

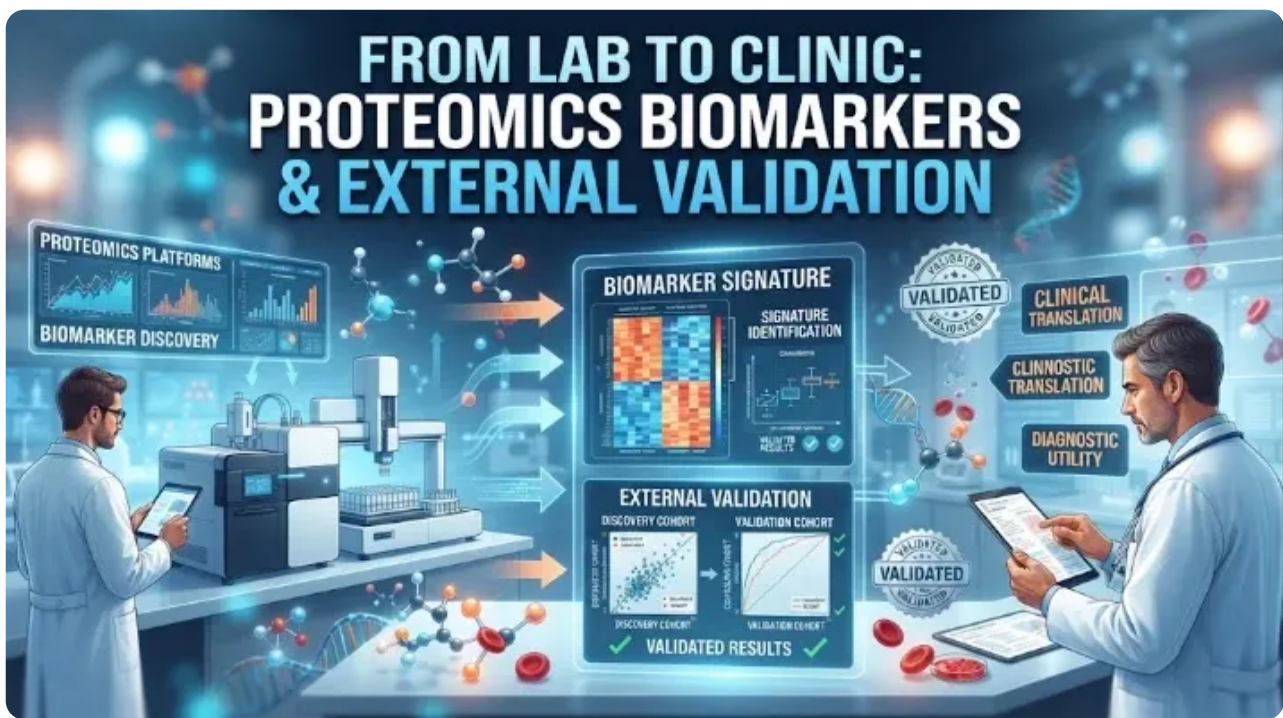
Proteomics Platforms for Biomarker Signatures in Clinical Translation Using External Validation

Sadia Tasmeem

Jul 26, 2026

Abstract

The article explains why proteomic biomarker discovery requires rigorous external validation before clinical translation. Proteins are biologically responsive and technically complex, with concentrations and modifications affected by disease state, treatment, handling, storage, and analytical platform. Mass spectrometry, affinity platforms, and programmable protein arrays such as NAPPA are discussed as complementary technologies for discovering candidate signatures, including tumor-associated autoantibodies. The article stresses that strong performance in a discovery cohort can reflect overfitting, batch effects, sampling bias, or unrealistic case–control contrasts. Credible translation therefore requires locked assays, independent populations, analytical validation, realistic clinical controls, calibration, confidence intervals, subgroup testing, and decision-focused evaluation. Orthogonal confirmation across measurement platforms can strengthen biological confidence. The article concludes that a useful proteomic signature must be reproducible, transferable, operationally feasible, and able to improve a defined clinical decision rather than merely produce an attractive statistical performance measure.



Although thousands of disease associated signals can be identified by proteomics, only a small number of these signals turn out to be clinically useful biomarkers. It's not only analytical sensitivity that is the problem. The problem with translation is that discovery samples are not representative, preanalytical variation is not controlled, models are overfit, or validation is done on cohorts that are too similar to the discovery cohorts. Thus, a platform strategy based on external validation from the outset rather than as a post hoc maneuver to add an appealing result is essential for a credible proteomic signature.

The measurement of proteins is a challenging task due to the variability of the protein composition of different tissues, different times, disease state, treatment, environment, etc. Proteins also range in concentration from very high to very low and can be cleaved, phosphorylated, glycosylated, oxidized and subject to other modifications. A genomic variant is relatively stable, while a protein signal may represent a biological condition when the sample was taken. That responsiveness is also clinically valuable and can be the source of confounding.

Mass spectrometry is a fundamental proteomic technology since it allows the identification and quantification of large numbers of peptides with high molecular specificity. Affinity based platforms are another option to measure multiple targets at scale with antibodies or engineered binding reagents. Protein microarrays can be used to detect binding events, such as patient autoantibodies to tumour associated antigens. The platforms represent different aspects of biology and the choice of platform should be based on the clinical question and not on fashion.

Nucleic Acid Programmable Protein Arrays illustrate a creative solution to one long standing protein array problem. NAPPA does not require purification and storage of thousands of proteins prior to printing, but instead prints DNA templates and capture reagents, and produces proteins in situ using cell free expression. Freshly expressed proteins can be then tested with patient serum. This design enables high throughput immunoproteomics, while also alleviating some of the storage and inconsistent purification yield issues.

The promise and challenge of autoantibody discovery platforms. Cancer can elicit humoral responses to altered or overexpressed proteins, resulting in antibody signals that can amplify relatively small quantities of tumor antigen. Programmable protein arrays have been used in studies to identify cancer panels for breast and ovarian cancers. However, many individual autoantibodies have limited sensitivity and panels can have variable performance when transferred from discovery sets to independent samples.

That is not a technical embarrassment—it's an important test of truth. Discovery models learn both biology and accidental properties of initial cohort. A number of hidden predictors may be introduced into the analysis, such as batch effects, sample handling, age distribution, medication effects, referral pathways, and disease prevalence. External validation poses the question: Is the signature retained when those accidental relationships change? Markers that are generalizable have a stronger case for the markers to be informative across institutions.

The validation must be external, it can't be self validation. Randomly splitting one biobank into training and test sets is very helpful for model development; however, it has nothing to do with the uncertainty of a new clinical environment. As of now, a more robust design will be binding on the assay, the preprocessing pipeline, the list of biomarkers, thresholds, and the statistical model, prior to evaluating an independent cohort. Ideally, the validation population should be geographically, pathologically, clinically, and laboratory workflow distinct from the intended use, but still representative of the intended use.

Sample selection is another part that needs to be considered. A biomarker for early detection of cancer should not be identified by comparing high grade, metastatic cancer with an unusual healthy cohort. This contrast can yield significant discrimination in early tumor biology but in inflammatory, cachectic, and/or treatment exposed or general physiological conditions. Benign conditions that mimic the signs or imaging findings should be included as clinically realistic controls. This complicates the validation process and makes it much more pertinent.

Analytical validations come before clinical claims. Researchers have to have evidence for precision, reproducibility, limit of detection, linearity, interference, stability and lot to lot consistency. If the ten protein signatures change from run to run or are strongly influenced by sample storage then they are not a usable clinical tool. The performance of the platform should be evaluated by the operators and laboratories when deploying beyond one research center. Instability in the measurement cannot be overcome by biological sophistication.

The same holds true for the statistical discipline. With high dimensional proteomic datasets there are numerous possibilities for false discovery, as the number of features measured can be significantly higher than the number of participants. Feature selection should be a part of the training process and should not be based on information from validation samples. Calibration should be evaluated in tandem with discrimination as a model can sort cases correctly but yield incorrect absolute probabilities. Headline performance metrics should be accompanied with confidence intervals.

A second question is raised by the clinical utility: What decision is changed as a result of the test? So, a biomarker signature for a pulmonary nodule, for instance, would only have a positive impact on the number of unnecessary invasive procedures if it would perform well in the patients who fall somewhere between the extremes of current risk assessment. If it is clear that you have a cancer problem or that you are in good health, the benefit of a test is fairly minimal. Decision curve analysis and prospective workflow studies can measure if the signature will enhance action or only a statistically significant area under a curve.

Comparison of platforms can improve the quality of translation. Protein Arrays can be used to identify a candidate which can then be verified by an orthogonal method, such as targeted mass spectrometry or immunoassay. The convergence of the various measurement principles decreases the risk of platform dependency. Discordance can also be informative, hinting at epitope effects, post translational modifications or assay interference. It is, therefore, no longer a quality control activity, rather, orthogonal confirmation should be considered as a mechanism.

The best way to grow is to be staged, and conservative. Discovery can be open and unrestricted, and verification can be more focused and can do this using a set of predetermined criteria. External validation should be with locked procedures and clinically representative samples. Prospective evaluation should then follow to determine if results are still accurate when the test is performed in real time in clinical practice with

the same missing data, comorbidities and workflow pressures.

The accuracy of the biomarkers also depends on reference standards. A cancer signature based on a histopathologic diagnosis has a different evidentiary base than a cancer signature based on a clinical diagnosis, which may be not be certain. Diagnostic studies should indicate the method of disease determination and if disease status was blinded to the proteomic results. Misclassifications of the reference standard can limit the range of the measurements of performance or produce false relationships.

After the statistical selection of the composition of a multianalyte panel, the composition should be given biological examination. A highly correlated set of proteins may contain some redundant information and add little robustness, while a small set of proteins that cover different mechanisms can lead to better robustness. Pathway mapping can demonstrate if a panel is a representation of tumor burden, host inflammation, tissue injury, or multiple process. This interpretation will help identify conditions which could produce similar signatures in the discovery cohort and thus cause confusion.

The costs and complexity of the assay are elements of translational evaluation. A model with hundreds of proteins can get a bit higher accuracy to the retrospective when compared to a compact panel, but is impractical for routine laboratory use. By comparing the performance with reduction, one can calculate how much of the performance is lost with the reduction. A more easily standardized, faster turnaround, smaller, locked assay with stable reagents may generate more value than a technologically wonderful platform that is not easily standardized.

External validation should be done using biobanks with sufficient documentation of collection history to allow for the identification of hidden batch structure. The inclusion of randomization of cases and controls across plates, technicians, and assay days helps to minimize technical variables aligning with diagnosis. Blind quality review can then detect anomalies without awareness of disease status, maintaining independence needed for validations of merit.

Proteomics offers unprecedented opportunities because proteins are closer to phenotype and treatment response. The nearness also implies that protein signals are context dependent. Successful clinical translation requires studied and designed platform capability, independent replication, analytical stability, and decision focused evaluation. The external validation step is not the end of a biomarker pipeline; it's the moment where a compelling signature finds its clinical acceptance. ☒