Building a Robust and Reliable Proteomics Biomarker Pipeline: NCI’s CPTAC Network

Christopher R. Kinsinger, Tara Hillke, Mehdi Mesri, Henry Rodriguez
National Cancer Institute, Bethesda, MD

Abstract

Translation of candidate proteins as indicators of cancer status from discovery to validation biomarkers of clinical utility is a major barrier in cancer research. While the scientific literature annually reports many cancer biomarkers, few prove sufficiently resilient, reproducible, and real to translate into clinically useful assays. Consequently, the cancer research community is in need of a systematic approach to streamline the biomarker translation process from discovery to validation. We propose here a strategy to verify the biological signal of protein biomarker candidates. This biomarker verification step would provide a means to demonstrate that a biomarker’s discriminative ability is due to biology as opposed to chance or bias in the samples or technology platforms.

Such a pipeline requires three components: discovery, verification, and validation. NCI’s Clinical Proteomics Technology Assessment for Cancer (CPTAC) is providing support for each step of this strategy through various means. Within the discovery phase, CPTAC is developing metrics and reagents to assess the variability and reliability of technology platforms used to analyze candidate protein biomarkers. CPTAC is conducting analogous studies within the verification step in order to benchmark MRM-based platforms for high-throughput targeted molecules. Additionally, CPTAC is collecting clinical samples for verification studies. These samples were collected through a novel protocol that minimizes bias. Finally, to support the validation step, CPTAC is producing reagents critical to assay development. While still in process, CPTAC has made significant progress in all three pipeline components for a proteomics technology platform that targets protein-based biomarkers.

This verification strategy integrates the essential elements of biomarker development far ahead of clinical validation. Implementation of this pipeline will enhance the biomarker development process, shorten the timeline, and improve the success rate to bring validated cancer biomarkers into the clinic.

Path to Discovery

Goal - Provide an assessment of run-to-run and lab-to-lab variability for a complex mixture of proteins
- Provide a measure of the sensitivity of a mass-spectrometry platform with such a mixture

Materials - Yeast lysate stock with protein mixture (Sigma LPS)

Method - A carefully designed, multi-lab experiment using replicate analyses of the yeast lysate
- Over 40 performance metrics already available in instrument output files
- Through a team-based network, assess variability for a complex mixture of proteins

Results - Over 40 performance metrics, already available in instrument output files, may now be used to determine a discovery platform’s condition for a given experiment (below)
- A whole new level of quality assurance for discovery proteomics

Path to Verification

Goal - Through a team-based network, assess variability of the MRM analysis platform
- Determine the sensitivity of the MRM analysis platform
- Seven target proteins spiked into stock solution

Method - A carefully designed, multi-lab experiment using replicate analyses of yeast lysate

Results - Over 40 performance metrics, already available in instrument output files, may now be used to determine a discovery platform’s condition for a given experiment (below)
- A whole new level of quality assurance for discovery proteomics

Path to Validation

Goal - Supply research community with robust, highly characterized monoclonal antibodies to cancer-relevant targets

Acknowledgments

This work was supported by grants from the National Cancer Institute (NCI) (U01CA133945, U01CA126477, U01CA126479, U10CA132550, HG003063, and HG005023), part of NCI’s Clinical Proteomic Technologies for Cancer ([https://proteomics.cancer.gov]). A component of this initiative are the Clinical Proteomic Technology Assessment for Cancer (CPTAC) teams that include the Broad Institute of MIT and Harvard, Memorial Sloan-Kettering Cancer Center, Pacific Northwest National Laboratory, Proteome Institute of the Pacific, and Stanford University School of Medicine.

Dr. Paul Rudnick of NIST for analysis and visualization of CPTAC data

Biospecimen plan – phase I

CPTAC center biospecimen collection:
- Plasma samples for technology validation
- 5 collection sites
- Common cohort: pre-biopsy women with a scheduled breast biopsy
- Age: All racial/ethnic groups
- Resulting sample set with 25%/75% cancerous/benign

Clinical CDE Database (MedSciNet)

Clinical common date elements stored in MedSciNet database
- In phase II, samples will be shipped to proteomic analysis labs to test the ability of technology platforms to reproducibly detect cancer-relevant features.

Table for sample preparation performed at NIST:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Recovery (ng/mL)</th>
<th>Recovery (pmol/μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>GYSIFSYATK</td>
<td>2.05</td>
<td>3.1</td>
</tr>
<tr>
<td>PSA</td>
<td>3.05</td>
<td>4.9</td>
</tr>
<tr>
<td>IVGGWEC</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>MBP</td>
<td>3.05</td>
<td>4.9</td>
</tr>
<tr>
<td>HGFLPR</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>MYO</td>
<td>3.05</td>
<td>4.9</td>
</tr>
<tr>
<td>LFTGHPETLEK</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>APR</td>
<td>3.05</td>
<td>4.9</td>
</tr>
<tr>
<td>AGLC</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>R</td>
<td>3.05</td>
<td>4.9</td>
</tr>
<tr>
<td>LEPIINDISHTQSVSAK</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>MBP</td>
<td>1.05</td>
<td>1.7</td>
</tr>
<tr>
<td>YLASASTMDHAR</td>
<td>1.05</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Data available at:
- [http://cancer.gov](http://cancer.gov)
- [http://clinicaltrials.gov](http://clinicaltrials.gov)