Performing relative quantitation of high-plex NULISAseq data

TECH NOTE

Sensitivity, dynamic range and reproducibility

The advancement of precision medicine requires the continued identification and validation of blood-based biomarkers to improve patient care through early identification of disease, therapeutic selection and disease monitoring. Blood-based protein biomarkers are currently used to tailor treatment in cancer, neurodegenerative and autoimmune diseases, among others. A significant barrier to the discovery and clinical application of new biomarkers is the minute concentration (pg/mL) of the disease-specific proteins in the presence of highly abundant proteins (mg/mL) in the same matrix.

Alamar Biosciences developed the NULISA™ platform, an immunoassay technology with ultra-high-sensitivity and dynamic range for protein biomarker discovery from biological fluids. The platform has been developed to have extremely low background resulting in industry-leading dynamic range (~10 logs) with attomolar sensitivity (fg/mL) in both single-plex and highly multiplexed assays. Visit www.alamarbio.com for more information on the NULISA assay and the ARGO™ HT instrument on which the workflow is completely automated.

Absolute Versus Relative Quantification

Absolute quantification, which is quantification in absolute concentration units, such as pg/ml, is not always necessary for protein assays. The need for absolute quantification depends on the specific research or clinical application and the goals of the study.
In many cases, relative quantification is sufficient for protein assays. Relative quantification is quantification relative to a chosen standard, such as a control group, and values are often expressed in units of fold change. Thus, relative quantification allows comparison of protein levels between different samples or conditions without providing an absolute measurement of protein concentration. The NULISAseq multiplex assay employs relative quantification. Many well-established data analysis methods exist for characterizing expression of analytes quantified on a relative scale, and these are particularly well suited for high-plex analysis. For example:

A. Differential expression analysis to identify analytes that are differentially expressed between distinct biological groups.
B. Longitudinal differential expression analysis to discover analytes whose expression changes in relation to a biological perturbation or stage.
C. Correlation analysis to identify analytes that correlate with another relevant quantitative factor.
D. Identification of biological pathways related to differential expression across biological conditions of interest.
E. Identification of predictive analyte signatures that are associated with a biological condition of interest.
F. Identification of biological subgroups of samples that are associated based on their molecular expression through unsupervised clustering.

However, there are situations where absolute quantification is required. The NULISAqpcr™ single-plex assays, using conventional standard curve methodology, provide precise measurement of protein concentration in absolute terms, typically expressed in units such as fg/mL or U/mL. This level of quantification is necessary when the actual amount of protein present is crucial for the study objectives, such as in clinical diagnostics, pharmacokinetics studies, or when comparing protein levels across different platforms. One clinically significant example is the Neuro-filament light (NfL) protein, a blood-based marker for neuroinflammation and neuronal injury which requires absolute quantification for clinically meaningful application (https://www.labcorp.com/tests/140555/neurofilament-light-chain-plasma). Levels above a pre-defined quantity are considered clinically high risk and no reference sample is required. This clinical test is run as a single-plex assay.

Relative quantification is the most commonly used method for mid- and high-plex assays and has been

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

demonstrated to work well in the data analysis methods described above (A-F). The adoption of absolute quantification for mid- and high-plex assays presents multiple challenges including the sourcing of quality antigens for standard curves, the potential introduction of experimental error in generating the standard curve, potential model fit error when fitting data to a standard curve, the additional costs of incorporating these standards and their associated measurements, and the reality that not all quantitative assays for the same analyte yield the same results. These challenges, while significant, can and must be overcome to translate higher plex assays to clinical application.

Whether using the absolute quantification of the NULISAqpcr assay or the relative quantification of the NULISAseq assay, the data are highly correlated, as demonstrated for a series of six different proteins (Figure 1).

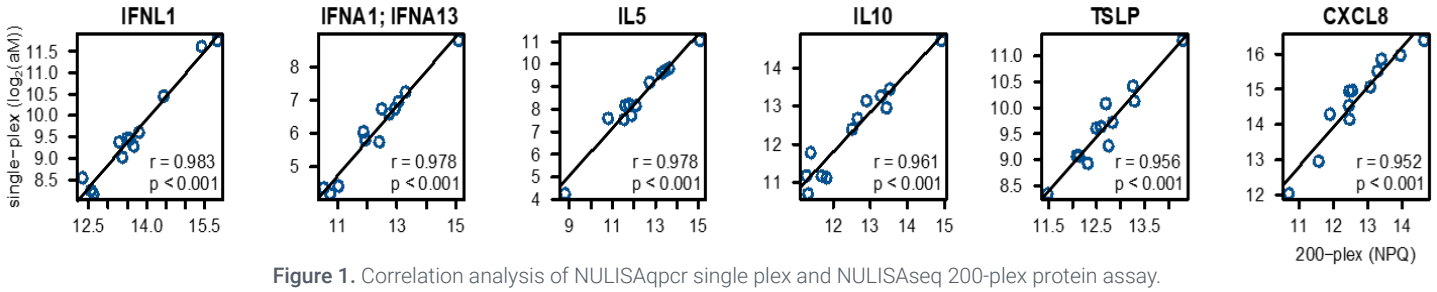

Figure 1. Correlation analysis of NULISAqpcr single plex and NULISAseq 200-plex protein assay.

How is NULISAseq multiplex data normalized?

Prior to differential expression or other types of data analysis, the data must be normalized to account for assay-specific technical variation. For NULISAseq, the first normalization step uses an internal control (IC) which is spiked into every sample in the same amount. In order to control for well-to-well variation, the value of each analyte for a given sample/well are divided by the IC value. The second normalization step is performed using interplate control (IPC) samples. There are typically 3 IPCs run on each plate. IC-normalized values for each analyte are divided by the analyte-specific median IC-normalized counts for the interplate control samples. This step helps to control for plate-to-plate variation. After these normalization steps, additional transformation (rescaling and log2-transformation) is carried out on the data to ensure distributions more amenable to statistical analysis (e.g., positive, normally distributed values). Overall, these normalization and data transformation steps yield the NULISA Protein Quantification (NPQ) unit.

Expressed in mathematical notation, the NULISAseq data normalization process is as follows:

The IC-normalized counts are $x_{it}^{IC} = x_{it}/x_{i,t=IC}$, where x_{it} are the raw sequencing read counts for sample i and target t, and $x_{i,t=IC}$ are the IC raw counts for sample i.

The IPC-normalized counts are calculated as $x_{it}^{IPC} = x_{it}^{IC}/median(IPC_t)$, where $median(IPC_t)$ represents the median IPC IC-normalized counts for target t. Finally, data is rescaled and log2 transformed to obtain $NPQ_{it} = \log_2(10{,}000 \cdot x_{it}^{IPC} + 1)$. As demonstrated in Figure 2, when applied, normalization results in data with reduced intra- and inter-plate variation. Relative quantification data generated in different experiments can also be compared as long as there are samples to bridge different runs, which is typically achieved by inter-plate controls.

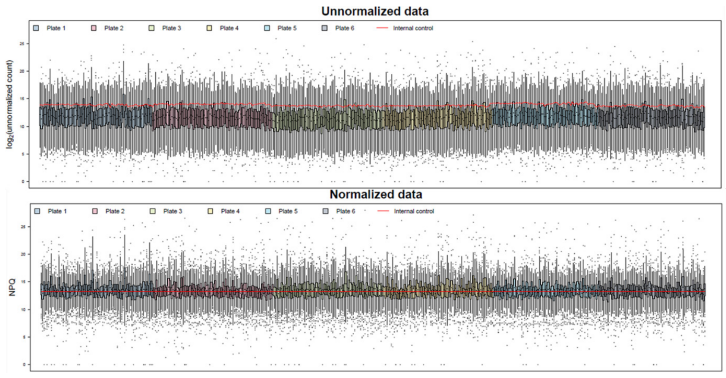

Figure 2. Impact of IC and IPC normalization on data variance across a six-plate experiment.

Visit **alamarbio.com** to learn more.

ALAMAR BIOSCIENCES

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Identification of biologically relevant changes in NULISAseq data

The best data analysis strategies for identifying biologically relevant changes in protein abundance depend on the experimental design and scientific questions of interest (for example, see A-F above). Specific strategies are provided in these two example case studies (further details provided in our pre-print):

SARS-CoV-2: The robustness and sensitivity of NULISAseq to detect changes in the relative expression of low abundant plasma proteins is highlighted in the longitudinal differential expression analysis of inflammatory protein expression in patients with mild cases of SARS-CoV-2. The expression of 200 inflammatory-associated proteins in plasma was measured for 9 mild SARS-CoV-2 patients across a 28-day course of disease progression and for 16 healthy controls. Differential expression of proteins at each time point relative to healthy controls was assessed using linear mixed-effect models with age and sex as covariates and subject as a random effect. For heat map analysis, both target and sample clustering were performed using cosine correlation distance and complete linkage (Figure 3). Distinct patterns of expression of low abundant circulating proteins such as interferons and cytokines, including IFNB1, IFNG, CXCL10, IL6 and IL10, were detected, increasing our understanding of the inflammatory response in patients with even mild cases of SARS-CoV-2.

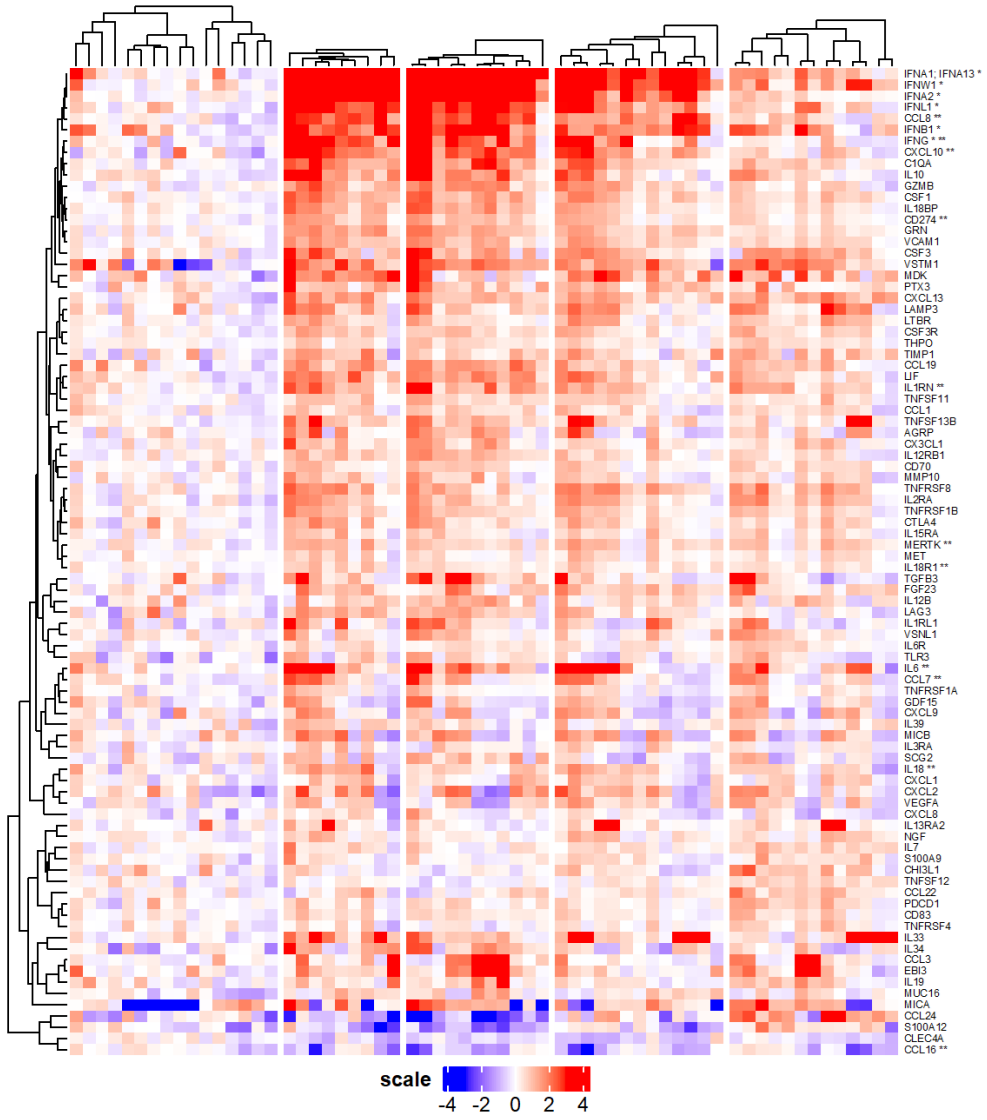

Figure 3. Changes in expression of key inflammatory markers across a 28-day disease progression of mild SARS-CoV-2. T0 represents the time point with the highest levels of the nucleocapsid (N) protein detected in the blood.

Visit **alamarbio.com** to learn more.

ALAMAR BIOSCIENCES

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Inflammatory Disease: The immune system becomes dysregulated in inflammatory diseases such as rheumatoid arthritis and lupus. Differential expression was performed comparing the relative abundance of circulating proteins between a group of inflammatory disease patients (n=21) and healthy controls (n=79) using a linear regression model that was fit to measure the impact of disease status, adjusted for age, sex and plate. Benjamini-Hochberg false discovery rate (FDR) correction was applied and statistical significance was defined using a 5% FDR cutoff. The analysis identified significant increases in expression of chemokines, cytokines and other proteins in the inflammatory disease group (Figure 4). The proteins in the upper right quadrant are upregulated in the inflammatory disease samples.

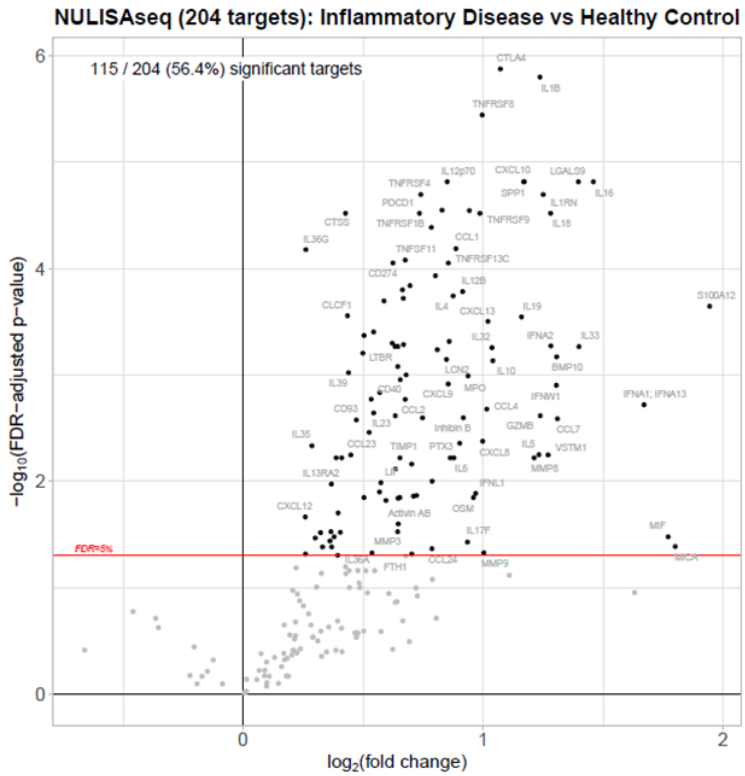

Figure 4. Differential expression of proteins between inflammatory disease and control plasma samples

Conclusion

The biomarker discovery pipeline might employ both relative and absolute quantitative methods for analyte analysis. The strategy will be influenced by many factors including the sample and analyte type, the requirements for the biomarker, and the nature of the application. Biomarkers may come in the form of single analytes or can require the measurement of multiple analytes. With the unrivaled sensitivity and accuracy of the NULISAseq multiplexed protein assay, even the most low-abundant biomarkers for cancer, neurodegenerative, and auto-immune diseases can be discovered and translated to clinical application.

Visit **AlamarBio.com** to learn more.

Alamar Biosciences, Inc.
47071 Bayside Parkway
Fremont, CA 94538
T +1 (510) 626-9888
E info@alamarbio.com

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.
©2024 Alamar Biosciences, Inc. All rights reserved. Alamar, NULISA, NULISAseq and ARGO are trademarks of Alamar Biosciences, Inc. All other trademarks and/or service marks not owned by Alamar that appear are the property of their respective owners.
M1005-1224

ALAMAR BIOSCIENCES