WHITEPAPER

Introduction to GeoMx® Normalization: Protein

nanoString®

Overview

This vignette introduces basic QC and normalization considerations for DSP protein data.

We will:

- Demonstrate QC of ROIs/AOIs and Probe Targets
- Show QC plots for normalization factors
- Assess concordance among the different factors
- Choose the appropriate normalization method

This small study investigates DSP Protein data from two colorectal cancer slides. Multiple ROIs were selected per slide and each ROI was segmented into Tumor and Tumor Microenvironment (TME) AOIs.

(Note: we present here data that has passed the initial QC step in the GeoMx® DSP Data Analysis Suite)

Let's first assign colors to the AOI types and to the different tissues/slides to help track them in our analyses:

AOI TYPE	TISSUE ID
Tumor	CRC Slide 1
TME	CRC Slide 2

AOI QC

The purpose of AOI-level QC is to identify AOIs with poor data that should be removed. We should look at both signal strength and background.

First, we will compute 2 metrics of AOI technical performance:

- Housekeeper geomean: this captures signal strength.
- IgG geomean: this captures background (negative controls), but in most experiments also reflects signal strength, as AOIs with more on-target signal also have more background.

To evaluate signal, we take various metrics of signal strength and look for outliers with low signal. AOIs with extremely low signal strength (i.e. many standard deviations away from the mean signal) compared to the rest may be unreliable and should be considered carefully.

Here's a look at signal strength as measured by the housekeepers. We use a log scale. (Note that log2(geomean) = mean(log2 expression).)

There are no clear outlier AOIs with low signal here, so we can safely use all AOIs. We do not see preferential clustering based on AOI Type or Tissue ID (**FIGURE 1**).

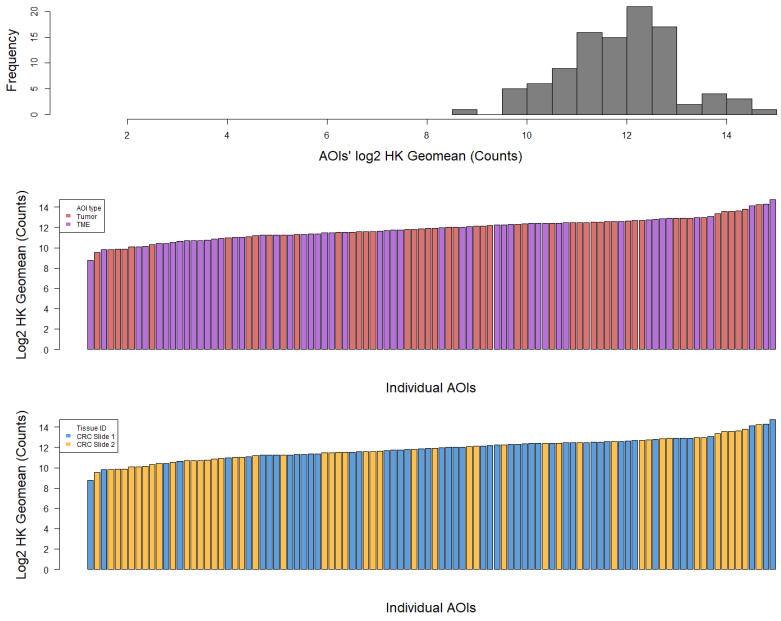


FIGURE 1

Now let's perform the identical analysis for background.

We do see some preferential clustering by Tissue ID. Keep this between mind after normalization (**FIGURE 2**).

Now let's take a look for any discrepancies in signal and background (**FIGURE 3**):

We see no points with dramatic departures from the line of best fit (i.e.. points with unusual signal/background ratios). Based on this and the previous plots, we can use all AOIs in this study.

(Note: typically signal (HK) and background (IgG) are concordant with each other).

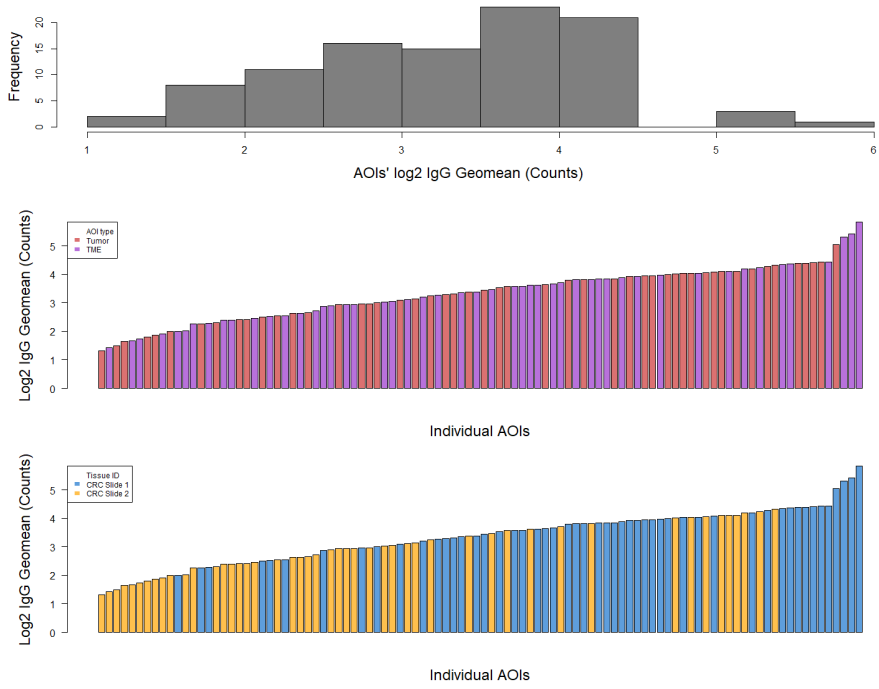


FIGURE 2

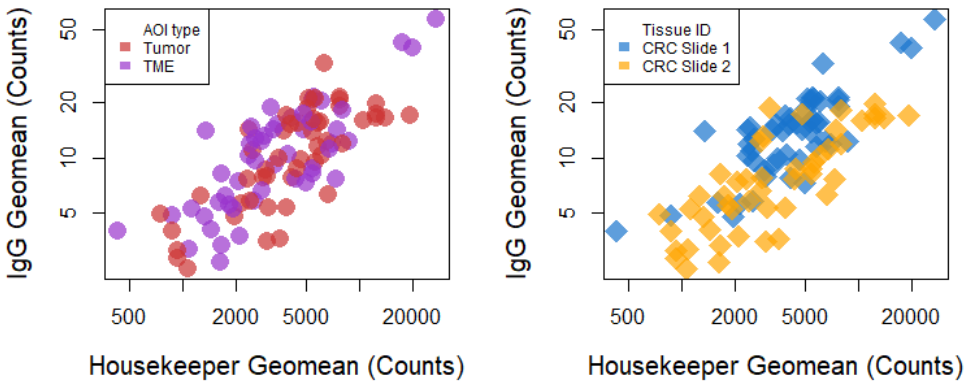


FIGURE 3

Probe QC

Probes that never rise above background should be interpreted carefully. The plot below shows a convenient way to identify poorly-performing probes. We first compute the "signal-to-background" ratio per target, which is simply each AOI's data divided by its IgG geomean (**FIGURE 4**).

We should interpret the figure by target. For example: ER-alpha has above-IgG counts (log signal-to-background > 0), but the fact that it is lower than Ms IgG1 and shows such a limited expression range could mean its signal is all in the background. This target should be interpreted carefully and/or potentially excluded from the study.

Note: within a given target, it is valuable to assess the distribution of AOIs to determine if certain AOIs are above signal while others are not; such differences can be the result of differing AOI type and/or Tissue ID and be of biological importance.

Normalization

This dataset presents four choices for normalization: housekeeper normalization, IgG normalization, area normalization, and nuclei normalization.

A hybrid approach of background-subtraction followed by housekeeper normalization is also possible, but it is not recommended unless the signal and background are highly divergent from each other with very poor correlation over a wide range of values.

Before choosing a normalization method, we must QC the potential normalization factors.

The theory is simple: if multiple probes all accurately measure signal strength, they should be highly correlated with each other.

Let's first investigate individual housekeeper counts in depth to learn more. Note the use of logscale in the plots below (**FIGURE 5**).

There are multiple important findings from the plots (**FIGURE 6**):

- S6 and GAPDH are consistent with each other.
- S6 and GAPDH do not display concerning trends, such as differing housekeeper ratios based on AOI type or Tissue ID.
- Histone H3 behaves divergently when comparing the two tissues. It should be removed as a housekeeper if we would like to make comparisons between the two tissues.

Next, lets investigate background (negative IgGs) in depth (**FIGURE 7**).

There are important findings from the plots (**FIGURE 8**):

- There is only weak correlation between the 3 IgGs.
- IgG counts are universally low (<30). This suggests their log-scale counts are statistically unstable, which explains their poor concordance with each other.

We conclude that negative IgGs should not be used as a normalization factor in this study.

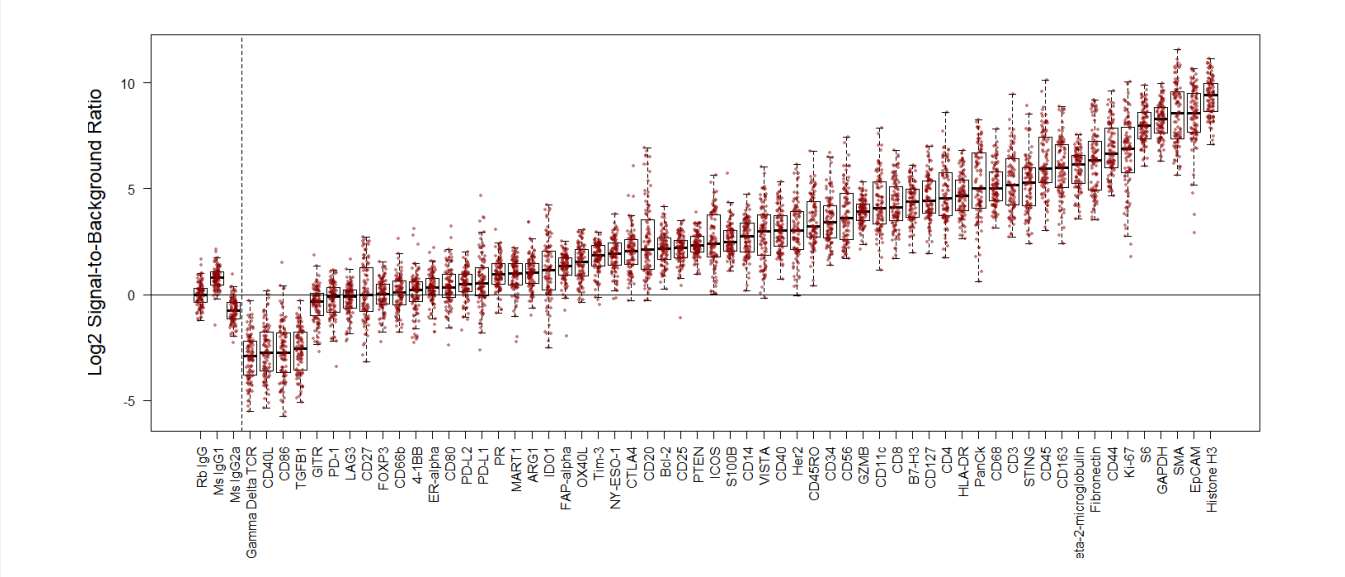
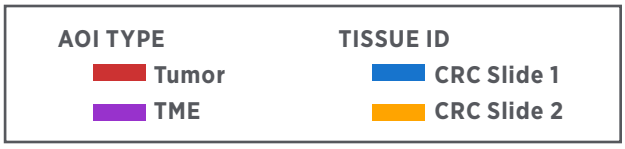


FIGURE 4



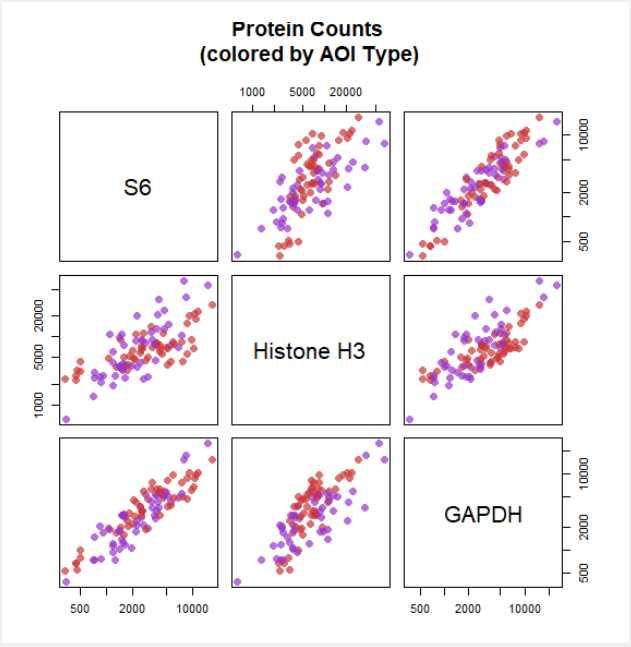


FIGURE 5

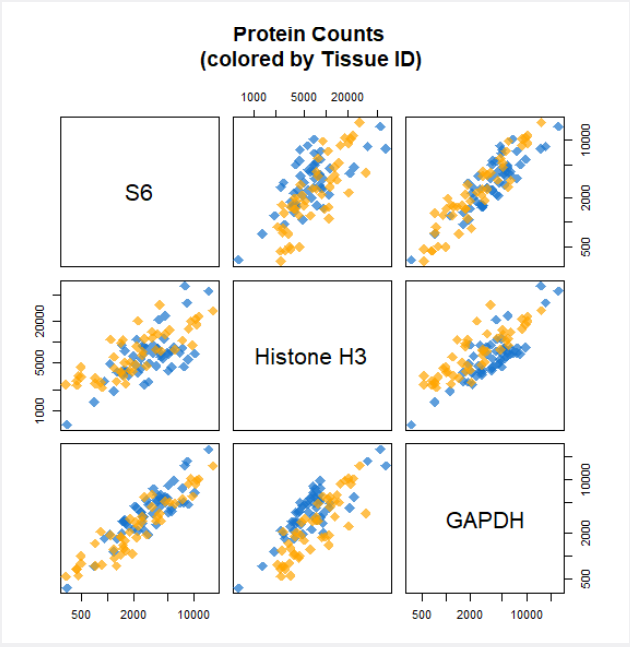


FIGURE 6

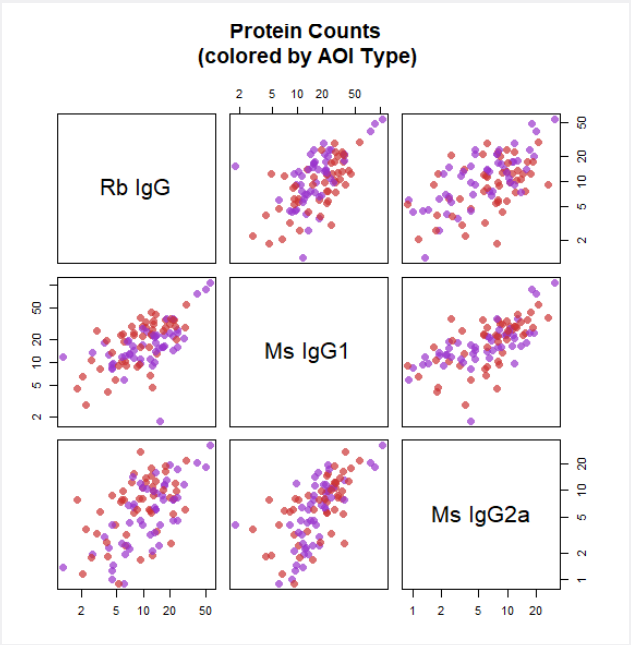


FIGURE 7

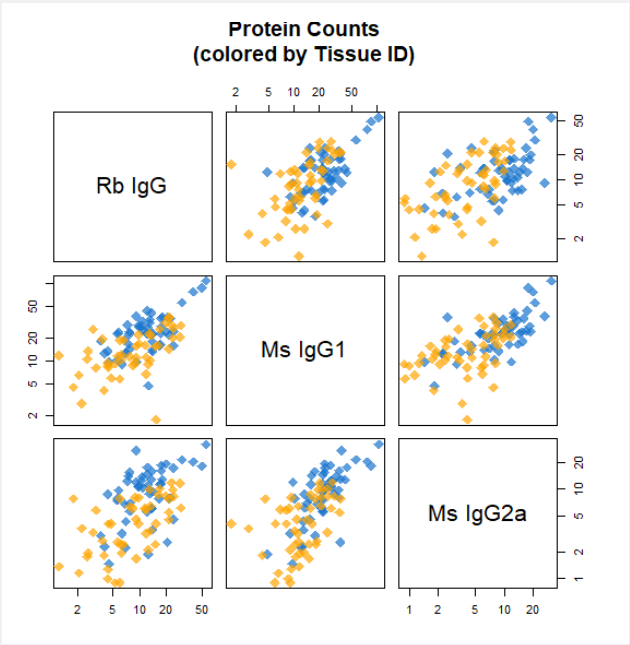


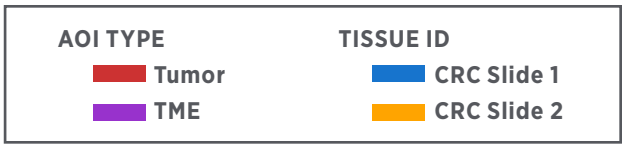
FIGURE 8

Finally, we look at the concordance of all possible normalization factors: the housekeeper geomean, the IgG geomean, Area, and Nuclei. Note: as described above, we have removed Histone H3 from the calculation of housekeeper geomean; this step is important to best reflect the normalization factor's performance. We are also careful in how we interpret the IgG geomean.

We observe a mild correlation between signal/background. This is expected given the low IgG counts for this particular experiment (**FIGURE 9**).

Area and nuclei are strongly correlated but are only mildly concordant with signal strength as measured by housekeepers (**FIGURE 10**).

Given the strong correlation between the selected housekeepers (S6 and GAPDH), the low IgG counts/concordance, and the weak correlation between area/nuclei vs. HKs, we would prefer HK normalization for this study.



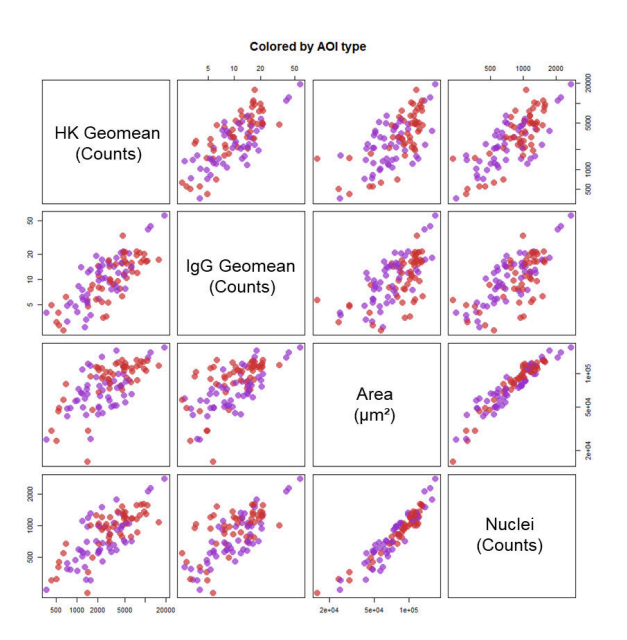


FIGURE 9

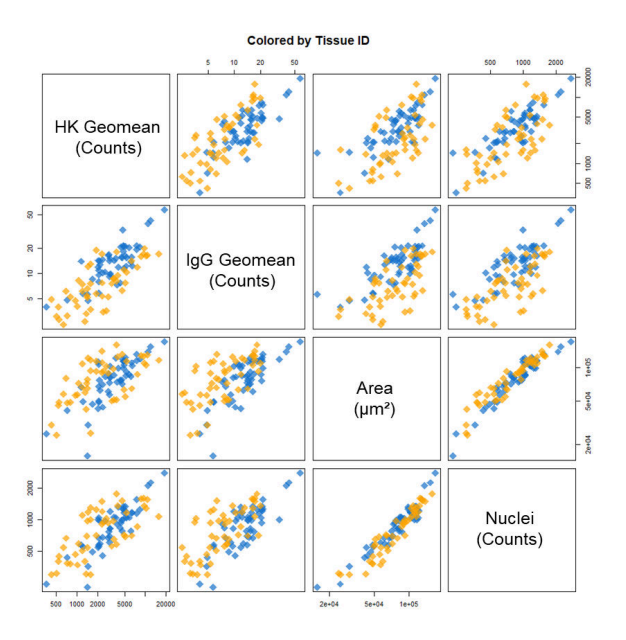


FIGURE 10

For more information, please visit nanostring.com

NanoString Technologies, Inc.
530 Fairview Avenue North
Seattle, Washington 98109
T (888) 358-6266
F (206) 378-6288
nanostring.com
info@nanostring.com

Sales Contacts
United States us.sales@nanostring.com
EMEA: europe.sales@nanostring.com
Asia Pacific & Japan apac.sales@nanostring.com
Other Regions info@nanostring.com

JUL 2020 MK2593