Exposing the criminal record of every blood sample: Use of SOMAmer™ technology and sample mapping vectors to mitigate false biomarker discoveries

Stephen Williams1, Rachel Ostroff1, Ed Brody1, Alex Stewart1, Harvey Pass2, William Rom2, William Bigbee3, Jill Siegfried3, Joel Weissfeld3, Mike Mehain1

1SomaLogic, Inc., Boulder, CO; 2NYU School of Medicine, New York, NY; 3University of Pittsburgh Medical Center, Pittsburgh, PA

Abstract
Biomarker discovery studies may fail to translate to the clinic because the study population does not match the intended clinical use or because hidden preanalytic variability in the discovery samples contaminates the apparent disease-specific information in the biomarkers. This can arise from differences in blood sample processing between study sites or in samples collected differently at the same study site. To better understand the effect of different blood sample processing procedures, we evaluated protein measurement bias in a large multi-center lung cancer study using the >1000 protein SOMAmer™ assay. These analyses revealed that perturbations in serum collection and processing result in changes to families of proteins from known biological pathways. We subsequently developed a protein biomarker signature of cell lysis, platelet activation and complement activation and assembled these preanalytic signatures into quantitative multi-dimensional Sample Mapping Vector (SMV) scores. The SMV score provides critical evaluation of the quality of every blood-based sample used in discovery and also enables the evaluation of candidate protein biomarkers for resistance to preanalytic variability. Despite uniform processing protocols for each clinic, the SMV analysis revealed unexpected case/control bias arising from collecting case and control serum from different clinics at the same academic centers, an effect that created false or bias-contaminated disease markers. We therefore used the SMV score to remove bias-susceptible analytes and to define a well-collected, unbiased training set. An improved classifier was developed, resistant to common artifacts in serum processing. The performance of this classifier to detect lung cancer in a high risk population is more likely to represent real world diagnostic results. We believe this approach is generally applicable to clinical investigations in all fields of biomarker discovery and translational medicine.

SOMAmer: Enabling Proteomic Discovery
Preanalytic variability can be modeled by deep interrogation of the proteome

Biomarker Discovery Study Design
1,326 serum samples from 4 clinical sites assayed with SOMAmer proteomic technology
Cases: 293 NSCLC Stages I-IIII, 60% adenocarcinoma, 30% squamous cell carcinoma
Controls: 1035 Benign pulmonary nodules & long term smokers with no evidence of cancer
75% used for biomarker discovery, algorithm training and cross validation
25% of samples tested as a blinded verification set
- Classifier trained for consistency across all sites
- 12 protein classifier with AUC of 0.9 in both training & verification
- Sensitivity ~90% for Stage I disease

BUT...more conventional approaches to biomarker validation would have failed

Quantifying Sample Processing Artifacts in Every Sample
Principal Component Analysis Reveals Relationship Between Sample Bias and Biological Processes
- Study sites can be separated by proteins clustering in discrete cellular pathways
- Representative SOMAmers are assembled into orthogonal Sample Mapping Vectors (SMVs)
- Individual samples can be mapped with each SMV to derive a sample quality score
- Bias between sites or between cases and controls within a site can now be quantified

Lessons Learned
Responsible biomarker discovery: mitigation of pre-analytic effects & hidden bias
- High throughput multiplex SOMAmer measures enable precise characterization of biological samples and discovery of robust disease biomarkers.
- Poor control of sample collection protocols can lead to useless sample sets and false biomarkers.
- Unintentional changes in processing biological samples may differentially activate proteomic pathways and produce case/control or site biases, often misinterpreted as disease signatures.
- Applying quantitative measures of preanalytic variability, we developed a tool for evaluating bias in clinical studies before proceeding to biomarker discovery.
- Choosing biomarkers with not just the best case/control discrimination but that are also resistant to sample processing bias increases the likelihood that a biomarker panel will perform well in the clinic.
- Work is continuing to make SMVs truly quantitative across different clinical studies.

This study includes samples from:
- SOMAmers provided to SomaLogic's SOMAmer Discovery Project (SOMAdx) by the Center for Clinical Proteomics, University of Pennsylvania
- SOMAmers provided to SomaLogic's SOMAmer Discovery Project (SOMAdx) by the Center for Clinical Proteomics, University of Pennsylvania

Discovery of Robust Biomarkers
Solutions for responsible biomarker discovery: (1) remove bias-susceptible proteins
- Reduces risk of future improper classifier

Solutions for responsible biomarker discovery: (2) remove affected samples
- BUT...more conventional approaches to biomarker validation would have failed

Cell Lysis SMV
Complement Activation SMV

SOMAmers are assembled into orthogonal Sample Mapping Vectors (SMVs)
SMVs map individual samples

PCA analysis of controls across 4 sites identifies proteins responsible for sample bias

SMV values map individual samples

Cell Lysis SMV
Complement Activation SMV
PLTs
NCS
NSCLC SMV Projection

Application of SMV scores to both control samples and case samples; a cutoff (shaded box) is defined for acceptable training samples which does not contain bias between sites or between cases and controls.

Retraining a robust lung cancer classifier
- The analysis was re-done on the unbiased, well collected fraction of the original sample set
- Resulting 7 marker classifier has an AUC of 0.86 for detection of lung cancer in a high risk population
- Classifier is robust across sites and insensitive to sample processing artifacts
- Biomarkers discovered in the un-biased samples also work well when applied to the entire study set