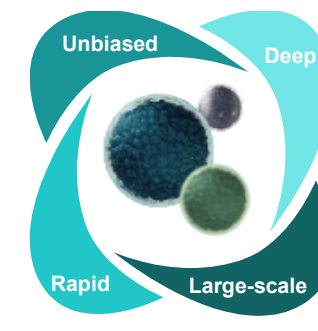


A Cloud-Scalable Software Suite for Large-Scale Proteogenomics Data Analysis and Visualization

Mara Riley*; Taylor Page; Margaret K.R. Donovan; Harsharn Auluck; Yuandan Lou; Theo Platt; and Serafim Batzoglou



The Proteograph™ Analysis Suite (PAS) is an intuitive and scalable data informatics solution

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

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

Introduction

Integrating multi-omic data can reveal the molecular consequences of genetic variation in human disease. Combining proteomics and genomics data can be challenging for researchers due to complex workflows and various tools involved. The Proteograph Analysis Suite (PAS) software application, included in Proteograph™ Product Suite¹, is a dedicated, cloud-based software solution that simplifies the processing, analyzing, and visualization of proteomics datasets from liquid chromatography-mass spectrometry (LC-MS).



A seamless journey from data to biological insight

PAS includes an experiment data management system, analysis protocols, analysis setup wizard, and result visualizations. PAS supports both Data Independent Analysis (DIA) and Data Dependent Analysis (DDA) workflows and is compatible with variant call format (.vcf) files for personalized database searches. Assess data quality with metrics for identified peptides and protein groups. Monitor experimental trends with visualizations such as principal component analysis, hierarchical clustering, and heatmaps. Uncover biological insights with differential abundance analyses using volcano plots, protein interaction maps, and protein-set enrichment.

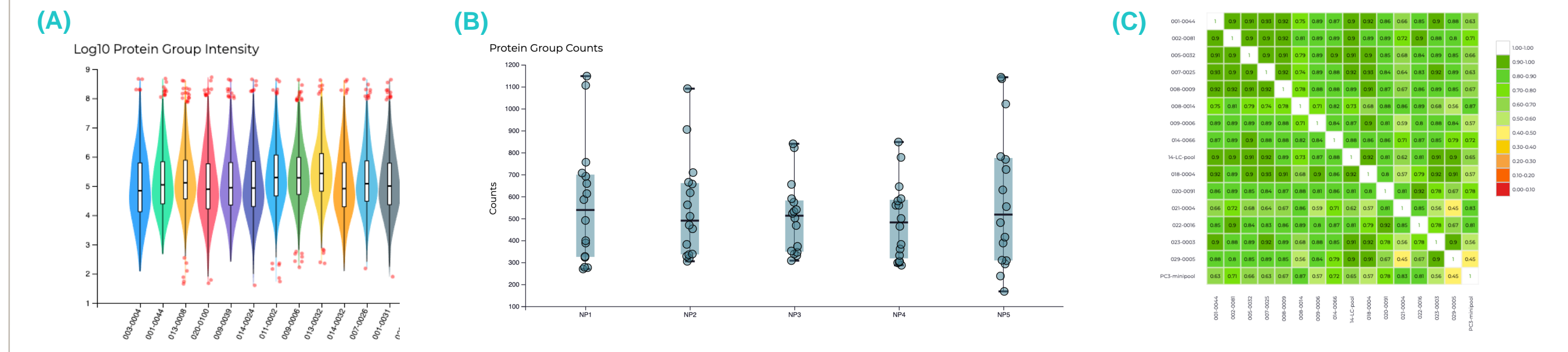


Figure 2. Analysis Summary and Metrics. (A) Distributions of protein group intensities and CVs across samples. (B) Box plots showing the number of protein groups identified across NPs. Hovering over a dot reveals the peptide or protein count, file, and sample name. Hovering over a box shows the quantile for the NP. (C) A color-coded matrix displays sample comparability data using PCC. Samples on the green end of the spectrum have high correlation, while samples on the red end of the spectrum have low correlation.

QC tools automate process control assessment and enable rapid run-to-run performance evaluation

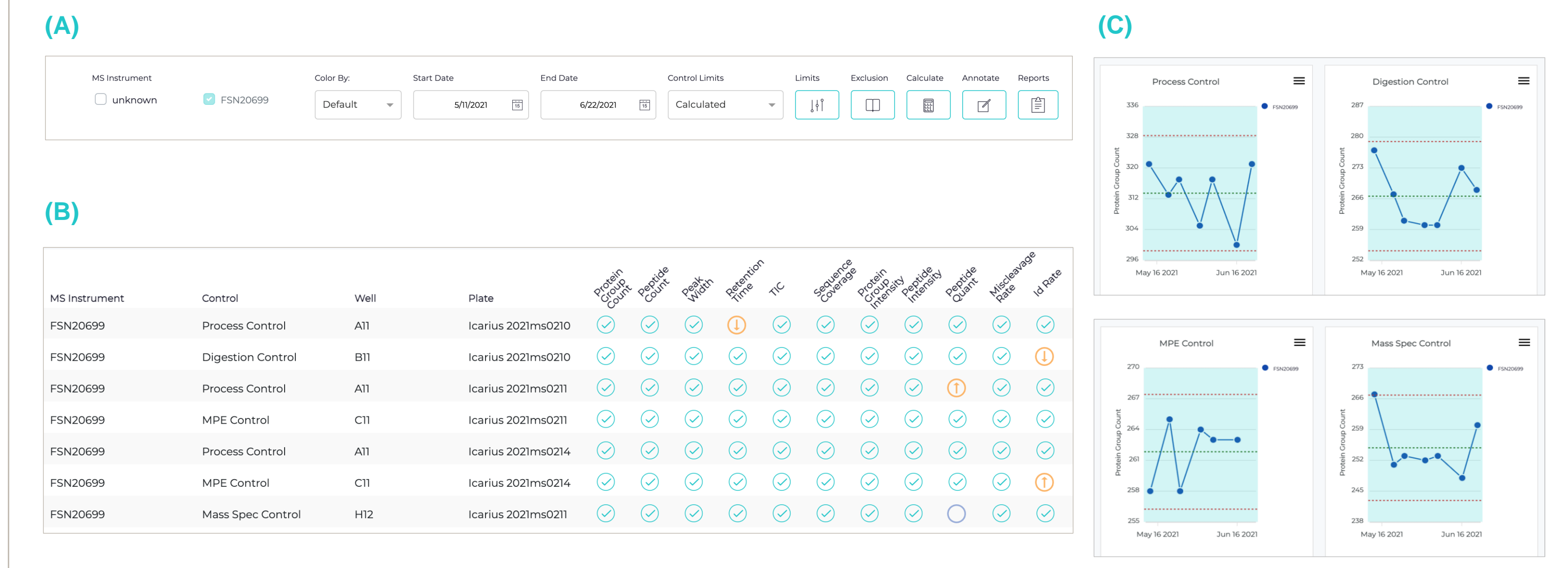


Figure 3. Control Results. (A) Toolbar to select or filter control data and to define expected limits. (B) Summary of control data for the selected analysis time frame. (C) QC charts with metrics for each control.

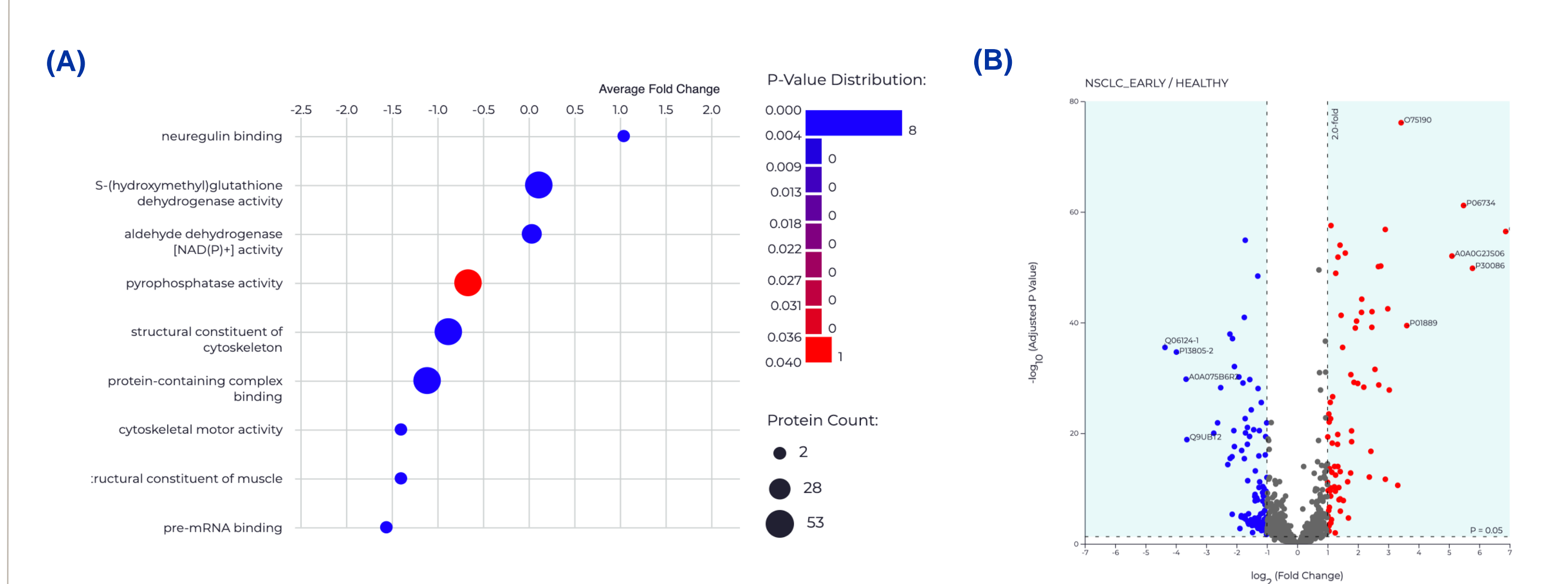


Figure 4. Group Analysis Results. (A) GO Enrichment: Explore how proteins associated with a group differ functionally. (B) Sample group analysis visualized with volcano plot. Additional visualizations available in PAS include

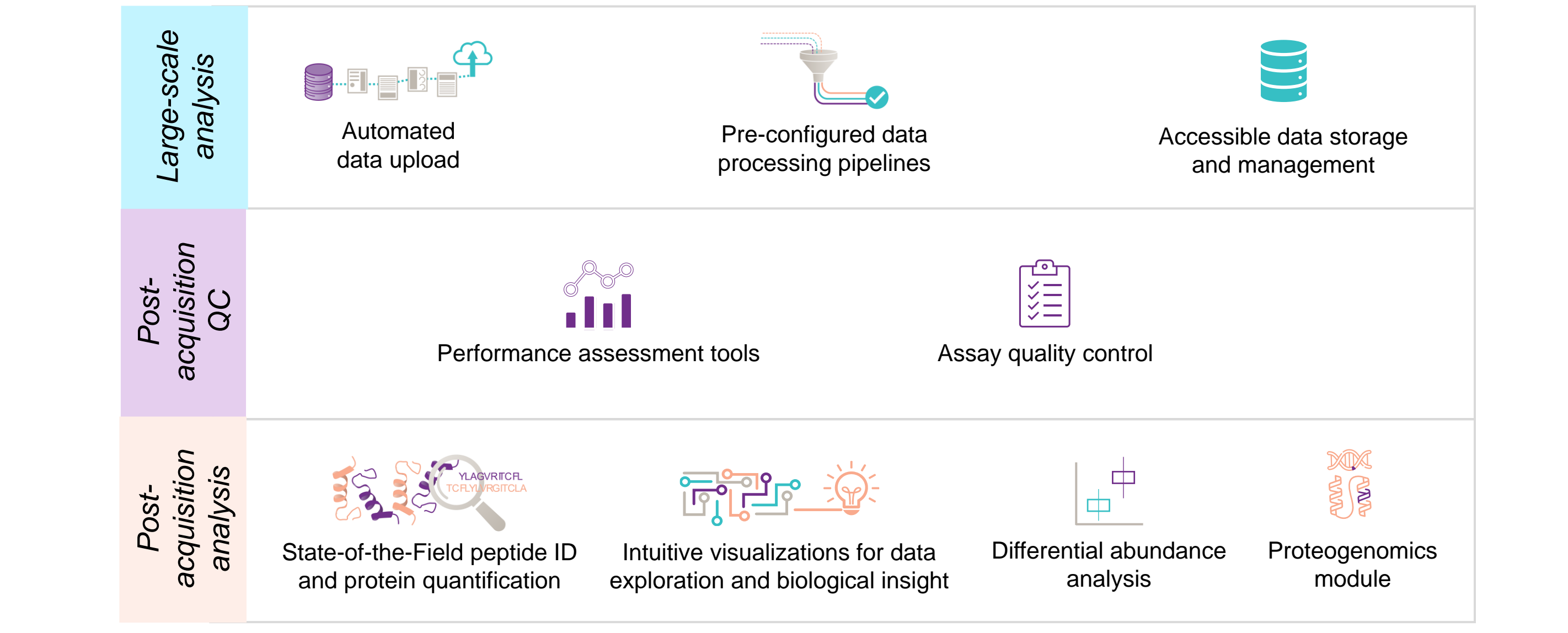
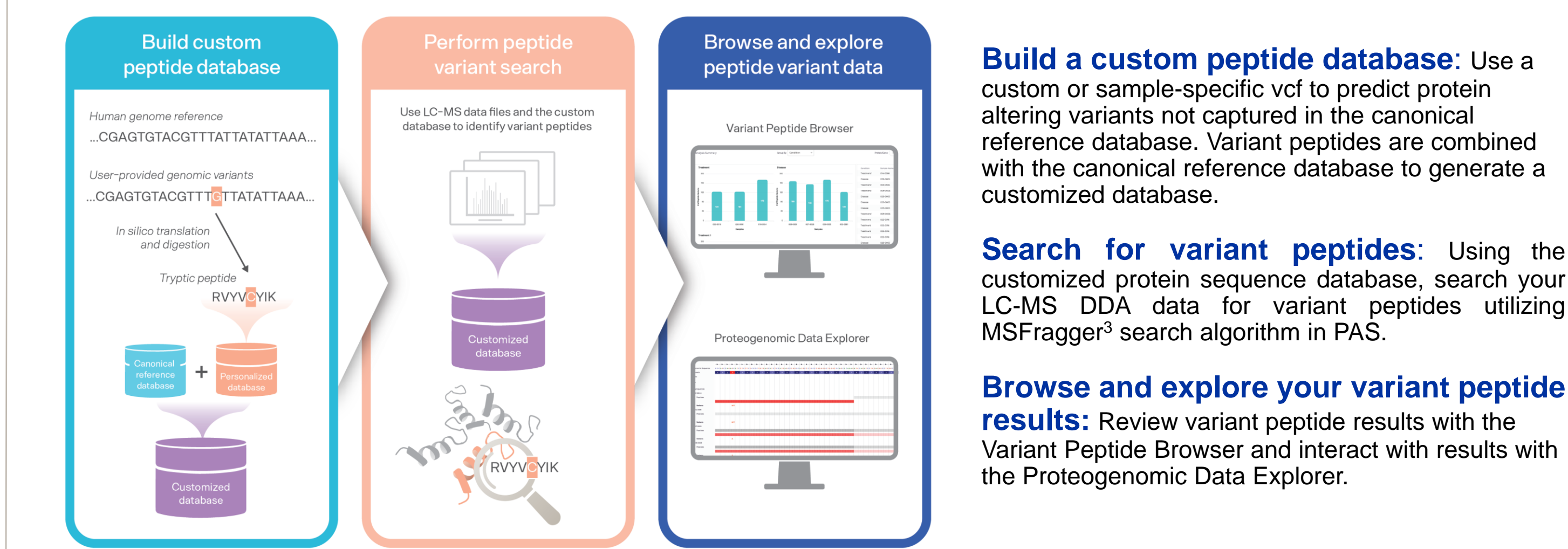


Figure 1. Data is seamlessly transferred from MS computer to PAS without manual intervention using the AutoUploader tool in PAS. PAS features multiple tools addressing large-scale proteogenomics analysis, post-acquisition QC and data visualization.

Proteogenomics workflow links genomic variants with the proteome for variant peptide identification²



Build a custom peptide database: Use a custom or sample-specific vcf to predict protein altering variants not captured in the canonical reference database. Variant peptides are combined with the canonical reference database to generate a customized database.

Search for variant peptides: Using the customized protein sequence database, search your LC-MS DDA data for variant peptides utilizing MSFragger³ search algorithm in PAS.

Browse and explore your variant peptide results: Review variant peptide results with the Variant Peptide Browser and interact with results with the Proteogenomic Data Explorer.



Conclusion

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

References

1. Blume et al. Nat. Comm. (2020)
2. Searle, B.C et al. Nat Commun (2018)
3. Demichev, V et al. Nat Methods (2020)