Development and validation of customized PEA biomarker panels with clinical utility

Introduction

Several studies have demonstrated the value of proteomics in identifying relevant biomarkers for non-congenital diseases, and to monitor disease progression, treatment responses, and efficacy. While the genome can largely be considered as static within an individual (at least in terms of DNA sequence), the proteome varies considerably in response to a wide range of physiological and pathological processes, making it a more immediate barometer of the state of the body. Proteomics is important because proteins represent the functional effector molecules in the cell. Drugs elicit their effects most often through interactions with proteins. Gene expression analyses have discovered numerous genes that are differentially expressed between malignant and benign tissues (*1*), but few have proven to be suitable as biomarkers, mainly because the mRNA levels do not always correlate well with protein abundance (*2*).

A higher power of discrimination can be obtained by combining more than one biomarker (*3-6*) to create a protein signature. However, large-scale studies of protein levels have been hampered by a lack of high-throughput methods. Using protein signatures could transform the future of disease diagnosis, treatment, and our understanding of health, but only if these technical limitations can be overcome using improved proteomics technologies. Recent advances such as Olink Proteomics' Proximity Extension Assay (PEA) technology (*7*) have enabled highly multiplexed analysis using many protein biomarker assays. This has allowed scientists to cast a wider net for new protein signatures that can be used to stratify patients, predict disease and treatment outcomes, and understand pathophysiology or discover new drug targets (*8*).

Standard Olink panels enable the simultaneous relative quantification of 92 proteins, using only 1 µL of sample. The assays have been thoroughly validated and panel composition is designed to focus on specific diseases or biological processes, and is optimized for the expected dynamic range of the target protein concentrations in clinical samples. To tailor healthcare to individual patients, the next step is to transpose protein signatures from discovery into the clinic using low-plex custom panels. To meet these needs and to support flexible and efficient customization, we have developed a new PEA protocol using an optimized molecular design and new conjugation chemistry, creating an assay in which high- and low-abundant analytes could be combined into a single panel. In the proof-of-principle study described below, calibrators were used for normalization and to enable absolute quantification of the measured proteins.

Background

In an on-going project, more than 400 proteins were screened as potential markers to identify different diseases or disease stages, using standard Olink 92-plex panels. From the candidate biomarkers identified during this screening phase, 19 were selected to build a custom panel using the new PEA protocol. These customized panels were then used to verify the initial findings in additional patient samples. In the study presented here, we have scrutinized the technical performance of the 19-plex panel with the aim of guiding a decision on whether the technology is fit-for-purpose for future clinical utility, such as in early-phase clinical trials. The study design was inspired by a white paper generated by the AAPS Biomarker Discussion group (*9*).

Transition from screening to verification

Scalability

A clinically useful biomarker signature may include both high- and low-abundant proteins. If assays for high-abundant proteins that normally require pre-diluted samples are combined with those for lower abundant proteins that require undiluted samples, the dynamic

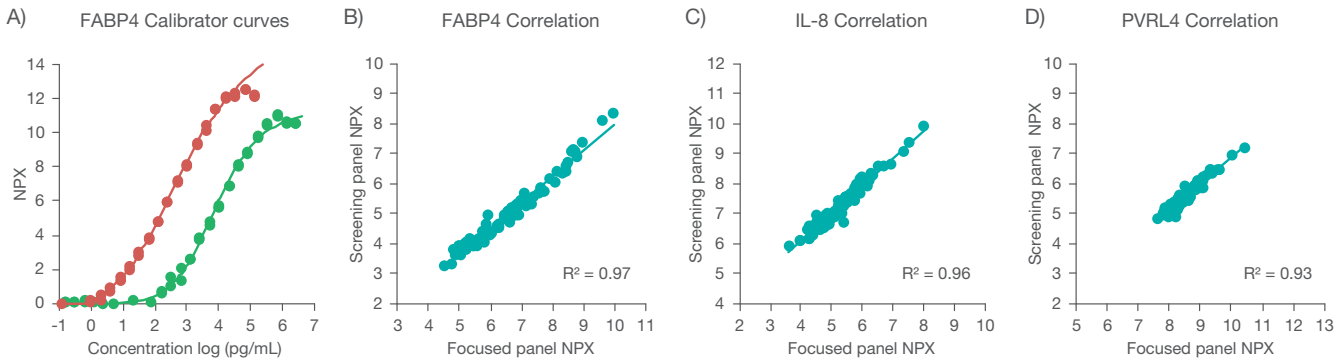


Fig 1. The proximity extension assays show good scalability, enabling transition from screening to verification. (A) FABP4 standard curves before (red) and after (green) shifting the dynamic range to be able to combine assays for high- and low-abundant targets; (B-D) Plasma samples were analyzed with either the 19-plex custom panel or the corresponding 92-plex screening panels. Scatter plots show the correlation between NPX values derived from the different analyzes for FABP4, IL-8, and PVRL4, respectively.

range must be shifted to cover the endogenous concentrations. In the current panel, an assay for FABP4 that normally uses samples diluted 100-fold was included, and therefore required dynamic range optimization. Figure 1A demonstrates the ~10-fold shift of the dynamic range for the FABP4 assay from the original 92-plex screening panel to the 19-plex panel. The other assays targeted were: CDH3, CPE, Dkk-4, EN-RAGE, FABP4, FGF-23, FR-alpha, Gal-1, IL-10, IL-17C, IL-8, KLK11, MK, MMP-7, NTRK3, PARP-1, PRSS8, PVRL4, and SOD2.

To verify the results from the screening phase, a set of 80 samples were analyzed both with the 19-plex panel and the corresponding screening panels. Relative quantification, NPX (Normalized Protein eXpression) values were compared among the different analyses, plotted, and correlated. Figure 1B-D shows example plots for FABP4, IL-8, and PVRL4. Note the high correlation obtained for FAPB4 even after applying a shift in dynamic range on top of other protocol changes. The average coefficient of determination (R^2) across the 19 assays was 0.90. The lowest correlation (R^2 = 0.73) was observed for the SOD2 assay, reflecting that the samples tested were within a narrow range (< 2 NPX). These data demonstrate that PEA assays can be combined in different configurations and panel formats while maintaining the same performance. This supports previous findings demonstrating the high scalability of the PEA technique (*7*). Together, these findings demonstrate the suitability of the technique for developing more focused panels for verification of screening data.

Sensitivity

Owing to the qPCR readout utilized in PEA, high sensitivity and wide dynamic range can be obtained. Standard curves were generated in multiplex for all assays included in the panel, using recombinant antigens for all assays. All assays were analyzed as triplicate measurements in two consecutive experiments. Data were normalized and a four-parameter logistic regression (4-PL) curve fitting was applied. The EN-RAGE standard curve had a slight plateau in the middle, possibly due to variable multimerization levels (10), and was therefore unsuitable for absolute quantification and excluded from further analyses. A 4PL non-linear curve fitting was applied to all data points (2 runs, 2 replicates, and 30 concentrations at 2-fold dilutions) and used to determine LOD, LLOQ, and ULOQ for each assay (Fig 2; data not shown).

Each run also included triplicate measurements of four different calibrators (Hi, Mid, Low, and Blank) for normalization and absolute quantification, as well as four control sample pools (healthy individuals or patients with different disease types) for precision calculations. Overall, the assays gave precise measurements and high sensitivity. The most sensitive assay in this panel was IL-8 with an LOD and LLQQ of 30 fg/mL. Another four assays (IL-10, FR-alpha, PARP-1, and PVRL4) exhibited very high sensitivity with LLOQ ≤ 1 pg/mL. The median LOD and LLOQ were 7.6 pg/mL and 11.4 pg/mL, respectively. Figure 2 displays the antigen standard curves for all assays.

Dynamic range

To visualize if the dynamic range was sufficient to accurately quantify proteins in a clinical context, a set of 36 relevant samples was analyzed, quantified, and plotted in the context of LOQ. Figure 3 shows the distribution of the three different groups of plasma samples, with the upper and lower LOQ indicated. The dynamic range obtained covered samples from both healthy and diseased subjects very well, with only a few exceptions. The mean log10 range was 3.7 (2.1–5.4). IL-8 had the widest dynamic range of all assays with a log10 range of 5.4. Altogether, the assays spanned 6.4 logs in concentration from the lowest to the highest

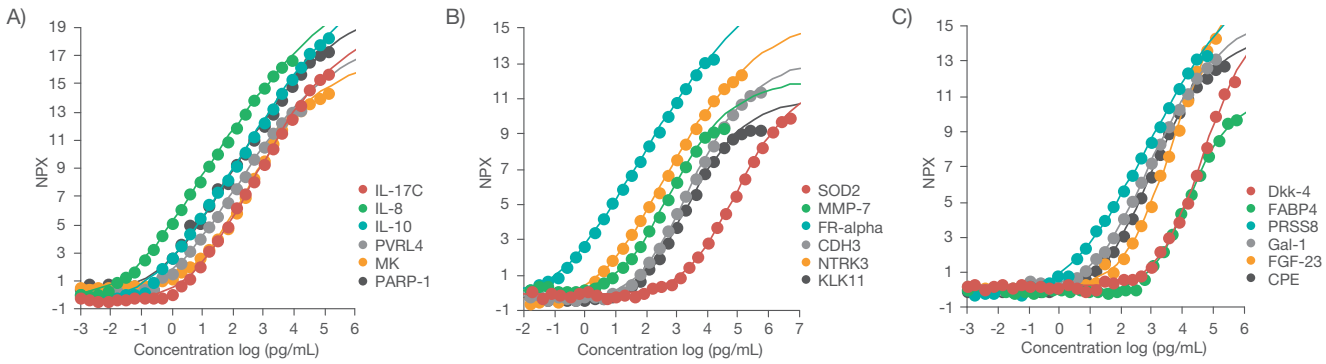


Fig 2. Antigen standard curves were generated for all 18 assays and analyzed in multiplex. Data were normalized against an internal control assay and calibrators, and curve-fitting was performed using 4-PL. Circles show the average of all data points from two runs and lines show the fitted curve.

control sample. The shifting of the dynamic range for FABP4 proved to be successful, resulting in samples being distributed around the middle of the dynamic range (second assay from the left).

Further assessment of suitability for clinical utility

Matrix interference

Several plasma components are known to potentially interfere with immunoassays. In a second study, the potential impacts of bilirubin, lipids, and hemolysate were evaluated at different spiked concentrations of analytes. These additions represent different patient health conditions and/or sample collection irregularities. For all assays, bilirubin and lipids could be added to concentrations corresponding to at least 8 (630 mg/mL) or 10 times (20 mg/mL) normal values (*11, 12*), respectively,

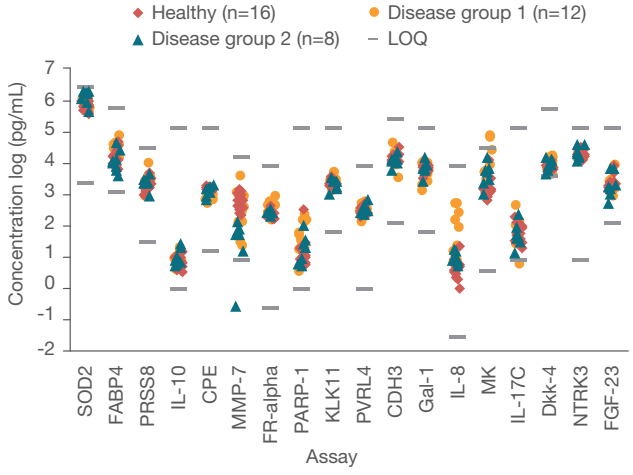


Fig 3. Dynamic range and sample distribution. Gray lines: ULOQ and LLOQ derived from the sensitivity experiments. Circles: Concentration values determined for plasma samples derived from three different group of individuals.

without disturbing assay performance (data not shown). In two out of 20 assays (PARP-1 and IL-8), a slight signal increase was observed by the addition of high concentration (15 and 7.5 mg/mL, respectively) of hemolysate (calculated from the hemoglobin level in the original blood sample). A concentration of 15 mg/mL of hemolysate represents 10% hemolysis of a sample. The reason for this is most likely due to analyte-specific leakage from disrupted erythrocytes rather than technical interference. These data were in good agreement with the results obtained for corresponding assays using Olink's screening panels (*13*).

Cross-reactivity

Cross-reactive events are a common problem for multiplex immunoassays, such as sandwich ELISA. The dual-recognition, DNA-coupled readout provided by PEA, however, can overcome this and provides exceptional specificity even at high multiplexing levels (*7*): Multiple blocking reagents are also included in the immunoassay step to avoid nonspecific binding. PEA probes are designed for pair-wise hybridization and detection requires double recognition of two specific primers to give a signal. This degree of specificity is a hallmark of PEA. However, since the current panel is built on a new PEA design, including entirely new DNA sequences, cross-talk was reassessed by testing each assay for recognition of the other 18 antigens, all at endogenous concentrations. Only one assay (SOD2) gave rise to a weak signal with non-significant contribution to the specific signal (0.5%; data not shown). This study confirmed that the new PEA protocol did not bring about any unexpected cross-talk events.

Previous data have demonstrated that PEA could distinguish between human and corresponding chicken proteins (*14*). However, homologous human proteins have not been studied in the context of cross-reactivity. In a second approach, a set of highly homologous

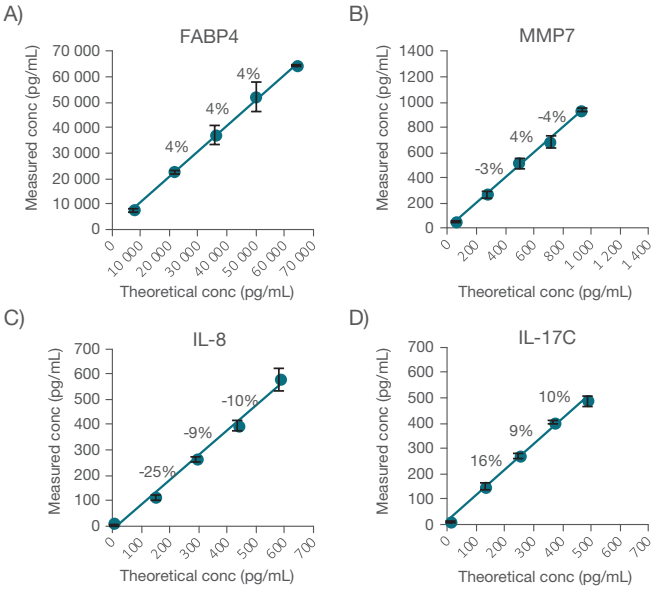
Table 1. Cross-reactivity assessment using homologous proteins demonstrating a striking high-level specificity of the PEA. Some recognition of FR-beta by FR-alpha was observed, although at a non-significant level in plasma.

Assay	Related protein	Coverage (%)	Identity (%)	Cross-reactivity (%)
FAPB4	FABP9	99	64	0.0
FR-alpha	FR-beta	87	77	0.1
KLK11	KLK8	90	49	0.0
EN-RAGE (S100A12)	S100P	98	45	0.0
CDH3	CDH1	97	54	0.0
CDH3	CDH2	87	46	0.0
CDH3	CDH4	87	45	0.0
MK	PTN	71	51	0.0
Dkk-4	Dkk-3	90	27	0.0

proteins were used to search for cross-reactive recognition of related proteins and to further challenge specificity. Homologous proteins from Olink's antigen library (n ≈ 1500) were included if they had an amino acid sequence coverage ≥ 90% and/or identity ≥ 50% according to Protein BLAST (Table 1). Endogenous levels (*15;* and data not shown) were used for both specific and homologous proteins as indicated in the table. Despite testing the most related proteins, cross-reactivity was not observed. FR-alpha showed some recognition of its highly homologous relative FR-beta (77% identity and 87% coverage), although at a non-significant level in plasma (0.1%). This systematic approach demonstrated that the assays can distinguish between very similar human proteins, and yet again highlights the high specificity of PEA.

Linearity

Linearity of dilution is often assessed by diluting a native sample with assay buffer, and thereby determining the minimum required dilution (MRD). However, in the current 21-plex protocol, crude samples were analyzed so that linearity was studied under true matrix conditions. This was done by mixing a sample containing a relatively high endogenous level of the protein analyte with a sample containing a low level at different ratios, to give 5 equally spaced concentrations (*16*). Native samples were chosen to give as wide a range as possible, requiring several different sample combinations to be included in the test, all depending on the endogenous concentrations. Analysis was performed by calculating the expected concentrations of the three intermediate points (based on the highest and lowest samples) and plotting the measured concentration against the expected (theoretical) concentrations. Figure 4A-D shows the results for four assays, and Figure 4E presents the R^2, range (fold-change: high/low) and maximum accuracy for all assays. Assays were linear in general (both at high and low ranges), with somewhat poorer accuracy observed for the MK assay, for which where the highest sample concentration was close to the estimated ULOQ.



E)

Assay	Range (max/min)	Average CV (%)	R^2	Max (abs) accuracy (%)
SOD2	2.1	6	0.95	13
FABP4	8.4	6	1.00	4
PRSS8	10.0	11	1.00	5
IL-10	6.5	12	0.97	18
CPE	2.1	4	0.95	12
MMP7	16.1	7	1.00	4
FR-alpha	1.7	10	0.97	8
PARP-1	24.4	8	1.00	11
KLK11	2.8	11	1.00	5
PVRL4	3.7	6	0.98	11
CDH3	9.1	6	0.96	24
Gal-1	7.3	4	0.99	8
IL-8	98.5	7	0.99	25
MK	9.1	4	0.95	38
IL-17C	47.1	6	0.99	16
Dkk-4	3.5	8	0.99	9
NTRK3	1.6	6	0.82	16
FGF-23	9.2	3	0.99	15

Fig 4. Examples and results from linearity of dilution study. Samples with high endogenous levels were diluted with samples with low levels at different ratios and quantified. (A-D) Graphs display measured and theoretical concentrations. Bars indicate standard deviation and the percentage represents accuracy for each point. (E) Table present the results for all 18 assays as range of concentrations tested, average CV% for the replicates, and R^2 for the regression, and maximum absolute accuracy.

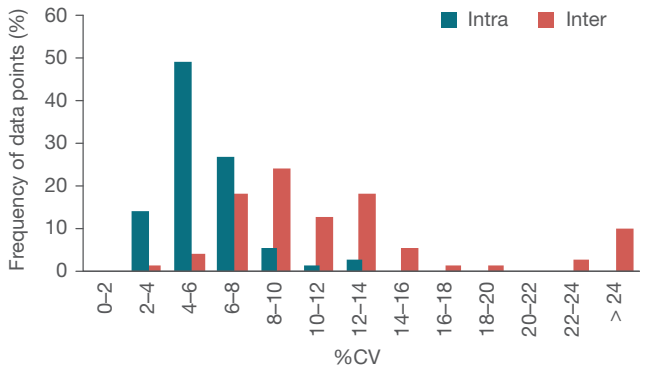


Fig 5. Intra- and interassay precision was determined from reference samples included in triplicates on 8 different runs. Histograms show the distribution of variation for all data points.

Precision

Precision is a key parameter to study, since this will directly dictate the size of both screening and clinical studies. As mentioned above, triplicate measurements were performed using four reference sample pools on each plate and these were used for precision calculations. Calculations were based on eight runs performed by two different operators. Average values from within each run were used for interassay CV calculations. Intraassay CV varied from 4% to 8% and interassay CV between 1% and 20% for the different assay. Average intra- and interassay CV were 5.8% and 13.5%, respectively, which is in parity with many commercial single ELISA's. These data can also be compared to the 5–10 CV% and 10–15 CV% obtained with Olink's commercial 92-plex screening panels. Figure 5 displays the fraction of assays/samples giving different CV% levels.

Stability

A robust diagnostic test requires that both biomarkers and reagents are stable throughout the protocol and are preferably also resistant to suboptimal handling. In a stability study, both whole-kit and reference samples were exposed to either 3 or 5 cycles of freeze-thawing, or storage at room temperature for 24 hours to assess stability compared to untreated references. Samples and kits were found to be insensitive (criteria: ± 30% deviation) to both freeze-thawing and RT storage. The exceptions were MMP-7, which was sensitive to RT storage both in plasma and in buffer (calibrator), and MK, which showed a decrease (38%) in calibrator signal after 24 h at RT (data not shown). As a follow-up, MMP-7 and MK will be studied along with the other biomarkers in both short-term and long-term storage studies to better define their utility as potential biomarkers.

Conclusions

The technical verification studies described here, performed on a custom 19-plex PEA panel, demonstrate that multiplex PEA is a highly scalable technique that is compatible for both screening and verification studies.

Several key immunoassay parameters, including sensitivity, dynamic range, specificity, linearity, precision, and stability, have been studied, and the results presented here will serve as a guide in determining whether PEA is fit for your research or clinical purposes.

To move further along the path towards clinical decision making and *in vitro* diagnostics, long-term reagent supply and readout platforms are also factors under active investigation. To those ends, PEA assays are being developed using antibodies developed in-house by Olink, and agnostic readout using standard qPCR machines is also being explored. Both these important developments show great promise and early results indicate that compatibility with PEA methodology is excellent.

References

1. J. D. Brooks *et al.*, *Genome Res.* **22**, 183 (2012).
2. C. Vogel *et al.*, *Nat. Rev. Genet.* **13**, 227 (2012).
3. G. Mor *et al.*, *PNAS* **102**, 7677 (2005).
4. T. Xiao *et al.*, *Mol. Cell Proteomics* **4**, 1480 (2005).
5. R. E. Brand *et al.*, *Clin. Cancer Res.* 17, 805 (2011).
6. S. T. Chang *et al.*, *J. Transl. Med.* **7**, 105 (2009).
7. E. Assarsson *et al.*, *PLOS ONE* **22**, 9 (2014).
8. www.olink.com/publications
9. D. Jani *et al*., *AAPS J.* **18**, 1 (2016).
10. O. V. Moroz *et al*., *BMC Biochem*. **10**, 11 (2009).
11. http://emedicine.medscape.com/article/2074115-overview
12. http://www.nlm.nih.gov/medlineplus/ency/article/003479.htm
13. http://www.olink.com/products/document-download-center/
14. M. Lundberg *et al*., *Mol. Cell Proteomics* **10**, M110.004978 (2011).
15. http://plasmaproteomedatabase.org
16. EP06-A. Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach, 1st Edition. Clinical and Laboratory Standards Institute

For research use only. Not for use in diagnostic procedures.
This product includes a license for non-commercial use. Commercial users may require additional licenses. Please contact Olink Proteomics AB for details.
There are no warranties, expressed or implied, which extend beyond this description. Olink Proteomics AB is not liable for property damage, personal injury, or economic loss caused by this product.
Olink® is a registered trademark of Olink Proteomics AB.
© 2017 Olink Proteomics AB. All third party trademarks are the property of their respective owners.

1083, v1.0, 2017-12-14

Olink Proteomics
Dag Hammarskjölds v. 52B
SE-752 37 Uppsala, Sweden
www.olink.com