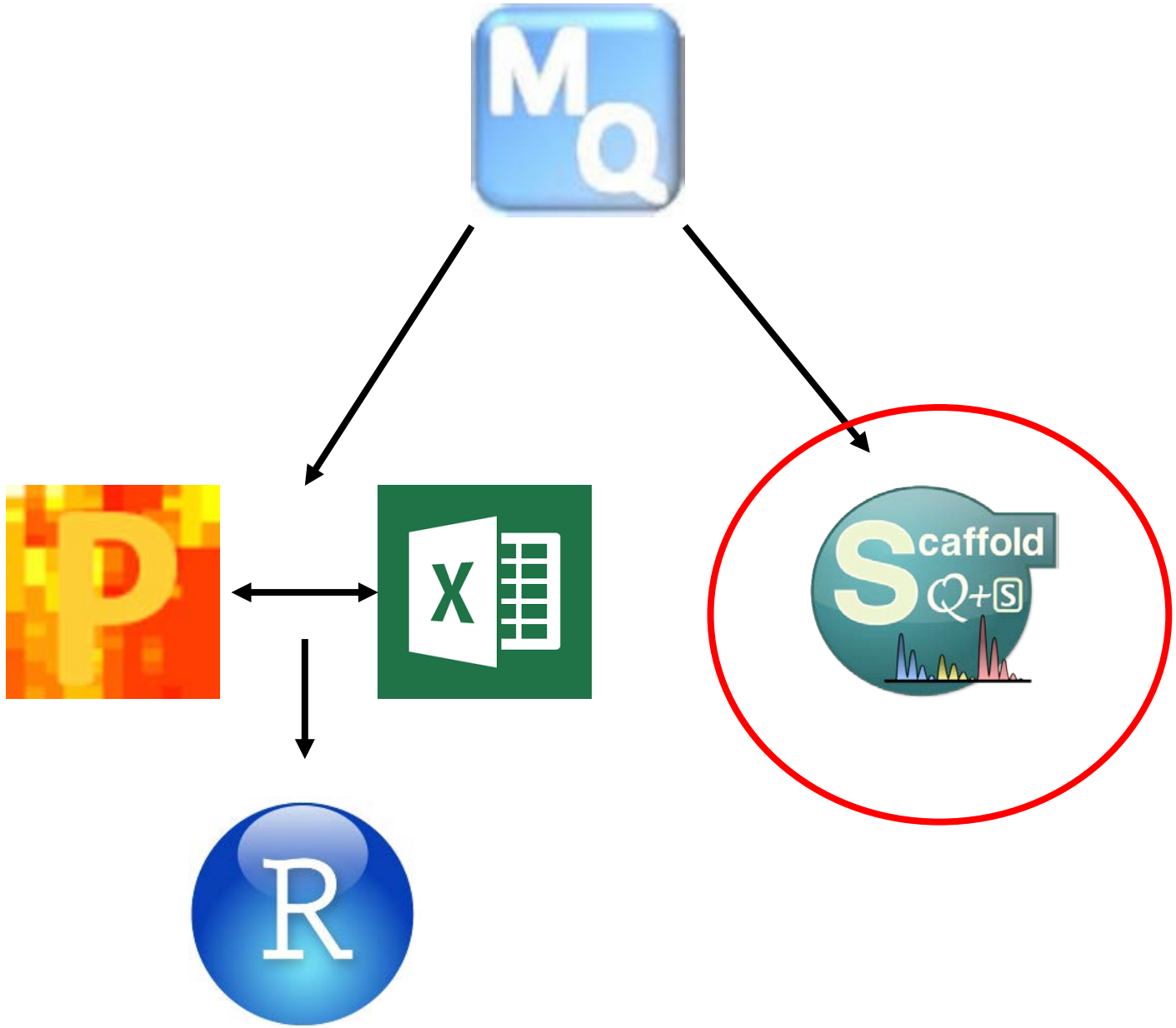
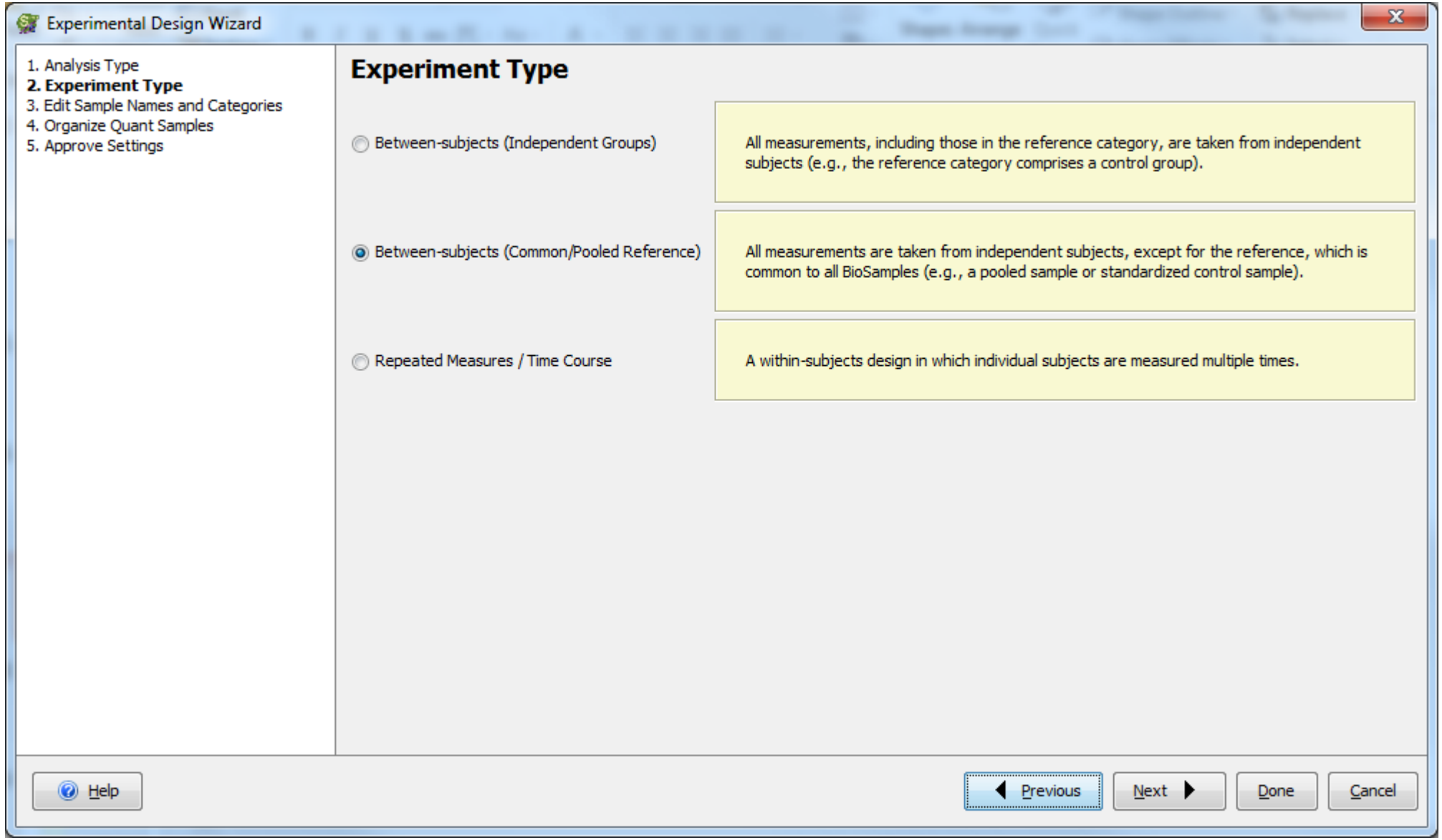
Basics of Using Scaffold and Visualizing Data

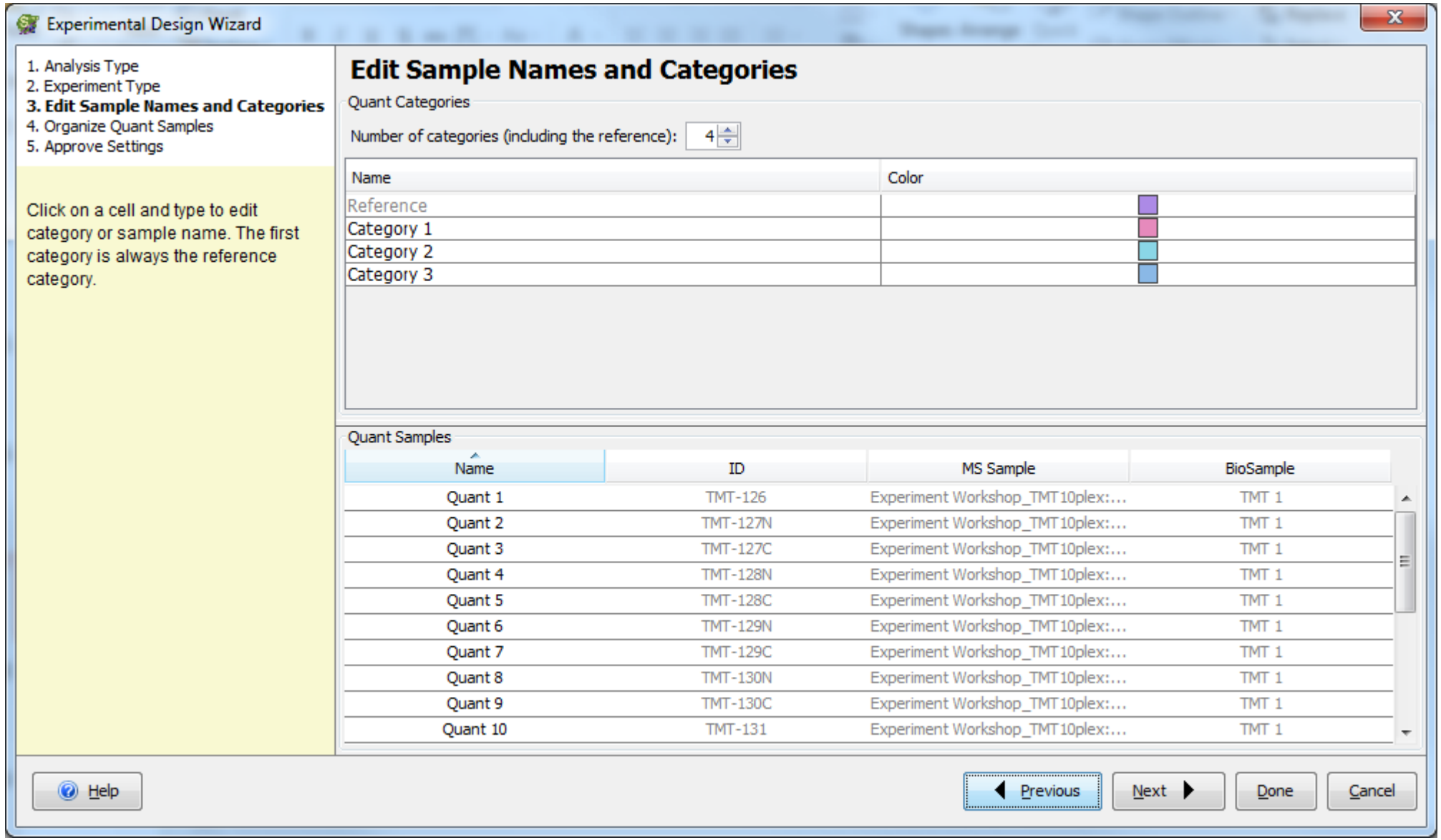
Stephanie Byrum, PhD


MQ
P
X
R
Scaffold
Q+S

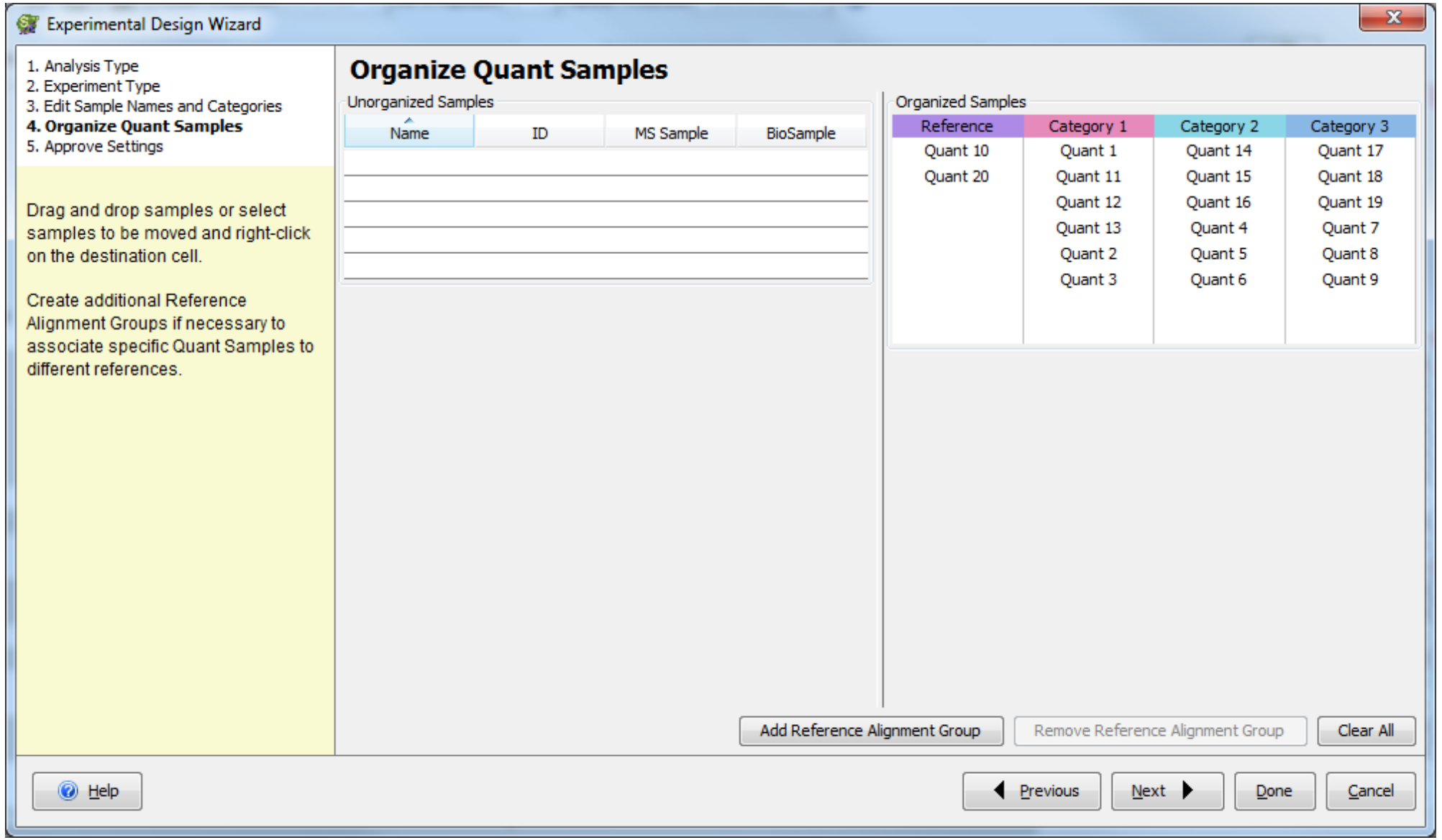
Scaffold Q+S Quant Module



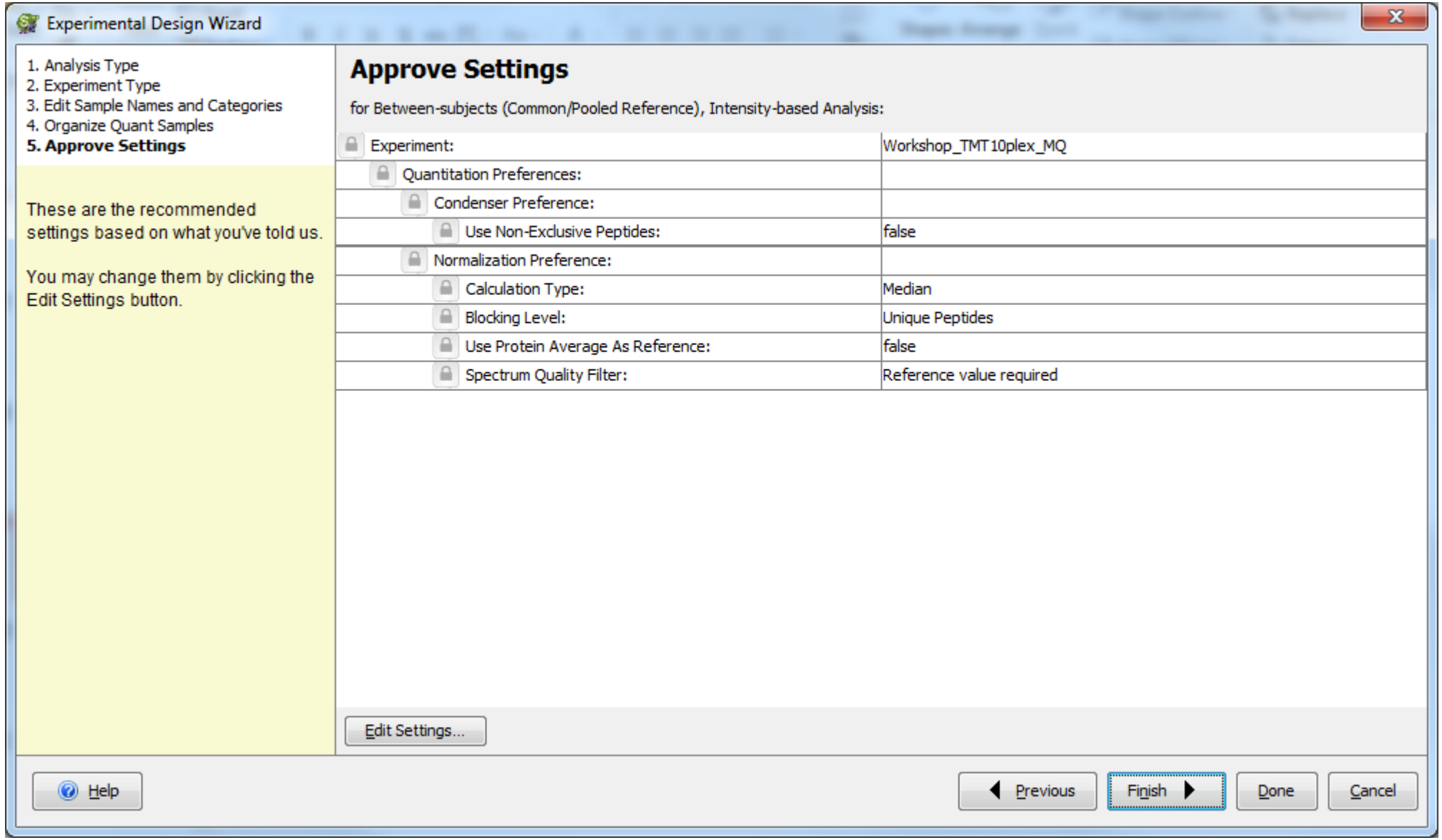
Scaffold Q+S Quant Module



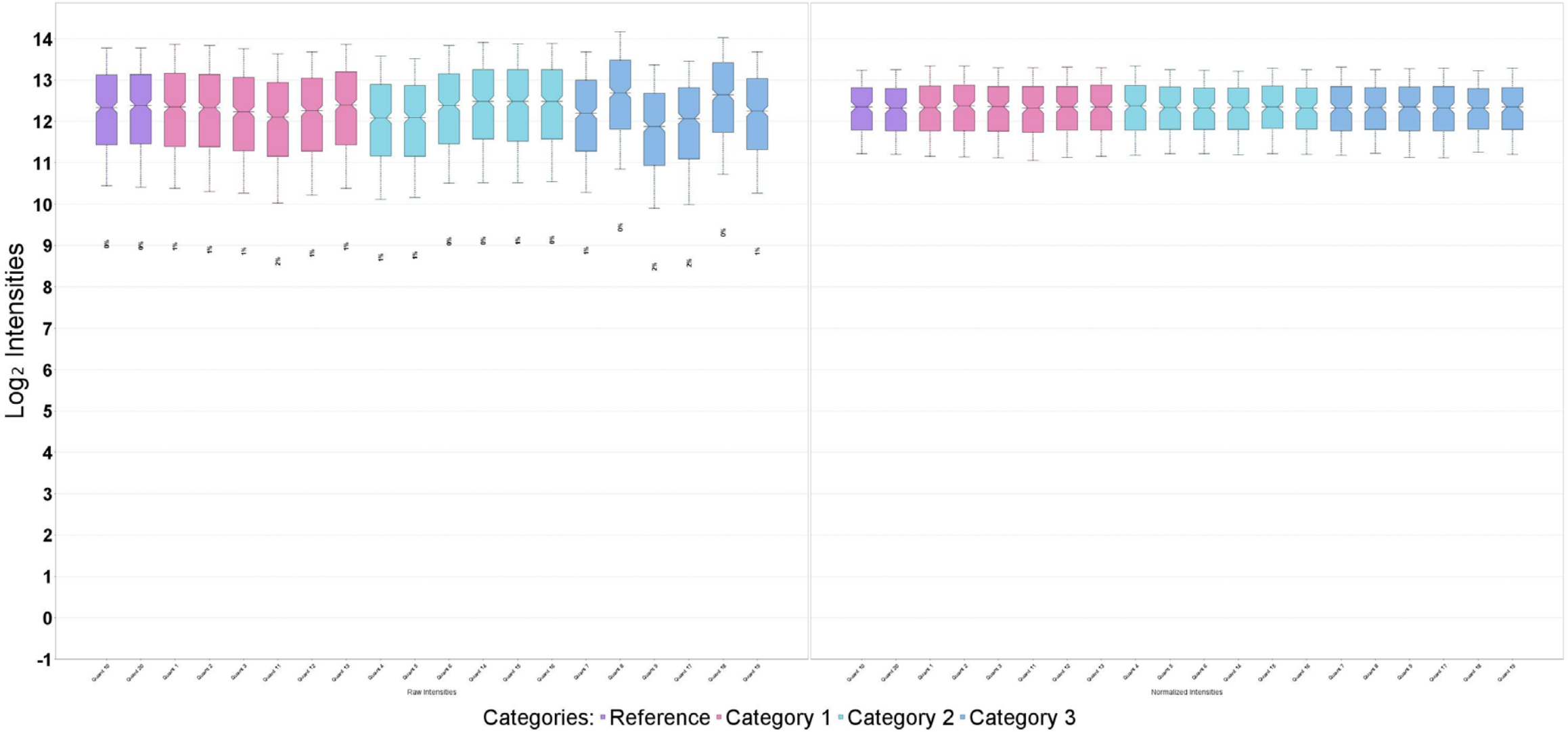
Scaffold Q+S Quant Module



Scaffold Q+S Quant Module



Scaffold Q+S Normalization



Scaffold Q+S Normalization

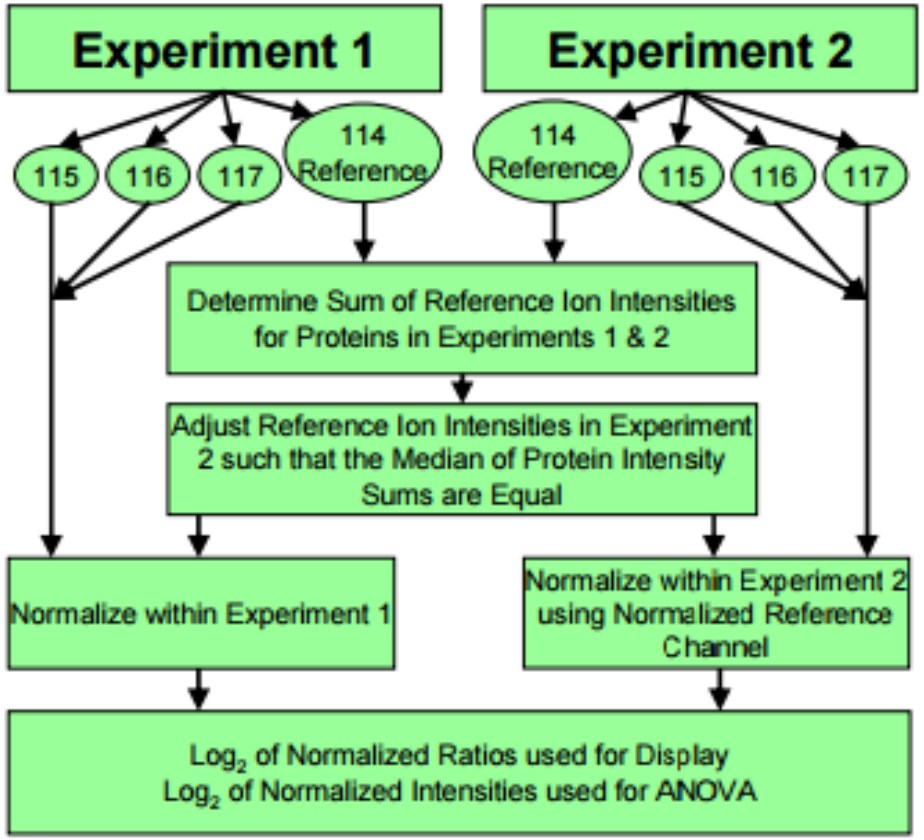


Figure 4: When combining multiple iTRAQ experiments Q+ first normalizes between all experimental reference channels. The normalized references are then used to further normalize ratios within each individual experiment.

Reporter ion channels are typically median normalized with respect to a user-defined reference sample. Employing median ratio normalization ensures that the calculated ratios are no longer dependent on missing or suspect ion intensities because those ratios are not used in the calculation.

An experimental design that daisy chains multiple iTRAQ experiments together using a single reference (but possibly different quantitative samples) can be normalized with an additional step. Before intra-experiment normalization is done, the separate experiments are linked together. The normalization factor is calculated as the median of the summed reporter ion intensities for each protein in the linked reference channels. Normalizing across reference samples in multiple iTRAQ experiments reveals the technical variance between those samples, which can be considered in the final analysis.

Log_2 values of the reporter ion to reference ratios are used to force symmetry between up and down regulated proteins. Statistical test such as the t-Test and ANOVA in Scaffold Q+ are based on Log_2 of the underlying median normalized reporter ion intensities. Testing using the ratios alone ignores variance in the reference channel. Testing using the reporter ion intensities ensures that reference channel variation is considered, while log normalization forces the tests to be magnitude independent.

https://proteome-software.wikispaces.com/file/view/quantitation-across-multiple-iTRAQ-samples

Scaffold Q+S Normalization

Appendix B.Normalization

Quantitative Raw data imported into the Q+ Quantitation Module are processed for Bias removal with two possible different normalizations methods:

- Intensity-Based Normalization
- Ratio-Based Normalization

Intensity-Based Normalization

The method used is a non-parametric version of a statistical model proposed by Ann Oberg and others for analyzing iTRAQ data[1]. A technical description of the different steps included in the normalization algorithm developed in Scaffold Q+ or Scaffold Q+S is provided below.

Pre-Processing

1. The raw intensity data is acquired from the spectra and purity-corrected as appropriate.
2. Spectra not assigned exclusively to one protein are discarded (unless the user has indicated otherwise).
3. The data is transformed by applying a logarithm (base 2).
4. Within each MS-Sample a "missing value" is assigned as the larger between (a) the minimum logged intensity acquired and (b) the value whose z-score is -4 for the distribution of all logged values in the MS-Sample. These missing values are applied to all intensities which either had raw value zero, or fall below the "Minimum Dynamic Range."
5. If the option is selected, spectra with missing values in the reference channel are removed.

Scaffold Q+S Normalization

Iterative Median Polish

A version of Tukey's median polish is applied iteratively to normalized the data. In what follows, all computations are on the logged data, and "average" denotes either median or mean depending on the mode the user has selected. (Median is the default mode.) The normalization has four steps:

1. Inter-Sample Normalization: A normalization factor consisting of the global average minus the within-MS-Sample average is added to each data point.
2. Intra-Sample Normalization: A normalization factor consisting of the within-MS-Sample average minus the within-channel average is added to each data point.
3. Peptide/Spectrum Normalization: For each protein, the averages across the values in each spectrum are brought into alignment by adding a per-spectrum normalization factor (average of these averages minus the particular spectrum's average).
4. Intensity Weighting: A weight is assigned to each spectrum based on a t-statistic derived from percent deviations from channel averages.

Steps (1)-(4) are repeated three times.

If the user has indicated that their reference consists of a pooled or common sample, a protein reference normalization step is applied to bring these reference samples into alignment.

Post-Processing

A standard deviation estimate is derived for each spectrum, based on smoothed within-protein deviations of spectral values from averages binned by total intensity.

The weight of each spectrum is divided by the total number of spectra matched to its peptide, providing a form of intermediate peptide-level averaging when subsequently computing protein-level values.

1. Elizabeth G. Hill, John H. Schwacke, Susana Comte-Walters, Elizabeth H. Slate, Ann L. Oberg, Jeanette E. Eckel-Passow, Terry M. Therneau, and Kevin L. Schey, A STATISTICAL MODEL FOR iTRAQ DATA ANALYSIS, J Proteome Res. 2008 Aug; 7(8): 3091–3101. Published online 2008 Jun 26. doi: 10.1021/pr070520u