# Development and Validation of Five Idiopathic Pulmonary Fibrosis (IPF) Biomarker Panels Comprising 12 Immunoassays: Advancing Clinical Diagnostic, Pharmacodynamic and Disease Activity Assessments in IPF.

Eisinger, D.[1], Kemp, J.[1], Melin, J.[1], Nitka, K.[1], Bencher, R.[1], Rassa, J.[2], Seyda, A.[2], Hersey, S.[2], and Schafer, P.[2]

1. Rules-Based Medicine, A Q² Solutions Company, Austin, TX.
2. Bristol-Myers Squibb, Summit, NJ.

RULES BASED MEDICINE
a Q² Solutions Company

## Abstract

Multiple peer-reviewed publications from various groups consistently report a reoccurring set of blood-based protein biomarkers linked to clinical utility and the course of IPF disease progression. Herein, we describe the analytical validation of the 12 most relevant IPF biomarkers related to: epithelial damage (CYFRA 21-1, SP-D, CA-125, CA-19-9, KL-6), Fibrosis (MMP-7, TN-C, Periostin), Inflammation (PARC, BLC, ICAM-1), and Thrombosis (PAI-1). The 12 assays were optimized into three multiplex panels and two singleplex panels on the Luminex platform. The assays were analytically validated under formal protocols with design controls and pre-defined acceptance criteria. Results will be presented with respect to the Limit of Detection, Sensitivity, Accuracy, Precision, Parallelism, Matrix Interference, Freeze-Thaw Stability, Short-term Analyte Stability, and Sample Reproducibility with supporting data across multiple reagent lots to ensure consistency of results.

## Background

In collaboration with other organizations, The Pulmonary Fibrosis Foundation (PFF) created a consortium called PROLIFIC (Prognostic Lung Fibrosis Consortium) to develop well-qualified immunoassays to measure peripheral blood proteins in patients with IPF. The team of consortium biomarker experts settled on 12 blood-based proteins that have substantial evidence of prognostic or pharmacodynamic value. The scientific biomarker literature is replete with instances of different assays and platforms demonstrating a directional change of a particular analyte's concentration in relation to a disease state or treatment. However, it is most often impossible to make absolute comparisons of analyte levels between different assays and platforms that are not calibrated with reference standards or analytically compared in a robust manner. Drug development in IPF has been extremely challenging, and despite the approval of two pharmacological agents in the past ten years, there has been little progress in reducing mortality (Lederer et al.). A well-qualified, consistently manufactured set of biomarker assays for broad use by clinical researchers will accelerate therapeutic efforts in treating IPF.

## Material and Methods

All 12 immunoassays, either in singleplex or multiplex format, utilized the Luminex® xMAP® platform and consisted of antigen-specific antibodies optimized in a capture-sandwich format. All incubations take place at room temperature. Three-level quality controls were prepared using either human serum with the appropriate amount of endogenous analyte or by spiking analyte into human serum. Healthy human serum and EDTA plasma were collected from consented donors at RBM, whereas serum and EDTA plasma from IPF patients were procured from external biorepositories. All samples were stored at -80°C until assayed. Three levels of control (low, mid, and high), each plated in duplicate, were tested alongside samples in each run to determine acceptability of the assay. All reagents, materials, and equipment needed to perform the validation studies, except for sample collection materials, were used exclusively at the RBM single reference laboratory site. All equipment is maintained at RBM per Equipment Control SOPs.

## Results

Assay development, manufacturing, and validation was an iterative process divided into three phases. Phase 1 encompassed reagent selection, feasibility, and generation of a Pilot lot (Lot #1). During Phase 2 Development (Lot #2) and Verification (Lot #3), lots were made and rigorously tested for performance with design control principles. Phase 3 was the final manufacturing and validation of a large lot (Lot #4 or Validation lot). The 12 immunoassays comprising a given manufacturing lot are composed of five panels, with panels one through three being multiplex panels and panels four and five as singleplex assays.

### Table 1. LDD, LLOQ, ULOQ, and Dynamic Range - Validation Lot #4 (Sample Type: Serum/EDTA Plasma)

| Panel | Analyte | Units | Sample Dilution | LDD | LLOQ | ULOQ | Dynamic Range |
|---|---|---|---|---|---|---|---|
| IPF1 | ICAM-1 | ng/ml | 1:50 | 40 | 32 | 17,300 | 3.4 - 17,300 |
| | KL-6 | IU/ml | 1:50 | 0.80 | 1.7 | 2,450 | 0.49 – 2,450 |
| | PAI-1 | ng/ml | 1:50 | 2.6 | 1.5 | 1,680 | 0.34 – 1,680 |
| | PARC | ng/ml | 1:50 | 0.91 | 0.85 | 1,730 | 0.34 – 1,730 |
| | SP-D | ng/ml | 1:50 | 7.3 | 11 | 15,000 | 3.0 – 14,950 |
| | TN-C | ng/ml | 1:50 | 49 | 154 | 228,000 | 45 – 228,000 |
| IPF2 | CA-125 | IU/ml | 1:5 | 6.6 | 9.7 | 8,150 | 1.6 – 8,150 |
| | CYFRA 21-1 | ng/ml | 1:5 | 0.31 | 0.51 | 326 | 0.065 - 326 |
| IPF3 | BLC | pg/ml | 1:5 | 9.2 | 14 | 18,000 | 3.6 – 18,000 |
| | Periostin | ng/ml | 1:5 | 3.7 | 18 | 18,500 | 3.7 – 18,450 |
| IPF4 | MMP-7 | ng/ml | 1:10 | 0.031 | 0.050 | 153 | 0.031 - 153 |
| IPF5 | CA-19-9 | IU/ml | 1:5 | 1.83 | 2.1 | 2,940 | 0.59 – 2,940 |

- **Least Detectable Dose (LDD)** was determined by adding three standard deviations to the average of the signal for 20 replicate determinations of the standard curve blank. This value is converted to concentration as interpolated from the standard curve and multiplied by the dilution factor used for testing samples for the matrix of interest.
- **Lower Limit of Quantitation (LLOQ)** was determined by performing serial dilutions of standard for a total of eight dilutions to be tested in triplicate over three runs. The percent Coefficient of Variation (%CV) is calculated for each of the dilution replicates. The LLOQ was determined as the concentration at which the %CV for the dilution is 30%.
- **Upper Limit of Quantitation (ULOQ)** was determined by running standards in triplicate over three runs. The %CV was calculated for each standard, and the ULOQ was determined as the highest concentration at which the %CV is 30% and multiplied by the dilution factor used for testing samples for the matrix of interest. Values of samples above the ULOQ are reported as >ULOQ.
- **Dynamic range** is defined as the range of standard used to produce the standard curve multiplied by the dilution factor used for testing samples.

### Table 2. Summary of Analytical Performance: Validation (Lot #4)

| Validation Parameter | Acceptance Criteria | | ICAM-1 | KL-6 | PAI-1 | PARC | SP-D | TN-C | CA-125 | CYFRA 21-1 | BLC | Periostin | MMP-7 | CA-19-9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Control Accuracy | ≤ 20% different from the nominal value for each control level average | Level 1 | -2% | 0% | -2% | -2% | -2% | -1% | -1% | 3% | 0% | 0% | -1% | -7% |
| | | Level 2 | 1% | 3% | 1% | 1% | 1% | -3% | 6% | 6% | 0% | 0% | 0% | -3% |
| | | Level 3 | -2% | -2% | 0% | -1% | 0% | 0% | 4% | 4% | 0% | 0% | 0% | -2% |
| Control Inter-Assay Precision | Inter-assay %CV of ≤ 25% for Level 1 and ≤ 20% for Level 2 and Level 3 | Level 1 | 6% | 9% | 7% | 7% | 8% | 5% | 13% | 11% | 12% | 19% | 13% | 9% |
| | | Level 2 | 9% | 7% | 7% | 5% | 12% | 11% | 9% | 9% | 10% | 20% | 9% | 6% |
| | | Level 3 | 12% | 10% | 8% | 7% | 9% | 6% | 10% | 10% | 11% | 11% | 9% | 6% |
| Freeze-Thaw Stability | 70-130% recovery for at least two out of three samples | | All analytes are acceptable up to five freeze-thaws | | | | | | | | | | | |
| Matrix Interference | Average percent recovery between 70-130% | | Tested interferents (bilirubin, hemoglobin, triglyceride) did not interfere with results | | | | | | | | | | | |
| Parallelism | Average percent recovery between 70-130% for at least two out of three samples | | All analytes are acceptable | | | | | | | | | | | |
| Sample Reproducibility | The %CV of six out of nine samples must be ≤ 20% | | All analytes are acceptable | | | | | | | | | | | |
| Short-term Analyte Stability | 70-130% recovery for at least two out of three samples | | All analytes are acceptable | | | | | | | | | | | |
| 24-Month Real Time Stability | Four out of six of the control results must be within 2.5 standard deviations of the established control values for the assay to pass acceptance | | Ongoing | | | | | | | | | | | |

- **Control Precision and Accuracy.** Control precision was determined by measuring all levels of controls in replicate over a minimum of three days using at least nine runs with at least two operators and at least two instruments. Precision samples provide the CV within (Intra-assay) and between (Inter-assay) multiple runs of an assay. The acceptance criteria for precision are an Inter-assay and Intra-assay %CV of < 25% for Level 1 and < 20% for Level 2 and Level 3. The acceptance criterion of accuracy is a % relative error (RE) of < 20% difference from the nominal value for each level of control.
- **Freeze-Thaw Stability** was determined by measuring the stability of samples for up to five freeze-thaw cycles. Percent recovery is calculated by comparing the results of samples that have undergone each freeze-thaw event to the corresponding freshly thawed samples. The acceptance criteria are percent recoveries within 70-130% for at least two out of three samples.
- **Matrix Interference** was determined by spiking hemoglobin, bilirubin, and triglyceride into samples and determining the percent recovery of the analyte of interest. The acceptance criterion is a percent recovery within 70-130%.
- **Parallelism** is measured by the ability to obtain results proportional to the analyte concentration in a sample when serially diluted 1:2 from the MRD (minimal required dilution). It was determined for three human EDTA plasma and three human serum samples. The % recovery was calculated as observed vs. expected concentration. The acceptance criterion is an average percent recovery between 70-130% for at least two out of three samples.
- **Sample Reproducibility** is the assessment of the assay's ability to reproduce a consistent measurement of a sample. It was determined by measuring nine serum and nine EDTA plasma samples, plated in duplicate, over a minimum of three days using at least five runs, two operators, and two instruments. The acceptance criterion is %CV < 20% for six out of nine samples.
- **Short-Term Analyte Stability** was determined by storing samples at ambient temperature and 2-8°C for up to 24 hours. Percent recovery is calculated by comparing the results of serum and plasma samples that have been stored at non-frozen temperatures to corresponding freshly thawed samples. The acceptance criteria are percent recoveries within 70-130% for at least two out of three samples.
- **24-Month Real Time Stability** is currently being performed to determine the amount of time the kit components can be left at the intended storage condition before the assay produces inaccurate results. Three levels of control, each plated in duplicate, will be tested on kits that have been stored for six, 12, 18, and 24 months.

### Table 3. Cross-Reactivity for Multiplex Panels IPF1, IPF2, and IPF3: Average Percent Cross-Reactivity Between Single Antigen and Other Assays

| IPF1 Simplex Standard (Ag) | ICAM-1 | KL-6 | PAI-1 | PARC | SP-D | TN-C |
|---|---|---|---|---|---|---|
| ICAM-1 | 100.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| KL-6 | 0.0% | 100.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| PAI-1 | 0.0% | 0.0% | 100.0% | 0.0% | 0.0% | 0.0% |
| PARC | 0.0% | 0.0% | 0.0% | 100.0% | 0.0% | 0.0% |
| SP-D | 0.4% | 0.5% | 0.0% | 0.0% | 100.0% | 0.5% |
| TN-C | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 100.0% |

| IPF2 Simplex Standard (Ag) | CA-125 | CYFRA-21-1 |
|---|---|---|
| CA-125 | 100.0% | 0.0% |
| CYFRA-21-1 | 0.0% | 100.0% |

| IPF3 Simplex Standard (Ag) | BLC | Periostin |
|---|---|---|
| BLC | 100.0% | 0.0% |
| Periostin | 0.0% | 100.0% |

Cross-reactivity is the ability of an assay to differentiate and quantify the analyte of interest in the presence of other analytes in the sample. It is determined by testing high concentrations of each single standard in the multiplex assay. Cross-reactivity is calculated by subtracting the standard curve diluent results from each of the simplex standard results, then dividing the simplex standard result by the actual standard result and multiplying by 100. Assays are accepted if the average S8 and S5 standard cross-reactivity is <10%.

### Table 4. Assay Concentration Ranges and Detectability (Lot #4)

| Validation (Lot 4) | ICAM-1 | KL-6 | PAI-1 | PARC | SP-D | TN-C | CA-125 | CYFRA 21-1 | BLC | Periostin | MMP-7 | CA-19-9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LLOQ | 32 | 1.7 | 1.5 | 0.85 | 11 | 154 | 9.7 | 0.51 | 14 | 18 | 0.05 | 2.1 |
| Min Result | 88 | 2.3 | 4.6 | 11 | 11 | 174 | 6.7 | 0.60 | 30 | 34 | 0.34 | 3.4 |
| Max Result | 760 | 96 | 316 | 213 | 637 | 2360 | 121 | 15 | 573 | 530 | 9.4 | 108 |
| Average Concentration | 323 | 15 | 99 | 56 | 124 | 763 | 24 | 2.7 | 106 | 222 | 2.0 | 15 |
| Total # of Data Points | 186 | | | | | | 279 | | 285 | | 252 | 214 |
| % Data Points Detectable Above LLOQ | 100% | 99% | 100% | 100% | 98% | 99% | 85% | 96% | 92% | 99% | 100% | 75% |
| # of IPF Data Points | 95 | | | | | | 121 | | 126 | | 64 | 71 |
| % IPF Data Points Detectable Above LLOQ | 100% | 100% | 100% | 100% | 100% | 100% | 95% | 100% | 95% | 100% | 100% | 85% |
| # of Normal Data Points | 91 | | | | | | 158 | | 159 | | 188 | 143 |
| % Normal Data Points Detectable Above LLOQ | 100% | 99% | 100% | 100% | 97% | 99% | 76% | 92% | 89% | 97% | 100% | 71% |

All assays detect ≥97% for both healthy and IPF samples, except for CA-125 and CA-19-9. For the CA-125 assay, most samples that fell below the LLOQ were normal, healthy samples. 80% of IPF samples were detectable, which is similar to the development lot detectability. For the CA-19-9 assay, most samples that fell below the LLOQ were normal, healthy samples. 86% of IPF samples were detectable, which is similar to the pilot and development lot detectability.

### Table 5. Kit Qualification: Reproducibility Between Different Lots

| Kit Qualification Summary | | ICAM-1 | KL-6 | PAI-1 | PARC | SP-D | TN-C | CA-125 | CYFRA 21-1 | BLC | Periostin | MMP-7 | CA-19-9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pilot Lot #1 vs. Development Lot #2 | Sample Correlation - R(must be ≥ 0.8) | 0.97 | 0.95 | 0.99 | 0.97 | 0.98 | 0.97 | 0.96 | 0.97 | 0.99 | 0.94 | 0.96 | 0.99 |
| | Sample Slope - post-adjustment(must be between 0.9 - 1.1) | 1.0 | 1.0 | 1.0 | 1.0 | 0.99 | 0.99 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.98 |
| Pilot Lot #1 vs. Verification Lot #3 | Sample Correlation - R(must be ≥ 0.8) | 0.96 | 0.97 | 0.96 | 0.96 | 0.97 | 0.95 | 0.95 | 0.97 | 0.99 | 0.95 | 0.97 | 0.99 |
| | Sample Slope - post-adjustment(must be between 0.9 - 1.1) | 0.99 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.99 | 0.99 | 1.0 |
| Pilot Lot #1 vs. Validation Lot #4 | Sample Correlation - R(must be ≥ 0.8) | 0.98 | 0.98 | 0.97 | 0.97 | 0.99 | 0.97 | 0.94 | 0.97 | 0.98 | 0.90 | 0.99 | 0.99 |
| | Sample Slope - post-adjustment(must be between 0.9 - 1.1) | 0.99 | 1.0 | 0.99 | 0.99 | 0.98 | 1.0 | 1.0 | 0.98 | 1.0 | 1.0 | 1.0 | 1.0 |

Kit qualification is the process of determining the relationship between two groupings of reagents (kits), where one kit has been previously accepted, and the other kit is being evaluated for acceptability. The process includes performing assays, generating slopes that demonstrate the relationship between the kits, calculating new standard values, and reviewing the outcome for acceptability. The final slope average must be between 0.9 – 1.1 for all analytes for the kit to pass acceptance.

## Conclusions

The five IPF panels comprising 12 blood-based protein analytes meet the acceptance parameters described.

## References:

Lederer et al. (2018) *N Engl J Med* 378:1811-1823. DOI: 10.1056/NEJMra1705751