SimpliFi™ in a Box:

Multiomics Data Analysis on Private Servers

PROTIFI™

Bringing precision omics to life!™

https://protifi.com | info@protifi.com

Jim Palmeri[1], Darryl J.C. Pappin[2]; John Wilson[1]

INTRODUCTION

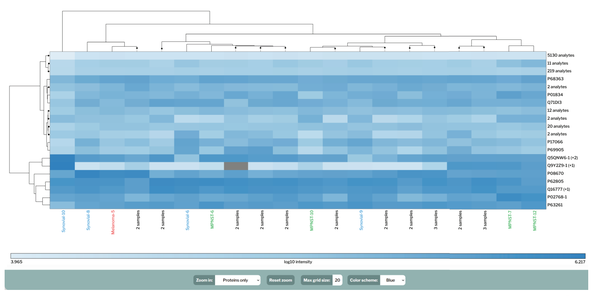
The rapid expansion of omics data poses a significant insights. This challenge is compounded by the increasing sample sizes and diversity of omics analyses within studies.

To address this challenge, speed the path to understanding, and open the power of omics to non-experts, we developed SimpliFi™, an interactive, intuitive data-to-meaning engine. Employing non-parametric statistics derived solely from inherent data structure, SimpliFi™ handles data without making incorrect assumptions. It uses resampling techniques to calculate confidence intervals for all metrics, including p-values, highlighting often-overlooked uncertainties of omics data. SimpliFi™ handles all kinds of omics data and seamlessly integrates data from various omics analyses. Sharing, exploration, or publication is as simple as sending a simple URL, which can be public or private. SimpliFi™ is highly optimized and runs on GPUs to provide results in minutes despite using computationally-intensive non-parametric approaches.

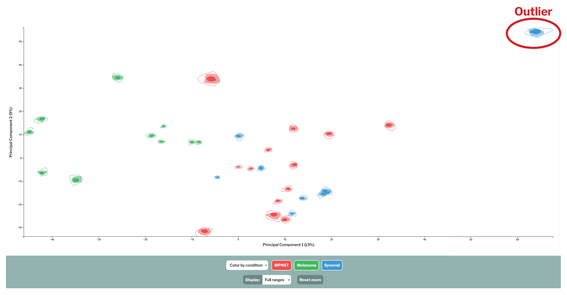
Ease of Use: Many data analysis tools take days or months to understand and use, and then they can break. SimpliF™i is accessible even to non-omics experts: if you can use an iPhone, you can use SimpliFi™!

QUALITY CONTROL & VISUALIZATION

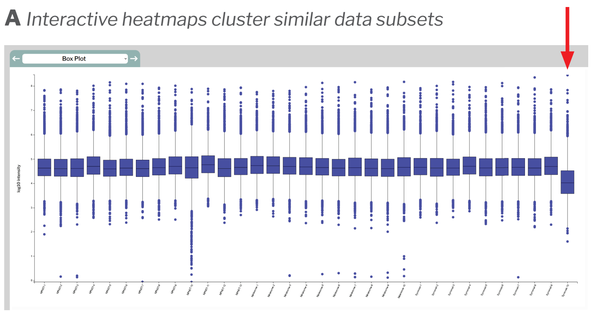
Ensuring quality control is imperative in all analytical workflows. Particularly with extensive sample cohorts, mitigating batch effects and ensuring reproducibility becomes paramount. SimpliFi's™ latest features are designed to efficiently visualize extensive datasets and dynamically alert users to any anomalies ***(Fig. 1)***.



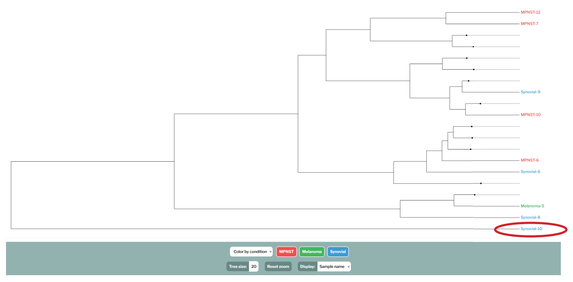
A *Interactive heatmaps cluster similar data subsets*



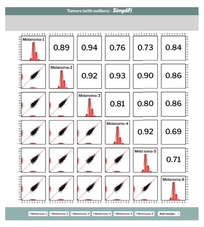
B *PCA identifies outliers, missing data variance is accounted for*



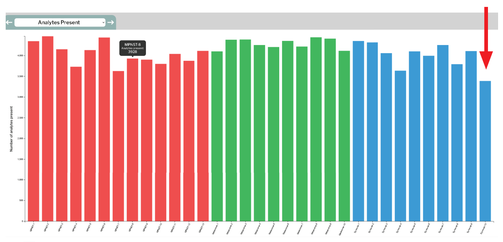
C *Spot outliers of signal intensity*



D *Interactive hierarchial clusterings*



E *Multisample plots*



F *Analytes present*

G *Interactive correlation plots and tables*

Fig. 1

RIGOROUS ANALYSIS

SimpliFi™ analyses often non-Gaussian biological measurements using nonparametric statistics, allowing sample replicates to define their own unique distributions, which can lead to errors in traditional statistical methods. Key to accuracy is its ability to correctly adapt to variances that change with measurement intensity: at low intensity, stochastic sampling expectedly gives high variation, while effects like saturation can occur at high intensity. Importantly, SimpliFi™ always reports p-values and fold- changes with accompanying confidence intervals.

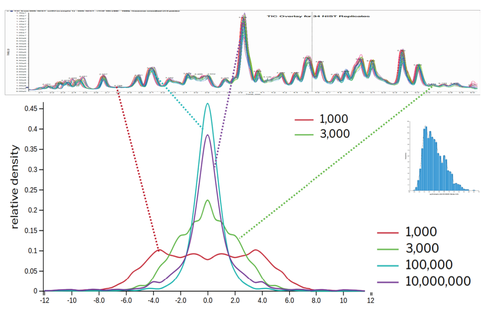


Fig. 2 *Red Trace: The likelihood of observing a 16x difference between any two replicates within the same class at low-level observations around 1000 counts is about the same as observing the expected 1x difference.*

Effect of intensity of observation: In comparing samples of the same kind *(like control to control)*, one expects and sees that the ratios between any two samples is likely to be around 1, and that variability will exist ***(Fig. 2)***. The breadth of this curve, i.e. how often ratios between the same kind of sample will significantly deviate from 1, are strongly a function of intensity of observation.

Impact of Replicate Consistency: As the number of biological replicates increases, consistent changes between states boost p-value certainty. However, inconsistencies or fewer replicates reduce this certainty. More agreeing replicates strengthen confidence, especially when changes align, while discrepancies or fewer replicates weaken it.

Note that SimpliFi™ p-values represent the chance that control was chosen at random from experimental (or the inverse). This is not the same as the probability that two normally distributed populations have the same mean.

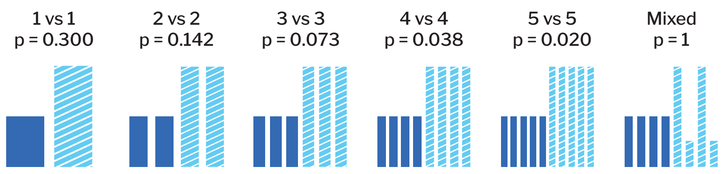
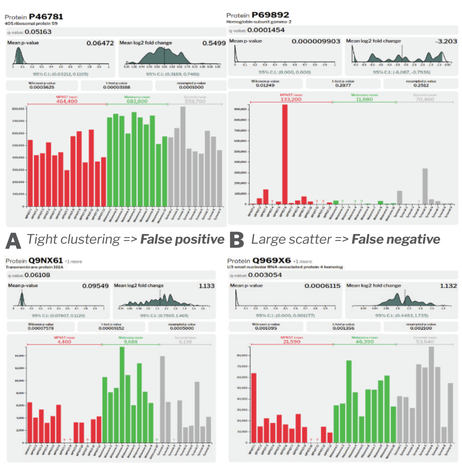


Fig. 3

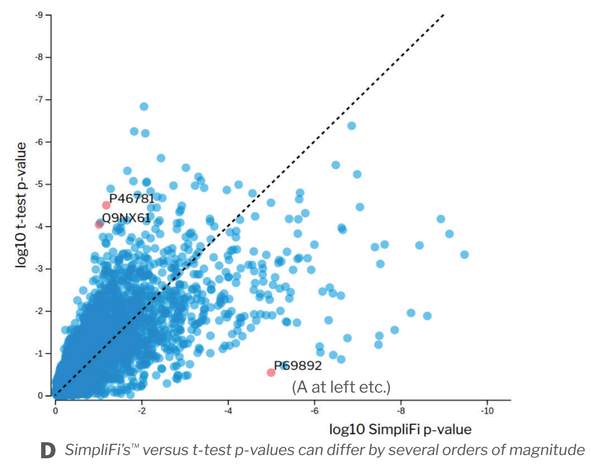
RIGOROUS ANALYSIS

Limitation of T-Tests: False positives and negatives result from under sampling of variability ***(Fig. 4A)*** or outliers ***(Fig. 4B)***. The effect of intensity on certainty of measurement can be substantial ***(Fig. 4C)***: Q9NX61 and Q969X6 have nearly identical fold changes, however the former observed at low intensity *(4k - 10k)* is not significant *(p = 0.095)*, while the latter at the 20k – 50k intensity range is highly significant *(p = 0.0006)*.



A *Tight clustering =>* ***False positive*** **B** *Large scatter =>* ***False negative***

C *Same ratio but lower signal:* ***p = 0.0006 vs 0.096!***



D *SimpliFi's™ versus t-test p-values can differ by several orders of magnitude*

Fig. 4 *Note that SimpliFi™ p-values represent the chance that control was chosen at random from experimental (or the inverse). This is not the same as the probability that two normally distributed populations have the same mean.*

EFFORTLESS EXPLORATION: BUCKETS

Create custom lists of analytes you wish to explore further with the Buckets feature. Choose them from the analyte lists or select all analytes within a box or freeform region of any diagram. Then enable the filter to perform analysis on just that set, or what that set has in common (or not) with other buckets. Below is the intersection of orange and purple buckets ***(Fig. 5)***.

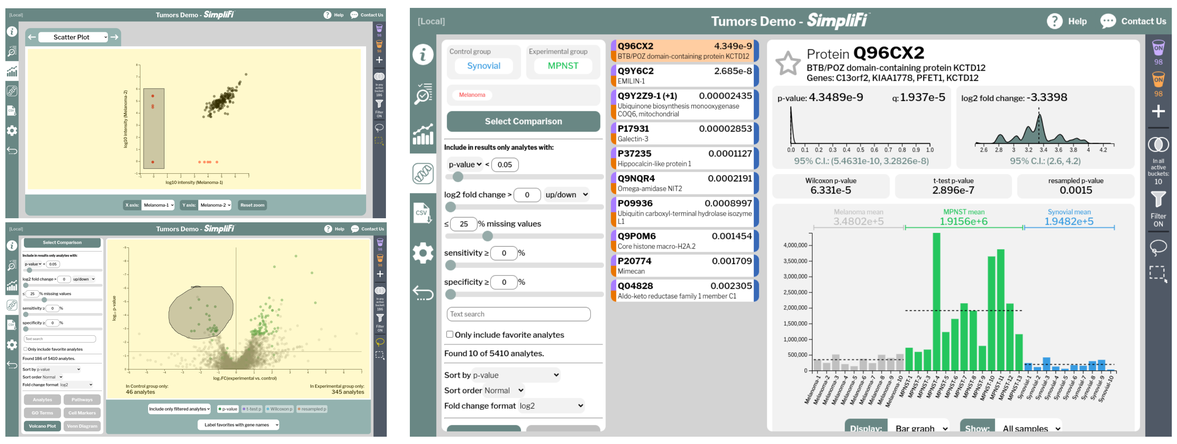
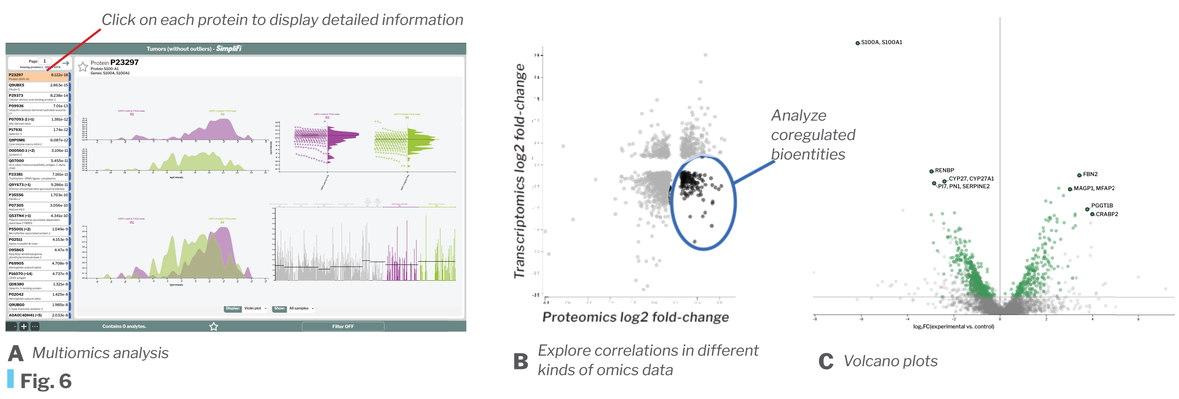


Fig. 5

INTERACTIVE PLOTTING

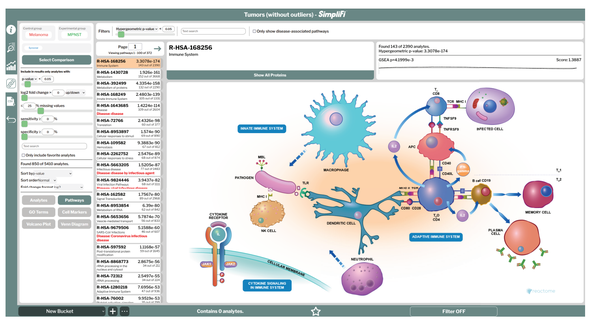
Data can be explored and visualized with multiple interactive tools including volcano plots, distribution plots, heat maps, etc. SimpliFi's™ on-the-fly response yields a machine-human interface where human intuition guides data exploration guided. Users of all skill levels can take deep dives into the data and share projects via a simple URL.



A *Multiomics analysis* **B** *Explore correlations in different kinds of omics data* **C** *Volcano plots*

Fig. 6

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, as well as using the Reactome pathway database to quickly understand biological effects.

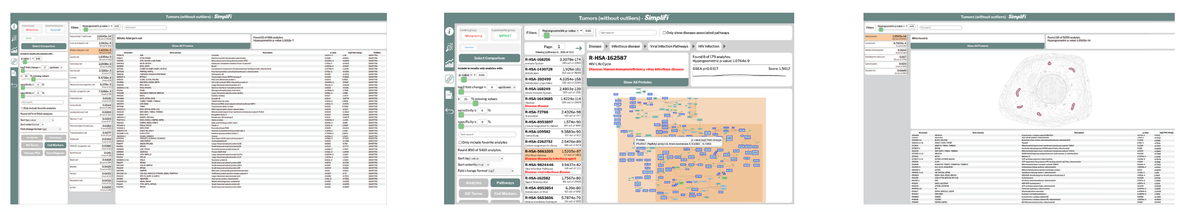


Fig. 7 *Pathway and cellular compartment maps in conjunction with GO annotation allow for deep exploration of proteomic, metabolomic, lipidomic, transcriptomic, and genomic datasets.*

[1]ProtiFi LLC, Fairport, NY, United States; [2]Cold Spring Harbor Laboratory, Long Island, United States