Identify and Explore Variant Peptides with the Proteograph™ Analysis Suite Proteogenomics Workflow

Genomics Meets Proteomics

A key step in bottom-up proteomics data analysis is peptide identification. In proteomics data analysis of liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS), peptides are identified by searching MS/MS spectra against a reference protein database *in silico* fragmentation pattern. While existing canonical reference databases, like Ensembl, RefSeq, and UniProt capture most protein-coding sequences in the genome, it is possible that novel gene models and their associated protein products, alternative splice forms, or protein altering genetic variants may not be annotated in these reference databases.

To consider novel variant peptides, a proteogenomics workflow can be used to incorporate genomic sequence information into the proteomic data analysis workflow by adding protein sequences predicted from genetic variants or

Proteomics: The study of proteins and how they function biologically through the construction of complete and quantitative protein maps.

Proteogenomics: Incorporation of genomic information with proteomic data analysis to identify variant peptides not captured in canonical reference protein databases.

alternative gene models to the canonical reference database, to create a customized database. Proteograph™ Analysis Suite version 2.0 (PAS), a cloud-based, proteomics data analysis software, now includes a Proteogenomics workflow to identify and explore peptide variants arising from allelic variation or other user-defined protein sequence altering variants, through the integration of Proteograph proteomics data with NGS variant information (Figure 1).
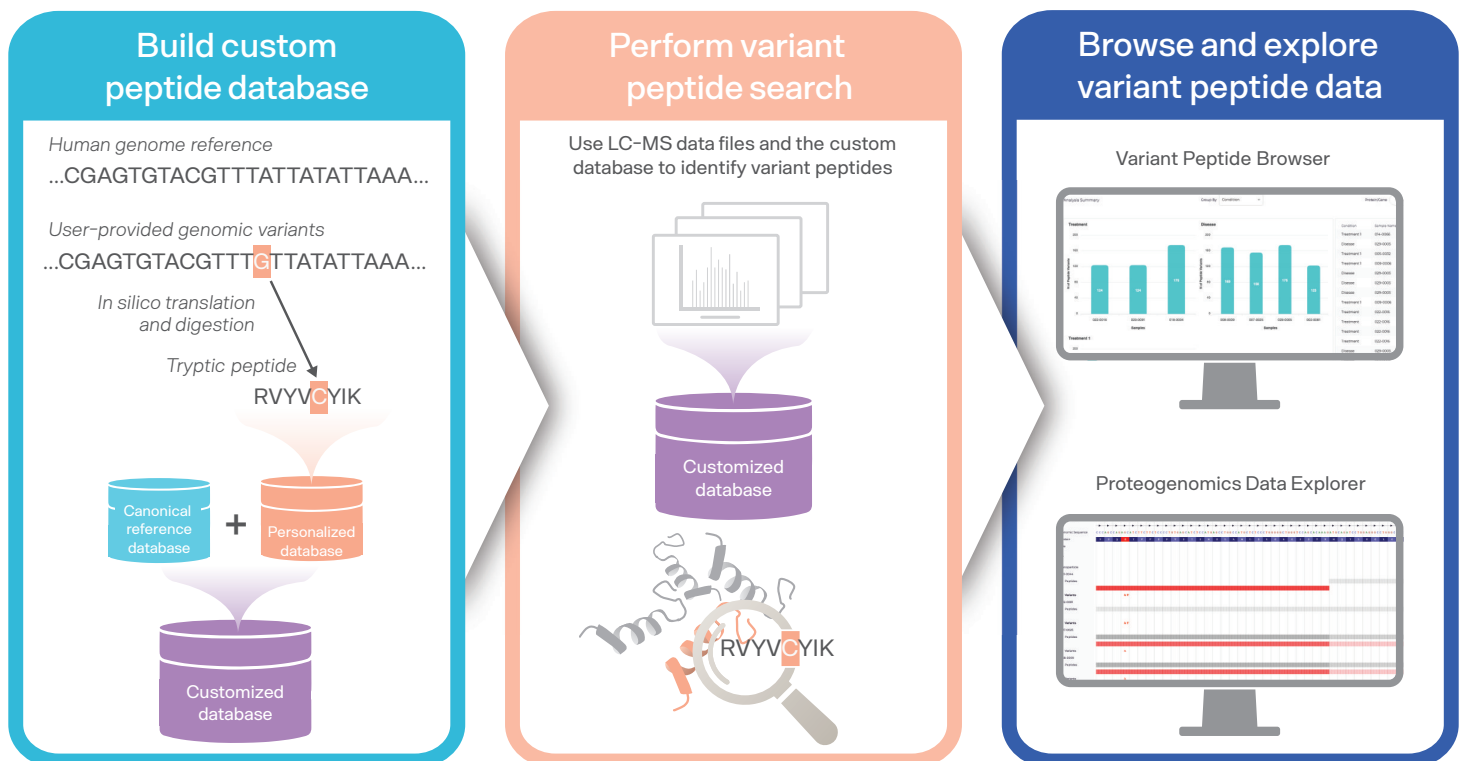

Figure 1. The PAS Proteogenomics workflow enables users to build a customized protein sequence database, perform a variant peptide search, browse variant peptide results in the Variant Peptide Browser, and explore variant peptides in genomic space with the Proteogenomics Data Explorer.
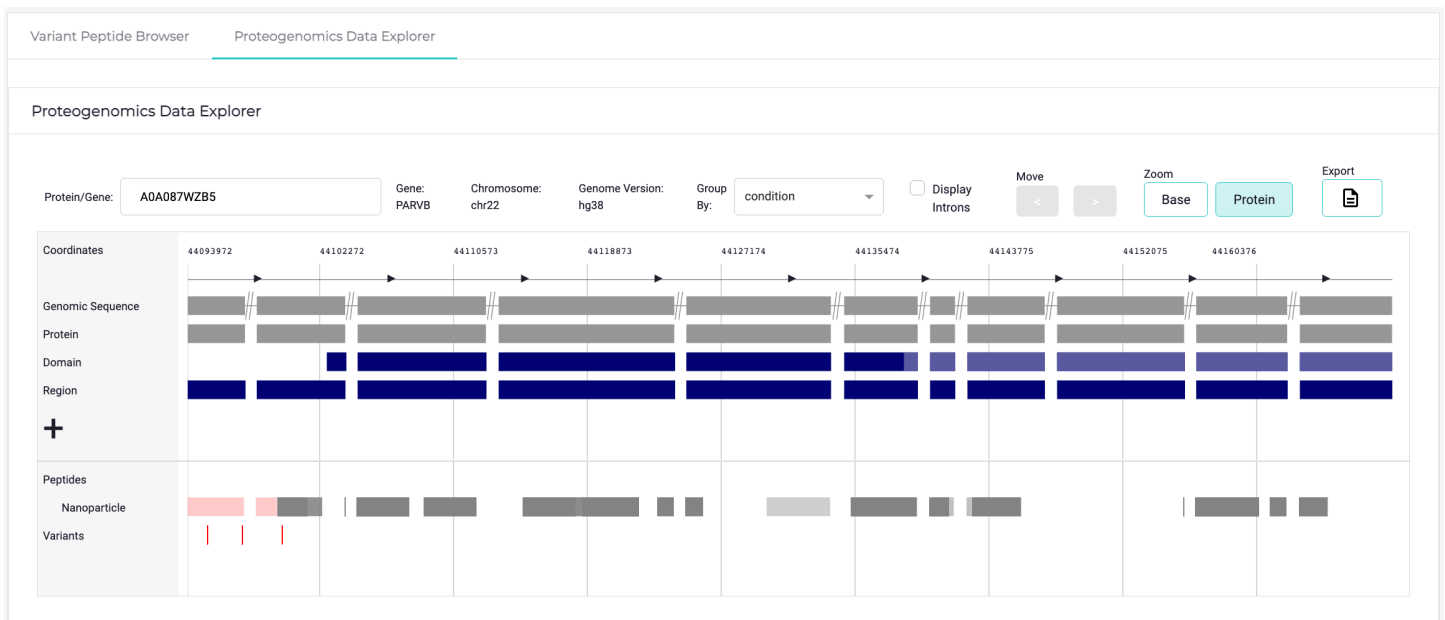
Figure 2. Get a birds-eye view of both genomic variant coordinates and relationship to peptide and protein sequencing data using the Proteogenomics data explorer.

Discover Proteoforms Using PAS Proteogenomics Integration

The PAS Proteogenomics workflow offers a scalable and easy-to-use solution to a typically computationally intensive bioinformatics pipeline, which includes:

- **Building a custom peptide database**: Upload a custom or sample-specific variant call file (VCF) to predict protein altering variants not captured in the canonical reference database. Personalized variant peptides are automatically combined with the canonical reference database to generate a customized database.

- **Performing a search for variant peptides**: Using the customized protein sequence database, search your LC-MS/MS Data Dependent Acquisition (DDA) data for variant peptides utilizing MSFragger search algorithms in PAS.

- **Browsing and exploring your variant peptide results**: Variant peptide results are summarized in an interactive table and plots. Browse peptide and variant peptide data maps in genomic space at nucleic acid/amino acid resolution (Figure 2). Visualize gene structure, protein domain information, and region information with respect to identified peptides.

Level Up Your Biological Discoveries

Discover sample-specific variant peptide not captured in the canonical reference database and gain a more comprehensive view of the proteome:

- Identify disease-specific peptides arising from genetic aberrations.

- Refine gene models.

- Better understand the molecular basis of human disease.

- Improve understanding of actionable proteins that can be used in targeted disease therapies.

Find out more at
seer.bio/products/proteograph-analysis-suite