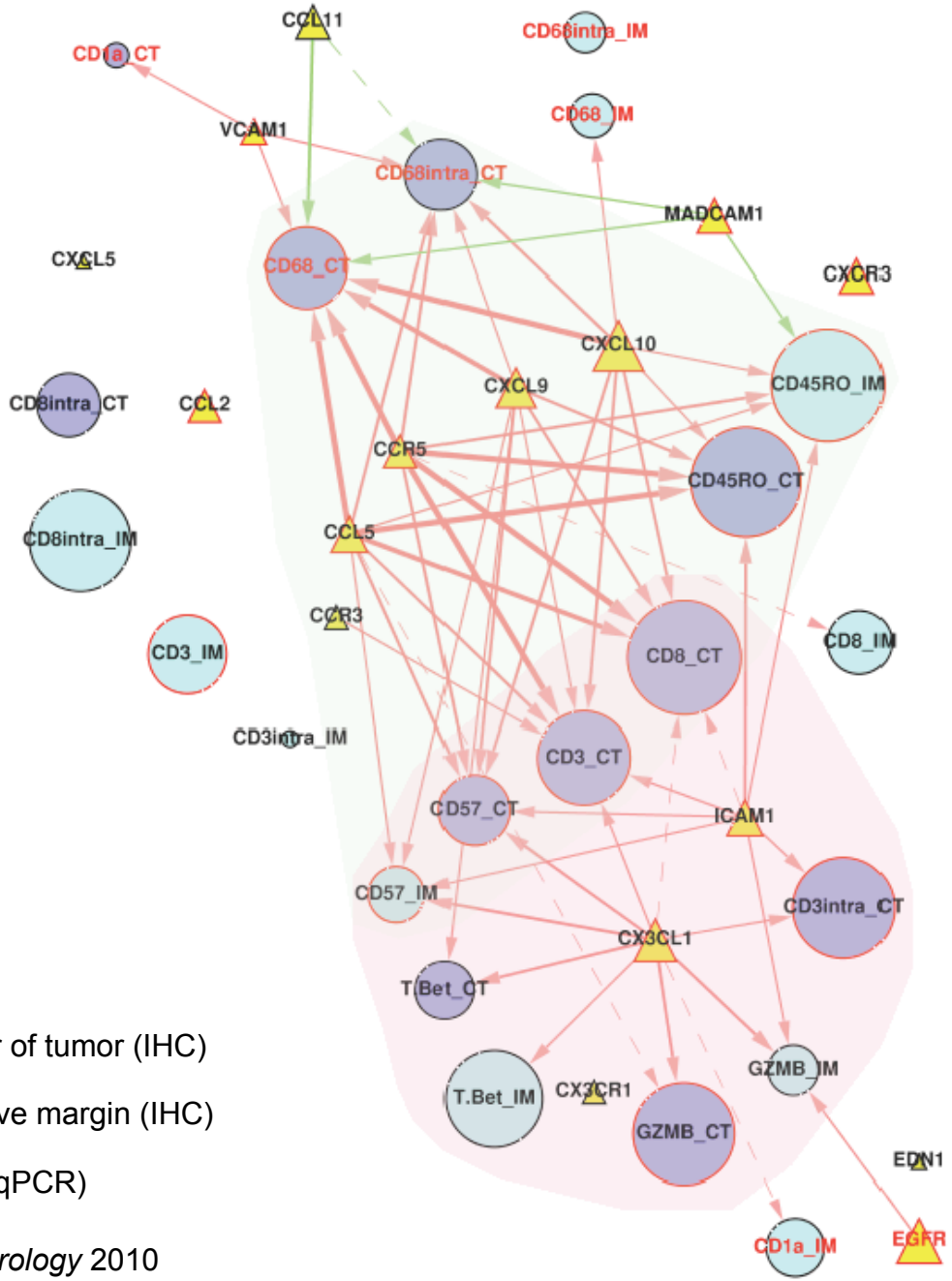
P-value ≤ 0.01

Genes with significant logrank for DFS

Predicted genes

Mlecnik et al. *Gastroenterology* 2010

Chemokines attract specific phenotypes of T-cells



cell density, center of tumor (IHC)

cell density, invasive margin (IHC)

gene expression (qPCR)

Mlecnik et al. *Gastroenterology* 2010

Data integration uncovers molecular mechanisms

- Chemokines CX3CL1, CXCL9, CXCL10 attract specific subsets of T cells within the tumor
- Chemoattraction and the presence of an addaptive immune reaction within the tumor are critical parameters influencing the outcome of colorectal cancer
- Can we predict therapeutic usefulness of targeting specific molecules and pathways (immunotherapy)?

Outlook: Computational Challenges

- Data integration
  - Molecular data (deep-sequencing, expression, proteomics)
  - Cytogenetic data
  - Imaging data
- Modeling biomolecular networks
  - Pathways and networks
  - Multi-scale modeling:
    - Spatial: nm to m (from molecules to cells to organs)
    - Temporal: min to yrs

Computers and Medicine

Acknowledgements

- Innsbruck Medical University/Graz University of Technology

  Pornpimol Charoentong, Hubert Hackl

  Bernhard Mlecnik, Gabriela Bindea, Robert Molidor, Fatima Sanchez-Cabo

- INSERM, Paris, France

  Jerome Galon, Franck Pages

- Funding

  bm:wf (GEN-AU Programme)

  FWF (SFB 021)

Integrating Biomolecular and Clinical Data for Cancer Research: Challenges and Concepts

Zlatko Trajanoski
Biocenter, Division for Bioinformatics
Innsbruck Medical University
Innrain 80, 6020 Innsbruck, Austria
Email: zlatko.trajanoski@i-med.ac.at
http://icbi.at