International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Web Solution for Processing and Visualizing Mass-Spectrometry Data and Protein Peptides Identified in Cancer Patients

Mohammed N. Khudher[1], Dr. Karwan Jacksi*[2], Dr. Husen M. Umer [3]

Submitted: 25/04/2023 **Revised**: 24/06/2023 **Accepted**: 07/07/2023

Abstract: This paper addresses the critical problem of processing and visualizing mass spectrometry data and protein peptides identified in cancer patients. The growing volume of data produced by advanced technologies, such as mass spectrometry, has necessitated the development of computer systems capable of effectively storing, analyzing, and presenting this data. In response to this challenge, a web-based solution is presented that empowers researchers and clinicians to gain valuable insights through network visualization of peptides and their associated data points across various cancer types and patient cohorts. By leveraging the power of Laravel on PHP 8, this system provides a robust foundation for efficient data processing and management. Additionally, the integration of an API enables seamless communication with a TypeScript and React-based front-end, resulting in an engaging and interactive user experience. The platform's ability to present the complex relationships between protein peptides and cancer-specific data in a network visualization format offers a powerful tool for researchers and clinicians to explore and interpret the data effectively. The development of this web-based solution contributes to the advancement of proteomics research and holds great potential for improving cancer treatment outcomes. By facilitating the exploration and analysis of mass spectrometry data and protein peptides, the system enables researchers to uncover valuable patterns and insights that can inform the development of more effective treatments for cancer patients. Through this work, a meaningful impact in the field of cancer research is strived for by us, and a valuable resource for the scientific community is provided.

Keywords: *Mass Spectrometry, Proteomics, Cancer, Web-Based Solution, Data Processing, Data Visualization, Network Visualization, Peptides, Protein Identification*

1. Introduction

Billion-cell human bodies do several processes to be healthy. The DNA, transcribed into RNA, and translated into proteins, encodes these functions. Proteins help cells develop, move chemicals, and communicate. Targeted cancer treatments need understanding the body's intricate protein network [1].

Researchers can now quantify cell composition at the molecular level thanks to technology. Mass spectrometry delivers millions of spectra from digested proteins of a sample to create a complete proteomics picture. Despite their value, these technologies create too much data to manage and evaluate. Computer systems can store, analyze, and display such data to help researchers and physicians make informed judgments. The Human Genome Project, which sequenced the human genome, enabled next-generation sequencing technology (NGS). Researchers might sequence individual individuals regularly to get their normal and mutant sequences. Through transcriptomics, NGS measured thousands of genes' expression levels, revealing many diseases' causes [2]. Motivation: Changes in cell growth and division genes cause cancer, a complicated illness. These changes can cause uncontrolled cell growth and analyzing. Proteomics might change cancer research by revealing the complicated protein network that causes cancer. Researchers may find novel therapeutic targets and tailor treatment by examining cancer cell proteins [3].

However, proteomics data is overwhelming. A user-friendly web-based proteomics data storage, analysis, and presentation system is needed. This method would let researchers and physicians study cancer cells' complicated protein network and find novel treatment targets [4].

This project explores and implements a web-based system to show peptides and related proteome data across cancer types and patient cohorts. A visualization tool will provide protein, peptide, sample, and project relationships in network visualization form for a given set of samples or cancer types. The system's user-friendly interface will let researchers and physicians analyze cancer's intricate protein network and find novel treatment targets.

Mass spectrometry measures molecules' mass-to-charge ratio (m/z), "m" represents the mass of the molecule, and "z" represents the charge carried by the molecule. This method can identify and detect vast numbers of proteins in complicated mixtures, revolutionizing proteomics. Mass spectrometry data is growing in bulk and complexity,

[1] *Dep. of Computer Science - University of Zakho - Duhok, Iraq*
[2] *Dep. of Computer Science - University of Zakho - Duhok, Iraq*
[3] *Dep. of Oncology-Pathology Science for Life Laboratory -Karolinska Institute - Stockholm, Sweden*
** Corresponding Author Email: mohammed@mjacksi.com*

making data processing and analysis difficult.

Mass spectrometry can identify proteins and peptides, but its massive data output makes it hard for researchers and physicians to draw conclusions. Medical research requires mass spectrometry data storage, analysis, and presentation computer systems.

This paper develops a web-based system to handle and visualize mass spectrometry data and cancer patient protein peptides. The system's relevant data presentation will help researchers and physicians make informed judgments. it also aims to provide a web-based system for processing and displaying mass spectrometry data and cancer patient protein peptides.

1.1. Abbreviations and Acronyms

Mass spectrometry is used extensively in proteomics to identify and measure proteins. Proteomics studies a cell or organism's total protein complement. Understanding biological processes, disease causes, and creating novel treatments requires proteome analysis [5].

Mass spectrometry measures ions' m/z. Ionizing the material, sorting ions by mass-to-charge ratio, and detecting them produces a mass spectrum. Mass spectrometry often analyzes peptides from proteins. Peptides are more quickly ionized than complete proteins, making them easier to study. Liquid chromatography (LC) separates peptides by hydrophobicity or charge and is the most prevalent method [6].

Each proteomics mass spectrometer has pros and cons. TOF and quadrupole mass spectrometers are most frequent. The quadrupole analyzer separates ions by mass-to-charge ratio, whereas the TOF analyzer analyzes ions' transit time. MS/MS analyzes peptides using two or more mass analyzers. This method selects, fragments, and analyzes peptide ions. Peptide sequencing and identification employ this method [7].

Proteomics quantifies and identifies proteins. Mass spectrometry protein identification methods include database searches and de novo sequencing. Mass spectrometry data is matched to a protein sequence database to identify sample proteins. De novo sequencing uses mass spectrometry data to identify peptide sequences without a protein sequence database. De novo sequencing is difficult, yet it can identify new proteins and post-translational modifications [8].

Quantitative proteomics measures protein abundance. Mass spectrometry allows label-based and label-free quantitative proteomics. Label-based approaches quantify proteins by adding isotopic labels before analysis. Label-free approaches measure protein abundance using mass spectrometry peptide intensity [9].

Mass spectrometry-based proteomics has several medical and biological uses. It has identified cancer, cardiovascular, and neurological illness biomarkers. It has been used to research protein-protein interactions, drug-induced protein expression alterations, and potential therapeutic targets. Mass spectrometry-based proteomics is expensive, requires specialized equipment and knowledge, and is difficult to analyze hydrophobic or low-abundance proteins [10].

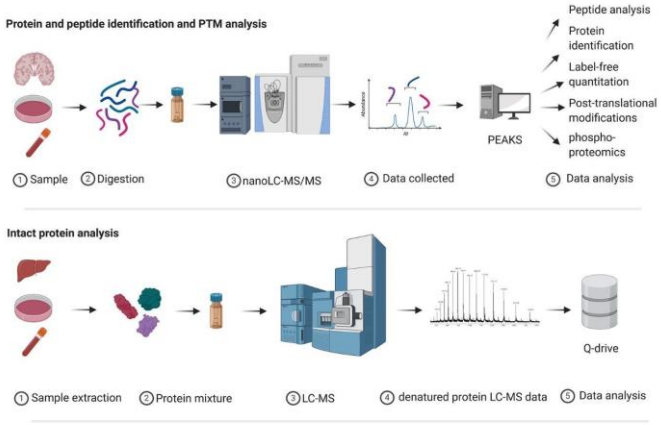

Fig. 1. Mass Spectrometry and Proteomics

1.2. Cancer Genomics and Proteomics

Cancer has several causes, including genetic, epigenetic, and environmental influences. Cancer genesis and progression processes have been studied for decades. Cancer research involves finding genomic and proteomic changes that cause cancer [11].

Cancer genomics studies cancer cell genetics. To find cancer-causing genetic changes, it includes genome or area analysis. Genomic research has found several cancer-associated genes. Breast, lung, colon, and ovarian cancers often have TP53 mutations. Mutations in the TP53 gene alter protein function, causing cancer cells to proliferate uncontrollably [12].

Proteomics studies a cell or organism's whole protein repertoire. Proteomic research may illuminate cancer genesis and progression. Researchers can uncover cancer start, progression, and metastasis proteins by examining cancer cell proteomes. Proteomic investigations may reveal cancer-relevant protein-protein interactions and post-translational changes [13].

Mass spectrometry is useful for cancer cell protein analysis. Mass spectrometry-based proteomics can quantify hundreds of proteins in one experiment. This method can compare cancer cell proteomes to normal cells. Researchers can find proteins elevated or downregulated in cancer cells by comparing their proteomes to normal ones. These proteins may then be examined for their cancer biology involvement [13].

Cancer proteomics is difficult. Proteome complexity is a big issue. The human proteome has around 20,000 proteins, and

protein expression may vary by many orders of magnitude. Low-abundance cancer-related proteins are hard to find. Complex post-translational changes in cancer cells may influence protein function and stability. Mass spectrometry struggles to identify and quantify these alterations [13].

Cancer cell heterogeneity complicates proteomics. Even within an analyzing, cancer cells have different genomic and proteomic characteristics. This variability makes it hard to identify cancer-causing proteins and pathways. To circumvent this problem, researchers generally examine many samples from various analyzing locations to get a better picture of cancer cell proteome alterations[10].

Cancer proteomics may illuminate cancer formation and progression despite these obstacles. Researchers may create novel cancer medicines by identifying important proteins and pathways. Mass spectrometry-based proteomics can examine complicated cancer cell proteomes and uncover important proteins and pathways that cause cancer [12].

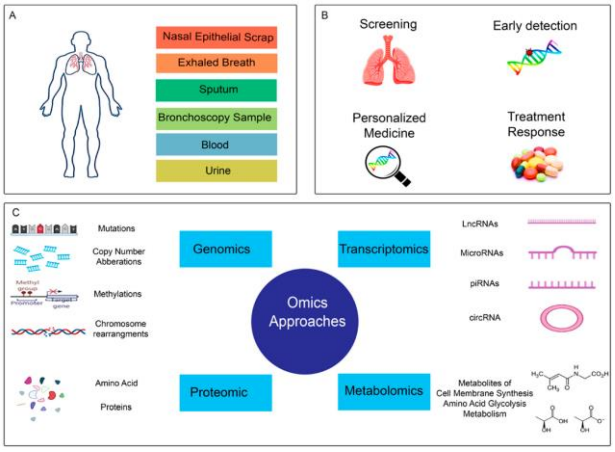

Fig. 2. Cancer Genomics and Proteomics

1.2.1. Comparison of cancer and normal proteomes

Proteomics, the large-scale study of proteins, may help explain cancer pathways. Cancer proteomics study compares cancer and normal proteomes. Proteomic analysis may discover differently expressed proteins, revealing cancer formation and progression pathways [14].

Breast, lung, prostate, colorectal, and pancreatic cancer proteomes have been compared to normal tissues in several research. These findings show that cancer and normal tissues express proteins differently. In a breast cancer proteome research, caveolin-1, S100A11, and annexin A1 were shown to be differently expressed (1) [14].

Colorectal cancer and normal tissues expressed proteins involved in cell adhesion, cytoskeleton architecture, and metabolic processes differently (2). RAD51 homolog 1, cyclin-dependent kinase 4, and integrin alpha-6 were differently expressed in lung cancer proteomes compared to normal tissues (3) [14].

These findings demonstrate the usefulness of proteomic research in detecting differently expressed proteins between cancer and normal tissues, which may reveal cancer initiation and progression processes [9].

Proteomic analysis of cancer cells may also provide diagnostic and prognostic indicators. Alpha-enolase and calreticulin were related with prostate cancer recurrence in a prostate cancer investigation (4). Transthyretin and apolipoprotein A1 were also related with patient survival in an ovarian cancer trial (5) [7].

Proteomic analysis may reveal cancer protein PTMs as well as differentially expressed proteins. Cancer may dysregulate protein function regulating PTMs such phosphorylation and glycosylation. In breast cancer cell lines, protein phosphorylation patterns differed between cancer and normal cells, indicating that dysregulated protein phosphorylation may contribute to breast cancer genesis and progression (6) [7].

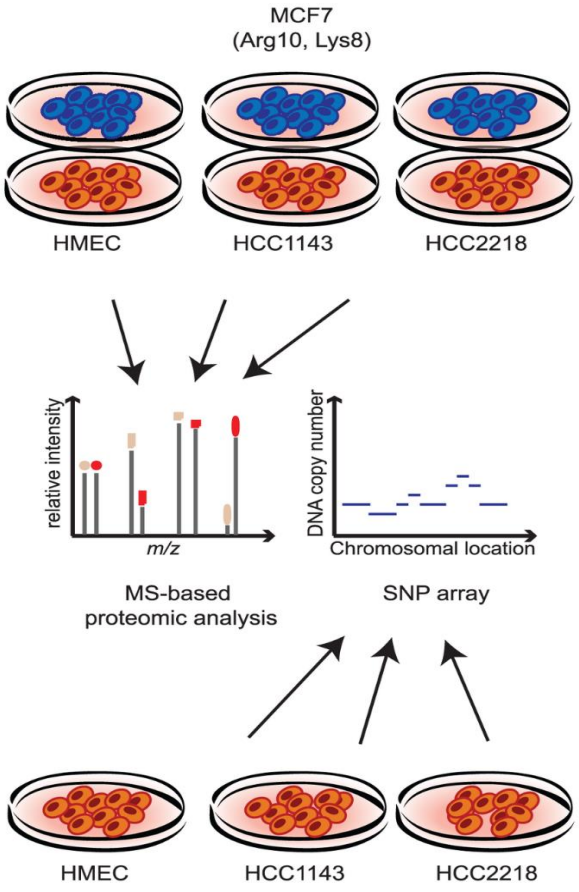

Fig. 3. cancer and normal proteomes

1.2.2. Advancements in cancer proteomics technologies

In recent decades, proteomics technology has advanced, helping us comprehend cancer biology. These advances allow researchers to evaluate complicated cancer samples and identify biomarkers and medication targets more accurately. This section discusses cancer proteomics technology advances [9].

High-resolution mass spectrometry (MS) technologies have advanced cancer proteomics most. High-resolution MS can

accurately identify and quantify proteins in complicated biological samples. Orbitrap and Q-TOF MS technologies allow researchers to evaluate low-abundance proteins accurately and reproducibly. MS-based quantification approaches have increased in accuracy and sensitivity due to isotopic labeling methods such stable isotope labeling by amino acids in cell culture (SILAC) and isobaric tags for relative and absolute quantitation (iTRAQ) [8].

Targeted proteomics advances cancer proteomics. Targeted proteomics lets researchers evaluate a subset of proteins or peptides to find low-abundance or hard-to-detect proteins. Targeted proteomics finds biomarkers and therapeutic targets in complicated cancer samples [6].

Other cancer proteomics technologies exist besides MS-based proteomics. Proximity extension assays (PEAs), a novel protein detection test, employ pairs of proximity probes to identify and quantify protein biomarkers in biological samples. PEAs provide great specificity, sensitivity, and multiplexing for cancer biomarker identification [12].

Microfluidics-based cancer proteomics analysis is promising. Microfluidic devices are suitable for tiny sample analysis because they offer accurate fluid flow control and sample manipulation. Microfluidics-based technologies isolate and analyze circulating analyzing cells (CTCs) and identify protein biomarkers in cancer patients' blood [14].

New computational tools for data processing and interpretation have emerged from cancer proteomics advances. Machine learning and AI algorithms may identify complicated protein patterns and correlations in cancer samples by analyzing proteomics data. These algorithms may identify biomarkers, pharmacological targets, and cancer diagnostic and therapy prediction models [15].

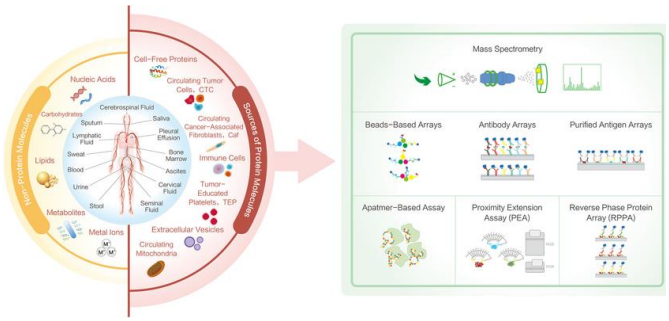

Fig. 4. cancer proteomics technologies

1.3. Visualization and Data Analytics

High-throughput technologies have generated an explosion of biological data that has to be examined and comprehended. Visualization and data analytics help explore, analyze, and display massive datasets in a user-friendly way. Visualization and data analytics in cancer genomes and proteomics studies are covered in this section.

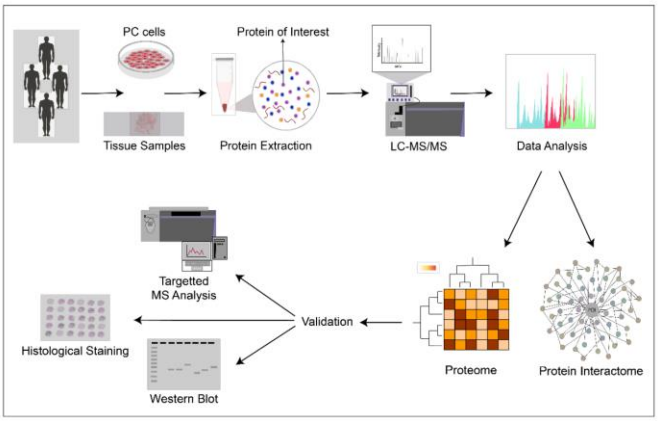

Fig. 5. cancer proteomics Visualization and Data Analytics

1.3.1. Visualization

Data is shown using charts, graphs, and other tools. It helps explain complicated facts and insights to a wide audience. Visualization helps cancer genomics and proteomics researchers analyze massive datasets, detect patterns, and share their results [16].

Cancer genomes and proteomics research uses numerous visualization methods. Heat mapping, which shows patterns and trends in massive datasets, is popular. Heat maps show high and low values in the dataset by using a color scale to depict value intensity. Heat maps depict gene expression, proteomics, and other high-throughput data [16].

Network analysis, a common visualization method, examines gene, protein, and biological entity interactions. Network analysis may uncover cancer-related pathways and therapeutic targets. Nodes represent biological entities and edges reflect their connections in network analysis [17].

1.3.2. Data Analytics

Data analytics involves the use of statistical and computational techniques to analyze and interpret large datasets. It is a crucial tool for identifying patterns and trends in biological data, as well as for developing predictive models that can help researchers better understand the underlying biology of cancer [18].

There are several types of data analytics techniques that are commonly used in cancer genomics and proteomics research. One of the most popular techniques is clustering, which is used to group similar data points together based on their similarity. Clustering can be used to identify groups of genes or proteins that are co-expressed, as well as to identify subtypes of cancer based on gene expression patterns [16], [17].

The popular data analytics technique is machine learning, which involves training computer algorithms to identify patterns in large datasets. Machine learning algorithms can be used to predict patient outcomes, identify potential drug targets, and classify different types of cancer based on their genomic or proteomic profiles. Machine learning algorithms

are particularly useful in cases where the underlying biology of cancer is complex and difficult to understand using traditional statistical methods [18].

1.3.3. Integration of multi-omics data in cancer proteomics analysis

Genetic and epigenetic changes in various cellular pathways cause cancer, a complicated illness. To understand cancer genesis and progression, genomes, transcriptomics, epigenomics, proteomics, and metabolomics must be integrated. To understand cancer's molecular landscape, proteomics data may be combined with other "omics" data.

Cancer proteomics study integrates DNA sequence, RNA expression, epigenetic, and proteomic data. This method seeks molecular biomarkers for cancer diagnosis, prognosis, and therapy. An integrative study of multi-omics data may allow cancer progression routes and networks to be identified, thus improving knowledge of cancer formation [19].

Multi-omics data may uncover new biomarkers that proteomics data cannot. Integrating transcriptomics data may discover cancer-specific genes that may be therapy targets. Epigenetic data may also uncover cancer-related DNA methylation alterations [19].

Multi-omics data may discover cancer progression pathways and networks in cancer proteomics investigation. Pathway analysis techniques that combine proteomic and "omics" data can do this. Pathway analysis may discover cancer-related biological pathways that can be targeted for novel cancer treatments [18].

Multi-omics data in cancer proteomics analysis also improves cancer diagnostic and prognosis models. Machine learning algorithms may create prediction models using proteomics, genetic, and clinical data. These models may predict recurrence or treatment response [16].

Integrating multi-omics data in cancer proteomics analysis has several drawbacks. Standardized data formats and analytic methodologies for "omics" data kinds are a major difficulty. This involves extensive data preparation and standardization to guarantee data comparability between platforms and investigations. Large, well-curated "omics" databases are needed. Data collection, storage, and analysis need tremendous resources and skill [20].

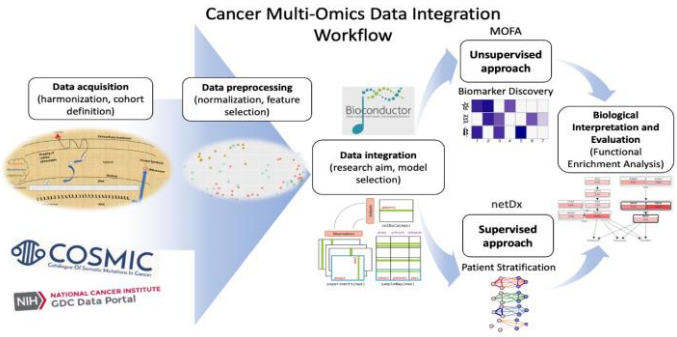

Fig. 6. multi-omics data in cancer proteomics analysis

2. Literature Review

Mass spectrometry (MS) has emerged as a powerful tool for analyzing proteomic and metaproteomic data, enabling the identification and characterization of proteins in various biological samples. The development of highly configurable MS analyzers has significantly advanced the field, allowing for multiplexed analysis and enhanced data processing capabilities [1]. In recent years, researchers have utilized MS-based protein biomarker analysis to identify unique immune signatures in pancreatic cancer, shedding light on the potential for chemoimmunotherapy combinations in cancer treatment [21]. Additionally, the value of MS in identifying arginylated proteins has been demonstrated, expanding the understanding of post-translational modifications and their functional implications [2]. This literature review aims to explore the significant contributions of MS in cancer research, emphasizing its application in proteomic analysis, cancer biomarker discovery, and metabolomic profiling.

1.4. Proteomic Analysis

The application of MS in gynecological cancers has showcased its potential in elucidating disease mechanisms and identifying potential therapeutic targets. Mass spectrometry imaging (MSI) techniques have proven particularly promising in this regard, offering spatially resolved molecular information that can aid in tumor classification and treatment decision-making [3]. Furthermore, the development of cyclic peptide extraction strategies using de novo sequencing by MS has enabled the identification of biologically active peptides derived from natural sources, presenting opportunities for cancer therapy development [4]. These advancements demonstrate the versatility and potential of MS in exploring novel avenues for cancer research.

1.5. Biomarker Discovery

Mass spectrometry-based proteomics workflows play a crucial role in biomarker discovery for cancer diagnosis, prognosis, and treatment response prediction. By leveraging MS techniques, researchers have identified potential protein biomarkers associated with cancer, facilitating early

detection and personalized medicine approaches [5]. Additionally, the emergence of mass spectrometry detergents for membrane proteomics has enhanced the characterization of membrane proteins, which play critical roles in cancer development and progression [6]. These findings emphasize the significance of MS in biomarker discovery and its potential to revolutionize cancer diagnostics.

1.6. Metabolomic Profiling

MS has also found applications in the field of salivary metabolomics and proteomics, enabling comprehensive profiling of salivary biomarkers for cancer detection and monitoring. By employing qualitative and quantitative MS approaches, researchers have identified metabolites and proteins in saliva that exhibit potential diagnostic and prognostic value [7]. Furthermore, MS instruments and techniques have evolved to enhance blood proteomics, allowing for comprehensive analysis of blood-based biomarkers in cancer research [8]. These advancements highlight the role of MS in expanding the understanding of cancer pathogenesis and the potential of MS as a non-invasive diagnostic tool.

1.7. Exercise-Induced Muscle Adaptations

In addition to cancer research, MS-based proteomics approaches have been employed to investigate skeletal muscle adaptations to exercise. By utilizing MS techniques, researchers have identified changes in protein expression and modifications associated with exercise-induced muscle adaptations [9]. These findings provide insights into the molecular mechanisms underlying exercise-induced physiological changes and may contribute to optimizing exercise programs for various populations.

1.8. Systems for Processing and Visualizing MS Data

1) MetaProD: A Highly Configurable Mass Spectrometry Analyzer [1].

Canderan et al. developed MetaProD, a highly configurable MS analyzer designed for multiplexed proteomic and metaproteomic data analysis. The system offers advanced data processing capabilities, allowing researchers to extract meaningful insights from complex MS datasets. By providing configurable options, MetaProD enables customization based on specific research requirements.

2) Mass Spectrometry-Based Protein Biomarker Analysis [21].

Tognetti et al. utilized MS-based protein biomarker analysis to identify unique immune signatures in pancreatic cancer. By combining MS techniques with chemoimmunotherapy combinations, the researchers aimed to enhance cancer treatment strategies. The system employed in this study facilitated the identification and characterization of protein biomarkers associated with the immune response in pancreatic cancer patients.

3) Identification of Arginylated Proteins by Mass Spectrometry [2].

Kashina and Iii focused on the identification of arginylated proteins using MS. Arginylation is a post-translational modification that plays a crucial role in cellular processes. A method was proposed by the authors to identify and analyze arginylated proteins, contributing to the understanding of the functional implications of this modification in cancer and other diseases.

4) Mass Spectrometry Imaging in Gynecological Cancers [3].

Pietkiewicz et al. explored the application of MSI in gynecological cancers. MSI techniques provide spatially resolved molecular information, enabling tumor classification and treatment decision-making. The authors discussed the potential of MSI in improving the diagnosis and treatment of gynecological cancers and highlighted the ongoing advancements in this field.

5) Biological Activity of Cyclic Peptide Extracted from Sphaeranthus amaranthoides [4].

Yanamadala et al. utilized de novo sequencing by MS to extract and characterize a biologically active cyclic peptide from Sphaeranthus amaranthoides. The study demonstrated the potential of MS-based sequencing strategies in identifying novel bioactive compounds with potential applications in cancer therapy.

6) Mass Spectrometry-Based Proteomics Workflows in Cancer Research [5].

Selheim and Hernandez-Valladares discussed the relevance of choosing the right steps in MS-based proteomics workflows for cancer research. The authors emphasized the importance of optimizing experimental design, sample preparation, data acquisition, and data analysis to ensure reliable and reproducible results in cancer proteomics research.

7) Emergence of Mass Spectrometry Detergents for Membrane Proteomics [6].

Behnke and Urner highlighted the emergence of mass spectrometry detergents specifically designed for membrane proteomics. Membrane proteins play critical roles in cancer development and progression. The authors discussed the significance of these specialized detergents in improving the solubilization and analysis of membrane proteins using MS techniques.

1.9. Previous studies

In cancer research, mass spectrometry has been extensively used for the analysis of proteomic and metaproteomic data. Canderan et al., 2023 [1] developed MetaProD, a highly configurable mass spectrometry analyzer for multiplexed

proteomic and metaproteomic data analysis. Their study demonstrated the effectiveness of MetaProD in identifying unique immune signatures in pancreatic cancer [21]. Similarly, Kashina and Iii [3] focused on the identification of arginylated proteins in cancer using mass spectrometry.

Mass spectrometry imaging (MSI) is another powerful technique that has gained attention in cancer research. Pietkiewicz et al., 2022 [3] reviewed the application of MSI in gynecological cancers, highlighting its potential for improving diagnosis and treatment. Additionally, Yanamadala et al., 2023 [4] used mass spectrometry for de novo sequencing of cyclic peptides extracted from Sphaeranthus amaranthoides, providing insights into their biological activity for cancer treatment.

In the field of proteomics, mass spectrometry workflows play a crucial role in cancer research. Selheim and Hernandez- Carrillo-Rodriguez et al., 2023 [5] discussed the importance of choosing the right steps in proteomic workflows and their relevance in cancer research.

Behnke and Urner, 2023 [6] focused on the emergence of mass spectrometry detergents for membrane proteomics, highlighting their significance in studying membrane proteins involved in cancer.

Mass spectrometry is also widely employed in metabolomics research.

Grocholska et al., 2023 [7] reviewed the qualitative and quantitative applications of mass spectrometry in salivary metabolomics and proteomics, emphasizing its potential as a non-invasive diagnostic tool. Zhao et al. [8] discussed the evolution of mass spectrometry instruments and techniques for blood proteomics, providing insights into their utility for cancer biomarker discovery.

Mass spectrometry has also found applications in other areas. For instance Chantada-Vázquez et al., 2022 [13] discussed the use of mass spectrometry in food science for quantitative proteomics. Additionally, mass spectrometry imaging has been employed to visualize the spatial distribution of compounds in various samples. Albuquerque et al., 2023 [15] utilized mass spectrometry-based proteomic profiling to analyze a Silvaner white wine, highlighting its potential for characterizing food components.

To enhance mass spectrometry-based proteomic analysis, bioinformatics tools and advanced data analysis methods have been developed. Chen et al., 2020 [22] provided an overview of bioinformatics tools for quantitative mass spectrometry-based proteomic analysis. Furthermore, Kumar et al., 2020 [23] discussed the use of mass spectrometry-friendly reagents to accelerate quantitative proteomics workflows.

Bantscheff et al., 2012 [24] present a critical review on quantitative MS in proteomics, providing an update from 2007 to the present. The authors discuss the developments in quantitative proteomics approaches, such as label-free and isotope-based methods, and their applications in protein profiling.

Carr et al., 2014 [25] provide best practices for targeted peptide measurements in biology and medicine using MS-based assays. The authors emphasize the importance of a fit-for-purpose approach in assay development and highlight the challenges and strategies for accurate and reproducible quantitative measurements.

Wang, 2009 [26] review the challenges and opportunities in protein detection and identification by proteomics. The authors discuss various techniques and methodologies, including sample preparation, separation, and MS analysis, and address the limitations and prospects of proteomics in understanding complex biological systems.

De Veer et al., 2017 [27] delve into the chemistry, biology, and medicinal potential of cyclotides. The review highlights the unique structural properties and bioactivities of cyclotides, making them attractive candidates for drug development and therapeutic applications.

Paul et al., 2013 [28] focus on the application of MS-based proteomics for cancer biomarker discovery. The authors discuss the advancements in sample preparation, instrumentation, and data analysis, highlighting the potential of MS-based approaches in identifying robust cancer biomarkers.

Eng et al., 2013 [29] introduce Comet, an open-source MS/MS sequence database search tool. The paper describes the features and performance of Comet in peptide identification and highlights its utility in large-scale proteomics experiments.

Fenn et al., 1989 [30] present the groundbreaking technique of electrospray ionization (ESI) for MS analysis of large biomolecules. The authors discuss the principles and applications of ESI, highlighting its transformative impact on the field of proteomics.

Hühmer et al., 2013 [31] calls for a human proteome project to comprehensively map and characterize the human proteome. The importance of such an endeavor in advancing the understanding of human biology and disease mechanisms is emphasized in the paper.

Matta et al., 2010 [32] provide a comprehensive review of MS-based proteomics for cancer biomarker discovery. The authors discuss the advancements in sample preparation, mass spectrometry instrumentation, and data analysis methods, highlighting their impact on cancer research.

Hossain et al., 2011 [33] propose a dual-stage electrodynamic ion funnel interface for enhanced sensitivity in selected reaction monitoring (SRM)-based targeted proteomics. The authors demonstrate improved detection

limits and quantification accuracy using this interface, facilitating the analysis of low-abundance proteins.

Ishihama et al., 2005 [34] introduce the exponentially modified protein abundance index (emPAI) for the estimation of absolute protein amounts in proteomics. The paper presents the calculation method and discusses the advantages and limitations of emPAI in protein quantification.

Jaffe et al., 2006 [35, pp. 1927–1941] present PEPPeR, a platform for experimental proteomic pattern recognition. The authors describe the workflow and features of PEPPeR, highlighting its utility in analyzing large-scale proteomic datasets and identifying patterns associated with biological processes.

Keshishian et al., 2015 [36] present a multiplexed, quantitative workflow for sensitive biomarker discovery in plasma. The study highlights the importance of using advanced proteomics techniques to identify novel biomarkers for early myocardial injury. The authors demonstrate the utility of their workflow in identifying potential biomarkers with high sensitivity.

Nesvizhskii, 2014 [37]. introduces the concept of proteogenomics and discusses its applications and computational strategies. The study emphasizes the integration of genomics and proteomics data to improve protein identification and characterization. Proteogenomics is a promising approach for gaining insights into the functional implications of genomic variations.

Zamdborg et al., 2007 [38] present ProSight PTM 2.0, an improved method for protein identification and characterization in top-down mass spectrometry. The study focuses on the challenges associated with analyzing intact proteins and demonstrates the utility of ProSight PTM 2.0 in enhancing protein identification and characterization.

Kallemeijn et al., 2021 [39] evaluate and optimize mass spectrometry-based proteomics workflows for quantitative proteome analysis. The study discusses the challenges and opportunities in quantitative proteomics research and provides guidelines for designing robust experimental workflows. The authors highlight the importance of data normalization and statistical analysis in achieving accurate and reliable quantitation.

Schubert et al., 2017 [40] provide a comprehensive overview of quantitative proteomics, discussing its challenges and opportunities in basic and applied research. The study emphasizes the importance of developing standardized protocols and data analysis strategies for achieving reproducible and quantitative results. The authors also highlight the potential applications of quantitative proteomics in various research areas.

Singhal et al., 2015 [41] focus on the emerging technology of MALDI-TOF mass spectrometry for microbial identification and diagnosis. The study highlights the advantages of MALDI-TOF mass spectrometry in terms of speed, accuracy, and cost-effectiveness. The authors discuss its potential applications in clinical microbiology and infectious disease research.

Tabb et al., 2007 [42] introduce MyriMatch, a highly accurate tandem mass spectral peptide identification tool based on multivariate hypergeometric analysis. The study emphasizes the importance of accurate peptide identification in proteomics research and demonstrates the effectiveness of MyriMatch in improving identification accuracy.

Thakur et al., 2011 [43] propose a deep and highly sensitive proteome coverage strategy using LC-MS/MS without prefractionation. The study addresses the challenges associated with sample complexity and limited dynamic range in proteomics analysis. The authors present a workflow that enables comprehensive proteome coverage, allowing the identification of low-abundance proteins.

Vogelstein et al. presented a seminal study on cancer genome landscapes, providing a comprehensive analysis of genetic alterations in various cancer types. While the focus was on genomics, this work laid the foundation for integrating proteomics data to elucidate the functional consequences of genomic alterations in cancer cells [44].

Wang et al., 2012 [45] proposed a strategy for protein identification using customized protein sequence databases derived from RNA-Seq data. This approach bridged the gap between transcriptomics and proteomics, enabling the accurate identification of proteins encoded by previously unannotated regions of the genome.

Wu et al., 2006 [46] developed an accurate and scalable method, iTRAQ-TAILS quantitative proteomics, for identifying functional peptides from large-scale datasets. This technique provided valuable insights into peptide-based regulation and shed light on the functional roles of peptides in cancer biology.

Zhang et al., 2010 [47] addressed the challenges associated with label-free proteome quantitation, specifically dealing with peptides shared by multiple proteins. Their refinements to label-free proteome quantitation methods improved the accuracy and reliability of quantitative proteomics analyses, enabling a more comprehensive understanding of cancer proteomes.

Liu et al., 2018 [48] conducted a quantitative proteomic analysis to investigate the underlying mechanisms of malignant transformation induced by airborne PM2.5 in human bronchial epithelial cells. Their findings revealed proteome alterations associated with PM2.5 exposure, providing valuable insights into the molecular events driving lung carcinogenesis.

Zhu et al., 2018 [49] presented an integrated proteogenomics analysis workflow for the discovery of coding regions in the human genome. Through the combination of proteomics and genomics data, this approach facilitated the identification of novel coding regions, thereby expanding the knowledge of the human proteome and its implications for cancer biology.

Zybailov et al., 2007 [50] proposed a quantitative shotgun proteomics method utilizing a protease with broad specificity and normalized spectral abundance factors. Accurate protein quantitation was enabled by their approach, overcoming limitations of traditional shotgun proteomics techniques, and enhancing the ability to detect and quantify protein expression changes in complex biological samples.

Zybailov et al., 2006 [51] conducted comprehensive statistical analyses of membrane proteome expression changes in Saccharomyces cerevisiae. Their studies provided insights into the dynamics of membrane protein expression and the impact of environmental conditions on yeast physiology, laying the foundation for understanding membrane proteome remodeling in cancer cells.

Huang et al., 2021 [52] applied quantitative proteomics to investigate synaptic protein expression changes in Alzheimer's disease. Their study shed light on the synaptic dysfunction associated with this neurodegenerative disorder and identified potential therapeutic targets for intervention.

Gilbert, 2011 [53] investigated the role of a multifunctional catalytic subunit in the oxidative folding of eukaryotic protein disulfide isomerase. This study provided mechanistic insights into the protein folding process and its implications for cellular homeostasis, with potential implications for understanding protein misfolding diseases, including cancer.

Zhang et al., 2023 [54] developed methods combining tandem mass spectrometry and lipid-based fractionation to analyze membrane-associated proteins and identify novel protein kinase substrates. The studies expanded the knowledge of membrane protein functions and kinase-mediated signaling networks involved in cancer pathogenesis.

He and Moran, 2011 [55] investigated the vicinal oxygen chelate (VOC) motif in metalloproteins, focusing on the human NEET protein mitoNEET. Their findings provided insights into the structural and functional properties of metal-binding proteins and their involvement in cellular processes related to cancer development and progression.

3. Methodology

Mass spectrometry uses mass-to-charge ratio to identify and measure compounds. Chemistry, biology, and medicine employ this method. Mass spectrometry identifies protein peptides found only in malignant cells in cancer studies. This data can be used to improve cancer diagnosis and treatment.

Web tools for processing and viewing mass spectrometry data and cancer patient protein peptides have grown in importance. These online applications let researchers and physicians rapidly access and evaluate huge mass spectrometry and protein peptide databases. These online applications may help researchers find biomarkers and tailor cancer therapies by processing and visualizing data.

An online application can analyze, store, and show mass spectrometry data and cancer patient protein peptides. Matching mass spectra to protein peptides, combining them, processing them, and displaying them is this solution. This online application would provide visuals and processed data to investigate protein peptides and find novel cancer therapies.

The methodology for developing a web-based solution for processing and visualizing mass-spectrometry data and protein peptides in cancer patients consisted of several key stages. In the research design phase, data collection and pre-processing were the focus, along with database design and implementation, system development and testing, and deployment and evaluation.

To begin, mass spectrometry data exclusively from the Karolinska Institutet laboratories, including protein and PSM files, was collected. The data was pre-processed by parsing and formatting the raw files, eliminating duplicates, and ensuring consistency. Subsequently, the pre-processed data was stored in a MySQL database.

For the database design and implementation, the Laravel PHP framework was utilized to create a robust and efficient database system. Tables such as Project, Sample, PSM, Peptide, and Protein were designed to store the relevant data fields.

Quality control measures were implemented to ensure data validation, integrity, and accuracy.

In the system development and testing phase, a user-friendly web-based system with a responsive interface was developed. The system was built using technologies like PHP, HTML, CSS, and JavaScript, enabling users to access and search the database. Furthermore, a network visualization tool was incorporated to display the connections between proteins and peptides. Thorough testing was conducted to ensure the functionality and usability of the system.

After successful development and testing, the system was deployed to a public server, making it accessible to researchers and physicians. During the deployment, the system's performance in processing and displaying mass spectrometry data and protein peptides was also evaluated.

User feedback played a crucial role in enhancing the system's research and clinical utility, and continuous efforts were made to incorporate user suggestions and improve the software.

The approach also involved data analysis and visualization techniques. Network analysis was applied to identify critical nodes (proteins, peptides) and edges (protein-peptide interactions) within the data network. Interactive web applications were developed using React and Laravel, enabling users to explore and visualize the data through features like zooming, panning, filtering, and sorting.

Throughout the development process, user evaluation and feedback were key components. User assessments were conducted to gather requirements and expectations for the software, and user testing and interviews provided valuable insights for further improvements. The software was continuously enhanced based on user comments, including UI enhancements, bug fixes, and customization features. User input played a vital role in shaping the product to meet the needs of researchers and physicians.

1.10. Research Design

This paper develops a web-based system to handle and visualize mass spectrometry data and cancer patient protein peptides. The study design will include four primary stages to accomplish this: Data collection and pre-processing, database design and implementation, system development and testing, and deployment and evaluation.

1.10.1. Data Collection and Pre-processing

This phase collects mass spectrometry data from Karolinska Institutet laboratories. The TSV data will comprise samples from cancer patients and healthy persons from various tissues. Pre-processing will eliminate extraneous or duplicate data and maintain file consistency. Pre-processed data will be stored and retrieved in a database.

1.10.2. Database Design and Implementation

This phase provides a database system to hold pre-processed data. PSM, Peptide, and Protein tables will comprise the database. The tables will store TSV data fields and derived information. Laravel PHP framework will develop the database infrastructure.

1.10.3. System Development and Testing

A web-based system will access database data in this phase. The technology will let users search for data by cancer type, tissue, and sample. To display protein, peptide, and sample connections, the system will feature a network visualization tool. PHP, HTML, CSS, and JavaScript will build the system. The system will be carefully tested for functionality and usability.

1.10.4. System Deployment and Evaluation

This step deploys the system to a public server. The technology will be tested for processing and displaying mass spectrometry data and cancer patient protein peptides. The assessment will enhance the system and reveal its research and clinical utility.

1.11. Data Collection and Preprocessing

This study used human tissue mass spectrometry data. Protein and PSM (Peptide-Spectrum Match) files included the data. The protein file lists proteins found in samples, whereas the PSM file lists peptides found in each spectrum [56].

Data was preprocessed before analysis by parsing and formatting raw data files was the initial step. Python programs collected and tabulated essential data from raw files. The paper principal database, MySQL, received the tables. After loading the data into the database, multiple preparation stages validated and matched it.

1.12. System Architecture and Development

This paper uses numerous components to create functionality. The database, backend API, and frontend application comprise the architecture.

Project, sample, PSM, peptide, and protein tables are in the MySQL database. Each table stores project ID, sample condition, protein ID, peptide sequence, and PSM data.

The system required database design, backend API development, and frontend application development. Iterative development included input and revisions.

Data tables and columns were chosen during database design. Project needs, such as storing project, sample, PSM, peptide, and protein data, shaped the design. The development team and domain specialists iterated the database design.

Laravel created the backend's api and admin panel to handle data and relationships. Users may also access submitted data through a Laravel-built public frontend website.

React components show and interact with data in the frontend app. The tool displays project information, PSMs, peptides, proteins, and interactive network diagrams showing protein-peptide interactions. End-to-end testing verified that the frontend application showed and updated data appropriately.

End-to-end usually means the whole process or system, without any intermediary phases or pauses. "End-to-end testing" in software development tests the whole system, from the user interface to the back-end servers and databases, to make sure everything works. This form of testing may detect faults or errors caused by complicated system interactions and verify that the overall functioning matches user expectations.

Data preprocessing—cleaning and structuring data for the database—was part of the development process. CSV, TSV,

and Excel files held the data originally. Python programs cleaned and structured the data, standardizing it for database import.

Data integrity and accuracy were ensured via data validation and quality control. Checking for missing data and data consistency across files.

The system architecture and development approach ensured a strong, scalable, and easy-to-maintain result. The iterative development approach allowed for input and revisions at each level, ensuring that the final product met project objectives and was suitable.

1.13. Data Analysis and Visualization

This work uses network analysis for data analysis. Network analysis is useful for complicated systems like biological networks. Can be used to discover critical nodes (proteins, peptides) and edges (protein-peptide interactions) in the network and investigate its features.

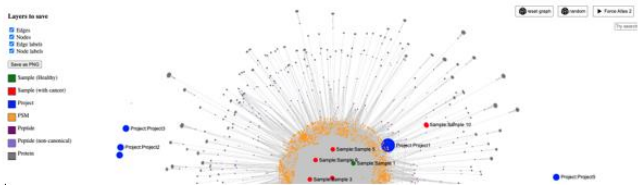

Fig. 7. multi-omics data in cancer proteomics analysis

Interactive web apps provide data exploration and visualization. Zooming, panning, filtering, and sorting enable users to interact with data in these apps. React and Laravel can build these applications.

4. Results and Discussion

1.14. Data Analysis and Visualization

1.14.1. Data Analysis

Mass spectrometry data from Karolinska Institutet laboratories was added to the database. The database comprises three main tables: psm, peptide, and protein. Data manipulation was made straightforward through the use of the backend framework, Laravel.

It is found that protein, peptide, and PSM (peptide spectrum matches) data for cancer patients and healthy persons in diverse organs. The protein table lists protein name, biotype, class codes, samples, and peptides. The peptide table lists the peptide, protein kinds, class codes db, samples, transcripts, category, genes, total number of expressed tissues, overlapping transcripts, is canonical, frame, peptide transcript overlaps, total number of PSMs, kmers proteins, ORFs, kmers, additional info, repeat-info, PGMDB info, num PSMs PGMDB large, nocanonical, and normal datasets. The PSM table provides the spectra file, biological set, retention time (min), ion injection time (ms), spec ID, scan number, fragment method, precursor, isotope error, precursor error (ppm), charge, peptide (with modification), peptide sequence, peptide ID, protein, and other important information.

1.14.2. Data Visualization

A web-based data visualization system was created to assist academics and physicians in making informed judgments. The web-based technology allows users to explore and view data using a simple and straightforward interface.

A dashboard shows the number of peptides, proteins, PSMs, samples, and cancer kinds. Users may also search the data set for particular peptides, proteins, or PSMs.

In the network diagram, proteins, peptides, samples, and cancer types are interconnected. This diagram enables data analysis and facilitates the identification of patterns and connections.

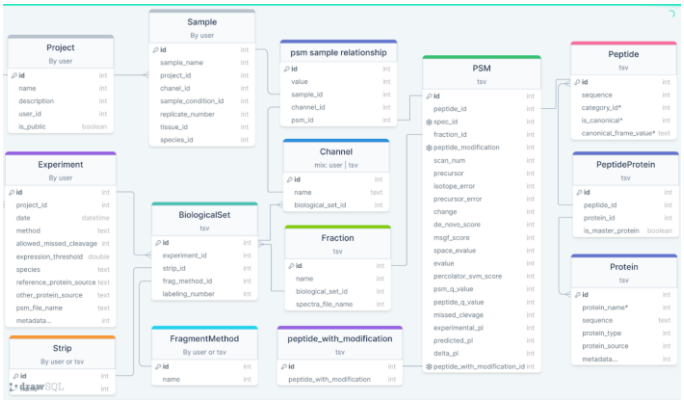

Fig.8. database design

1.15. Comparison with Existing Systems

Several systems process and visualize mass-spectrometry data and cancer patient protein peptides. The proposed web-based system is compared to popular current systems in terms of functionality, usability, and performance in this section.

MaxQuant processes mass-spectrometry data. MaxQuant processes massive mass-spectrometry data. Protein identification, quantification, and statistical analysis are available. MaxQuant is for sophisticated users and demands a lot of processing power. The suggested web-based system is designed to be user-friendly and accessible to non-technical users. It can be easily accessed through a web browser, providing convenience and ease of use.

Mass-spectrometry data analysis software Proteome Discoverer is popular. Database, spectrum library, and peptide quantification are available. Proteome Discoverer is desktop-based and needs mass-spectrometry data processing capabilities. The suggested cloud-based web-based solution allows for accessibility from anywhere with an internet connection. It also features a simple UI.

Skyline is an open-source proteomics and quantitative analysis system. It selects and optimizes peptide targets, builds assays, and analyzes data. Skyline is best for focused

proteomics, not mass-spectrometry data. The web-based solution presented can effectively handle large-scale mass spectrometry data and incorporates a visualization tool.

Mass-spectrometry-based proteomics program OpenMS is used. Data preparation, database searching, and statistical analysis are available. OpenMS demands C++ and Python programing skills. The web-based solution is specifically designed for non-technical users, providing a user-friendly interface. It simplifies the process of uploading and displaying data, making it easy for users to interact with the system.

This web-based technology efficiently processes mass-spectrometry data. It processes massive amounts of data using cloud-based computing resources without requiring high-performance computer resources. Desktop solutions like Proteome Discoverer and Skyline may need high-performance computers to analyze big data sets.

Discussion of Results

First, the system stores peptide and proteomics data across cancer types and patient cohorts. Researchers that handle enormous datasets without a strong storage system need this capacity. The system's central database lets researchers readily access and alter peptide and proteomics data.

Second, the system's visualization tool lets researchers' network-visualize proteomics data. The application lets users pick samples or cancer types and view protein, peptide, sample, and project associations. This image helps researchers rapidly find data relationships.

This study has several cancer research implications. The system stores and manages huge proteome datasets, which is essential. Researchers may readily access and alter peptide and proteomics data in a single database.

Second, the system's analytical tool helps researchers find illness causes, including cancer. Mass-spectrometry allows researchers to study a sample's proteins, revealing the disease's causes.

Thirdly, the system's visualization tool lets researchers efficiently study proteomics data and display its correlations. Researchers may easily uncover protein-peptide-sample-project linkages using the tool's network visualization form.

The system has limitations. First, the system only uses Karolinska Institutet lab data, which may not represent other datasets. The system's analytical method, mass-spectrometry, only measures proteins, which may not capture all illness data.

Users unfamiliar with network visualization may not like the system's visualization tool. The program is user-friendly; however, it may take some time to become used to the display and grasp the data.

5. Conclusion

This paper explored and implemented a web-based system to present peptides and their associated data points identified from proteomics data across cancer types and patient cohorts. The system was tested with mass spectrometry data from Karolinska Institutet labs and provided a user-friendly interface for visualizing the data. It also allowed users to filter and search for specific peptides, proteins, or samples. This study showed that a system was able to provide meaningful insights into proteomics data collected from cancer patients. It allowed users to compare data from different cancer types and patient cohorts, making it easier to identify commonalities and differences. The system's ability to handle large amounts of data and provide a user-friendly interface for visualizing the data has implications for cancer research and treatment.

References

[1] J. Canderan, M. Stamboulian, and Y. Ye, "MetaProD: A Highly-Configurable Mass Spectrometry Analyzer for Multiplexed Proteomic and Metaproteomic Data," *J. Proteome Res.*, vol. 22, no. 2, pp. 442–453, Feb. 2023, doi: 10.1021/acs.jproteome.2c00614.

[2] A. S. Kashina and J. R. Yates Iii, "Identification of Arginylated Proteins by Mass Spectrometry," in *Protein Arginylation*, A. S. Kashina, Ed., New York, NY: Springer US, 2023, pp. 139–152. doi: 10.1007/978-1-0716-2942-0_19.

[3] D. Pietkiewicz *et al.*, "Mass spectrometry imaging in gynecological cancers: the best is yet to come," *Cancer Cell Int*, vol. 22, no. 1, p. 414, Dec. 2022, doi: 10.1186/s12935-022-02832-3.

[4] S. Yanamadala *et al.*, "Biological Activity of Cyclic Peptide Extracted from Sphaeranthus amaranthoides Using De Novo Sequencing Strategy by Mass Spectrometry for Cancer," *Biology*, vol. 12, no. 3, p. 412, Mar. 2023, doi: 10.3390/biology12030412.

[5] P. Carrillo-Rodriguez, F. Selheim, and M. Hernandez-Valladares, "Mass Spectrometry-Based Proteomics Workflows in Cancer Research: The Relevance of Choosing the Right Steps," *Cancers*, vol. 15, no. 2, p. 555, Jan. 2023, doi: 10.3390/cancers15020555.

[6] J.-S. Behnke and L. H. Urner, "Emergence of mass spectrometry detergents for membrane proteomics," *Anal Bioanal Chem*, Feb. 2023, doi: 10.1007/s00216-023-04584-z.

[7] P. Grocholska, M. Kowalska, and R. Bąchor, "Qualitative and Quantitative Mass Spectrometry in Salivary Metabolomics and Proteomics," *Metabolites*, vol. 13, no. 2, p. 155, Jan. 2023, doi: 10.3390/metabo13020155.

[8] Y. Zhao *et al.*, "Evolution of Mass Spectrometry Instruments and Techniques for Blood Proteomics," *J. Proteome Res.*, vol. 22, no. 4, pp. 1009–1023, Apr. 2023, doi: 10.1021/acs.jproteome.3c00102.

[9] [9] D. T. Cervone, R. Moreno-Justicia, J. P. Quesada, and A. S. Deshmukh, "Mass spectrometry-based proteomics approaches to interrogate skeletal muscle adaptations to exercise," *Scandinavian Med Sci Sports*, p. sms.14334, Mar. 2023, doi: 10.1111/sms.14334.

[10] Y.-W. Au *et al.*, "Abstract 3974: Discovery and validation of therapeutic targets in immune cells by mass spectrometry-based proteomics," *Cancer Research*, vol. 83, no. 7_Supplement, pp. 3974–3974, Apr. 2023, doi: 10.1158/1538-7445.AM2023-3974.

[11] C. Punzalan, L. Wang, B. Bajrami, and X. Yao, "Measurement and utilization of the proteomic reactivity by mass spectrometry," *Mass Spectrometry Reviews*, p. mas.21837, Mar. 2023, doi: 10.1002/mas.21837.

[12] N. Takemori and A. Takemori, "In-depth structural proteomics integrating mass spectrometry and polyacrylamide gel electrophoresis," *Front. Anal. Sci.*, vol. 2, p. 1107183, Jan. 2023, doi: 10.3389/frans.2022.1107183.

[13] [13] M. D. P. Chantada-Vázquez, C. Núñez, and S. B. Bravo, "Quantitative proteomics by mass spectrometry in food science," in *Food Proteomics*, Elsevier, 2022, pp. 15–48. doi: 10.1016/B978-0-323-90889-4.00006-3.

[14] H. Webel *et al.*, "Mass spectrometry-based proteomics imputation using self supervised deep learning," Bioinformatics, preprint, Jan. 2023. doi: 10.1101/2023.01.12.523792.

[15] W. Albuquerque *et al.*, "Mass Spectrometry-Based Proteomic Profiling of a Silvaner White Wine," *Biomolecules*, vol. 13, no. 4, p. 650, Apr. 2023, doi: 10.3390/biom13040650.

[16] S. Parikh, R. Patel, D. Khunt, V. P. Chavda, and L. Vora, "Data Analytics and Data Visualization for the Pharmaceutical Industry," in *Bioinformatics Tools for Pharmaceutical Drug Product Development*, V. Chavda, K. Anand, and V. Apostolopoulos, Eds., 1st ed.Wiley, 2023, pp. 55–76. doi: 10.1002/9781119865728.ch4.

[17] Y. Yoshimura and N. Zaima, "Application of Mass Spectrometry Imaging for Visualizing Food Components," *Foods*, vol. 9, no. 5, p. 575, May 2020, doi: 10.3390/foods9050575.

[18] H. Aikawa *et al.*, "Abstract 350: Visualizing alectinib distribution in mice brain by using MALDI mass spectrometry imaging technique," *Cancer Research*, vol. 76, no. 14_Supplement, pp. 350–350, Jul. 2016, doi: 10.1158/1538-7445.AM2016-350.

[19] Y. Fujimura and D. Miura, "MALDI Mass Spectrometry Imaging for Visualizing In Situ Metabolism of Endogenous Metabolites and Dietary Phytochemicals," *Metabolites*, vol. 4, no. 2, pp. 319–346, May 2014, doi: 10.3390/metabo4020319.

[20] M. Lorenz *et al.*, "Co-Registered Application of Matrix Assisted Laser Desorption/Ionization Mass Spectrometry and Time-of-Flight Secondary Ion Mass Spectrometry Images for Visualizing Signaling Molecules," *Microsc Microanal*, vol. 25, no. S2, pp. 1170–1171, Aug. 2019, doi: 10.1017/S1431927619006585.

[21] M. Tognetti, N. Beaton, K. Sklodowski, R. Bruderer, and L. Reiter, "602 Mass spectrometry-based protein biomarker analysis in chemoimmunotherapy combinations identifies unique immune signatures in pancreatic cancer," *J Immunother Cancer*, vol. 10, no. Suppl 2, Nov. 2022, doi: 10.1136/jitc-2022-SITC2022.0602.

[22] C. Chen, J. Hou, J. J. Tanner, and J. Cheng, "Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis," *International Journal of Molecular Sciences*, vol. 21, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/ijms21082873.

[23] V. Kumar, S. Ray, S. Ghantasala, and S. Srivastava, "An Integrated Quantitative Proteomics Workflow for Cancer Biomarker Discovery and Validation in Plasma," *Front. Oncol.*, vol. 10, p. 543997, Sep. 2020, doi: 10.3389/fonc.2020.543997.

[24] M. Bantscheff, S. Lemeer, M. M. Savitski, and B. Kuster, "Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present," *Anal Bioanal Chem*, vol. 404, no. 4, pp. 939–965, Sep. 2012, doi: 10.1007/s00216-012-6203-4.

[25] S. A. Carr *et al.*, "Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-based Assay Development Using a Fit-for-Purpose Approach," *Molecular & Cellular Proteomics*, vol. 13, no. 3, pp. 907–917, Mar. 2014, doi: 10.1074/mcp.M113.036095.

[26] H. Wang, "Data Mining in Protein Identification by Tandem Mass Spectrometry:," in *Encyclopedia of Data Warehousing and Mining, Second Edition*, J. Wang, Ed., IGI Global, 2009, pp. 472–478. doi: 10.4018/978-1-60566-010-3.ch074.

[27] S. J. De Veer, J. Weidmann, and D. J. Craik, "Cyclotides as Tools in Chemical Biology," *Acc. Chem. Res.*, vol. 50, no. 7, pp. 1557–1565, Jul. 2017, doi: 10.1021/acs.accounts.7b00157.

[28] D. Paul, A. Kumar, A. Gajbhiye, M. K. Santra, and R. Srikanth, "Mass Spectrometry-Based Proteomics in Molecular Diagnostics: Discovery of Cancer Biomarkers Using Tissue Culture," *BioMed Research International*, vol. 2013, pp. 1–16, 2013, doi: 10.1155/2013/783131.

[29] J. K. Eng, T. A. Jahan, and M. R. Hoopmann, "Comet: An open-source MS/MS sequence database search tool," *Proteomics*, vol. 13, no. 1, pp. 22–24, Jan. 2013, doi: 10.1002/pmic.201200439.

[30] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, "Electrospray Ionization for Mass Spectrometry of Large Biomolecules," *Science*, vol. 246, no. 4926, pp. 64–71, Oct. 1989, doi: 10.1126/science.2675315.

[31] A. F. R. Hühmer *et al.*, "The Chromosome-Centric Human Proteome Project: A Call to Action," *J. Proteome Res.*, vol. 12, no. 1, pp. 28–32, Jan. 2013, doi: 10.1021/pr300933p.

[32] A. Matta, R. Ralhan, L. V. DeSouza, and K. W. M. Siu, "Mass spectrometry-based clinical proteomics: Head-and-neck cancer biomarkers and drug-targets discovery," *Mass Spectrom. Rev.*, vol. 29, no. 6, pp. 945–961, Nov. 2010, doi: 10.1002/mas.20296.

[33] M. Hossain *et al.*, "Enhanced Sensitivity for Selected Reaction Monitoring Mass Spectrometry-based Targeted Proteomics Using a Dual Stage Electrodynamic Ion Funnel Interface," *Molecular & Cellular Proteomics*, vol. 10, no. 2, pp. S1–S9, Feb. 2011, doi: 10.1074/mcp.M000062-MCP201.

[34] Y. Ishihama *et al.*, "Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein," *Molecular & Cellular Proteomics*, vol. 4, no. 9, pp. 1265–1272, Sep. 2005, doi: 10.1074/mcp.M500061-MCP200.

[35] J. D. Jaffe, D. R. Mani, K. C. Leptos, G. M. Church, M. A. Gillette, and S. A. Carr, "PEPPeR, a Platform for Experimental Proteomic Pattern Recognition," *Molecular & Cellular Proteomics*, vol. 5, no. 10, pp. 1927–1941, Oct. 2006, doi: 10.1074/mcp.M600222-MCP200.

[36] H. Keshishian *et al.*, "Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury," *Molecular & Cellular Proteomics*, vol. 14, no. 9, pp. 2375–2393, Sep. 2015, doi: 10.1074/mcp.M114.046813.

[37] A. I. Nesvizhskii, "Proteogenomics: concepts, applications and computational strategies," *Nat Methods*, vol. 11, no. 11, pp. 1114–1125, Nov. 2014, doi: 10.1038/nmeth.3144.

[38] L. Zamdborg *et al.*, "ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry," *Nucleic Acids Research*, vol. 35, no. Web Server, pp. W701–W706, May 2007, doi: 10.1093/nar/gkm371.

[39] W. W. Kallemeijn *et al.*, "Proteome-wide analysis of protein lipidation using chemical probes: in-gel fluorescence visualization, identification and quantification of N-myristoylation, N- and S-acylation, O-cholesterylation, S-farnesylation and S-geranylgeranylation," *Nat Protoc*, vol. 16, no. 11, pp. 5083–5122, Nov. 2021, doi: 10.1038/s41596-021-00601-6.

[40] O. T. Schubert, H. L. Röst, B. C. Collins, G. Rosenberger, and R. Aebersold, "Quantitative proteomics: challenges and opportunities in basic and applied research," *Nat Protoc*, vol. 12, no. 7, pp. 1289–1294, Jul. 2017, doi: 10.1038/nprot.2017.040.

[41] N. Singhal, M. Kumar, P. K. Kanaujia, and J. S. Virdi, "MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis," *Front. Microbiol.*, vol. 6, Aug. 2015, doi: 10.3389/fmicb.2015.00791.

[42] D. L. Tabb, C. G. Fernando, and M. C. Chambers, "MyriMatch: Highly Accurate Tandem Mass Spectral Peptide Identification by Multivariate Hypergeometric Analysis," *J. Proteome Res.*, vol. 6, no. 2, pp. 654–661, Feb. 2007, doi: 10.1021/pr0604054.

[43] S. S. Thakur *et al.*, "Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation," *Molecular & Cellular Proteomics*, vol. 10, no. 8, p. M110.003699, Aug. 2011, doi: 10.1074/mcp.M110.003699.

[44] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer Genome Landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, Mar. 2013, doi: 10.1126/science.1235122.

[45] X. Wang *et al.*, "Protein Identification Using Customized Protein Sequence Databases Derived from RNA-Seq Data," *J. Proteome Res.*, vol. 11, no. 2, pp. 1009–1017, Feb. 2012, doi: 10.1021/pr200766z.

[46] W. W. Wu, G. Wang, S. J. Baek, and R.-F. Shen, "Comparative Study of Three Proteomic Quantitative Methods, DIGE, cICAT, and iTRAQ, Using 2D Gel- or LC−MALDI TOF/TOF," *J. Proteome Res.*, vol. 5, no. 3, pp. 651–658, Mar. 2006, doi: 10.1021/pr050405o.

[47] Y. Zhang, Z. Wen, M. P. Washburn, and L. Florens, "Refinements to Label Free Proteome Quantitation:

How to Deal with Peptides Shared by Multiple Proteins," *Anal. Chem.*, vol. 82, no. 6, pp. 2272–2281, Mar. 2010, doi: 10.1021/ac9023999.

[48] S. Liu *et al.*, "TMT-Based Quantitative Proteomics Analysis Reveals Airborne PM2.5-Induced Pulmonary Fibrosis," *IJERPH*, vol. 16, no. 1, p. 98, Dec. 2018, doi: 10.3390/ijerph16010098.

[49] Y. Zhu *et al.*, "Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow," *Nat Commun*, vol. 9, no. 1, p. 903, Mar. 2018, doi: 10.1038/s41467-018-03311-y.

[50] B. L. Zybailov, L. Florens, and M. P. Washburn, "Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors," *Mol. BioSyst.*, vol. 3, no. 5, p. 354, 2007, doi: 10.1039/b701483j.

[51] B. Zybailov, A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens, and M. P. Washburn, "Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*," *J. Proteome Res.*, vol. 5, no. 9, pp. 2339–2347, Sep. 2006, doi: 10.1021/pr060161n.

[52] Q. Huang *et al.*, "Quantitative Proteomics Reveal the Protective Effects of the Combined Chinese Herbal Extracts Against Alzheimer's Disease via Modulating the Expression of Synapse-associated Proteins in Mouse Model," In Review, preprint, Oct. 2021. doi: 10.21203/rs.3.rs-956329/v1.

[53] H. F. Gilbert, "Protein Disulfide Isomerase and the Catalysis of Oxidative Protein Folding," in *Folding of Disulfide Proteins*, R. J. Y. Chang and S. Ventura, Eds., New York, NY: Springer New York, 2011, pp. 133–149. doi: 10.1007/978-1-4419-7273-6_7.

[54] G. Zhang *et al.*, "Identifying Membrane Protein-Lipid Interactions with Lipidomic Lipid Exchange-Mass Spectrometry," Biophysics, preprint, Jun. 2023. doi: 10.1101/2023.06.02.543293.

[55] P. He and G. R. Moran, "Structural and mechanistic comparisons of the metal-binding members of the vicinal oxygen chelate (VOC) superfamily," *Journal of Inorganic Biochemistry*, vol. 105, no. 10, pp. 1259–1272, Oct. 2011, doi: 10.1016/j.jinorgbio.2011.06.006.

[56] E. M. Templeton et al., "Comparison of SPEED, S-Trap, and In-Solution-Based Sample Preparation Methods for Mass Spectrometry in Kidney Tissue and Plasma," IJMS, vol. 24, no. 7, p. 6290, Mar. 2023, doi: 10.3390/ijms24076290.

[57] Yadav, R. ., & Singh, R. . (2023). A Hyper-parameter Tuning based Novel Model for Prediction of Software Maintainability. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 106–113. https://doi.org/10.17762/ijritcc.v11i2.6134

[58] Muhammad Khan, Machine Learning for Predictive Maintenance in Manufacturing: A Case Study , Machine Learning Applications Conference Proceedings, Vol 1 2021.

[59] Kothandaraman, D., Praveena, N., Varadarajkumar, K., Madhav Rao, B., Dhabliya, D., Satla, S., & Abera, W. (2022). Intelligent forecasting of air quality and pollution prediction using machine learning. Adsorption Science and Technology, 2022 doi:10.1155/2022/5086622