New Diagnosis Systems based on Biomarker Pattern Recognition

June 2004

Jonathan Heller, Ph.D.
Vice President, Information and Project Planning
Predicant Biosciences, Inc.

The Vision

Physicians seek a clinically reliable method of detecting, diagnosing and monitoring complex disease states.

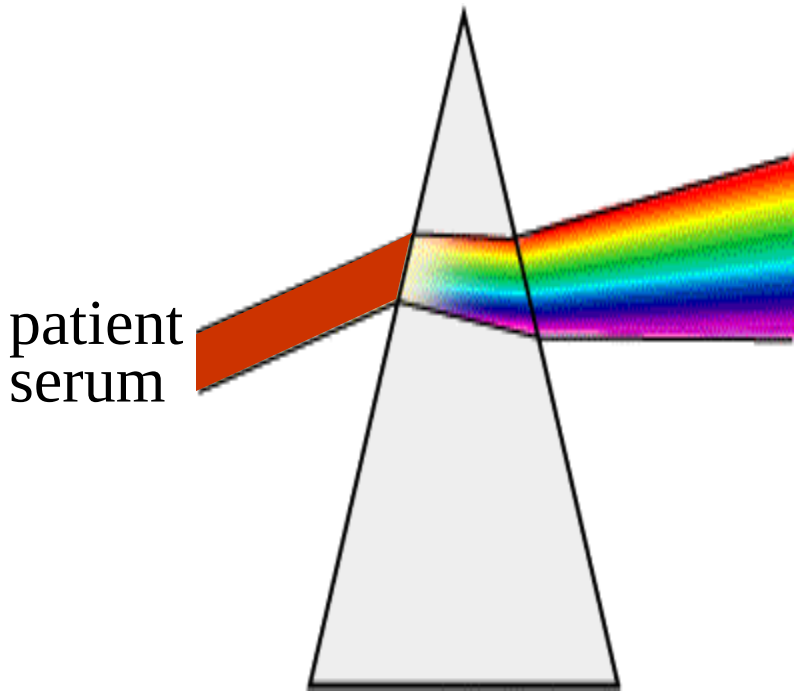
A revolutionary class of analyses is being developed to transform clinical patient care based upon the identification of protein patterns in serum that determine and differentiate biological state. This offers new hope for improved disease outcomes.

The ability to define and monitor biological states will revolutionize medicine.

Predicant BIOSCIENCES

Defining Biological States

The complement of proteins, protein fragments and peptides present at any specific time defines who and what we are at that moment, as well as our state of health or disease.



Healthy
Early stage cancer

Benign tumor
Malignant tumor

Drug efficacy
Drug non-response

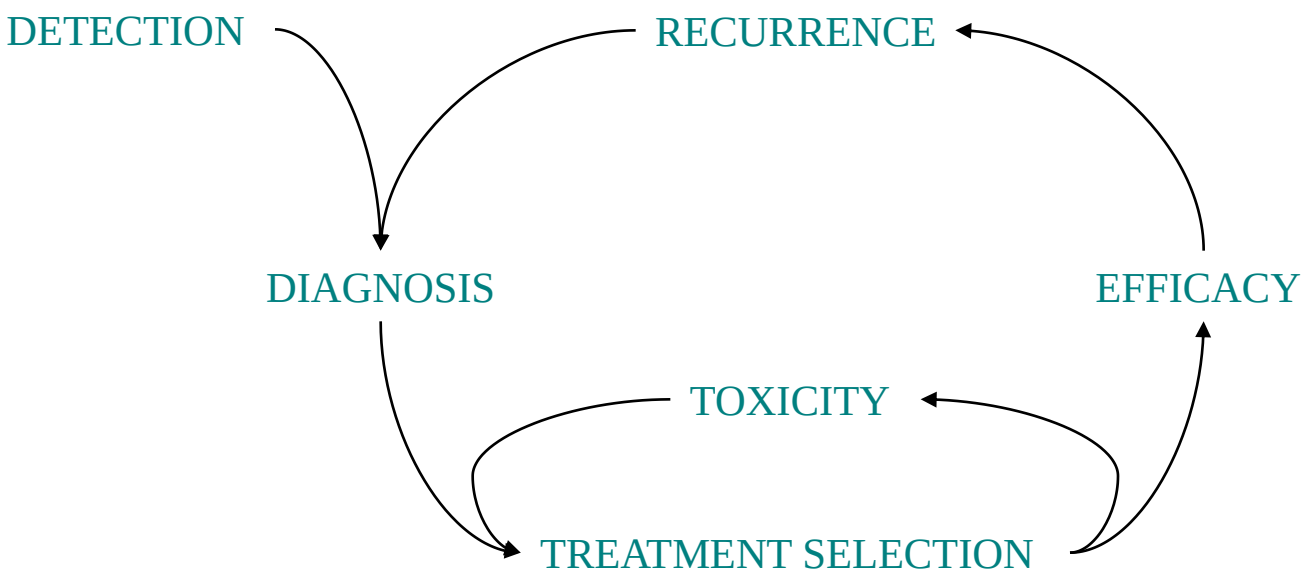
Remission
Recurring cancer

Concentrations of proteins, protein fragments and peptides in blood fluctuate in disease.

Predicant BIOSCIENCES

Applications Throughout Disease Management

Early detection, improved diagnosis and treatment selection, and better monitoring will improve and lengthen the lives of those with disease.

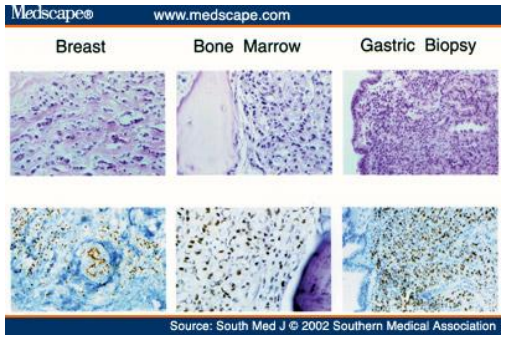


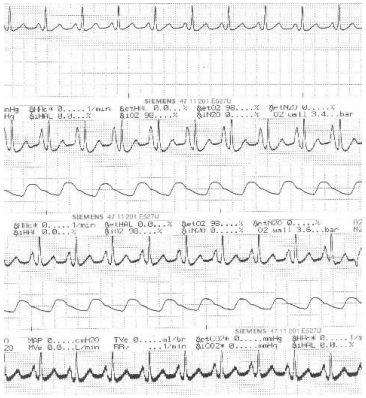
Improved methods in cancers and cardiovascular disease alone would have an enormous impact.

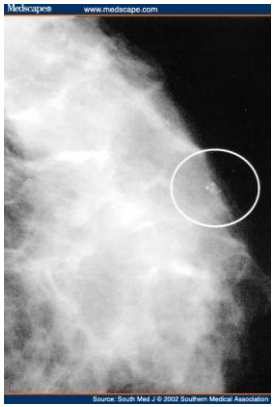
Better methods are needed throughout the disease management process to enhance the lives of patients.

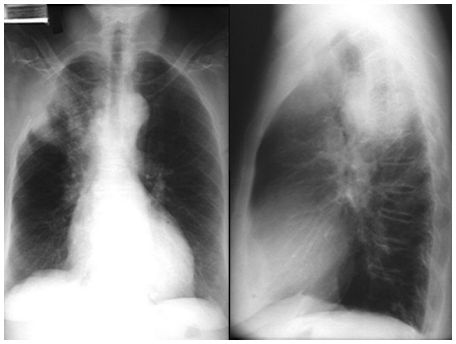
Predicant BIOSCIENCES

Current Disease Diagnosis Methods









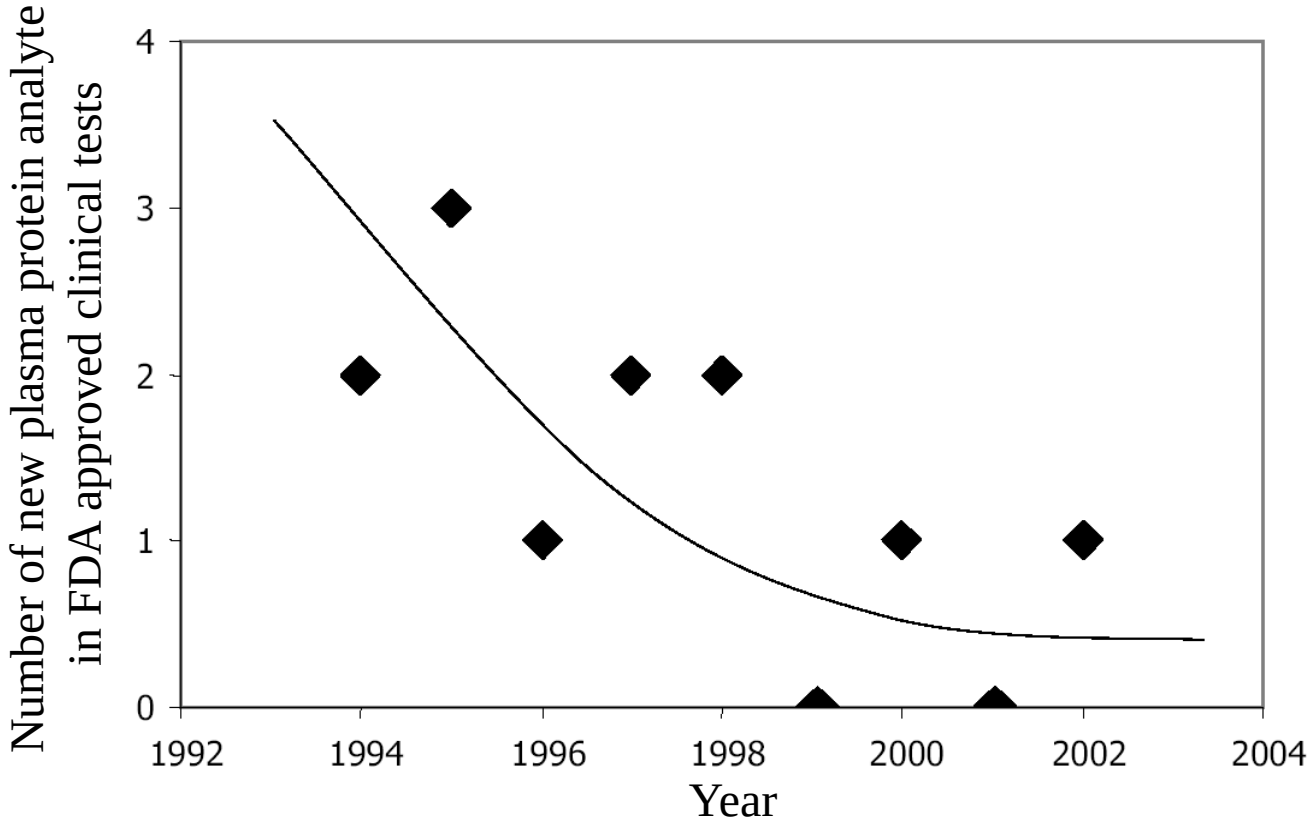
Doctors have many sources of data: medical/family history, signs and symptoms, imaging, blood test results, biopsy…

Doctors recognize patterns using large amounts of data to make diagnoses.

Predicant BIOSCIENCES

Single Protein Tests are Inadequate

Existing single protein tests tend to lack sensitivity and/or specificity. For example PSA is ~90% sensitive, but only ~25% specific, and therefore is not a very accurate test.



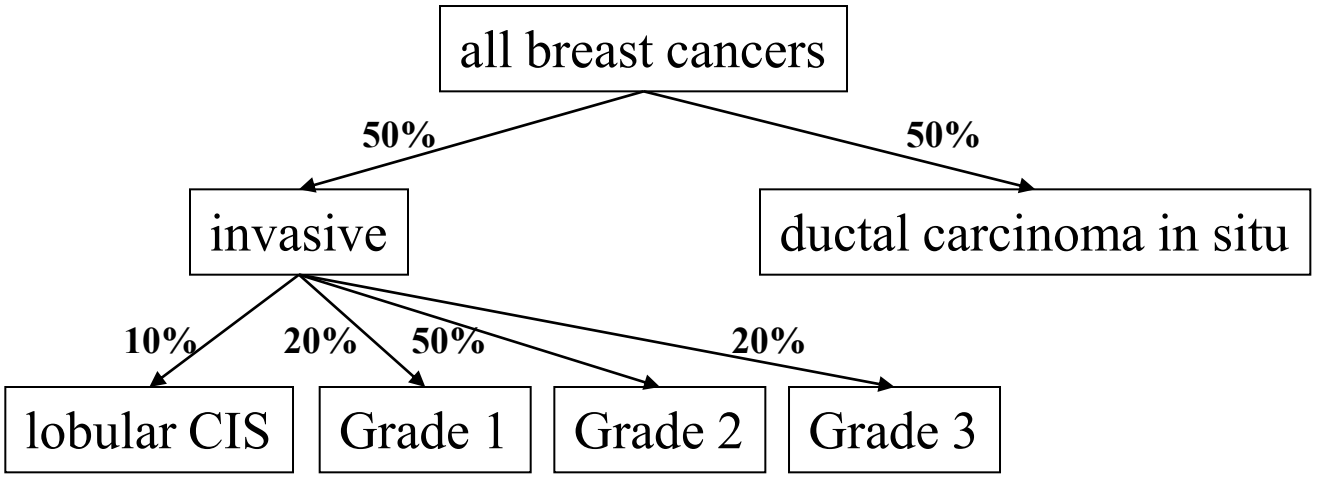
The rate of introduction of new single protein tests is falling and existing single marker tests are insufficient. No smoking guns.

Predicant BIOSCIENCES

Reference: NL Anderson, et al., Molecular & Cellular Proteomics(2002) v1 p845

Cancer is not a single disease

Breast cancer survival rates as a function of spread:
localized: 97%
regional : 78%
distant: 23%



Additional breast cancer subtypes used to specify treatment:

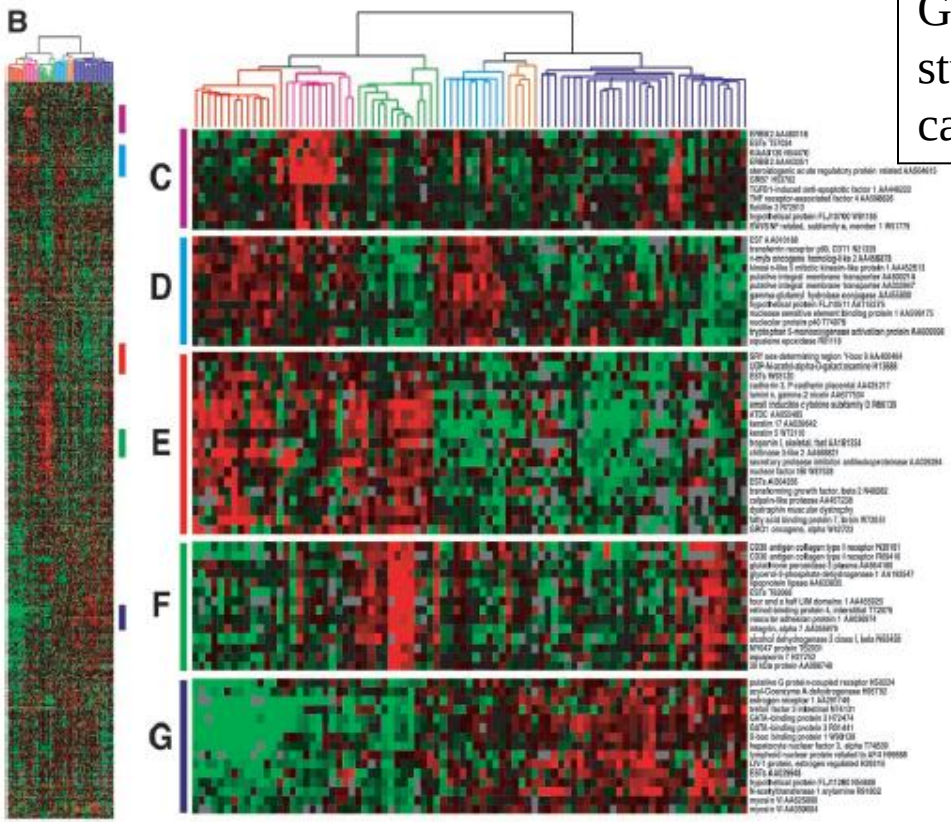
- Estrogen/progesterone receptor overexpression
- Her-2/neu overexpression
- Plasminogen activator levels

Multiple forms of breast cancer exist. Each form has different treatment options and survival rates.

Predicant BIOSCIENCES

Complex Diseases have Complex Signatures

Subsets of genes can be used to help diagnose and guide treatment of disease. Large amounts of data have become available.



Gene expression studies of breast cancer patients

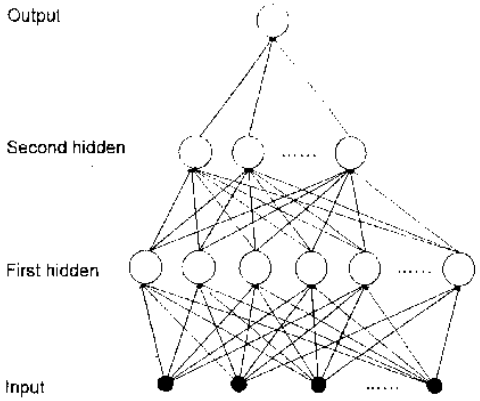
Disease is complex & heterogeneous. The body changes in many ways in disease. New approaches to diagnostics are needed.

References: T Sorlie, et al., PNAS(2001) v98 p10869.
L van't Veer, et al., Nature(2002) v215 p530.

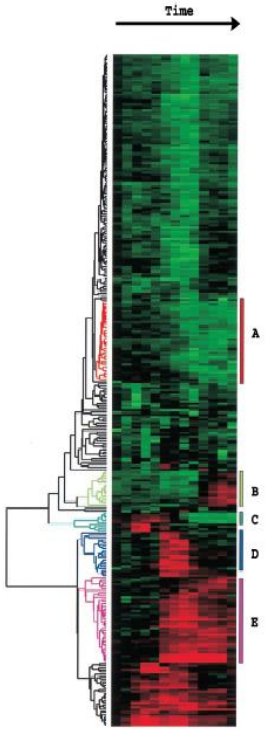
Predicant BIOSCIENCES

Pattern Recognition on Large Biological Data Sets

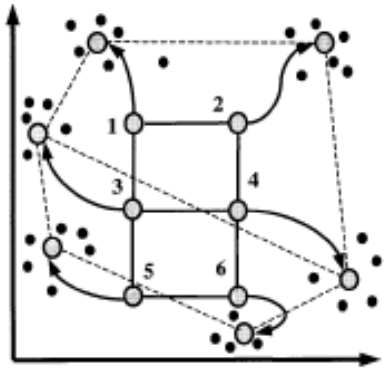
Neural Networks



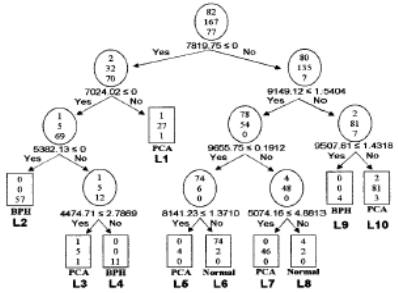
Cluster Analysis



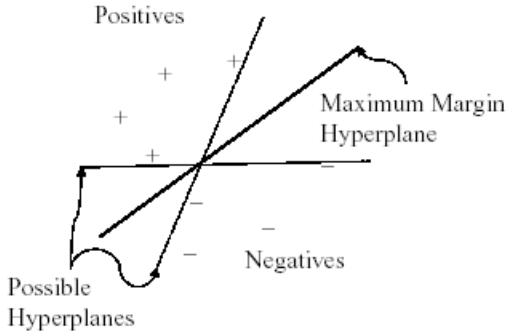
Self Organizing Maps



Decision Trees



Support Vector Machines



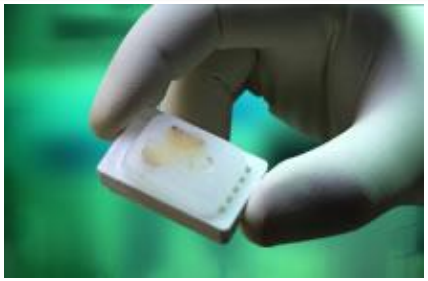
Pattern recognition has become an necessity to find complex patterns in large data sets.

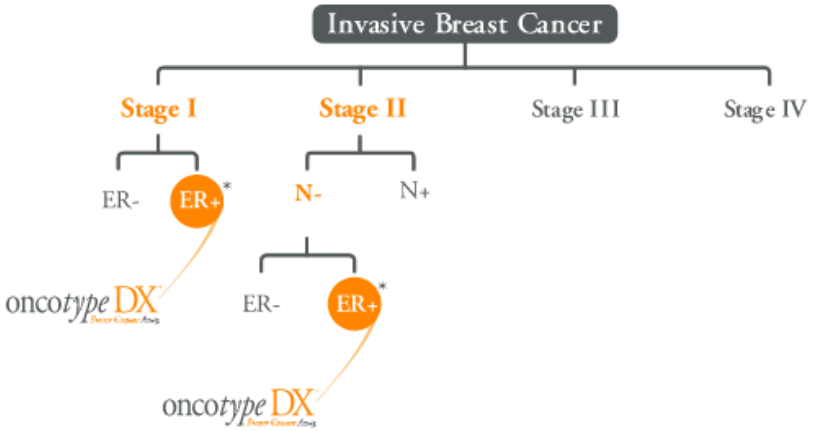
References: J Vohradsky, Electrophoresis (1997) v18 p2749.
MB Eisen et al., PNAS(1998) v95 p14863.
P Tamayo et al., PNAS(1999) v96 p2907.
MP Brown et al., UCSC Technical Report UCSC-CRL-99-09 (1999)

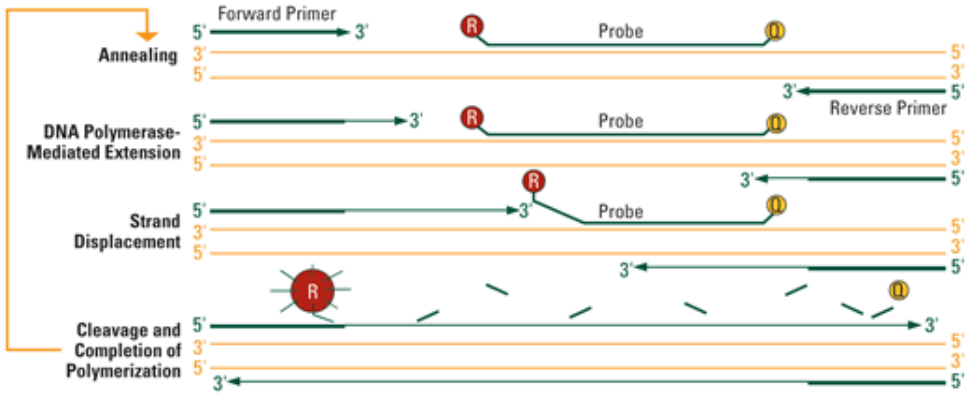
Predicant BIOSCIENCES

Multivariate Gene Expression Based Tests

OncotypeDX™ is a new test for breast cancer recurrence risk from Genomic Health Inc. that is based on the expression of 21 genes.





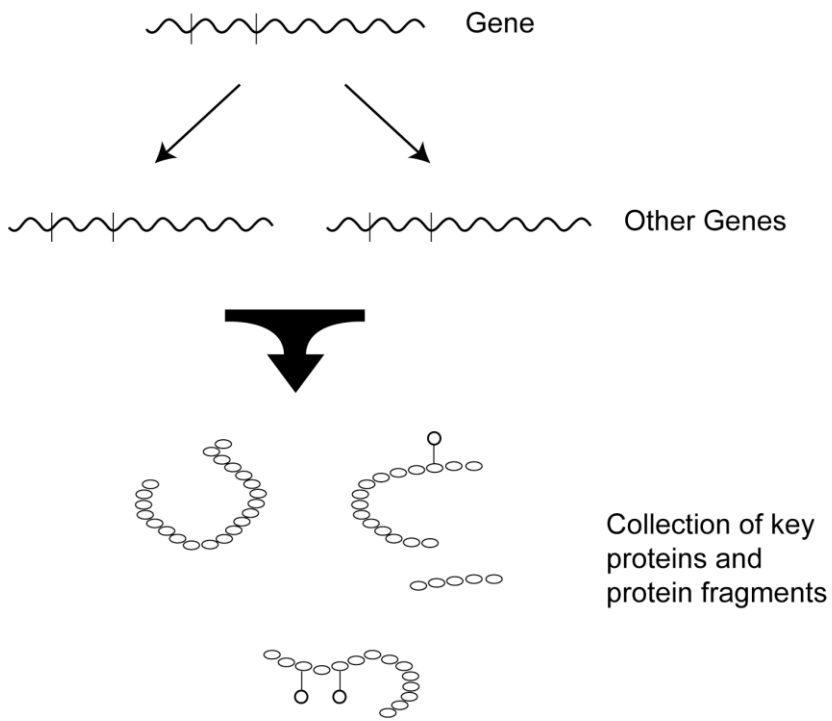


Patterns of small numbers of genes in tissue have been found to be informative.

Predicant BIOSCIENCES

Proteins are the Actors; Blood is the Stage

Biological state is best measured by analyzing the collection of key proteins and protein fragments from serum.



Genotype Analysis: "Behind the Scenes"

Phenotype Analysis: "Direct Picture"

Number of genes: ~50,000

Number of proteins (splice variants, PTM): ~2,500,000?

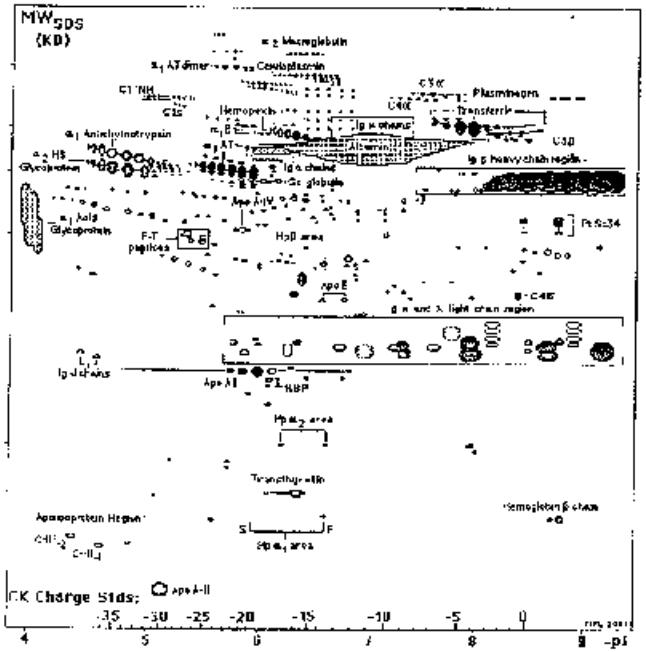
Measurement of protein fragments gives a more direct picture of biological state compared with genetic analyses.

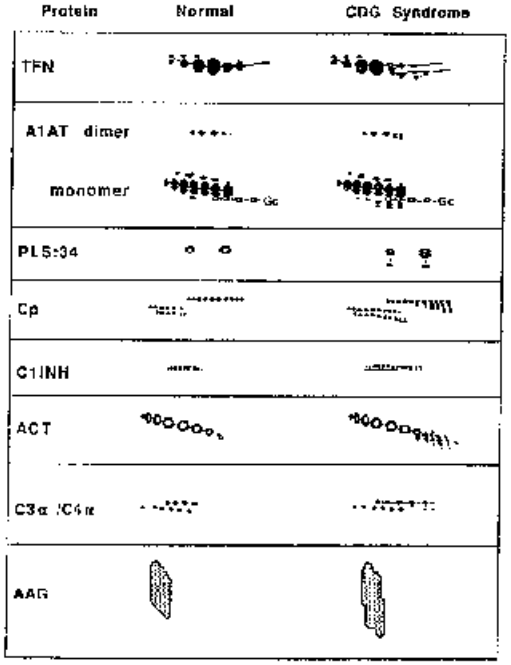
Predicant BIOSCIENCES

Protein Data Quantity ≠ Clinical Utility

Prior to gene expression studies, 2DGE showed that selected features from large data sets could be used in diagnosis, but the methods are not clinically viable.

2 dimensional gel electrophoresis of serum from a patient with carbohydrate-deficient glycoprotein syndrome and from a control





Vast amounts of data have become available, both protein and nucleic acid. That data must be mined to select features and find patterns.

Predicant BIOSCIENCES

Reference: HH Harrison et al., Clin Chem. (1992) v38 p1390.

A New Methodology for Diagnostic Discovery

Sample Preparation & Separations

↓

Ionization and Mass Spectrometry Detection

↓

Signal Processing and Pattern Recognition

Serum profiling with mass spec has many advantages:

- Serum
- Direct measurement of key species, e.g. proteins & peptides
- Sensitivity
- Complex patterns can be measured

THE LANCET ¥Vol 359 ¥February 16, 2002 ¥www.thelancet.com

Mechanisms of disease

Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

[CANCER RESEARCH 62, 3609–3614, July 1, 2002]

Advances in Brief

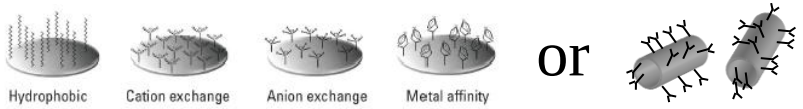
Serum Protein Fingerprinting Coupled with a Pattern-matching Algorithm Distinguishes Prostate Cancer from Benign Prostate Hyperplasia and Healthy Men[1]

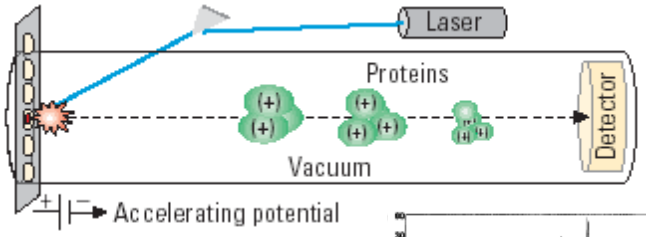
Bao-Ling Adam,[2] Yinsheng Qu,[2] John W. Davis, Michael D. Ward, Mary Ann Clements, Lisa H. Cazares, O. John Semmes, Paul F. Schellhammer, Yutaka Yasui, Ziding Feng, and George L. Wright, Jr.[2,3]

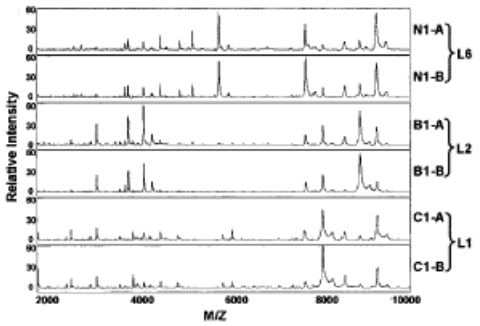
Mass Spectrometry is being pursued as a diagnostic discovery and clinical assay platform.

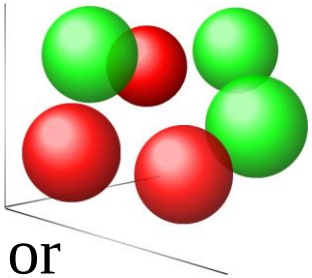
Predicant BIOSCIENCES

MALDI/SELDI Biomarker Discovery

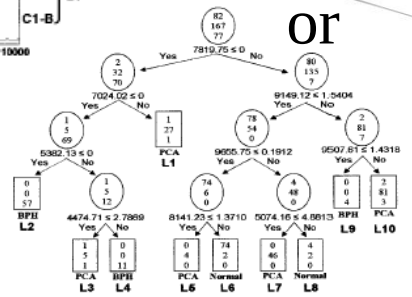








or



Protein capture and laser based ionization has been used for biomarker discovery.

Predicant BIOSCIENCES

Reference: HJ Issaq et al., Anal. Chem.(2003) v75 p148A.

Limitations of Current Methods

- Ion suppression limits the number of proteins observed
- Mass spec resolution in some studies leads to multiple proteins per observed peak
- Poor reproducibility (day-to-day, sample-to-sample, lab-to-lab) due to chip variability and ionization method
- Protein identification difficult/impossible with original instrumentation
- Lack of robust signal processing methods supply pattern recognition tools with noisy, low confidence data
- Feature selection and pattern recognition tools being used are incapable of looking at pattern consisting of combinations of large numbers of proteins

- Validation of published results has not been possible

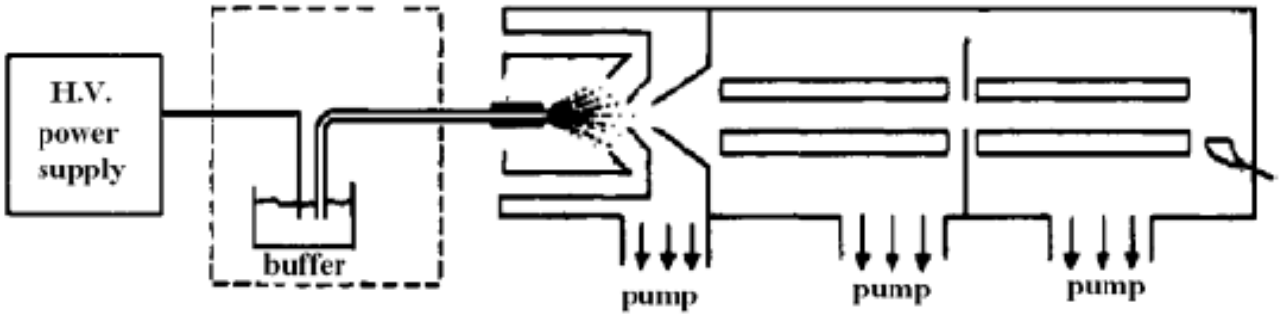
Recently published biomarkers have not been validated and current technologies are not clinically viable.

Predicant BIOSCIENCES

Existing Alternatives to MALDI/SELDI

Off-line separations & fraction collection w/ LDI or ESI or On-line CE/LC-ESI-MS

While these methods are well established and have been used for serum & bodily fluid analysis, they have inherent limitations.



Advantage: reduces ion suppression, more components revealed

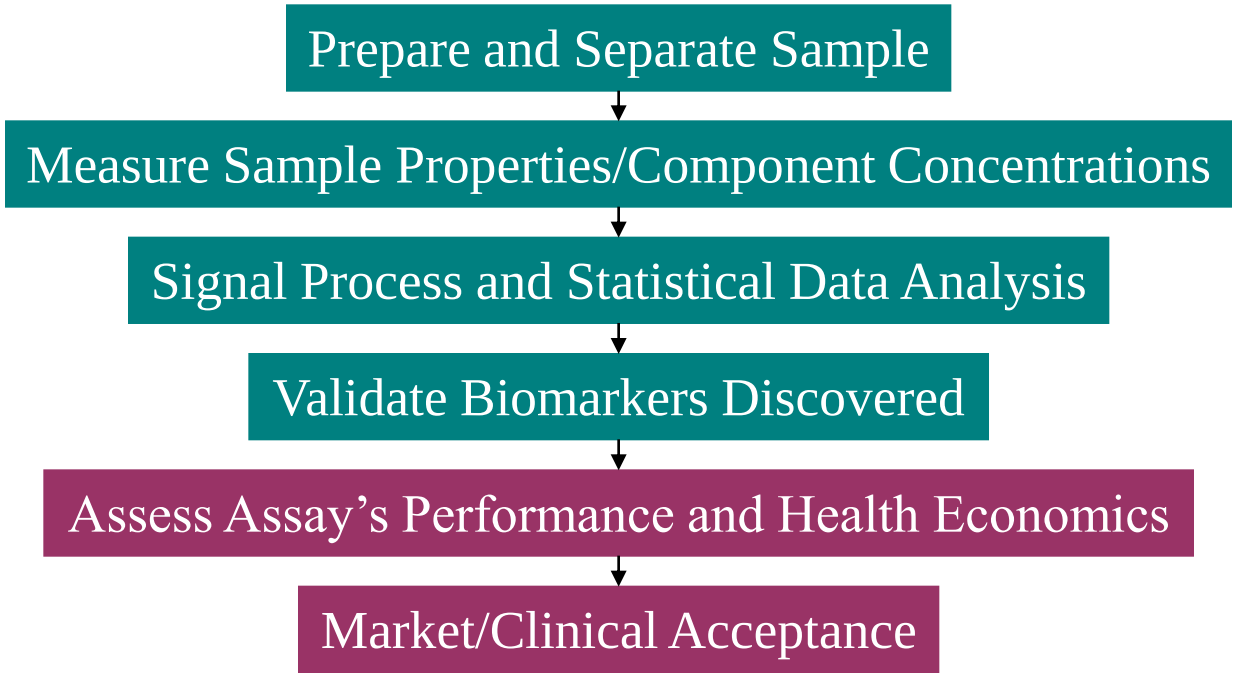
Disadvantages: complexity and length of expt, carry-over, not an integrated solution, reproducibility

New technologies are needed to create a clinically viable mass spec diagnostic system.

References: Schmitt-Kopplin et al., Electrophoresis(2003) v23 p3837.
Tomlinson et al., Electrophoresis(1994) v15 p62.
Tomlinson et al., J Am Soc Mass Spec(1997) v8 p15.
Cao et al., J Am Soc Mass Spec(1998) v9 p1081.

Predicant BIOSCIENCES

A New Solution is Necessary



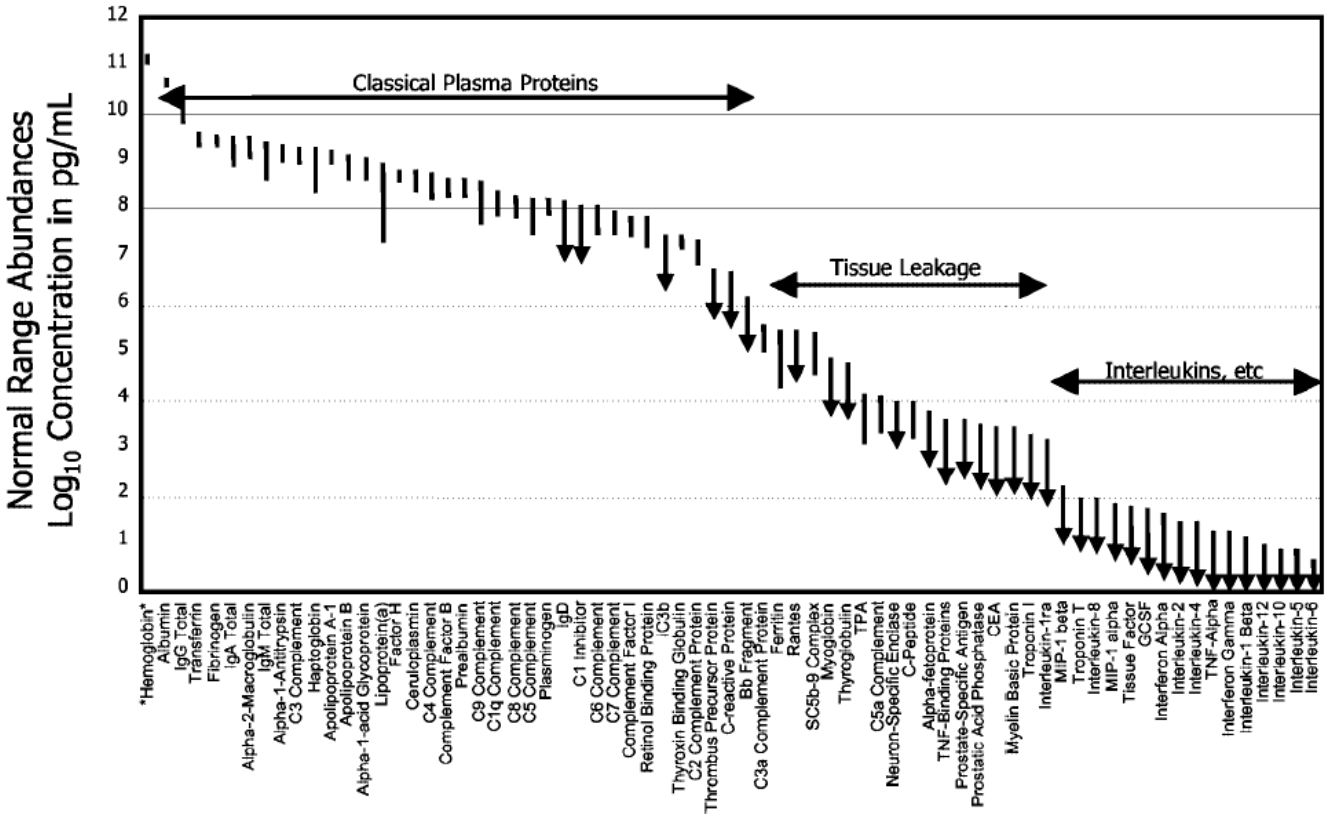
An integrated solution includes technical components and business solutions for discovery and clinical assay.

Predicant BIOSCIENCES

Sample Preparation is Critical

Sample preparation requirements:

- Remove abundant proteins
- Free species bound to high abundance proteins
- Final sample must be in mass spec compatible buffers at low volume in high concentration
- Reproducibility

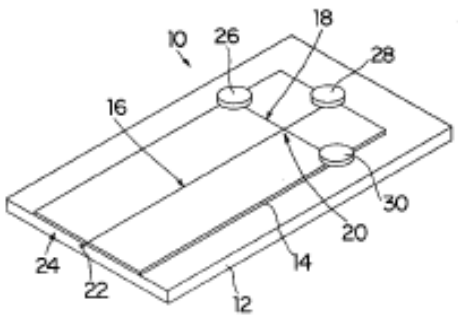


Given the huge dynamic range of proteins in blood, careful sample preparation is needed to reveal the informative proteins.

Predicant BIOSCIENCES

Reference: NL Anderson, et al., Molecular & Cellular Proteomics(2002) v1 p845

Advantages of Microfluidics Technologies

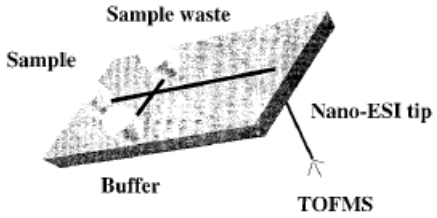


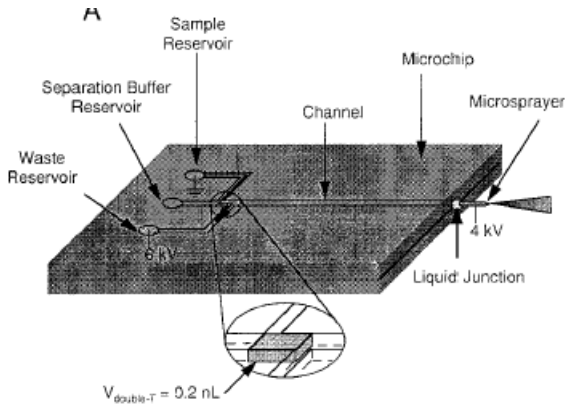
- Well established technology
- Disposable chips; no sample carry-over
- Increases reproducibility by minimizing sample & liquid handling steps
- Fast sample analysis
- Small sample and reagent volumes
- Resulting devices can be easy to use
- Minimal biohazard

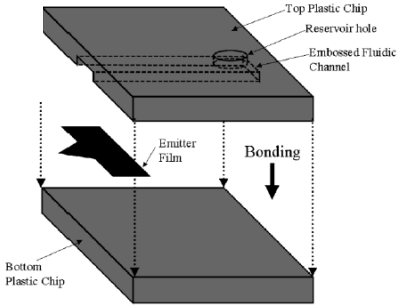
Microfluidics technologies have many of the characteristics desired for clinical diagnostics.

Predicant BIOSCIENCES

On-line Separations with Microfluidics ESI-MS







Several issues with these methods need to be addressed:

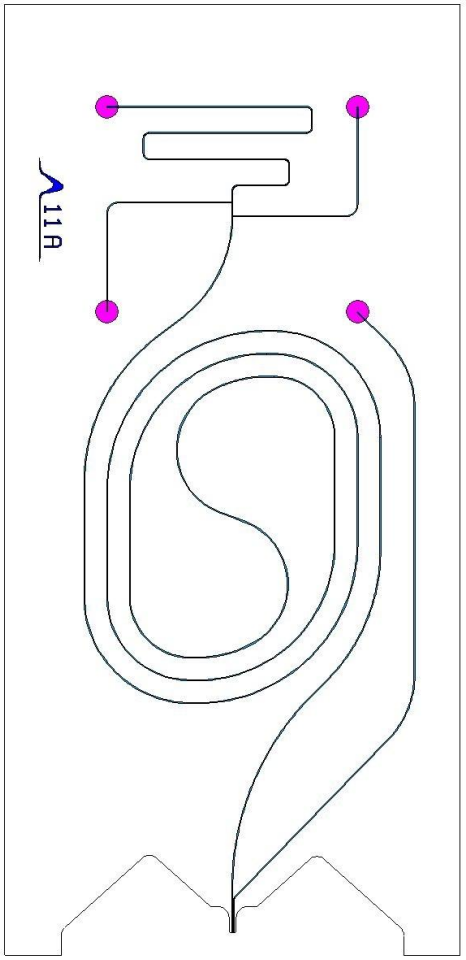
- manufacturability
- separations quality
- reproducibility
- ease of use
- fragility
- automation
- throughput

Separations with Microfluidic ESI-MS have been demonstrated, but need to be improved for use in clinical diagnostics.

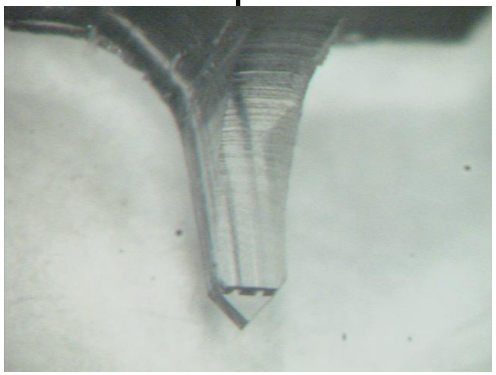
References in Anal Chem: Q Xue et al. (1997) v69 p426. RS Ramsey et al. (1997) v69 p1174. D Figeys et al. (1997) v69 p3153. B Zhang et al. (1999) v71 p3258. J Kameoka et al. (2002) v74 p5897.

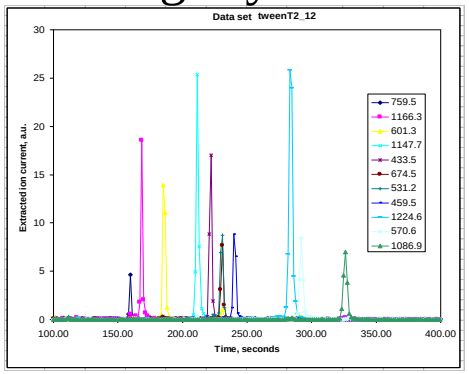
Predicant BIOSCIENCES

A Microfluidic Device for CE-ESI-MS



- Separation channel geometry maximizes separation quality and minimizes hydrodynamic flow
- ESI tip, chip and coatings are manufacturable and give reliable, reproducible spray and high sensitivity
- Multiple channels open onto the tip to provide electrical contact – no bubble formation at tip or arcing
- Reproducible injection to form reproducible sample plug
- Recessed tip minimizes biohazard and fragility

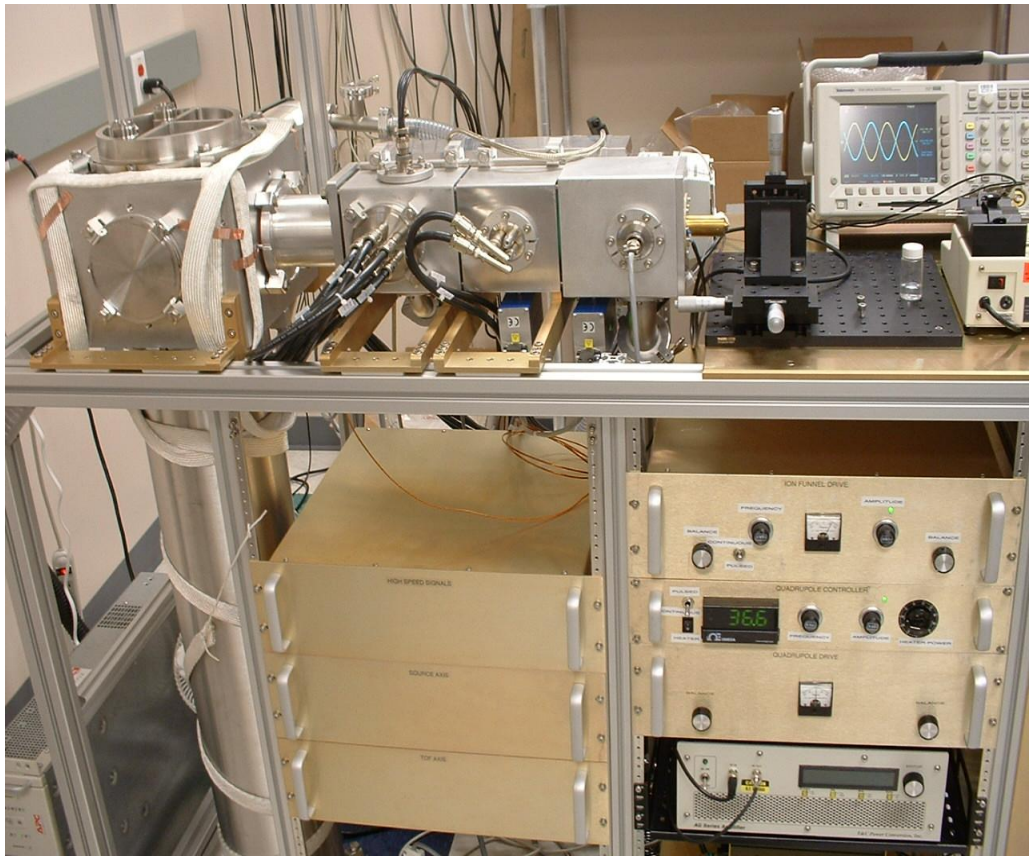




The chip that has been developed has the necessary characteristics to be a clinical diagnostic tool.

Predicant BIOSCIENCES

High Sensitivity Mass Spectrometry



Goals:
- High sensitivity
- High resolution
- High dynamic range
- Easy to use
- Small size
- Low cost

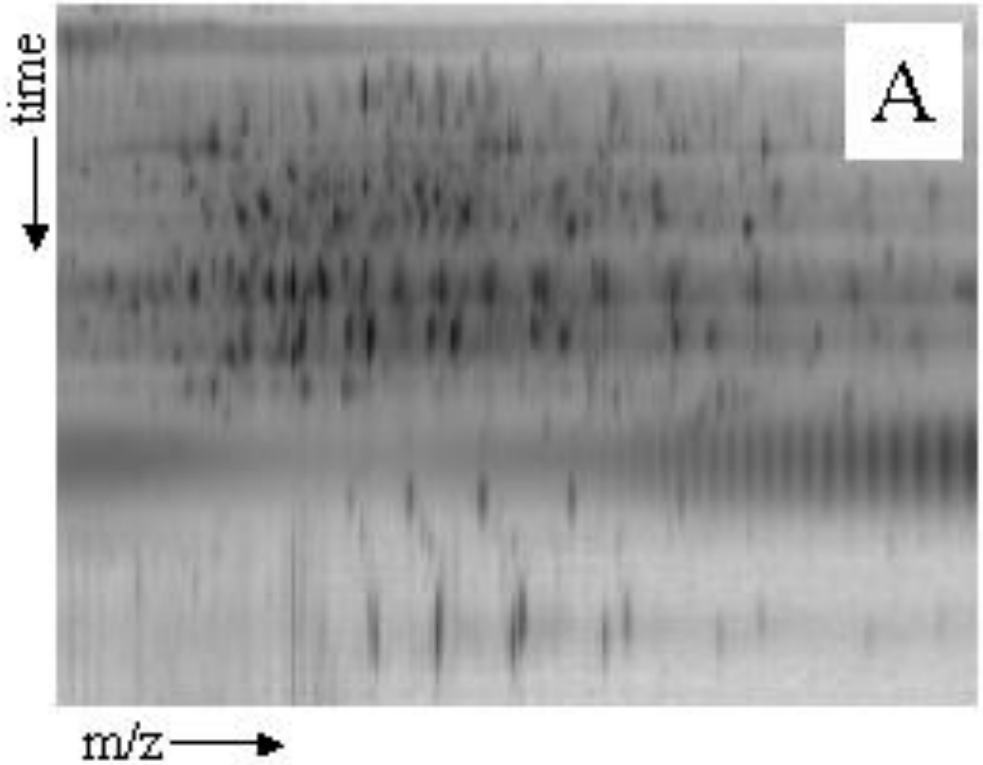
Improvements:
- Multiplexing or pipelining
- Quad filtering of high abundance components
- High Transmission

Improvements in mass spectrometry will increase our ability to find protein patterns.

Predicant BIOSCIENCES

Data Analysis Challenges

- ‘Curse of dimensionality’
- Signal v. noise
- Data points correlated
- Robust patterns



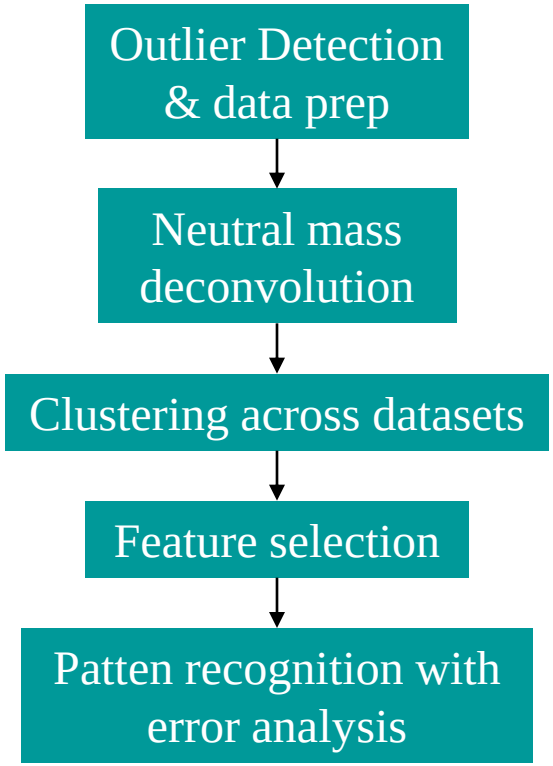
While pattern recognition is well established, data is not generic and specialized signal processing tools are needed.

Current biomarker discovery either uses each point independently or uses heuristic peak picking. Neither solution is adequate.

Predicant BIOSCIENCES

Data Analysis Pipeline

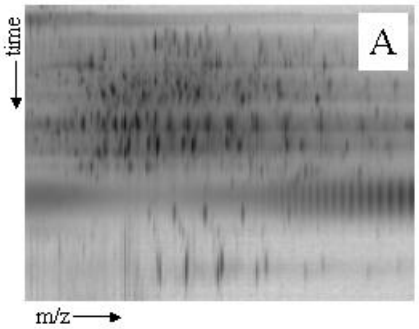
Rigorous error analysis with cross validation and false discovery rate methods increase the likelihood that meaningful patterns are discovered.



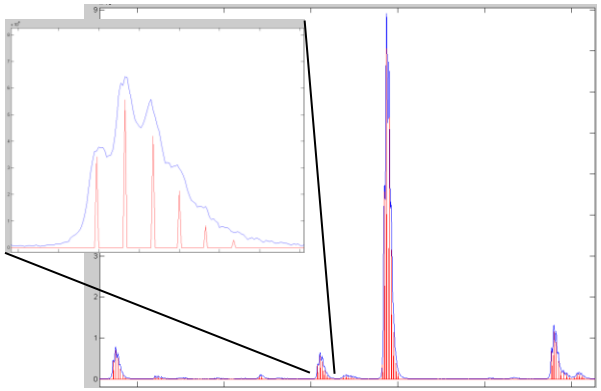
A robust data analysis pipeline can yield robust and accurate biomarker patterns.

Predicant BIOSCIENCES

Signal Processing Goals and Results



72 million data points → List of peaks: (m/z, $\bar{t}$, intensity) → List of components: (m, $\bar{t}$, intensity) → ~10^3 component intensities



Advantages:

- solves 'curse of dimensionality'
- removes noise
- collapses correlated data
- increases likelihood that patterns found are robust
- component intensities and patterns are analytically meaningful

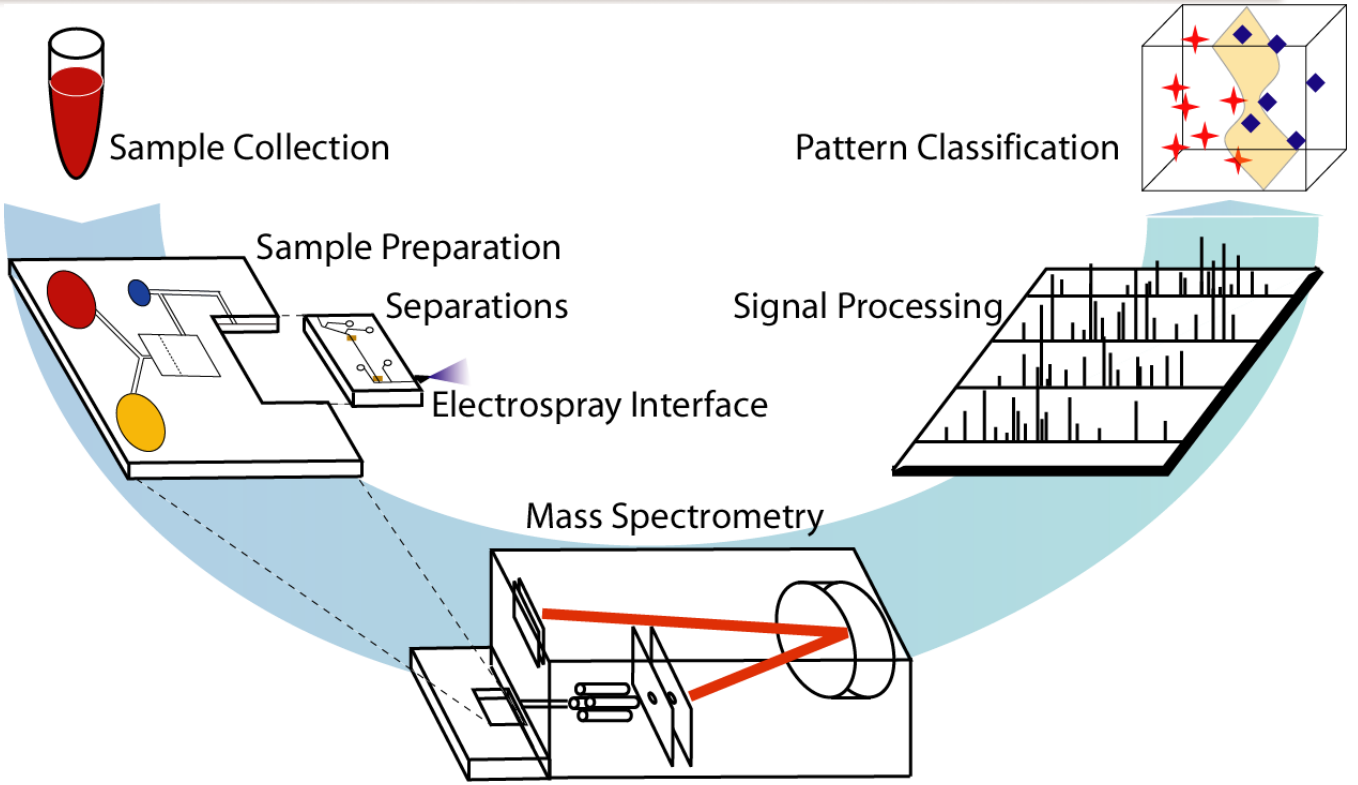
Better signal processing methods vastly improve one's ability to discover robust biomarker patterns.

Predicant BIOSCIENCES

Basis Pursuit Reference: Donoho et al., J Royal Stat Soc B (1992) v54 p41.

An Integrated System for Complex Pattern Recognition

The ideal solution is a clinically useful integrated system that includes microfluidic sample prep/separations, electrospray ionization, sensitive mass spectrometry, and specialized data analysis tools.



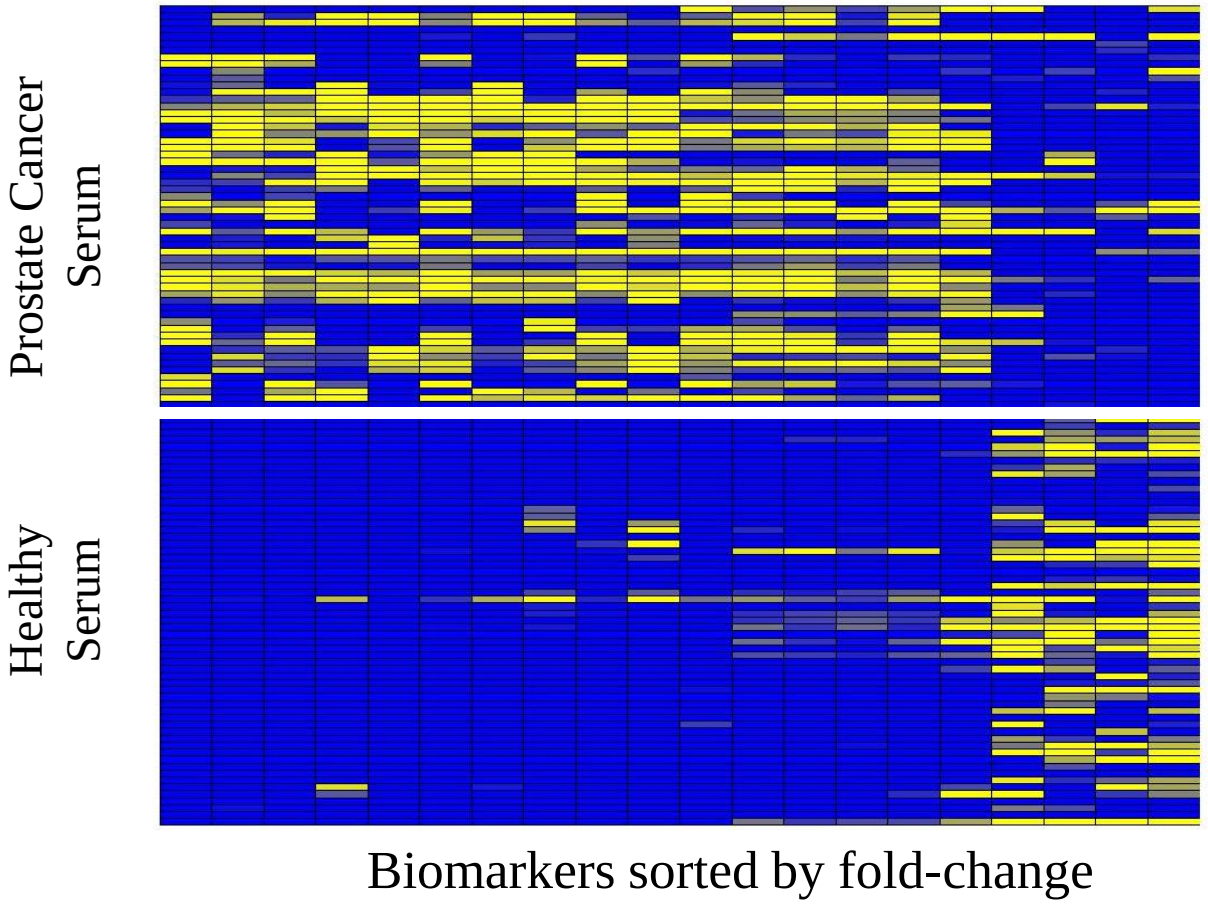
An integrated solution is vital: matched volumes & flow rates, instrument specific challenges & lineshapes, etc.

New technologies have been developed to deliver an integrated, clinically viable diagnostic system.

Predicant BIOSCIENCES

Example: Prostate Cancer Diagnosis

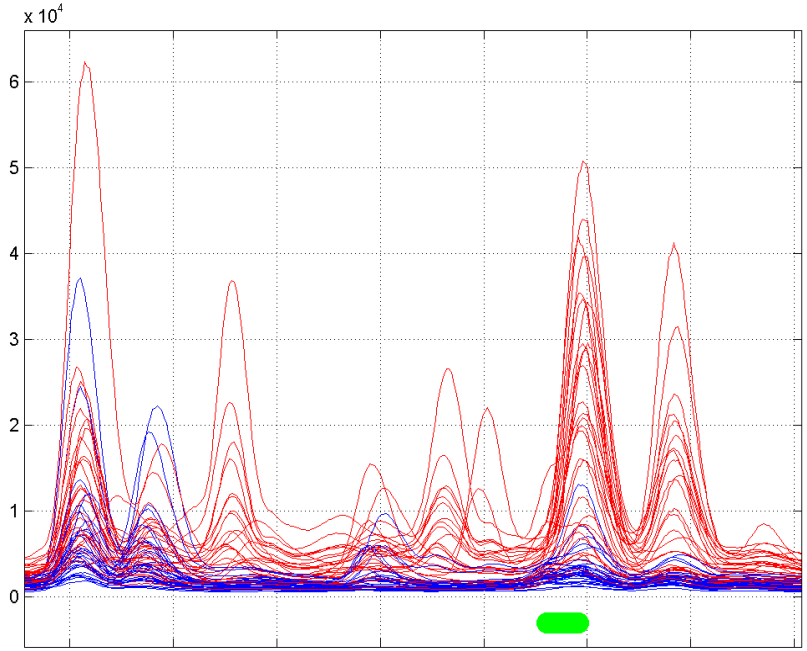
We analyzed 25 prostate cancer patients and 25 healthy controls. A pattern of 25 markers gave ~92% sensitivity and ~88% specificity. This is a much more accurate test than, e.g., PSA.



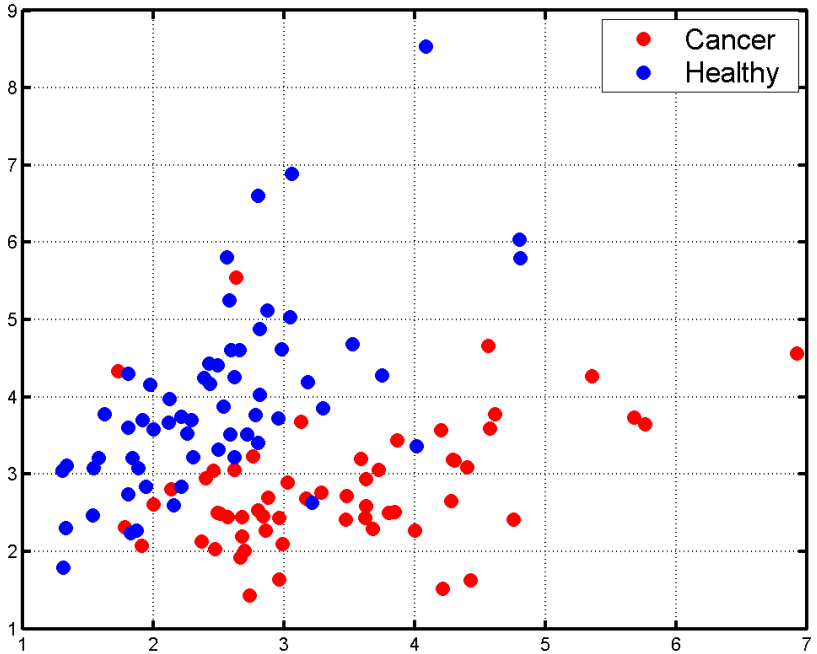
An integrated CE-ESI-MS with customized signal processing has allowed us to identify complex serum biomarker patterns in prostate cancer. Validation of results is underway.

Predicant BIOSCIENCES

Prostate Cancer Data



Cancer serum shown in red; healthy serum shown in blue; selected peak shown in green.



Two markers differentiate patients reasonably well; multi-marker diagnostics necessary.

High reproducibility is essential for both discovery and clinical application.

Predicant BIOSCIENCES

An Example Prostate Cancer 'Pattern'

Biomarker	M/Z	Separation Time (seconds)	MW	Charge State	Levels in Cancer
1	255.055				up
2	257.062	366.00	256	1	up
3	269.053	300.00	268	1	up
4	295.031	300.00	294	1	up
5	297.039	300.00	295	1	up
6	298.07				up
7	347.093				up
8	361.115				down
9	395.281				up
10	396.18				up
11	405.142	300.00			down
12	411.152				up
13	419.174				down
14	425.167	300.00	424.17	1	up
15	427.154				up
16	591.183	570.00	5901.00	10	down
17	602.083	477.00	4209	7	down
	702.342	477.00	4209	6	down
	842.768	477.00	4209	5	down
18	929.567	666.00	9287	10	down
	1032.72	666.00	9287	9	down
19	813.352	837.00	8123	10	up
	903.343	837.00	8123	9	up
	1016.15	837.00	8123	8	up
	1161.79	837.00	8123	7	up
20	614.853	474.00			up
21	810.319	513.00	13763	17	down
	918.263	513.00	13763	15	down
22	887.85	483.00	10645	12	up
	968.47	483.00	10645	11	up
	1065.31	483.00	10645	10	up
23	665.466	513.00	4655	7	up
24	698.136	432.00	4818	7	up
	813.352	432.00	4818	6	up
25	1143.86	618.00		13	up

L = (NBI_1 * 0.00287486039568013 + NBI_2 * 0.0152718769589517 + NBI_3 * 0.00209473797462831 + NBI_4 * 0.000787430789421225 + NBI_5 * 0.0207066586062057 + NBI_6 * -0.0219823090617778 + NBI_7 * 0.014879856480702 + NBI_8 * -0.0271510213710407 + NBI_9 * 0.00298210732138475 + NBI_10 * -0.00604246370468643 + NBI_11 * 0.0336368775785866 + NBI_12 * 0.0237998629569305 + NBI_13 * -0.0174155959784781 + NBI_14 * 0.00312940149063507 + NBI_15 * -0.00453966244716806 + NBI_16 * -0.014386456060012 + NBI_17 * -0.00742832069108851 + NBI_18 * -0.0143882694420193 + NBI_19 * -0.00485154255850317 + NBI_20 * 0.000702160371199416 + NBI_21 * -0.0177099514748433 + NBI_22 * 0.00317406284571304 + NBI_23 * 0.0119606323436643 + NBI_24 * 0.0071983115720181 + NBI_25 * 0.0293623597760145) - 0.00127399357851983

Where NBI_1 is the intensity of biomarker 1 after normalization and other pre-processing steps.

If L > 0, then prostate cancer; if L < 0, then healthy.

A SVM analysis of our prostate cancer data resulted in a diagnostic pattern much more accurate than current tests like PSA.

Predicant BIOSCIENCES

Applications

Disease areas of high interest:

- cancer (lung, breast, prostate, bladder)
- stenosis
- rheumatoid arthritis
- pre-eclampsia

Specific applications within those disease areas:

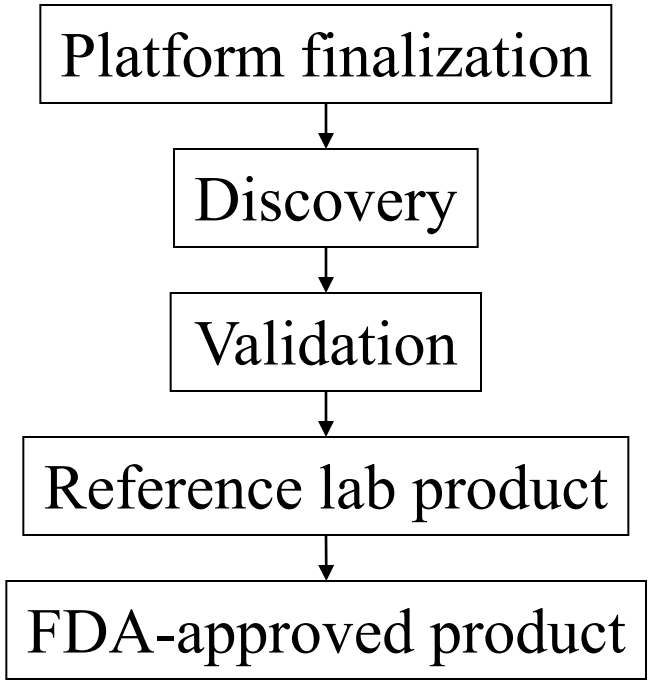
- adjunct to existing screening test (e.g. follow-on to mammography)
- drug response prediction (e.g. anti-TNF therapy – Enbrel, Remicade, Humira)
- early screening in high risk populations (e.g. lung cancer screening in heavy smokers)

Several applications have been selected as our first entry points. We are in the process of running discovery and validation studies .

Predicant BIOSCIENCES

Product Introduction through a Reference Lab

We expect to have a first product available by early ’06. Using the same platform, follow-on products will be available quickly.



Introducing our product through a government-regulated reference lab will speed our time to market. Proper validation studies are essential for this approach.

Predicant BIOSCIENCES

The Vision

Physicians seek a clinically reliable method of detecting, diagnosing and monitoring complex disease states.

A revolutionary class of analyses is being developed to transform clinical patient care based upon the identification of protein patterns in serum that determine and differentiate biological state. This offers new hope for improved disease outcomes.

A novel microfluidic & mass spectrometry based platform has the ability to transform disease management.

Predicant BIOSCIENCES