NULISA™ multiplex normalization & quality control

TECH NOTE

NULISA Assays

The scalable, high sensitivity chemistry of the NULISA assay combined with the ARGO™ HT instrumentation provides the ability to focus in on a single target or multiplex hundreds to potentially thousands of targets using one platform for the analysis of protein biomarkers.

NULISAseq™ multiplex assays leverage next generation sequencing (NGS) to measure the relative abundance of the protein specific barcodes in each sample. Sequencing reads (FASTQ files) are normalized and converted into NULISA Protein Quantification (NPQ) values within the Alamar Control Center analysis software. This technical note covers the workflow of NPQ calculation, assay controls, quality control (QC) metrics and additional data normalization strategies. Additional information is provided within the ARGO HT User's Guide.

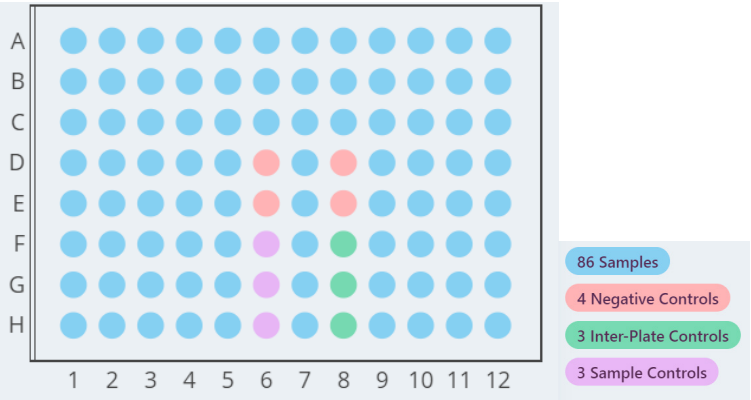


Figure 1. NULISAseq assay plate setup with controls.

Each NULISAseq assay plate includes four negative controls (NCs), three inter-plate controls (IPCs) and three sample controls (SCs), leaving space for up to 86 samples per assay plate. Please note that samples will be transferred from sample plate onto the assay plate by the ARGO system after starting the instrument run. The positions of the samples loaded into the sample plate do not reflect the final position in the assay plate on the ARGO run.

While NULISA uses a proprietary workflow to remove background signal, there may still be low levels of background from binding of capture/detection antibodies in absence of antigen. The four negative controls do not include any analyte and are used to measure background signal resulting from non-specific antibody binding. These buffer-only samples are assayed with the full panel of antibodies and therefore measure a background signal for each analyte that is used to calculate the limit of detection (LOD) and detectability (see section on Detectability).

Inter-plate controls (IPCs): The three IPCs are replicates of a pooled plasma control and used to normalize samples between experiments. Since IPCs are the same on every assay plate, IPC normalization of samples within a plate also serves as an inter-plate normalization.

Sample controls (SCs): SCs are replicates of a pooled plasma control from an independent pooled plasma source that is different from IPCs. SCs are used for calculating intra- and inter-plate coefficient of variation (CV). They are necessary for obtaining an unbiased measure of inter-plate CV when IPC normalization is used.

Internal control (IC): Each sample on the plate is spiked with the same concentration of an IC protein. An antibody pair that recognizes the IC protein is present in the antibody panel. The IC is used both as part of a well normalization procedure and as a method for assessing the uniformity of the assay run.

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

Calculation of NPQ & data normalization

NPQ data is calculated from raw sequencing counts for each analyte in each well of the assay plate. To control for well-to-well variation, the raw count value of each analyte for a given sample/well is divided by that well's IC raw count value. To control for plate-to-plate variation, the IC-normalized values for each analyte are divided by the analyte-specific median IC-normalized counts from the 3 IPCs on the plate. Data is then rescaled and log2-transformed to ensure distributions more amenable to statistical analysis (i.e., positive, normally distributed values). Overall, these normalization and data transformation steps yield the data in NPQ units. This process automatically occurs upon import of NGS sequencing data into the ARGO Control Center (ACC) software.

NPQ Calculation:

1. Each sample-specific target count is divided by that sample's IC count
2. After IC normalization, the target-specific median of the three IPC replicates is calculated
3. IC normalized counts are then divided by these target-specific IPC medians
4. Data is then rescaled by multiplying by a factor of 10,000 to shift the log-transformed data into a positive range, and the value of 1 is added to avoid taking log of zero
5. Log2 transformation then yields NPQ values

Sample	Target	Target Reads	IC Reads	IC normalized Reads	IPC Median	IC normalized Sample / IPC Median	Multiply by factor of 10,000, add 1	Log2 transform = NPQ value
Sample 01	IL4	631	33607	0.0188		0.0600	601	9.2312
IPC1	IL4	11592	36993	0.3134	0.3134			
IPC2	IL4	9666	32303	0.2992				
IPC3	IL4	11029	35064	0.3145				

Table 1. Example NPQ calculation

Quality control metrics

Several QC metrics serve to assess assay performance for specific plates, samples, and targets and to help identify potential outliers or biases. These QC metrics are displayed after the import of data into the NULISA Analysis Software (NAS) and are calculated on a per-plate basis.

			Sample QC Summary			Run QC Summary				
	Run ID	Run QC	Wells	Passed	Warning	IC CV	IPC CV	IPC Target CV	Run Detectability	Reads
○	082416ba-a2c1-4e2c-a4cd-0101ea7527ed	Passed	90	87	3	12.4%	4.7%	5.8%	97.6%	258,755,274
○	18265347-bb0c-48f7-88e7-7696277e3f43	Passed	96	93	3	14.6%	1.8%	5.1%	97.6%	266,586,275

Figure 2. Plate-specific QC criteria in NULISA Analysis Software

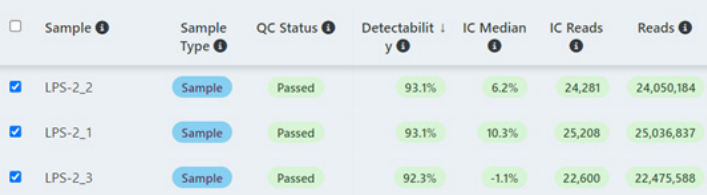
Coefficient of variation

CV is used for various QC metrics and to measure assay precision for each target. CV is calculated as the standard deviation divided by the mean for the raw counts (IC CV and IPC CV) or IC/IPC normalized counts (IPC Target CV and SC CV). Since NPQ is on a log2 scale, data should first be unlogged and the value of 1 subtracted before calculating CV. To measure assay precision, intra-plate and inter-plate CV is calculated for each target using the IC/IPC-normalized SCs.

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

Plate-specific QC criteria
To assess plate-specific performance, the following QC criteria are calculated per plate. If a specific plate exceeds the maximum recommended IPC or IC CV, we recommend re-running those samples:

- **IC CV:** CV of IC reads across all samples (maximum = 25%)
- **IPC CV:** CV of the total reads across IPCs (maximum = 25%)
- **IPC Target CV:** Across-target median CV of IC-normalized reads for IPCs (maximum = 10%)
- **Run Detectability:** Percentage of detectable targets (target is considered detectable if > 50% samples have signal > LOD) (minimum 90%). When running other sample types, this may be flagged as the sample types may have normally lower detectability.
- **Reads:** Total reads for the experiment (minimum = 100 M)



Sample	Sample Type	QC Status	Detectability	IC Median	IC Reads	Reads
LPS-2_2	Sample	Passed	93.1%	6.2%	24,281	24,050,184
LPS-2_1	Sample	Passed	93.1%	10.3%	25,208	25,036,837
LPS-2_3	Sample	Passed	92.3%	-1.1%	22,600	22,475,588

Figure 3. Sample-specific QC criteria in NULISA Analysis Software

Sample-specific QC criteria
For each sample, the following sample-specific QC criteria are calculated. If any thresholds are not met, a QC warning flag is applied. Flagged samples are not automatically excluded, but it is recommended to individually assess the performance of each flagged sample using principal components analysis (PCA) and clustered heatmaps, for example.

- **Detectability:** Percentage of targets above LOD for a sample. For the validated sample types, plasma, serum, and CSF, the minimum threshold is 90%, 90% and 70%, respectively. Although other sample types like dried blood, urine, tissue lysates have been reported to yield excellent results, these do not have a defined minimum value (minimum = 0%).
- **IC Median:** Percent deviation from the overall plate median IC reads (within +/- 40% of plate median)
- **IC Reads:** Number of IC reads for a sample (minimum = 1,000)
- **Reads:** Total reads for a sample (minimum = 500,000)

Detectability

Target detectability is the percentage of samples that are above the LOD for that target. A target is considered "detectable" if it is above LOD in greater than 50% of samples. Sample detectability refers to the percentage of targets that are above LOD for a given sample.

LOD calculation
LOD is calculated for each target as the mean plus three standard deviations of the 4 negative control wells' IC-IPC-normalized reads. This value is then rescaled, one is added, and log2-transformation is carried out to obtain LOD in NPQ.

Sample	Target	Target Reads	IC Reads	IC normalized reads	IPC Median	IC normalized Sample / IPC Median	NC Mean + 3 SDs	LOD NPQ
IPC1	IL4	11592	36993	0.3134	0.3134			
IPC2	IL4	9666	32303	0.2992				
IPC3	IL4	11029	35064	0.3145				
NC1	IL4	61	35774	0.0017		0.0054	0.0094	6.5699
NC2	IL4	20	36544	0.0005		0.0016		
NC3	IL4	0	27274	0		0		
NC4	IL4	0	32874	0		0		

Table 2. LOD calculation. IPC1-3 are triplicates of inter plate controls and NC1-4 are 4 negative controls (buffer only) included on every plate.

Note: Upon calculation of LOD, a maximum of 1 outlier NC value may be removed by NAS.

ALAMAR BIOSCIENCES

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

Data normalization strategies

As described above, the standard NPQ value is a result of both IC and IPC normalization. IC normalization minimizes technical variation between wells on the same plate, and when IPCs are from the same lot, IPC normalization minimizes technical variation between plates. One advantage of IPC normalization is that samples are not required to be randomized to different plates.

However, the lot of IPCs can change across studies conducted over a long period of time. In order for an IPC to be able to correctly normalize multiple plates, it is imperative that the lot (and thus, the normalizing signal) are consistent. In such scenarios, additional inter-plate normalization approaches may be required, each with its own advantages.

Intensity normalization

Intensity normalization allows multiple plates to be combined for analysis by setting the median NPQ value for each target to be equal across plates. Since intensity normalization relies on the assumption that the target-specific medians are identical across plates, sample randomization plays an important role. When samples are randomly assigned to plates, it is more likely that the biological medians will be similar across plates. Randomization strategies are discussed in the Sample randomization section below.

The two options are statistically equivalent (i.e., will give essentially the same results when t-tests or linear models are applied), but Option 1 might be preferred if you expect to normalize additional future plates together with Plates A and B, to avoid re-calculating global medians each time.

Intensity normalization steps:

1. Calculate the median NPQ value for each target and each plate. Then proceed with Option 1 or Option 2.

Option 1. Normalize Plate B to a specific reference plate, Plate A:

2. For each target, calculate median NPQ for Plate A minus median NPQ for Plate B.
3. Then, for each target, add this quantity to the Plate B NPQ values.

Option 2. Normalize Plate A and Plate B to the global median:

4. For each target, calculate the global median NPQ value (i.e., pool the data from all plates and calculate the median).
5. Then, for both Plate A and Plate B NPQ values, subtract the plate-specific median and add the global median.

Bridge sample normalization

The use of bridge samples may be a more robust method for normalizing multi-plate projects, as sample randomization is not required for bridge sample normalization. Rather, a minimum of 6-8 samples are needed as bridging samples. This chosen bridge sample set should ideally reflect the biological conditions of the experiment, with at least 1-2 samples from each of the major subgroups. For example, if a study involves a disease state with varying severity, the bridge samples should capture this range of severity plus any controls. If there are five or more conditions, it may be necessary to increase the number of bridge samples to accurately capture the dynamic range.

Bridge sample normalization can also be applied to bridge two batches of plates that have themselves been internally normalized to each other via IPC or intensity normalization. In this case, bridging samples are not required on every plate but rather only one set of bridge samples is required per batch.

Bridge normalization steps:

1. Choose one plate as the reference plate, Plate A.
2. For all other plates (e.g., Plate B), for each target, and for each bridge sample, calculate the within-sample differences in NPQ, Plate A minus Plate B.
3. For each target, calculate the median of these differences across all the bridge samples.
4. For each target, add the median difference to the Plate B NPQ values.
5. Repeat the steps for Plate C, Plate D, etc., if applicable.

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

Sample randomization

Even though IPC and bridge normalization do not require sample randomization, randomization of samples to plates and wells is always good practice for several reasons. First, randomization helps prevent the association (i.e., confounding) of technical variation with biological variation, and thus prevents the misattribution of true biological NPQ differences to technical effects and vice versa. For example, if disease and control samples were run on separate plates, and intensity normalization were applied, biological as well as technical variation would be eliminated from the data. Second, since disease samples often show elevated concentrations and thus elevated sequencing reads for many markers, randomization helps to balance sequencing read distributions across plates and thus helps to achieve more similar assay precision across plates and sample groups. Finally, if there are any unanticipated technical issues with IPCs or bridge samples, randomization enables the application of intensity normalization if needed.

The use of a random number generator is the easiest way to randomize samples. For example, each sample tube may be assigned a random number, tubes ordered following these numbers and placed onto the sample plate accordingly. Random number generation can be done using software such as Microsoft Excel using the RAND() and RANDBETWEEN() functions or in R using sample().

The effectiveness of randomization should be evaluated to ensure that important biological variables are balanced across plates. This can be done by simply tabulating the sample group numbers and calculating mean values of continuous variables across plates and verifying that they are sufficiently similar, or through formal statistical tests (e.g., Fisher's exact or chi-squared test for categorical variables, Wilcoxon, t-test, or ANOVA for continuous variables). In some situations, more advanced randomization methods may be useful. Stratified randomization ensures roughly equal distribution of subgroups across the plates. This is done by pre-allocating a proportional number of each subgroup to each plate, and then performing random plate assignment within each subgroup. In the case of longitudinal studies where multiple samples have been collected from the same individual, randomization should occur at the level of the individual, rather than at the sample level. This ensures that the samples from a given individual will not be subject to any plate effects and thus minimize technical confounding for within-individual differences. If individuals are members of different biological groups, these should be balanced across the plates at the level of the individuals.

In summary, sample randomization is always recommended whenever possible to minimize the impact of technical variation so that NPQ differences reflect true biological variation. Sample randomization is required for intensity normalization, since unequal distributions of biological subgroups can lead to removal of true biological variation. However, if sample randomization is not possible for some reason, such as when plates are run at much different points in time or at different laboratories, IPC or bridge sample normalization is recommended.

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.
M1016-0924

ALAMAR BIOSCIENCES