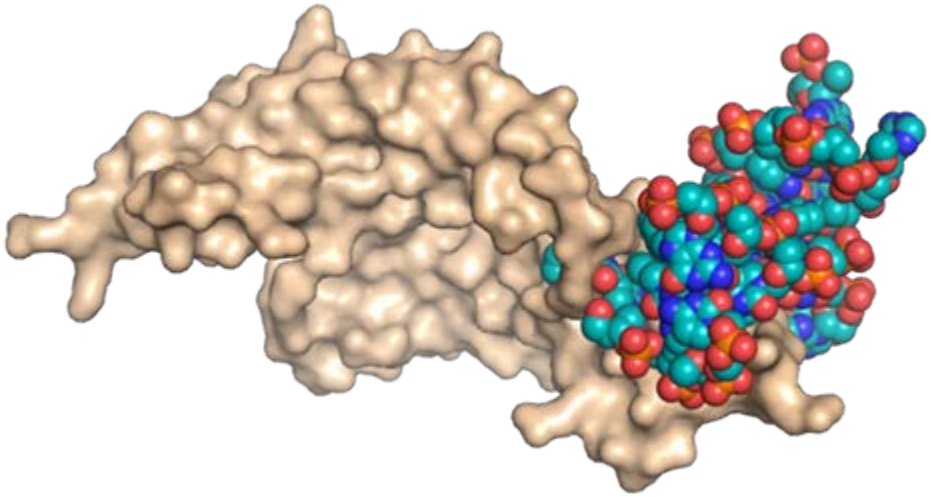


Exposing the Criminal Record of Every Blood Sample: Use of SOMAmer Technology and Sample Mapping Vectors to Mitigate False Biomarker Discoveries

Rachel Ostroff, PhD
Clinical Research Director, SomaLogic Inc.
rostroff@somalogic.com

© 2000-2012, SomaLogic, Inc.

Blood is the Cornerstone of Biomarker Discovery

- Noninvasive window into disease changes throughout the body
- Large biorepositories established for cohort and case/control biomarker discovery studies
- Yet many biomarker discoveries fail to validate
  - Study population doesn't match final clinical indication
  - Hidden preanalytic variability in the discovery samples contaminates the apparent disease markers

© 2000-2012, SomaLogic, Inc.

Blood is a Dynamic Fluid

- Profound post-collection changes can occur
  - Platelet activation
  - Complement activation
  - Cell lysis and release of intracellular proteins
  - Proteolysis
- Sample processing effects can masquerade as disease markers
- Proteomic evidence reveals this bias as a sample's "criminal record"

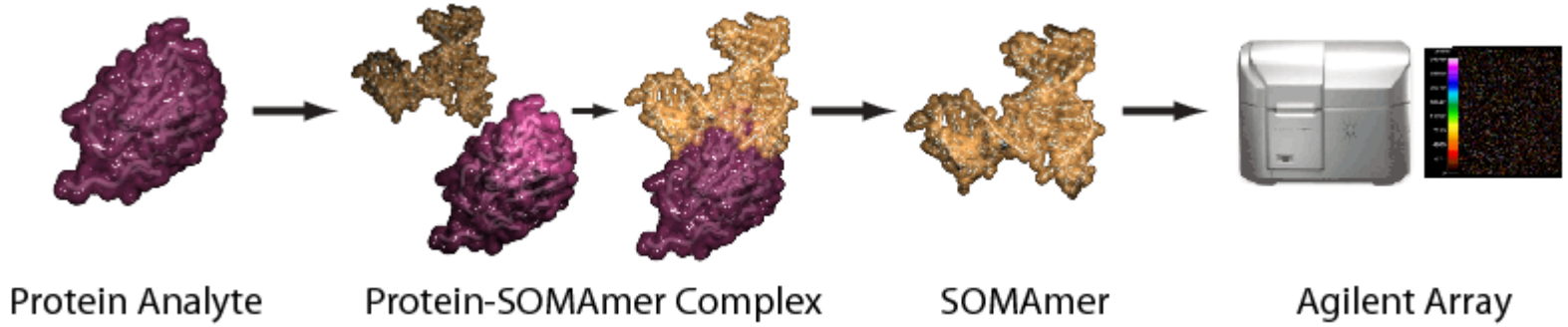
© 2000-2012, SomaLogic, Inc.

Untangling Preanalytic Bias and Biomarker Discovery

Sample mishandling causes entire pathways to change

SOMAscan assay delivers deep interrogation of the proteome

- *Sensitive and precise: LOD <1pM with <5% CV*
- *Highly multiplexed platform: >1000 analytes from 20 μL sample*
- *High throughput: >300 samples/day*
- *Broad coverage of cellular pathways and disease pathophysiology*
- *Works in many biological samples (e.g., tissue, blood, urine, CSF)*

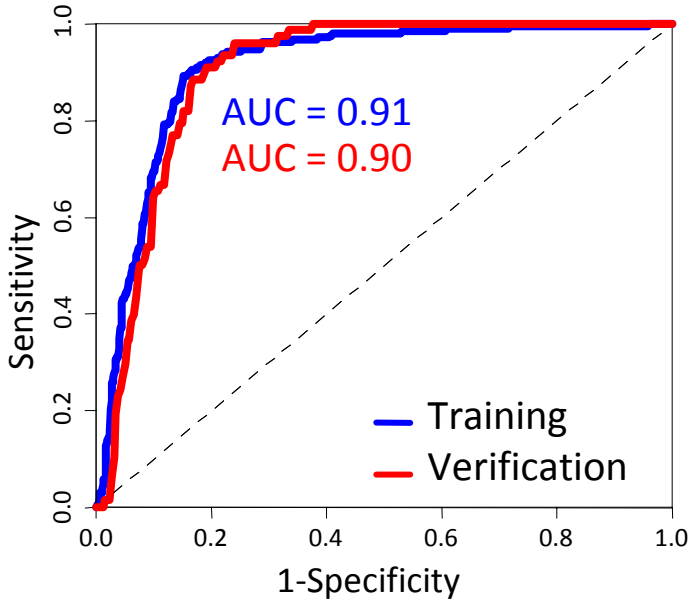


© 2000-2012, SomaLogic, Inc.

Lung Cancer Case Study: Revealing Sample Bias

Biomarker Discovery Study Design

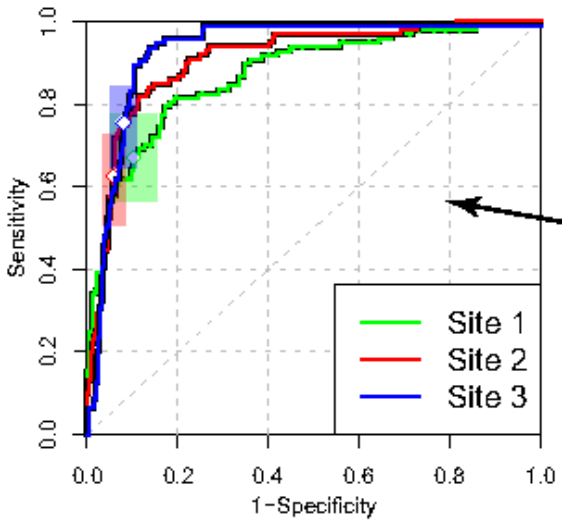
- 1,326 serum samples from 4 biorepositories with the same sample collection protocol
- Exact target population for clinical diagnostic application
  - 30% NSCLC cases (half were Stage I) & 70% High risk smoker controls
- Training set of ~900 samples and a blinded verification set of ~400 samples
- Excellent performance, consistent across sites
  - AUC of ~90%
  - Sensitivity 89% specificity 84%
  - Performance independent of stage



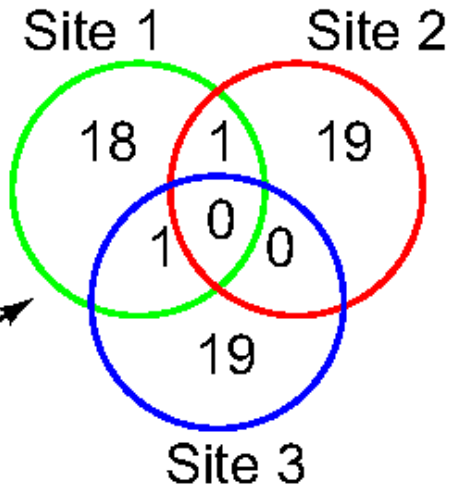
Ostroff et al. (2010) PloS ONE 5, e15003

© 2000-2012, SomaLogic, Inc.

But… Top Biomarkers at Each Site are Different



A single classifier performs well on each site

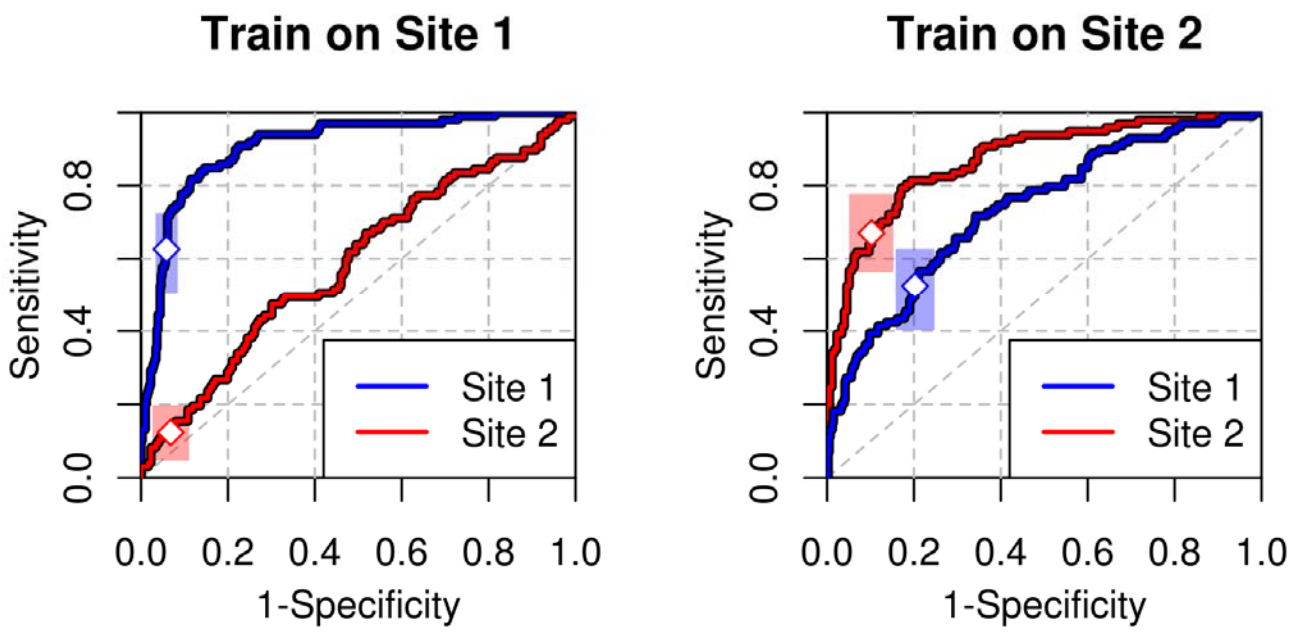


But the top biomarkers at each site are different

© 2000-2012, SomaLogic, Inc.

And… Translation Across Sites is Poor

Blinded verification would have failed if we had trained on any one site and tested on any other

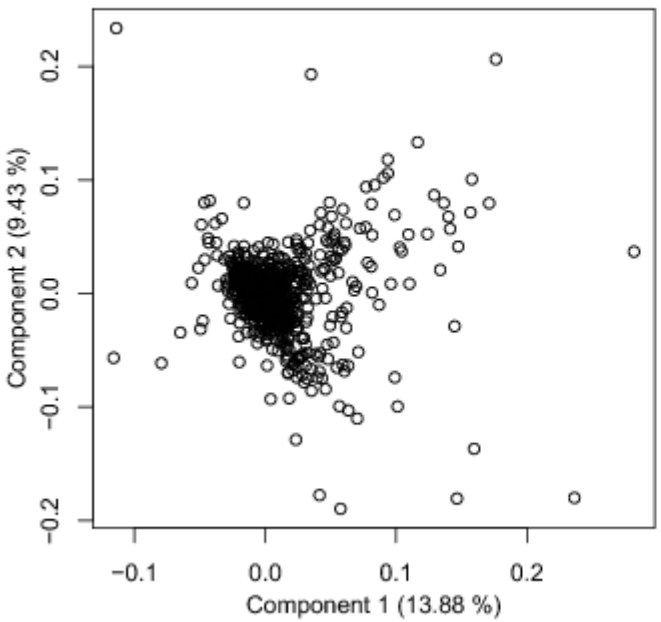


© 2000-2012, SomaLogic, Inc.

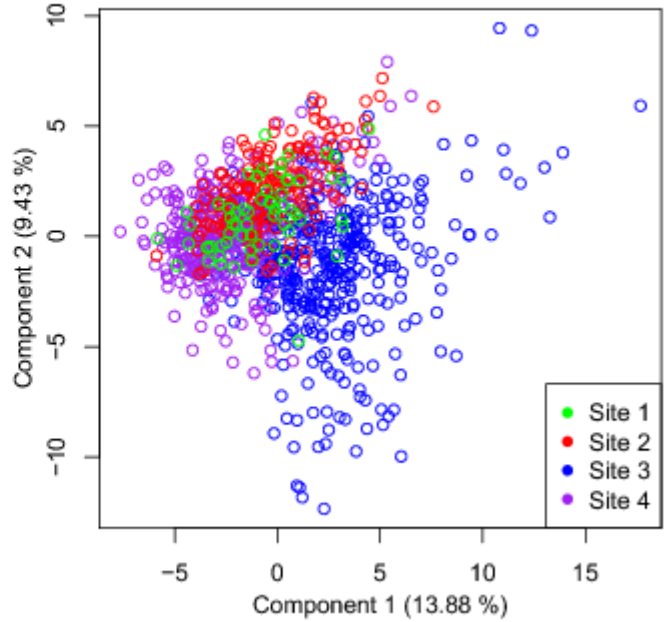
So… Let's Deconstruct Site Differences (1): The Controls Are Different

Principal Components Analysis was used to uncover the hidden patterns in the data

Individual Proteins



Individual Samples

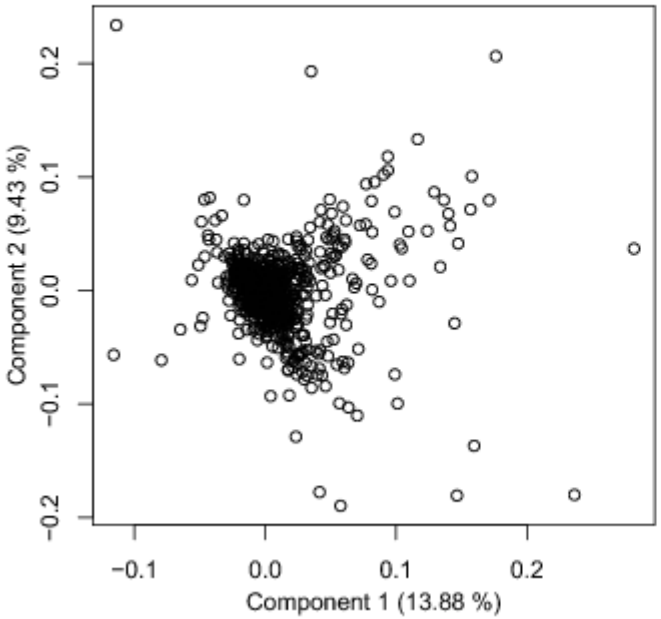


© 2000-2012, SomaLogic, Inc.

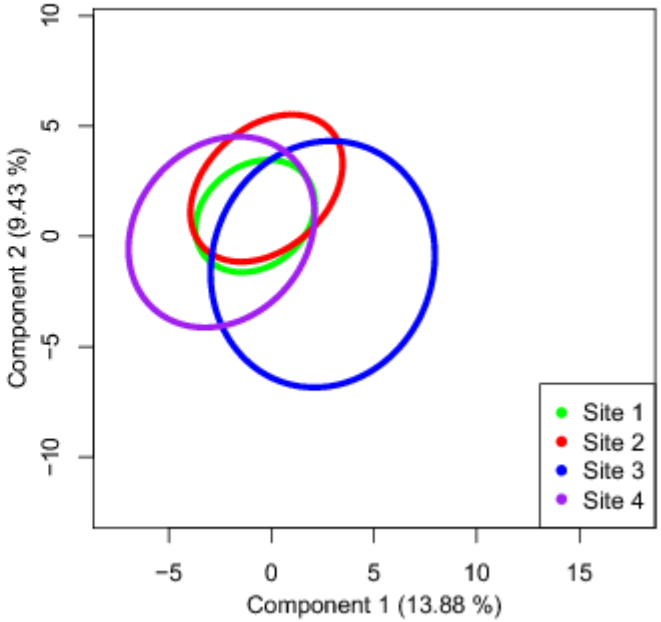
So… Let's Deconstruct Site Differences (1): The Controls Are Different

Samples from different sites occupy different spaces in the first 2 principal components

Individual Proteins

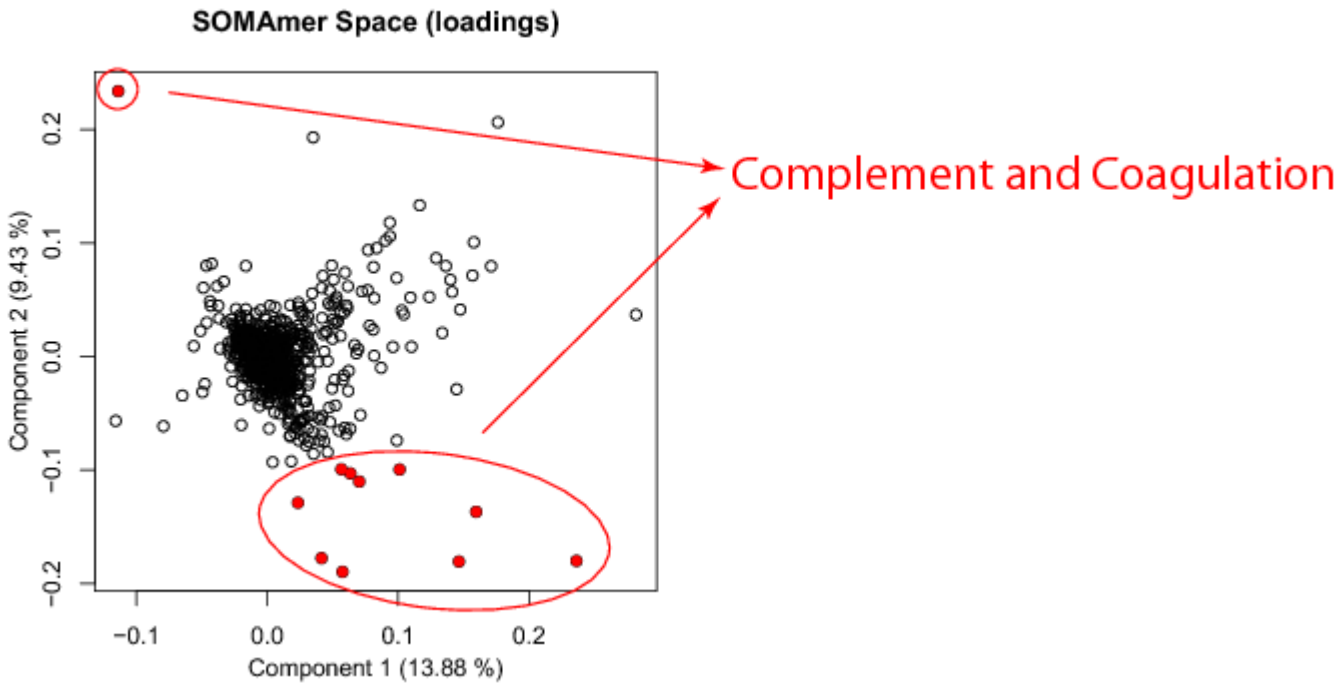


Individual Samples



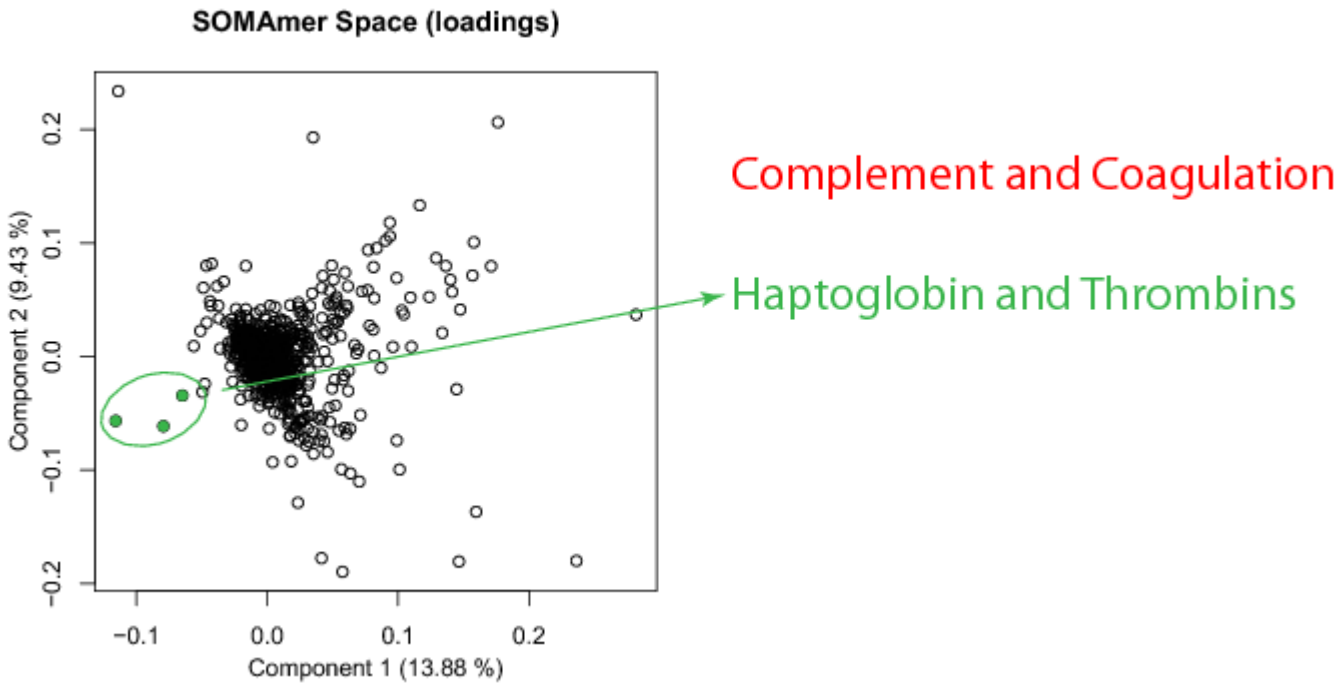
© 2000-2012, SomaLogic, Inc.

So… Let's Deconstruct Site Differences (2): Why The Controls Are Different – Biological Causes



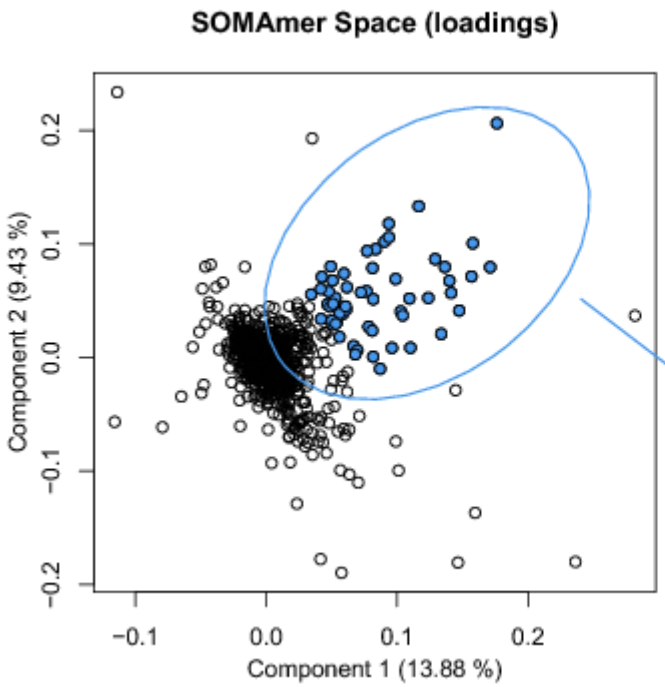
© 2000-2012, SomaLogic, Inc.

So… Let's Deconstruct Site Differences (2): Why The Controls Are Different – Biological Causes



© 2000-2012, SomaLogic, Inc.

So… Let's Deconstruct Site Differences (2): Why The Controls Are Different – Biological Causes



Complement and Coagulation

Haptoglobin and Thrombins

Significantly enriched for intracellular proteins (via GO)

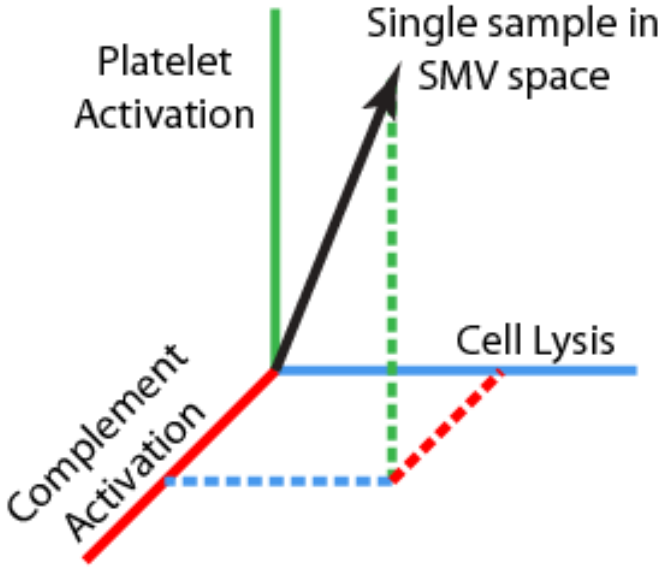
Variability could be replicated in the lab by manipulating processing time, temperature, centrifuge speed, or inducing platelet activation

Ostroff et al. (2010) J Proteomics 73:649

© 2000-2012, SomaLogic, Inc.

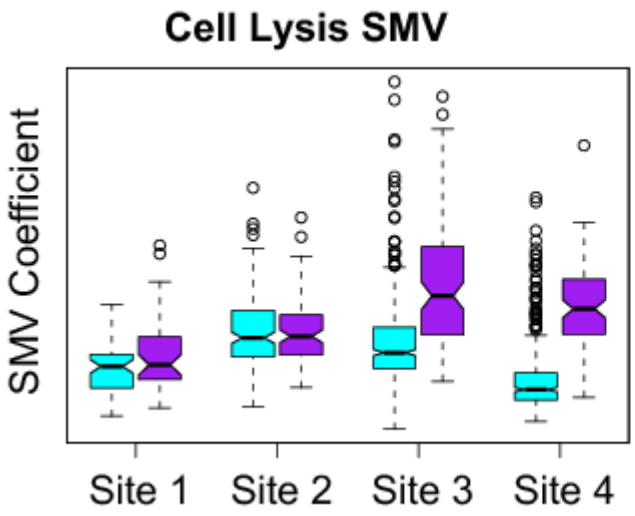
Quantifying Site Differences: Creating a Criminal Record for Each Sample

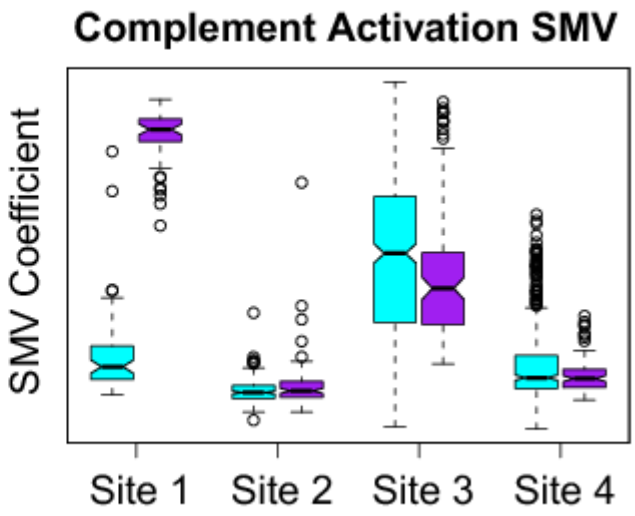
- From the 1000 proteins measured we selected small sub-panels which were purely affected by only one type of pre-analytic effect
- We created multidimensional vectors of the effects which are applied to each sample
- We then went back and applied them to the lung cancer discovery set



© 2000-2012, SomaLogic, Inc.

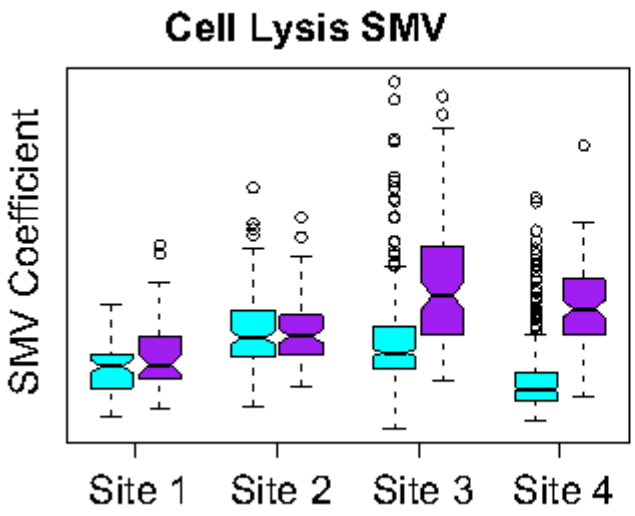
Quantifying Site Differences: Evaluating The Criminal Record for Each Lung Cancer Sample

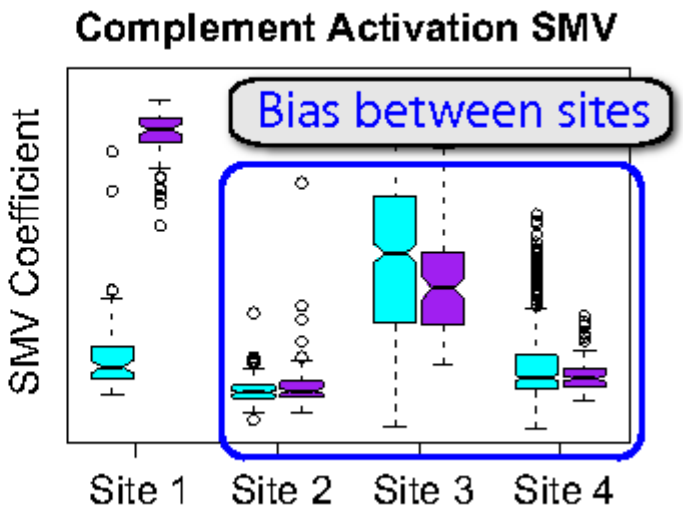




© 2000-2012, SomaLogic, Inc.

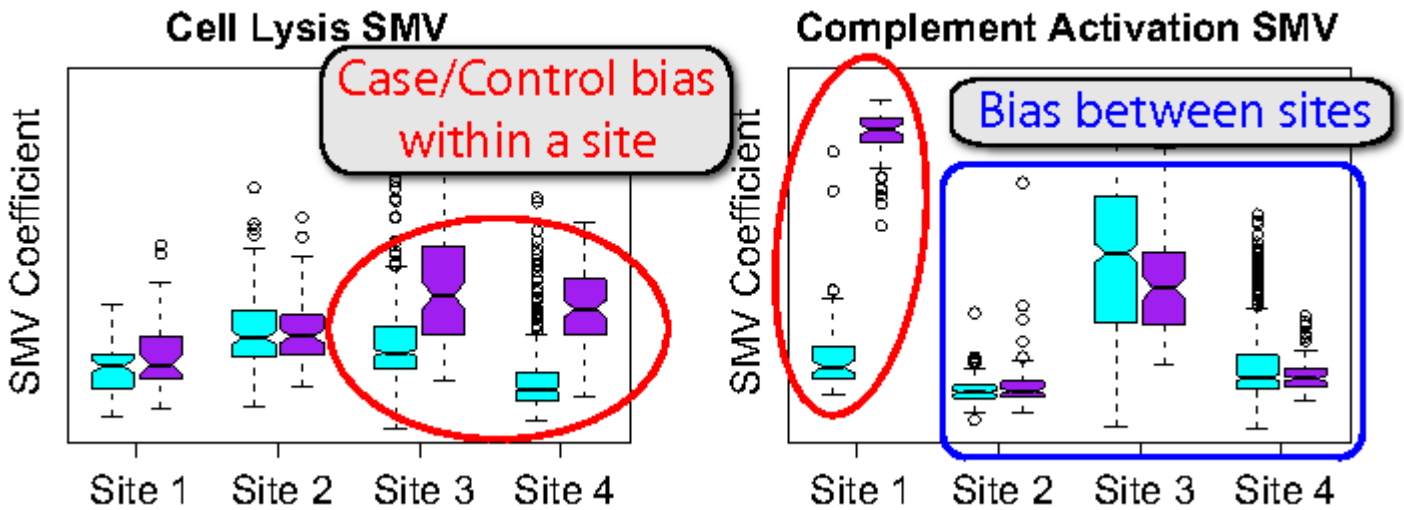
Quantifying Site Differences: Evaluating The Criminal Record for Each Lung Cancer Sample





© 2000-2012, SomaLogic, Inc.

Quantifying Site Differences: Evaluating The Criminal Record for Each Lung Cancer Sample



© 2000-2012, SomaLogic, Inc.

Overcoming Site Differences: Possible Solutions

Remove Affected Analytes

- Pros: reduces risk of improper classification
- Cons: potential loss of true biomarkers

Remove Affected Samples

- Pros: removes bias from initial discovery
- Cons: loss of power, may be impractical for a product

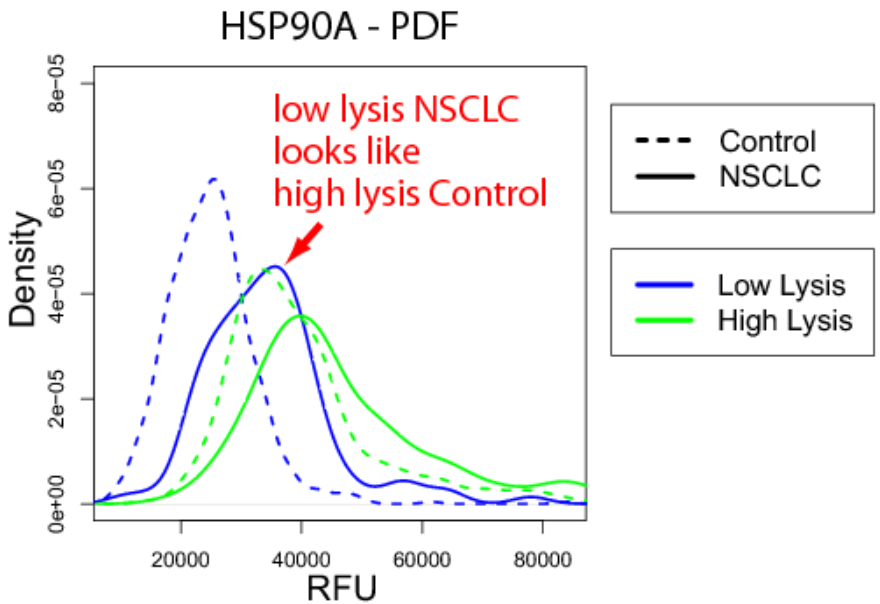
Mathematically Correct the Effects

- Pros: keeps all samples and analytes
- Cons: requires a lot of experimentation, difficult to validate, may not be possible for some biological effects

© 2000-2012, SomaLogic, Inc.

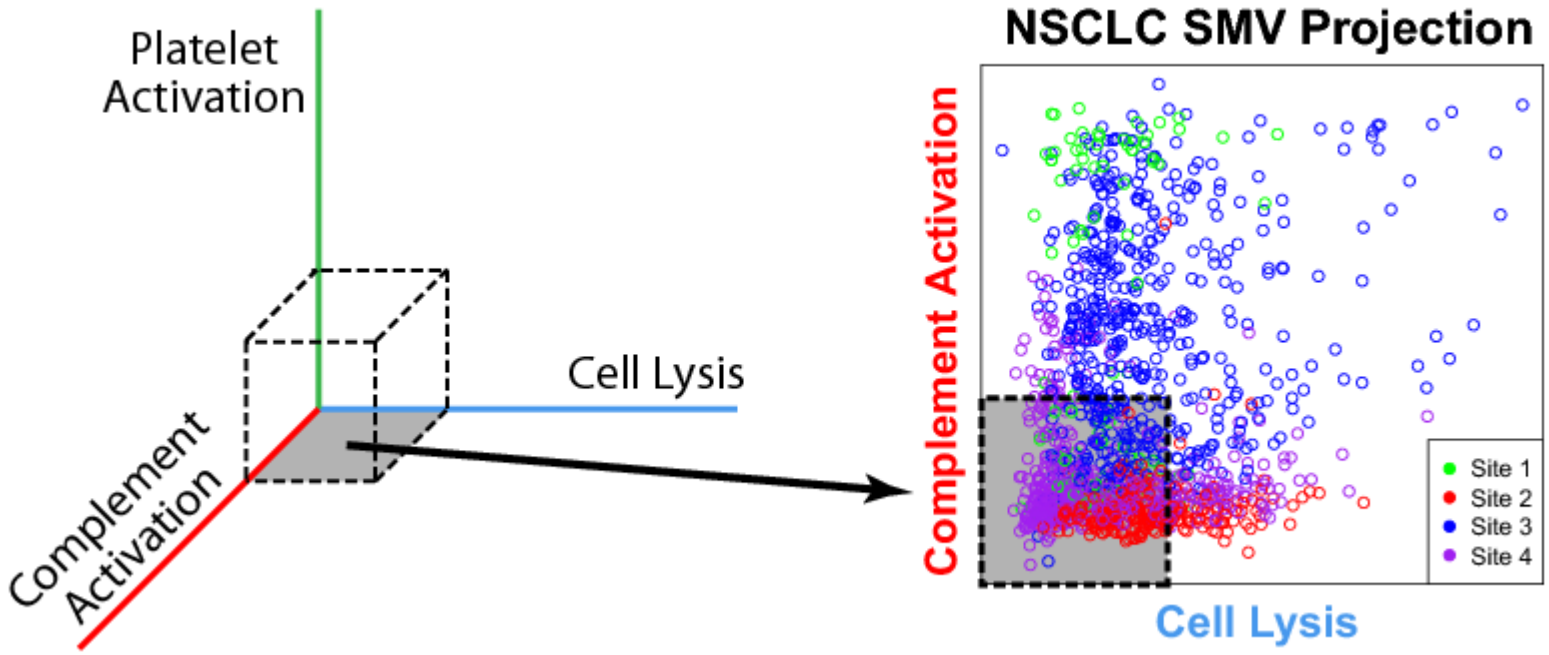
Overcoming Site Differences
1. Remove Disease Markers Susceptible to Bias

Example: HSP90 is a known marker of lung cancer and is a therapeutic target. However, as a diagnostic, it is adversely affected by leukocyte lysis and therefore susceptible to bias and variability of collection.



© 2000-2012, SomaLogic, Inc.

Overcoming Site Differences:
2. Remove Samples From Biomarker Discovery

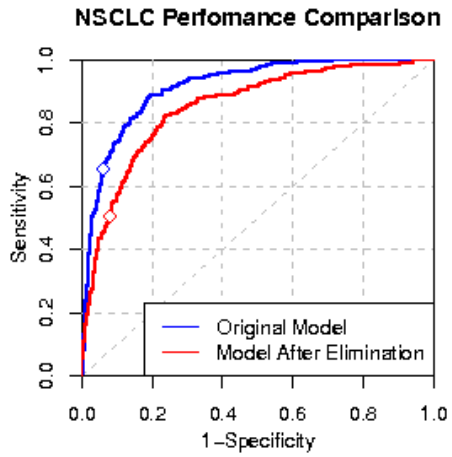


We identified an unbiased, uncontaminated fraction of samples in the lung cancer discovery set and re-did discovery on those samples

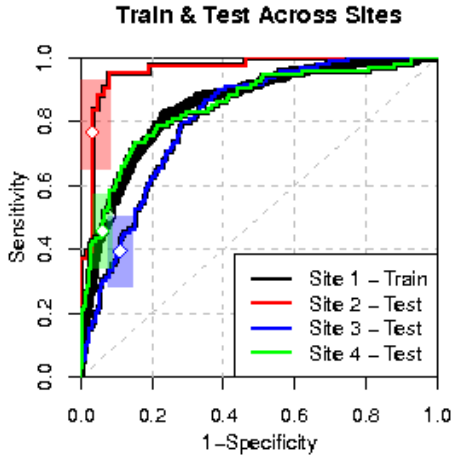
© 2000-2012, SomaLogic, Inc.

Solution: Mitigating False Discovery by Removing Criminal Analytes and Samples

- Half the "top 12" analytes were eliminated
- AUC decreased from 0.9 to 0.84
- Now see a clear relationship between disease stage and performance
  - Explains previous observations: Case-control bias markers aren't stage related!



- The classifier did not lose performance if training on one site and testing on others



© 2000-2012, SomaLogic, Inc.

Responsible Biomarker Discovery: Overcoming Preanalytic Effects and Hidden Bias

- Mishandling of samples results in activation of biological pathways and produces unintentional bias
  - Leads to false discovery of apparent disease signatures
- Choose markers wisely
  - Not just the best case/control discrimination
  - Must be resistant to processing variability
  - Increases the likelihood that a biomarker panel will translate to the clinic
- The SOMAmer-based SMV tool quantifies sample bias
- SMV analysis can be applied broadly to reveal trustworthiness of samples before proceeding to biomarker discovery

© 2000-2012, SomaLogic, Inc.

Acknowledgements

SomaLogic

Ed Brody

Mike Mehan

Alex Stewart

Stephen Williams

Assay Team

Biorepository Staff

University of Pittsburgh Medical Center

William Bigbee

Jill Siegfried

Joel Weissfeld

NYU School of Medicine

Harvey Pass

William Rom

© 2000-2012, SomaLogic, Inc.