

**Permutation Achieved Classification Error  
(PACE): A Significance Test for Assessing the  
Statistical Significance of Classifiers for Peptide  
Profiling via MALDI/SELDI-TOF Mass  
Spectrometry**

The “Pitt Crew”:

James Lyons-Weiler,

Milos Hauskrecht, Richard Pelikan,

David Malehorn, Doug Landsittel, Herb Zeh,

David C. Whitcomb, Bill Bigbee

**Benedum Oncology Informatics Center**

**Computer Science Department**

**Clinical Proteomics Facility**

**University of Pittsburgh Cancer Institute**

EDRN Supplemental Funding UO1 CA84968-04S1

BOIC Faculty Recruitment Funds (JLW)

CS Department Faculty Recruitment Funds (MH)

# Outline

---

- PACE Analysis
- Pancreatic Cancer Preliminary Classifier
- Negative Results
- Published Data
- Caveats
- Biological Significance
- Parting Shots

## Published SN and SP from SELDI-TOF-MS Studies

---

- Ovarian Cancer: 100%, 95% (Liotta et al., 2002)
- Prostate Cancer: 100%, 100% (Qu et al., 2003)
- Breast Cancer: 90%, 93% (Vlahou et al., 2003)
- Breast Cancer: 91%, 93% (Li et al., 2002)
- Head & Neck: 83.3%, 90% (Wadsworth et al., 2004)
- Lung Cancer: 93.3%, 96.7% (Xiao et al., 2003)
- Pancreatic Cancer: 78%, 97% (Koopmann et al., 2004)

# Are these REAL?

- Diamandis
- Baggerly et al.
- New York Times...
- Etc...

# Dr. Diamandis

- Diamandis EP. 2004. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst. 96(5):353-6.
- Diamandis EP. 2004. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: Opportunities and potential limitations. Mol Cell Proteomics. 2004 Feb 28. [Epub ahead of reprint]
- Diamandis EP. 2004. Re: diagnostic potential of serum proteomic patterns in prostate cancer. J Urol. 171(3):1244-5; author reply 124-5-64.
- Diamandis EP. 2004. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: Opportunities and potential limitations. Mol Cell Proteomics. 2004 Jan 30 [Epub ahead of print]

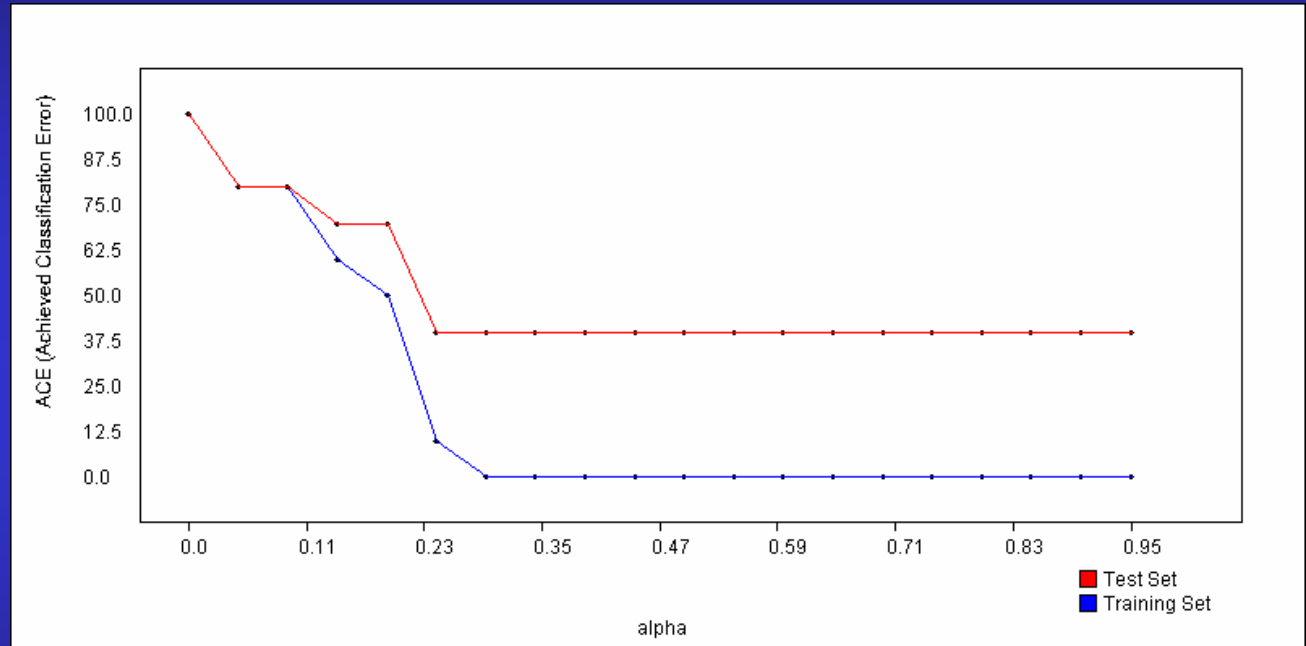
# Practical significance of some of these concerns

- Overtraining (model overfit)
- Biased estimate of Achieved Classification Error (ACE)
- Low External Generalizability of Model

# What is Overtraining?

ACE

Random  
Resampling  
 $N1=N2=8$   
5000 random  
features



stringency of feature selection method

# Bias in ACE

- Achieved Classification Error = the proportion of cases that are incorrectly predicted.
- High biased estimates can be either too optimistic *or* pessimistic (conservative)
- Use of TEST cases protects against this bias
- Various learning evaluation schemes provide low-biased estimates of ACE in general population:
  - Leave-one-out Validation
  - *k*-fold Validation
  - Hold-out *m*% method
  - Random Resampling hold-out method



# PACE: Permutation Achieved Classification Error

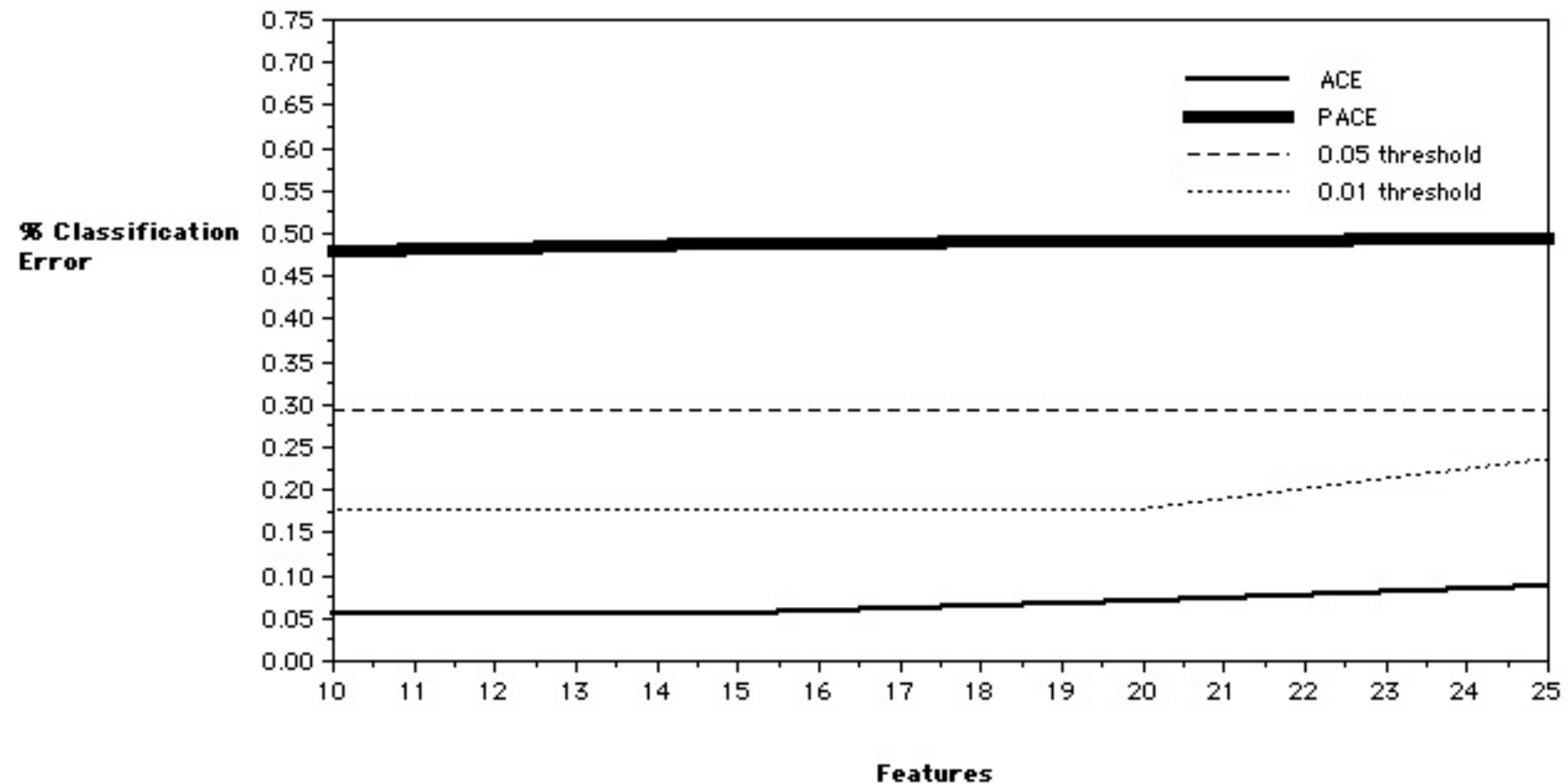
- Expected classification error under the null model (one homogeneous population, stochastic variation in profiles reflect random differences only, no signal)
- This is 50% (a coin toss) in case vs. control studies

# Approach

- Determine ACE using a method  $x$
- Create 1,000 null data sets via random sample class labeling
- Perform analysis on all 1,000 null data sets
- Determine Mean Achieved Classification Error (MACE) and
- 95th, 99th Percentile of the PACE distribution
- Compare ACE to PACE distribution

# Pancreatic Cancer Result

Achieved Classification Error, Permutation Achieved Classification, 95th and 99th PACE percentile



# Adenocarcinoma of the Pancreas

---

- 30,300 cases in the U.S. in 2002
- Ninth most common cancer but...**fourth** leading cause of cancer deaths (5-6 % of all cancer-related deaths)
- Overall 5-year survival is <5%
- For the minority of patients with resectable disease 5-year survival is ~20%, however 80% of these patients will recur within 2 years and die of their disease

# Survival after Resection of Adenocarcinoma of the Pancreas

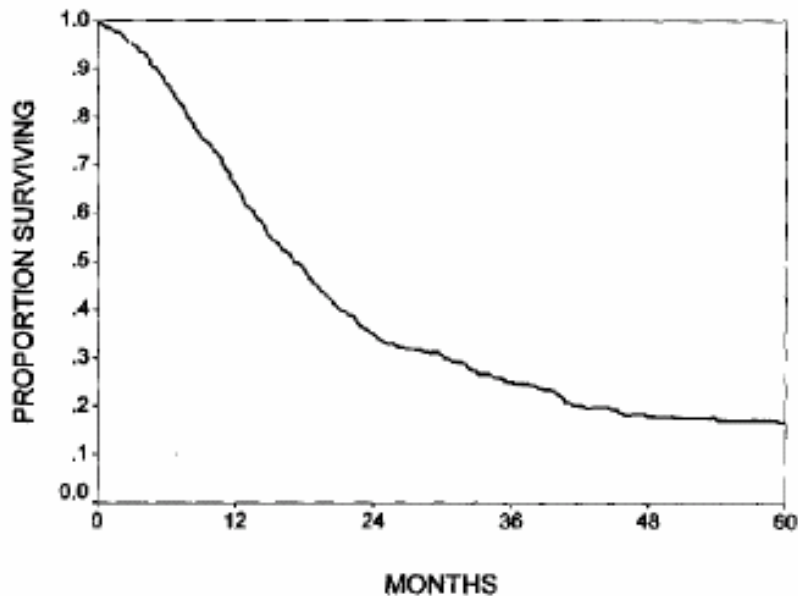
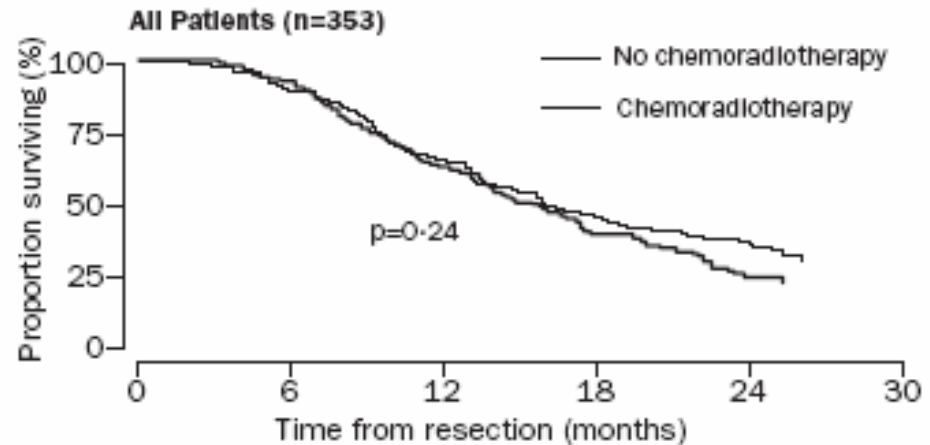


Fig. 1. Kaplan-Meier actuarial survival for entire cohort of 616 patients with resected pancreatic adenocarcinoma.

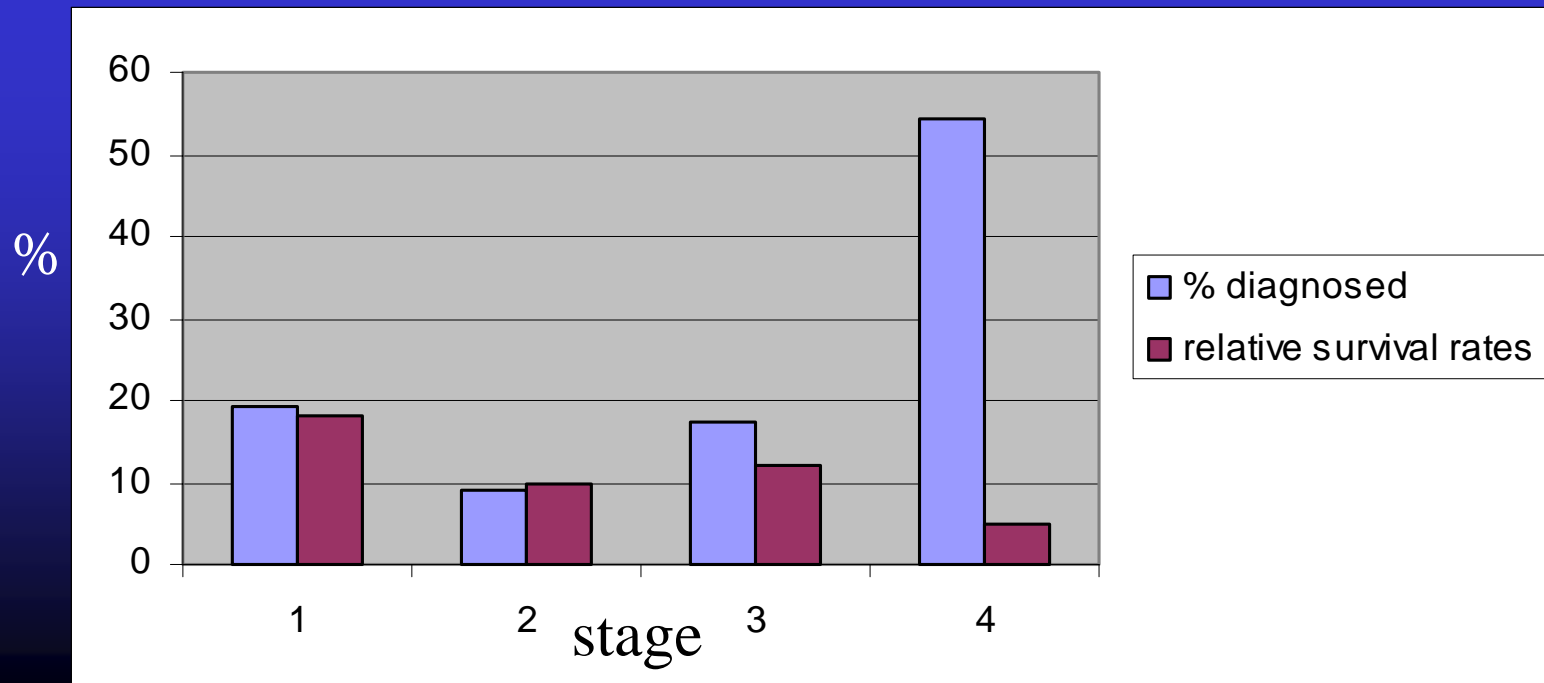


Neoptolemos et al., 2001 *Lancet*

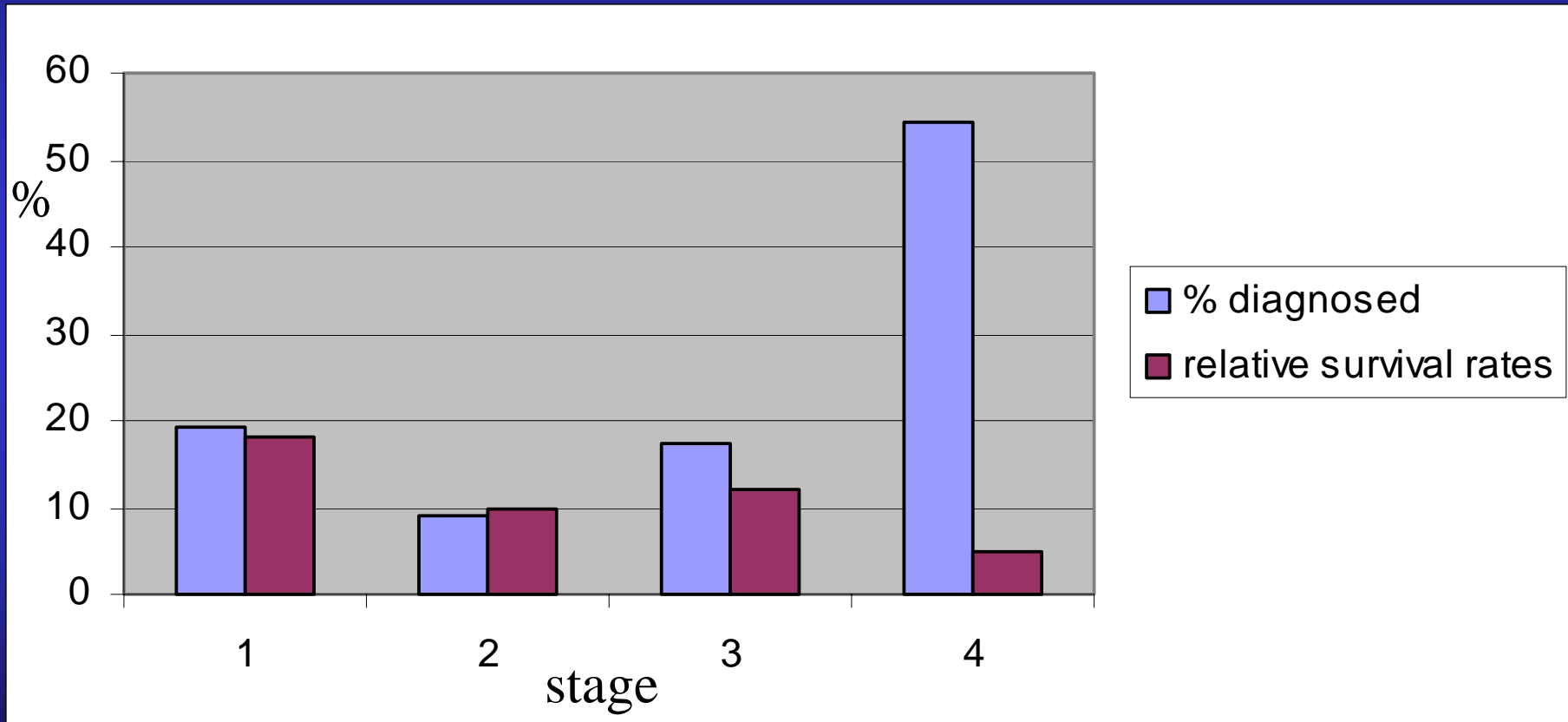
Sohn et al Resected adenocarcinoma of the pancreas-616 patients: results, outcomes, and prognostic indicators. *J Gastrointestinal Surgery* 4: 567-579, 2000.

# Neoptolemos JP, et al. Ann Oncol. 2003 May;14(5):675-92.

- Overall median survival from diagnosis < 3-5 mos
- 12-mo survival rate of ~ 10%
- 5 – year survival rate of 0.4-3%



# Pancreatic Cancer Stage at Diagnosis and Survival



# Retrospective data support the concept that early detection and treatment will lead to improved survival in patients with pancreatic cancer

---

- Patients with smaller tumors had better overall survival,  $< 3.0$  cm ( $P = 0.004$ )
- Survival of a select group ( $N = 75$ ) of patients with small tumors ( $< 3.0$  cm), negative margins, and negative lymph nodes:
  - 1 Y - 81%
  - 3 Y - 46%
  - 5 Y - 31%



## Ariyama et al 1990. *Int J Pancreatol.* 7:37-47.

- Reported 100% 5-year survival in patients undergoing resection of pancreatic tumors detected very early (<1 cm)

Risk of cancer	Specificity/Sensitivity	Years gained
5%	0.80	0.11
10%	0.80	0.49
20%	0.80	1.24
30%	0.80	1.98
5%	0.90	0.28
10%	0.90	0.69
20%	0.90	1.52
30%	0.90	2.34

Kern S,H.R.H.M. 2001. **A white paper: the product of a pancreas cancer think tank.**  
*Cancer Res* 61:4923-4932.

# Available Serum Tumor Biomarkers for Pancreatic Cancer

---

---

- CA19-9 levels are correlated with the presence of pancreatic cancer
  - Sensitivity and specificity for CA19-9 between 70-90% and 90%, respectively
  - Unfortunately, there is a high degree of overlap between pancreatic cancer and a variety of benign inflammatory conditions of the pancreas, limiting the clinical applicability of CA19-9 as a specific early detection/screening marker
- Tissue polypeptide specific antigen (TPS) (a breakdown product of the extracellular matrix) is another recently described serum tumor marker that has been reported to provide improved discrimination between pancreatitis and pancreatic cancer over CA19-9
  - Sensitivity and specificity of 97% and 98%, respectively vs versus normal controls
  - However, TPS was also found to be elevated in 17-20% of patients with chronic pancreatitis

# Study Design and Subject Accrual

---

- **Study Design and Target Accrual**
  - 30 newly-diagnosed pancreatic cancer cases,
  - 30 case-matched healthy controls,
  - 30 chronic pancreatitis
  
- **Currently Accrued**
  - 120 cases since October 2002,
  - 80 pancreatic and 40 non pancreatic “Whipple controls”,
  - Over 100 acute and chronic pancreatitis,
  - Over 100 matched controls

# Patients and Controls (see Herb)

- Preoperative serum samples from 32 pancreatic cancer cases (17 female, 15 male)
- Patient samples were obtained in the operative room using fast-stabilizing protocols to protect against trypsin (Whitcomb, MD, PhD)
- 23 non-cancer age-, gender-, and smoking history-matched controls were analyzed.
- Ages ranges 34-87; mean age PCa 64; controls 67 (p=0.19)
- 16 were resected; 6 patients had locally advanced unresectable disease, 10 had metastatic disease.

# SELDI-TOF-MS Profiling (see Dave)

- The serum samples were denatured and processed in duplicate on a single type (IMAC3-Cu) ProteinChip® Array (CIPHERGEN Biosystems, Fremont CA).
- Samples were processed in random order to avoid confounding sample type with temporally autocorrelated laboratory conditions
- Biomek2000 liquid-handling robotic workstation (Beckman Instruments, Inc., Fullerton CA).
- Whole serum samples ('neat spotting')
- ProteinChips were read in a PBSIIc mass spectrometer (CIPHERGEN) using positive ion mode, with time delay focusing, from 0-100 kDa.
- Mass calibration was performed externally, using a mixture of seven peptide species from 1-7 kDa (CIPHERGEN).
- Spectra were preprocessed by baseline subtraction with smoothing; filtering (averaging, by 0.2x expected peak width) and normalized by total ion current.

# Analysis

- CART Analysis (CIPHERGEN BPS software; Dave)
- MatLab functions developed by Milos Hauskrecht, Richard Pelikan, CS Department, JL-W
- Feature Selection
  - Fisher Score, w, w/out Decorrelation
  - Principal Components Analysis (aggregate feature)
- Classification
  - Naïve Bayes
  - Support Vector Machine (SVM)

# CART

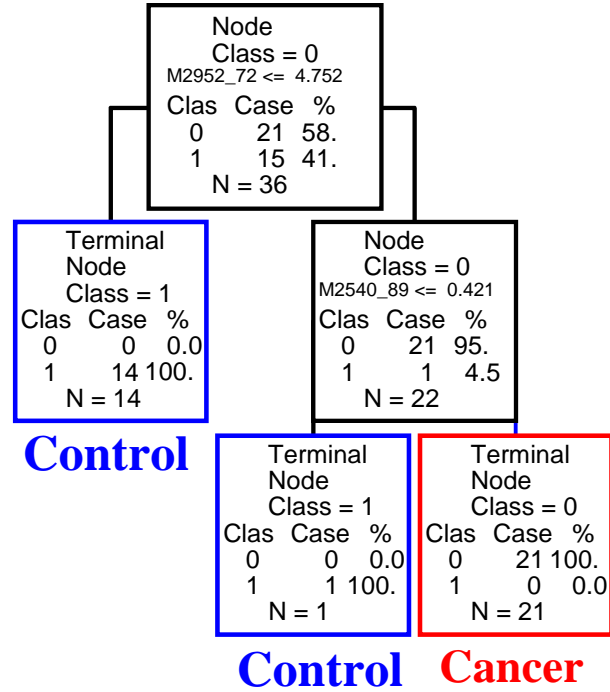
---

- ‘training set’ of 21 cases/15 controls,
- ‘testing set’ of 11 cases/8 controls.
  
- 64 most significant peaks from a pair-wise comparison using CART analysis (CIPHERGEN Biomarker Patterns<sup>®</sup> Software).



# SELDI-TOF MS Serum Profiling Discriminates Pancreatic Cancer Cases from Matched Healthy Controls

**2-Variable Classification Tree  
 Pancreatic Cancer vs Control**



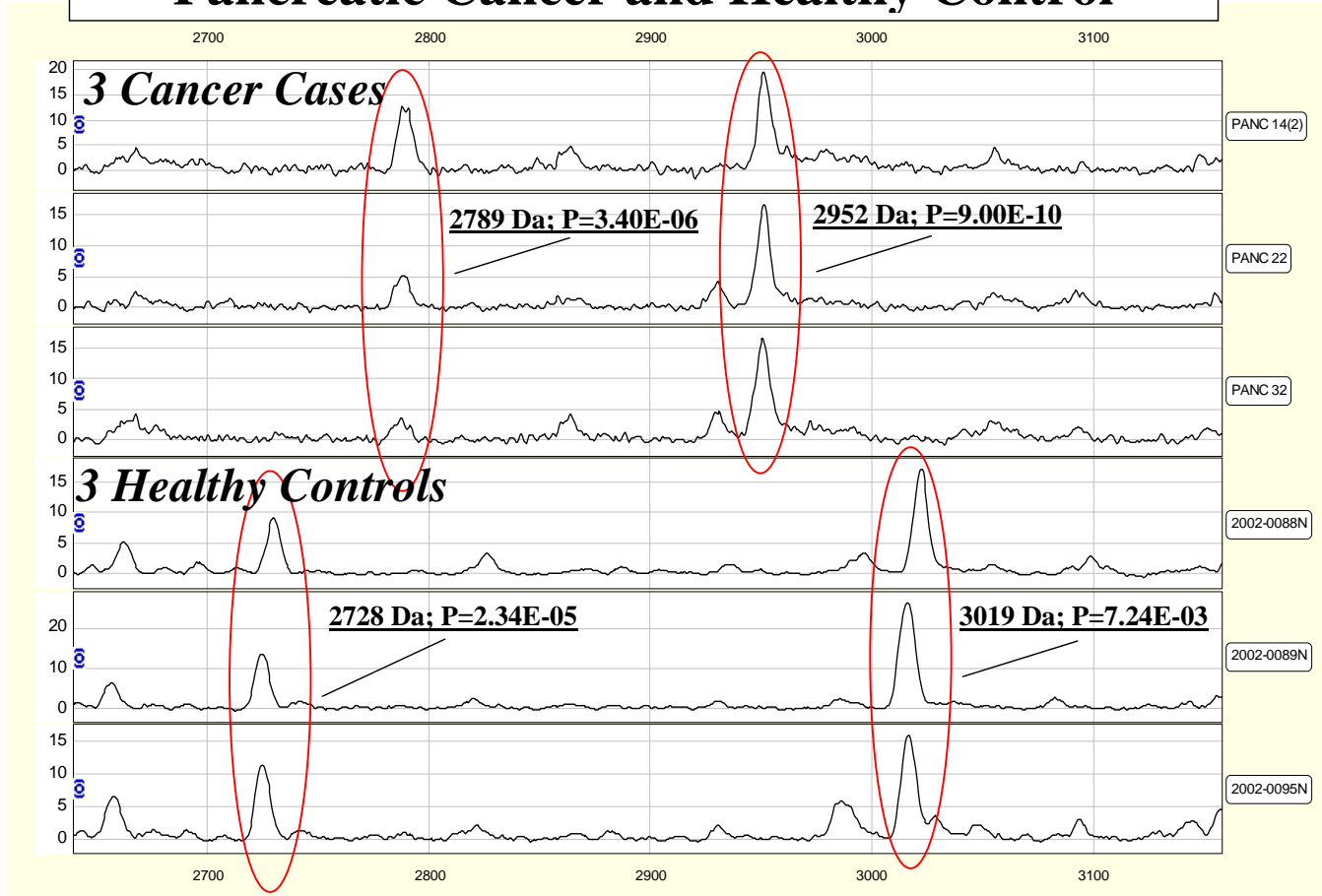
<b><u>Training Data:</u></b>	<b><u>Testing Data:</u></b>
<b>100% Sens (21/21)</b>	<b>91% Sens (10/11)</b>
<b>100% Spec (15/15)</b>	<b>100% Spec (8/8)</b>





# SELDI-TOF MS Serum Profiling Discriminates Pancreatic Cancer Cases from Matched Healthy Controls

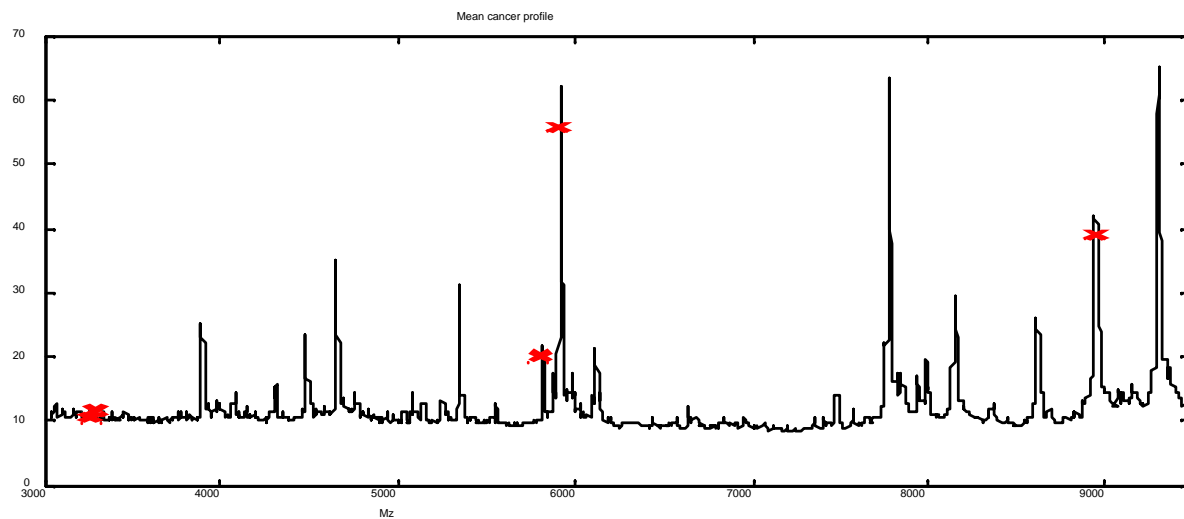
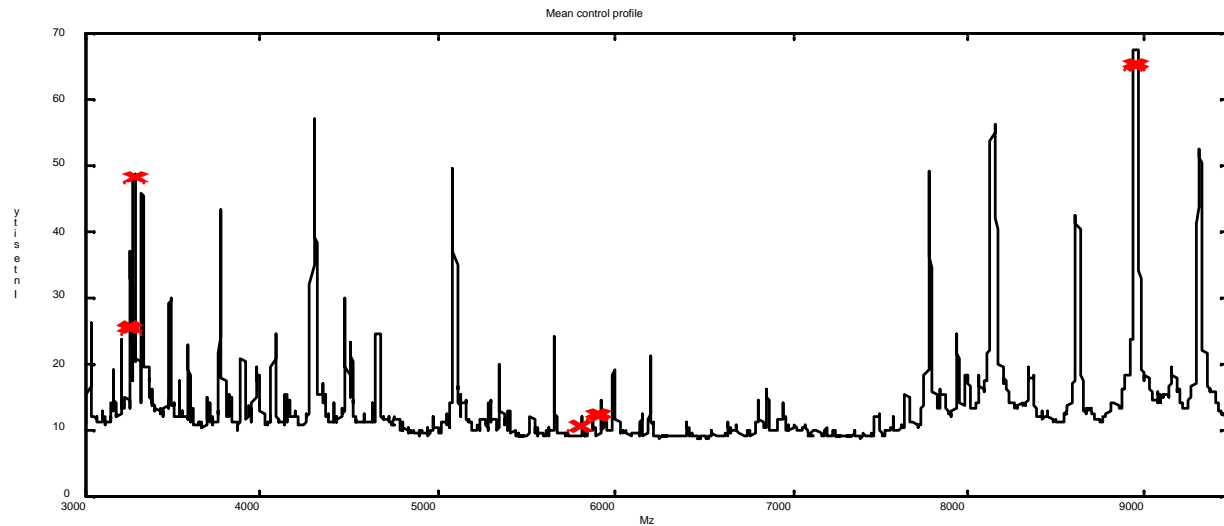
**Positive and Negative Differences Between Pancreatic Cancer and Healthy Control**



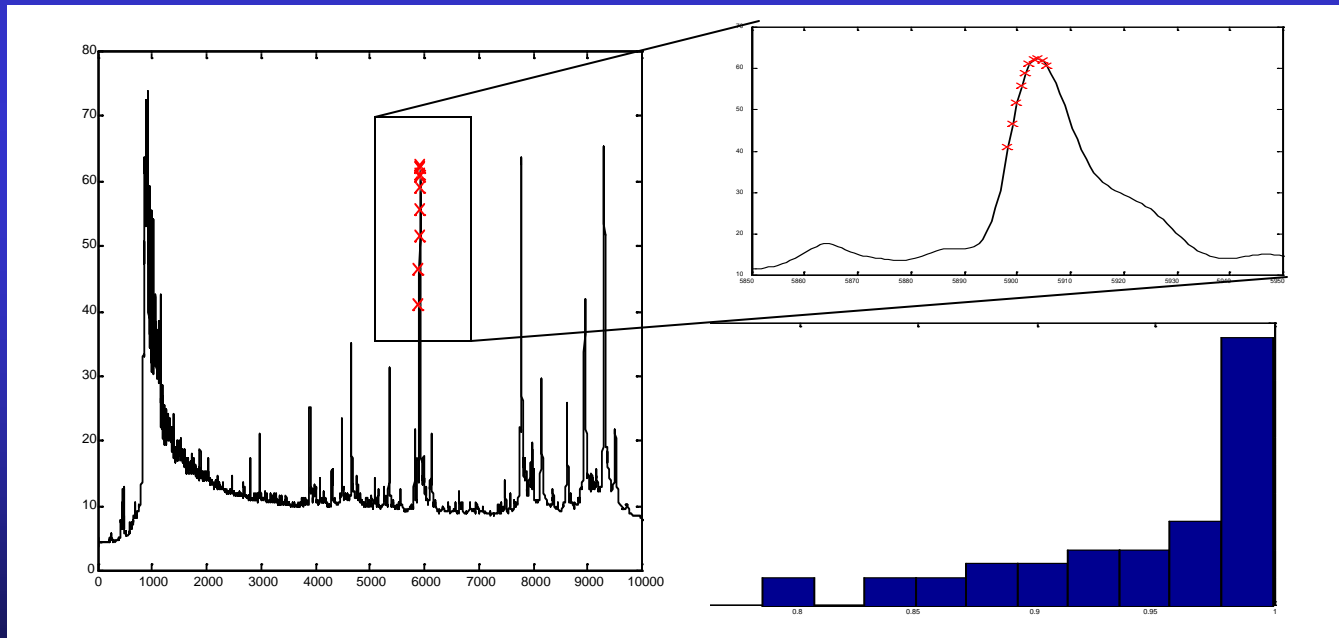
# Supervised Learning

- Post-Feature Identification
- Find Informative Features
  - “Differentially expressed features”
    - t-test, modified t-test (e.g., SAM), AUC, Fisher score
    - Decorrelation filtering (Décor)
  - Aggregate Features (PCA)
- Predict Class Labels
  - Classification predictions via Naïve Bayes or linear SVM
- See Milos, Richard or poster for details...

# Mean Cancer and Control Profiles (Top 5 Fisher Score Positions)



- Decorrelation after feature ranking is similar to PCA
- Useful because the most significant features are not independent
- Likely due to mass drift among the profiles
- Weakens the apparent statistical significance of the single feature
- Loss of correlation among features as a biomarker?**



PEAK	# Features	SN	SP	Test Error
No	5	0.8697	0.669	0.2132
No	10	0.9073	0.9039	0.0941
No	15	0.9298	0.9537	0.0603
No	20	0.9499	0.9502	0.05
Yes	5	0.9123	0.726	0.1647
Yes	10	0.9424	0.9644	0.0485
Yes	15	0.9474	0.968	0.0441
Yes	20	0.9348	0.9751	0.0485

Risk of cancer	Specificity/Sensitivity	Years gained
5%	0.80	0.11
10%	0.80	0.49
20%	0.80	1.24
30%	0.80	1.98
5%	0.90	0.28
10%	0.90	0.69
20%	0.90	1.52
30%	0.90	2.34

Kern S,H.R.H.M. 2001. A white paper: the product of a pancreas cancer think tank.

*Cancer Res* 61:4923-4932.

$$p(\text{cancer}|\text{biomarker}) = p(\text{cancer}) \frac{p(\text{biomarker}|\text{cancer})}{p(\text{biomarker})}$$

$$p(\text{cancer}|\text{biomarker}) = p(\text{cancer}) \frac{p(\text{biomarker}|\text{cancer})}{p(\text{cancer})p(\text{biomarker}|\text{cancer}) + p(\text{cancer})p(\text{biomarker}|\text{cancer})}$$

$$p(\text{cancer}|\text{biomarker}) = \text{prior} \frac{\text{sensitivity}}{(\text{prior})(\text{sensitivity}) + (1-\text{prior})(1-\text{specificity})}$$

Using prevalence = 0.008 as the prior,

Equations derived from Alan Schwartz, UIC  
**Medical Decision Making, MHPE 494, Spring 1999**  
<http://www.sjdm.org/~alansz/courses/mhpe494/week3.html>

marker	SN	SP	expected false alarms/100k	Reference
<b>invasive techniques (biopsy)</b>				
MUC-1 expression (patients w/masses)	0.96	0.94	5995	Chheng et al., 2003
cytology	0.63	1	0	Mu et al., 2003
p53 expression	0.44	0.8	19992	Mu et al., 2003
cytology + p53	0.78	1	0	Mu et al., 2003
cytology + CA19.9	0.67	0.8	19989	Mu et al., 2003
cytology + p53 +CA19.9	0.78	0.8	19987	Mu et al., 2003
EUS-FNA	0.843	0.97	2997	Eloubadie et al., 2003
k-Ras mutations	0.27	1	0	Castells et al., 1999
<b>minimally invasive (serum or circulating DNA)</b>				
CA 242 in serum	0.75	0.855	14491	Ozkan et al., 2003
CA 19-9 in serum	0.8	0.675	32476	Ozkan et al., 2003
CEA in serum	0.4	0.73	26991	Ozkan et al., 2003
k-Ras mutations	0.27	1	0	Castells et al., 1999
k-Ras2 mutations	0.46	0.87	12995	Maire et al., 2002

<b>SELDI-TOF-MS SVM+F (overall classifier)</b>	<b>0.95</b>	<b>0.94</b>	<b>5995</b>	<b>this study**</b>
<b>SELDI-TOF-MS SVM+PCA (overall classifier)</b>	<b>0.9348</b>	<b>0.9751</b>	<b>2488</b>	<b>this study**</b>

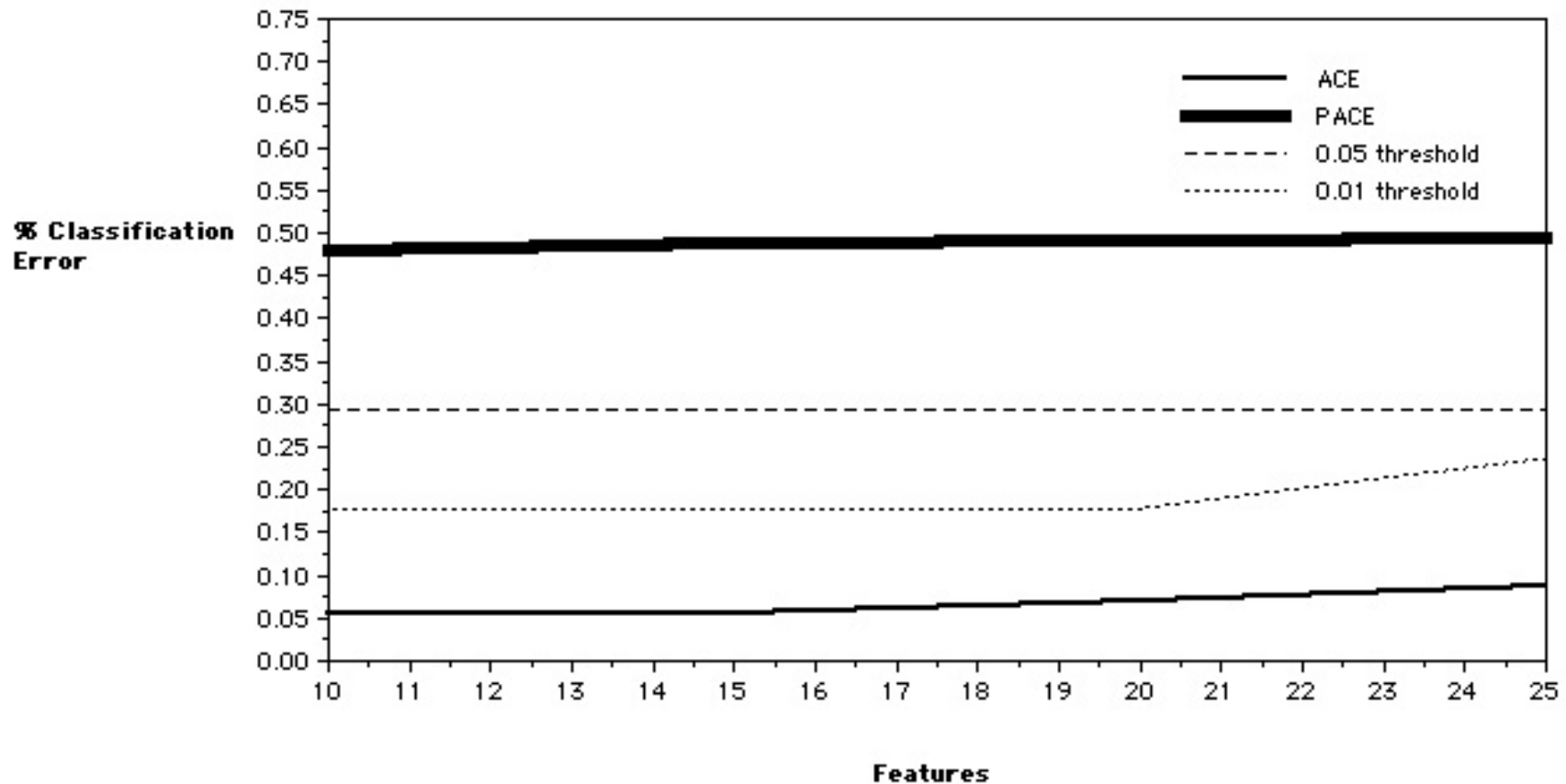
SVM = support vector machine

F = Fisher-score selected features

PCA = principal-component selected features

# Preliminary Pancreatic Cancer Result

Achieved Classification Error, Permutation Achieved Classification, 95th and 99th PACE percentile



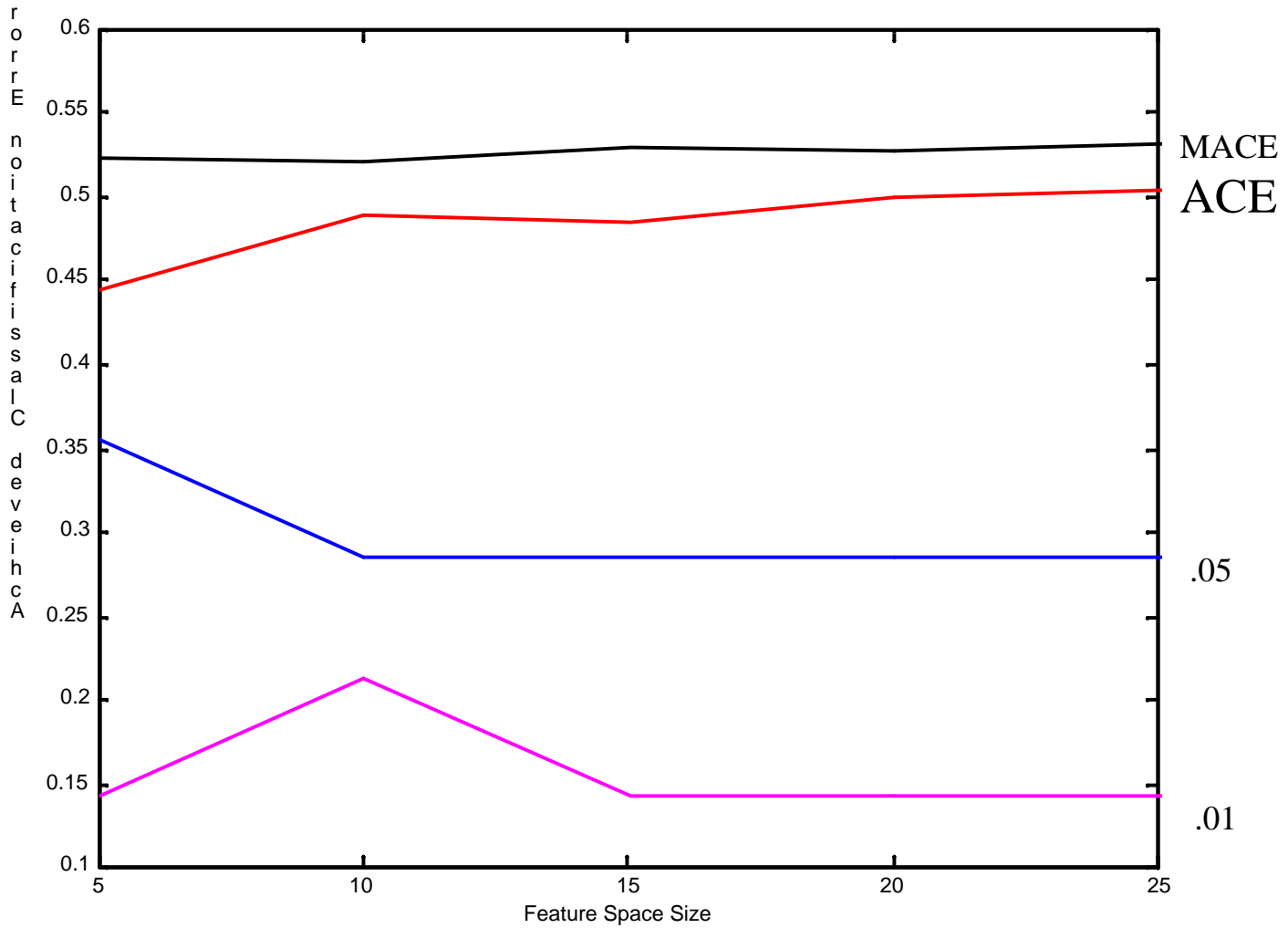


*I have a statistical basis  
for confidence...*

- *Yea, though I walk through the  
valley of the shadow of death, I  
will fear no evil*

# *Much more is at stake than our careers*

- *Be concerned about patient care*
- *Errors in either direction (overly optimistic or overly pessimistic) could be costly.*



AUC, No Decorr, Naive Bayes

# Reason for No Significance?

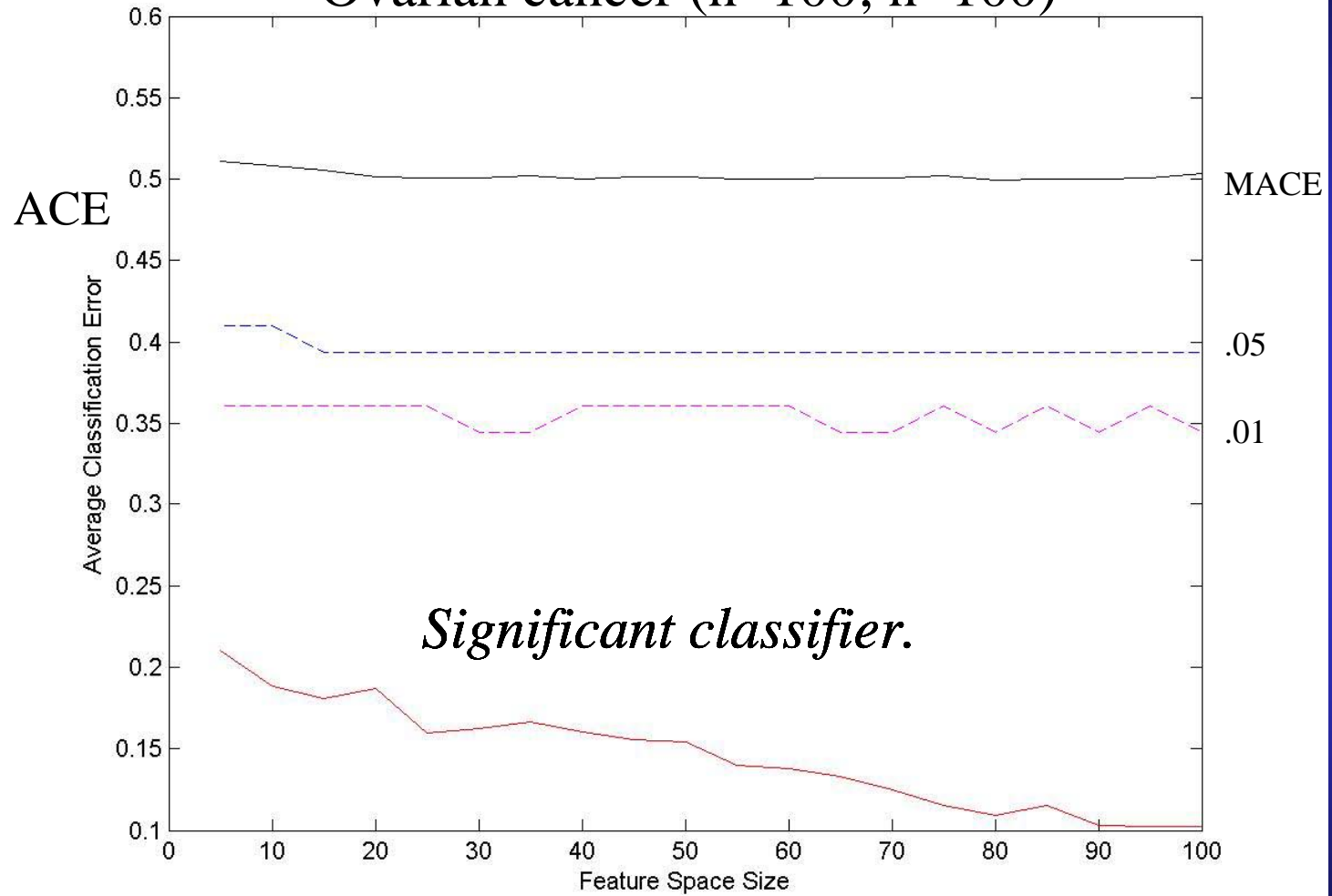
- No biological signal
- Poor study design
- Need for improved technology
- Weak signal (insufficient samples, power of the test)

# What of Ovarian Cancer?

Use of proteomic patterns in serum to  
identify ovarian cancer.

Lancet. 2002 Feb 16;359(9306):572-7.

# Ovarian cancer (n=100; n=100)



*Significant classifier.*

SVM + PCA

# Claim, Conjecture, or Tautology?

- Perform methods  $x_1, x_2, x_3 \dots x_n$
- Which ever steps in any method reduce ACE are not helpful
- Which ever method(s) exhibit significant ACE are justified
- Which ever methods achieve lowest ACE is preferred
- Different features identified by different methods w/significant ACE are all interesting

## What about your SOP's?

- Could PACE be used to optimize laboratory protocols?
- Could PACE be used to optimize preprocessing?
- Could PACE be used to optimize analysis strategies?



# Caveats

- PACE **does not** protect against improper study design (e.g., confounding)
- Signal detected is tentative and requires
  - (a) evaluation of classifier with new, unseen samples
  - (b) replication of the study\*
  - (c) identification of peptide complexes
- Ultimate validation derives from shift in the proportion of cases detected at earlier stages and increases in survivorship.

# Biological vs. Statistical Significance



# Universal Biomarker Myth

- Each tumor is a unique biological event
- “Cancer” often includes various stages
- Unique genes/proteins may be lost in each patient
- Patient classes may be unique to genes (this is not merely disease subclasses)
- Proportion of samples in which a feature is informative
- Biomarker Panels:
  - Not protein 1 AND protein 2 AND protein 3... $N$
  - Protein 1 OR 2 OR 3... $N$
- Accommodate Individualized Medicine

# Biological vs. Statistical Significance 2: Systematic Juvenile Rheumatoid Arthritis

- Raphael Hirsch
- Takako Miyamae
- Shumpei Yokota
- Bonnie Lemster
- David Malehorn
- Bill Bigbee

Please contact Dr. Hirsch

# Required Reading

- Mehta AI, Ross S, Lowenthal MS, Fusaro V, Fishman DA, Petricoin EF 3rd, Liotta LA. Biomarker amplification by serum carrier protein binding. *Dis Markers*. 2003-2004;19(1):1-10.

# Open Areas

- (1) Why Peaks Only?
- (2) Workflow/Protocol Optimization

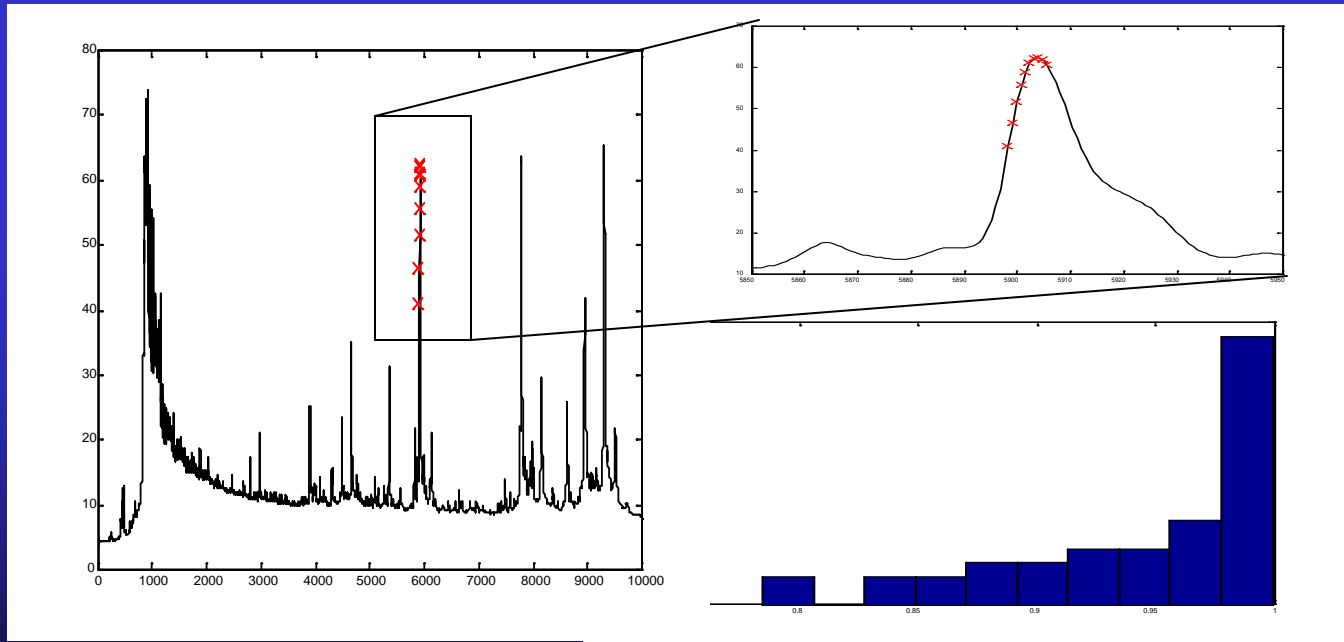
# Peaks Etc?

- Peaks (Directed Approach)
- Troughs (Directed Approach)
- Slopes (Directed Approach)
- Binning (Partially Directed Approach)
- Whole-Profile Analysis (Non-directed Approach)
- Anti-Peaks (All  $m/z$  values - peaks) (Anti-directed Approach)
  
- Which one has more information?

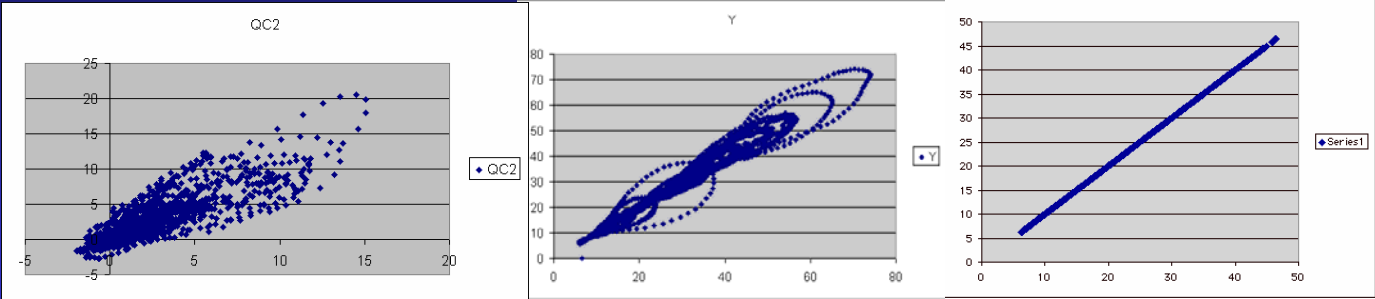


# Open Areas

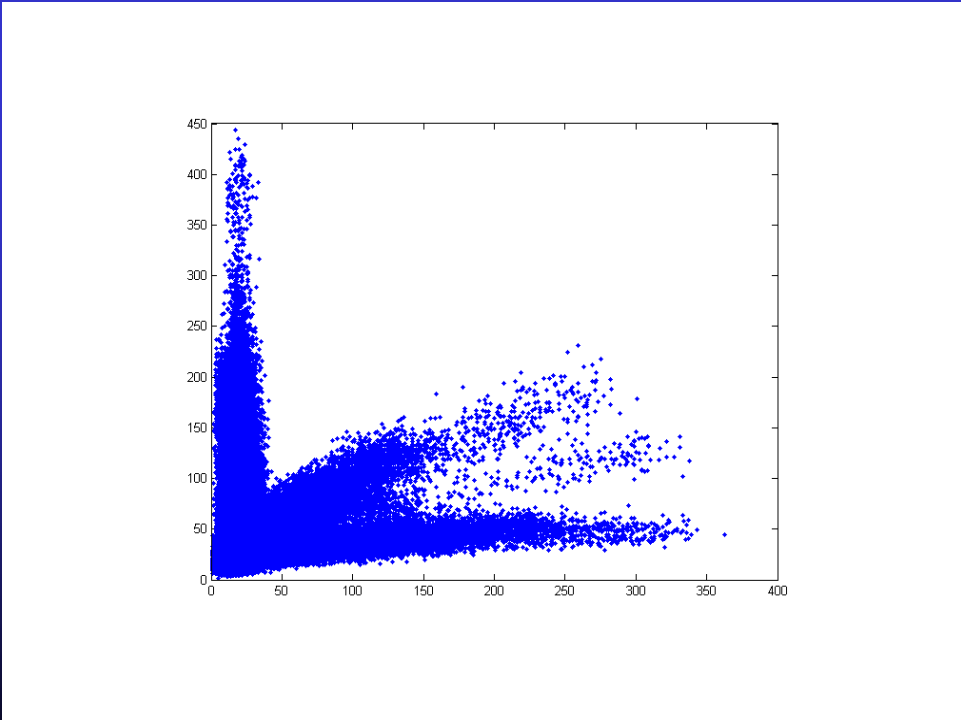
- (1) Why Peaks Only?
- (2) Workflow/Protocol Optimization
- (3) Profile alignment/mass calibration+

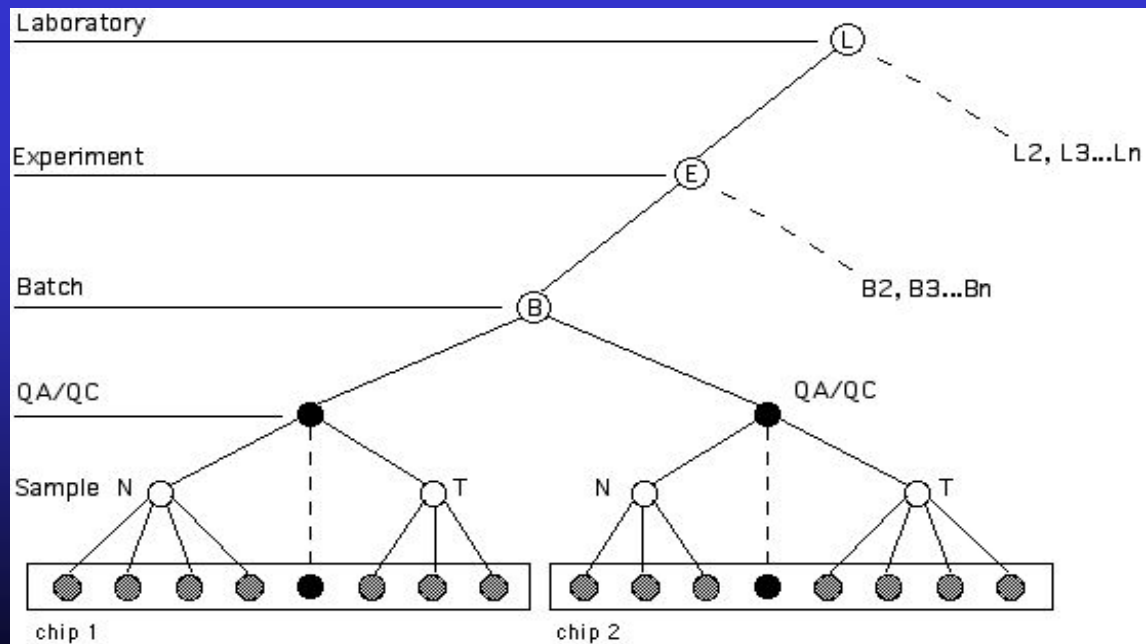
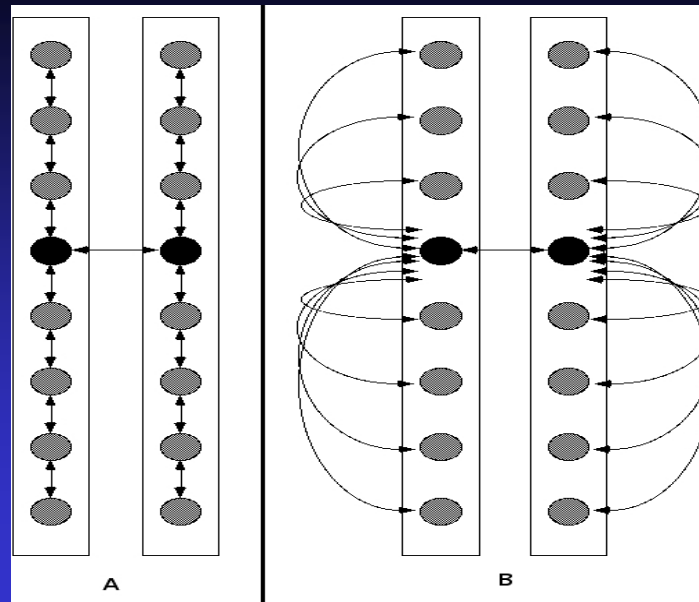


# Mass Drift as Seen in Pooled Reference Serum



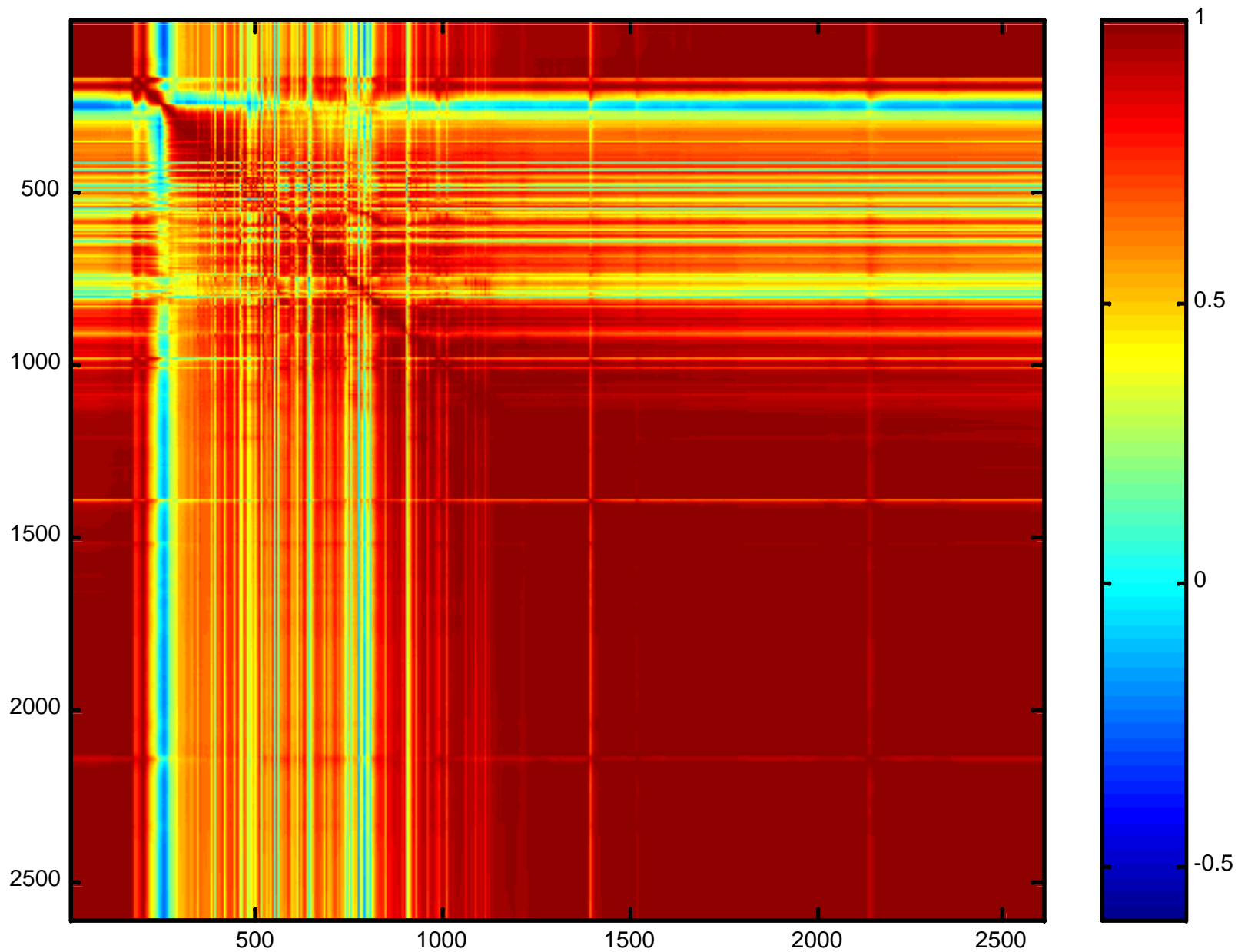
Samples  
on  
 $Q^*$

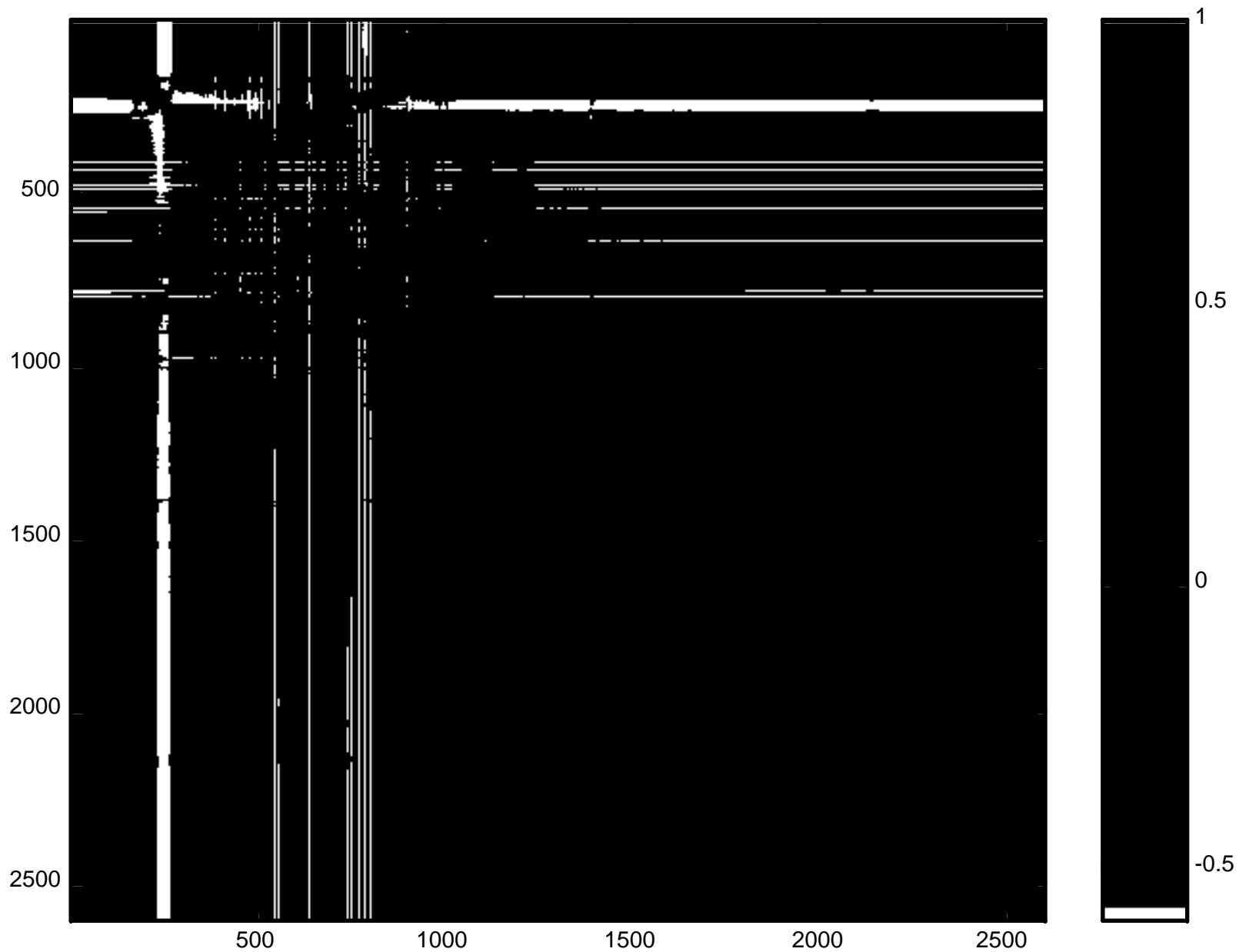


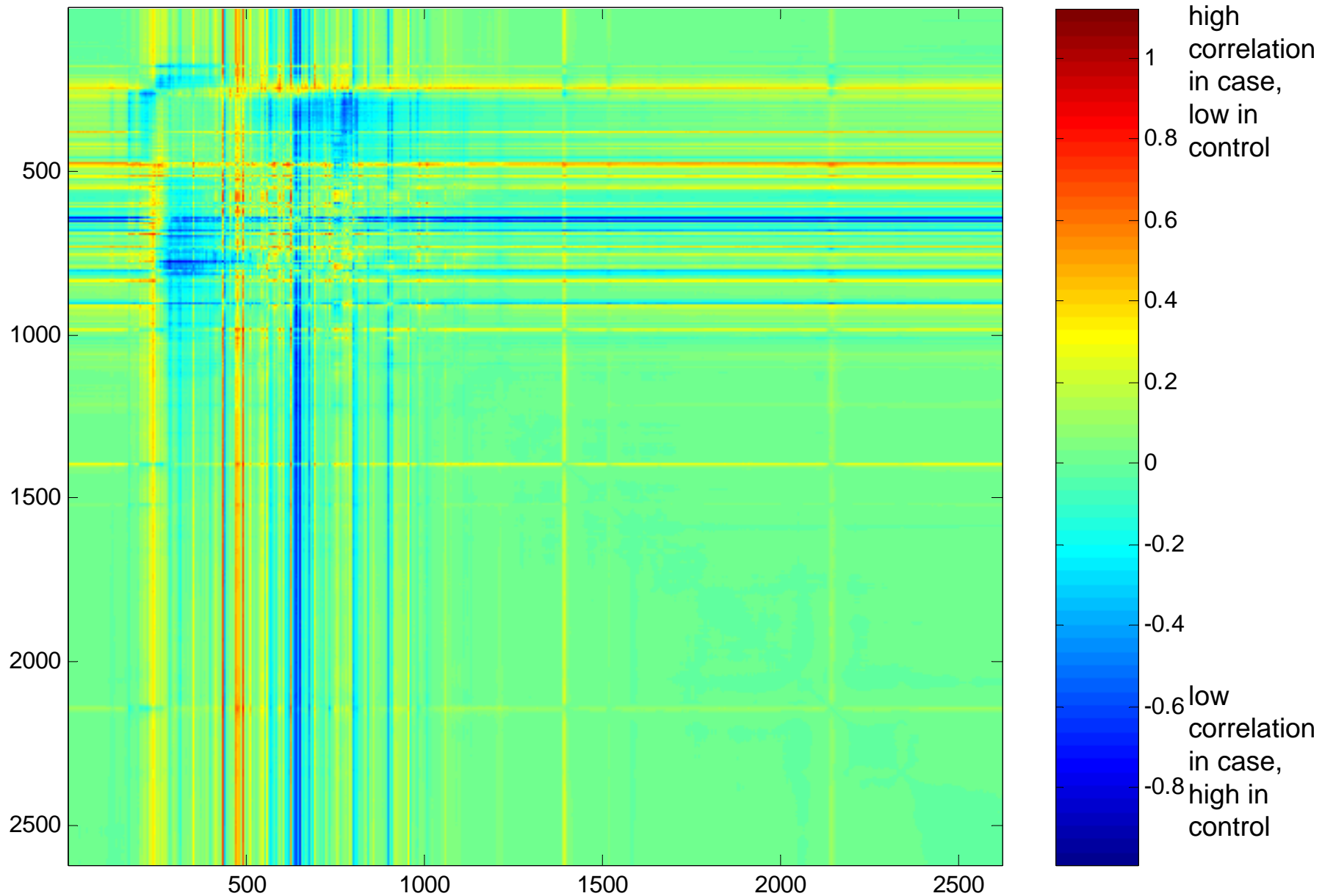


# Open Areas

- (1) Why Peaks Only?
- (2) Workflow/Protocol Optimization
- (3) Profile alignment/mass calibration+
- (4) Anti-correlated features as cancer biomarkers?







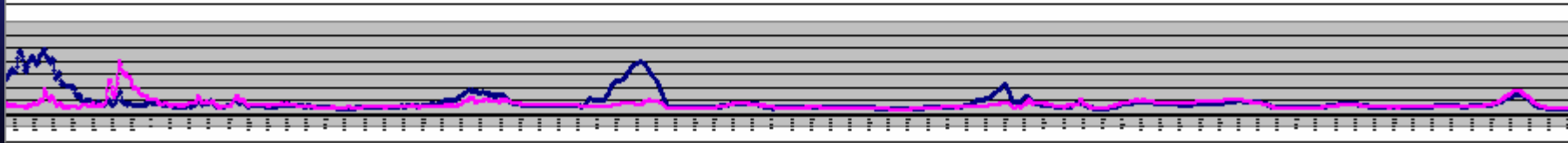
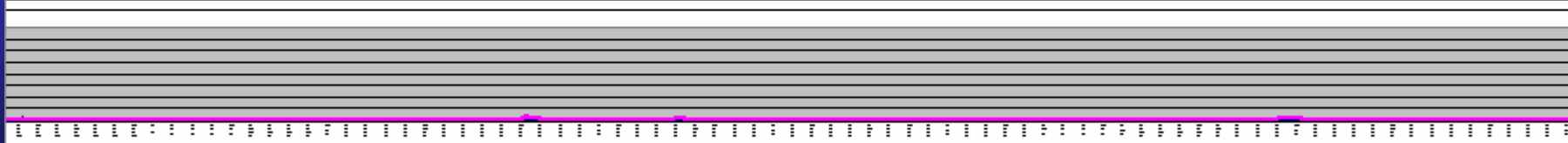
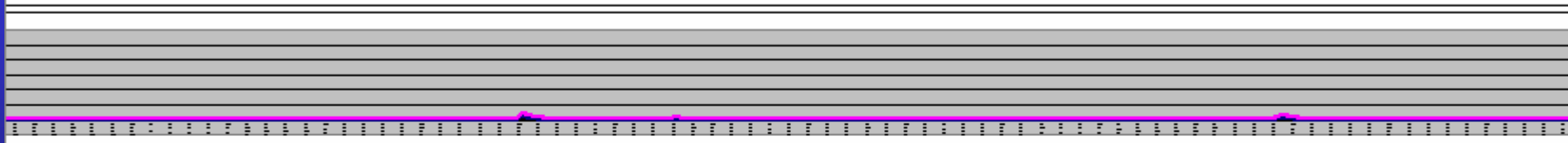
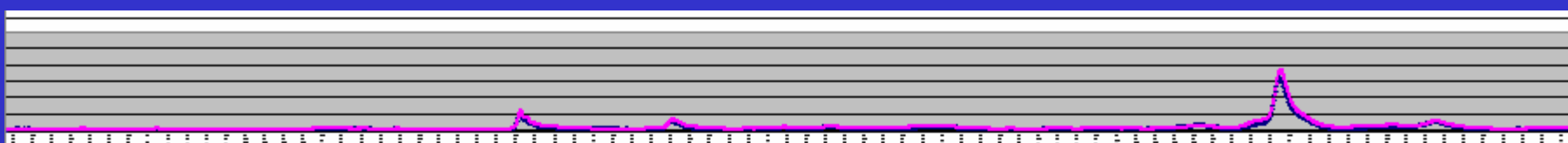
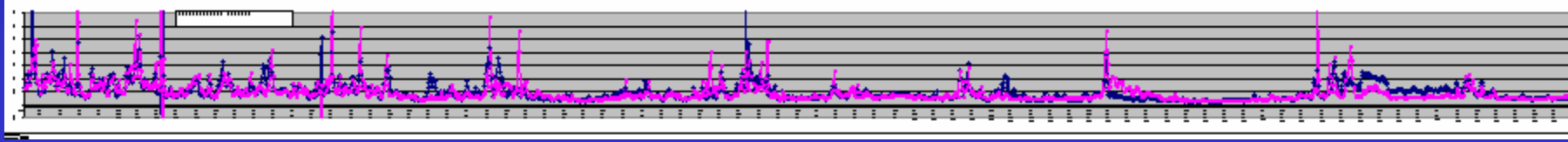
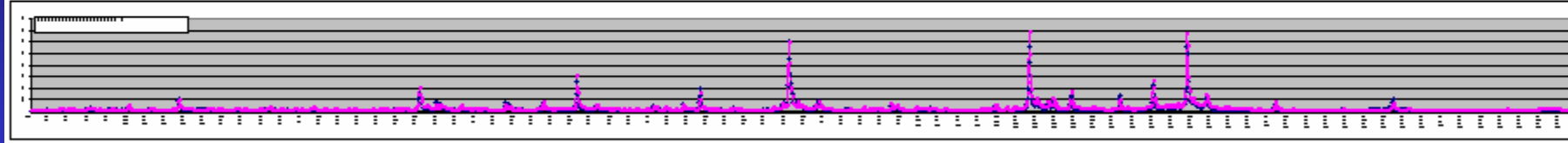
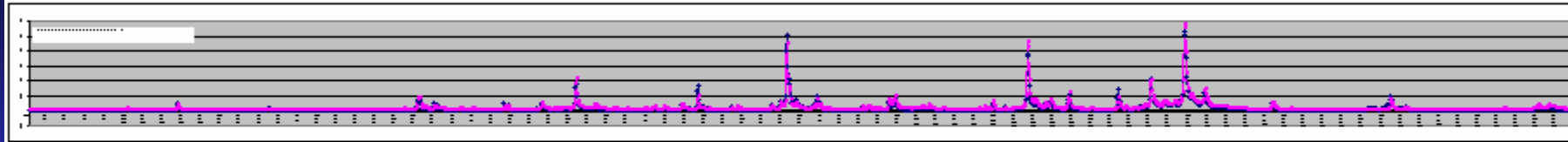
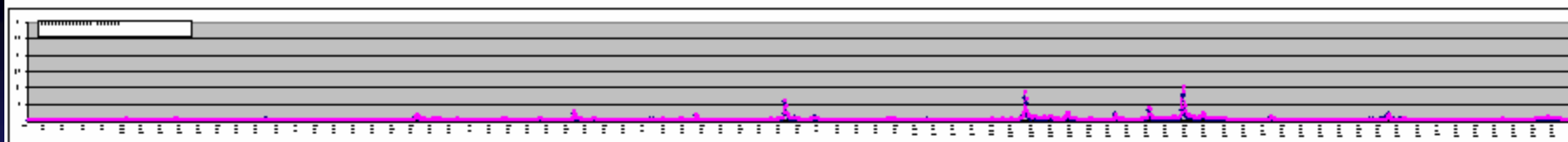


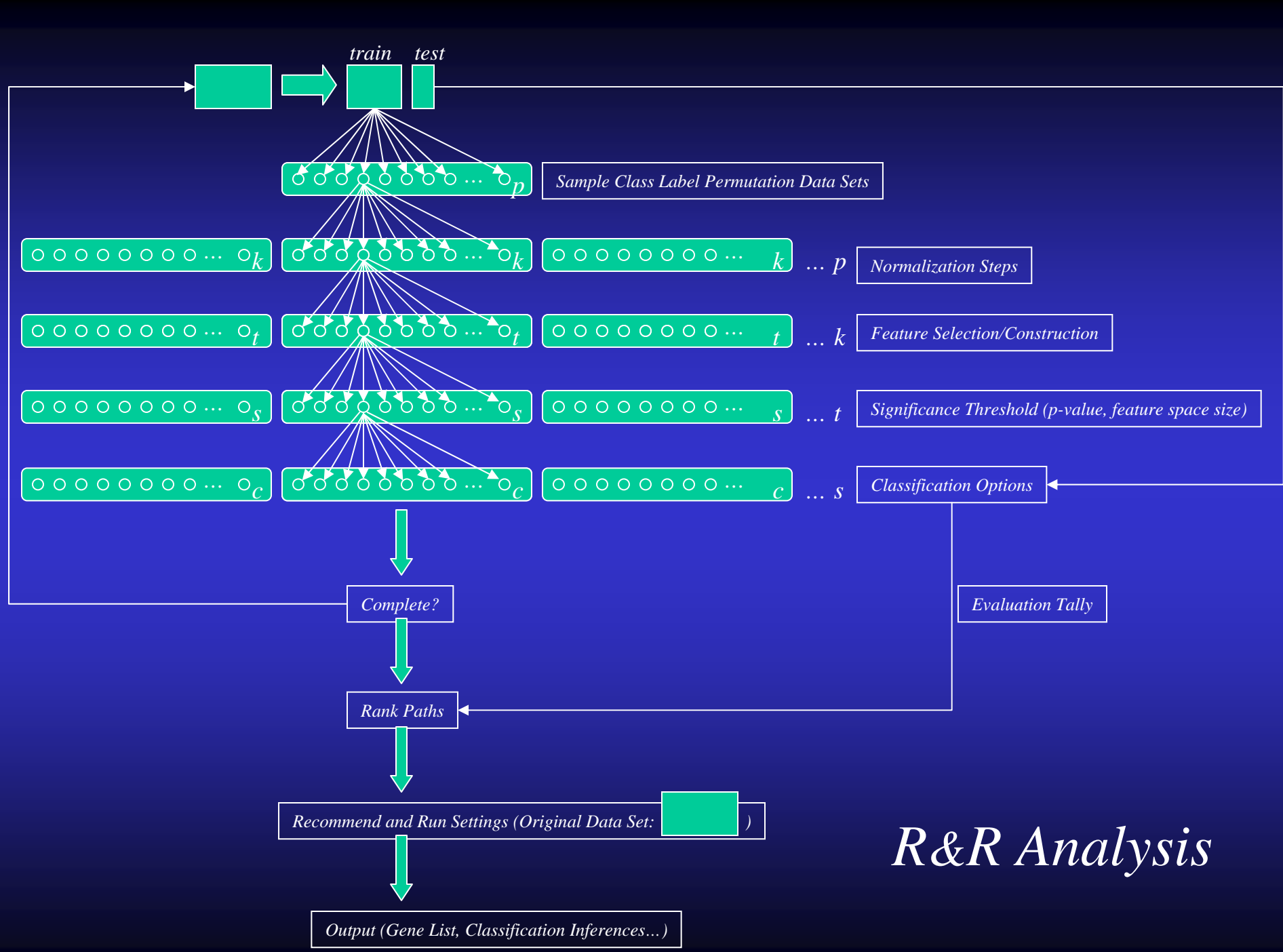
# Open Areas

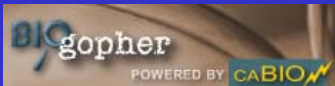
- (1) Why Peaks Only?
- (2) Workflow/Protocol Optimization
- (3) Are the markers cancer-specific?
- (4) Profile Alignment (mass calibration) +
- (5) Anti-correlated features as cancer biomarkers?
- (6) Relating Cellular Proteome to the Transcriptome

# Open Areas

- (1) Why Peaks Only?
- (2) Workflow/Protocol Optimization
- (3) Are the markers cancer-specific?
- (4) Profile Alignment (mass calibration) +
- (5) Anti-correlated features as cancer biomarkers?
- (6) Relating Cellular Proteome to the Transcriptome
- (7) Permutation-based feature selection
- (8) Network (grid) of Computational Clusters  
Dedicated to HTP Proteomic Analysis



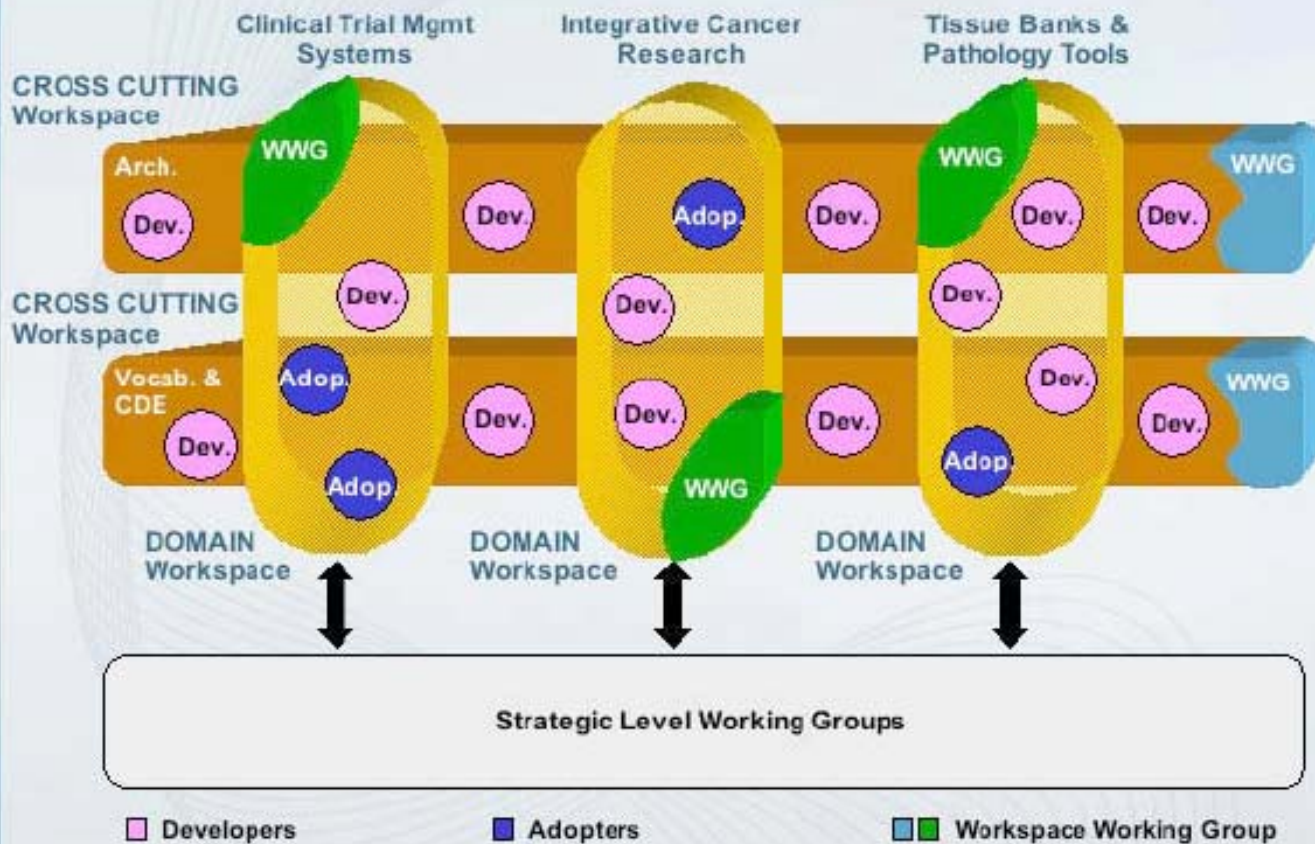




.  
. .  
n

### caBIG Facts and activities – Structure

The overall structure of the caBIG pilot is shown in the diagram below. An overview of these elements is given subsequent to the diagram.



# Thing (ca) BIG!

- Open Source
- Open Access
- Standardized Development
- Grid Computing?
- Develop & Participate
- National Integrative Cancer Research  
Databases

# Open Areas

- (1) Why Peaks?
- (2) Workflow/Protocol Optimization
- (3) Are these markers cancer-specific?
- (4) Profile Alignment (mass calibration)+
- (5) Look for anti-correlated features as cancer biomarkers
- (6) Relating Cellular Proteome to the Transcriptome
- (7) Permutation-based feature selection
- (8) Network (grid) of Computational Clusters Dedicated to HTP Proteomic Analysis

# Conclusions - and a Challenge

- Room for Improvement
- We recommend that all results achieved with new data sets - and prior results achieved with published data sets - be evaluated with PACE.



(A)



$\Sigma$

(B)



(C)

(D)



# Turns of the Cycle (Future Applications)

- (1) Cancer Detection
- (2) PROGRESSION/Identification of Targets
- (3) Recurrence
- (4) Theranosis (Prediction of Therapy Outcome)
- (5) Therapy Dose/Scheduling (real-time)
- (6) Unraveling the immunological stories...

# CPRN?

- (1) Encourage and enable studies of the mechanisms of cancer PROGRESSION
- (2) Re-analysis of published microarray and SELDI/MALDI-TOF data sets
- (3) Identification and characterization of PROGRESSION Targets
- (4) Include invitations to primary authors on original studies
- (5) Animal Model Focus
- (6) Lay up for Clinical Trials

# Example: Astrocytoma

- 3 Published Data Sets
- Fairly poor analysis
- Goldmine for consumers
- 18 genes differentially expressed in 2 data sets
- 8 genes differentially expressed in 3 data sets
- These are now validated progression markers.
- So?





- There are fewer producers than consumers - not for long.
- Learn Slowly
- Understand fundamental characteristics of the data
  - Lab workflows, sources of variability
  - Statistical Properties
- Communicate clearly
  - With each other
  - With our clients!
- Focus on clinical translation of knowledge acquired by our efforts.
- Effort to ID peptides at each m/z value detected by each chip

# APIII 2004

Advancing Practice, Instruction and Innovation through Informatics

---

October 6-8, Pittsburgh PA

- Association for Pathology Informatics
  - Early Detection Research Network
- Microarray Research Coordination Network
  - caBIG

Pathology Informatics  
Oncology Informatics  
Bioinformatics

Data Analysis & Statistical Methods  
Proteomics Bioinformatics  
Networks & Pathways

# Applied Bioinformatics

---

EDRN themed volume, late  
2004/early 2005

Please submit additional articles as  
well!

(lyonsweilerj@upmc.edu)