Performance Lab:

Human Immune Monitoring Center
Holden T. Maecker, PhD, Director
1651 Page Mill Road, Palo Alto, CA 94304

Analyte (s)	250 human cytokines
Technical Platform (s)	NULISAseq
Positive and negative controls, calibrators, and reference standards	Internal controls in the NULISA platform (see **Material and Methods**)
Quality control parameters for specimens/analytes	As performed by NULISA analysis software (see **Material and Methods, Appendix 1**)
Any critical pre-analytic variables	Intended for serum or plasma analysis

Table 1. Summary of analytical performance characteristics for NULISA

Accuracy	Equivalent to or better than Olink and Luminex (See **Figure 1**)
Precision: Inter-assay	CV (%) for each analyte from 9 runs showed highly consistent results for most of the analytes. (see **Figure 2**)
Precision: Intra-assay	CV (%) for each analyte from 4 runs (6 replicates in each run) showed highly consistent results for most of the analytes.(see **Figure 3, Table 2**)
Analytical sensitivity	Comparable to or better than Olink and Luminex (see **Figure 4**)
Analytical specificity including interfering substances	Antibody specificity determined by vendor. We rely on the vendor's specification.
Reportable range	~5-25 NPQ(see **Figure 4A and Figure 5**)
Reference interval (normal range)	Varies by population; see **Figure 5** for example
Standardization, harmonization, reproducibility, and ruggedness	- Ruggedness: Freeze-thaw experiment (**Figure 6**) indicates very minimal difference between 1 and 2 freeze-thaw cycles for serum samples.

	- Reproducibility: See **Figure 2** (inter-assay precision). - Standardization: An automated workflow from sample to data leads to minimal variabilities (**Appendix 1**)
Turn-around time	Setup is achievable in under 30 minutes. Pooling of NGS libraries can be completed in less than 6 hours, albeit NGS runs separately (**Appendix 1**)
Failure rate of the assay as it is to be performed in the trial	0% (8 runs with no failure, as of 1/29/2024)
Quality control and improvement procedures	- Internal controls are included in each run (**Material and Methods**) - Communicating with the vendor to improve procedures - Healthy donor controls are included
Actual number of samples of the reproducibility study	74 samples (see **Figure 2**)
How run-to-run variation (Coefficient of Variation; CV) was assessed and handled	- Inter-plate controls are included - Automated workflow from sample to data leads to minimal run-to-run variation
How inter-laboratory variability in the measurements was assessed and how these sources of variation were minimized to maintain performance at all sites within acceptable limits and to prevent drift or bias in the assay.	Inter-laboratory variability was assessed by 2 operators with 2 lots of panel kit on 2 instruments at 2 different sites, showing minimal differences (see **Figure 2**). Automated workflow resulted in minimal inter-laboratory variability.
Describe proficiency testing and results	We had 4 runs during the training. Two runs were the demo from the vendor (1 for single-plex and 1 for multiplex). The other two runs were completed by our lab under the supervision of the vendor. (**see Figure 7** for results)
Scoring procedures and type of data to be acquired: **• quantitative/continuously distributed**	- Data is continuously distributed - No absolute quantification (no standard provided)

• semi-quantitative/ordered categorical **• qualitative/non-ordered categorical**	- See **Material and Methods** for type of data (NPQ unit)
Criteria and metrics for defining significant changes (e.g., between timepoints, between responders and non-responders)	Multivariate modeling is provided by CIMAC statistician
Throughput (estimated number of samples in a given time period)	Up to 3 plates in 24 h (+ about 5 days sequencing time)
Any other performance data	Differential expression in COVID-19 patients versus healthy controls showed slightly higher overall fold changes by NULISA compared to Olink (see **Figure 8**)

Related Manuscript

Most data shown in this report are related to a manuscript, which was published in Journal of Immunology (1). Total number of available protein targets in NULISA was 203 for Figure 1, 4, 5 and 8. Current commercial assay has 250 targets.

Material and Methods

Cohort:

For most data shown in this report, we used samples from the NIH IMPACC study (IMmunoPhenotyping of A COVID-19 Cohort). The IMPACC study was conducted to investigate the clinical and immunological characteristics of COVID-19 in hospitalized patients (2,3). This observational cohort study involved a network of expert clinicians, geneticists, and immunologists with the objective of examining the relationship between the clinical course of COVID-19 and the immune response in racially, ethnically, and geographically diverse adult patient populations across the United States. Serum samples were collected at multiple time points during hospitalization, including enrollment, days 4, 7, 14, 21, and 28. Throughout the hospitalization period, clinical outcome data such as mortality and level of care were documented.

Sample:

The IMPACC sample subset used in this analysis included 23 COVID-19 patients with samples collected at up to 6 time points each, giving a total of 78 serum samples (mean age 59.3 years, 6 female/17 male). An additional 8 serum samples were obtained from 8 healthy adult donors (3 female/5 male). All samples were stored at -80°C prior to thawing for each assay. These samples were employed without any filtration or modification. This study was approved by the Stanford IRB.

Protocol:

Alamar's NULISAseq 250-plex Inflammation panel targets mostly inflammation and immune response-related cytokines and chemokines (4). First, serum samples were centrifuged at 10,000g for 10 min. Supernatant from the serum samples were then analyzed using Alamar's NULISAseq proteomic platform. For NULISA, the capture antibody is conjugated with partially double-stranded DNA containing a poly-A tail and a target-specific barcode, whereas the detection antibody is conjugated with another partially double-stranded DNA containing a biotin

group and a matching target-specific barcode. When both antibodies are incubated with a sample containing the target molecule, an immune complex is formed. The formed immune complexes are captured by added paramagnetic oligo-dT beads and subsequent dT-polyA hybridization, and the sample matrix and unbound detection antibodies are removed by washing. As dT-polyA binding is sensitive to salt concentration, the formed immune complexes are then released into a low-salt buffer. After removing the dT beads, a second set of paramagnetic beads coated with streptavidin is introduced to capture the immune complexes in the solid phase a second time. Subsequent washes remove free unbound capture antibodies, resulting in essentially pure immune complexes on the beads. Then, a ligation mix containing T4 DNA ligase and a specific DNA ligator sequence is added, allowing ligation of the proximal ends of DNA attached to the paired antibodies and thus generating a new DNA reporter molecule containing unique target-specific barcodes. The levels of the DNA reporter are then quantified by Next Generation Sequencing. Data normalization was done using an internal control spiked into each sample well to remove potential technical variation introduced during data collection. Then data was rescaled and log2-transformed to yield NULISA Protein Quantification (NPQ) units for downstream statistical analyses.

Luminex assay (EMD Millipore, Burlington, MA) and Olink immunoassay (Olink Bioscience, Uppsala, Sweden) were performed by the Human Immune Monitoring Center at Stanford University Immunoassay Team. For Luminex assay, Human 80 Plex kits were purchased from EMD Millipore Corporation, Burlington, MA., and run according to the manufacturer's recommendations with modifications described as follows: H80 kits include 3 panels: Panel 1 is Milliplex HCYTA-60K-PX48. Panel 2 is Milliplex HCP2MAG-62K-PX23. Panel 3 includes the Milliplex HSP1MAG-63K-06 and HADCYMAG-61K-03 (Resistin, Leptin and HGF) to generate a 9 plex.

For Olink immunoassay, the samples were subjected to Olink inflammatory panel multiplex assay (Target 96 Inflammation), according to the manufacturer's instructions. The inflammatory panel includes 92 proteins associated with human inflammatory conditions. NPX, Normalized Protein eXpression, is Olink's arbitrary unit which is in log2 scale.

Detectability:

To determine detectability, limits of detection (LODs) recommended by the respective manufacturers were employed. For all three platforms, LOD is defined as the mean plus 3 times

the standard deviation of the blank samples. Detectability was defined as the percentage of measured values out of the total samples above the LOD for each protein target.

Accuracy

Note: NULISA data shows relative values, currently with no calibration.

Conclusion: Performance of NULISA was equivalent to or better than existing platforms.

Correlation with 33 common targets in 3 assays:

We compared Kendall's coefficient of concordance for 33 common targets using baseline COVID samples (Figure 1A). Alamar and Olink showed moderate to high level of concordance (Kendall's coefficient mean: 0.56, median: 0.71, range: -0.14 to 0.91). Alamar and Luminex showed a moderate to low level of concordance (Kendall's coefficient mean: 0.48, median: 0.41, range: -0.30 to 0.85), which was lower compared to Alamar and Olink. Olink and Luminex showed a low level of concordance (Kendall's coefficient mean: 0.36, median: 0.42, range: -0.16 to 0.82), which was lower compared to both Alamar and Olink, and Alamar and Luminex. Overall, the results suggested varying levels of concordance between different measurement methods. Alamar and Olink exhibited the highest concordance, followed by Alamar and Luminex, and then Olink and Luminex.

Correlation in pairwise cross-platform comparisons for common targets:

Figure 1B displays the distribution of Kendall's coefficient in each platform comparison for all of the shared targets. Specifically, the comparison between Alamar and Olink (56 targets) indicated a moderate to high level of concordance (Kendall's coefficient mean: 0.56, median: 0.70, range: -0.15 to 0.93; Figure 1B, left). The Alamar and Luminex comparison (58 targets) suggested a moderate to low level of concordance (Kendall's coefficient mean: 0.36, median: 0.40, range: -0.30 to 0.85; Figure 1B, middle), which was lower compared to Alamar and Olink. Moreover, findings indicated a similar level of concordance between Olink and Luminex (38 targets; Kendall's coefficient mean: 0.35, median: 0.40, range: -0.16 to 0.82; Figure 1B, right) when compared to Alamar and Luminex. Taken together, these results highlighted varying levels of concordance between different measurement methods, mirroring the observation of the highest concordance by Alamar and Olink in comparison involving 33 shared targets.

Association between detectability and correlation:

The association between the difference in detectability and the correlation is depicted in Figure 1C. Notably, in each cross-platform comparison, strong correlations were observed when the differences in detectability were minimal. Conversely, larger disparities in detectability generally corresponded to lower correlation. Furthermore, Alamar NULISAseq exhibited greater detectability than Olink Target 96 and Luminex H80 for targets exhibiting lower overall correlation (more targets to the right of the dotted vertical line at zero in Figures 1C left and middle).

Inter-Assay Precision

Conclusion: CV (%) for each analyte from 9 runs showed highly consistent results for most of the analytes (Figure 2).

Intra-Assay Precision

Conclusion: Replicates of serum control (N=6 in each run) in 4 runs showed highly consistent CV (%) for most cytokine targets (Figure 3, Table 2).

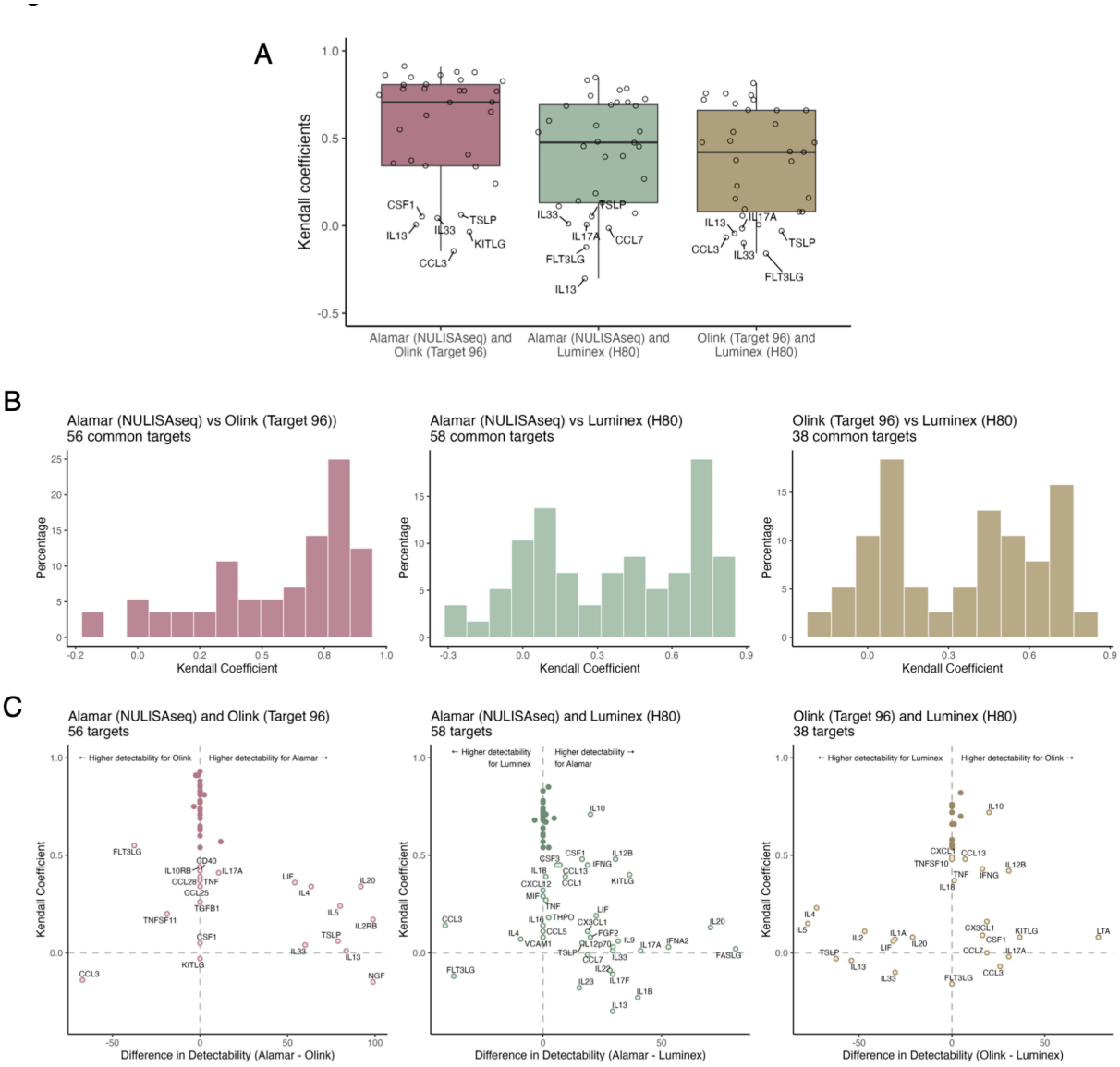

Figure 1. Correlation of measured values across platforms using baseline (Visit 1) samples for COVID-19 (n=23) and Healthy Controls (n=8). (A) Kendall coefficients of each platform pair for 33 common targets. (B) Kendall coefficient distributions for all common targets in each platform pair. (C) Association between the difference in detectability (x-axis) and Kendall correlation (y-axis) for common targets in each platform pair using baseline samples. Targets falling on the vertical dashed line showed zero difference in detectability; horizontal dashed line indicates correlation coefficient of zero. Labels highlight protein targets with Kendall coefficient below 0.50 or a difference in detectability greater than 15%. For the most part, larger differences in detectability between platforms corresponds to lower correlation. However, some targets with no detectability difference still showed poor correlation, for example, CSF1 and KITLG for Alamar-Olink, and FLT3LG for Olink-Luminex.

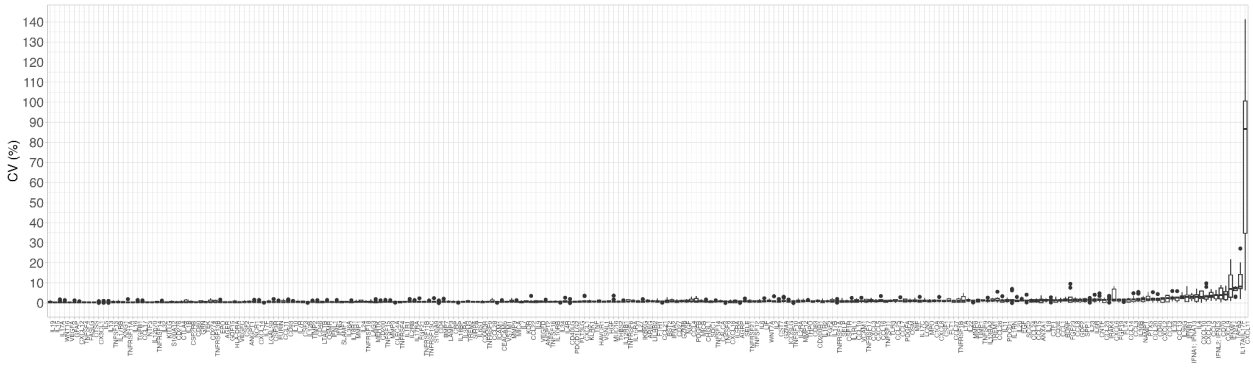

Figure 2. Inter-assay precision. CV (%) for each analyte using data with assay control from 9 runs. Analytes are sorted from left to right based on the median.

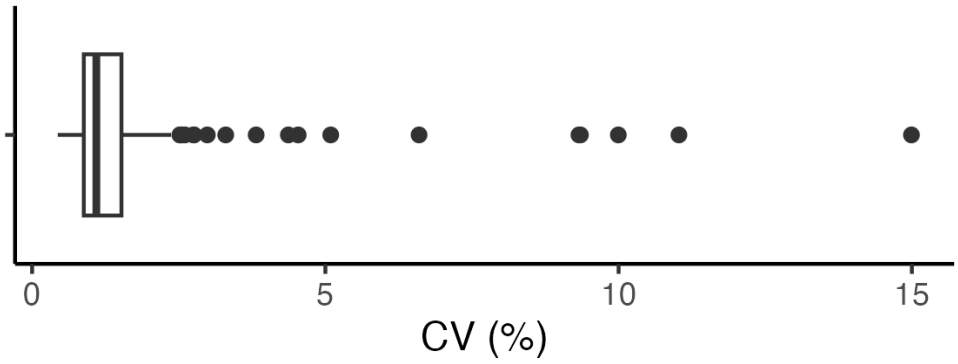

Figure 3. Intra-assay precision. CV (%) calculated for each of the 250 targets using serum control. Each run has 6 replicates. 4 runs were performed.

Table 2. Summary of CV (%)

	CV (%)
Mean (SD)	**1.5 (1.6)**
Median	**1.1**
Minimum	**0.4**
Maximum	**15.0**

Analytical sensitivity

Note: NULISA data shows relative values, currently with no calibration.

Conclusion: NULISA demonstrated highest overall detectability, followed by Olink and then Luminex (Figure 4).

Overall detectability:

Overall across-target mean detectability (i.e. percent of signal values above LOD) for the platforms was 94.7% for Alamar NULISAseq (203 targets), 86.5% for Olink Target 96 (92 targets), and 84.1% for Luminex H80 (80 targets).

Detectability with 33 common targets in 3 assays:

The detectability of protein targets varied considerably across the three assays for the 33 shared targets (Figures 4A, 4C). In the case of Alamar NULISAseq and Luminex H80, only 1 of 33 protein targets (CCL3 and IL20, respectively) exhibited ≤ 50% detectability. Olink Target 96 had 7 targets with ≤ 50% detectability (IL13, IL20, IL33, IL4, IL5, LIF, TSLP), with five proteins (IL13, IL20, IL4, IL5, and TSLP) consistently falling below the LOD in at least 75% of the samples. For these 33 targets, mean detectabilities were 96.1% for Alamar, 83.5% for Olink, and 87.2% for Luminex.

Detectability in pairwise cross-platform comparisons for common targets:

Alamar NULISAseq and Olink Target 96 (Figure 4B, left): Of 56 common targets, Alamar showed one target with ≤ 50% of samples above LOD (CCL3), while Olink showed 9 (IL13, IL20, IL2RB, IL33, IL4, IL5, LIF, NGF, TSLP). Alamar mean detectability was 97.3%; Olink mean was 86.5%.

Alamar NULISAseq and Luminex H80 (Figure 4B, middle left): Of 58 common targets, Alamar NULISAseq showed one target with ≤ 50% of samples above LOD (CCL3), while Luminex showed 4 (FASLG, IFNA2, IL20, IL22). Alamar mean detectability was 96.4%; Luminex mean was 85.6%.

Olink Target 96 and Luminex H80 (Figure 4B, middle right): Of 38 common targets, Olink showed 9 targets with ≤ 50% of samples above LOD (IL13, IL1A, IL2, IL20, IL33, IL4, IL5, LIF, TSLP), while Luminex showed 4 (IL1A, IL2, IL20, LTA). Olink mean detectability was 80.5%; Luminex mean was 83.7%.

Figure 4D is a supplemental data showing signal relative to a reference control sample. This data also shows that Alamar NULISAseq has less number of targets with ≤ 50% of samples above the value from control than the other two platforms.

In summary, in pairwise cross-platform detectability comparisons for common targets, Alamar showed the highest detectability, followed by Luminex and then Olink.

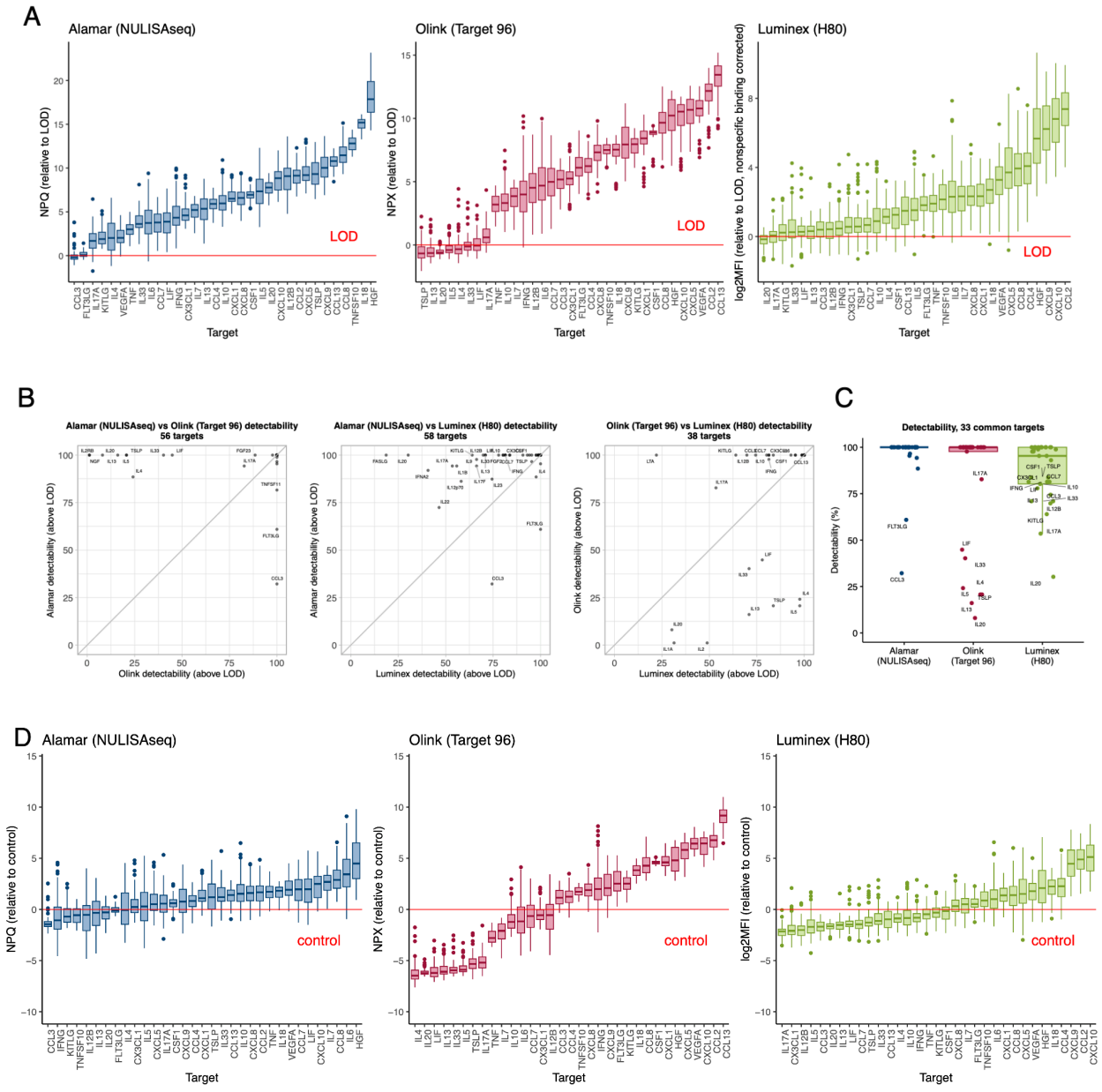

Figure 4. Detectability across platforms. (A) Measured values relative to limit of detection (LOD) for 33 common targets assessed in 78 COVID-19 serum samples from 23 patients and 8 healthy control serum samples. Targets are sorted from left to right based on the median. (B) Comparison of detectability for common targets in each platform pair. (C) Detectability of three platforms for 33 common targets. (D) Measured values relative to a reference control sample.

Normal Ranges

Conclusion: Normal range varies by population. Examples (Figure 4A and Figure 5) demonstrated ~5-25 NPQ.

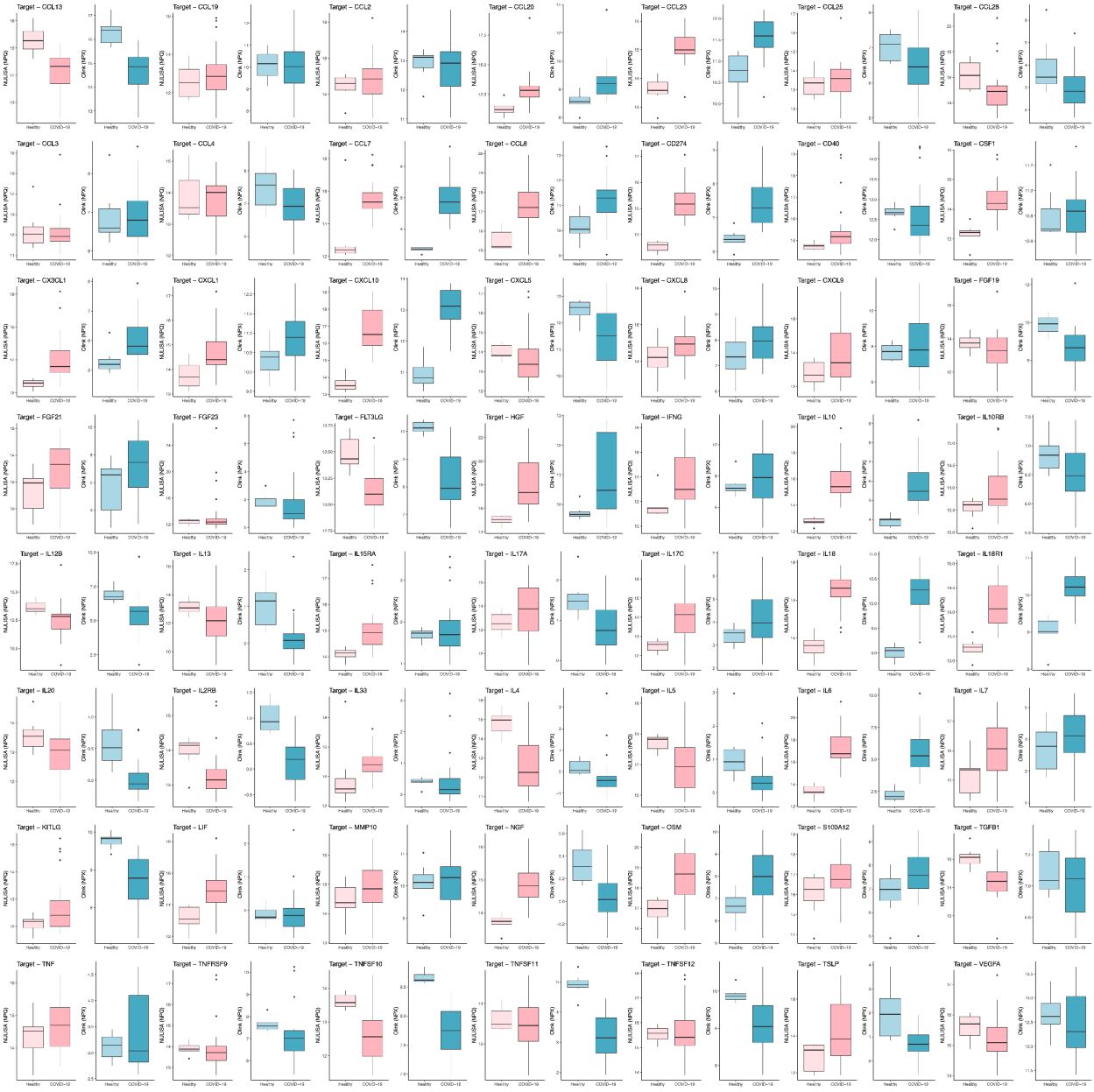

Figure 5A. Boxplot: Ranges of mean NPQ (NULISA, pink) and mean NPX (Olink, blue) per protein target at baseline (first visit) with samples from healthy donors (light color) compared to IMPACC COVID-19 patients (dark color).

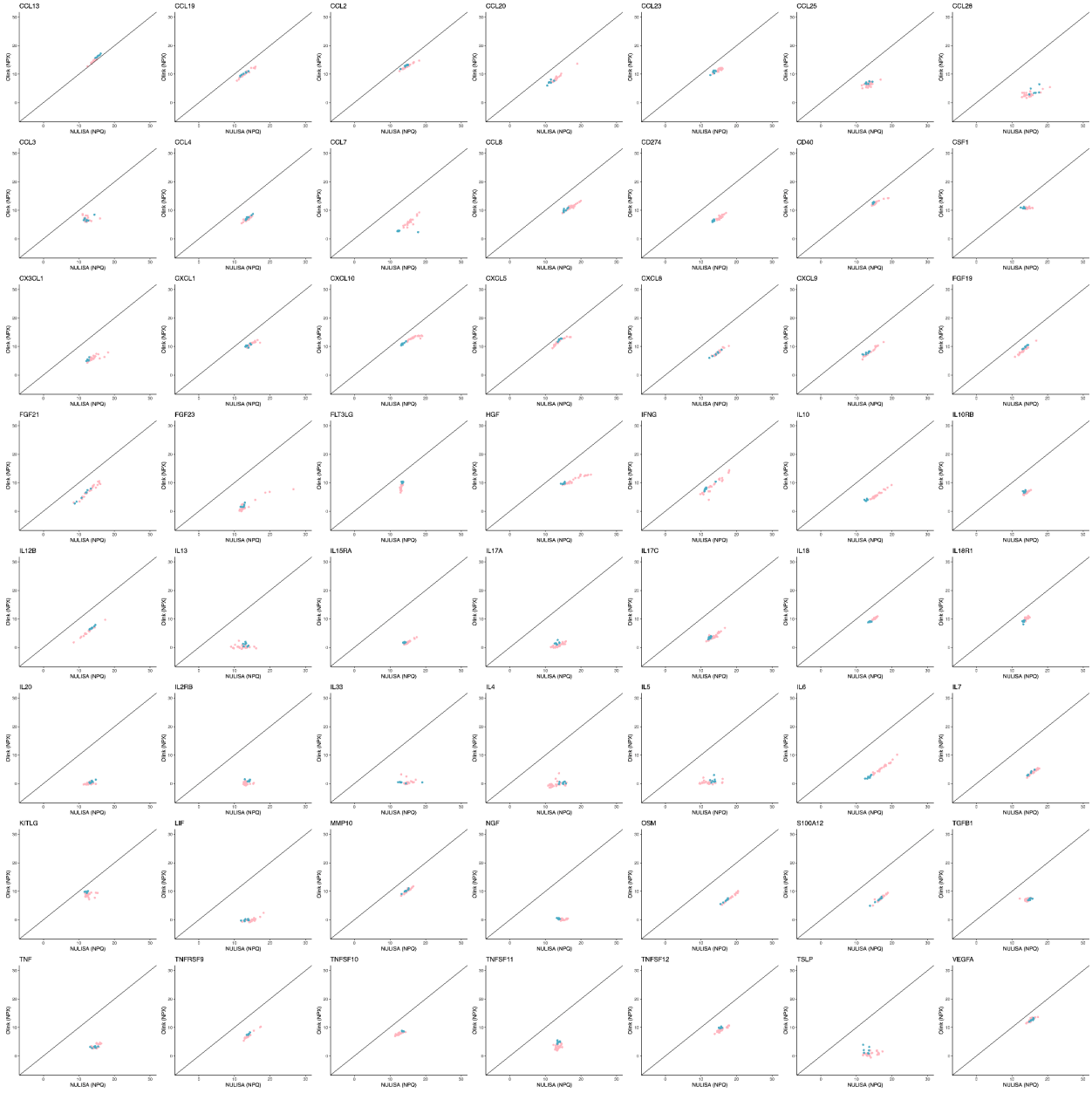

Figure 5B. Correlation plot: Ranges of mean NPQ (NULISA, x axis) and mean NPX (Olink, y axis) per protein target at baseline (first visit) with samples from healthy donors (blue) compared to IMPACC COVID-19 patients (pink).

Standardization, reproducibility, and ruggedness

Ruggedness:

Freeze-thaw experiment indicates very minimal difference between 1 and 2 freeze-thaw cycles for serum samples in the NULISA (Figure 6).

Standardization:

NULISA is available in a fully automated workflow from sample to data, which leads to minimal variabilities during sample preparation and assay run (Appendix 1).

Reproducibility:

High inter-assay precision demonstrated high reproducibility (Figure 2).

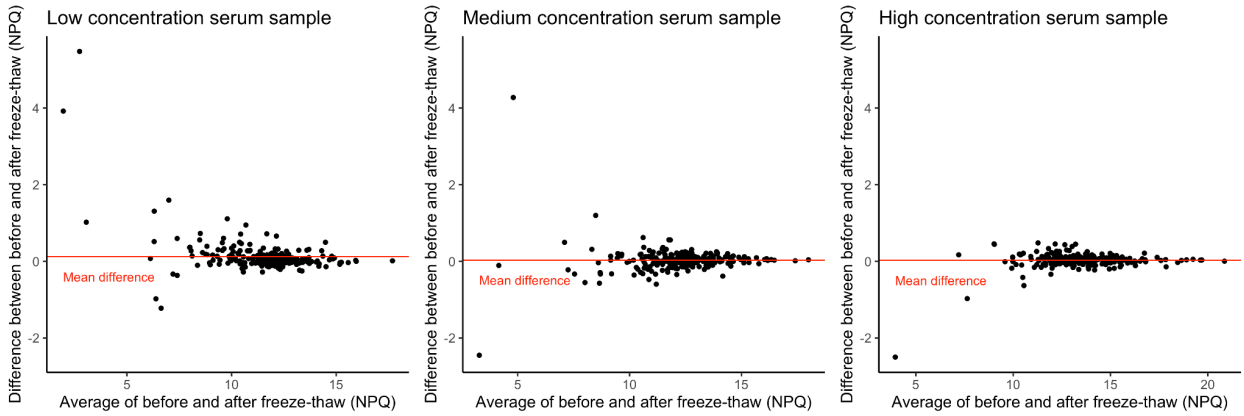

Figure 6. Bias plot for Alamar (NULISAseq) NPQ before and after freeze-thaw cycle for 3 serum samples. Since original samples were frozen, this is 1x vs. 2x freeze-thaws comparison. The freeze for the second freeze-thaw cycle was at -80°C for 24 hours. Serum samples had additional antigen spiked in to form low, medium, and high concentration samples. Each sample had 2-3 replicates on the plate. Data are the mean of the replicates for each concentration and condition. Data shown is from the NULISAseq inflammation panel with 248 targets.

Training runs and results

NULISA training and results were provided by the vendor in October - November 2023, including the following contents: setting up experiments in NULISA Control Center (NCC), instrument operation, instrument maintenance, NULISApcr assay, NULISAseq assay, data workflow, QC and analysis. Figure 7 shows the minimum difference of NPQ obtained by trainer and trainee using pooled plasma controls.

Other performance data

Differential expression in COVID-19 patients versus healthy controls:

Of 56 common targets, Alamar and Olink identified 17 common significant targets (Figure 8). Alamar and Olink log2(fold change) estimates for the 56 targets were positively correlated (r = 0.83). Alamar identified an additional 6 upregulated (CSF1, CX3CL1, FGF21, IL15RA, IL17C, LIF) and 2 downregulated (IL4, TGFB1) targets, and Olink identified 11 additional targets, all of which were downregulated (CXCL5, FGF19, IL12B, IL13, IL20, IL2RB, IL5, KITLG, TNFSF11, TNFSF12, TSLP).

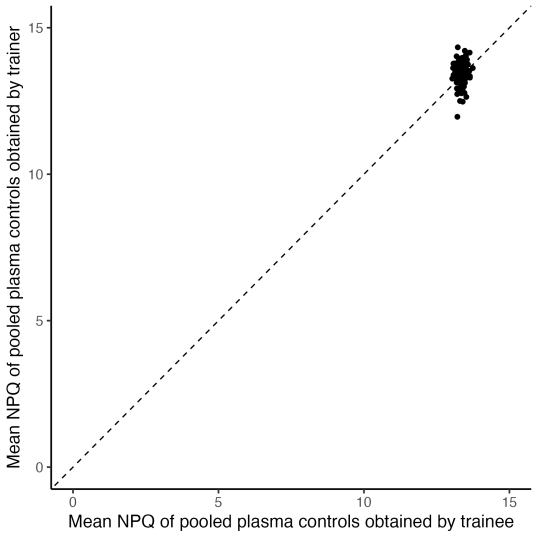

Figure 7. Training run results. Report and data were provided after training runs. The first run was the demo from the vendor, and the second run was performed by our lab under the supervision of the vendor. Pooled plasma control was used to calculate mean NPQ for each target.

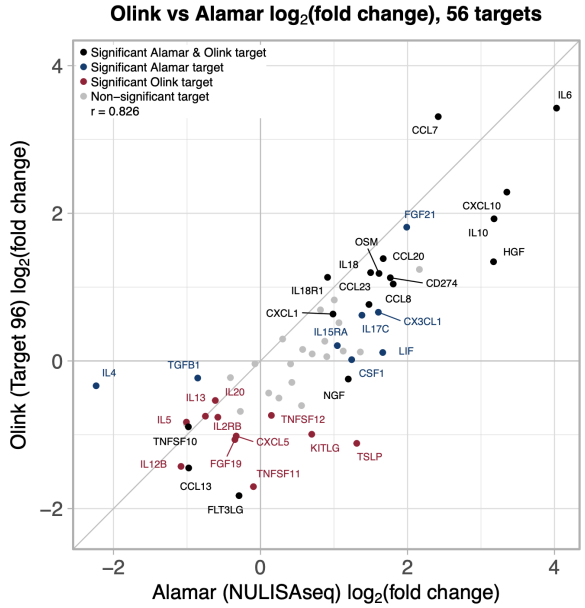

Figure 8. Differential expression comparing inflammatory protein target expression levels for baseline (Visit 1) COVID-19 patients (n=23) and Healthy Controls (n=8). Scatterplots show the relationship of log2 fold changes for all common targets in the platform pair. Overall fold changes observed by NULISA were slightly higher than Olink.

Pre-analytical sample collection or handling:

Preanalytical variables are the same as other immunoassay platforms, including Olink. Sample collection and processing are described in the CIMAC-CIDC Specimen Collection Umbrella Protocol.

Assay Chex beads:

Assay Chex beads were used for nonspecific binding correction in the Luminex assay. The procedure encompassed log transforming the mean fluorescence intensity (MFI) values of both target proteins and CHEX4 beads, followed by mean-centering the log-transformed CHEX4 MFI. Subsequently, a linear regression was performed, regressing protein MFI values against centered CHEX4 MFI values. The resulting nonspecific binding-corrected values were derived by adding the regression residuals to the intercept estimate. NULISA is not a bead-based assay, so Assay Chex beads are not possible to use. However, the multiple capture-and-release steps in the NULISA protocol serve to greatly reduce non-specific binding relative to other platforms.

When to apply NULISA:

We feel the NULISA assay is a cost-effective and high-throughput alternative to Olink, and with increased sensitivity as demonstrated in our JI publication. We think it could be implemented broadly across trials, even in addition to Olink Explore, to provide a higher-sensitivity look at pro-inflammatory cytokines and checkpoint molecules of greatest interest to our CIMAC studies.

Also, even with the much broader coverage of Olink Explore, there are ~35 non-overlapping analytes that are only in NULISA.

Appendix 1: ARGO instrument and software

The ARGO HT System is an automated, high-throughput platform for high sensitivity analysis across a range of multiplex levels (Figure 9, https://alamarbio.com/). This system provides the following features: all-in-one instrument with integrated system (<30 minutes total hands-on time), high throughput capacity processes up to 288 samples per run batch, rapid results in <8 hours for single-plex assays and <15 hours for multiplex NGS assays, integrated software and data analysis. Figure 10 shows an example of quality control results obtained from the software.

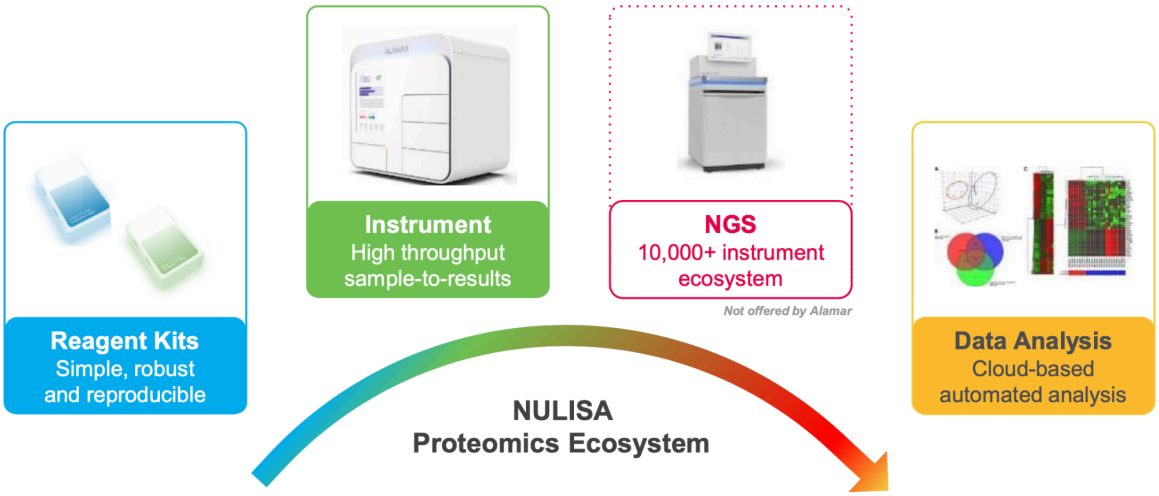

Figure 9. Workflow from sample to data. Simplicity of the ARGO HT System leads to quick technical acquisition by operator and reproducible results.

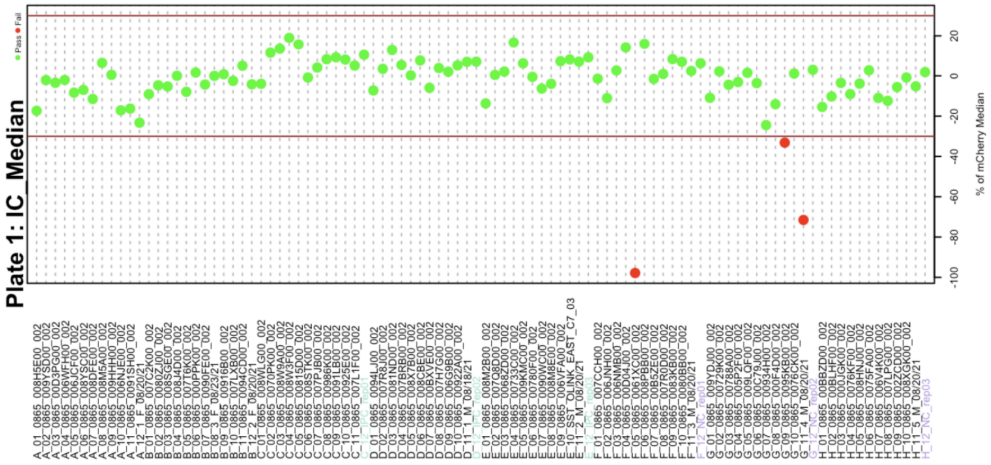

Figure 10. Example of quality control results. Internal control reads are checked with threshold (e.g. 30% of the plate median) and outliers are detected to be excluded for further analysis.

References

1. Abe K, Beer JC, Nguyen T, Ariyapala IS, Holmes TH, Feng W, et al. Cross-Platform Comparison of Highly Sensitive Immunoassays for Inflammatory Markers in a COVID-19 Cohort. J Immunol. 2024 Feb 9;ji2300729.
2. Ozonoff A, Schaenman J, Jayavelu ND, Milliren CE, Calfee CS, Cairns CB, et al. Phenotypes of disease severity in a cohort of hospitalized COVID-19 patients: Results from the IMPACC study. eBioMedicine. 2022 Sep;83:104208.
3. Team IMW, Committee on behalf of the INS, Rouphael N, Maecker H, Montgomery RR, Diray-Arce J, et al. Immunophenotyping assessment in a COVID-19 cohort (IMPACC): A prospective longitudinal study. Sci Immunol. 2021 Aug 10;6(62):eabf3733.
4. Feng W, Beer JC, Hao Q, Ariyapala IS, Sahajan A, Komarov A, et al. NULISA: a proteomic liquid biopsy platform with attomolar sensitivity and high multiplexing. Nat Commun. 2023 Nov 9;14(1):7238.