OPEN ACCESS Freely available online

Journal of Proteomics & Bioinformatics

Editorial

A Brief Note on Proteogenomics

Nafiseh Yavari*

Department of Bioresource Engineering, McGill University, Sainte-Anne-de-Bellevue, QC, Canada

EDITORIAL NOTE

Proteogenomics is a branch of biology that uses a mix of proteomics, genomics, and transcriptomics to help in peptide discovery and identification. Proteogenomics is a technique for discovering novel peptides by comparing MS/MS spectra to a protein database built from genomic and transcriptome data. Proteogenomics is a term that refers to research that uses proteomic data, which is frequently acquired from mass spectrometry, to enhance gene annotations. Genomics is the study of complete organisms' genetic code, whereas transcriptomics is the study of RNA sequencing and transcripts. To discover and analyse the activities of proteins, proteomics uses tandem mass spectrometry and liquid chromatography. Proteomics is a technique for identifying all of the proteins expressed by an organism, which is referred to as its proteome. The problem with proteomics is that it assumes that current gene models are right and that proper protein sequences can be obtained using a reference protein sequence database; however, this is not always the case, as certain peptides are not detected in the database. Mutations can also result in the emergence of new protein sequences. Proteomic, genomic, and transcriptomic data can be used to address these difficulties. Proteogenomics, which combines proteomics and genomics, was established as a separate science in 2004.

1. Single-cell proteogenomics is a term that has recently been used to describe the profiling of surface proteins and mRNA transcripts from single cells using techniques such as CITE-Seq, despite the fact that the aim of these investigations are not linked to peptide discovery. Since the year 2019, these techniques have been referred to as multimodal omics or multi-comics.

Methodology: The fundamental idea behind the proteogenomic technique is to find peptides by comparing MS/MS data to protein databases with predicted protein sequences. The protein database is created in a number of methods using genomic and transcriptomic data.

Applications: Proteogenomics can be used in a variety of ways. Gene annotation enhancement in multiple species is one application. Gene annotation entails the identification of genes and their activities. In prokaryotic species, proteogenomics has proven particularly effective in the identification and refinement of gene annotations. Various microorganisms, such as *Escherichia coli*, *Mycobacterium*, and numerous species of *Shewanella* bacteria, have had their genomic annotation investigated using the proteogenomic technique. Proteogenomic investigations can reveal the existence of planned frameshifts, N-terminal methionine excision, signal peptides, proteolysis, and other post-translational changes, in addition to enhancing gene annotations. Proteogenomics has medical benefits, particularly in oncology research. Methylation, translocation, and somatic mutations are all examples of genetic alterations that cause cancer. According to research, understanding the molecular changes that contribute to cancer require both genomic and proteomic knowledge. This has been assisted by proteogenomics, which has identified protein sequences that may have functional functions in cancer. A recent example of this was a research involving colon cancer that led to the discovery of prospective cancer therapeutic targets. Proteogenomics has also led to customised cancer targeting immunotherapies, in which antibody epitopes for cancer antigens are predicted using proteogenomics, allowing for the development of drugs that target the patient's individual tumour. Proteogenomics may give insight into cancer diagnosis in addition to treatment. Proteogenomics was used to discover somatic mutations in colon and rectal cancer investigations. The detection of somatic mutations in patients might aid in the diagnosis of cancer.

Proteogenomics may provide peptide identification techniques without the drawbacks of proteomics, such as inadequate or erroneous protein databases; nonetheless, the proteogenomic approach comes with its own set of obstacles. The sheer amount of protein databases created is one of the most difficult aspects of proteogenomics. According to statistics, a big protein database is more likely to result in inaccurate protein database data matching to MS/MS data, which can obstruct the discovery of novel peptides.

Correspondence to: Nafiseh Yavari, Department of Bioresource Engineering, McGill University, Sainte-Anne-de-Bellevue, QC, Canada, E-mail: Nafiseh.yavari@12mail.mcgill.ca

Received: October 07, 2021; **Accepted:** October 21, 2021; **Published**: October28, 2021

Citation: Yavari N (2021) A Brief Note on Proteogenomics. J Proteomics Bioinform.14:e130.