Proteomics

Background and clinical utility

H.H. Helgason MD
Antoni van Leeuwenhoek Hospital
The Netherlands Cancer Institute
Amsterdam

NKI-AVL
Het Nederlands Kanker Instituut
Antoni van Leeuwenhoek Ziekenhuis

Introduction

- Background
  - Definitions
  - Protein biomarkers
- Technical aspects
  - Serum proteomic profiling
- Our results
- Clinical utilities
  - Importance of validation
- Future aspects
  - Tissue proteomic profiling and imaging

Background

February 1953
DNA structure elucidated by Watson and Crick

April 2003
Completion of the full human genome sequence

Background

“Any global analysis of changes in the quantities and posttranslational modifications of all the proteins in an organism”

normal cell

diseased cell

Definitions (1)

- **Proteome:**
  - It defines the entire protein content in a given cell, tissue or organism.
  - Proteome depict the protein complement of a genome and represents the end product of the genome.
  - Although the cellular genome is relatively constant, the proteome changes constantly.

Proteomics

The genes are the same,

but the proteins are not!

Definitions (2)

- **Proteomics:**
  - Proteome analysis or proteomics is a system-wide study of proteins and can be defined as the systematic determination of protein sequence, quantity, modification state, interaction partners, activity and structure in a given cell type at a particular time.
  - Any global analysis of changes in the quantities and post-translational modifications of all proteins in an organism

Definitions (3)

- **Proteomic pattern:**
  - The discriminating pattern formed by a small key subset of proteins or peptides buried among the entire repertoire of thousands of proteins represented in the sample spectrum.
  - The pattern is defined by the peak amplitude values only at key mass/charge (M/Z) positions along the spectrum horizontal axis.

Protein biomarkers (1)

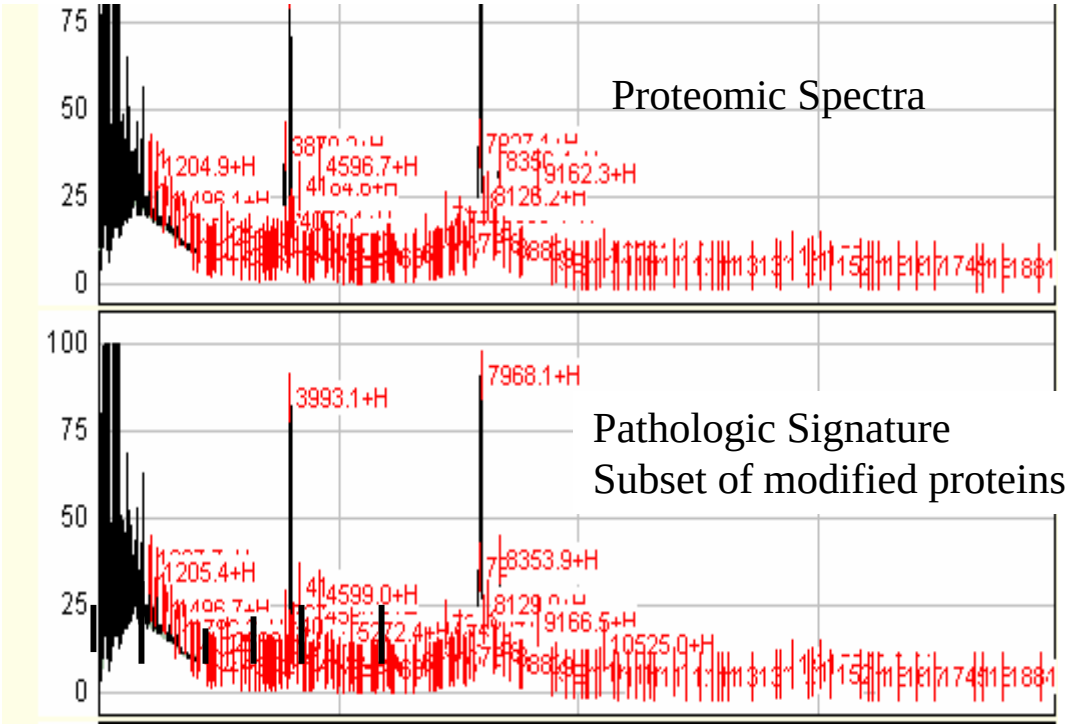
- Change in the expression level of a single protein

Biomarker	Mw	Disease
Prostate Specific Antigen (PSA)	28 kDa	prostate ca
α-Fetoprotein (AFP)	70 kDa	germ cell ca
Carcino Embryonic Antigen (CEA)	200 kDa	colon ca
Carcinoma Antigen-125 (CA 125)	> 200 kDa	ovarian ca

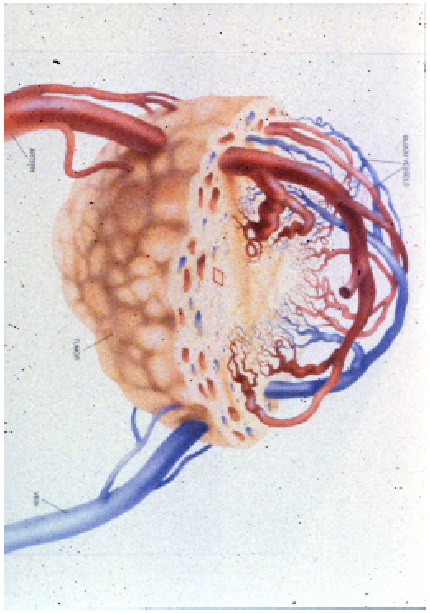
  - Specificity and sensitivity are suboptimal

- Complex pattern of several proteins / peptides with different expression levels

PATTERNS OF PROTEOMIC INFORMATION IN SERUM

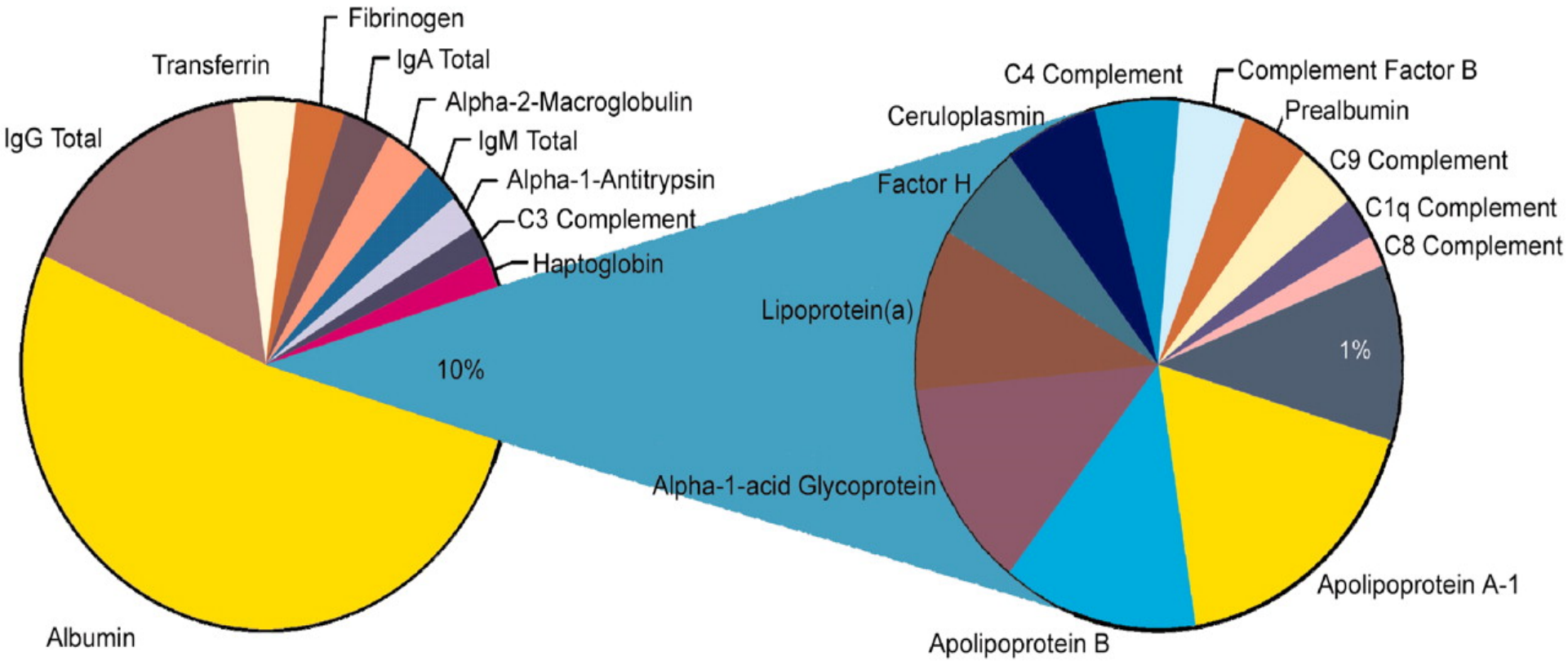


Perfused Tissue



- **Serum proteome:**
  a population of thousands of complexed proteins and peptides
- **Tissues are continuously perfused by the serum proteome:**
  their physiologic state may be reflected in serum proteomic patterns

Protein biomarkers (2)



Pie chart representing the relative contribution of proteins within plasma

Tirumalai, R. S. (2003) Mol. Cell. Proteomics 2: 1096-1103

Protein biomarkers (3)

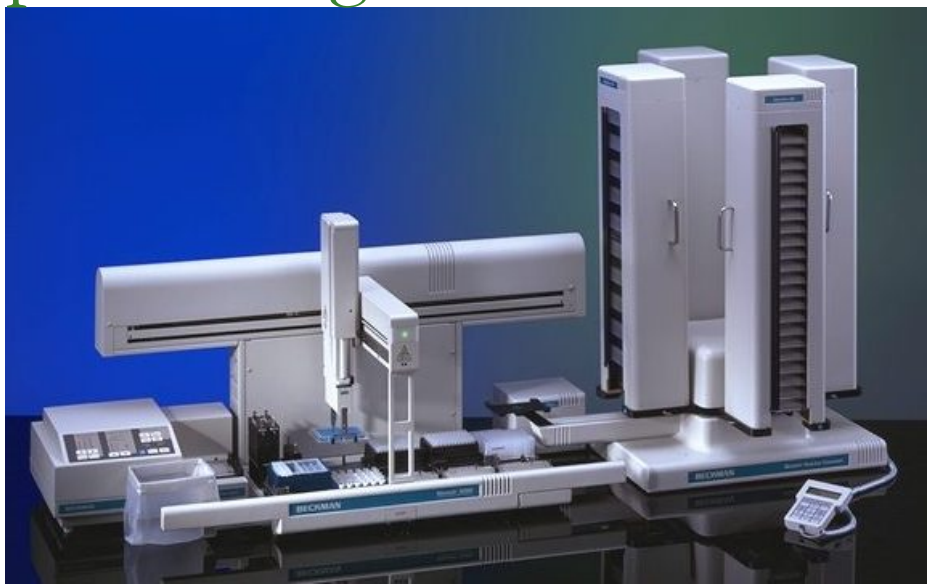
- "Other" proteins:
  - ≤1% of plasma protein content
    - 400.000 - 500.000 proteins
- Sophisticated analytical techniques are required, e.g.:
  - 2D gel-electrophoresis
  - LC-MS/MS (combined with tryptic digestion)
  - 2D HPLC
  - *SELDI-TOF MS*

Serum proteomic profiling

- SELDI – TOF – MS
  - Surface Enhanced Laser Desorption / Ionisation – Time of Flight Mass Spectrometry
  - Sensitive bio-analysis of low molecular weight proteins
- System requirements:
  - Protein-Chip Array
  - Protein-Chip Reader
  - Bioinformatics Software

Serum sample loading

One Microliter of Serum



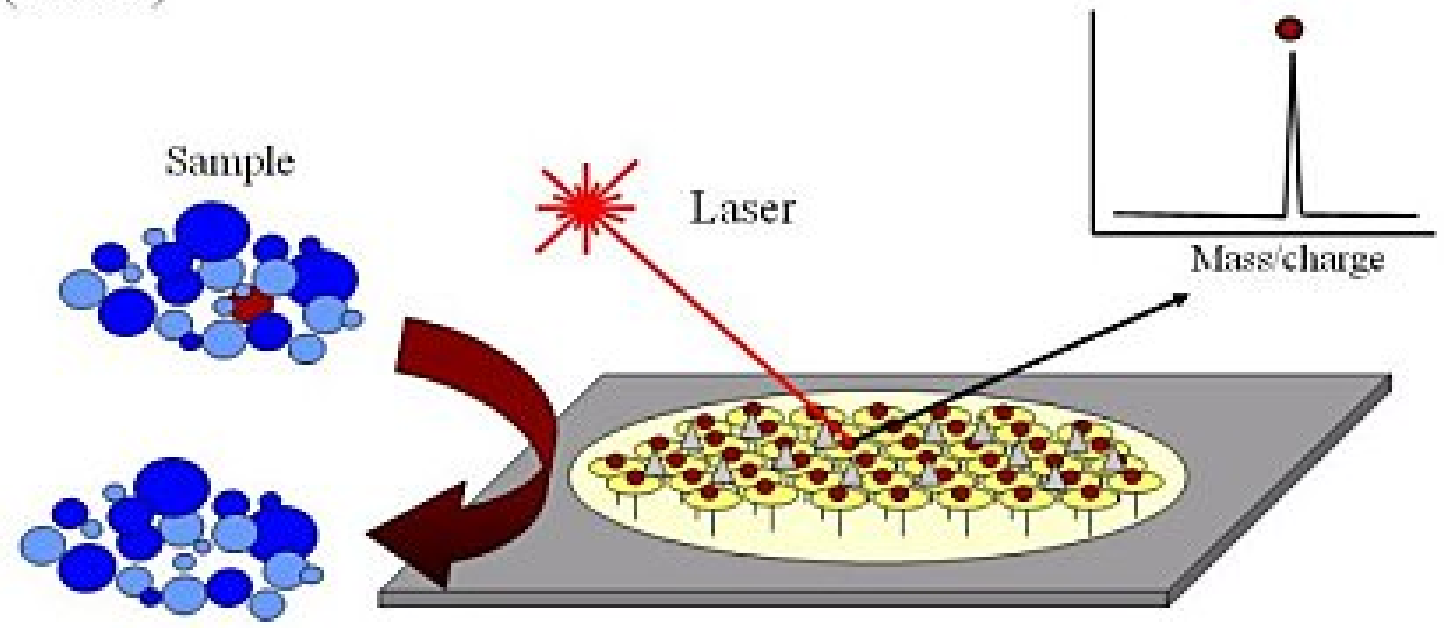
Robotic handling

- Improved Reproducibility
- Better Sample Handling
- Increased Throughput
- Reduce Cross Contamination

Vincent Fusaro and Sally Ross

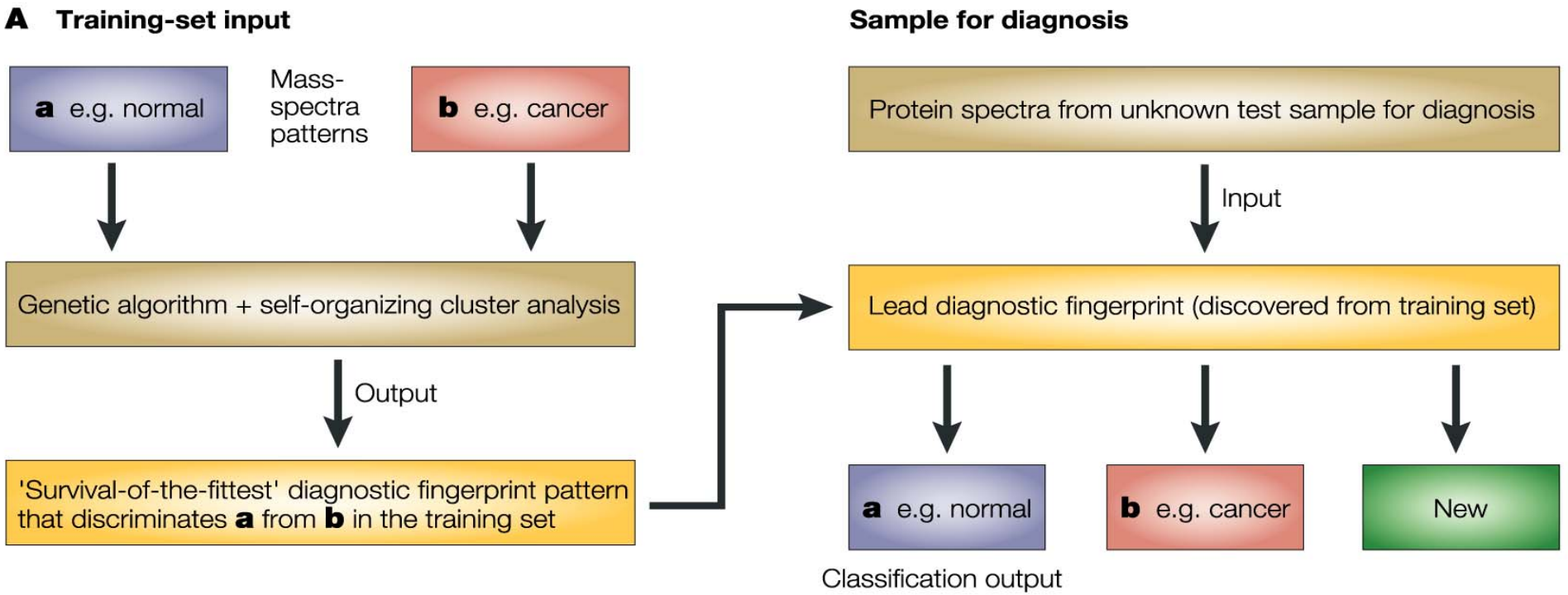
SELDI ProteinChip™ Technology Process

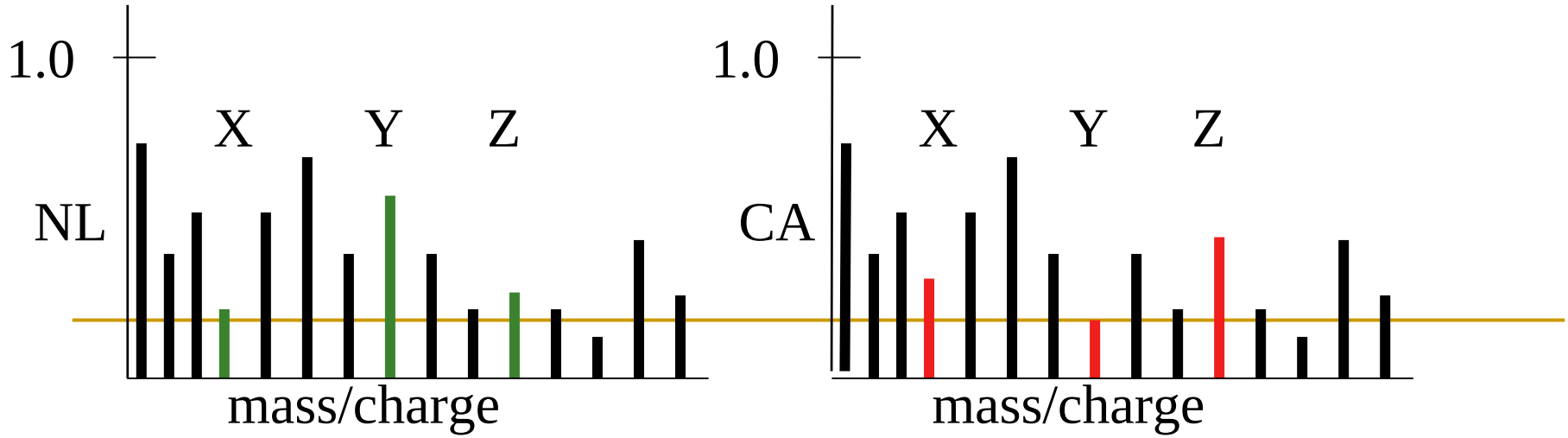
- Sample goes *directly* onto the ProteinChip™ Array
- Proteins● are captured and *retained* on the chip (affinity capture)
- EAM△ is added to the chip
- Retentate map is "read" by Surface-Enhanced Laser Desorption/Ionization (SELDI)



SELDI - TOF: Surface-Enhanced Laser Desorption/Ionization – Time of flight

Bioinformatics Discovery Tool





Clinical proteomics

- Miscellaneous:
  - Rheumatoid arthritis — J. Proteome Res ’02;1:495
  - HIV-infection — Science ’02;298:995
  - Infectious diseases — Proteomics ’03;3:273
  - Alzheimer — Proteomics ’03;3:1486
- Oncology:
  - Ovarian ca — Lancet ’02;359:572
  - Prostate ca — Cancer Res ’02;62:3609
  - Lung ca — Lung Cancer ’03;40:267
  - Pancreatic ca — Cancer Res ’05;65:10613

Proteomics in breast cancer

- Difficulties in diagnosis of breast cancer:
  - Limitations of current methods for (early) detection
  - Lack of adequate follow-up parameters

- Study objective:
  - To find new serum protein profiles that distinguish breast cancer patients from healthy controls

Study design

- **Total population: 140 breast cancer patients (BC)**
  110 healthy matched controls (HC)

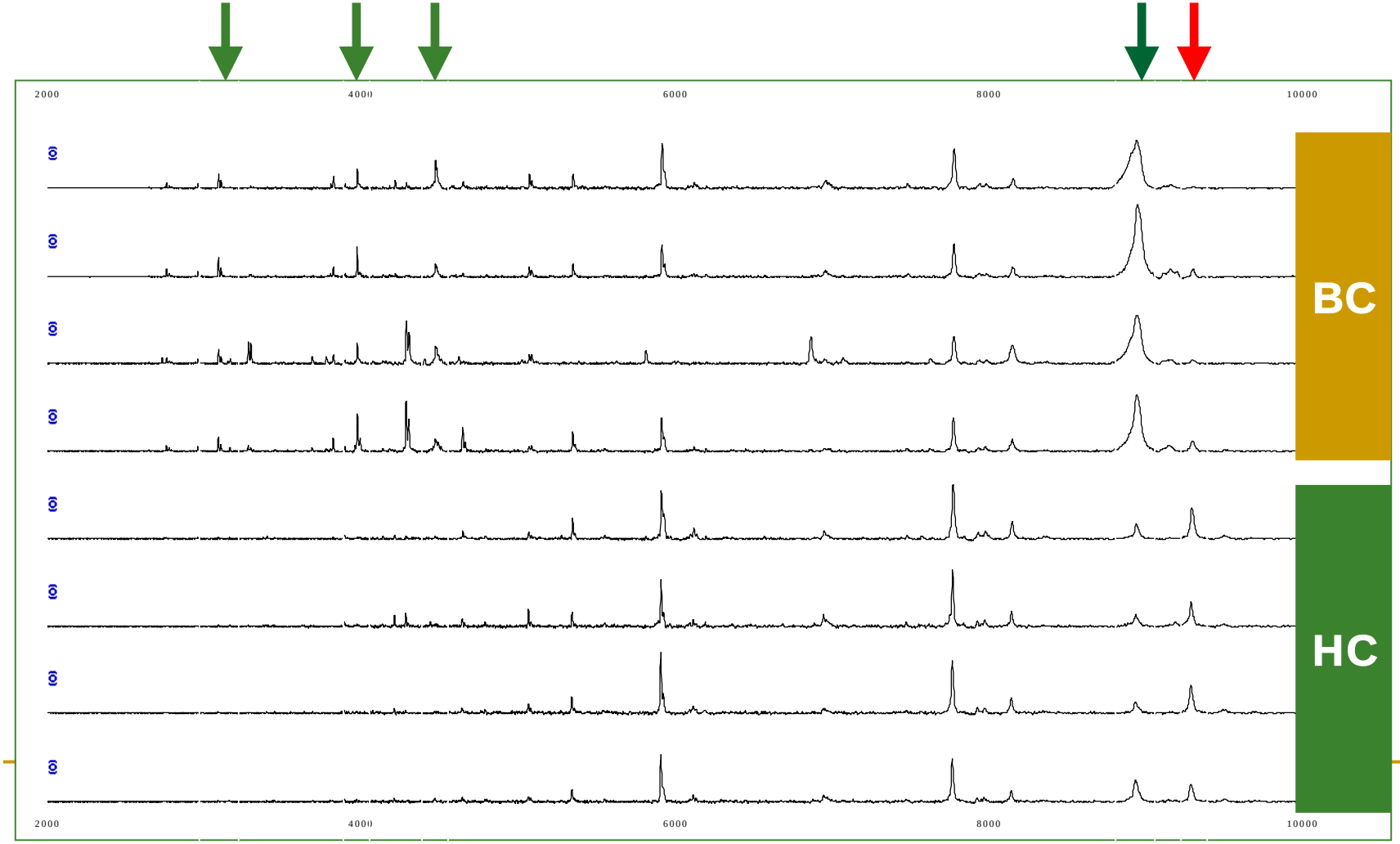
- **Serum samples are divided in 4 groups (BC vs HC):**

group A: 4 vs 4 — **assay development**

group B: 39 vs 39
group C: 47 vs 47
group D: 54 vs 24

- decision tree construction
(using e.g. group D)
- prospective validation
(using groups B and C)

Protein profile



Biostatistics

> 25 significant proteins

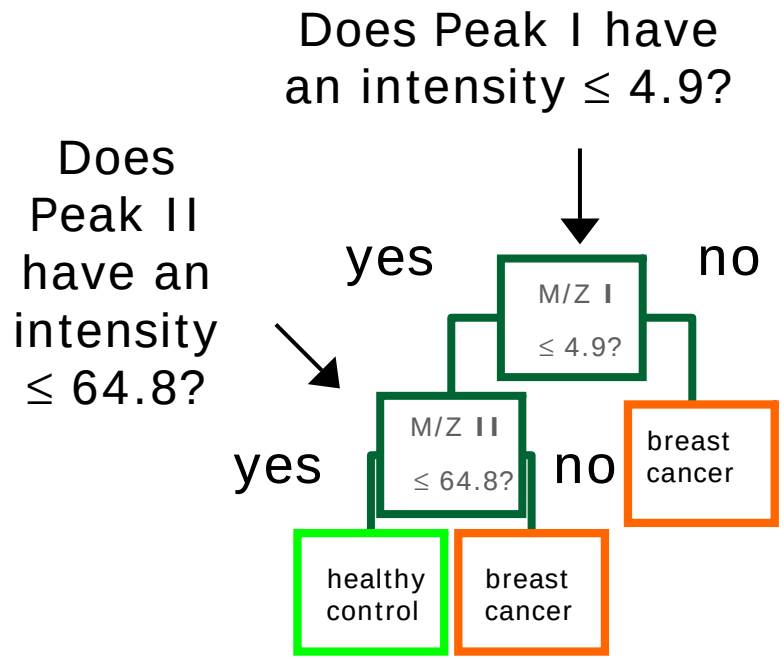
↓

5 proteins were selected for use in decision trees

I	0,0000000000
II	0,0000000000
III	0,0000000000
IV	0,0000000002
V	0,0000000002
6	0,0000001160
7	0,0000004404
8	0,0000005519
9	0,0000011993
10	0,0000079700
11	0,0000119020
12	0,0000168053
13	0,0000176464
14	0,0000194503
15	0,0000204169
16	0,0000214291
17	0,0000502253
18	0,0000632240
19	0,0001543835
20	0,0001835788
21	0,0004079332
22	0.0006616940
23	0.0022242915
24	0.0046378632
25	0.0073711691
26	0.0086531986

Decision tree - example

Construction (group D)



Prospective validation

group B		
Class	Cases	%correct
Control	39	**97.4**
Cancer	39	**74.4**

sensitivity: 93.0%
specificity: 94.2%

Conclusion

- Breast cancer patients can be distinguished from healthy controls by their serum protein profile.
- Demographic co-variables tested did not have influence on the discrimination between breast cancer and healthy
  - (data not shown)

What are these peptides / proteins?

Does is matter?

Tissue proteomics

- Laser Capture Micro Dissection
- MALDI – TOF MS
  - Matrix-assisted laser desorption/ionization
- Tissue imaging
  - Caprioli RM (Vanderbilt University)
- Protein microarrays

Future perspectives

- Potential applications in malignancy:
  - Screening (cave: specificity and sensitivity)
    - Early diagnosis of patients at risk
  - Monitoring progression or relapse of disease
  - Prognostic factor
    - Disease free or overall survival
  - Treatment monitoring and follow-up
  - Predictive factor
    - Treatment response (surrogate end point)
  - Pharmacodynamics
    - tissue distribution by proteomic imaging
  - Insight into pathogenesis of diseases
  - Possible new leads for drug targeting

Take home message - proteomics

- Global study of the proteome of a sample
- High – throughput
- Profiling – thus using changes in protein content
  - Expression, post-transcriptional modifications, degradation
- Technically difficult
  - Reproducibility
  - Subject to bias
- Diverse clinical utilities
- Difficulties in validation or standardized quality control
  - Differences in sample preparation and calibration

Acknowledgement

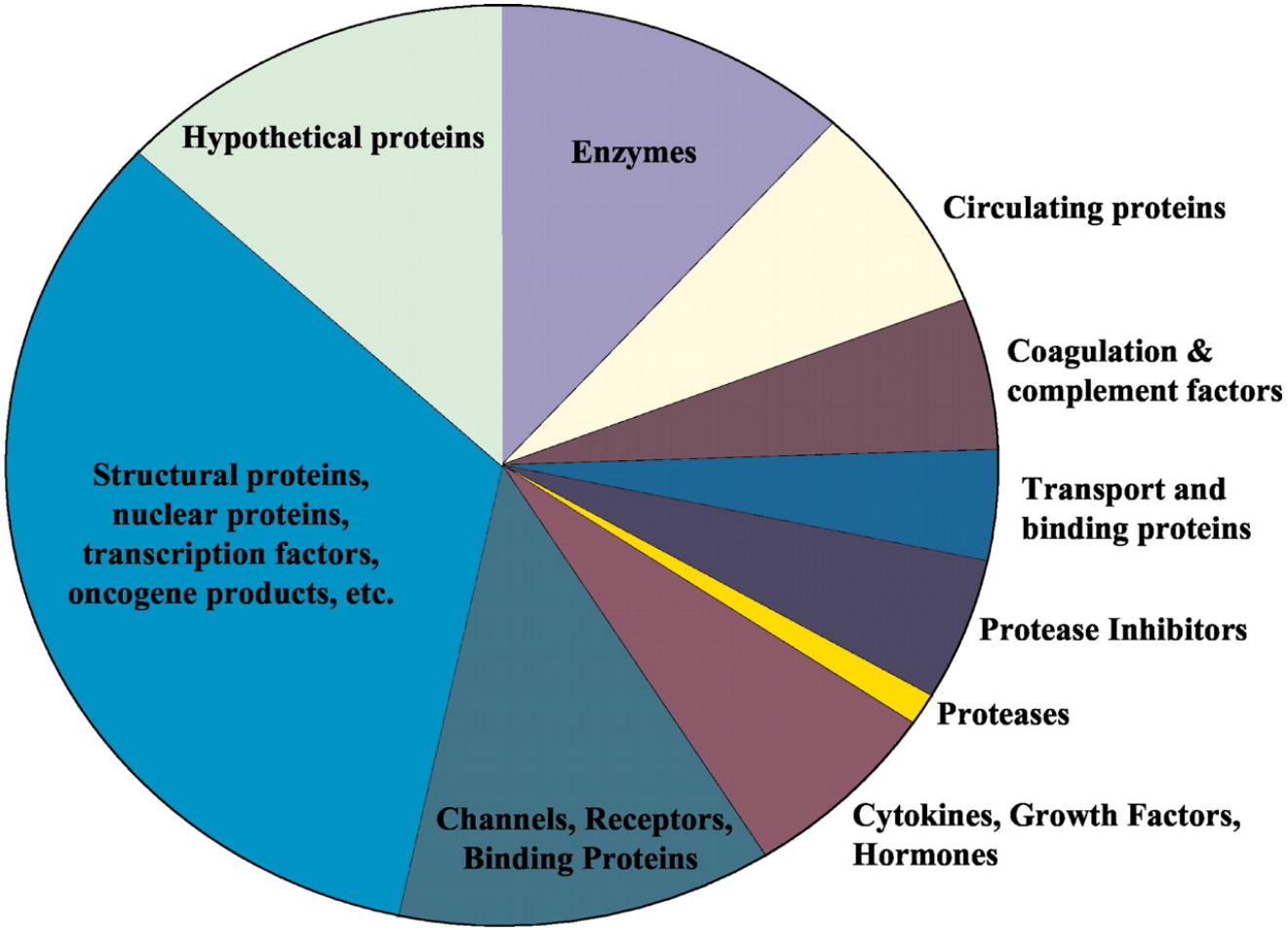
- Department of Medical Oncology

  The Netherlands Cancer Institute
  - Prof. Dr. Jan H.M. Schellens

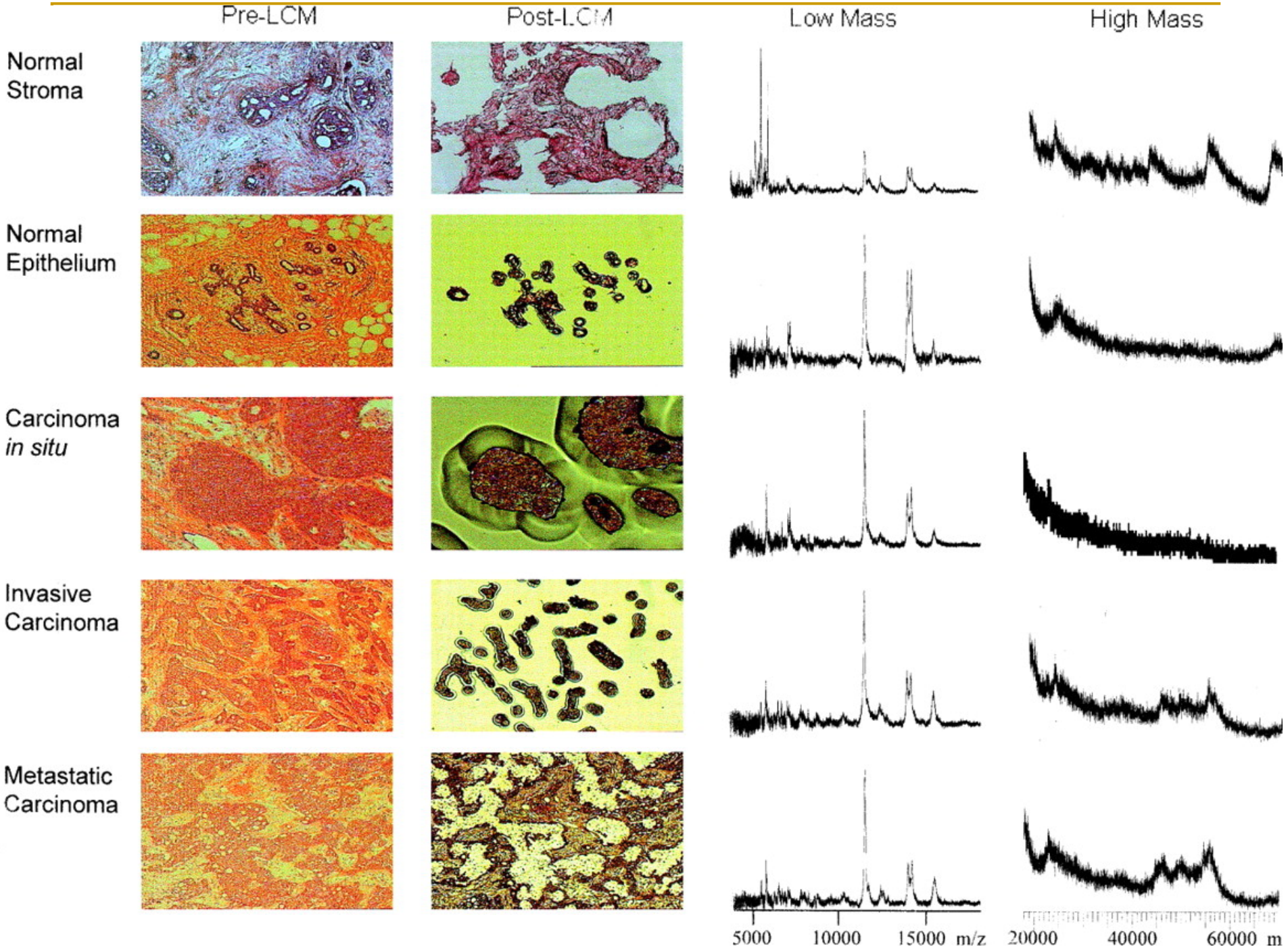
- Department of Pharmacology

  Slotervaart Hospital
  - Prof. Jos Beijnen Pharm
  - M.C. Gast

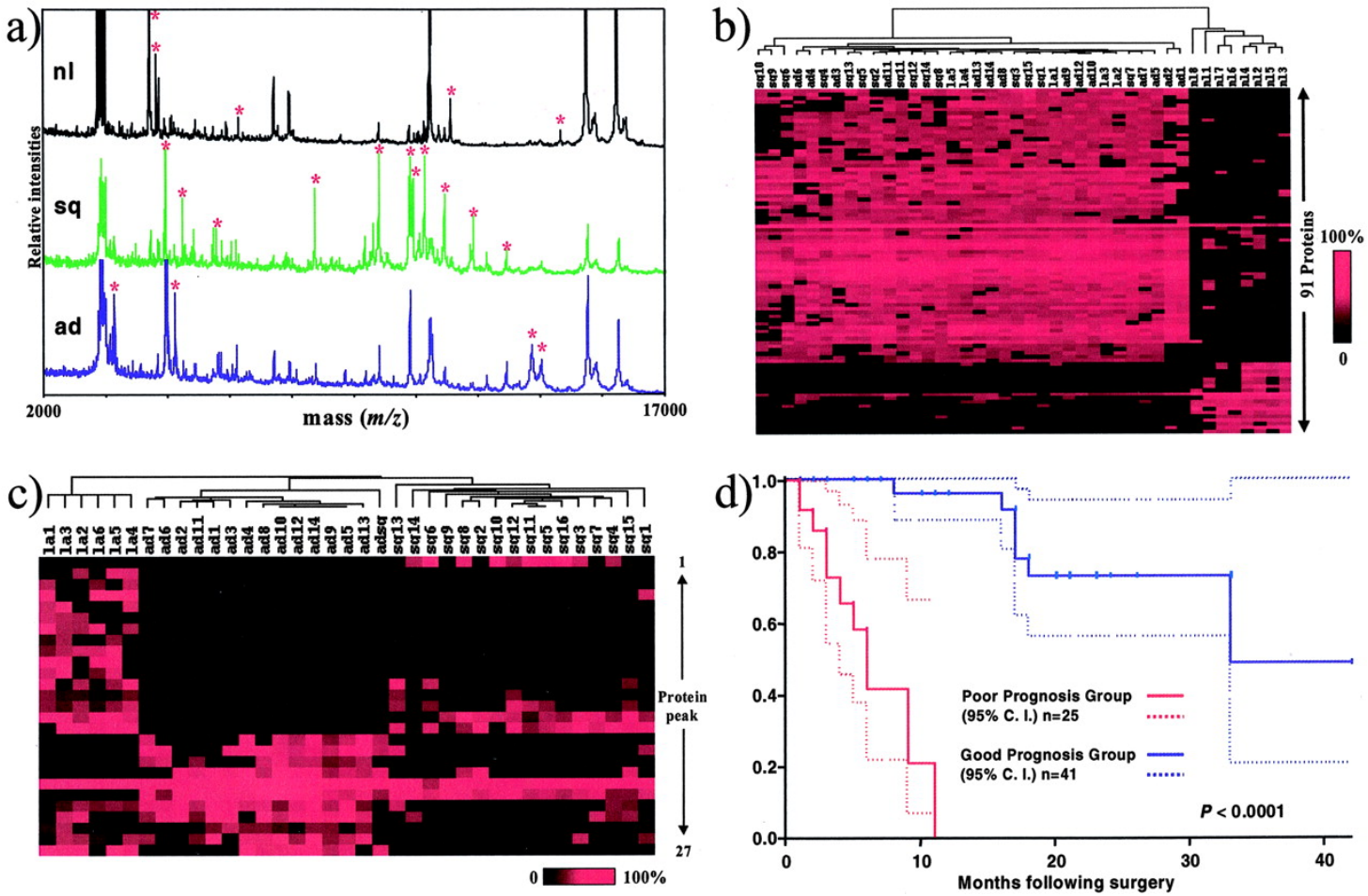
Relative numbers of proteins identified within the LMW serum proteome



Tirumalai, R. S. (2003) Mol. Cell. Proteomics 2: 1096-1103


Pre-LCM
Post-LCM
Low Mass
High Mass
Normal Stroma
Normal Epithelium
Carcinoma in situ
Invasive Carcinoma
Metastatic Carcinoma
5000
10000
15000 m/z
20000
40000
60000 m

Tissue profiling - prognosis



R. Caprioli et al

Assay development

- Group A: 4 BC vs 4 HC

  - **selection of ProteinChip Array (SAX, WCX, H4, H50, IMAC Ni/Cu)**
  - **optimization of sample pretreatment (denaturation in ureum / CHAPS / DTT)**
  - **selection of binding- and wash-procedure (% acetonitrile, buffer pH 4–9)**
  - **criteria: high, distinctive protein levels and protein profiles**

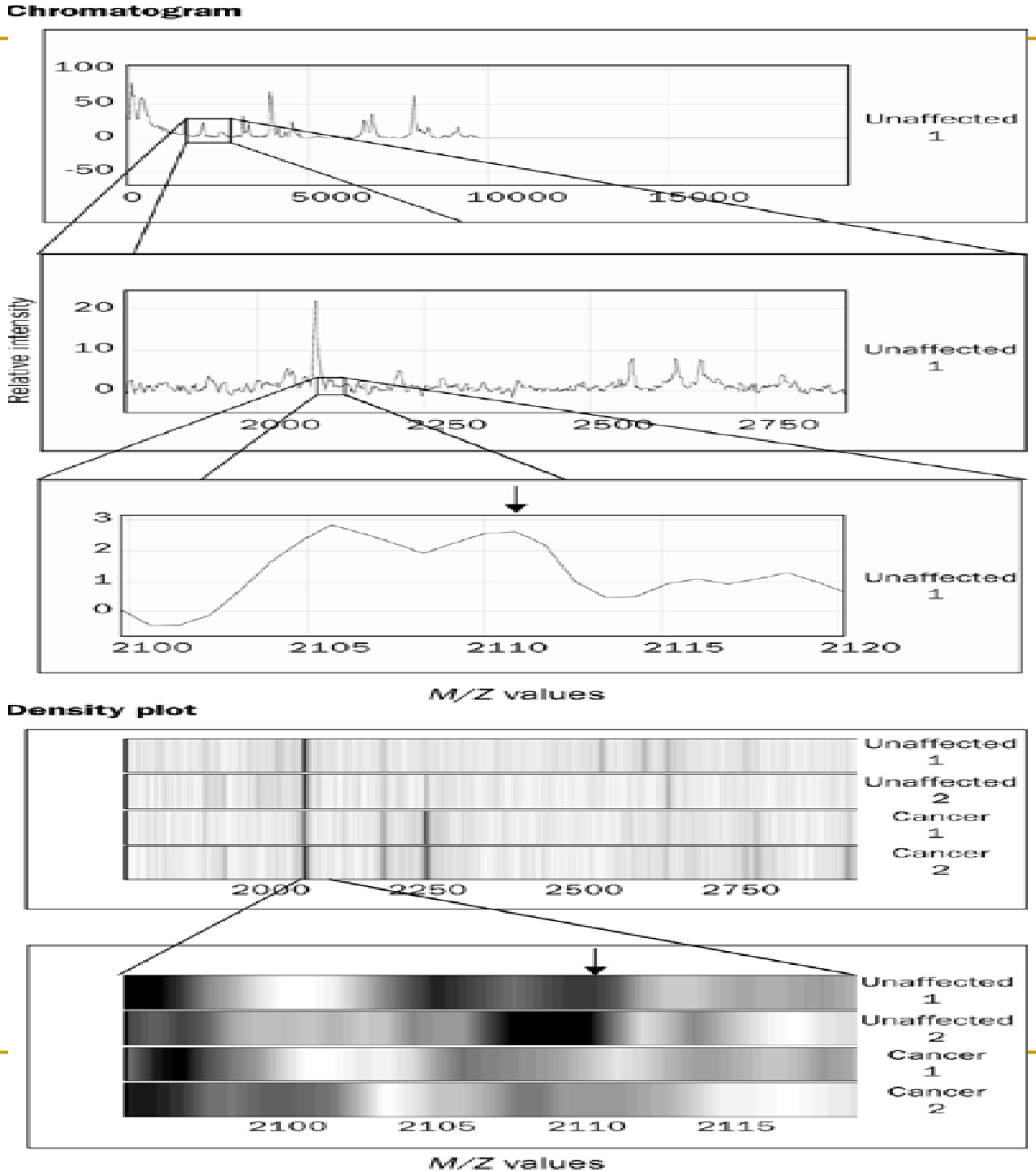
  - **<u>results</u>: IMAC30-Ni array**
    Ureum 9M / CHAPS denaturation
    PBS pH7.4 / 0.5M NaCl / TritonX-100

Proteomic pattern to identify ovarian cancer

	Training	Masked
Unaffected women	50	66
(4 subgroups)		
Cancer	50	50
St 1	6	18
St II, III, IV	44	32

- Spectrum (patient): 15.200 data points
- Cluster analysis
- Comparison of diverse profiles
- Optimum discrimination pattern
  - 5 M/Z values (534, 989, **2111**, 2251, 2465)

Petricoin et al. Lancet 2002 ; 359: 572-77


Chromatogram
100
50
0
-50
0
5000
10000
15000
Unaffected 1
Relative intensity
20
10
0
2000
2250
2500
2750
Unaffected 1
3
2
1
0
2100
2105
2110
2115
2120
Unaffected 1
M/Z values
Density plot
Unaffected 1
Unaffected 2
Cancer 1
Cancer 2
2000
2250
2500
2750
Unaffected 1
Unaffected 2
Cancer 1
Cancer 2
2100
2105
2110
2115
M/Z values

Proteomic pattern to identify ovarian cancer

Classification by proteomic pattern			
	Cancer	Unaffected	New
Unaffected women	3/66	47/66	16/66
Cancer			
St 1	18/18	0/18	0/18
St II, III, IV	32/32	0/32	0/32

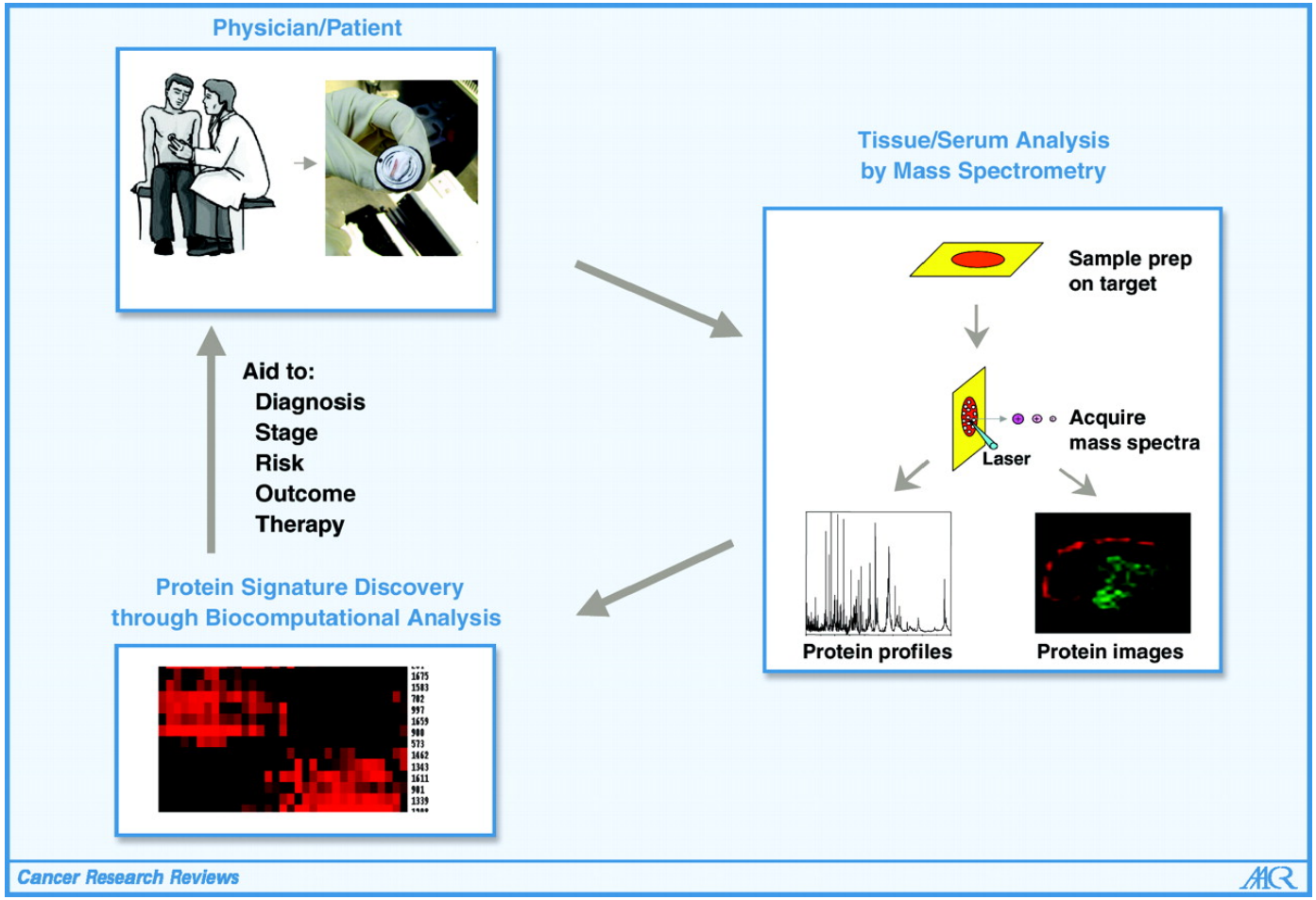
Petricoin et al. Lancet 2002 ; 359: 572-77

Proteomic pattern to identify ovarian cancer

	Proteomics	CA 125 -/+ Ultrasound
Sensitivity	100%	50-80%
Specificity	95%	
Positive predictive value	94%	10 / 20%

Petricoin et al. Lancet 2002 ; 359: 572-77

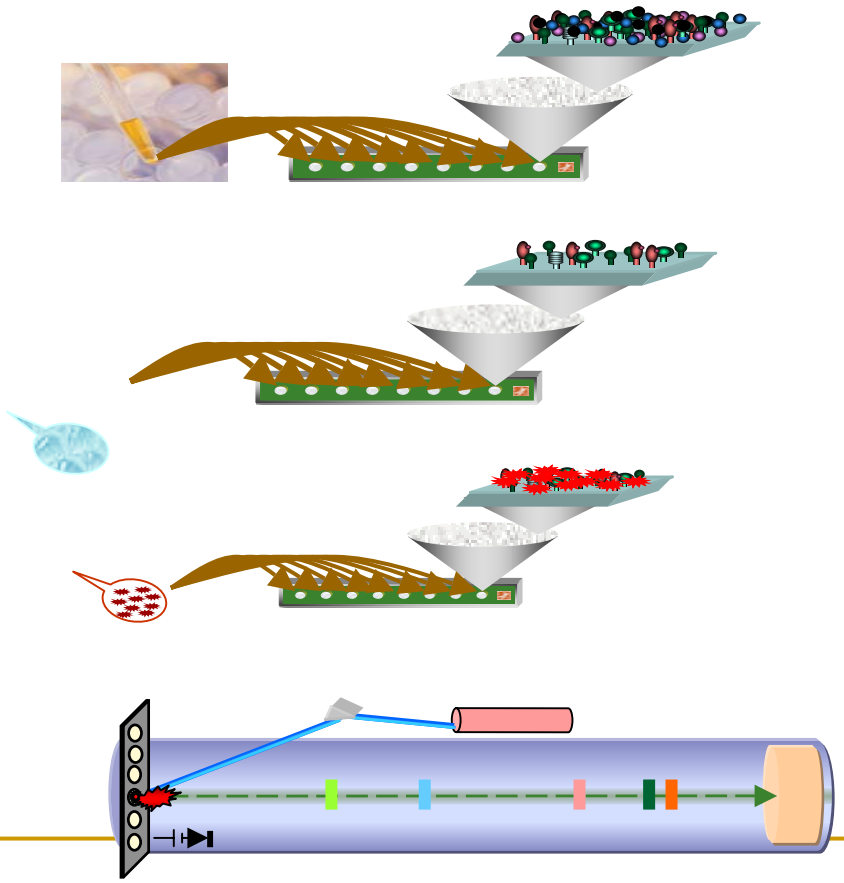
Requirements for molecular signatures



Caprioli, R. M. Cancer Res 2005;65:10642-10645

Sample processing (1)

1. Apply sample
2. Wash ProteinChip Array
3. Add Energy Absorbing Molecules or “Matrix”
4. Analyze in a ProteinChip Reader



Protein biomarkers (2)

22 proteins constitute ~99% of the plasma protein content

80 %	19 %	1 %
Albumin	C8-complement	Others
IgG total	C1q-complement	
Transferrin	C9-complement	
Fibrinogen	Prealbumin	
IgA total	Complement Factor B	
α-2-Macroglobulin	C4-complement	
IgM total	Ceruloplasmin	
α-1-Antitrypsin	Factor H	
C3-complement	Lipoprotein (a)	
Haptoglobin	α_1-Acid Glycoprotein	
	Apolipoprotein B	
	Apolipoprotein A-1	

Mol Cel Prot 2003, 2: 1096-1103