

The Proteograph Analysis Suite is an intuitive, scalable, data informatics solution

PAS enables automated results generation and intuitive, easy to interpret proteomics visualizations

Differential expression analysis tools simplify data interpretation allowing easy determination of biological insights

Researchers are increasingly adopting multi-omics approaches to understand the complex biological processes that underlie human diseases. Next-generation sequencing (NGS) is widely used for identifying genetic variants and gene function while mass spectrometry is used to quantify protein abundances, modifications, and interactions from complex samples like plasma. A new plasma profiling platform called the Proteograph™ Product Suite was developed that leverages multiple nanoparticles with distinct physiochemical properties to provide deep plasma proteomic analysis at scale¹. The analysis of proteomics and genomics data typically requires a wide collection of different tools, which is further complicated by the prevalence of command-line interfaces and operating system-specific requirements that can act as a barrier for researchers to adapt new data analysis tools due to their steep learning curve and implementation costs.

In this abstract, we present a cloud-based analysis software platform called Proteograph Analysis Suite (PAS) that analyzes proteomics data derived from the Proteograph workflow along with genomic variant results imported from NGS experiments. The main features of the software suite include an experiment data management system, analysis protocols, an analysis setup wizard, and tools for reviewing and visualizing results. PAS can support both Data Independent Analysis (DIA)^{2,3} and Data Dependent Analysis (DDA) proteomics workflows⁴, and is compatible with widely accepted variant call file formats from NGS workflows. For each analysis run, users can view various quality control metrics like peptide/protein group intensity, protein sequence coverage, relative protein abundance distribution, peptide and protein groups stratified by nanoparticle. Various visualizations such as principal component analysis, hierarchical clustering, and heatmaps allow intuitive identification of dataset trends. Quantitative differential expression tools such as volcano plots, protein interaction maps and protein-set enrichment simplify data interpretation and enable functional insights.

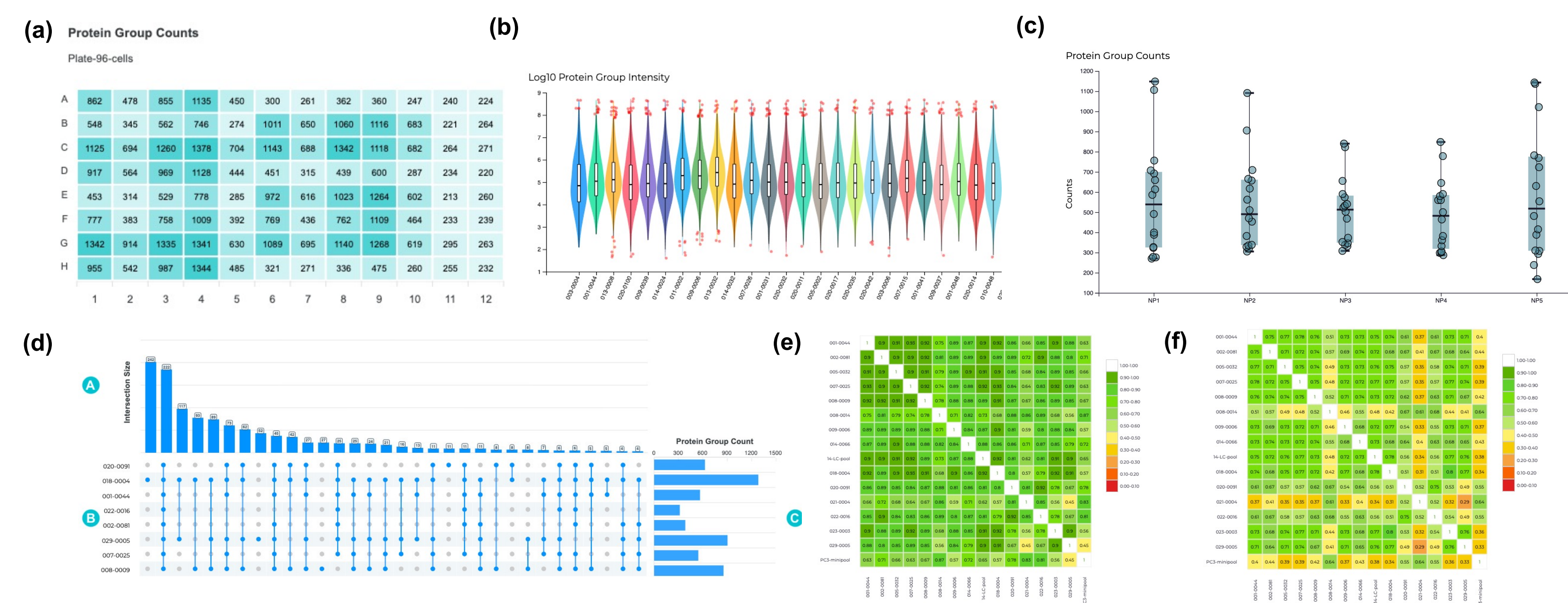


Figure 2. Analysis Summary and Metrics (a) View results for protein groups (shown) and peptide counts, quant mass, miscleavage rate, oxidation ratio and ID rate in a simple and intuitive plate format. (b) Distributions of protein group intensities and CVs across samples. (c) Box plots showing the number of protein groups identified across NPs. Hovering over a dot reveals the peptide or protein count, file, and sample name. Hovering over a box shows the quantile for the NP. (d) Graphs and a matrix show protein group overlaps; Intersection Size bar graph (B) Protein Group Count bar graph (C) Matrix (e, f) A color-coded matrix displays sample comparability data using PCC (left) or the Jaccard index (right). Samples on the green end of the spectrum have high correlation, while samples on the red end of the spectrum have low correlation.

QC Charts automate statistical process control assessment allowing rapid run-to-run performance evaluation

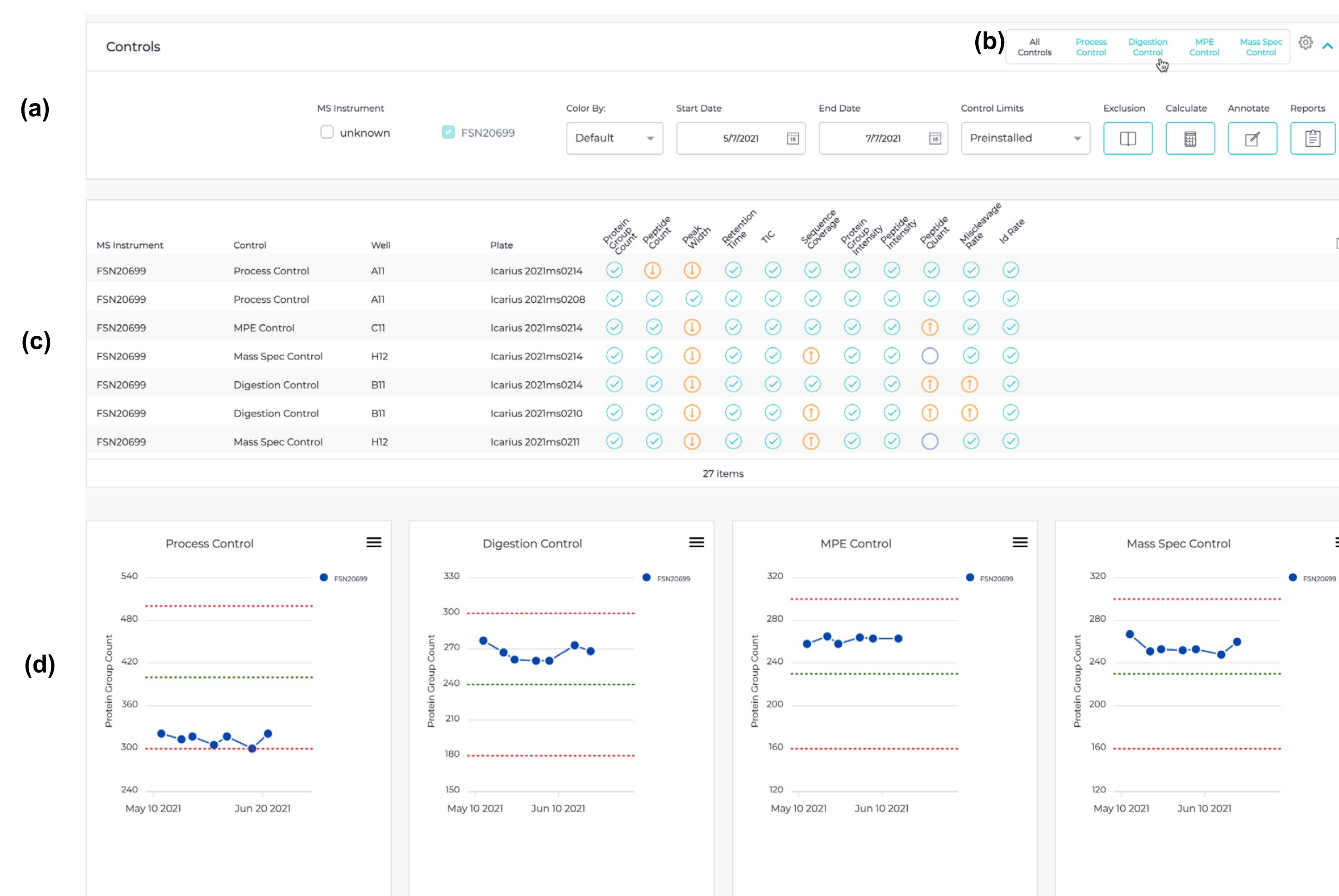


Figure 3. Control Results: (a) Filters for viewing charts for all controls or a selected control type. (b) Toolbar with additional filters and functions. (c) Summary of control data for the selected analysis time frame. (d) QC charts with metrics for each control.

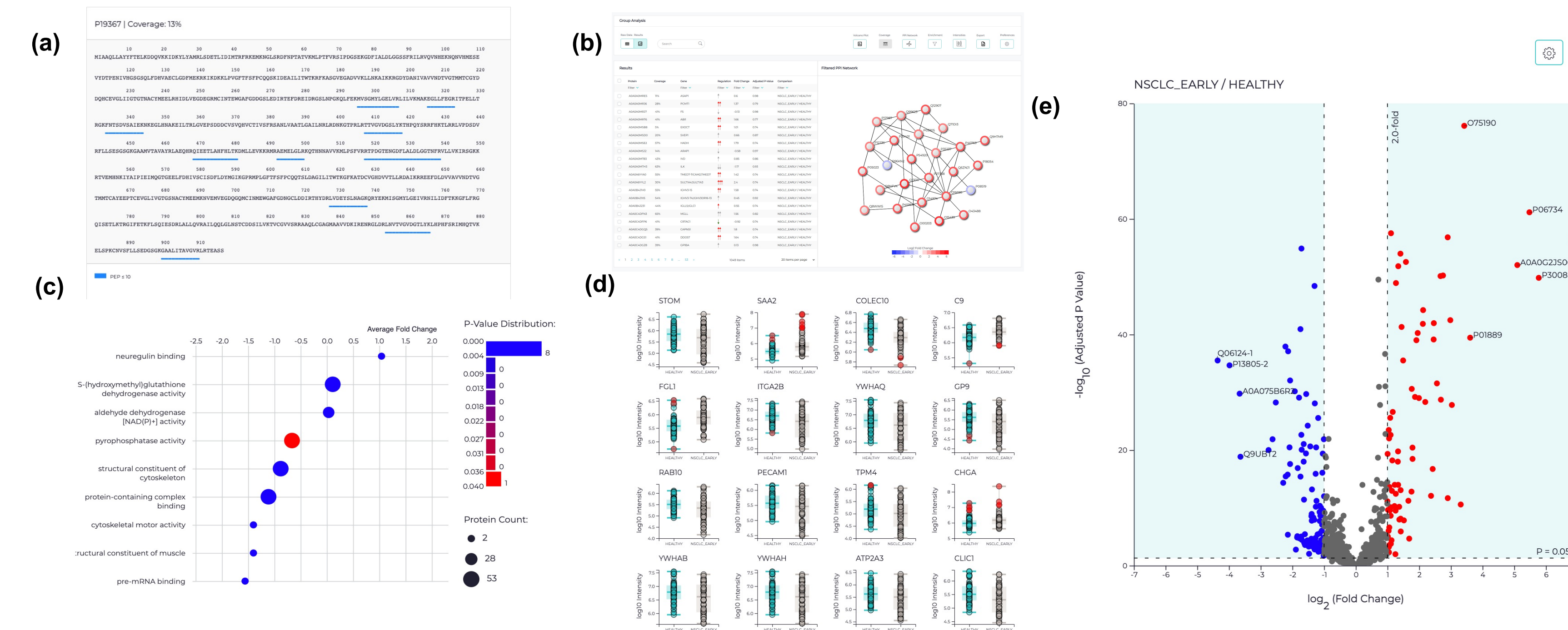
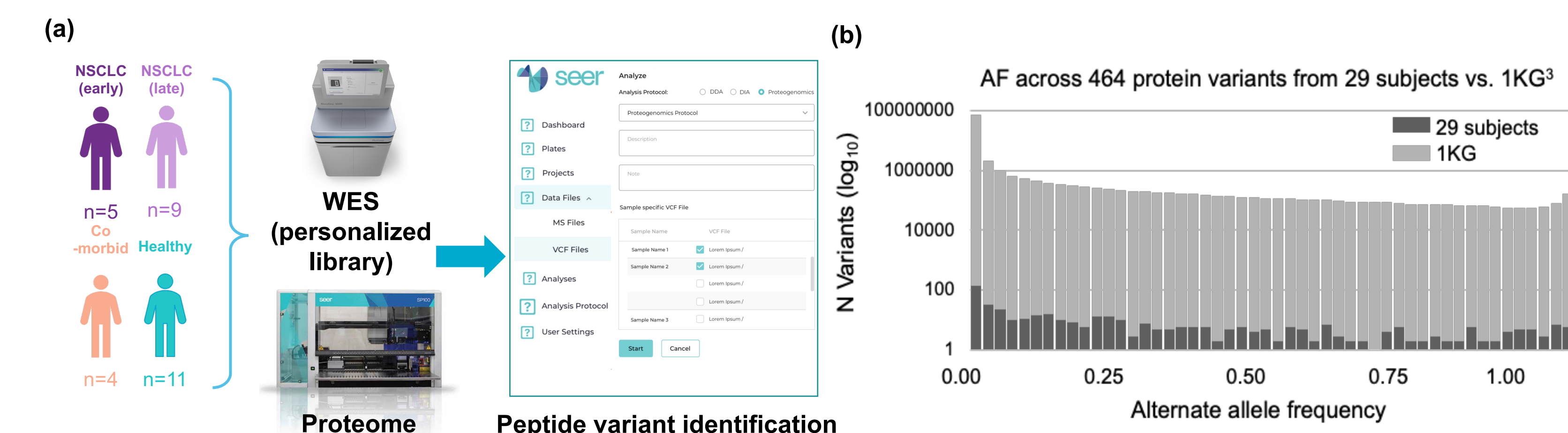


Figure 4. Group Analysis Results: (a) Sequence Coverage: Visualize where peptides map relative to the protein sequence. (b) Protein-Protein-Interactions Comparison: Build a STRING-based PPI network to identify differences in protein interactors. (c) GO Enrichment: Explore how proteins associated with a group differ functionally. (d) Intensity Comparison: View how the intensity of a protein of interest differs between groups. (e) Sample group analysis visualized with volcano plot.

Proteogenomics functionality allows integration of multi-omics datasets such as linking genomic variant results with the proteome



Peptide variant identification using personalized libraries

Figure 5. (a) PAS can analyze VCF files generated from NGS pipelines in combination with mass spec data to identify peptide variants using personalized libraries. (b) An example of the allele frequency of the variants found in the 29 individuals against the background of allele frequencies in the 1000 genomes project shows the distributions are similar, demonstrating the unbiased nature of the Proteograph solution.

Conclusions

We present a comprehensive proteogenomic analysis software suite to enable user-friendly and reproducible multi-omics analyses of proteomic and genomic data.

References:

- Blume et al. Nat. Comm. (2020)
- Searle, B.C et al. Nat Commun (2018)
- Demichev, V et al. Nat Methods (2020)
- Tyanova S et al. Nat Protocols (2016)

