Use of the “Evaluate-Normalization-Options” DSP DA Script

Intended use

This script was designed for data from the GeoMx protein assay. It will also work with data from the GeoMx RNA assay, but some plots will be extraneous. If the nCounter is used to count probes, then only ERCC-normalized data should be run through this script.

This script does the following:

1. Automatically identifies relevant variables from your segment annotations for plotting
2. Arbitrarily assigns colors to the segment annotations identified in step 1
3. Computes multiple normalization factors, including negative control IgGs, housekeepers, are, and nuclei, for comparison
4. Produces a plot used to QC the negative control IgGs
5. Produces a plot used to QC the housekeeping proteins
6. Produces a plot used to QC the normalization factors computed in step 3

Relevant resources

White Paper “Introduction to GeoMx Normalization: Protein” https://blog.nanostring.com/geomx-online-user-manual/Content/PDF_downloads/MK2593_GeoMx_Normalization-Protein.pdf

Setting User Parameters

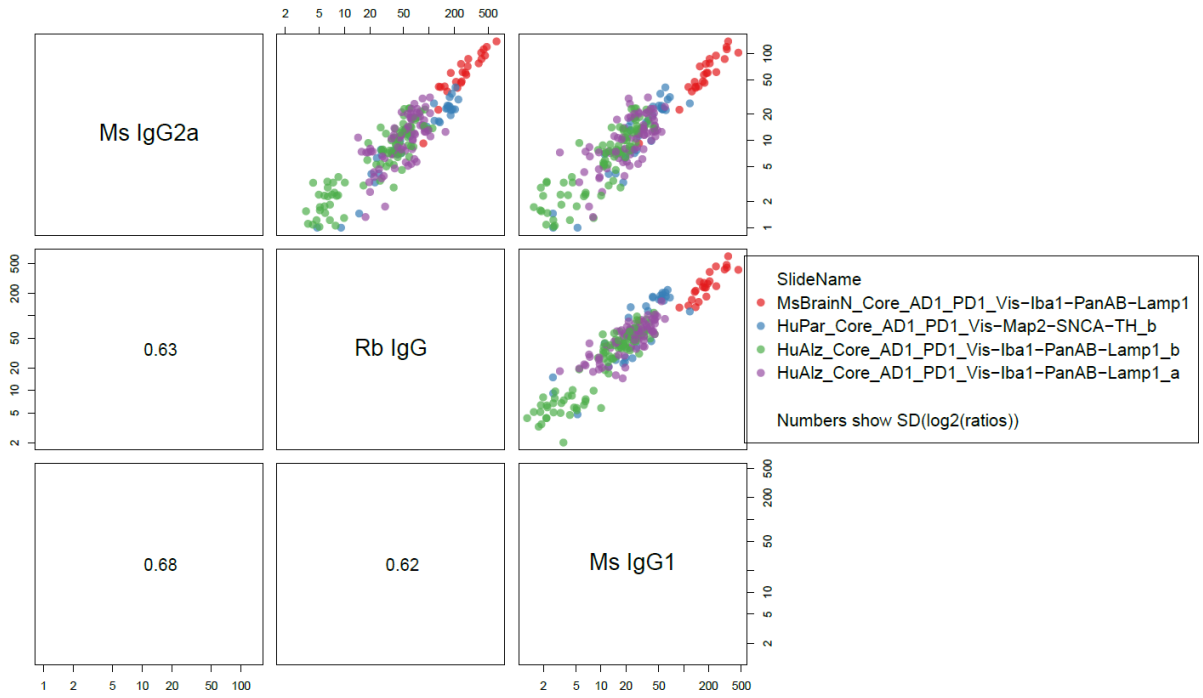
Here is a setting that can be easily adjusted by the user in the script:

plot_factor – this allows the user to color by their annotation factor of interest for the QC plots.

Interpretation of results

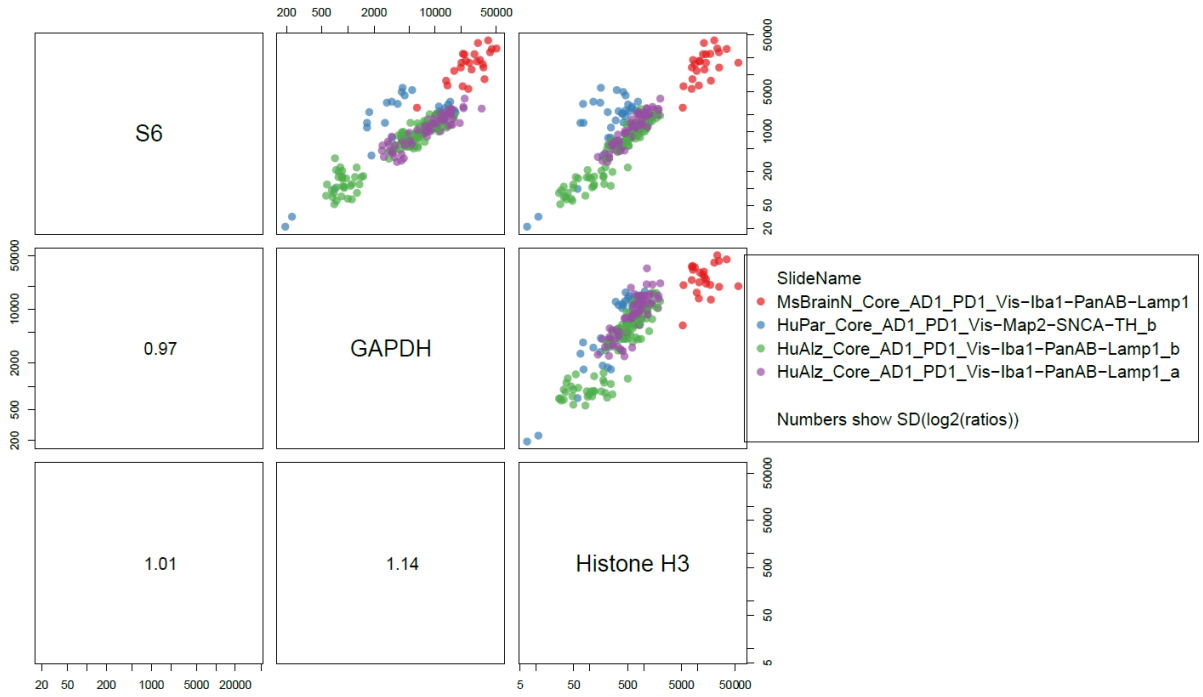
First, we will look at the QC plots for housekeepers and negative control IgGs. Our motivating theory is simple: if several probes all accurately measure signal strength, they should be highly correlated with each other. More precisely, the log-ratios between them should have low SDs (this latter criterion is similar in spirit to the geNorm algorithm).

Below is an example plot QC-ing the IgGs:



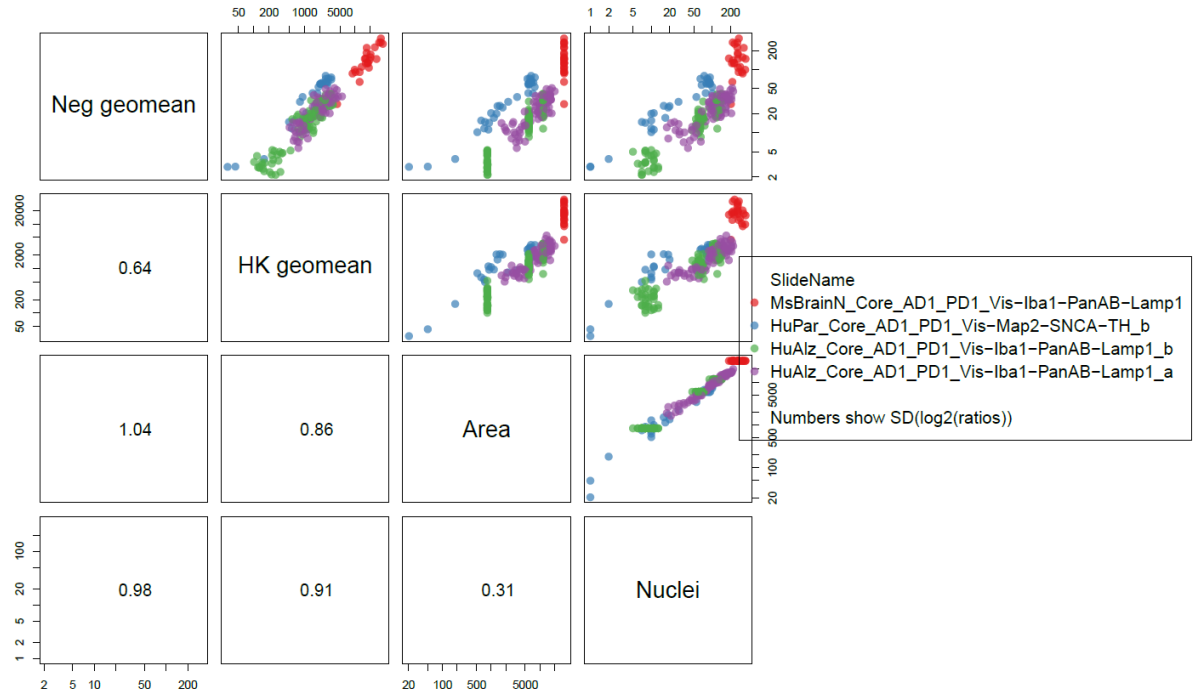
Above we see good concordance amongst the IgGs, confirming they all can be used. Numbers in the lower-left panels show the SD of the log2-ratios between IgGs. Importantly, we do not see a tendency for one IgG to be offset from the others, suggesting there's no between-slide bias in calculation of background.

Now let us look at the same plot drawn for the housekeepers, shown below:



Above we see a tendency for the blue-colored slide to over-express S6. Housekeeper normalization might be better without this protein. Though the offset of the red points in the middle-right cluster casts some doubt on GAPDH as well.

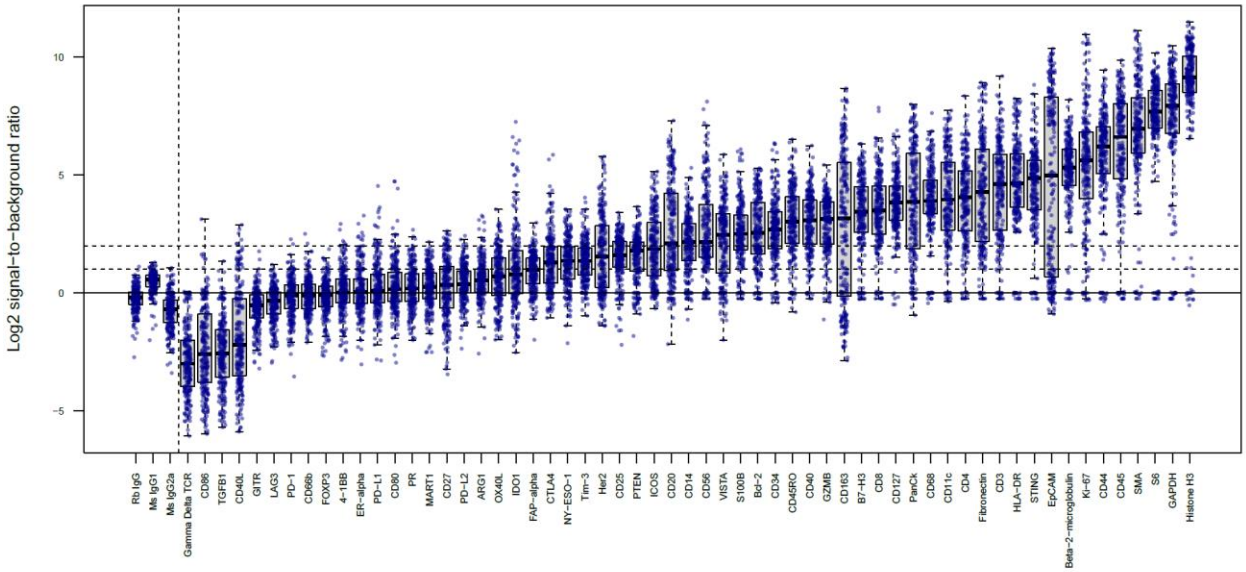
Finally, let us look below at the overall agreement of the housekeeper factors:



Observations and conclusions we can make:

- The IgGs and the housekeepers agree nicely, suggesting that if we normalize using one of them, the other will leave little artifactual signal in the data. If these factors diverged strongly, we would know that normalization with one of them would fail to account to the other, leaving an artifact in the data that must be accounted for in downstream analysis.
- Area and nuclei are highly consistent with each other (SD log2 ratio of just 0.31).
- Area and nuclei diverge somewhat from the probe-based normalization factors Neg geomean and HK geomean. This suggests that signal strength is not purely a result of area/cell count, or alternatively, that the neg and HK geomeans are noisy metrics.
- The concordance of Negs/HKs suggests their performance is adequate, leading to the conclusion that area/nuclei are noisy measurements of signal strength in this data.

This script also produces a QC plot for protein expression, shown below:



The above plot helps us identify proteins with no useful signal. For example:

- IgGs are plotted on the far left of the plot.
- LAG3 hovers around background in all segments and should probably be excluded from analysis.
- PD-L1 is mostly near-background, but it has meaningfully high signal in a handful of segments.
- CD40L seems to have lower background than the negative controls. But its long range, and especially the existence of points well above background, suggests this protein has interpretable data.