

About ROCs...

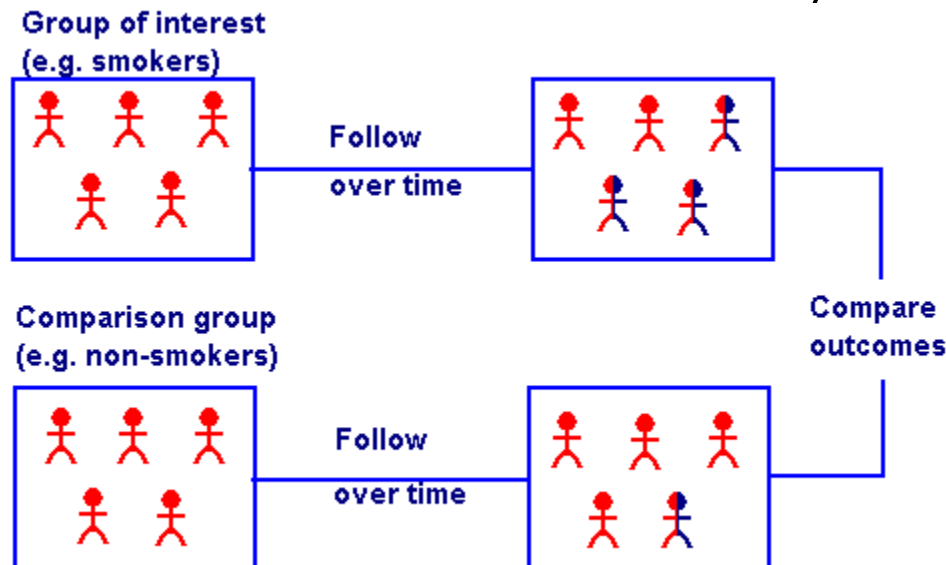
athanas@vastscientific.com

April 2010

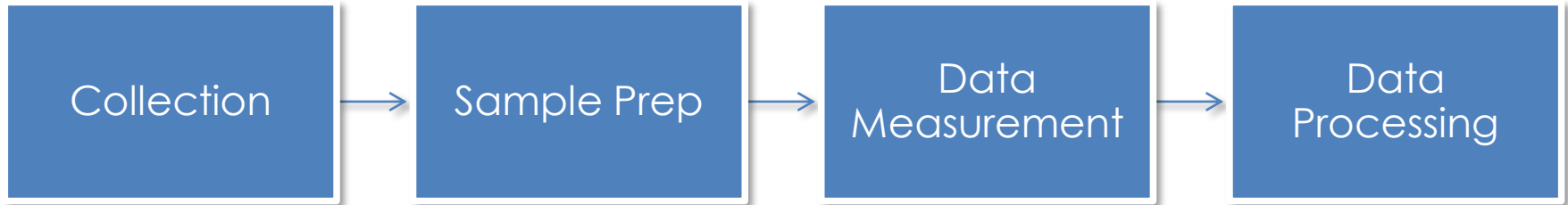


Context

- Biomarker Discovery / Confirmation Experiment
- Two class cohort study with ~10+ patients in each class
 - Example: Disease vs Normal
- Question:
 - Is there a biomarker or set of biomarkers that distinguish the two classes in the study?



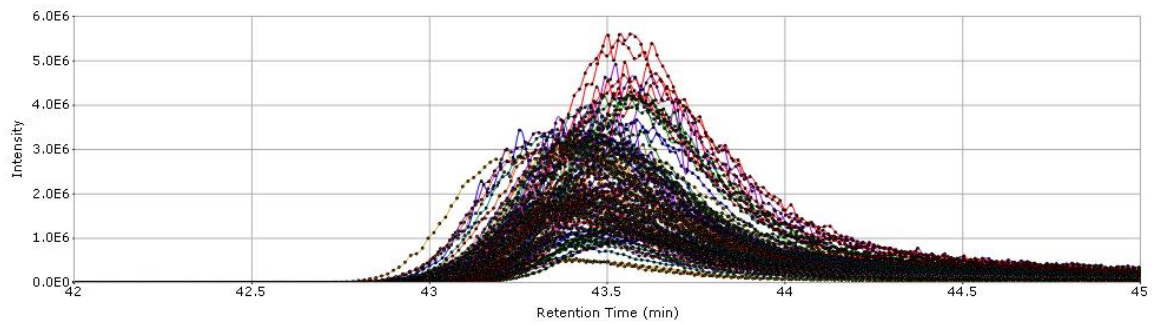
Discovery Experiment



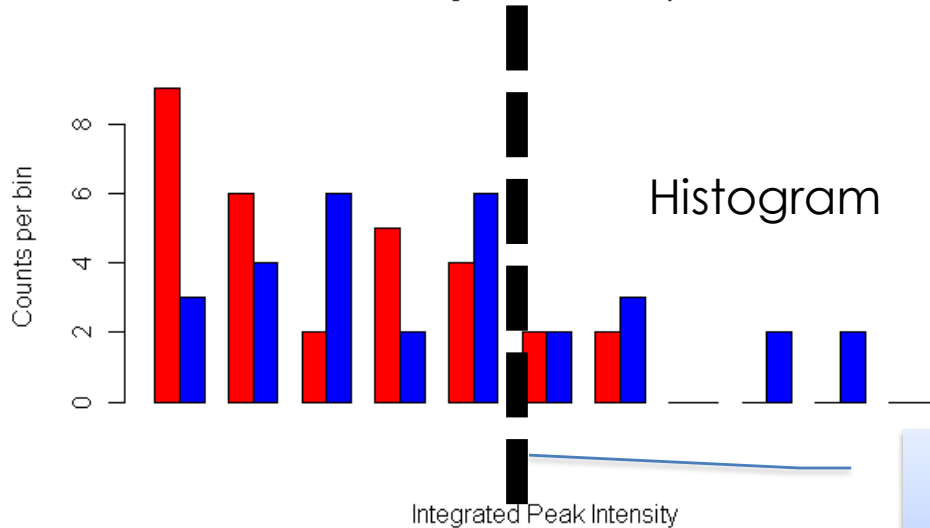
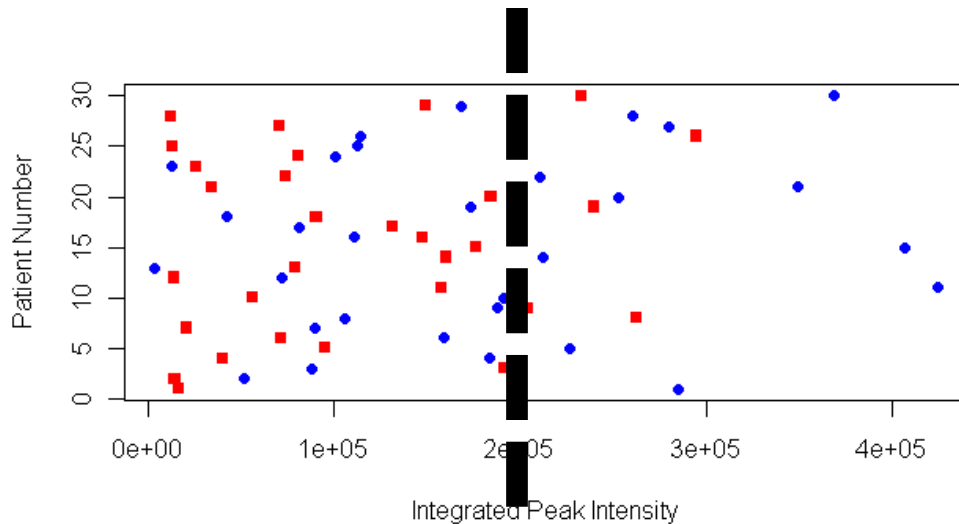
Set of candidate markers

Each marker candidate is a set of integrated peak intensities which corresponds to expression levels (and other related quantities)

BRIMS / Mass General heart study comparing plasma extracted from chambers before and after PFO closure – 246 samples.



Example Measurement – Marker Candidate



Simulated Experiment

One marker

30 Control
30 Treatment

Ratio = 2
StdDev=2.2
Pvalue = 0.013

What is the best threshold to distinguish Control vs Treatment?



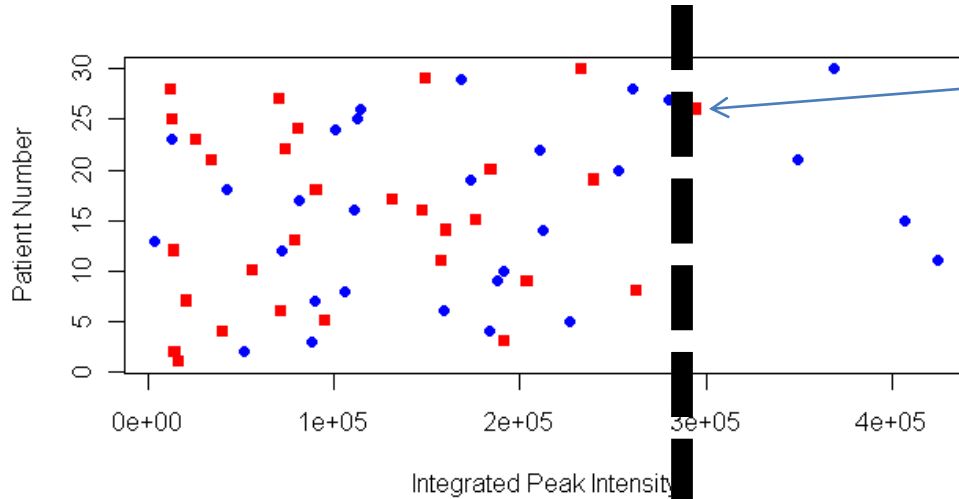
Why Ratios are Not So Interesting

- Ambiguous:
 - What is the best ratio cutoff value to designate a marker candidate?
 - 1.2 fold? 2 fold? 10 fold? ?????
- A ratio value alone is insufficient:
 - A marker could have a high fold value but large standard deviation
- (Ratios)+(Standard Deviations)+(Pvalues) are a complicated way of describing the strength of a marker candidate
- Clinicians who gauge the “strength” of a marker do not care about ratios

How to Gauge a Marker's Effectiveness

- Sensitivity
 - The probability that a test result will be positive when the disease is present
 - True positive
- Specificity
 - The probability that a test result will be negative when the disease is not present
 - True negative

Sensitivity and Specificity

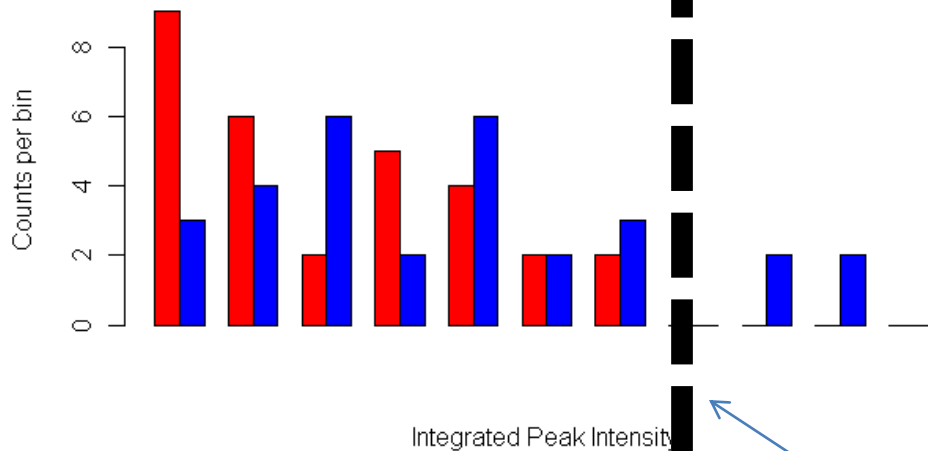


One **treatment** (red) measurement above threshold

Four **control** measurements (blue) above threshold

Sensitivity:
True Positives = $\frac{4}{30}$

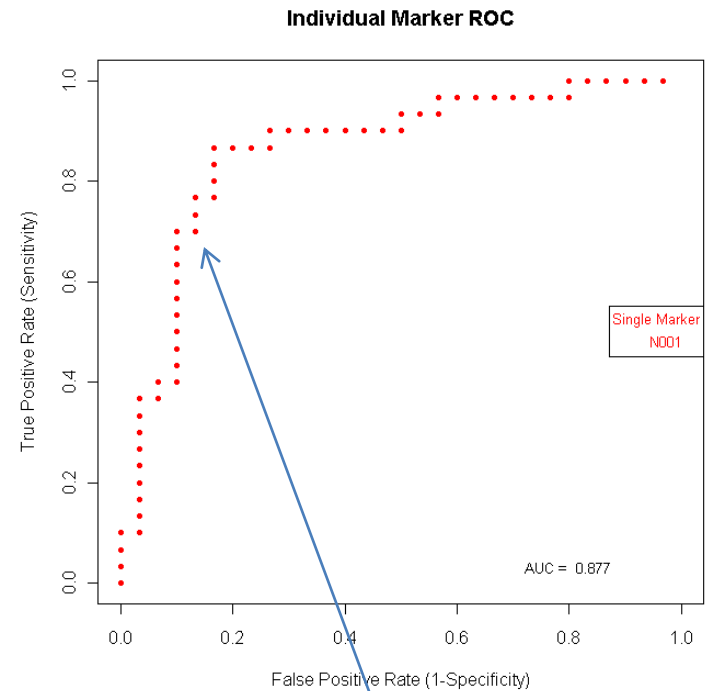
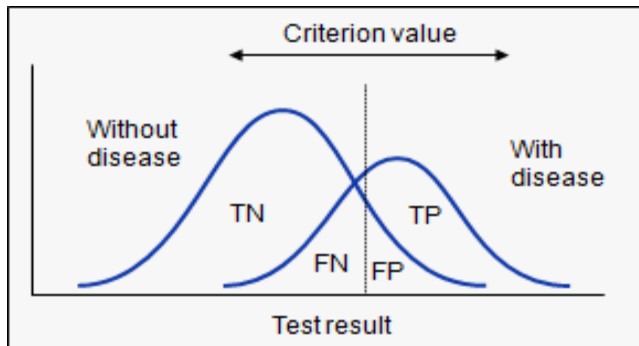
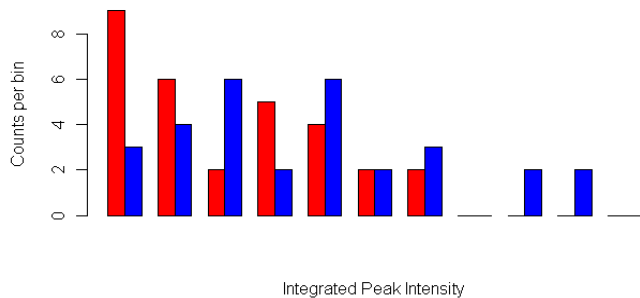
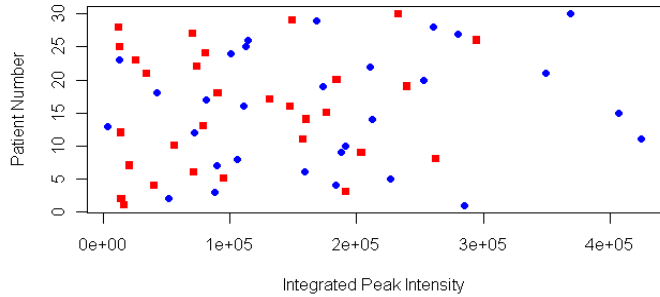
1-Specificity:
False Positives = $\frac{1}{30}$



Pick a threshold value (arbitrary)

ROC Plot

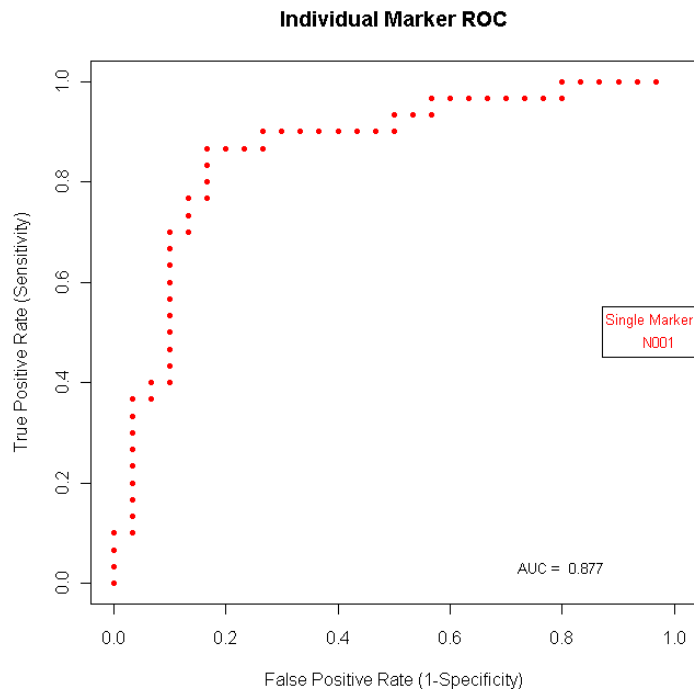
Plot Sensitivity versus (1-Specificity) for all possible thresholds



This point corresponds to a threshold that gives 75% Sensitivity with a false positive rate of 20%

Area Under the Curve AUC

Reduce the core content of a ROC plot to a single number



A ROC plot has dimensions of 1x1

- Maximum area under the curve is 1
- AUC close to 1 → strong marker
- AUC close to 0.5 → weak marker

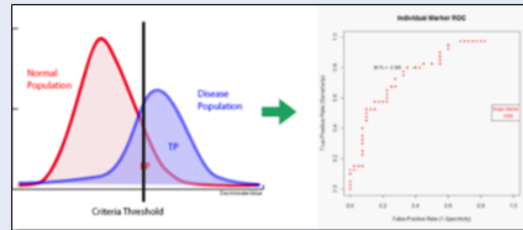
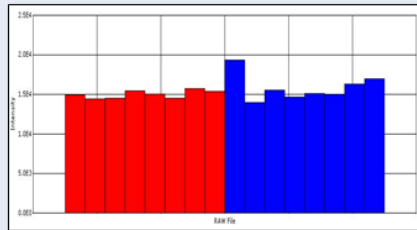
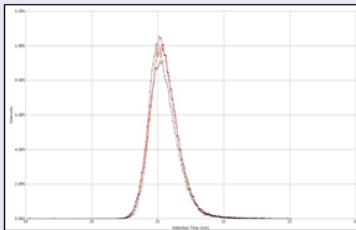
BRIMS

TWO PASS DISCOVERY WORKFLOW

SIEVE ROC Critical to Workflow

In order to realize the quantitative power of SIEVE, data collection must be very robust

BRIMS Two Pass Discovery Workflow using SIEVE and Orbitrap



Design and Optimization

- Robust, commercially available nanoflow LC
- Focus on stable spray
- Focus on high reproducibility of peak intensities, CV<8%

methods

Pass 1: Quantification

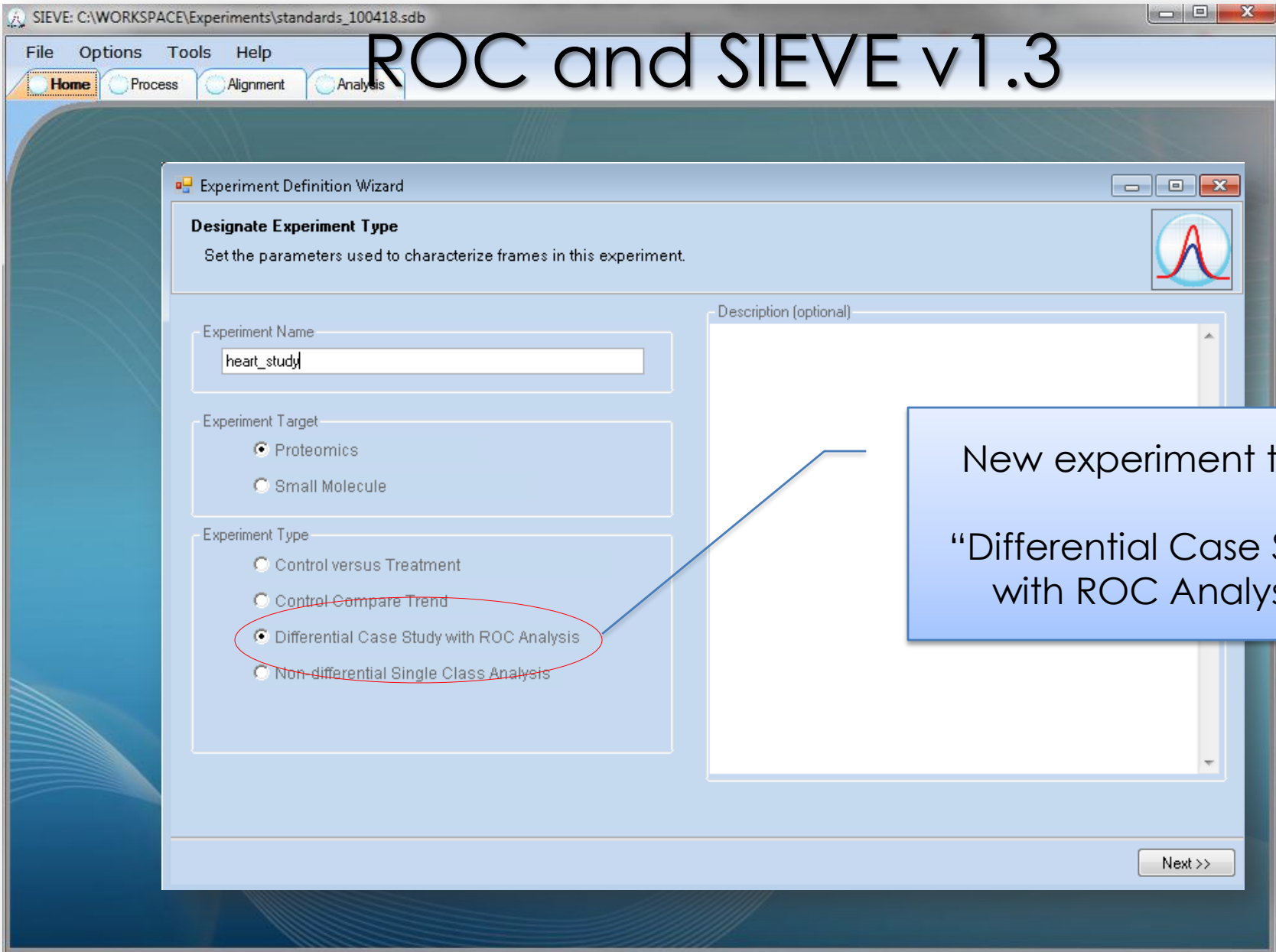
- Chromatographic alignment
- Uncompromised full scan measurements
- Each sample is measured once – no need for replicates
- No MS2's acquired

Inclusion list

Pass 2: Identification

- Targeted fragmentation by inclusion list
- Not all samples measured – subset as determined from SIEVE analysis
- Marker stratification using multi marker and single ROC AUC (SIEVE 1.3)
- Export to Ingenuity pathway analysis

Slide 14 from Amy's presentation



ROC and SIEVE v1.3

Designate Experiment Type

Set the parameters used to characterize frames in this experiment.

Experiment Name

heart_study

Experiment Target

- Proteomics
- Small Molecule

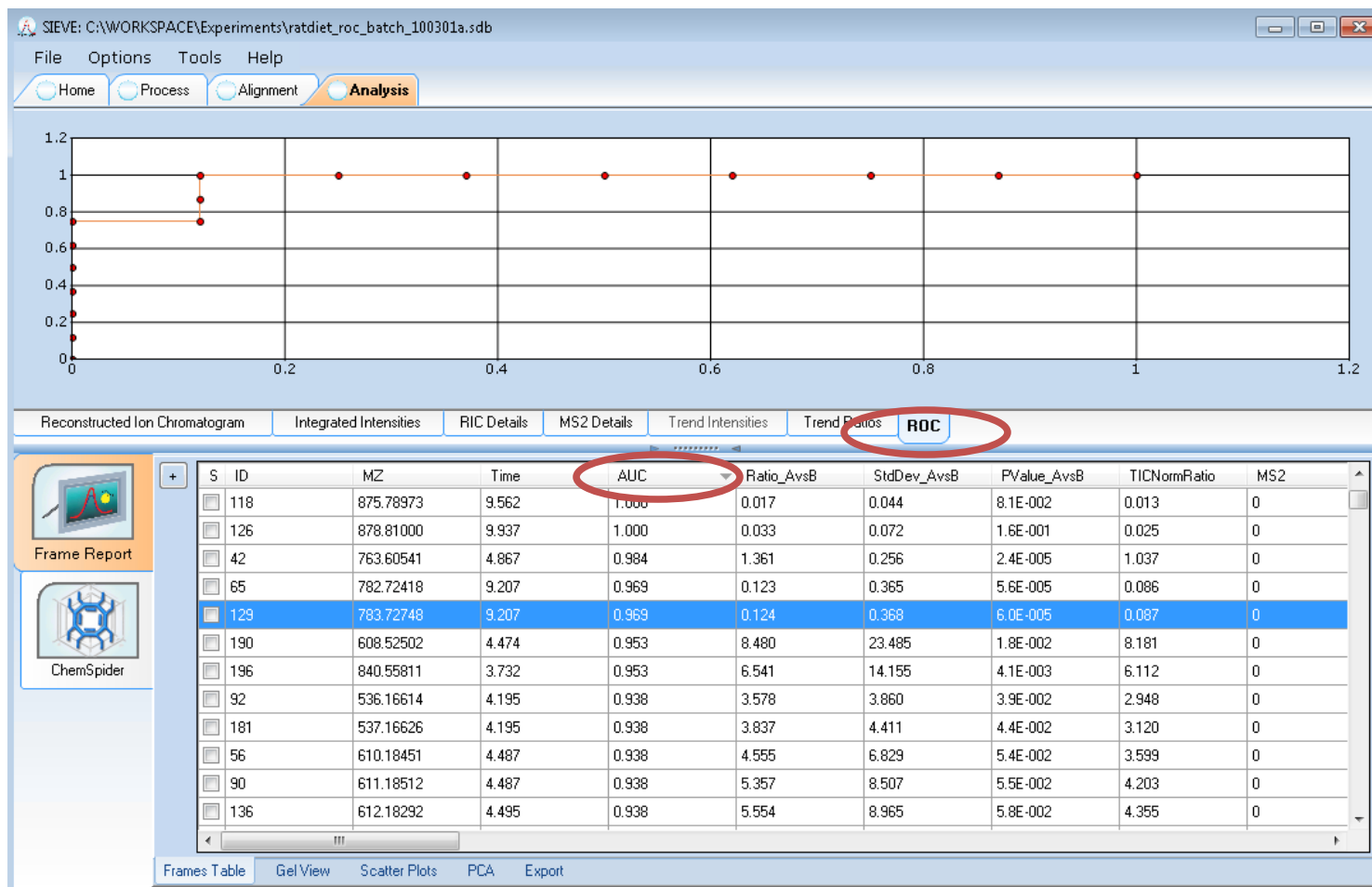
Experiment Type

- Control versus Treatment
- Control Compare Trend
- Differential Case Study with ROC Analysis
- Non-differential Single Class Analysis

New experiment type:
"Differential Case Study
with ROC Analysis"

Next >>

ROC Area Under Curve

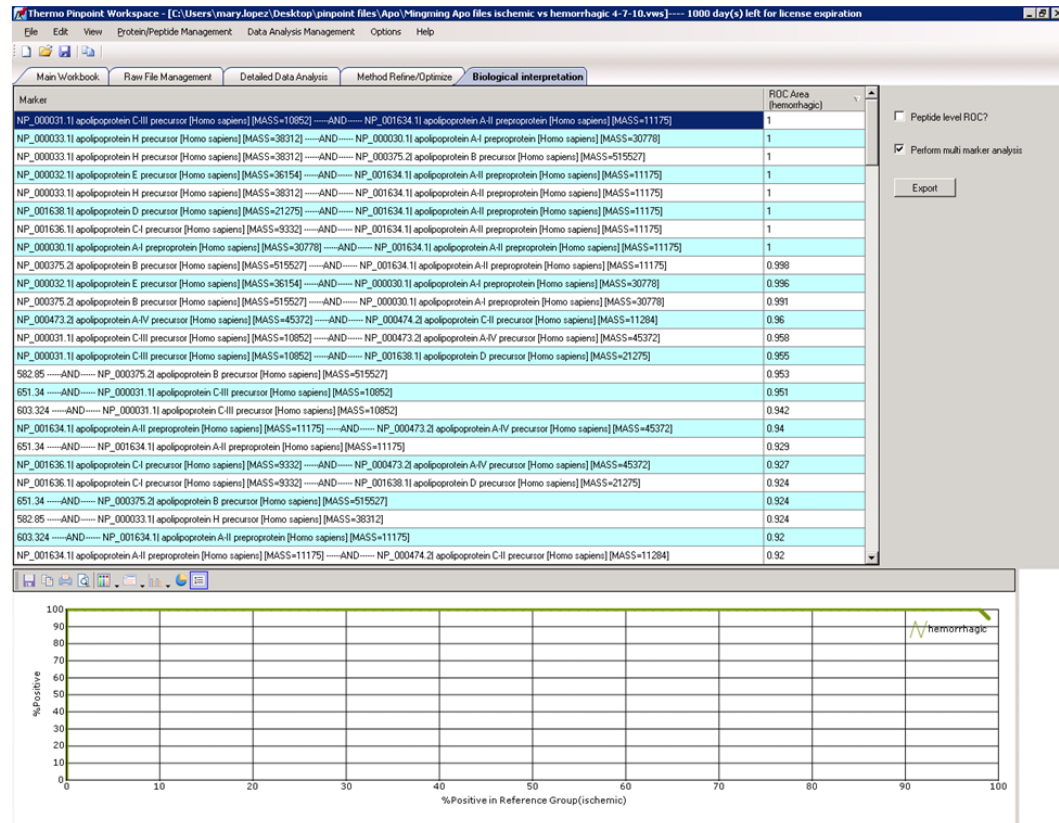
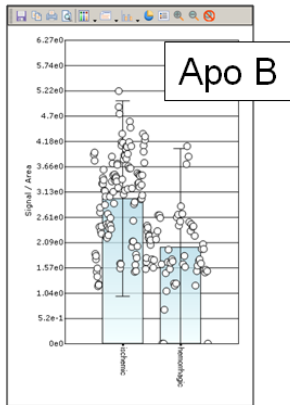


ROC in Pinpoint

ROC analysis of apolipoprotein levels in hemorrhagic vs ischemic stroke patients

Top AUC

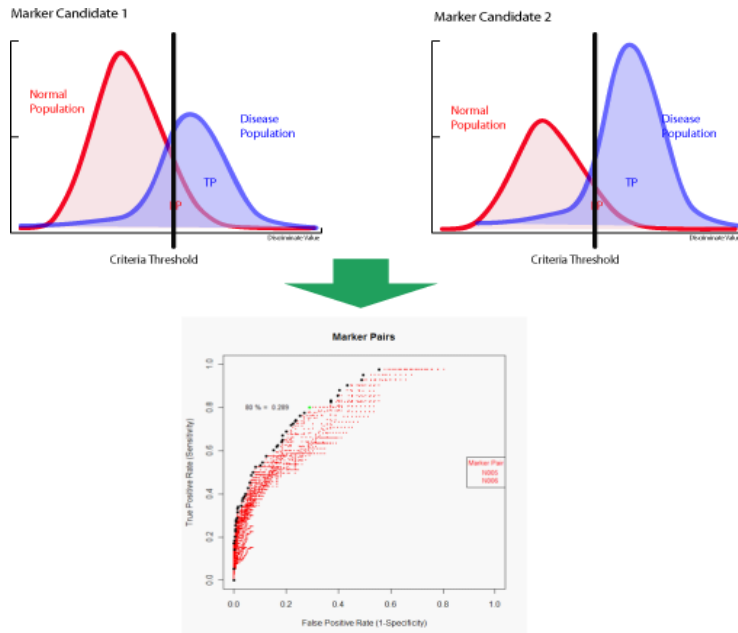
- Apo C-III and Apo AII 1.0
- Apo H and Apo B 1.0
- Apo E and Apo AII 1.0
- Apo H and Apo AII 1.0
- Apo D and Apo AII 1.0



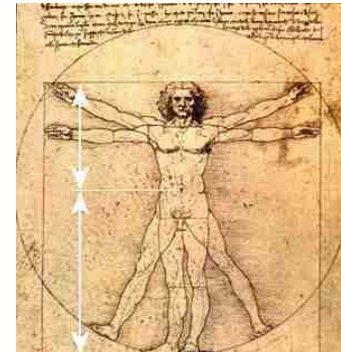
Novel Algorithms to Combine Markers

Expression patterns between multiple markers can enhance the ability to distinguish between disease and normal classes.

Combining two or more markers:



ROC from marker ratios:



Why do we ROC?

- Ratios are not best way to assess a marker candidate.
- ROC is used for marker confirmation.

Additional Resources

- MedCalc
 - <http://www.medcalc.be/manual/roc.php>
- “Discovery analysis of multiple protein marker panels: Optimizing sensitivity and selectivity”, Athanas, et. al.
 - MSACL 2010 Poster