

Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis

Authors

Lotta Wik, Niklas Nordberg, John Broberg, Johan Björkesten, Erika Assarsson, Sara Henriksson, Ida Grundberg, Erik Pettersson, Christina Westerberg, Elin Liljeroth, Adam Falck, and Martin Lundberg

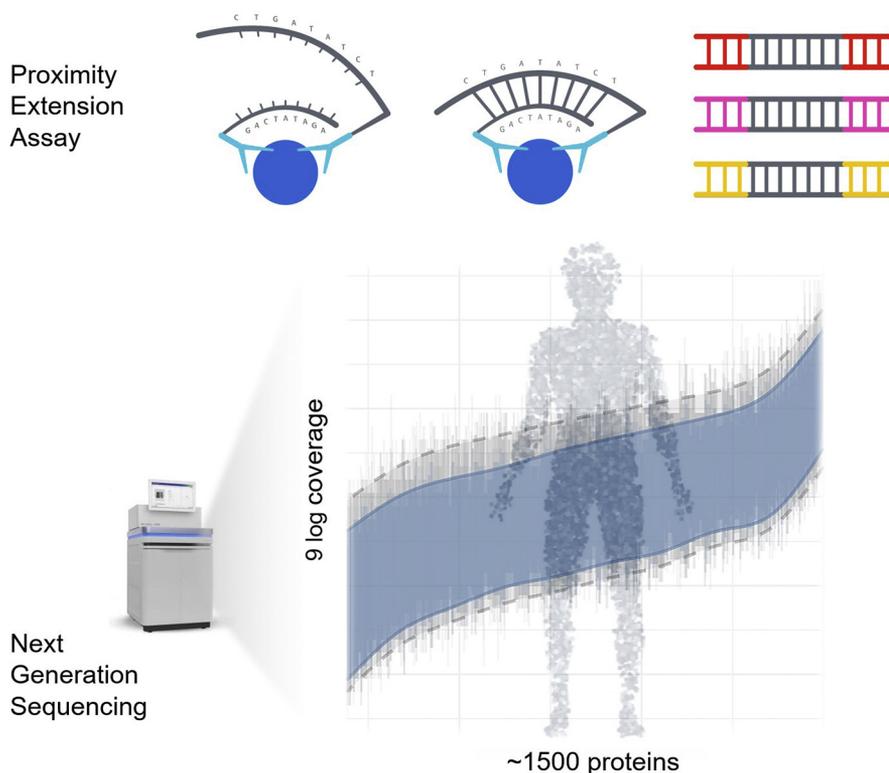
Correspondence

lotta.wik@olink.com

In Brief

Wik *et al.* presents a groundbreaking method based on the PEA technology in combination with NGS readout to measure nearly 1500 validated proteins in parallel. The protocol contains an innovative and automated sample preparation workflow generating close to 150,000 data points per run while consuming minimal amount of sample. Using dual DNA tags as readout enables high multiplex detection with exceptional specificity. This protocol meets all the requirements for a proteomic tool of the 21st century healthcare.

Graphical Abstract



Highlights

- Novel tool for protein discovery and quantification.
- PEA technology in combination with high-throughput sequencing read-out.
- Parallel measurement of nearly 1500 proteins.
- Innovative, miniaturized, and automated sample preparation protocol.

Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis

Lotta Wik^{*}, Niklas Nordberg, John Broberg, Johan Björkesten, Erika Assarsson, Sara Henriksson, Ida Grundberg, Erik Pettersson, Christina Westerberg, Elin Liljeroth, Adam Falck, and Martin Lundberg

Understanding the dynamics of the human proteome is crucial for developing biomarkers to be used as measurable indicators for disease severity and progression, patient stratification, and drug development. The Proximity Extension Assay (PEA) is a technology that translates protein information into actionable knowledge by linking protein-specific antibodies to DNA-encoded tags. In this report we demonstrate how we have combined the unique PEA technology with an innovative and automated sample preparation and high-throughput sequencing readout enabling parallel measurement of nearly 1500 proteins in 96 samples generating close to 150,000 data points per run. This advancement will have a major impact on the discovery of new biomarkers for disease prediction and prognosis and contribute to the development of the rapidly evolving fields of wellness monitoring and precision medicine.

Proteins are described as the building blocks of life and are required for the structure and function of all cells in the body. They are the main targets in drug development and in diagnostic testing and are commonly monitored over time as they represent the interaction between phenotype and environmental and lifestyle factors. Proteins can be used as strong predictors for diseases as well as for patient stratification based on disease subtyping, and they may also act as surrogate markers in clinical trials to predict clinically meaningful endpoints. However, proteins are far more complex to measure than DNA, and most current proteomics technologies have failed to deliver on critical performance parameters such as specificity, sensitivity, throughput, and dynamic range. The dynamic range of protein concentration in plasma can span more than ten orders of magnitude and is one of the greatest challenges in analyzing the plasma proteome (1, 2). The targeted approach of immunoassays has multiple advantages over untargeted approaches such as mass spectrometry, the most important of these being assay sensitivity, throughput, and reproducibility following sample preparation (3). As mass

spectrometry bias in favor of highly abundant proteins and affinity-based approaches usually cover the medium to low abundant proteins, a combination of methodology will increase the depth of proteome coverage and facilitate more comprehensive conclusions (4). The Proximity Extension Assay (PEA) is a technology developed for the analysis of secreted proteins in serum and blood plasma. The technology has been proven to possess exceptional readout specificity and sensitivity (sub-pg/ml), enabling high multiplex assays with coverage across a broad dynamic range (~9 log) while consuming a minimal amount of sample. In PEA, matched pairs of oligonucleotide-labeled antibodies will bind to their target antigens in a pairwise manner (Fig. 1A). Upon antibody binding, the matched oligonucleotides are brought into proximity and with the use of a DNA polymerase, a PCR target sequence is created, amplified, detected, and quantified. This downstream process is usually carried out by qPCR (5, 6); however, to increase the capacity for high-throughput screening of biological samples and to expand our assay library, we decided to take our PEA technology to the next level via the use of automation, miniaturization, and next-generation sequencing (NGS). NGS has rapidly evolved during the last decades, and today Illumina is the market leader in massively parallel sequencing of short reads. Combining our PEA technology with an NGS readout makes an important milestone for the new era of protein identification and quantification. Here we present Olink Explore, comprising nearly 1500 validated protein assays arranged over four 384-plex panels run in parallel, utilizing a miniaturized and automated library preparation protocol to provide unprecedented throughput (Fig. 1C).

EXPERIMENTAL PROCEDURES

Probe Generation

Polyclonal antibodies (pAb) split in two or monoclonal antibodies (mAb) were resuspended to 2 mg/ml in PBS according to the

From the Olink Proteomics, Uppsala, Sweden

^{*}For correspondence: Lotta Wik, lotta.wik@olink.com.

concentration stated by the manufacture. The concentrations were measured by NanoDrop and the antibodies further diluted to 1 mg/ml in PBS. Antibodies with lower concentration than 1 mg/ml were first concentrated and then diluted to 1 mg/ml. Two different oligonucleotides were diluted and connected to their respective pair of antibodies creating one forward and one reverse probe. Ten microgram antibody was used in the conjugation reaction. Oligonucleotide performance assessment experiments were run using IL6 (MAB206 and AF-206-NA, R&D Systems) and HE4 (Agrisera) as a test system to determine the optimal design of the assay-specific oligonucleotides. For each disease panel the forward and reverse probes were pooled separately into four blocks (blocks A–D) representing proteins of similar sample concentration, resulting in a total of 16 different forward probe tubes and 16 matched reverse probe tubes. The forward and reverse probes were diluted and stored at 4 °C. All oligonucleotides were from Integrated DNA Technology.

Sample Preparation and Immunoreaction Setup

Ten microliters each of 88 plasma samples, two plasma control samples, three plate controls, and three negative controls were transferred to a 384-well sample source plate (Eppendorf twin.tec PCR plate 384). The sample dilution plate was filled with 9 µl sample diluent using SPT Labtech's DragonFly according to the layout in [Figure 1C](#). The SPT Labtech's Mosquito was then used to transfer 1 µl from the sample source plate to the first six columns in the sample dilution plate to make a 1:10 dilution. The plate was then sealed, mixed, and centrifuged before proceeding with the 1:100 and 1:1000 dilution where 1 µl of the 1:10 and 1:100 dilution was respectively transferred to the next six columns in the sample dilution plate. Between the 1:10 and 1:100 dilution the plate was sealed, vortexed, and centrifuged.

The immunoreaction was set up with the SPT Labtech's Mosquito using a miniaturized protocol where 0.6 µl incubation mix was mixed with 0.2 µl sample. Before this reaction was set up, incubation mixes were prepared by mixing 80 µl incubation solution with 10 µl of forward probes and 10 µl reverse probes resulting in 100 µl incubation mix of each block. The incubation mixes were manually transferred to a 384-well plate (Eppendorf twin.tec PCR plate 384) using an 8-well multichannel pipette and reverse pipetting. The Mosquito was fed with one sample source plate containing undiluted plasma samples as well as control samples, one sample dilution plate containing samples diluted 1:10, 1:100, and 1:1000, one reagent source plate containing incubation mixes, and two empty plates for the immunoreaction setup. The SPT Labtech's Mosquito transferred 0.6 µl incubation mix into the empty plates according to layout and then 0.2 µl of sample from the undiluted and diluted samples where appropriate. As the SPT Labtech's Mosquito deck has only five positions, the immunoreaction setup must be run twice for all four panels. The plates were then sealed, centrifuged, and incubated in 4 °C for 16 to 24 h.

PCR Amplification and PCR1 and PCR2 Pooling

In the first PCR amplification step (PCR1 step), 19 µl PCR1 mix containing MilliQ-water, 10% PCR1 solution (Olink Proteomics), forward and reverse universal amplification primers (IDT technologies), 10% PCR1 enhancer (Olink Proteomics), and PCR1 Enzyme (Olink Proteomics) were added to each reaction using the SPT Labtech's Dragonfly. After mixing, the four plates were transferred to two thermocyclers (ProFlex Dual 384-well PCR system, Applied Biosystems) where an initial extension step (50 °C, 20 min) was run followed by 95 °C for 5 min and then 25 amplification cycles (95 °C 30 s, 54 °C 1 min, 60 °C 1 min). The forward primer used in the PCR1 step was

complementary to Illumina's P5 adapter sequence, and the reverse primer was complementary to a common sequence in the connected oligo. After PCR1 amplification, the four blocks from each sample were pooled resulting in 20 µl pooled PCR1 product containing amplicons at equal concentration from blocks A, B, C, and D. In the second PCR amplification step (PCR2 step), the same forward primer was used as in the PCR1 step. 96 different reverse primers containing the P7 adapter sequence, and a sample specific index were added to the reaction using the following PCR program: 95 °C for 3 min and then ten amplification cycles (95 °C 30 s, 68 °C, 1 min). The resulting amplicon was 148 base pairs long and contained the P5 adapter sequence, the Rd1SP site sequence, FBC, RBC, sample index sequence, and the P7 adapter sequence. After the PCR2 step, the epMotion 5075lc (Eppendorf) was used to pool all samples in each individual panel resulting in four libraries ready for purification and sequencing.

Bead Purification and Sample QC

To efficiently remove primers, primer dimers, dNTPs, and other contaminants, the samples were purified using SPRI paramagnetic beads (Beckman-Coulter Agencourt AMPure XP beads) before sequencing. The purification was conducted according to manufacturer's instructions using 50 µl unpurified library and 80 µl AMPure XP beads. The libraries were eluted in 50 µl water and immediately frozen or further prepared for sequencing. Before sequencing, all libraries were checked on a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies) as per the manufacturer's instructions to verify the correct size of the library.

NGS

The multiplex library pools were sequenced on the Illumina NovaSeq 6000 system using the S1 flow cell with the Xp workflow following the NovaSeq 6000 Sequencing System Guide. The NovaSeq Xp workflow enables libraries to be loaded onto individual lanes of the flow cell; therefore the four different library pools were loaded onto one lane each. As the S1 flow cell only contains two lanes, dual S1 flow cells were required to run all four libraries at the same time. Sequencing was performed using single-read sequencing with the read length of 66 base pairs covering the FBC, the RBC, and the sample specific index. To reduce the risk of run-to-run carryover a maintenance wash was performed before every sequencing run. As the diversity in the sequences of our libraries is very low, a PhiX control v3 library was added to the sample library before sequencing to balance the fluorescent signal. Briefly, the bead purified samples were diluted 1:100 and 45 µl of the diluted samples were mixed with 5 µl 1 nM PhiX each. Eighteen microliters of the library/PhiX mixture was then mixed with 4 µl of 0.2 N NaOH and incubated for 8 min. After 8 min, 5 µl of 400 mM Tris-HCl pH 8.0 was added to each tube. For the Xp workflow an ExAmp master mix was prepared and added to each sample, and the ExAmp/library/PhiX mixture was added to each NovaSeq Xp manifold well according to instructions. The two flow cells were loaded onto the NovaSeq 6000 system, and the number of cycles for Read 1 was set to 66.

Ethical Consideration

As for the paired plasma and serum samples analyzed in this study ([Performance of PEA in Serum and Plasma](#)), a standardized protocol was used for blood collection and anthropometric measurements. The protocols were approved by the Ethics Committee of the Karolinska Hospital and in accordance with the Declaration of Helsinki principles.

All individuals gave informed consent to their participation and no identifying information is included in the text.

RESULTS

PEA Probe Design for NGS Readout

To achieve an efficient PEA probe design, we included Illumina adapter sequences, barcodes for assay (protein) identification, indices for sample identification, and a sequence for pairwise hybridization of the two oligonucleotides when in proximity (Fig. 1B). The Illumina adapter sequences, P5 and P7, enable library amplicons to bind to the Illumina sequencing flow cells, whereas the read 1 sequencing primer (Rd1SP) site is used for single-read sequencing. Sample indexing is a common approach to distinguish pooled samples in multiplex sequencing, and in our case, we also incorporated matched assay barcodes to enable the identification of each protein assay via dual recognition. During data analysis the sample specific indices and the assay specific barcodes are used to demultiplex sample reads and determine the number of reads for each specific protein assay (n = 1472) in each sample (n = 96).

Probe Generation—PEA probes are generated from two paired antibodies, either matched monoclonal antibodies

(mAb), one polyclonal antibody (pAb) split in two, or a mix of both (one mAb and one pAb). The two matched antibodies are coupled to unique sequences in which one contains Illumina's P5 and Rd1SP sequence and the other contains a common sequence used as a primer binding site in the PCR-based preamplification process (Fig. 1C). To distinguish the different assays, the matched PEA probes contain assay specific barcodes and a hybridization site used for pairwise annealing between the two probes. The performance of the hybridized and paired oligonucleotides was tested by determining their reporting efficiencies and signal-to-noise ratio resulting in selection of the top performing sequences (data not shown).

Protocol Description—In the first step of the protocol, biological samples are sequentially diluted and then mixed with PEA probe pairs that will bind to target proteins in the sample. Binding of the PEA probe pairs with their protein brings the attached oligonucleotides into proximity and enables pairwise DNA hybridization. Addition of a DNA polymerase in the second step extends the oligonucleotides to form unique DNA sequences, which are amplified with PCR using universal primers (PCR1, see Fig. 1C). In the third step, the sample-specific indices and Illumina's adapter sequence P7 are

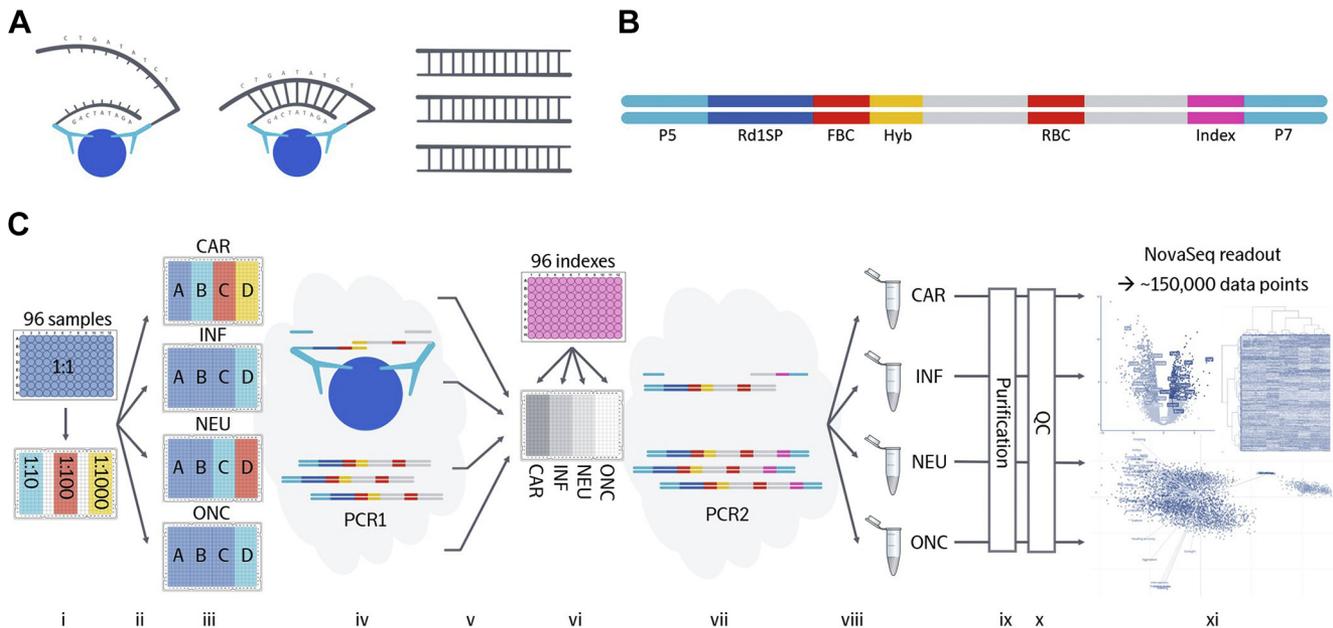


FIG. 1. Schematic overview of Olink Explore. A, matched PEA probes (pairs of specific oligonucleotide-coupled antibodies) generate detectable amplicons only upon pairwise binding to a target protein. B, sequence elements of PEA amplicons for NGS readout: Illumina P5 and P7 adapters and read 1 sequencing primer site (Rd1SP), assay specific forward and reverse barcodes (FBC and RBC), hybridization site between probe arms (Hyb), and a sample-specific index. C, the highly automated workflow of Olink Explore is divided into the following steps: (i) A serial dilution of 96 samples (different sample dilutions are used depending on the concentration of the target proteins). (ii) The undiluted and diluted samples are distributed into four 384-well plates and combined with probes from each of the four 384-plex protein panels generating four abundance blocks (A–D) for each panel. (iii) The combined samples and probes are incubated over night at 4 °C. (iv) Amplicons are generated and preamplified from proximal binding probe pairs (PCR1). (v) Abundance blocks are pooled into a single 384-well plate and (vi) combined with unique sample index primers. (vii) The sample index sequences are incorporated via PCR (PCR2). (viii) PCR amplicons are pooled into four sequence libraries. (ix) The libraries are purified using AMPure XP beads and (x) quality controlled using the Agilent 2100 Bioanalyzer system before (xi) sequenced on an Illumina NovaSeq 6000 instrument to generate close to 150,000 data points.

added *via* a second PCR (PCR2). The resulting library DNA sequences contain all required Illumina sequence adapters, primer binding sites, assay specific barcodes and sample specific indices.

Assays and Controls

Since the development of the single-plex (6), 24-plex (6) and 96-plex (5) protein assays, we have now increased the degree of multiplex to 384 as well as introduced NGS as a readout method. The Olink Explore product consists of four 384-plex panels with a focus on inflammation (INF), oncology (ONC), cardiometabolic (CAR), and neurology (NEU) protein biomarkers. As PEA relies on internal built-in quality controls (QCs), each panel consists of up to 372 human protein assays and 12 internal controls. Three of the Olink Explore assays (IL6, IL8, and TNF) are also included in each of the four panels for quality assurance purposes. The total number of unique protein assays in Olink Explore is 1463.

Olink's library content is based on low-abundant inflammation proteins (e.g., TNF and IL6), actively secreted proteins (e.g., MMP2 and sCD14), organ-specific proteins that have leaked into circulation (e.g., KLK4 and SAA4), established drug targets and candidates from ongoing clinical trials (e.g., KIT and IL1R1), and proteins previously detected in blood by other methods (supplemental Appendix 1). The library also includes proteins used in clinical diagnostic settings, such as NT-proBNP, Troponin I, CA125, coagulation factors VII and IX, and Prolactin. Selection, classification, and categorization of proteins were based on various databases (e.g., Gene Ontology and Reactome), the Blood Atlas—the human secretome (www.proteinatlas.org), a collaboration with the Institute for Systems Biology (ISB), Seattle, WA, for tissue-specific proteins, www.clinicaltrials.gov for mapping of drug targets, Human Plasma Proteome Project (HPPP) and collaborators internal mass spectrometry data to enrich for detection of proteins in blood by other methods, and finally a literature search for protein biomarkers. Our protein biomarkers vary in a wide range of properties, e.g., various sizes (5–3800 kDa), intracellular and secreted, heavily glycosylated proteins (e.g., MUC16), and cleaved protein fragments (e.g., NT-proBNP). Mapping of the Olink Explore library to the Reactome database (www.reactome.org) shows coverage of all major pathways (supplemental Appendix 2). The library also provides coverage of the majority of the more detailed sub-pathways within each category, with 86% and 74% of the first and second subpathway levels covered respectively. Particularly comprehensive coverage with several proteins per pathway represented is the categories' immune system, signal transduction, and programmed cell death. All antibodies used in the panels were either commercially available or custom-developed. Custom-developed antibodies were produced in rabbits who received a total of four immunizations. The primary injection was done using Freund's complete adjuvant and the other using Freund's incomplete adjuvant. Two weeks

after the third immunization, a test bleed was done, and an ELISA titer determination was performed. Two weeks after the fourth immunization, blood was collected and affinity purified against protein-coupled agarose. The antigens used in the immunization and affinity purification process are soluble, glycosylated, and correctly folded when applicable, to preserve the tertiary structure of the protein.

Internal and External Controls

Data normalization ensures that the measured changes in assay signal levels reflect actual changes in protein levels and not experimental artifacts. Successful normalization minimizes the variability and will generate more reproducible and precise data. To monitor, control, and normalize key steps in the protocol, from immunoreaction to detection, three specifically engineered internal controls are included in each incubation reaction (Fig. 2). The incubation (immuno) control (Inc Ctrl) is used as a QC and comprises PEA probes measuring a fixed concentration of nonhuman green fluorescent protein (GFP). The Inc Ctrl is added in the immunoreaction step. The extension control (Ext Ctrl) is used for normalization of the data and is added at a fixed concentration in the immunoreaction step. It is composed of two paired oligonucleotides coupled to the same antibody molecule, thereby keeping the two oligonucleotides in constant proximity and allowing direct hybridization independent of antigen binding. The amplification control (Amp Ctrl) consists of a synthetic double-stranded DNA template and is also used in QC to monitor the PCR steps in the protocol.

During PCR amplification, all extension products (including the Ext Ctrl) will be amplified at the same rate, and the resulting number of amplicons will be relative to the starting concentration of the Ext Ctrl in all samples. In a sample with a high protein concentration, the resulting yield of amplicons will be high relative to the Ext Ctrl and vice versa; in a sample with low protein concentration, the resulting number of amplicons will be low relative to the Ext Ctrl. During data normalization, the number of amplicons from each assay will be normalized against the number of amplicons for the Ext Ctrl to enable comparison of protein levels between samples.

In addition to the internal controls, each plate also includes external controls. A negative control sample (buffer) run in triplicate is utilized to set background levels and calculate the limit of detection (LOD), a plate control sample (pooled plasma) is run in triplicate to adjust levels between plates, and a biological sample control (optional, not used for normalization) is included in duplicate to estimate precision within and between runs.

Abundance Blocks

The dynamic range of proteins in plasma is extensively wide, spanning more than ten orders of magnitude (1, 2). Consequently, the highest abundant proteins can give rise to a very high number of amplicons that might potentially

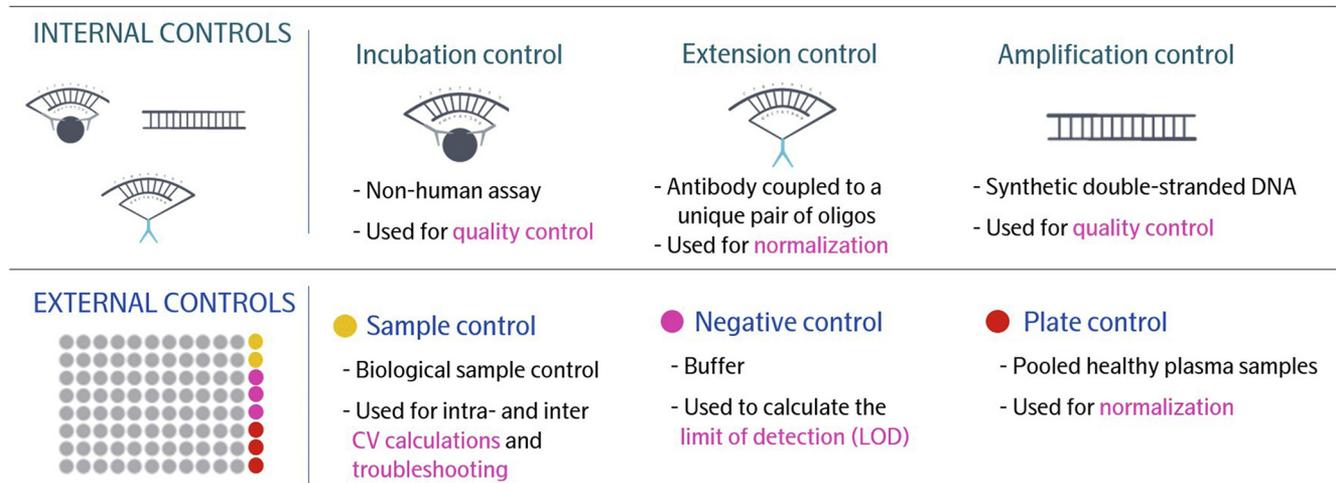


FIG. 2. **Internal and external controls.** Olink Explore uses three internal and three external controls that are used for quality control and data normalization.

outcompete all other amplicons on the flow cell during sequencing. To prevent this, and for optimal readout quality, assay probes are divided into different probe pools (blocks) based on their determined abundance in plasma samples. For an optimal concentration of target proteins, plasma samples are diluted and incubated separately with each block and are extended and preamplified separately before they are pooled at sample level for sequencing. As a result, each sample will contain amplicons corresponding to both high and low abundance proteins.

Automation and Miniaturization

In the library preparation protocol, we introduce miniaturization and automation by use of liquid handling robots to minimize sample consumption, decrease reagent volumes (and therefore, lower the cost), ensure correct pipetting, minimize assay variation, and maximize the throughput of samples. The instruments used in the Olink Explore protocol are all commercially available. SPT Labtech offers several different liquid handling systems, and we use the Mosquito LV and Dragonfly Discovery in our protocol to perform the pre-PCR workflow (Fig. 1C). SPT Labtech’s Mosquito is a nanoliter liquid handling system that offers highly accurate and precise multichannel pipetting from 25 nl to 1.2 µl. Disposable tips are used to eliminate cross-contamination, and a true positive-displacement pipetting technology assures accurate and precise volumes. Using SPT Labtech’s Mosquito in the immunoassay step reduces the volume of sample down to 0.2 µl per reaction and 2.8 µl in total for the complete Olink Explore protocol (16 blocks with different dilution factors). For post-PCR pooling and setting up PCR2 reactions, we include the epMotion 5075lc system from Eppendorf. EpMotion 5075lc has 15 workable positions, and instrument pipetting is based on Eppendorf’s air cushion technology.

NGS

NGS has had a great impact on genomic research during the last couple of decades by increasing analysis throughput while decreasing costs. Today, Illumina’s sequencing by synthesis (SBS) is the most widely used technology in genomics research, and Illumina sequencers are the world leading NGS platforms. In SBS, single bases are detected as they are incorporated into the growing chain of fluorescently labeled nucleotides. The NovaSeq 6000 (Illumina) is the most powerful Illumina system today and offers an output of up to 6 Tb of data or 20 billion reads. It uses the latest patterned flow cell technology combined with Exclusion Amplification (ExAmp) chemistry (7), which significantly increases sequencing cluster density and therefore, data output. The two-channel chemistry used in the NovaSeq 6000 reduces imaging acquisition and increases data processing time as only two images per cycle instead of four are required to detect all four nucleotides.

The Olink Explore Protocol

Sample Preparation—The Olink Explore protocol was developed for measurement of proteins in plasma and serum samples, but other sample matrices, such as CSF and aqueous humor, have also been analyzed successfully (data not shown). As each of the four 384-plex panels consists of four abundance blocks requiring different dilutions, the first step in the protocol is to predilute the samples to appropriate dilutions for each of the 16 blocks (Fig. 1C). The sample dilution is done using SPT Labtech’s Dragonfly for dispensing sample diluent, and SPT Labtech’s Mosquito LV is used to dilute the samples. Each of the 96 samples is serially diluted to 1:10, 1:100, and 1:1000 in the Olink Sample Dilution buffer.

Immuno Reaction (Incubation)—To minimize reagent and sample consumption as well as costs while maximizing

throughput, we decided to miniaturize the protocol and use five times less reagent compared with the Olink Target 96 protocol (5). This results in a reduction in plasma sample volume from 1 μ l of plasma sample per reaction (96 assays) to only 0.2 μ l per reaction. As the different abundance blocks require different sample dilutions, the total volume consumed for all 16 abundance blocks is 2.8 μ l. A total of four 384-well plates, consisting of four blocks each, are used to set up all four panels for Olink Explore in 96 samples (Fig. 1C).

In the immunoreaction step, the samples (typically 88 plasma samples plus eight control samples) are separately mixed with an incubation mix consisting of incubation solution and PEA probes. In the miniaturized protocol we also decreased the incubation mix fivefold to 0.6 μ l instead of the 3 μ l used in the Olink Target 96 protocol, and the total incubation volume (plasma sample and incubation mix) per reaction in the miniaturized protocol is reduced to 0.8 μ l. The immunoreaction, whereby the paired PEA probes will bind to their respective protein and hybridize, takes place during overnight incubation at +4 °C.

PCR1—The incubation plate is brought to room temperature and centrifuged at 400g for 1 min. Using SPT Labtech's Dragonfly, a combined extension and preamplification mix containing PCR1 solution, PCR1 Enhancer, and PCR1 enzyme is added to each reaction. After sealing, vortexing, and centrifugation, plates are moved to the post-PCR room for extension and PCR amplification using two Proflex Dual 384-Well sample Block PCR systems (Thermo Fischer Scientific). The PCR1 program starts with an initial extension step (50 °C, 20 min) followed by a denaturation step (95 °C, 5 min) and 25 cycles of amplification (95 °C, 30 s; 54 °C, 1 min; 60 °C, 1 min). The first PCR is performed using forward and reverse universal preamplification primers for amplification of all extension products in the reaction (Fig. 1C).

Pooling of PCR1 Products—PCR1 products generated for the same sample in the four different abundance blocks from each panel are pooled together using Eppendorf's epMotion 5075lc. This results in one 384-plate containing all four panels (Inflammation, Oncology, Cardiometabolic, and Neurology) where amplicons from each plasma sample and abundance block are represented (Fig. 1C).

PCR2—Individual index sequences together with Illumina's P7-adapter sequence are added to each sample *via* a second PCR step. The PCR2 setup is handled by Eppendorf's epMotion 5075lc where the pooled and diluted PCR1 products are mixed with an index primer and a PCR2 mix containing PCR2 solution and PCR2 enzyme. The PCR2 program starts with a 95 °C incubation for 3 min followed by ten cycles of amplification (95 °C, 30 s; 68 °C, 1 min).

Pooling of PCR2 Products—All 96 samples run with the same panel are then pooled together using Eppendorf's epMotion 5075lc. The PCR2 pooling results in four different

libraries representing each of the four panels containing amplicons from all assays and all samples.

Bead Purification—As adapter dimers can bind to the flow cell and as free index primers may result in index hopping (<https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf>), the libraries are purified using the AMPure XP magnetic bead purification protocol (Beckman Coulter) directly after the PCR2-pooling step. The purified libraries are eluted in the same volume yielding approximately the same concentration for each library.

Quality Control of Library—Agilent's 2100 Bioanalyzer is used for troubleshooting and QC of the DNA libraries before sequencing. As the PCR2 products are made up of 148 base pairs, Agilent's High Sensitivity DNA kit will show a peak around 150 bp after bead purification. Additional peaks containing larger fragments, known as bubble products (<https://emea.support.illumina.com/bulletins/2019/10/bubble-products-in-sequencing-libraries-causes-identification-.html>), might exist but do not influence the data according to experimental verification (data not shown).

NGS—The Olink Explore protocol is PCR-based and incorporates P5 and P7 adapters together with the read 1 sequencing primer (Rd1SP) site to be used on Illumina's sequencers. The assay barcodes together with the in-line index are read from the Rd1SP as part of the sequence read. As off-the-shelf commercial qPCR-based kits failed to reproducibly quantify our amplicons and since a volume-based approach gave a much more consistent number of reads (data not shown), we decided to use a standardized protocol to ensure that correct dilutions were used to maximize the output data while at the same time minimizing failed runs. The different panels were diluted and sequenced on separate NovaSeq S1 flow cell lanes according to the NovaSeq Xp workflow.

Data Analysis

The Olink Explore protocol generates almost 150,000 quality-controlled protein data points per run (Fig. 3). Therefore, data analysis is key to generate meaningful insights and actionable conclusions from big data sets.

Each sequence read will contain both assay and sample information. The assay barcode and sample index sequences are designed so that there is no risk that matched barcodes and indices will be incorrectly assigned. Only exact matches are included for further analysis. Assay barcode sequences were designed using an in-house script owing to specifically select matched barcodes that have a hamming distance of at least three and do not form structural motifs. Due to the large amount of sequencing data obtained from each sequencing run, the data analysis is handled in steps. Briefly, the sequencing run data is first processed on a local server where the program bcl2counts transforms the BCL or CBCL files to a

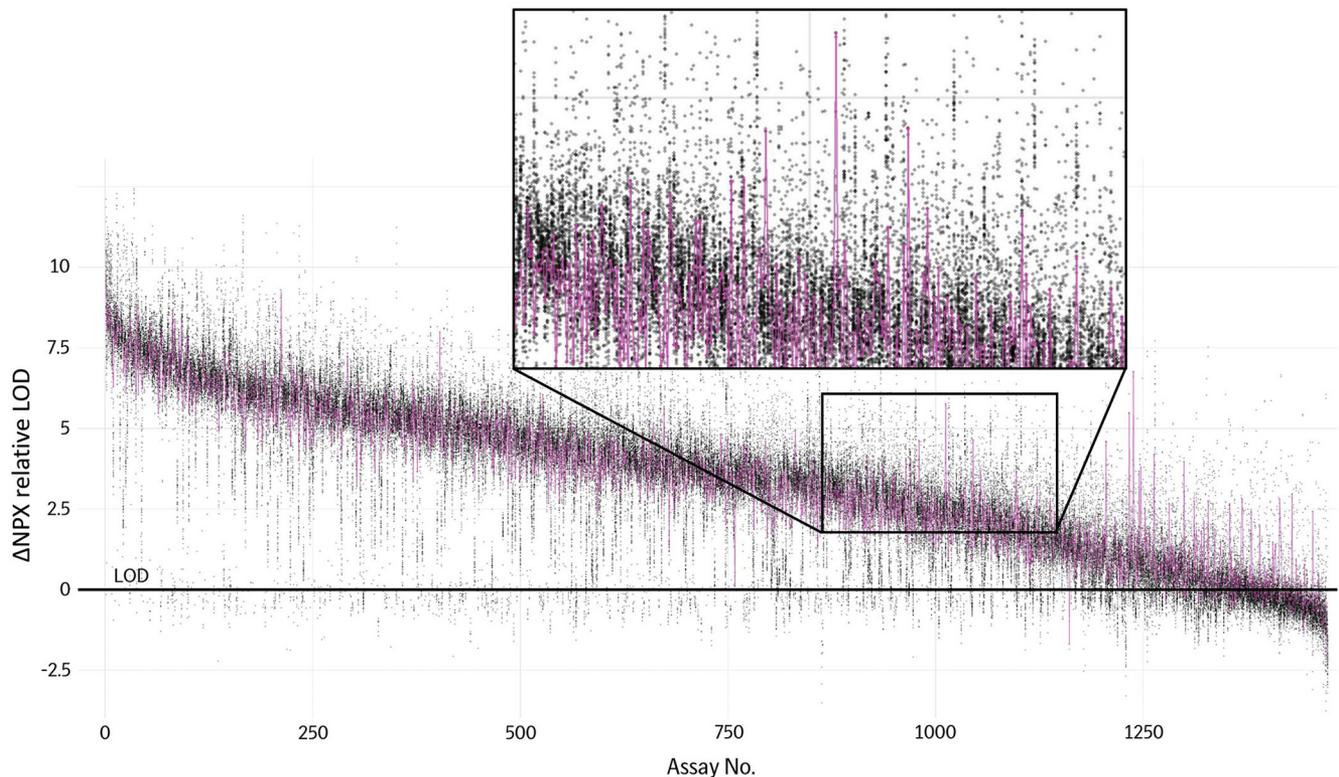


FIG. 3. The almost 150,000 data points generated from one Olink Explore run are visualized in the plot where protein abundances from 88 samples are plotted as Delta Normalized Protein eXpression (dNPX) values for all assays. The 88 samples are paired plasma and serum samples from 40 individuals, six of which are analyzed as technical triplicates. The assays are sorted based on decreasing mean dNPX as measured in plasma among all individuals. The 40 individuals could be stratified into two distinct groups based on BMI (normal or high), with 20 individuals each. The data points from the individual with the highest BMI as measured in plasma are connected by a pink line to visualize the shape of a protein profile fingerprint. The black horizontal line denotes limit of detection (LOD).

file containing counts from each sample index matched with the different assay barcodes. The data is then uploaded to a secure cloud-based application where normalization and QC of data are performed.

Data Normalization—Normalized Protein eXpression (NPX) is Olink's relative protein quantification unit on a log₂ scale and values are calculated from the number of matched counts from the sequencing data. Data generation of NPX consists of three main steps: normalization to the Ext Ctrl, log₂-transformation (to get a more normally distributed data and make it more interpretable), and level adjustment using the plate control (pooled plasma sample). First, matched sequence reads (counts) for each specific combination of assay barcode and sample index will be divided by the number of counts for the extension control with the same sample index. The resulting ratio will be log-transformed.

$$\text{ExtNPX}_{i,j} = \log_2 \left(\frac{\text{Counts}(\text{Sample}_i, \text{Assay}_j)}{\text{Counts}(\text{ExtCtrl}_j)} \right)$$

The number of counts for the plate controls is used to correct for variation between plates by subtracting the median

of the Ext Ctrl-normalized counts for the plate controls for the corresponding sample:

$$\text{NPX}_{i,j} = \text{ExtNPX}_{i,j} - \text{median}(\text{ExtNPX}(\text{plate ctrls}_i))$$

If more than one sample plate is analyzed, and the samples throughout the plates are adequately randomized, an optional intensity normalization step can be performed to further minimize inter plate variation where the global median of all NPX from a sequencing run for each assay is subtracted from the calculated NPX value:

$$\text{NPX}_{\text{IntNorm } i,j} = \text{NPX}_{i,j} - \text{plate median}(\text{NPX}_i) + \text{global median}(\text{NPX}_i)$$

where i = assay and j = sample.

Sample Quality Control and Calculation of Coefficient of Variation—The QC assessment is performed at two levels; run QC and sample QC. At the run QC level each of the four abundance blocks for each 384-plex protein panel and sample plate should fulfill the following criteria: (i) mean absolute deviation (MAD) in internal controls (Inc Ctrl and Amp Ctrl) may not exceed 0.3 NPX and (ii) deviation on the sample QC level is

allowed for a maximum of one out of six samples. Further, in each panel, the median of at least 90% of the assays in plate and negative control samples must be in the accepted range from predefined values set during validation. Apart from run QC, the performance of each sample is assessed individually by the internal controls that should be within ± 0.3 NPX from the median level across the abundance block. Additionally, the mean assay count for a sample may not be below 500 counts. Abundance blocks and samples that do not fulfill their respective QC criteria will receive a QC warning. Another important QC metric is precision evaluated as the coefficient of variation (CV). CV is a measure of technical variation for individual assays both within a plate (intra-CV) and across multiple plates (inter-CV). The assay intra- and inter-CVs are derived from a pooled plasma sample present in duplicates in each sample plate. The CVs are expressed in percent and defined according to the equation (8):

$$CV = 100\sqrt{e^{\sigma^2} - 1}$$

Where $\sigma = \sigma_{\text{assay NPX in duplicate control pools}} \ln(2)$.

The mean intra- and inter-CVs among all assays are recommended to be $\leq 15\%$ and $\leq 25\%$, respectively.

Validation and Robustness Tests

Analytical performance was carefully validated for each protein assay, the results of which are available at www.olink.com. Technical criteria include assessing sensitivity, dynamic range, specificity, precision, endogenous interference, and detectability in both healthy and pathological plasma and serum samples. Consistency in sample preparation is a key parameter for finding valuable biomarker candidates. To achieve this, a reproducible and robust protocol is paramount. Therefore, robustness tests were also performed to ensure that such was the case for this protocol. Results from the robustness investigations are summarized in the Supplemental data (supplemental Figs. S1–S9).

Dynamic Range and Precision—The dynamic range for each assay was determined by using a dilution series of recombinant antigen in multiplex (Fig. 4A). The values correspond to antigen concentration before dilution (used in the abundance blocks) resembling measurable levels from actual samples. At the low end of the series, the LOD is defined as three standard deviations (SDs) above background levels (negative control). Values below LOD are generally recommended to be included in the dataset for biomarker studies to increase the statistical power and to get a more normal distribution of the data. The high end of the series possesses an assay limitation because of the high dose hook effect (<https://www.olink.com/question/what-is-the-high-dose-hook-effect/>). We defined the highest signal of the antigen standard curve as the hook threshold. Within those points we further defined the quantifiable range of each assay by estimating the limit of quantification (LOQ) with the requirement that each back-calculated standard point

in the curve must have a %CV and absolute accuracy below 30%. On average for the 1472 assays, the quantifiable range, *i.e.*, the distance between the upper and lower LOQ (ULOQ and LLOQ) was 2.7 when expressed on a log₁₀ scale (corresponding roughly to 9 on a log₂ scale, or a 500-fold difference in antigen concentration). The average assay precision in terms of intra- and inter-CV was 7.8 and 10.6%, respectively (Fig. 4B) calculated on values above LOD.

Detectability—Detectability was defined for each assay as the percentage of samples that were detected on two plates above the maximum LOD over four plates (72 samples over two plates twice). This was assessed in EDTA plasma samples ($n = 72$, purchased from BioIVT) corresponding to the following five general categories: healthy controls ($n = 24$), oncology ($n = 12$), neurology ($n = 12$), inflammation ($n = 12$), and cardiovascular ($n = 12$). On average, 85% of the protein assays ($n = 1472$) were detected over LOD in at least 50% of the samples ($n = 72$). Of the remaining assays (217 proteins representing 15% of the protein assays), some had a robust signal in samples from a specific disease category, indicating a lower general detectability but a strong potential for protein biomarker relevant to specific diseases (see supplemental Appendix 3). However, most of the remaining assays had low detectability in the EDTA samples tested here but were included based on previous detectability testing or other sample matrices.

Specificity—The unique requirement for dual antibody recognition when using PEA technology overcomes the first long-standing and well-recognized challenge with immunoassay: unspecific binding. Furthermore, all assays go through predefined specificity testing. Before incorporation into Olink Explore, all antibodies were screened for cross-reactivity in a multistep procedure involving several pools of antigens. After removal of poorly performing antibodies, a second screen was performed using an expanded set of antigen pools. As a final product validation, the assays were then tested against a selected set of antigens ($n = 96$), comprised of well-known targets ($n = 15$) from each of the four 384-plex panels as well as several proteins ($n = 36$) from families exhibiting at least 50% sequence identity. In total, 99.8% of the protein assays (1469/1472) in Olink Explore showed no cross-reactivity according to the tests described. Of the remaining 0.02% protein assays, FOLR3 showed a nonspecific signal against the related FOLR2 protein (83% amino acid sequence identity), CCL3 against the related CCL4 protein (58% amino acid sequence identity), and LHB against CGB3 (85% amino acid sequence identity). The signal contribution for each of the protein assays was further investigated and more details are noted on the specific biomarker assay pages at www.olink.com.

Index Hopping

The introduction of ExAmp chemistry combined with patterned flow cell technology to NGS methods has

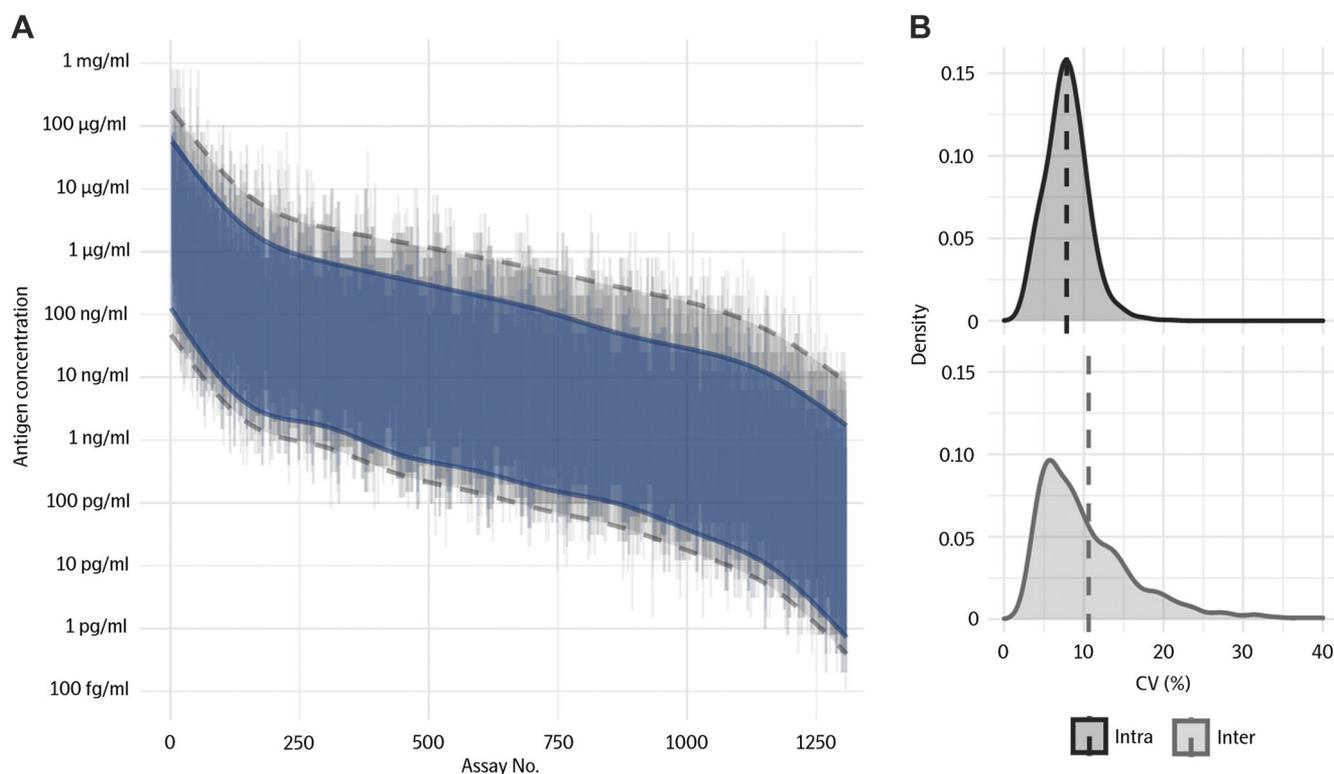


FIG. 4. Validation of range and precision for Olink Explore assays. *A*, standard curves were generated from a dilution series of the target recombinant antigens. Four different parameters: limit of detection (LOD), lower limit of quantitation (LLOQ), upper limit of quantitation (ULOQ), and hook were defined from the generated standard curves. LOD was defined as three SD above the signal generated from the Olink Explore negative control sample. LLOQ, ULOQ, and hook were defined using four parameter logistic regressions (details can be found on the Olink website, www.olink.com). The four parameters for each assay are visualized in the figure as a *gray line* indicating LOD and hook. The *gray line* is overlaid with a *blue line* indicating the quantifiable range. The assays are sorted based on decreasing mean of the quantifiable range. Four lines (*gray dashed lines* for LOD and hook and *blue solid lines* for LLOQ and ULOQ) of smoothing averages using generalized additive models (GAM) were plotted on top of the *vertical lines* of the individual assay values. The area between the smoothed averages for LLOQ and ULOQ is filled with a semitransparent *blue color* representing the quantifiable range. *B*, intra- and inter-CV values were calculated individually for each assay from linearized NPX values from samples run on the same and different plate(s), respectively, here visualized using density plots (details can be found on the Olink website, www.olink.com). The *dashed vertical lines* denote average values.

significantly increased data output and sped up protocols in the NGS technology field. However, a known phenomenon called index hopping (<https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf>) where sequence reads are assigned to wrong sample has been associated for this chemistry, which leads to inaccurate sequencing results (9). Some general guidelines for reducing the effect of index hopping are to store libraries at -20°C and to sequence libraries as soon as possible after pooling. As we use a PCR-based protocol with index primers, it is possible for excess free index primers to bind to hybridized library fragments *via* their unbound complementary 3' ends, extend and create new library molecules with the new index before binding to the patterned flow cell. To prevent this, bead purification was introduced as a step in the protocol to reduce the concentration of free index primers (supplemental Fig. S10). After purification, the effect of index hopping was rectified.

Correlation of IL6, IL8, and TNF

As previously mentioned, and as part of the built-in QC, three assays, IL6, IL8 (CXCL8), and TNF, were included in each of the four 384-plex panels (CAR, INF, NEU, and ONC). Interpanel correlations were assessed on a set of plasma samples ($n = 64$) consisting of a mix of healthy ($n = 21$) and disease states ($n = 43$) (Fig. 5). Analysis was performed on all samples with a signal above LOD. Regardless of panel comparison, the correlation was high for IL6 (coefficient of determination, r^2 , between 0.96 and 0.99, with an average r^2 of 0.98) and IL8 (r^2 between 0.98 and 0.99, with an average r^2 of 0.99). Due to lower signal in most healthy samples, correlation was slightly lower for TNF (r^2 between 0.80 and 0.94, with an average r^2 of 0.88).

Cross-platform Comparison

Variability in performance between immunoassay platforms using different technologies can contribute to results being misinterpreted. Therefore, it is important to compare platform

performance for several key analytical parameters such as accuracy, detectability, and interference. Considering these parameters, selected plasma samples (healthy and diseased donors) were used to perform an analytical comparison between four immunoassay platforms. The comparison was done using commercial kits that measure cytokines, and the results were compared with extracted results for the same cytokines included in the Olink Explore panel. The Bio-Plex Pro Human Cytokine 27-plex immunoassay from BioRad (run on a Luminex instrument at the SciLifeLab in Stockholm) and the V-PLEX Chemokine Panel 1 (human), V-PLEX Proinflammatory Panel 1 (human), V-PLEX Cytokine Panel 1 (human), V-PLEX Cytokine Panel 2 (human), and U-PLEX (10-plex custom panel G-CSF/CSF3, IL-17F, IL-33, FLT3L, TRAIL/TNFSF10, SDF-1 α /CXCL12, MIP-3 β , MCP-2/CCL8, MCP-3/CCL7, I-TAC) from Meso Scale Discovery (MSD) (run at the SciLifeLab in Uppsala) were used in the comparison. We also compared the results with Olink Target 48 Cytokine that utilizes qPCR as the readout method. Individual single replicates ($n = 33$) were used to analyze the detectability for each of the methods. Detectability was analyzed for 18 different cytokines and the results were compared with the quantifiable range for each method, where the LLOQ and the ULOQ were used to approximately normalize the quantifiable range for the different methods (Fig. 6A) IL13, IL2, and IL4 demonstrated low detectability in all methods except for Bio-Plex (Fig. 6A). No quantifiable range exists for Bio-Plex CXCL10 and U-plex CSF3; hence these data were excluded from the comparison. In the correlation analysis, only cytokines that were quantifiable in at least six individuals were included in the calculations (Fig. 6C). The correlation of 11 cytokines between Olink Explore and Olink Target 48 was high (r^2 between 0.882 and 0.997) and for the immunoassays CCL11 and IL7, the r^2 was 0.882 and 0.965, respectively. The correlation between Olink Explore and Bio-Plex or V- and U-Plex varied depending on the cytokine analyzed. Between Olink Explore and Bio-Plex or V-PLEX, the r^2 for CCL11 was 0.585 and 0.679, respectively, while the r^2 for IL7 was 0.039 for Bio-Plex and 0.920 for V-PLEX. Furthermore, the correlations between V- and U-plex and Bio-Plex were in general very modest, with only one protein assay presenting an r^2 value above 0.8 (supplemental Fig. S11). Technical triplicates were used to calculate intra-assay CV, which were lowest for the V- and U-Plex, but in general adequate for all platforms (Fig. 6B, upper). The median CV values were 7.3% (Olink Explore), 8.8% (Olink Target 48), 7.0% (Bio-Plex), and 3.8% (V- and U-PLEX). Pooled plasma samples with and without spiked anti-species IgG (donkey anti-goat, -rabbit, -mouse, -chicken) were used to address possible heterophilic interference for each method. The ratio of the signal between the two samples was calculated and presented as relative signal by addition of IgG (Fig. 6B, lower). The interference in Olink Explore was limited (all assays stayed within $\pm 16\%$ of the original level). The IgG interference for Olink Target 48 and V- and U-Plex was also in general limited,

except CCL3 for V-plex with a relative signal increase of 81%. The interference on Bio-Plex was more substantial, exhibiting relative increase in signal for VEGFA (117%), IL6 (61%), CXCL8 (55%), and IL4 (47%).

Performance of PEA in Serum and Plasma

Serum and Plasma—Most proteomic analyses used to identify biomarkers are performed in samples derived from blood and both plasma and serum are widely used matrices. Today there are few studies that have systematically compared the relative performance of multiplex proteomics methods across these matrices. Both matrices are obtained from the liquid part of blood but in plasma, an anticoagulant is added to prevent clotting of the blood prior to extracting the noncellular fraction *via* centrifugation. In serum, the blood is allowed to clot before centrifugation bringing additional components (such as fibrinogen, platelets, and various proteins) into the cellular fraction. The difference in sample preparation will affect the resulting protein profiles found in serum and plasma. As one example, it has been demonstrated that the concentration of VEGF varies more than eightfold between 30 pg/ml and 250 pg/ml when comparing plasma and serum (10). Here, the abundances of 1463 proteins were measured in paired plasma and serum samples collected from the same individuals ($n = 40$). The samples were selected from a larger study of 618 individuals (11) and were at the extreme ends of the body mass index (BMI) distribution from these 618 individuals. The two different sample types were collected at the same site (Karolinska hospital, Sweden) using a standardized protocol, and the samples from each individual were collected and centrifuged within 2 h (11). The samples obtained from one individual were analyzed as technical duplicates while the remaining samples were analyzed in single replicates. Analysis using Olink Explore demonstrated a detectability of 80% or more of the samples for 88% and 83% of the protein assays for plasma and serum, respectively. The measured protein abundances demonstrated a high correlation between technical duplicates for both matrices (r^2 of 0.994 and 0.996 for plasma and serum, respectively) (Fig. 7A). The correlation between the two matrices was clearly lower (average r^2 of 0.897). Paired t-tests were performed for all proteins with paired detectability (both plasma and serum) in at least five individuals to assess the number of significantly differing proteins between the two matrices. The result is summarized in a volcano plot where the significance (p -value) on the y-axis is plotted against the estimated difference in NPX on the x-axis (Fig. 7B). A large fraction ($n = 594$) of the total number of protein assays ($n = 1463$) were not significantly different between the two sample matrices. Protein assays showing a higher signal in serum *versus* plasma and vice versa are presented in the graph, and the nominal p -value and Bonferroni cutoff are added as horizontal lines. For plasma, 398 protein assays demonstrated significantly higher NPX compared with serum and, in contrast, 409 protein assays

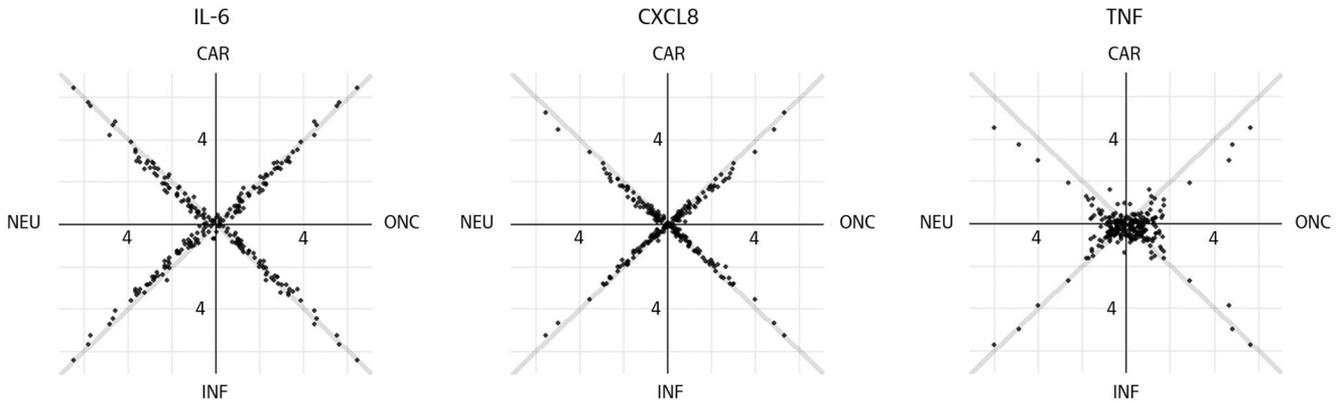


FIG. 5. **Three biological assays (IL6, IL8/CXCL8, and TNF) were used to assess interplate correlation.** These are present in all four 384-panels and used to verify the high quality of the data. The following plots represent the protein abundances (NPX) for all three assays measured in the different 384-assay panels. This was done for 64 individuals from one run demonstrating high correlation between panels.

demonstrated significantly higher NPX in serum compared with plasma. When using the Bonferroni adjusted significance level, 243 protein assays demonstrated higher NPX in plasma

and 192 protein assays demonstrated a higher NPX in serum. Six protein assays with higher levels in plasma compared with serum demonstrated estimated differences of more than four

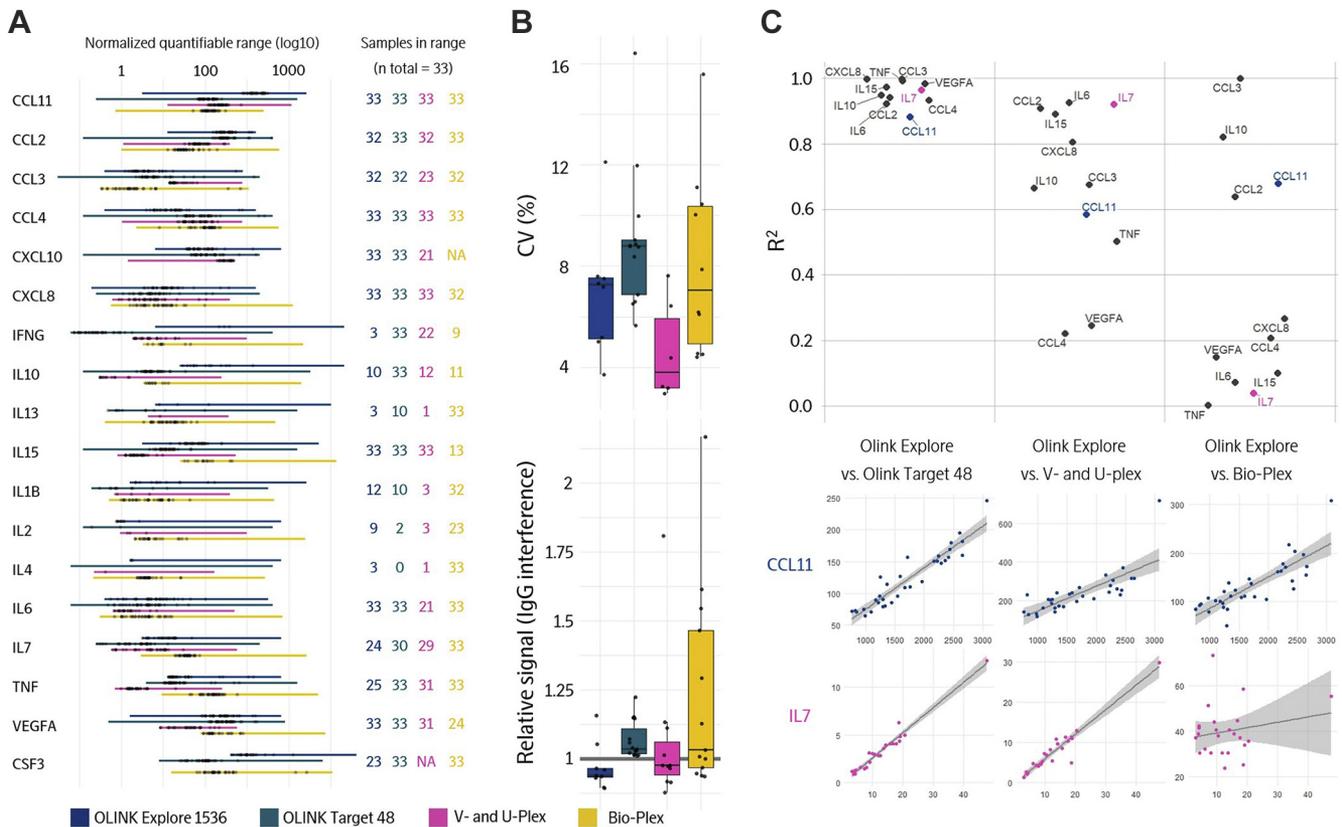


FIG. 6. **A small subset of Olink Explore assays targeting cytokines (n = 18) were compared to the corresponding assays from the Olink Target 48, MSD and Luminex platforms.** A, the number of quantifiable plasma measurements from 33 tested individuals. The quantifiable data generated from each method (black dots) are plotted within the corresponding normalized quantifiable range (colored line) on a log₁₀ scale. B, technical triplicates were used to calculate intraassay precision for the different methods (top) and a pooled plasma sample with and without additional IgG was used to calculate interference as the percentage difference in the sample with added IgG (bottom). Only quantifiable data were used for the evaluation. C, R² values were calculated between Olink Explore and the other methods for all assays with at least six pairwise quantifiable measurements. Two assays (CCL11 and IL7) are highlighted with colored dots together with the corresponding correlation plots including linear regressions and confidence intervals (Olink Explore values on the y-axis and the comparison method on the x-axis, linear scales).

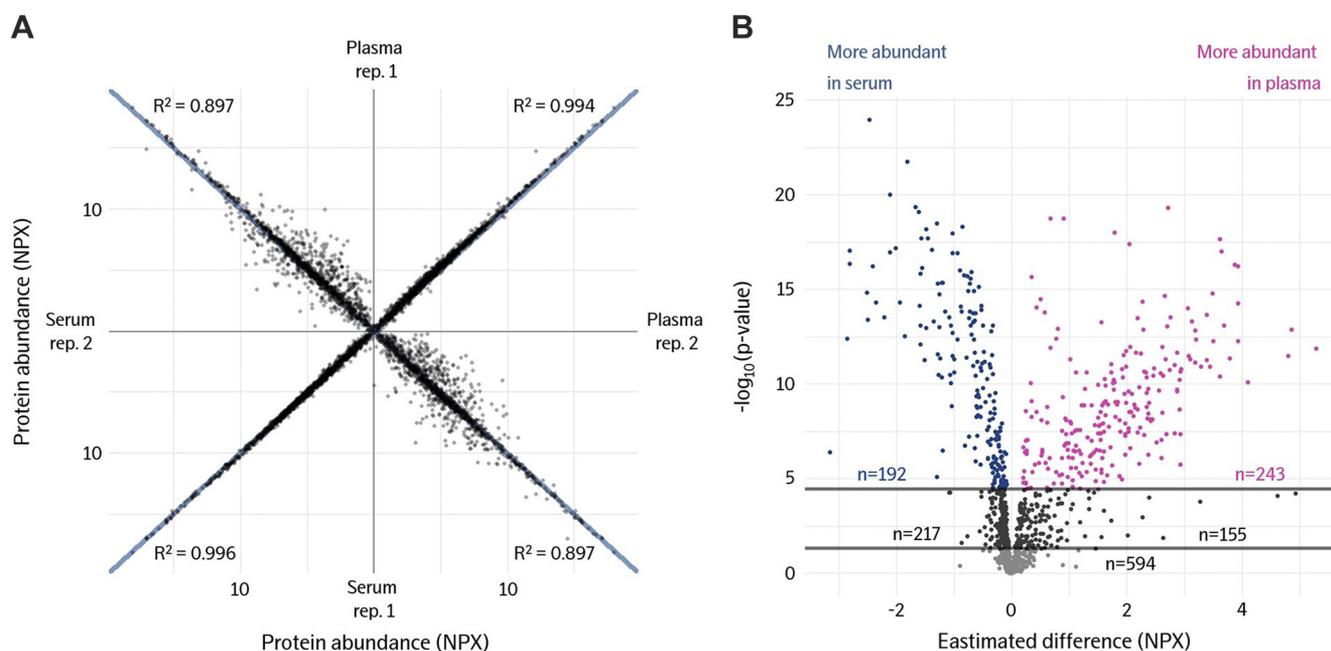


FIG. 7. Comparison between protein abundances measured from paired EDTA plasma and serum samples for a set of 40 individuals. A, the EDTA plasma and serum samples from one individual were analyzed in duplicates and are summarized in a four correlation figures. All axes are positive and represent the measurements in NPX from one of the four replicates. The correlation between technical replicates of the same sample type was high (r^2 of 0.994 and 0.996 for plasma and serum, respectively) as seen in the *upper right* and *lower left* quadrant. In contrast, a clear difference is observed between plasma and serum (r^2 of 0.897) as seen in the *lower right* and *upper left* quadrant. The *blue lines* are not linear regressions but denote theoretical equality between replicates (i.e., $x = y$). B, paired t-tests were applied to all the 1472 proteins measured in EDTA plasma and serum from the 40 individuals and are summarized in a volcano plot. The y-axis represents the probability of an actual difference between the two sample matrices and the x-axis represents the estimated difference. The *lower* and *upper horizontal lines* denote nominal ($p < 0.05$) and Bonferroni ($p < 0.05/1472$) significance, respectively. The *blue* ($n = 192$) and *pink* ($n = 243$) data points represent proteins that are increased in serum and EDTA plasma, respectively, using the Bonferroni significance.

NPX (approximately a 16-fold difference in protein concentration). All individual assays with corresponding p -value and estimated difference are presented in the supplemental material (supplemental Appendix 4). In summary, our data demonstrates that the choice of sample matrix will impact on downstream results, and observed differences between sample matrices should be interpreted with caution. Their comparability is highly dependent on the specific set of proteins investigated, and inconsistent use of the sample matrix can lead to inaccurate results and/or diagnosis. However, both plasma and serum can be used successfully for protein analysis using Olink Explore.

Obesity-related Biomarkers in Serum and Plasma—A high BMI is associated with several diseases such as heart disease, type 2 diabetes, and certain cancers (12). Comparison between overweight and obese individuals with BMI >30 kg/m² and individuals of normal weight and a BMI <22 kg/m² (Fig. 8B) shows 14 (plasma) and 17 (serum) Bonferroni significant protein assays ($p < 3.4 \times 10^{-5}$), where they are significantly increased in obese patients compared with normal weight individuals (Fig. 8A). Obese patients with BMI >30 kg/m² had significantly higher plasma and serum Leptin concentrations than patients with BMI <22 kg/m² ($p = 3.6 \times 10^{-12}$ and $p = 1.4 \times 10^{-10}$, respectively) (Fig. 8C). A similar set

of proteins across the two sample matrices showed the strongest trends toward association with BMI.

DISCUSSION

There is a clear need and growing demand for large studies that can exploit the diagnostic potential held within the low abundant plasma proteome. Such studies require robust measurement of as many proteins as possible in plasma samples. To date, no technology can measure the full proteome, but by applying sophisticated strategies, one can get a comprehensive view of the proteome that covers all important biological processes and protein pathways involved in health and disease. The use of biomarkers is essential when evaluating the most effective therapeutics for individual patients as well as to personalize lifestyle recommendations based on biomarker fingerprints. Accurate recommendations and decisions based on these biomarkers depend largely on the accuracy of the method used to measure analytes as well as the correct collection and handling of patient samples. There are many factors that can influence a blood sample, and it is highly important and strongly recommended to establish a standard operation procedure when collecting samples for analysis and diagnosis (13–15). Cayer *et al.* (16) accurately described the current state of proteomics technology

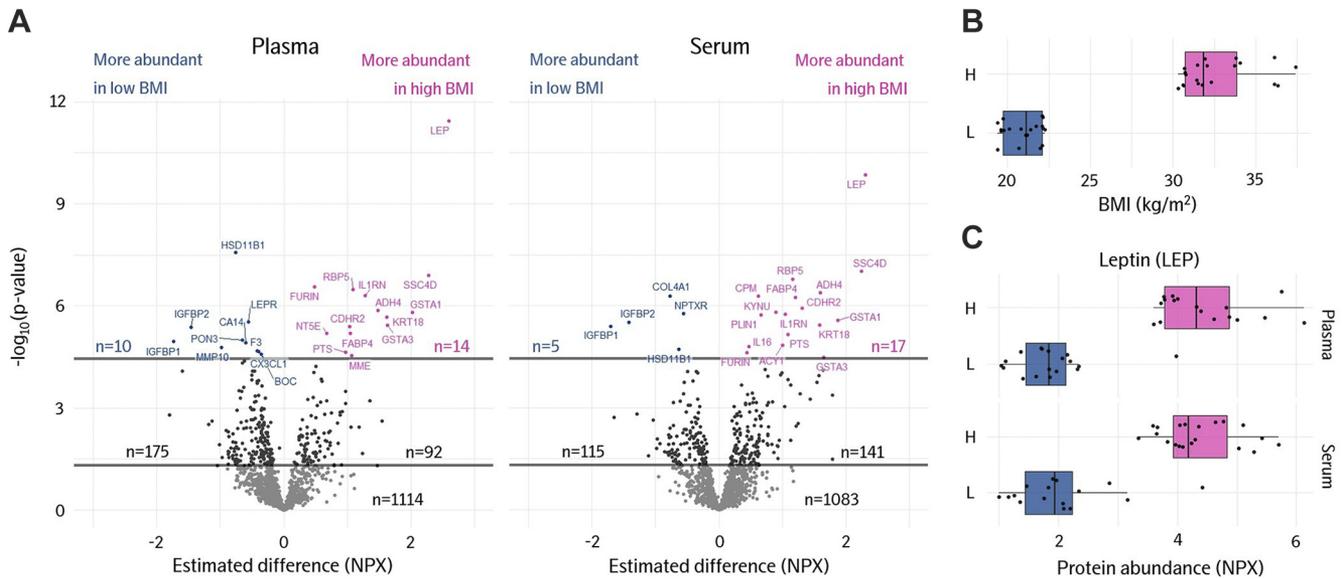


FIG. 8. Protein profile comparison between 20 obese (high BMI, 30–37.5 kg/m²) and 20 normal weight (low BMI, 18–22.5 kg/m²) individuals as measured from EDTA plasma and serum samples. A, t-tests assuming equal variance within the groups (high and low BMI) were individually applied to all the 1463 proteins screened in EDTA plasma and serum, presented here as two volcano plots. The y-axis represents the probability of an actual protein assay difference between the two BMI groups, whereas the x-axis represents the size of the estimated difference. The lower and upper horizontal lines denote nominal ($p < 0.05$) and Bonferroni ($p < 0.05/1472$) significance, respectively. The blue and pink data points represent assays that are increased in low and high BMI groups, respectively, using Bonferroni significance. B, distribution of BMI in the two groups (H: high BMI, L: low BMI). C, the distribution in protein level for the most significantly affected protein, Leptin (LEP), demonstrates no overlap between the high and low BMI groups in neither of the matrices.

research: that there is an urgent need for innovations in proteomic technologies that can compare to the advances in NGS, and that current proteomic methodologies suffer from issues related to reproducibility, sensitivity, sample requirements, and limited multiplexing capacity. With the launch of our latest product, Olink Explore, important pieces of that puzzle are now solved. Olink Explore is a powerful tool offering a high-multiplex protein biomarker platform with a tremendously high-throughput capacity in combination with minimal sample volume requirements and with high specificity and sensitivity. The technology is based on the already proven high-multiplex PEA technology that has primarily been applied for screening of proteins in blood (17–28) but is now also used for other sample matrices, e.g., CSF (29), cell (30) and tissue (31) lysate, saliva (32), urine (33), interstitial fluid (34), cell culture media (35), peritoneal fluid (36), breast milk (37), BALF (38) and synovial fluid (39). Moreover, as minimal sample volumes are required, the technology has also been applied to samples of minimal sample volume, for example, single cells (40), exosomes (41), dried blood spots (42), aqueous humor (43) and fine-needle biopsies (44). The platform has also successfully been applied in animal models (45–47). Unlike other immunoassay platforms, Olink's PEA technology has been demonstrated to be completely scalable and maintain the same exceptional data quality at a high level of multiplexing in, e.g., large-scale screening studies, as for smaller studies when validating specific protein signatures (48) using

Olink Focus, a unique custom-developed protein panel with potential for clinical implementation. Olink can provide a seamless transition from explorative to focused studies without the need to change platform.

Integration of multiomics data together with clinical information has become a significant aspect to the development of next-generation medicine, spurred on by the growth of many collaborations into large global consortiums collecting and organizing data. The possibility to integrate protein data with genetic information using protein quantitative trait loci (pQTLs) is a strong instrument for selecting novel drug targets based on proteins related to disease. The SCALLOP consortium (www.scallop-consortium.com) is a collaborative framework for discovery and follow-up of genetic associations to proteins with pQTL mapping using the Olink platform (49). To date, SCALLOP comprises summary level data for more than 65,000 patients and controls.

In 2020, COVID-19 resulted in a global crisis and according to the World Health Organization, the number of confirmed cases is now above 150 million worldwide. In a collaboration together with Massachusetts General Hospital (MGH) and the Broad Institute, Olink Proteomics quantified the abundance of proteins in plasma from the cohort of 306 COVID-19 patients and 78 symptomatic controls using Olink Explore (50). The analysis uncovered unique protein signatures for prediction of COVID-19 disease outcome and stratification of most severe patients (death or intubation) at the time of entry to the

emergency care unit and data provided important insights into underlying disease mechanisms. Raw data from the study (protein measurements and clinical parameters) is publicly available and accessible via the Olink website (<https://www.olink.com/mgh-covid-study/>).

By using NGS DNA tags as a readout, Olink Explore enables the high-throughput analysis of nearly 1500 proteins in multiplex while maintaining exceptional assay sensitivity, specificity, and the ability to measure protein abundance of a broad dynamic range required for the analysis of the plasma proteome as well as other sample matrices. The performance of the platform has already been demonstrated in various disease and wellness studies, detecting novel biomarkers for early diagnosis and disease monitoring, as well as providing a better understanding of the proteome in healthy as well as disease cohorts. Olink Explore meets all the requirements for a proteomics technology that is as advanced as NGS, but free from the issues plaguing proteomics methods that came before it. In summary, Olink Explore combines our established PEA technology with NGS to bring proteomics technology into the next generation where we will be able to deliver on the promises of the 21st century healthcare.

DATA AVAILABILITY

The performance data visualized in the technical assessment and validation studies are available from the corresponding author (Lotta Wik, lotta.wik@olink.com, Olink Proteomics) upon request. The plasma and serum sample data presented in [Performance of PEA in Serum and Plasma](#) is considered personal data and can be requested via a data processing agreement with Karolinska Institute (Anders Målarstig, anders.malarstig@ki.se, Karolinska Institute).

Supplemental data—This article contains [supplemental data](#).

Acknowledgments—We sincerely thank all people involved at Olink Proteomics for invaluable contributions. We would also like to thank Mathias Uhlén and the Human Blood Atlas for input on selection of secreted proteins in Olink's library, Leroy Hood at the Institute of Systems Biology for collaboration on tissue-specific proteins, and Angela Silveira at the Department of Medicine, the Karolinska Institute. Finally, we thank Alexandra Coutinho for valuable input and Anders Målarstig for providing samples and input on the plasma/serum correlation testing.

Author contributions—L. W., N. N., John Broberg, E. P., and M. L. conceptualization; L. W., N. N., John Broberg, Johan Björkesten, and M. L. formal analysis; L. W., N. N., John Broberg, Johan Björkesten, C. W., E. L., and M. L. investigation; L. W., N. N., John Broberg, E. A., and M. L. methodology; S. H., E. P., and M. L. project administration; L. W., N. N., John

Broberg, C. W., E. L., and A. F. resources; E. A. supervision; L. W., N. N., John Broberg, S. H., C. W., E. L., and M. L. validation; L. W. and Johan Björkesten. visualization; L. W., N. N., Johan Björkesten, S. H., and I. G. writing—original draft; L. W., Johan Björkesten, E. A., S. H., I. G., and M. L. writing—review and editing.

Conflict of interest—All authors are employees of Olink Proteomics AB commercializing the described method. An author may or may not be named as an inventor on current patents owned and controlled by Olink Proteomics AB, <https://www.olink.com/patents/>, as amended from time to time. All authors have financial interest in Olink Holding AB.

Abbreviations—The abbreviations used are: Ab, antibody; Ag, antigen; Amp ctrl, amplification control; bp, base pair; Cq, threshold cycle; CV, coefficient of variation; δ Cq, delta Cq; ExAmp, exclusion amplification; Ext ctrl, extension control; FBC, forward barcode; GFP, green fluorescent protein; Hyb, hybridization site; IL, interleukin; Inc ctrl, incubation control (immuno control); LLOQ, lower limit of quantification; LOD, limit of detection; mAb, monoclonal antibody; MAD, mean absolute deviation; MSD, meso scale discovery; NPX, normalized protein expression; nt, nucleotide; pAb, polyclonal antibody; pQTL, protein quantitative trait loci; PEA, proximity extension assay; qPCR, quantitative real-time PCR; RBC, reverse barcode; Rd1SP, read 1 sequencing primer; r^2 , coefficient of determination; RT, room temperature; SBS, sequencing by synthesis; SD, standard deviation; ULOQ, upper limit of quantification.

Received May 20, 2021, and in revised form, October 14, 2021
Published, MCPRO Papers in Press, October 27, 2021, <https://doi.org/10.1016/j.mcpro.2021.100168>

REFERENCES

- Hortin, G. L., and Sviridov, D. (2010) The dynamic range problem in the analysis of the plasma proteome. *J. Proteomics* **73**, 629–636
- Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867
- Smith, J. G., and Gerszten, R. E. (2017) Emerging affinity-based proteomic technologies for large-scale plasma profiling in cardiovascular disease. *Circulation* **135**, 1651–1664
- Joshi, A., Rienks, M., Theofilatos, K., and Mayr, M. (2021) Systems biology in cardiovascular disease: A multiomics approach. *Nat. Rev. Cardiol.* **18**, 313–330
- Assarsson, E., Lundberg, M., Holmquist, G., Björkesten, J., Thorsen, S. B., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S., Edfeldt, G., Andersson, A. C., Lindstedt, P., Stenvang, J., Gullberg, M., and Fredriksson, S. (2014) Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* **9**(4), e95192
- Lundberg, M., Eriksson, A., Tran, B., Assarsson, E., and Fredriksson, S. (2011) Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102
- [patent] Shen, M. R., Boutell, J. M., Stephens, K. M., Ronaghi, M., Gunderson, K., Venkatesan, B. M., Bowen, M. S., and Vijayan, K. (2016) *Kinetic exclusion amplification of nucleic acid libraries*. USPTO 20160053310:A1. US Patent, Filed October 9, 2015, and Issued February 25
- Canchola, J. A., Tang, S., Hemyari, P., Paxinos, E., and Marins, E. (2017) Correct use of percent coefficient of variation (%CV) formula for log-transformed data. *MOJ Proteomics Bioinform.* **6**, 316–317

9. Vodak, D., Lorenz, S., Nakken, S., Aashein, L. B., Holte, H., Bai, B., Myklebost, O., Meza-Zepeda, L. A., and Hovig, E. (2017) Sample-index misassignment could for example impact tumor exome sequencing. *Sci. Rep.* **8**, 5307
10. Benoy, I., Salgado, R., Colpaert, C., Weytjens, R., Vermeulen, P. B., and Dirix, L. Y. (2002) Serum interleukin 6, plasma VEGF, serum VEGF, and VEGF platelet load in breast cancer patients. *Clin. Breast Cancer* **2**, 311–315
11. Krapivner, S., Chernogubova, E., Ericsson, M., Ahlbeck-Glader, C., Hamssten, A., and van 't Hooft, F. M. (2007) Human evidence for the involvement of insulin-induced gene 1 in the regulation of plasma glucose concentration. *Diabetologia* **50**, 94–102
12. Dai, H., Alsaihe, T. A., Chalghaf, N., Riccò, M., Bragazzi, N. L., and Wu, J. (2020) The global burden of disease attributable to high body mass index in 195 countries and territories, 1990–2017: An analysis of the Global Burden of Disease Study. *PLoS Med.* **17**, e1003198
13. Enroth, S., Hallmans, G., Grankvist, K., and Gyllensten, U. (2016) Effects of long-term storage time and optimal sampling month on Biobank plasma protein concentrations. *EBioMedicine* **12**, 309–314
14. Shen, Q., Björkstén, J., Galli, J., Ekman, D., Broberg, J., Nordberg, N., Tillander, A., Kamali-Moghaddam, M., Tybring, G., and Landegren, U. (2018) Strong impact on plasma protein profiles by precentrifugation delay but not by repeated freeze-thaw cycles, as analyzed using multiplex proximity extension assays. *Clin. Chem. Lab. Med.* **56**, 582–594
15. Lee, J. E., Kim, S. Y., and Shin, S. Y. (2015) Effect of repeated freezing and thawing on biomarker stability in plasma and serum samples. *Osong Public Health Res. Perspect.* **6**, 357–362
16. Cayer, D. M., Nazor, K. L., and Schork, N. J. (2016) Mission critical: The need for proteomics in the era of next-generation sequencing and precision medicine. *Hum. Mol. Genet.* **25**, R182–R189
17. Bretherick, A. D., Canela-Xandri, O., Joshi, P. K., Clark, D. W., Rawlik, K., Boutin, T. S., Zeng, Y., Amador, C., Navarro, P., Rudan, I., Wright, A. F., Campbell, H., Vitart, V., Hayward, C., Wilson, J. F., et al. (2020) Linking protein to phenotype with Mendelian Randomization detects 38 proteins with causal roles in human diseases and traits. *PLoS Genet.* **16**, e1008785
18. Su, Y., Chen, D., Yuan, D., Lausted, C., Choi, J., Dai, C. L., Voillet, V., Duvvuri, V. R., Scherler, K., Troisch, P., Baloni, P., Qin, G., Smith, B., Kornilov, S. A., Rostomily, C., et al. (2020) Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* **183**, 1479–1495.e20
19. Consiglio, C. R., Cotugno, N., Sardh, F., Pou, C., Amodio, D., Rodriguez, L., Tan, Z., Zicari, S., Ruggiero, A., Pascucci, G. R., Santilli, V., Campbell, T., Bryceson, Y., Eriksson, D., Wang, J., et al. (2020) The immunology of multisystem inflammatory syndrome in children with COVID-19. *Cell* **183**, 968–981.e7.
20. Arunachalam, P. S., Wimmers, F., Mok, C. K. P., Perera, R. A. P. M., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., Wagh, D., Collier, J., Pellegrini, K. L., Kazmin, D., Alaaeddine, G., et al. (2020) Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220
21. Uhlén, M., Karlsson, M. J., Hober, A., Svensson, A. S., Scheffel, J., Kotol, D., Zhong, W., Tebani, A., Strandberg, L., Edfors, F., Sjöstedt, E., Mulder, J., Mardinoglu, A., Berling, A., Ekblad, S., et al. (2019) The human secretome. *Sci. Signal.* **12**, eaaz0274
22. Magis, A. T., Rappaport, N., Conomos, M. P., Omenn, G. S., Lovejoy, J. C., Hood, L., and Price, N. D. (2020) Untargeted longitudinal analysis of a wellness cohort identifies markers of metastatic cancer years prior to diagnosis. *Sci. Rep.* **10**, 16275
23. Fraser, D. D., Cepinskas, G., Patterson, E. K., Slessarev, M., Martin, C., Daley, M., Patel, M. A., Miller, M. R., O'Gorman, D. B., Gill, S. E., Pare, G., Prassas, I., and Diamandis, E. (2020) Novel outcome biomarkers identified with targeted proteomic analyses of plasma from critically ill coronavirus disease 2019 patients. *Crit. Care Explor.* **2**, e0189
24. Kalla, R., Adams, A. T., Bergemalm, D., Vatn, S., Kennedy, N. A., Ricanek, P., Lindstrom, J., Ocklind, A., Hjelm, F., Ventham, N. T., Ho, G. T., Petren, C., Repsilber, D., Söderholm, J., Pierik, M., et al. (2021) Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. *J. Crohns Colitis* **15**, 699–708
25. van der Heijden, E. H. M., Blokland, S. L. M., Hillen, M. R., Lopes, A. P. P., van Vliet-Moret, F. M., Rosenberg, A. J. W. P., Janssen, N. G., Welsing, P. M. J., Iannizzotto, V., Tao, W., Pandit, A., Barone, F., Kruijze, A. A., Radstake, T. R. D. J., and van Roon, J. A. G. (2020) Leflunomide–hydroxychloroquine combination therapy in patients with primary Sjögren's syndrome (RepurpSS-I): A placebo-controlled, double-blinded, randomised clinical trial. *Lancet Rheumatol.* **2**, e260–e269
26. Ferreira, J. P., Verdonschot, J., Collier, T., Wang, P., Pizard, A., Bär, C., Björkman, J., Boccanelli, A., Butler, J., Clark, A., Cleland, J. G., Delles, C., Diez, J., Girerd, N., González, A., et al. (2019) Proteomic bioprofiles and mechanistic pathways of progression to heart failure. *Circ. Heart Fail.* **12**, e005897
27. Bom, M. J., Levin, E., Driessen, R. S., Danad, I., Van Kuijk, C. C., van Rossum, A. C., Narula, J., Min, J. K., Leipsic, J. A., Belo Pereira, J. P., Taylor, C. A., Nieuwdorp, M., Rajmakers, P. G., Koenig, W., Groen, A. K., et al. (2019) Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease. *EBioMedicine* **39**, 109–117
28. Gruber, C. N., Patel, R. S., Trachtman, R., Lepow, L., Amanat, F., Krammer, F., Wilson, K. M., Onel, K., Geanon, D., Tuballes, K., Patel, M., Mouskas, K., O'Donnell, T., Merritt, E., Simons, N. W., et al. (2020) Mapping systemic inflammation and antibody responses in multisystem inflammatory syndrome in children (MIS-C). *Cell* **183**, 982–995.e14
29. Whelan, C. D., Mattsson, N., Nagle, M. W., Vijayaraghavan, S., Hyde, C., Janelidze, S., Stomrud, E., Lee, J., Fitz, L., Samad, T. A., Ramaswamy, G., Margolin, R. A., Malarstig, A., and Hansson, O. (2019) Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol. Commun.* **7**, 169
30. Johansson, P., Krona, C., Kundu, S., Doroszko, M., Baskaran, S., Schmidt, L., Vinel, C., Almstedt, E., Elgendy, R., Elfineh, L., Gallant, C., Lundsten, S., Ferrer Gago, F. J., Hakkarainen, A., Sipilä, P., et al. (2020) A patient-derived cell atlas informs precision targeting of Glioblastoma. *Cell Rep.* <https://doi.org/10.1016/j.celrep.2020.107897>
31. Árnadóttir, S. S., Mattesen, T. B., Vang, S., Madsen, M. R., Madsen, A. H., Birkbak, N. J., Bramsen, J. B., and Andersen, C. L. (2020) Transcriptomic and proteomic intra-tumor heterogeneity of colorectal cancer varies depending on tumor location within the colorectum. *PLoS One* **15**, e0241148
32. Majster, M., Lira-Junior, R., Höög, C. M., Almer, S., and Boström, E. A. (2020) Salivary and serum inflammatory profiles reflect different aspects of inflammatory bowel disease activity. *Inflamm. Bowel Dis.* **26**, 1588–1596
33. Fellström, B., Helmersson-Karlqvist, J., Lind, L., Soveri, I., Wu, P. H., Thulin, M., Ärnlov, J., and Larsson, A. (2020) Associations between apolipoprotein A1, high-density lipoprotein cholesterol, and urinary cytokine levels in elderly males and females. *J. Interferon Cytokine Res.* **40**, 71–74
34. Rojahn, T. B., Vorstandlechner, V., Krausgruber, T., Bauer, W. M., Alkon, N., Bangert, C., Thaler, F. M., Sadeghyar, F., Fortelny, N., Gernedl, V., Rindler, K., Elbe-Bürger, A., Bock, C., Mildner, M., and Brunner, P. M. (2020) Single-cell transcriptomics combined with interstitial fluid proteomics defines cell type-specific immune regulation in atopic dermatitis. *J. Allergy Clin. Immunol.* **146**, 1056–1069
35. Harden, J. L., Shih, Y. H., Xu, J., Li, R., Rajendran, D., Hoffand, H., and Chang, A. L. S. (2021) Paired transcriptomic and proteomic analysis implicates IL-1 β in the pathogenesis of papulopustular rosacea explants. *J. Invest. Dermatol.* **141**, 800–809
36. Perricos, A., Wenzl, R., Husslein, H., Eiwegger, T., Gstoettner, M., Weinhaeusel, A., Beikircher, G., and Kuessel, L. (2020) Does the use of the "Proseek® multiplex oncology I panel" on peritoneal fluid allow a better insight in the pathophysiology of endometriosis, and in particular deep-infiltrating endometriosis? *J. Clin. Med.* **9**, 2009
37. Rabe, T., Lazar, K., Cambroner, C., Goelz, R., and Hamprecht, K. (2020) Human cytomegalovirus (HCMV) reactivation in the mammary gland induces a proinflammatory cytokine shift in breast milk. *Microorganisms* **8**, 289
38. Bhargava, M., Viken, K., Wang, Q., Jagtap, P., Bitterman, P., Ingbar, D., and Wendt, C. (2017) Bronchoalveolar lavage fluid protein expression in acute respiratory distress syndrome provides insights into pathways activated in subjects with different outcomes. *Sci. Rep.* **7**, 7464
39. Solheim, N., Östlund, S., Gordh, T., and Rosseland, L. A. (2017) Women report higher pain intensity at a lower level of inflammation after knee surgery compared with men. *Pain Rep.* **2**, e595

40. Darmanis, S., Gallant, C. J., Marinescu, V. D., Niklasson, M., Segerman, A., Flamourakis, G., Fredriksson, S., Assarsson, E., Lundberg, M., Nelander, S., Westermark, B., and Landegren, U. (2016) Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* **14**, 380–389
41. Larssen, P., Wik, L., Czarnewski, P., Eldh, M., Löf, L., Ronquist, K. G., Dubois, L., Freyhult, E., Gallant, C. J., Oelrich, J., Larsson, A., Ronquist, G., Villablanca, E. J., Landegren, U., Gabrielsson, S., *et al.* (2017) Tracing cellular origin of human exosomes using multiplex proximity extension assays. *Mol. Cell. Proteomics* **16**, 502–511. Erratum in: *Mol Cell Proteomics*. 2017 Aug;16(8):1547
42. Björkstén, J., Enroth, S., Shen, Q., Wik, L., Hougaard, D. M., Cohen, A. S., Sörensen, L., Giedraitis, V., Ingelsson, M., Larsson, A., Kamali-Moghaddam, M., and Landegren, U. (2017) Stability of proteins in dried blood spot Biobanks. *Mol. Cell. Proteomics* **16**, 1286–1296
43. Wierenga, A. P. A., Cao, J., Mouthaan, H., van Weeghel, C., Verdijk, R. M., van Duinen, S. G., Kroes, W. G. M., Dogrusöz, M., Marinkovic, M., van der Burg, S. S. H., Luyten, G. P. M., and Jager, M. J. (2019) Aqueous humor biomarkers identify three prognostic groups in Uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* **60**, 4740–4747
44. Franzén, B., Kamali-Moghaddam, M., Alexeyenko, A., Hatschek, T., Becker, S., Wik, L., Kierkegaard, J., Eriksson, A., Muppani, N. R., Auer, G., Landegren, U., and Lewensohn, R. (2018) A fine-needle aspiration-based protein signature discriminates benign from malignant breast lesions. *Mol. Oncol.* **12**, 1415–1428
45. Bergholt, N. L., Olesen, M. L., and Foldager, C. B. (2019) Age-dependent systemic effects of a systemic intermittent hypoxic therapy *in vivo*. *High Alt. Med. Biol.* **20**, 221–230
46. Buitrago-Molina, L. E., Pietrek, J., Noyan, F., Schlue, J., Manns, M. P., Wedemeyer, H., Hardtke-Wolenski, M., and Jaeckel, E. (2021) Treg-specific IL-2 therapy can reestablish intrahepatic immune regulation in autoimmune hepatitis. *J. Autoimmun.* <https://doi.org/10.1016/j.jaut.2020.102591>
47. Li, Z., Rasmussen, T. S., Rasmussen, M. L., Li, J., Henríquez Olguín, C., Kot, W., Nielsen, D. S., and Jensen, T. E. (2019) The gut microbiome on a periodized low-protein diet is associated with improved metabolic health. *Front. Microbiol.* **10**, 709
48. Enroth, S., Berggrund, M., Lycke, M., Broberg, J., Lundberg, M., Assarsson, E., Olovsson, M., Stålberg, K., Sundfeldt, K., and Gyllenstein, U. (2019) High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer. *Commun. Biol.* **2**, 221
49. Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D. H., Hedman, Å. K., Schork, A., Page, K., Zernakova, D. V., Wu, Y., Peters, J., Eriksson, N., Bergen, S. E., Boutin, T. S., Bretherick, A. D., Enroth, S., *et al.* (2020) Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148
50. Filbin, M. R., Mehta, A., Schneider, A. M., Kays, K. R., Guess, J. R., Gentili, M., Fenyves, B. G., Charland, N. C., Gonye, A. L. K., Gushterova, I., Khanna, H. K., LaSalle, T. J., Lavin-Parsons, K. M., Lilly, B. M., Lodenstein, C. L., *et al.* (2021) Longitudinal proteomic analysis of plasma from patients with severe COVID-19 reveal patient survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep. Med.* <https://doi.org/10.1016/j.xcrm.2021.100287>