OPEN ACCESS Freely available online

ISSN: 2153-0602
Journal of
Data Mining in
Genomics & Proteomics

Commentary

Protein Biomarkers and their Relationship with Genomic, Transcriptomics and Metabolomics Data

Martin Batcho*

Department of Genetics, Palacký University, Šlechtitelů, Czech Republic

DESCRIPTION

Protein biomarkers are observable flags in a patient sample that correlate with important events or indicate specific states in a biological process, such as disease progression. They can be used for diagnosis, prognosis, prediction and exposure assessment of various human conditions. However, protein biomarkers are often influenced by multiple factors, such as genetic variations, gene expression, post-translational modifications, environmental stimuli and metabolic pathways. Therefore, integrating protein biomarker data with other omics data, such as genomic, transcriptomics and metabolomics data, can provide a more comprehensive and holistic view of the biological system and its perturbations.

Genomic data refers to the information encoded in the DNA sequence of an organism, such as Single Nucleotide Polymorphisms (SNPs), Copy Number Variations (CNVs), Insertions and Deletions (indels) and Structural Variations (SVs). These genetic variations can affect the function and regulation of genes and proteins, as well as their interactions with metabolites. For example, SNPs can alter the activity or specificity of enzymes involved in metabolic pathways, leading to changes in metabolite levels and profiles. Similarly, CNVs can result in gene dosage effects or gene fusion events that can affect protein expression and function. Integrating genomic data with protein biomarker data can help to identify the genetic determinants of protein biomarker variation and to stratify patients based on their genomic profiles.

Transcriptomics data refers to the information derived from the measurement of mRNA levels in a cell or tissue, reflecting the gene expression status of the organism. Transcriptomics data can be obtained using various technologies, such as microarrays, RNA-seq or single-cell RNA-seq. Transcriptomics data can provide insights into the transcriptional regulation of genes and proteins, as well as their response to external stimuli or internal signals. For example, transcriptomics data can reveal the differential expression of genes and proteins between normal and diseased tissues, or between different stages of disease progression. Integrating transcriptomics data with protein biomarker data can help to elucidate the molecular mechanisms underlying protein biomarker changes and to identify potential transcriptional regulators or targets of protein biomarkers.

Metabolomics data refers to the information derived from the measurement of small molecules (<1200 Da) and biochemical intermediates (metabolites) in a biological sample, reflecting the metabolic status of the organism. Metabolomics data can be obtained using various analytical platforms, such as Mass Spectrometry (MS) or Nuclear Magnetic Resonance (NMR) spectroscopy. Metabolomics data can provide insights into the metabolic pathways and networks involved in various biological processes, such as energy production, biosynthesis, signaling and detoxification. For example, metabolomics data can reveal the metabolic alterations associated with disease onset, progression or treatment response. Integrating metabolomics data with protein biomarker data can help to characterize the metabolic consequences or causes of protein biomarker changes and to identify potential metabolic markers or modulators of protein biomarkers.

There are various methods and tools for integrating protein biomarker data with genomic, transcriptomics and metabolomics data, depending on the research question and the type and quality of the data.

Approaches for integrating protein biomarker data with genomic, transcriptomics and metabolomics data

Correlation-based integration: This approach involves calculating the pairwise correlation coefficients between protein biomarkers and other omics variables across samples or groups. This can help to identify linear associations or co-variations between different omics domains. However, this approach may not capture nonlinear or complex relationships or causal effects.

Correspondence to: Martin Batcho, Department of Genetics, Palacký University, Šlechtitelů, Czech Republic, E-mail: martin@batch.as.edu

Received: 03-Mar-2023, Manuscript No. JDMGP-23-20750; **Editor assigned:** 06-Mar-2023, JDMGP-23-20750 (PQ); **Reviewed:** 20-Mar-2023, QC No. JDMGP-23-20750; **Revised:** 27-Mar-2023, Manuscript No. JDMGP-23-20750 (R); **Published:** 03-Apr-2023, DOI: 10.4172/2153-0602.23.14.290

Citation: Batcho M (2023) Protein Biomarkers and their Relationship with Genomic, Transcriptomics and Metabolomics Data. J Data Mining Genomics Proteomics. 14:290

Concatenation-based integration: This approach involves combining protein biomarker data with other omics data into a single matrix or table and applying multivariate statistical methods, such as Principal Component Analysis (PCA), cluster analysis or discriminant analysis. This can help to reduce dimensionality, identify patterns or clusters or classify samples based on multiple omics features. However, this approach may not account for heterogeneity, noise or missing values in the data.

Multivariate-based integration: This approach involves applying multivariate regression models, such as Partial Least Squares (PLS), Canonical Correlation Analysis (CCA) or Sparse Canonical Correlation Analysis (SCCA), to relate protein biomarker data with other omics data. This can help to identify latent variables or components that capture the common or shared variation between different omics domains.