



# Proteogenomics and its Methodology

Abraham Mesuere\*

Department of Biomolecular Medicine, University of Nicosia, Nicosia, Cyprus

## DESCRIPTION

Proteogenomics is a field of biological research that uses a combination of proteomics, genomics, and transcriptomics to assist in the discovery and identification of peptides. The proteogenomics is used to identify new peptides by comparing MS/MS spectra with protein databases obtained from genomic and transcriptomics information. Proteomics often refers to studies that use proteomics information often obtained from mass spectrometry to improve gene annotation. Using both proteomics and genomics data, advances in the availability and performance of spectroscopic and chromatographic technologies led to the emergence of proteomics as a separate discipline in 2004. Proteomics treats proteins in the same way that genomics studies the genetic code of the whole organism, while transcriptomics deals with RNA sequencing and transcript studies. While all three disciplines use mass spectrometry and chromatographic forms to identify and study the function of DNA, RNA, and proteins, proteomics is a correct genetic model for all related protein sequences. It is assumed to be in a reference database such as: Proteomics identification database. Proteogenomics helps eliminate reliance on existing limited genetic models by combining datasets from multiple fields to create a database of proteins or genetic markers. In addition, the emergence of new protein sequences due to mutations, often unexplained by traditional proteomics databases, can be predicted and investigated using the synthesis of genomic and transcriptomics data. The resulting studies can be applied to improve genetic annotation, study mutations, and understand the effects of genetic manipulation. Recently, co-profiling of surface proteins and mRNA transcripts from single cells by methods such as CITESeq and ESCAPE is called single-cell proteogenomics, but the purpose of these studies is independent of peptide identification these methods are more commonly referred to as multimodal omics or multi-omics.

## METHODOLOGY

The main idea behind the proteogenomic approach is to identify peptides by comparing MS/MS data with a protein database

containing predicted protein sequences. Protein databases are generated in a variety of ways using genomic and transcriptome data. Below are some of the ways in which protein databases are generated:

### Six-frame translation

Six-frame translations can be utilized to create a database that predicts protein sequences. The limitation of this method is that databases will be very large due to the number of sequences that are generated, some of which do not exist in nature.

### Ab initio gene prediction

In this method, protein bases are generated by gene prediction algorithms that allow identification of protein coding regions. The database is similar to the database generated by the 6-frame conversion in that the database can be very large.

### Expressed sequence tag data

For 6-frame translations, you can use an Expression Sequence Tag (EST) to generate a protein database. EST data provides transcription information to help you create a database. The database can be very large and has the disadvantage of having multiple copies of a particular sequence. However, this problem can be avoided by compressing the protein sequences generated by the computational strategy.

### Other methods

Protein databases can also be created using RNA sequence data, annotated RNA transcripts, and variant protein sequences. In addition, there are other, more specialized protein databases that can be built to properly identify the peptide of interest. Another way to identify proteins by proteogenomics is comparative proteogenomics. Comparative proteogenomics compares proteomics data from multiple related species at the same time and leverages the homology between those proteins to improve annotation with higher statistical certainty.

**Correspondence to:** Abraham Mesuere, Department of Biomolecular Medicine, University of Nicosia, Nicosia, Cyprus, E-mail: ssezginer@gmail.com

**Received:** 10-Jan-2022, Manuscript No. JDMGP-22-15134; **Editor assigned:** 12-Jan-2022, PreQC No. JDMGP-22-15134 (PQ); **Reviewed:** 24-Jan-2022, QC No. JDMGP-22-15134; **Revised:** 27-Jan-2022, Manuscript No. JDMGP-22-15134 (R); **Published:** 31-Jan-2022, DOI: 10.4172/2153-0602.22.13.e134.

**Citation:** Mesuere A (2022) Proteogenomics and its Methodology. J Data Mining Genomics Proteomics. 13:e134.

**Copyright:** © 2022 Mesuere A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.