Higher and higher N: SimpliFi for the masses now takes masses of samples

1ProtFi, LLC, Farmingdale, NY
2Cold Spring Harbor laboratory, Cold Spring Harbor, NY

Large cohorts? No worries; they’re great!
Ever-increasing amounts of omics data can limit our ability to understand those data and reach actionable decisions. Adding to the challenge, studies continue to grow both in the number of analyzed samples and in the kinds of omics analyses. Previously, we developed SimpliFi, a cloud-based, browser-driven data-to-meaning engine with an intuitive, user-friendly accessible to users of all experience levels. We have now recorded both the backend, speeding calculations by orders of magnitude, as well as the frontend, where summary displays now enable human understanding of information present even in large (>100s) sets of samples. SimpliFi accepts and integrates data from all omics analyses and allows results to easily be shared, explored or published by simply sending a URL.

Statistical correctness
SimpliFi models biology using nonparametric statistics in which sample replicates define their own distributions: such distributions always fit. These distributions are frequently non-normal (insert below) and thus do not satisfy the assumptions of many classic statistics tools, easily resulting in wrong conclusions. SimpliFi automatically accounts for increased data variance at low or high intensities: at low intensity, low stochastic sampling results in expectedly high variation, while at high intensity, effects like saturation become apparent. Importantly, p-values and fold-changes are always reported with confidence intervals.

Effect of intensity of observation on observed variability expressed as fold-changes between replicates of the same condition. In LC-MS-based proteomics, variance is a function of feature intensity wherein low intensity measurements are subject to large amounts of stochastic sampling error and high intensity measurements encounter effects such as saturation. SimpliFi accounts for the individual variance of every feature as a function of intensity.

Effect of number of replicates that agree, or don’t. With an increasing number of biological replicates, if observed changes between states are in the same direction, p-values become more certain; observations of different directions, or inclusion of fewer replicates, have the opposite effect. Certainly increases with increased numbers of agreeing biological replicates and p-values become more certain if observed changes between states are in the same direction. In contrast, observations of different directions or fewer replicate numbers, have the opposite effect.

Interactive plotting for understanding
Data can be explored and visualized with multiple interactive tools including volcano plots, distribution plots, heat maps, etc. SimpliFi’s on-the-fly response produces a machine-human interface that facilitates human intuition guiding data exploration. Users of all skill levels can take deep dives into the data and share projects via a simple URL.

Data-to-meaning via reactome integration
Understanding a dataset requires understanding of regulation within the biological systems. SimpliFi provides tools to map data to pathways and analyze cellular compartments and biological functions. SimpliFi uses the reactome pathway database to quickly understand biological effects.

In Summary: SimpliFi your data
• Straightforward analysis of any size omics datasets for all experience levels
• Non-parametric statistics defined by the data themselves yield statistically correct p-values confidence intervals for all features
• Quality control tools ensure identification of suspect samples
• Interactive visualization tools allow rapid exploration of datasets to quickly derive meaning from data
• Projects can be conveniently shared via URLs with any associate rendering it perfect for joint collaborations.