Comprehensive Single-Cell Proteomics Analysis of Cancer Cell-Line Model Systems

By
Ibrahim ElZahaby

Laboratory of Systems & Synthetic Biology

17th of May 2023

Supervisor 1: Cristina Furlan
Supervisor 2: Edoardo Saccenti

Abstract

The ability of phenotypic characterization through proteomics is vital at the single-cell level to enhance our understanding of the complexity of biological systems. while the traditional bulk approaches facilitated proteomic profiling from given samples, still some cell features are masked, such as heterogenicity. Single-cell omics technologies, particularly in single-cell transcriptomics, enabled a deeper understanding of cellular mechanisms in various biological phenomena. Despite the recent advancements in scRNA-seq, there is a poor correlation between the mRNA and protein levels, indicating an unreliable indicator when profiling the protein expression. Single-cell proteomics is one of the most promising approaches to unfold biomarkers and regulatory pathways, characterize the heterogeneity in cell populations, distinguish between healthy and diseased tissues, and unravel relationships between lineages. Computational studies on the single-cell proteomics level are poorly investigated. This report will comprehensively analyse Melanoma, Hela, Pancreatic Ductal Adenocarcinoma (PDAC), and U937 Monocytes cell line model systems. Gene set enrichment analysis, such as GO and KEGG pathways, were performed to observe the main characteristics of each cell type. Additionally, network analysis was constructed, and hub proteins were investigated. We will highlight the variation between cell groups found in the UMAP. Also, discussing the enrichment results that showed biomarkers between our cancer cell types and COVID-19.

Dedication & Acknowledgements

I would like to express my sincere gratitude to my supervisors, Cristina Furlan and Edoardo Saccenti, for their invaluable guidance and support throughout the process of completing my MSc thesis. Their patience and encouragement have been fundamental in shaping my academic pursuits. I am deeply grateful for their constructive feedback, insightful comments and their dedication that positively reflected on my personal and professional growth.

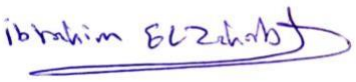
I warmly thank my study advisor, Robert Smith, for his tremendous effort and helpful advice in reaching this point.

I am also grateful for the wonderful working environment, Systems and Synthetic Biology (SSB) chair group, which made my research experience enjoyable and fulfilling. I enjoyed my time there and the research activities that helped shape my academic mindset.

I would also like to give special thanks to my family, Father (Mohsen), Mother (Nagwa), and Sister (Eman), for their continuous support and belief in my capabilities.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic awards. Except where indicated by specific references in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others is indicated as such. Any views expressed in this dissertation are those of the author.

Signed: 

Date: 17th of May 2023

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1- Phenotypic profiling through proteomics

The importance of molecules in supporting life has been recognized since the early stages of biological research, with Berzelius in 1838 coining the definition of "protein," which derived from the Greek word "proteios", meaning "the first rank" [1]. The "proteome" refers to the entire protein content of a cell, characterized by its localization, interactions, post-translational modifications, and turnover at a specific time [2]. Proteomics, coined by Marc Wilkins in 1996, is the study of the protein complement of a genome, providing valuable insights about genes. While the proteome of eukaryotic cells is complex, exhibiting extensive dynamic range, prokaryotic proteins play a significant role in pathogenic mechanisms, presenting challenges in the analysis due to their variations [3].

Proteomics has a critical role in early disease detection, prognosis, and drug development. It involves the characterization of the proteome, including the expression, structure, functions, interactions, and modifications of proteins at any level [4]. The proteome is dynamic and varies from cell to cell, fluctuating in response to external stimuli. In eukaryotic cells, proteomics is complex due to various post-translational modifications that occur at different sites [5]. Even though transcriptome analysis can reveal fluctuations in gene expression levels, it cannot directly measure increased mRNA synthesis [6]. Proteins, as biological effectors, are crucial to understanding a biological system, as their levels depend not only on mRNA levels but also on translational control and regulation. Therefore, proteomics is a more relevant technology for characterizing biological entities [7].

1.2- Quantitative proteomic characterization using bulk analysis

Various technologies have been implemented to quantify the total amount of proteins from the phenotype of interest in light of proteomics bulk analysis approaches. Therefore, we will expand our understanding by providing some of the MS and Non-MS quantification techniques below.

1.2.1- Non-MS techniques

Enzyme-Linked Immunosorbent Assay (ELISA) is an immunoassay widely used for diagnostic purposes since its first publication by Engvall and Pearlmann in 1971. The assay utilizes the antigen or antibodies on a solid surface and the addition of enzyme-conjugated antibodies to measure fluctuations in enzyme activities proportional to antibody and antigen concentration in the biological specimen [10]. Another quantification technique was Western blotting which separates proteins using electrophoresis and detects low-abundance proteins using enzyme-conjugated antibodies [9]. It has been used to diagnose infectious diseases such as Herpes Simplex Virus type 2 and visceral leishmaniasis [12].

One of the high throughput techniques is Nuclear Magnetic Resonance (NMR) which characterizes protein dynamics at the atomic resolution [55]. NMR determines protein structure and behaviour through various phases, sample preparation, measurements, and interpretive approaches. This information is essential in several research areas, including structure-based drug design and functional genomics [13].

1.2.2- Mass-Spectrometry (MS) techniques

To overcome the limitations of non-MS quantification approaches, such as narrow dynamic range, low sensitivity, and tedious costs of establishing high throughput technologies, Mass-Spectrometry (MS) is considered one of the most well-suited technologies that are utilized in various fields, such as clinical laboratories, virus research, plant science, and cancer diagnosis. MS is a technique used to determine the mass-to-charge ratio (m/z) of molecules and is particularly useful in determining the molecular weight of proteins. The process involves three steps; transforming molecules into gas-phase ions, separating ions based on their m/z values using electric or magnetic fields in a mass analyzer and measuring the separated ions and the amount of each species with a particular m/z value. Different ionization methods are used, including matrix-assisted laser desorption ionization (MALDI), surface-enhanced laser desorption/ionization (SELDI), and electrospray ionization (ESI) [8].

MS has been used in cancer diagnosis to characterize blood proteins such as IBP2, IBP3, IGF1, IGF2, and A2GL, which have been proposed as biomarkers for breast cancer diagnosis [14]. Additionally, MS has been used to analyze human serum for PSA, human growth hormone, and interleukin-12 and to detect the most abundant proteins of tomato xylem sap after *Fusarium oxysporum* infection using mass spectrometric sequencing and peptide mass fingerprinting [15]. Moreover, imaging MALDI mass spectrometry has been used to analyze whole-body tissues and the distribution of drugs and metabolites following drug administration, providing deeper insight into toxicological and therapeutic processes [16]. However, MS bulk-based methods can only capture the average of the protein expression profile, which leads to hiding cellular characteristics, masking biological features, and the inability to discover cellular heterogenicity.

1.3- Disadvantages of Bulk analysis approaches

Cell-to-cell variation is a fundamental characteristic of multi-cellular organisms that exhibit diverse cell types with distinct functions, morphology, and gene expression patterns. Also, various cells with different characteristics are present within seemingly uniform tissues. However, studying cellular populations or tissues is often limited because conventional methods of analysis that measure bulk tissue need to provide a clear picture of the intrinsic cellular heterogeneity that can mask essential information [17]. While individual tumour genetics can be precisely examined, studying bulk expression profiles provides limited information due to

averaging phenotypic determinants of cancer programs, tumour microenvironment (TME) influences, and genetic heterogeneity within the tumour [18].

Additionally, bulk transcriptomic analysis approaches cannot capture the significant heterogeneity within a cancer tissue, and valuable information related to sub-cell populations can be lost. Moreover, traditional methods for analysing bulk-level measurements are ineffective since they cannot differentiate between changes in the proportion of cell types and changes in the molecular program within a cell type [20]. However, single-cell analysis has been introduced as a robust tool for dissecting intratumorally heterogeneity. It can provide an unprecedented molecular resolution, allowing for generating a new disease taxonomy [19].

1.4- The Era of Single-Cell Omics

Recent years have witnessed the emergence of technologies that enable the acquisition of genome-scale molecular information at the level of single cells, offering unparalleled opportunities for systematic exploration of cellular heterogeneity across multiple molecular layers, including DNA [21], RNA [22], proteins [23], and metabolites [24]. These innovations have enabled the identification of new cell types and previously unknown biomarkers and facilitated the prediction of cellular developmental trajectories [25].

Single-cell analysis has revolutionized our understanding of cell differentiation and the mechanisms underlying disease development and progression [28]. thus, to expand the catalogue of mammalian cell states and identities, single-cell analysis has challenged the traditional view of cell-fate determination and identified new approaches to understand disease mechanisms better [27]. Single-cell DNA sequencing has revealed significant cellular heterogeneity within tumours and revised clonal evolution models. Furthermore, single-cell RNA sequencing has provided insights into the role of tumour microenvironments in disease progression and drug resistance [26].

In addition, single-cell analysis has been crucial in refining our understanding of cell differentiation [30]. It has been observed in several studies that megakaryocytes emerge at a high level, closer to that of hematopoietic stem cells, challenging the conventional model that placed megakaryocytic lineage late in the differentiation process [31]. The transcriptomic patterns of cell states are continuous rather than forming distinct groups, making cell annotation challenging and calling for significant revisions to current models of cell lineage hierarchy [29].

The comprehensive understanding of the complexity of cells in multi-cellular organisms demands improved experimental and computational methods that can derive meaningful insights from large-scale and diverse single-cell data [32]. Various challenges need to be addressed for the advancement of single-cell analysis. Recent papers have discussed these challenges [33], [34]. In

the following part, we will expand on these discussions to focus on current advancements in this field.

1.5- Single-Cell Transcriptomics

Single-cell RNA sequencing (scRNA-seq) is a primarily utilized technology for single-cell sequencing that has seen significant advancements in recent years, with various technologies for sensitive, highly multiplexed, or combinatorically barcoded profiling [44], [45], [46]. These advancements are complemented by other single-cell genomic, epigenomic, and proteomic profiling technologies, such as methods for measuring genome sequence, chromatin accessibility, DNA methylation, cell surface proteins, small RNAs, histone modifications, and chromosomal conformation [47], [48]. Additionally, recent efforts have been made to accurately develop methods to capture spatial or lineage information in single-cell studies [41], [42], [43].
Tumor tissue is a highly complex biological system encompassing cancer cells and the surrounding microenvironment. Single-cell transcriptomics technologies have emerged as powerful tools for evaluating the therapeutic effects of cancer immunotherapy and identifying novel therapeutic targets by systematically assessing tumor tissue and its microenvironment [36]. The heterogeneity of gene expression involved in immune response, surveillance, and escape can significantly impact the clinical outcomes of immunotherapy [37].

Recent studies have demonstrated the potential of single-cell RNA-seq in unravelling the genetic heterogeneity of lung cancer and its immune-related pathways, such as the cell cycle and antigen-presenting pathways. For example, neutrophils in the microenvironment have been found to mediate metastatic outgrowth stimulated by IL11-expressing minor subclones, providing a novel target for immunotherapy [38]. Moreover, it can also guide future practices in overcoming emerging treatment resistance, such as the infiltration of immuno-suppressive immature myeloid cells (IMCs) in resistant tumours that reduce the efficacy of combinatorial therapy [39].

ScRNA-seq can also be used to characterize immune cell populations, predict tumour markers, and develop novel immunotherapy methods. For instance, CyTOF, combined with scRNA-seq and histological multiple imaging techniques, has comprehensively assessed the specific immune responses of natural killer cells and myeloid cells in tumour tissue, normal lung tissue, and peripheral tissues, blood mononuclear cells of lung cancer patients. Reduction of CD8+ granzyme B+/CD39hiCD38hiPD-1hiCTLA4hiFoxp3hi regulatory T cells has been identified as a potential biomarker for lung cancer and can guide clinical treatment [40]. The findings from these studies highlight the potential of single-cell sequencing technologies in providing a better understanding of the tumour microenvironment and in developing personalized immunotherapies for cancer patients [35].

Bulk transcriptional analysis of tumour samples has revealed transcriptional variation that affects prognosis but is limited by its inability to identify small subsets of tumour tissues [52]. Single-cell RNA sequencing (scRNA-seq) is a revolutionary method widely used to characterise molecular properties and dissect cellular compositions of complex tissues [53]. scRNA-seq could help identify novel markers, rare subgroups, and evolutionary patterns, particularly in brain development. Despite its potential, the full range of differences between individual cells in tumour tissues and adjacent normal tissue has yet to be estimated using scRNA-seq [54].

1.6- Single-Cell proteomics Analysis

The human genome comprises around 20,000 protein-coding genes, but each protein can be translated from different splice variants and have post-translational modifications (PTMs), such as phosphorylation or glycosylation. This means the number of unique proteins comprising the proteome can reach billions [49]. However, not all proteins are expressed simultaneously in cells; in cancer cells, only around 12,000 proteins have been detected [51]. Several techniques have been developed that enable the analysis of proteins at a single-cell resolution, and these assays can be categorised based on the protein detection principle [50]. These techniques will be discussed in more detail below.

1.6.1- Non-MS quantification techniques (Single-cell-based)

The Heath group has developed the Single-Cell Barcode Chips (SCBC) method to enable comprehensive protein analysis in single cells. SCBC involves compartmentalizing individual cells into microchambers and applying an antibody barcode array in each chamber to capture secreted proteins from single cells. By staining the charged proteins with corresponding biotinylated antibodies and fluorescent streptavidin, the abundances of the captured proteins can be determined by comparing the fluorescent signals generated with the calibration curve. Mass spectrometry can directly identify and quantify the charged proteins [115]. The SCBC method allows the analysis of up to 20 different proteins secreted from a single cell. It enables the simultaneous quantification of secreted, membrane, and cytoplasmic proteins and metabolites from the same cell. SCBC has advantages over other single-cell proteomic technologies in that it can detect and analyse secreted proteins from live cells. However, the assay's multiplex capacity and detection sensitivity must be improved. Still, it leads to the enlarged volume of the chamber and the reduced concentrations of the secreted proteins, thus sacrificing the detection sensitivity of the assay.

In 2018, a new technique called multiplexed in situ targeting (MIST) was developed for single-cell secreted protein analysis [116]. This method uses a glass coverslip coated with a monolayer of microbeads encoded with single-stranded DNA. The microbeads are hybridised with antibodies conjugated with complementary DNA strands, which enables them to recognise specific protein

targets secreted from single cells. After the protein targets are detected, the locations of the microbeads are recorded, and the antibodies are washed off. The identity of the protein targets is then determined through successive cycles of hybridisation and DE hybridization of fluorescent cDNAs, which produce a unique encrypted code of ordered fluorescence colour sequence. This approach can detect tens to hundreds of proteins secreted from single cells. While MIST has a high multiplexing capacity, it must compromise its sensitivity to achieve it. Increasing the number of microbeads or decreasing their size will enhance the multiplexing capability of the assay but will also reduce the sensitivity. Additionally, the potential for cross-hybridization between the different ssDNA on microbeads and decoding probes could also limit the multiplexing capacity of this platform. Overall, MIST provides a promising approach for single-cell secreted protein analysis, but further improvements are needed to increase its sensitivity and capability.

The Immuno-SABER technique is a highly sensitive in situ proteomics approach developed in 2019 by Saka and his colleagues [117]. This approach involves staining protein targets with DNA-tagged antibodies, which are then hybridized to DNA concatemers generated by a primer exchange reaction. Multiple fluorescent oligonucleotides are subsequently hybridized to the binding sites on the DNA concatemers, amplifying the signal and increasing the detection sensitivity by up to 180-fold. After imaging, the fluorescent oligonucleotides are DE hybridized, and new ones are used to stain other protein targets, allowing for sequential imaging of multiple targets.

CyTOF is a novel technique that uses inductively coupled plasma time-of-flight mass spectrometry and max par metal-labelling technology to detect proteins at the single-cell level [56]. With its ability to measure over 40 markers per cell using 135 detection channels, CyTOF provides unprecedented resolution and high multiplicity of biomarker detection, absolute quantification, and simplified measurement protocols [57]. Unlike traditional flow cytometry, CyTOF has no detection channel overlap or sample matrix effects, making it a more efficient and accurate method [59]. While it has been successfully applied in human leukemia studies, its potential in the GC field has yet to be explored [58].

1.6.2- MS-based single-cell quantification approaches

Mass spectrometry is one of the promising approaches to identifying a large number of proteins in single cells to facilitate a wide-scale proteome analysis on a huge number of cells [60]. This section will explicate the latest single-cell proteomics quantification methods using mass spectrometry, also known as “SCoPE” protocols.

In 2018, Nikolai Slavov and his colleagues improved a high-efficient technology for Single Cell ProtEomics by Mass Spectrometry (SCoPE-MS). This method has eliminated the difficulties of LC-MS/MS, particularly minimizing the proteome loss in the MS machine by lysing the single cells using sonication and the carrier cells that increase the identified proteins in single cells. Also, it

jointly characterizes and measures the peptides through the innovative use of tandem mass tags (TMT) which enhance the multiplexing capacity. SCoPE-MS showed successful results in terms of cell clustering in single mammalian cells and identifying the covariant proteins during embryonic stem cell differentiation at three-time points [65].

SCoPE2 protocol was introduced as a second version of SCoPE-MS that avoids the limitation of the identified proteins in single cells and maximize the efficiency with a reasonable cost and time through utilizing the Minimal ProteOmic sample Preparation (mPOP) method with the use of isobaric carriers. With these features, SCoPE2 significantly increased the quantified peptides with decent quality. It can identify over 3040 proteins in over 1400 cells across two single-cell types, macrophage-like cells and monocytic cell lines (U937). In addition, it precisely classifies the cell types based on their differential protein expression levels and estimates the fold changes at the single-cell level [66].

The third proteomics platform developed in the Slavov laboratory is PlexDIA. This technology is meant to increase the depth and the quantification efficiency in small single-cell samples. Also, it raises the data accuracy, minimizing the sample losses and increasing the multiplexing capacity for the single-cell samples and peptides with the help of DIA-NN as a deep learning algorithm to quantify the peptides with a high-confidence level through computing the false positives using a decoy channel-based. This method was used to identify approximately 1000 proteins in two human single-cell samples, Jurket cells and monocytes cell lines (U937), with 98% accuracy. Moreover, it could not identify 643 differentially abundant proteins and analyze the important proteins involved in the cell division cycle (CDC) [67].

To increase the accuracy and symmetry of the identified proteins, pSCoPE was introduced. This protocol was made up to prioritize the analyzed peptides to raise the proteome depth. This strategy succeeded when implemented on pancreatic ductal adenocarcinoma (PDAC) and achieved up to 75% of the measured proteins per (PDAC) single-cell. Also, it was enabled to analyze the enriched protein groups in bone-marrow-derived-macrophage (BMDM) cells to investigate the endocytic activity [68]. Furthermore, pSCoPE was applied to characterize the correlated proteins that differ across the cell division phases in monocytes and melanoma single-cells. Eventually, pSCoPE investigated the protein variations between melanoma subpopulations during cell division [113].

Single-cell proteomics sequencing is anticipated to revolutionize the field of investigative pathology, especially when integrated with other multi-omics platforms, such as transcriptomics and proteomics, for the comprehensive analysis of cellular processes at the single-cell level [61].

1.7- Research Gap and Aim of the thesis project

Despite its numerous advantages, single-cell transcriptomics technology has some limitations that must be considered. One of the significant challenges is the sparse data obtained for particular molecular layers, which can impede the correlation analysis with other layers. Moreover, messenger RNA (mRNA) may not be a reliable indicator of protein levels. This has spurred the development of new techniques for the quantitative analysis of the proteome. Early studies have shown that mRNA levels correlate poorly with protein levels for all but the most abundant proteins [62]. Additionally, The field of single-cell proteomics is relatively new and has primarily focused on developing experimental methods, resulting in a need for more published literature on computational approaches for data analysis [63]. Regarding these limitations, this thesis is bridging the gap of the analysis tools by taking advantage of the available proteomics datasets from Slavov's lab and accessing the biological characteristics at the single-cell level through computational techniques.

This report comprehensively investigated four cancer cell-line model systems: Melanoma cell-line, Hela cells, Pancreatic Ductal Adenocarcinoma (PDAC), and Monocytes myeloid leukaemia (U937). We identified the distinct sub-groups through UMAP and PCA that helped us to differentiate between the sub-cellular populations between the same cell type that exhibit distinct proteomic profiles. GO enrichment analysis revealed biological insights into the cellular pathways such as RNA splicing, ribonucleoprotein complex biogenesis, and cytoplasmic translation among cancer cell types and between specific cell subpopulation analysis. Thus, KEGG pathway analysis showed shared protein biomarkers between cancer and COVID-19, indicating a possible relationship between coronavirus disease and cancer. With the help of biological network analysis, we identified hub-proteins signalling pathways and the protein complexes that have biological roles in cell communication.

Chapter 2: Methods

2.1- Data Acquisition

The single-cell proteomics datasets were acquired from the Slavov laboratory [64]. The lab innovated various methods to identify the proteins at the level of single cells using different Mass Spectrometry MS-based approaches such as SCoPE-MS [65], SCoPE2-MS [66], PlexDIA [67], and pSCoPE [68]. Four cell-line model systems have been used to be computationally analyzed. Because of their authenticity, the selected cell lines helped identify the biological characteristics that will be helpful in the era of cancer biology and biomedical research. More information about the single-cell data can be found in (Appendix 1).

2.2- Proteomics Data Collection

Four cancer cell-line model systems were used for the analysis. these were Monocytes (U937) that originally derived from histiocytic lymphoma [113], WM989-A6-G3 cells which is the Melanoma cell line [67], [HPAF-II cells (ATCC CRL-1997), CFPAC-I cells (ATCC CRL-1918), BxPC-3 cells (ATCC CRL-1687)]. These cells are derived from Pancreatic Ductal Adeno Carcinoma (PDAC) [68], and Hela cells are derived from cervical cancer [114]. An R package [69], a Venn diagram was used to extract the proteins of interest for the analysis. These proteins were selected from the overlapping proteins among the different protein quantification experiments used for the computational analysis. See, (Table 1).

Table 1: Single-Cell proteomic datasets that provide an overview of the cell types, number of experiments, and the overlapping proteins among different experiments.

Cell Type	No. Experiments	Intersected Proteins
Melanoma cell-line	**2**	**1083 (35.3%)**
Hela cells	**2**	**912 (36.5%)**
Pancreatic Ductal Adenocarcinoma (PDAC)	**2**	**961 (48%)**
Monocytes (U-937)	**5**	**584 (11.1%)**

2.3- Proteomics Data Processing

All proteomics datasets were formatted and stored in Excel files (.xlsx). Each dataset includes the rows (protein accessions) and columns (cell ID) and the protein abundance level in every cell. Furthermore, they were imputed, batch corrected, and the FDR is 1%.

2.4- Proteomics Identifiers Standardization

The protein identifiers used in the datasets were standardized to the UniProt Database [70], and the alias protein accessions have been removed. The unique accessions were only kept in each dataset.

2.5- Gene Identifiers Standardization

SYNGO “Synaptic Gene Ontologies " has generated all gene identifiers, " an online tool has been used to convert the protein accessions to HGNC gene symbols and names [71]. Using default settings, the output file from SYNGO was an Excel format containing the gene list, including Name, the symbol, the Alias, the tax id, the Entrez gene, and the protein query.

2.6- Statistical Analysis

2.6.1- Data transformation and Multiple Testing Correction

The protein abundance levels were log-transformed before the analysis. After that, the multiple testing correction was applied, where the p-values for genes and proteins were adjusted using the Fisher exact test for multiple testing hypothesis [72]. The differentially expressed genes and abundant proteins with a p-value cutoff >= 0.05 and a q-Value cutoff >= 0.03 were considered significant.

2.6.2- Multivariate Analysis

With the help of the R packages (FactoMineR) [73] and (factoextra) [75], Principal Component Analysis (PCA) was used to explore the high dimensions of the proteomics datasets on the cellular and sub-cellular levels to investigate the variation between the groups among cell types and within each cell type. PCA was applied to explore the datasets, GO terms analysis, and KEGG pathway to find the unique GO terms and biological pathways to distinguish between each cell type. Proteomics datasets were scaled to unit variance before the projection.

2.7- Cell and Sub-Cell Clustering

To cluster cancer cell types, Uniform Manifold Approximation and Projection (UMAP) is an unsupervised learning algorithm for dimensionality reduction and R package [74] that has been utilized to distinguish between cells and sub-cell populations based on their abundance levels.

2.8- Gene Set Enrichment Analysis (GSEA) for cell types

Other R packages, clusterprofiler [76] and enrichplot [77] were used for GSEA. Cluster profiler was used for GO enrichment analysis using the generated file from SYNGO entrezgene; the ontology term was Biological Process (BP), the organism database was homo sapiens with p-value cutoff = 0.05, and q-value cutoff = 0.03. In addition, it is used for the KEGG pathway with q-value

cutoff = 0.03 and default settings. Enrich plot was utilized to visualize the enrichment results from cluster profiler outputs such as Dot plots, Enrichment Maps, and Gene-Concept-Based Networks (Cnet plots).

2.9- Protein Set Enrichment Analysis (PSEA) for cell types and subpopulations

To differentiate between the sub-cell populations, the STRING database for functional protein annotation and association networks [78] for GO annotation using protein abundance levels in each sub-cellular cluster.

2.10- Network Analysis

To find the positive and negative correlated proteins, the corrplot package [79] uses Spearman correlation with default settings. Anvis (R package) was utilized for constructing the networks and exploring hub proteins.

2.11- Software

The thesis project code is available through the WUR-SSB GitLab link (https://gitlab.com/wurssb/StudentProjects/cancer-omics/-/tree/master) or on GitHub (https://github.com/IbrahimElzahaby/Single-Cell-Proteomics-MSc-Thesis).

Chapter 3: Results

3.1- The overlapped proteins were selected for analysis

In (Figure 1), the Venn diagram showed that 48% of the total proteins overlapped in PDAC datasets which were 961 proteins, followed by 36.5% in the Hela cells datasets, which equal 912 proteins. In comparison, Melanoma datasets contain the largest overlapping protein counts, 1083 representing 35.3%. Monocyte (U937) cells contain the least overlapping protein counts, 584, representing 11.1% among 5 datasets.

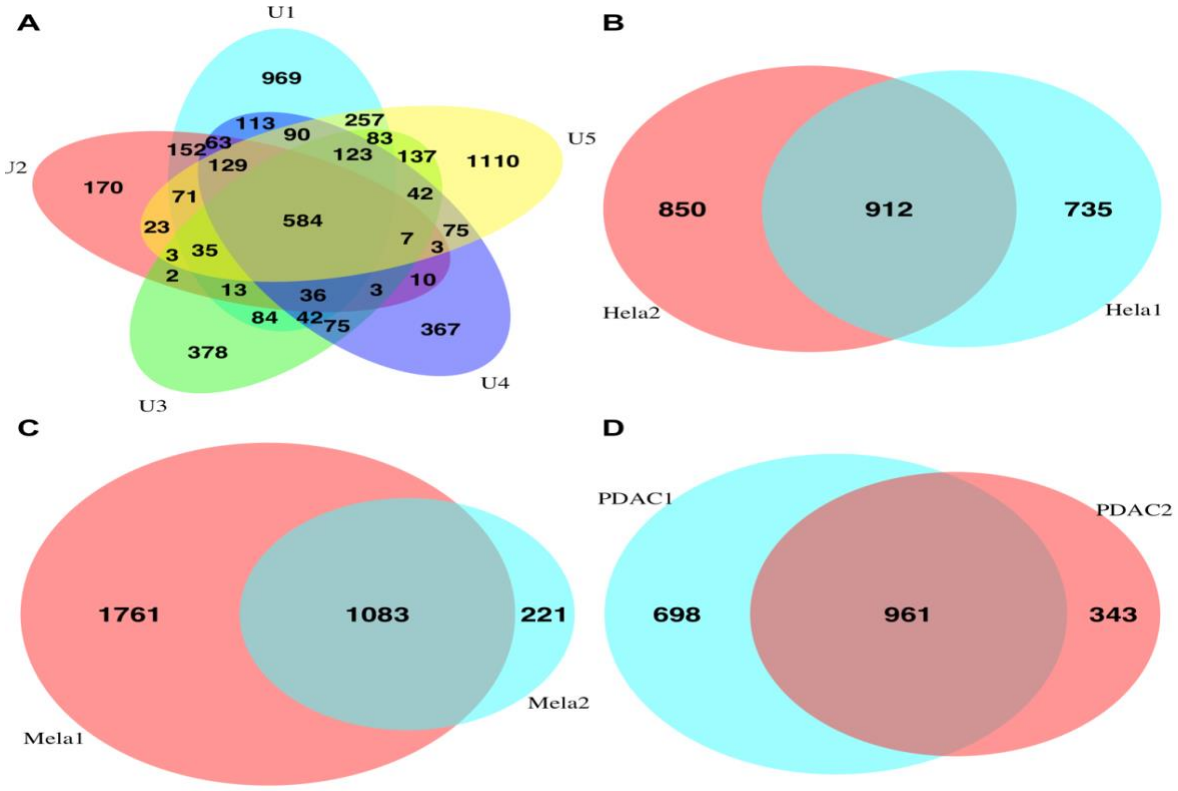


Figure 1:*Venn diagrams illustrate the overlapped proteins between different protein quantification experiments among four cancer cell line model systems.* ***(A)*** *Venn diagram shows 584 overlapped proteins among five protein quantification experiments in the Monocyte cell line (U937).* ***(B)*** *912 overlapped proteins among two protein quantification experiments in the Hela cells.* ***(C)*** *represents the overlapped proteins in the Melanoma cell line.* ***(D)*** *Venn diagram shows the overlapped proteins in the Pancreatic Ductal Adenocarcinoma (PDAC).*

3.2- Gene Set Enrichment Analysis (GSEA) revealed gene connectivity among cell types

We performed Gene Ontology (GO) and KEGG pathway enrichment analyses. GO enrichment analysis focused on the "Biological Process" ontology term. The PDAC cell line has the highest count of 698 genes associated with enriched terms, followed by 676 gene counts enriched in Melanoma. In contrast, Monocytes and Hela cells have 552, and 527 gene counts enriched respectively, while the Monocytes have the highest percentage, 95% of enriched terms out of the total gene counts 584. More details about the enrichment data are summarized in (Appendix 2). In addition, we performed the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis among all cancer cell types. Melanoma genes have the highest enriched counts (54) followed by PDAC genes found in 44 pathways in the KEGG database. While Hela and Monocytes genes were found in 35 and 34 pathways respectively. A summary of the enriched terms and pathways per cell type can be found in (Figure 2).

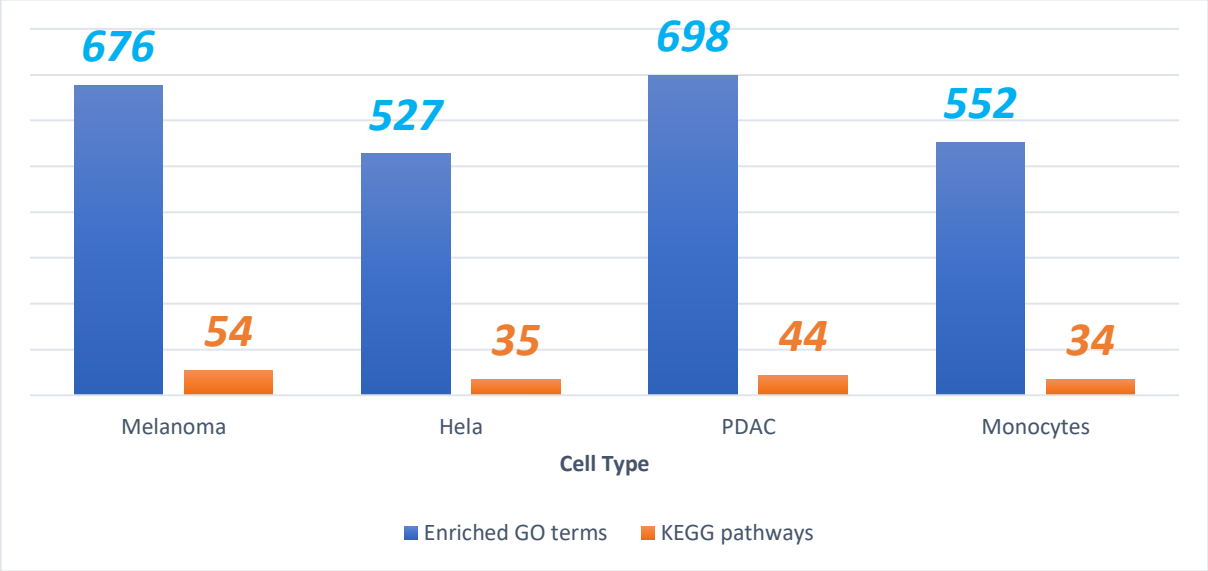


Figure 2: Bar Chart shows the enriched terms and KEGG pathways among four cancer cell types (Melanoma, Hela, PDAC, and Monocytes). The blue bars indicate the enriched GO terms, while the orange bars show the KEGG pathways associated with the gene lists in all cancer types. The statistical significance level was (padj < 0.05) and (q-val < 0.03).

3.2.1- *Translation processes were abundantly enriched in all cancer cell types*

2453 genes were subjected to Gene Ontology (GO) enrichment analysis among all cancer cell types focused on biological processes (BP) ontology terms to explore their biological roles and cellular characteristics. the dot plot shows that all cell types almost share the same enriched terms though at a different significance level as expressed by the p-value. Most of the gene counts that were significantly enriched in Monocytes (Figure 3A), Hela (Figure 3B), Melanoma (Figure 3C), and PDAC (Figure 3D) cell types were found in the transcription processes such as “RNA splicing” and “mRNA splicing via spliceosome” and also in the translation processes such as “Cytoplasmic translation”, “Translation initiation”, and “Ribonucleoprotein complex biogenesis”. Hela, Melanoma, and PDAC have the highest gene counts of the “Ribonucleoprotein complex biogenesis” term, while the “Cytoplasmic translation” has the highest gene counts in monocyte cells.

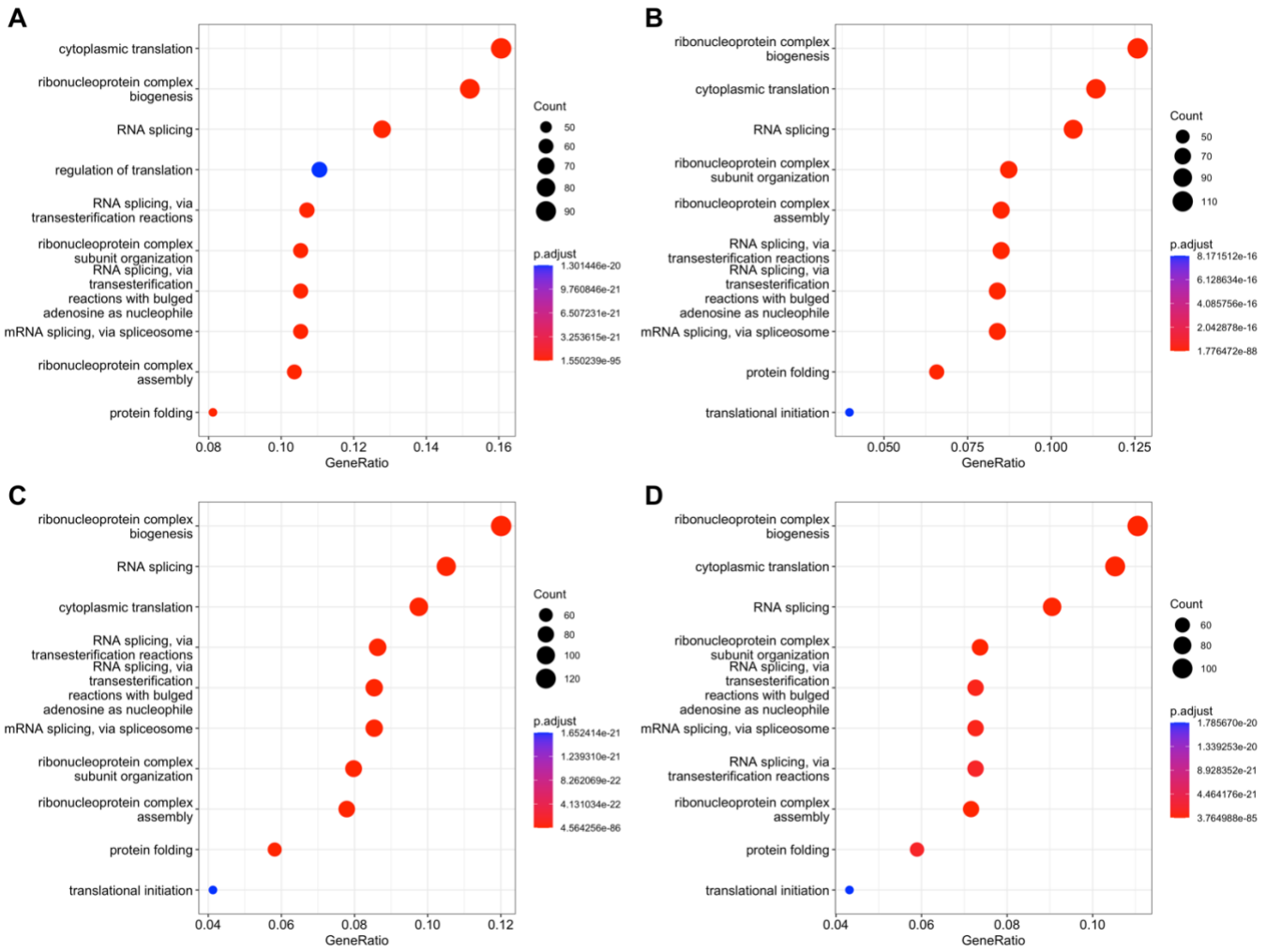


Figure 3: Dot plots provide the most 10 significant Gene Ontology (GO) enriched terms (padj < 0.05) in four cancer cell types. (A) indicates GO terms for Monocytes enriched terms, (B) shows Hela GO terms, (C) illustrates Melanoma enriched terms, and (D) indicates PDAC GO terms. The color dots indicate the significance level of each term, and the circle size indicates the gene counts. X-axis represents the gene ratio, and Y-axis represents the “biological process” enriched terms.

Additionally, we established an enrichment map (E-Map) for the top 10 significant GO terms as shown in (Figure 4) to investigate the connectivity of the shared genes among various biological processes. The transcription and translation processes, in particular “RNA splicing”, “mRNA splicing via spliceosome”, and “Ribonucleoprotein complex assembly & subunit organization” GO terms, have the highest connectivity among all cancer cell types. In contrast, the “Regulation of translation” term found in the Monocyte gene list (Figure 4A) and “Translational initiation” found in Hela (Figure 4B), Melanoma (Figure 4C), and PDAC (Figure 4D) gene lists have the least connectivity with other biological processes. Despite the “Cytoplasmic translation” GO term having the highest gene counts, its associated genes have limited connectivity with other biological processes’ genes. Moreover, the “Protein folding” term has no connected genes with any other biological processes in the enrichment map among all cell types.

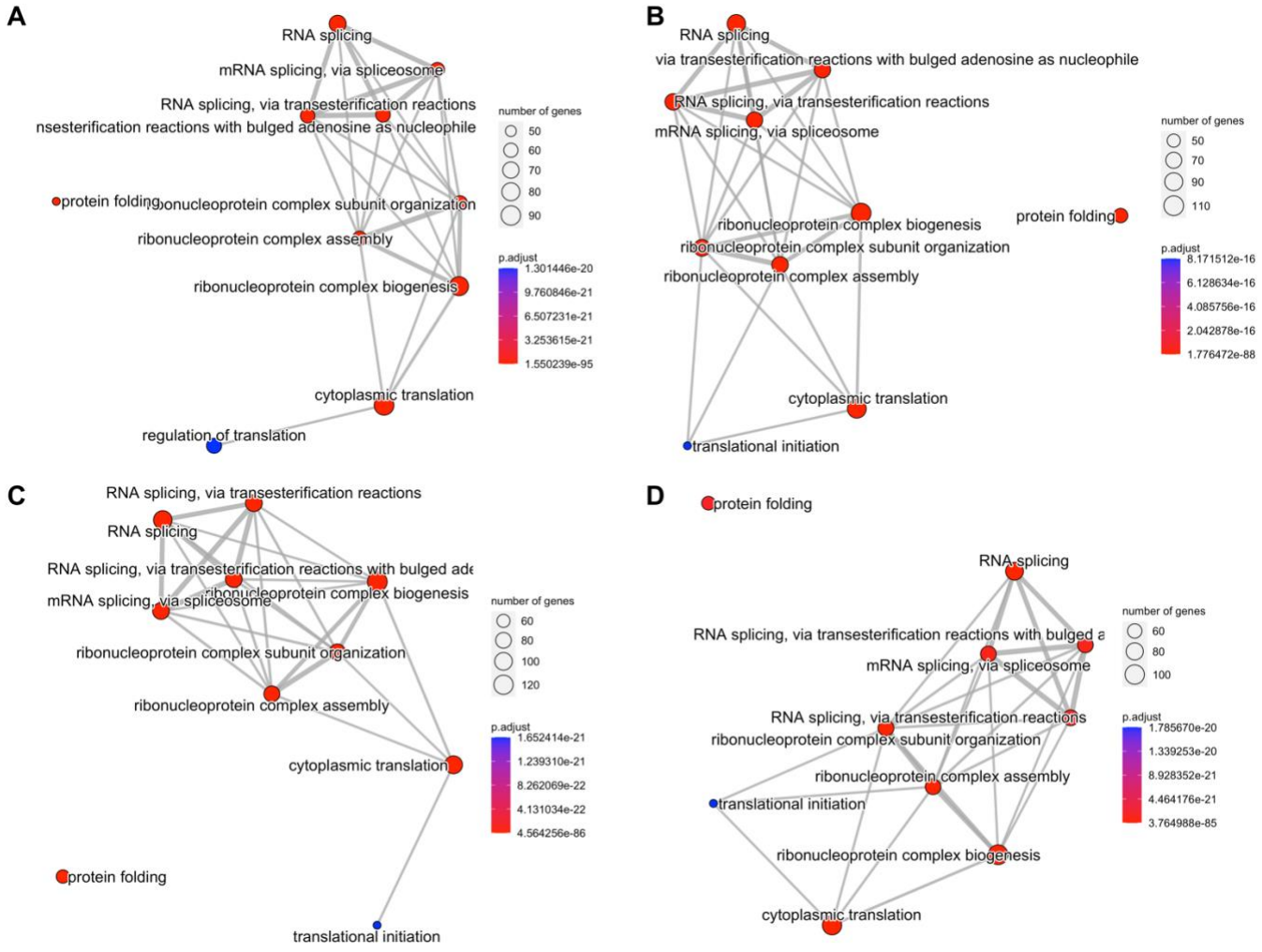


Figure 4: Enrichment maps (E-Maps) illustrate the connectivity of the genes among biological processes. (A) shows the enrichment map of GO terms in Monocytes, (B) shows Hela cells enriched terms, (C) provides the connected processes in Melanoma genes, and (D) shows PDAC shared genes among biological process terms.

To investigate the complexity of the same gene that probably belongs to different GO terms, we constructed Gene-Knowledge-Based Networks (C-Net plots). As illustrated in (Figure 5), we highlighted the top 5 hits in their biological process "BP" term based on the significant level (padj < 0.05). Among the large number of enriched genes found in all cancer cell types, five main gene families were abundant which are Eukaryotic Initiation Factor (EIF), Ribosomal Protein Large subunit (RPL), Ribosomal Proteins (RPS), DEAD-(DDX) and DEAH-(DHX) box gene families. The enriched genes belonging to those gene families were found abundantly in "ribonucleoprotein complex biogenesis" and "cytoplasmic translation" terms. YBX1, RPS13, SYNCRIP, HNRNBU, RPS26, and PABC1 genes were shared between cytoplasmic translation and RNA splicing terms among all cell types. Also, SETX, SRSF6, SRSF10, SF3B1, SNRPD1, and SF3A3 genes were connected between "RNA splicing" and "ribonucleoprotein complex subunit organization" processes in all cancer cell types. Moreover, DHX9 was the only gene connected to all biological process terms in all cancer cell types.

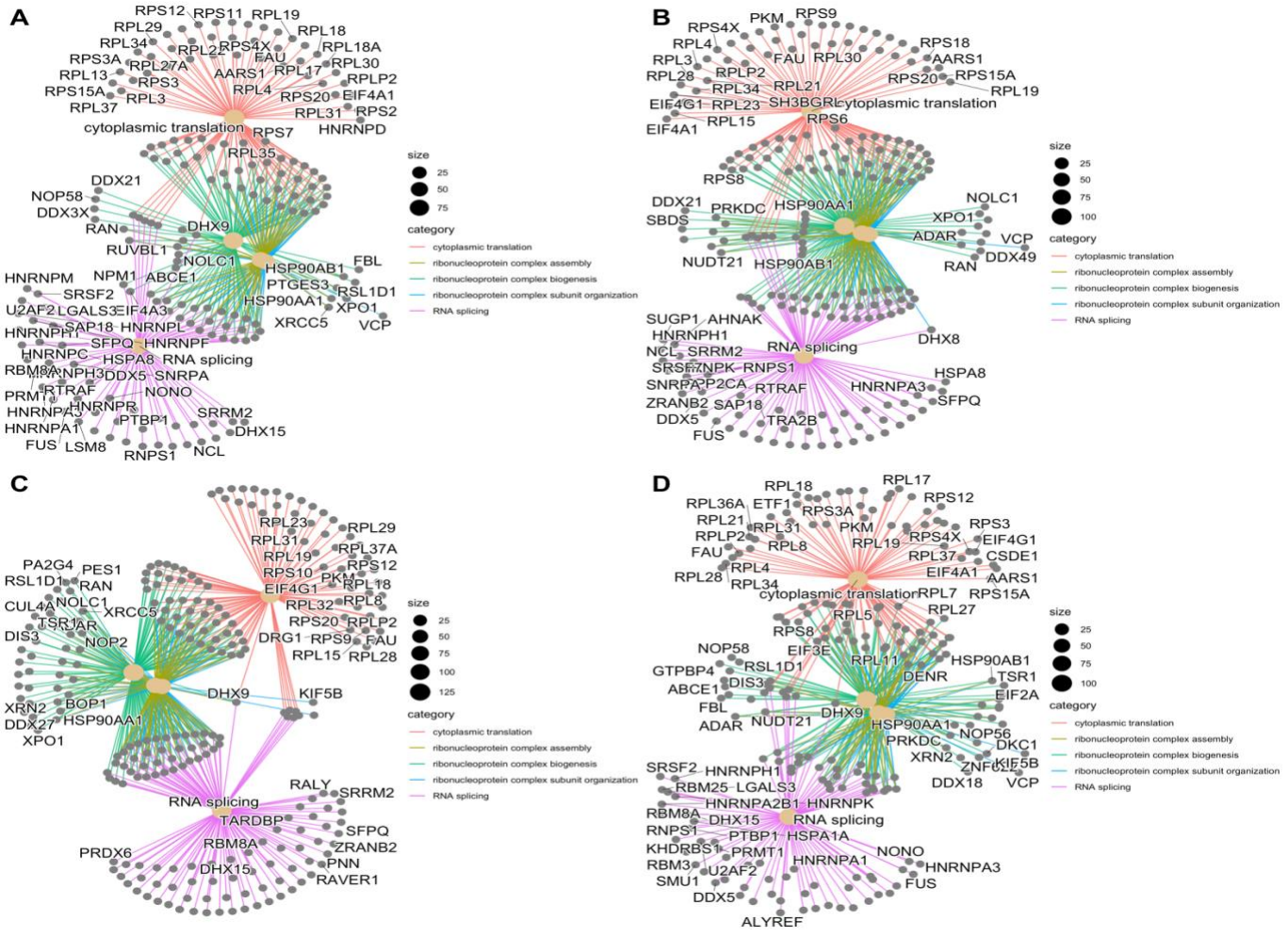


Figure 5: Knowledge-based network plots (C-Net) visualize the complexity of the same gene among different biological processes. (A) Monocyte cells C-Net plot. (B) Hela cells C-Net plot. (C) Melanoma cells C-Net plot. (D) PDAC C-Net plot. Each GO term has a specific colour and is unified among all cell types. The black dots (Nodes) represent the genes, and the brown dots represent the size of gene counts for each GO term while the (edges) represent the link between genes in one or more biological processes.

3.2.2- *Amyotrophic Lateral Sclerosis (ALS), COVID-19 and cancer share protein biomarkers*

We obtained 167 enriched pathways associated with the gene lists among all cancer cell types using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. KEGG pathway enrichment analysis showed similar pathways enriched terms, particularly in protein synthesizes, biochemical processes, and disease pathways. The “Ribosome”, “Amyotrophic lateral sclerosis”, and “Coronavirus disease - COVID-19” were the top three enriched pathways associated with all cancer types. As (Figure 6) illustrates, they included the highest gene counts in these pathways. In addition, the top three pathways have slightly different ranks among the dot plots of each cell type based on their p-value significance level.

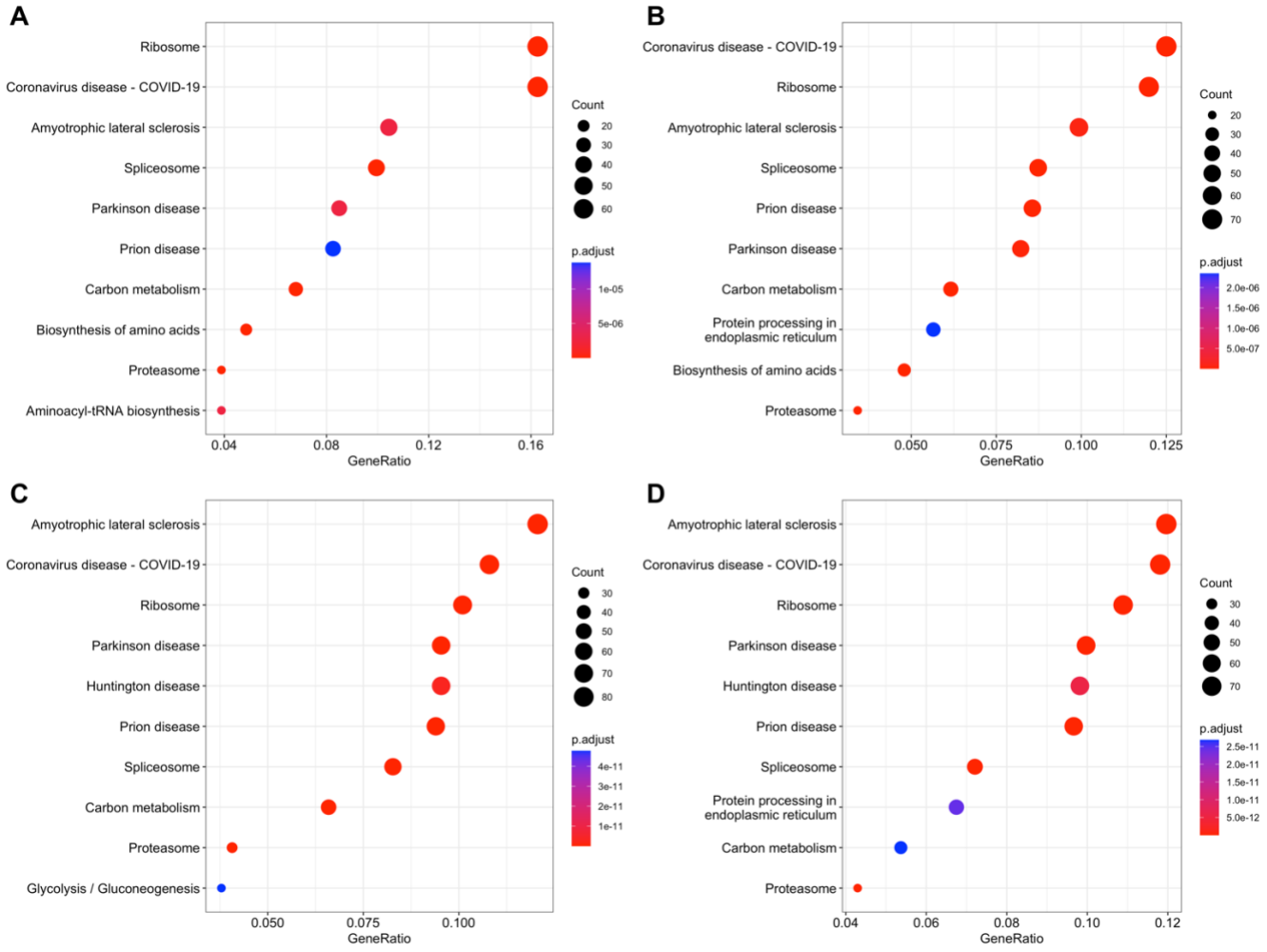


Figure 6: Dot plots provide the most 10 significant KEGG enriched pathways (padj < 0.05) in four cancer cell types. (A) indicates GO terms for Monocytes enriched pathways, (B) shows Hela KEGG pathways, (C) illustrates Melanoma enriched pathways, and (D) indicates PDAC pathways using the KEGG database. The colour dots indicate the significance level of each term, and the circle size indicates the gene counts. X-axis represents the gene ratio, and Y-axis represents KEGG-enriched pathways.

We constructed an enrichment map (E-Map) to investigate the connectivity of the genes in our cancer cell types datasets. As (Figure 7) illustrates, Monocytes, Hela, Melanoma, and PDAC genes were involved in other disease pathways. Prion, Parkinson, and amyotrophic lateral sclerosis were highly connected to proteasome genes. Also, carbon metabolism was connected with the biosynthesis of amino acids and gluconeogenesis terms. Moreover, according to the enrichment map of all cancer types, Coronavirus disease COVID-19 was highly connected to the ribosome genes.

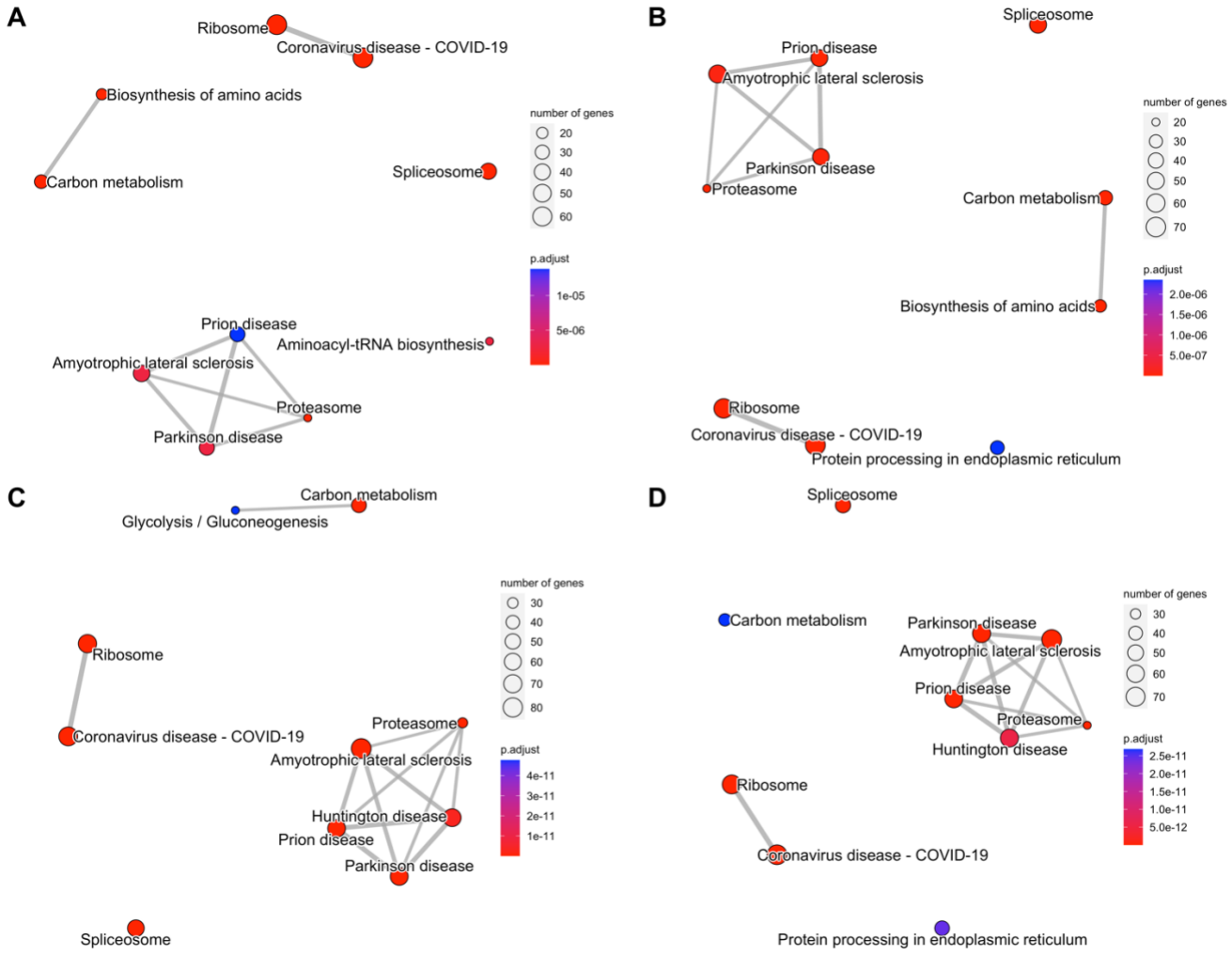


Figure 7: Enrichment maps (E-Maps) illustrate the KEGG pathways of the associated genes of four cancer cell types. (A) shows the enrichment map of KEGG pathways in Monocytes, (B) shows Hela cells enriched pathways, (C) provides the connected disease pathways in Melanoma genes, and (D) shows PDAC shared genes among KEGG pathway terms.

We constructed Gene-Knowledge-Based Networks (C-Net plots) for the most 5 significant pathways using the KEGG database to investigate the complexity of the same gene in our cancer cell types and their association with other mechanisms and disease pathways. In (Figure 8), we found highly connected genes shared between the Ribosome pathway and Coronavirus COVID-19 disease; those genes belong to RPL and RPS gene families and exist among all cancer cell types. Also, there were common genes between carbon metabolism and biosynthesis of amino acid pathways such as PGK1, GAPDH, GOT2, TKT, and PSAT1 found in Hela cells, see (Figure 8B). As shown in (Figure 8D), a robust connection exists between proteasome and Parkinson's disease genes, particularly in their proteasome subunit genes (PSMs).

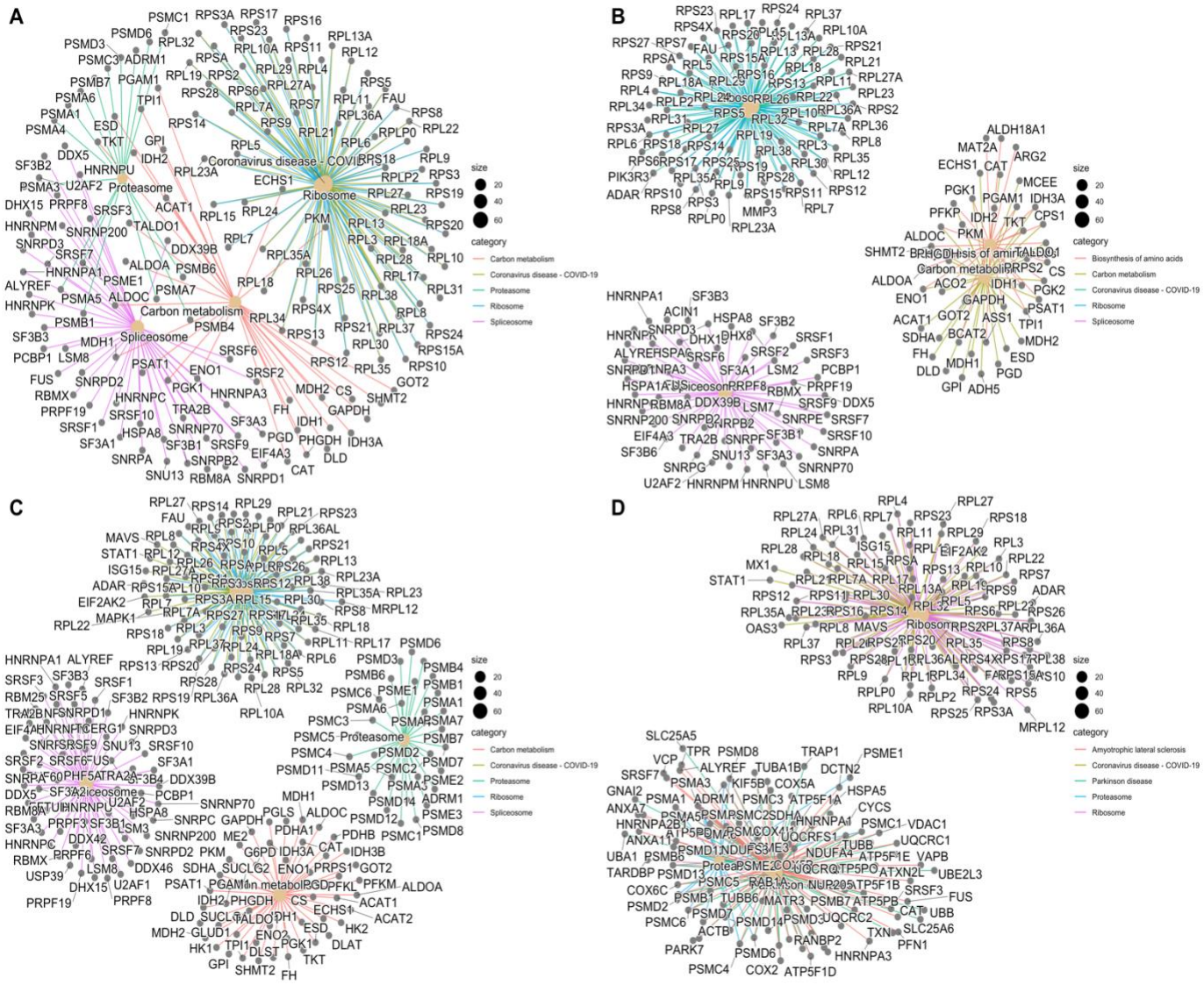


Figure 8: Knowledge-based network plots (C-Net) visualize the complexity of the same gene and its association with other KEGG pathways. (A) Monocyte cells C-Net plot. (B) Hela cells C-Net plot. (C) Melanoma cells C-Net plot. (D) PDAC C-Net plot. Each GO term has a specific colour and is unified among all cell types. The black dots (Nodes) represent the genes, and the brown dots represent the size of gene counts for each pathway while the (edges) represent the link between genes in one or more biological or disease pathways.

3.3- Protein correlation analysis showed variability among cancer datasets

To investigate the biological networks between the belonging proteins that are associated with Melanoma and Hela cell types, we constructed a network based on a correlation matrix of proteins using a specific correlation cutoff (the minimum threshold was 0.45) for each cell type.

As shown in (Figure 9A), there are 36 correlated proteins found in the Hela cells set4 when (0.75) is used as a cutoff for the correlation matrix. Two hub proteins are found in the network, which is O13283 (NAD(P)H-dependent D-xylose reductase I,II) and O95749 (Geranylgeranyl pyrophosphate synthase). In Hela set5 (Figure 9B), the hub proteins were Q15382 (GTP-binding protein Rheb), Q08J23 (RNA cytosine C(5)-methyltransferase NSUN2), and Q15239 (Pregnancy-specific beta-1-glycoprotein 5) out of total proteins 57 with a (0.87) correlation cutoff.

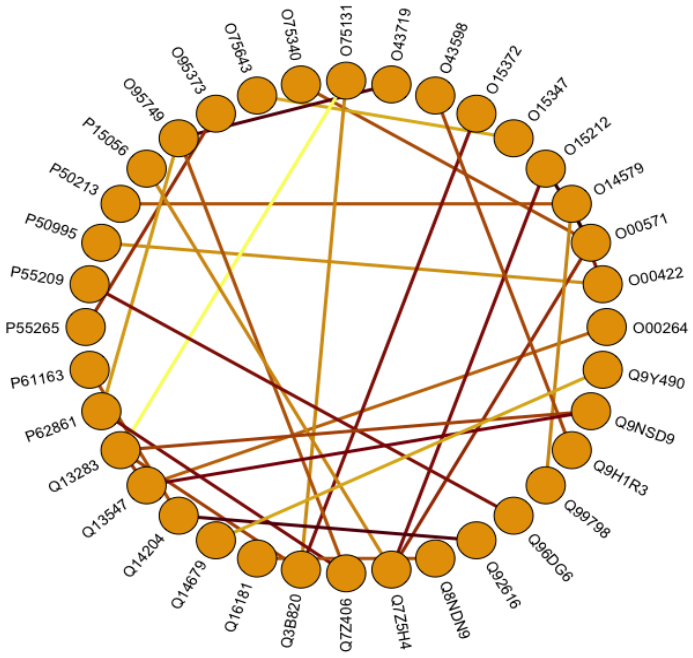


Figure 9A: Network analysis for the correlated proteins in Hela_set4.

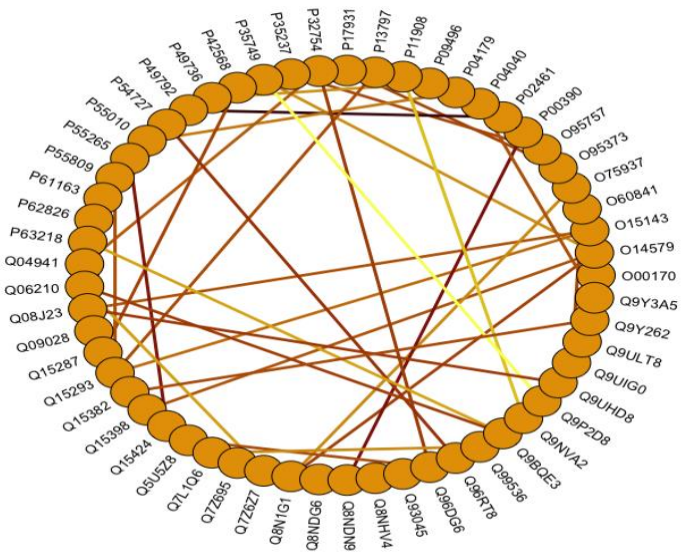


Figure 9B: Network analysis for the correlated proteins in Hela_set5.

When we used (0.45) as a cutoff in Melanoma set1, we obtained 44 correlated proteins, including six hub proteins with six or more connected edges, which are P26373 (60S ribosomal protein L13), P26805 (Gag polyprotein), P49207 (60S ribosomal protein L34), P63220 (40S ribosomal protein S21), P61254 (60S ribosomal protein L26), and P84098 (60S ribosomal protein L19).

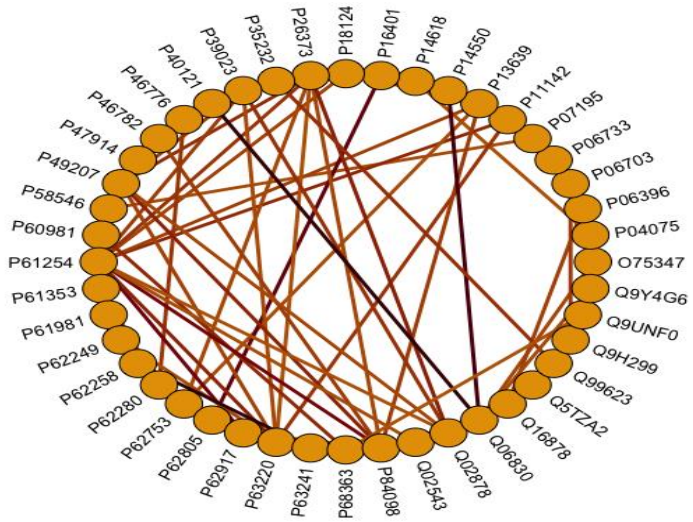


Figure 10: Network analysis for the correlated proteins in Melanoma_set1.

3.4- Unsupervised clustering unfolded distinct sub-cellular groups in cancer datasets

To have an overview of the structure of the proteomics datasets in each cell type, we projected the cancer cells using Principal Component Analysis (PCA) to investigate the variability between subcellular groups. We obtained high variability in Monocytes (Figure 11A), Hela (Figure 11B), Melanoma (Figure 11C), and PDAC (Figure 11D) subcellular populations. In contrast, we utilized UMAP to cluster each dataset better and differentiate between their GO terms.

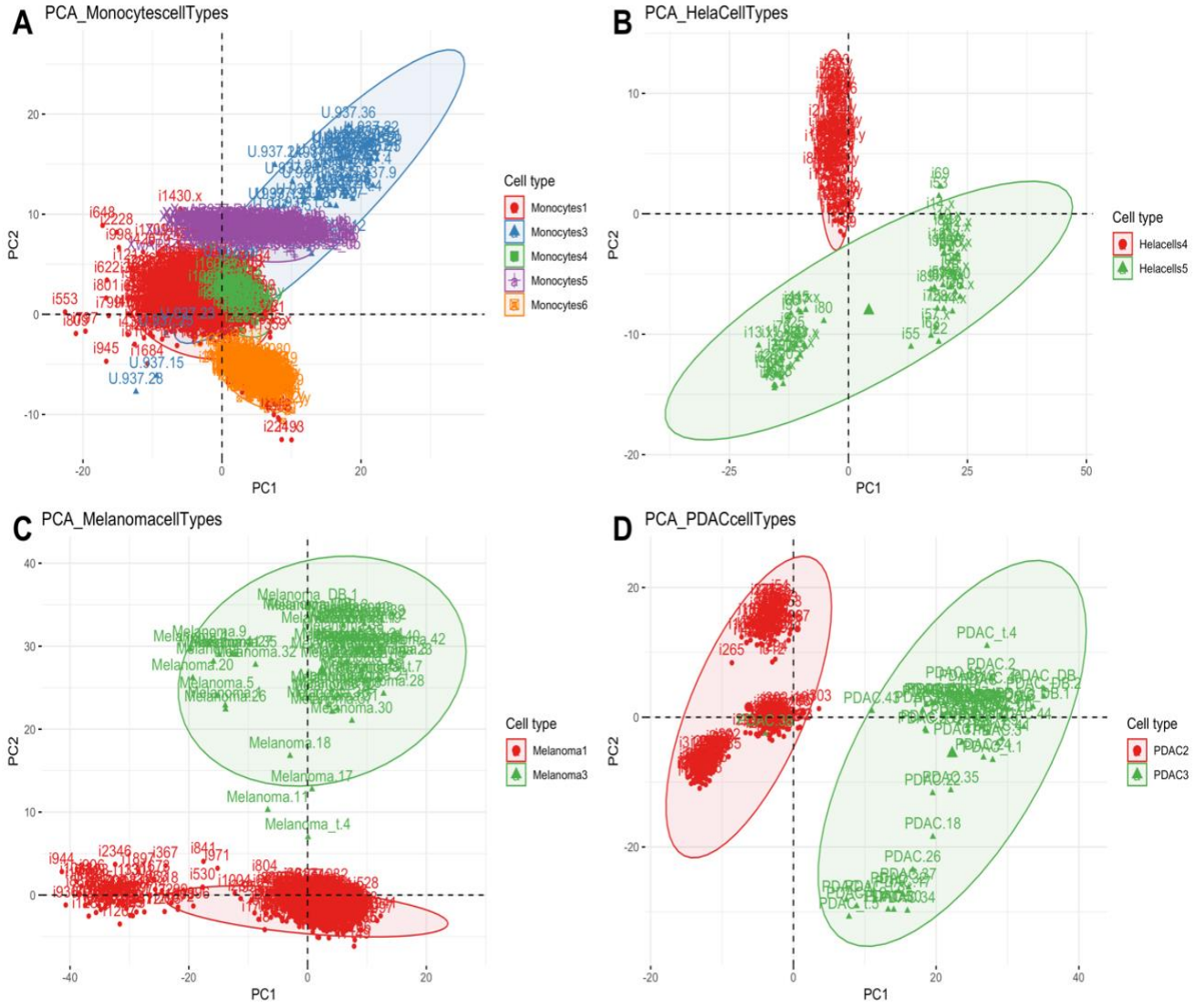


Figure 11: Principal Component Analysis (PCA) shows the variability between sub-cell groups in four cancer cell types. (A) illustrates the Monocyte cells, (B) shows Hela cells, (C) Melanoma cell populations, (D) PDAC cell types.

With the help of STRING-DB to annotate the subcellular proteins, we distinguished between subcellular populations using the protein set enrichment analysis for each sub-cell type. We started clustering each proteomic dataset that belongs to four cancer cell types using Uniform Manifold Approximation and Projection (UMAP) technique. Then we annotated sub-cellular groups to observe the differences in their biological roles and enrichment terms.

The annotation GO enrichment results have not detected significant enrichments in the first UMAP within both clusters (Figure 12A) in U3 subgroups (Figure 12B); the red and blue clusters also have no significant enrichments. In contrast, green cluster proteins were involved in Ion transmembrane transport and Calcium ion binding processes. They were detected in the mitochondrial membrane, collagen-containing extracellular matrix, and organelle inner membrane localizations. As (Figure 12C) illustrates, both red and light blue clusters have not scored significant enrichment results according to the STRING-DB search query. In contrast, U5 subgroups in (Figure 12D) showed no significant enrichment in the light blue, green, and red clusters. In contrast, proteins belonging to the purple cluster were enriched in cell chemotaxis and lymphocyte activation involved in the immune response.

The associated molecular function was structural molecule activity located in the cytosolic large ribosomal subunit. The last monocyte subgroups in (Figure 12E) were varied in their GO annotations. Red cluster proteins enriched in muscle filament sliding, blood coagulation, and positive regulation of supramolecular fiber organization. They played a role in S100 protein binding, calcium-dependent protein binding and actin binding. The localization was detected in the plasma membrane. Green cluster proteins were enriched in the actin filament and cytoskeleton organizations and response to wounding processes. They are located in the plasma membrane and play a role in phospholipase inhibitor activity. Blue cluster proteins were located in the extrinsic component of the plasma membrane and enriched in T-cell stimulation and muscle contraction ontology terms.

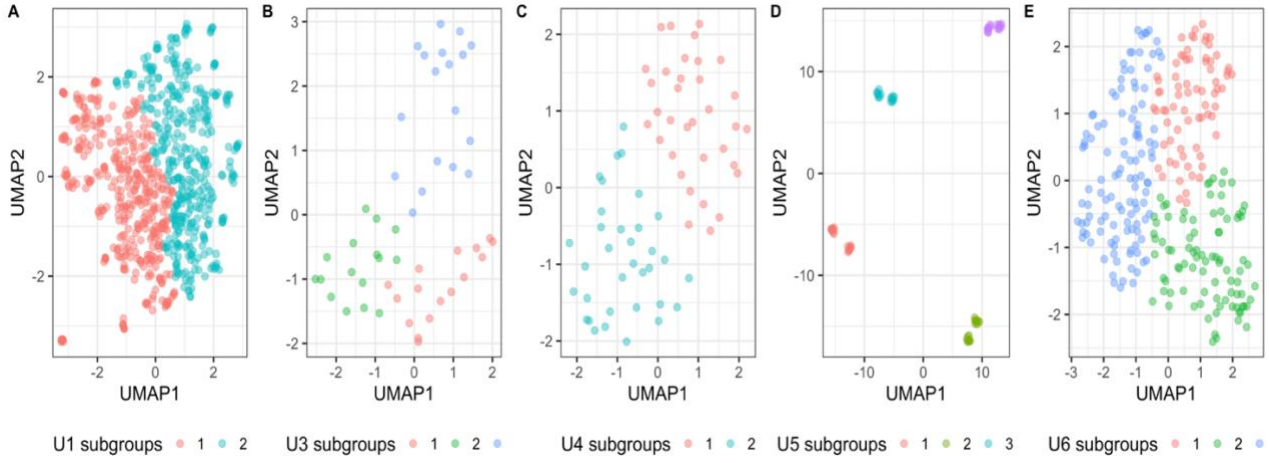


Figure 12: UMAP for Monocyte subgroups.

As (Figure 13B) illustrates, Hela5 subgroups include two distinct clusters. While enrichment results have similar GO terms in both clusters, Hela subgroup_1 (Cyan) has more enriched terms in its biological processes than Hela subgroup_2 (light red), such as Peptide, Amide, and Nitrogen compound transport processes. The molecular function terms were similar: structural constituent of the ribosome and structural molecule activity. Also, Cellular component terms were similar except for the chromatin term that only enriched in Hela subgroup_1. In (Figure 13A) we obtained four sub-cellular groups of the Hela4 cells. The enrichment analysis showed no significant enriched terms in the (light red) and (green) clusters, while (cyan) and (purple) groups share the same GO terms.

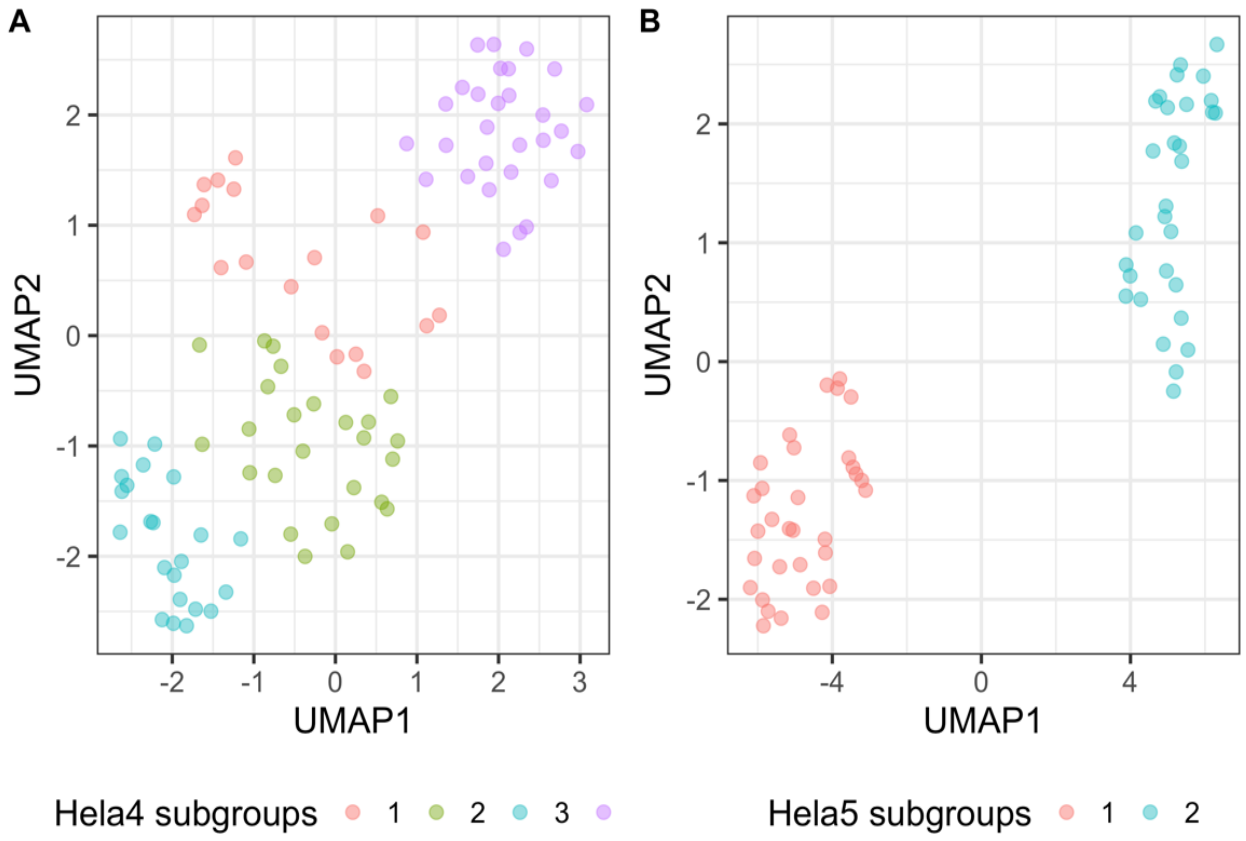


Figure 13: UMAP for Hela subgroups.

As (Figure 14A) illustrates, the Biological Process terms were not detected in the red and green clusters. At the same time, it was enriched in the blue cluster, involved mainly in the mitochondrial ATP synthesis coupled electron transport, oxidative phosphorylation, and cristae formation processes. Both reg and green clusters shared similar terms in their molecular function and cellular component, such as structural constituent of the ribosome and ribosomal subunits, respectively. The GO terms were slightly different in the blue cluster with a functional enrichment in the cytochrome-c oxidase activity. They were located in the cytochrome complex and mitochondrial respiratory chain complex IV localizations. All clusters in melanoma1 subgroups were involved in the Ribosome KEGG pathway.

In Melanoma3 subgroups (Figure 14B), the proteins belonging to the green cluster were enriched in regulating alternative mRNA splicing via spliceosome. At the same time, their molecular function and cellular component terms were not detected. The red cluster's proteins were enriched in regulating collagen biosynthesis and metabolic processes. It has a biological role in calcium-dependent protein binding and was localized in the actin cytoskeleton. Lastly, the cytoplasmic translation and oxidative phosphorylation processes enrich the blue cluster proteins. The belonging proteins were located in the endoplasmic reticulum lumen and mitochondrial membrane. It played various roles, including RAGE receptor binding, calcium-dependent protein binding, and proton transmembrane transporter activities.

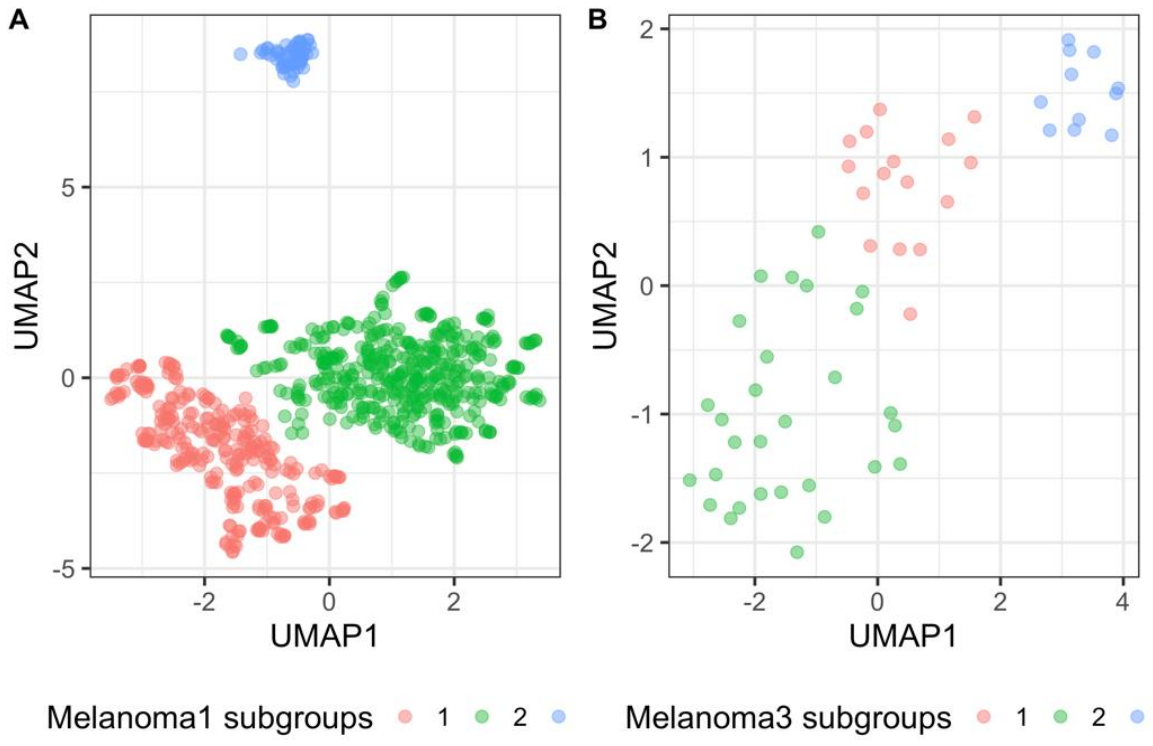


Figure 14: UMAP for melanoma subgroups

In (Figure 15A), we obtained three distinct clusters in the PDAC set2. The enrichment results showed similarity in the biological process terms between clusters 2 (Green) and 3 (Blue). In contrast, cluster 1 (Red) proteins mainly translate cellular amide metabolic processes. In addition, the cellular component in cluster 1 was in the transport vesicle membrane, and the molecular function terms were not annotated yet in the proteins that belong to that cluster. The cellular components and the molecular function terms in cluster 2 and cluster 3 were mainly in the endoplasmic reticulum, ribosomal subunit, and the structural constituent of ribosome, respectively. The enrichment analysis of the PDAC set3 (Figure 15B), particularly in cluster 1 (Red), was involved in cornification, keratinization, and epidermal cell differentiation processes. In addition, it is located in the intermediate filament and microtubule. Moreover, the molecular function terms show that the proteins that belong to this cluster were described in calcium-dependent protein binding and cell-cell adhesion mediator activities. Cluster 2 enrichment analysis was mostly described as uncharacterized annotation among their GO terms.

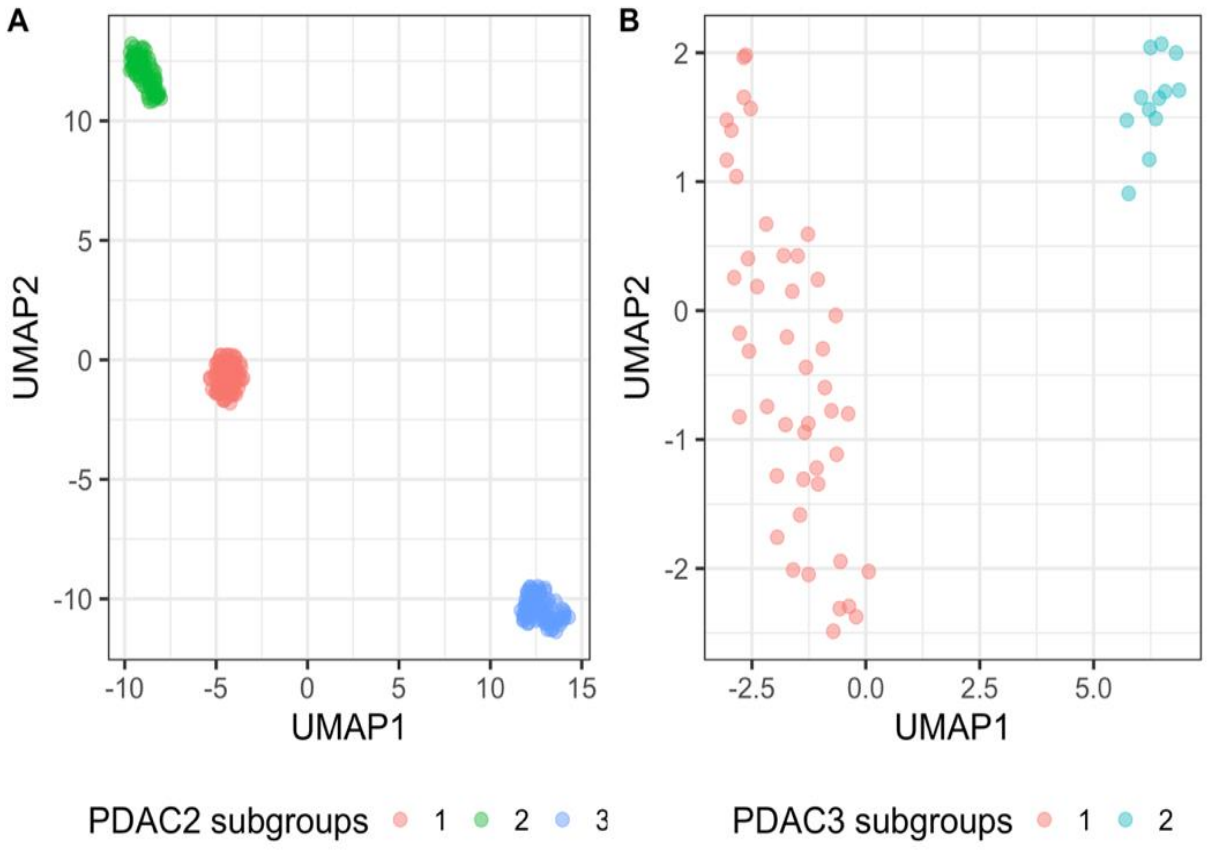


Figure 15: UMAP for PDAC subgroups

Chapter 4: Discussion

The comprehensive single-cell proteomics analysis of Hela, Melanoma, PDAC, and Monocyte cell-line model systems presented in this thesis have shown insights into the functional and regulatory mechanisms underlying cancer progression. By integrating various multi-omics approaches, including gene ontology (GO) enrichment, KEGG pathway analysis, cell clustering, and network analysis, we have better understood the key proteins and gene families involved in biological processes that are dysregulated in these cancer cell types. Additionally, we have found shared protein biomarkers between cancer and other diseases, such as Coronavirus disease. These biomarkers could be potential drug targets and therapeutic strategies. Also, the variation found in the PCA and UMAP clustering results indicated biological and technical variations of these cancer cells.

Moreover, hub proteins were found in the network analysis involved in complex ribosomal biogenesis and catalytic activities, which infer the diverse roles that cancer cells play during cancer development and progression. Overall, the thesis output findings shed light on the necessity of single-cell proteomics analysis in characterizing and differentiating the complex molecular landscape of cancer cells and its regulatory pathways, which could be potential drug targets for cancer treatments. The following paragraphs will provide a more explained interpretation of the results and their implications for cancer research and personalized medicine.

Our gene set enrichment analysis showed that most annotated genes were found in the translation processes. This annotation was expected because Dysregulated translation in cancer cells is associated with the phenotypic hallmarks of malignancy and allows cancer cells to adapt to stresses associated with the tumor microenvironment and therapies [80]. Most translation processes share similarities between genes among all cancer cell types. One of the gene families is Eukaryotic Initiation Factor (EIF). In BRAF-mutated tumors, increased expression of eIF4F is associated with resistance to anti-BRAF therapy and metastasis [81]. In addition, eIF4E, a key component of the translation initiation complex, promotes the translation of mRNAs that regulate reactive oxygen species (ROS), further promoting cancer cell survival [82].

Small-molecule inhibitors of eIF4E, such as eIF4E antisense oligonucleotides (ASOs), are currently being tested in clinical trials and have shown promise in inhibiting tumor growth [83]. Although many genes in the EIF gene family have been functionally annotated, the functions of eIF2α in cancer are complex and context-dependent, and direct inhibition of eIF2α phosphorylation needs to be carefully considered in the specific tumor microenvironment [84].
Another enriched protein family that resulted from our enrichment analysis is RNA-binding proteins (RBPs), widely studied regulators of post-translational processes in RNA biology,

including RNA metabolism, transport, and localization [85]. They are believed to reshape the structure of mRNA to alter its affinity for translation machinery, thereby affecting translation [86]. RBPs have been found to play a critical role in regulating RNA translation in tumorigenesis. For instance, RBPs such as Musashi-1 (MSI1) and Musashi-2 (MSI2) are highly expressed in various tumors and can inhibit or promote the translation of oncogenes and tumor suppressor genes [87]. Other RBPs that have been found in our investigations, such as eIF3, SYNCRIP, CELF1, and HuR, have been implicated in liver cancer, leukaemia, colorectal cancer, and breast tumors, respectively, by modulating the translation of specific target genes [88].

Additionally, CPEBs, IMP-3, GAIT complex, and hnRNP L, have been shown to influence tumorigenesis and metastasis by regulating the translation of genes involved in proliferation, angiogenesis, and immune response [89]. Furthermore, hub protein (Q15382), one of the GTP-binding proteins Rheb found in Hela correlated proteins, accelerates cervical cancer by activating mTOR signalling [119].

Despite the critical roles of (RBPs) in cancer and other diseases, the exact mechanisms through which RBPs regulate translation are not fully understood and require further exploration and confirmation. Nevertheless, understanding the functional significance of RBPs in translation regulation provides valuable insights into the complex molecular mechanisms underlying tumorigenesis, metastasis, and immune response. It may lead to developing novel therapeutic strategies targeting RBPs for cancer treatment [90].

The third gene family found in our GO terms findings is Ribosomal Proteins (RPs) which have been implicated in developing and progressing pancreatic cancer, a highly malignant tumor. Such an RPSA is found to be highly expressed in invasive human pancreatic cancer cells and is linked to Integrin α6 (ITGA6), a protein involved in cell-cell attachment regulation. ITGA6 activates the PI3K pathway through AKT phosphorylation, while RPSA activates the MAPK signalling pathway, promoting the invasion and metastasis of pancreatic cancer cells [95]. In addition, RPSA has been found to form a complex with transient receptor potential melastatin-related 7 (TRPM7), regulating pancreatic ductal cancer cell migration, suggesting that RPSA may function through multiple pathways and could be a promising target for pancreatic cancer treatment [96].

Our in-silico analysis showed RPL34 enriched in PDAC cell type that was confirmed by In vivo assays in affecting pancreatic cancer cells through the MAPK and p53 pathways, indicating that it could be a potential biomarker for detecting and treating pancreatic cancer [97]. Similarly, the Inhibition of RPS15A by miR-519d-3p leads to down-regulation of the Wnt/β-catenin pathway in pancreatic cancer, suggesting the potential of RPS15A as a target for pancreatic cancer treatment [98]. In the melanoma cell type enrichment results, we found RPL38 that has been implicated in translating a specific subset of Hox mRNA. This suggests that ribosomal proteins may have more

specific functions beyond being components of ribosomal subunits [91]. Also, the hub protein (P26373), which is RPL13, can be a potential therapeutic strategy to inhibit melanoma tumor development [92].
Hela cell line GO terms findings have RPL34 abundantly enriched, and a literature review revealed that RPL34 functions as a tumor suppressor, potentially inhibiting cervical cancer cell proliferation, migration, and invasion through the MDM2-P53 pathway. An antisense long non-coding RNA, RPL34-AS1, regulates the expression of RPL34. These findings provide valuable insights into the mechanisms underlying cervical cancer and suggest RPL34 as a potential target for early diagnosis and treatment [93].

Another gene found in Monocyte cells is RPS15A, which is abundantly enriched and significantly regulates carcinogenesis in various human cancers. Increased expression of RPS15A in tumor cells is associated with phenotype changes indicative of more aggressive malignancy. The findings of this study suggest that RPS15A may also have a crucial role in regulating cell growth in acute myeloid leukemia (AML) and that downregulation of RPS15A may induce apoptosis in AML cells. These results strongly indicate that RPS15A may play a central role in AML carcinogenesis and maintenance of malignant phenotypes, making it a potential therapeutic target in AML treatment [94].

Another important gene connected to the most enriched GO terms was DHX9, a member of the DExD/H-box family of helicases, which possesses a conserved helicase core domain, double-stranded RNA-binding domains, a nuclear transport domain, and a single-stranded DNA-binding RGG-box. DHX9 is essential for maintaining cellular homeostasis, but defects in DHX9 can adversely impact cell growth and viability, potentially leading to human diseases [101]. Notably, DHX9 is overexpressed in several cancer types, as demonstrated by two independent studies analyzing lung cancer samples, which showed higher DHX9 expression in tumour samples compared to normal lung tissues, and correlated DHX9 overexpression with poorer patient survival [102].

Current research has highlighted the potential of targeting DHX9 as a chemotherapeutic strategy. In a mouse lymphoma model resistant to ABT-737, a potent inhibitor of pro-survival BCL-2 family members, DHX9 was identified as a modifier of ABT-737 resistance through an shRNA screen. DHX9 suppression sensitized lymphoma cells to ABT-737 by activating a p53 response, inducing replicative stress, upregulating NOXA, inhibiting MCL-1, and ultimately leading to apoptosis [99]. The mouse lymphoma model used in this study recapitulates features of non-Hodgkin's lymphomas, suggesting DHX9 as a potential target in combination with ABT-737 or its derivatives. Furthermore, DHX9 knockdown has shown synergistic effects with dexamethasone, a glucocorticoid, in human multiple myeloma cell lines [100], indicating potential for targeting DHX9 in multiple myeloma.

KEGG pathway enrichment analysis provided insights about shared biomarkers between amyotrophic lateral sclerosis (ALS) disease and our cancer cell types. A recent study was conducted by Taguchi and his colleagues when they identified biomarkers to distinguish between patients with amyotrophic lateral sclerosis (ALS) and healthy individuals. They utilized a principal component analysis (PCA)-based unsupervised feature extraction (FE) approach on microarray gene expression data, they identified 101 probes as potential gene biomarkers for ALS with a high accuracy of 95% in discriminating between ALS patients and controls. Most of the genes associated with these probes were found to be related to various types of cancer. These findings provide valuable insights for developing new therapeutic options or drugs targeting ALS and cancer [103].

Additionally, research has demonstrated that prostate cancer survivors have a reduced risk of developing amyotrophic lateral sclerosis (ALS). In contrast, melanoma and tongue cancer survivors have a significantly increased risk of ALS. Despite differences in the clinical outcomes of ALS and cancer, both diseases share abnormalities in cellular processes such as cell survival, cell death, and cell cycle regulation. In ALS, these abnormalities lead to progressive motor neuron dysfunction, whereas in cancer, they result in uncontrolled cell survival and proliferation. Over thirty genes have been identified as causing ALS, and new ALS-causing genes continue to be discovered annually [104].

Our KEGG pathway analysis presents the significant pathways for cancer biomarkers, COVID-19 biomarkers, and common genes between them. Cancer biomarkers such as CCNB1 and CDK1 are found to be enriched in pathways like the p53 signalling pathway. Furthermore, cancer biomarkers are enriched in pathways related to coronavirus disease, Parkinson's disease, Huntington's disease, Alzheimer's disease, and pathways of neurodegeneration, including genes like RPS15, RPS9, RPS5, RPL35, RPL9, NDUFB7, and UQCRQ. These findings suggest that COVID-19 may exacerbate the pathways involved in other diseases, leading to comorbidity [112].

There are overlapping molecular mechanisms in the pathogenesis of COVID-19 and pancreatic tumorigenesis. The entry of SARS-CoV-2 into host cells is facilitated by ACE-2, TMPRSS2, and CTSL, which are also associated with pancreatic adenocarcinoma [108]. Cathepsin L is upregulated in various cancers, including pancreatic cancer, while TMPRSS2 is moderately upregulated, and ACE-2 is overexpressed in some pancreatic carcinomas [106]. ACE2 upregulation has been linked to favourable survival in pancreatic cancer, but SARS-CoV-2 reduces ACE2 expression[107]. RAGE, involved in neutrophil extracellular trap formation in pancreatic cancer, may also play a role in COVID-19. In conclusion, pancreatic cancer increases the risk of COVID-19 and its severity, and coronavirus infection may contribute to pancreatic cancer development [105].

Two studies have been conducted to investigate the impact of COVID-19 on cervical tissue; Vavoulidis and colleagues reported on a 32-year-old HPV-positive patient who developed CIN I after a COVID-19 infection [110]. In contrast, Ondič and colleagues analyzed SARS-CoV-2 RNA in cervical smears from 102 patients with asymptomatic COVID-19 infection and found that only one had a positive result in qPCR [111]. Both studies concluded that SARS-CoV-2 may directly affect cervical epithelium, leading to adverse outcomes and disease progression [109]. Overall, the correlation between COVID-19 and various cancer types is still poorly investigated, whether on the genome or the proteome levels, and more future research should be conducted in this field.

In Melanoma subgroup clustering, no biological variations have been found between melanoma subgroups. The original dataset generated by Leduc and his colleagues [113] has indicated a technical variation between subgroups regarding reproducibility across biological replicates and various MS protocols. The enrichment results of Melanoma3 subgroups showed biological variability between three clusters and various ontology terms among each sub-cellular group. This indicates that the PlexDIA protein identification approach increases the sensitivity of single-cell proteomics analysis, which improves the detection of different proteomes across sub-cell types [67].

In PDAC subgroup clustering, UMAP projections showed technical variation between PDAC2 subgroups. Thus, referring to three cell types were utilized for single-cell proteomic profiling, which are HPAF-II (green), CFPAC-I (blue), and BxPC-3 (red) cell types. While the enrichment results have a slight biological variation among PDAC subgroups, the quantification method (pSCoPE) increases proteome coverage. It maximizes the peptides' quantification consistency, increasing protein abundance across sub-cell types [68]. On the other hand, two subgroups in PDAC3 and differ in their annotation, which indicates different biological roles for each subgroup. Different protein identification platforms, TimsTOF SCP and Q-Exactive classic, also contributed to the technical variation between PDAC3 subgroups [67].

The UMAP projections of Hela datasets did not show biological variations between Hela4 and Hela5 subgroups. Hela5 subgroups showed technical variation regarding the batch effects and incomplete normalization of the SCoPE2 analysis pipeline [118]. Our enrichment results of monocyte cells among datasets do not have a significant annotation which indicates that the large-scale experiments of proteomics identification at the level of single-cell needs more improvements, particularly in the experimental aspects to better classify subpopulations based on their phenotypic profile (proteome characterization).

Chapter 5: Conclusion

In summary, our analysis found that the ribosomal proteins (RPs) and RNA-Binding Proteins (RBPs) were annotated as key players in the regulation of translation in cancer cell types. In addition, the Eukaryotic Initiation Factors (EIFs) gene family is crucial in promoting cancer cell survival and metastasis. KEGG pathways highlighted shared biomarkers and molecular mechanisms between COVID-19 and cancer. Biological and technical variations were observed in the proteomic profiling of subpopulations, highlighting the importance of considering these factors when expanding the proteome profile in large-scale single-cell experiments. Our findings provide valuable insights in understanding the complexity of cancer proteomics at the single-cell level and identifying potential drug targets for therapeutic interventions.

Chapter 6: Further Recommendations

Based on the analysis results, some potential recommendations would be helpful for further research investigations, such as optimizing the experimental protocols to expand to large-scale experiments. In addition, future research is required to better understand the molecular mechanisms of how RBPs regulate translation and investigate the downstream target RBPs genes that could be potential therapeutic targets. Also, the ribosomal proteins could provide a promising strategy for cancer treatment approaches, especially in PDAC and melanoma. Furthermore, future studies could be considered, such as utilizing advanced single-cell techniques like imaging mass cytometry and high-dimensional single-cell sequencing to better differentiate between sub-cellular populations. Finally, Given the computational observations of the potential comorbidity between COVID-19 and cancer, more research should be done to understand the regulatory pathways underlying the shared biomarkers to identify potential drug targets for treatment and investigate the possibilities of early diagnosis for cancer and COVID-19.

Appendix

Appendix 1: The table shows the selected data for computational analysis.

Sample type	Proteins	Single-cells	Cell type	Reference paper
pSCoPE	2844	1543	Melanoma and Monocytes (U937)	Leduc et al, 2022
pSCopE	1659	206	Pancreatic ductal adenocarcinoma (PDAC)	Huffman et al, 2022
	1123	373	Bone-Marrow-Derived Macrophages (BMDMs)	
plexDIA	1475	50	Monocytes (U937)	Derks et al, 2022
	1475	56	Melanoma	
	1475	58	Pancreatic ductal adenocarcinoma (PDAC)	
SCoPE2	1647	68	Monocytes (U937)	Leduc et al, 2021
	1647	95	Hela cells	
SCoPE2	1762	62	Hela cells	Petelski et al, 2021
	1762	64	Monocytes (U937)	
SCoPE2	2772	277	Monocytic cell line (U937)	Specht et al, 2019
	2772	741	Macrophage-like cells	

Appendix 2: The table provides total enriched GO terms and KEGG pathways per cell types

Cell Type	Genes	Enriched GO terms	% Enriched genes	KEGG pathways
Melanoma	1083	676	62%	54
Hela	912	527	58%	35
PDAC	961	698	73%	44
Monocytes	584	552	95%	34

Bibliography

[1] I. M. Cristea, S. J. Gaskell, and A. D. Whetton, "Proteomics techniques and their application to hematology," *Blood*, vol. 103, no. 10, pp. 3624–3634, May 2004, doi: https://doi.org/10.1182/blood-2003-09-3295.

[2] M. Traini *et al.*, "Towards an automated approach for protein identification in proteome projects," *Electrophoresis*, vol. 19, no. 11, pp. 1941–1949, Aug. 1998, doi: https://doi.org/10.1002/elps.1150191112.

[3] A. Pandey and M. Mann, "Proteomics to study genes and genomes," *Nature*, vol. 405, no. 6788, pp. 837–846, Jun. 2000, doi: https://doi.org/10.1038/35015709.

[4] B. Domon, "Mass Spectrometry and Protein Analysis," *Science*, vol. 312, no. 5771, pp. 212–217, Apr. 2006, doi: https://doi.org/10.1126/science.1124619.

[5] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, Feb. 2001, doi: https://doi.org/10.1038/35057062.

[6] R. D. Canales *et al.*, "Evaluation of DNA microarray results with quantitative gene expression platforms," *Nature Biotechnology*, vol. 24, no. 9, pp. 1115–1122, Sep. 2006, doi: https://doi.org/10.1038/nbt1236.

[7] J. Cox and M. Mann, "Is Proteomics the New Genomics?," *Cell*, vol. 130, no. 3, pp. 395–398, Aug. 2007, doi: https://doi.org/10.1016/j.cell.2007.07.032.

[8] Yates Iii and J. R, "A century of mass spectrometry: from atoms to proteomes," *Nature Methods*, vol. 8, no. 8, pp. 633–637, Aug. 2011, doi: https://doi.org/10.1038/nmeth.1659.

[9] B. KURIEN and R. SCOFIELD, "Western blotting," *Methods*, vol. 38, no. 4, pp. 283–293, Apr. 2006, doi: https://doi.org/10.1016/j.ymeth.2005.11.007.

[10] R. M. Lequin, "Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA)," *Clinical Chemistry*, vol. 51, no. 12, pp. 2415–2418, Dec. 2005, doi: https://doi.org/10.1373/clinchem.2005.051532.

[11] M. A. Fernández-Baldo *et al.*, "Development of an indirect competitive enzyme-linked immunosorbent assay applied to the Botrytis cinerea quantification in tissues of postharvest fruits," *BMC Microbiology*, vol. 11, no. 1, Oct. 2011, doi: https://doi.org/10.1186/1471-2180-11-220.

[12] W. Hogrefe, X. Su, J. Song, R. Ashley, and L. Kong, "Detection of Herpes Simplex Virus Type 2-Specific Immunoglobulin G Antibodies in African Sera by Using Recombinant gG2, Western Blotting, and gG2 Inhibition," *Journal of Clinical Microbiology*, vol. 40, no. 10, pp. 3635–3640, Oct. 2002, doi: https://doi.org/10.1128/jcm.40.10.3635-3640.2002.

[13] S. Wiese, K. A. Reidegeld, H. E. Meyer, and B. Warscheid, "Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research," *PROTEOMICS*, vol. 7, no. 6, pp. 1004–1004, Mar. 2007, doi: https://doi.org/10.1002/pmic.200790019.

[14] G. Such-Sanmartín, N. Bache, A. K. Callesen, A. Rogowska-Wrzesinska, and O. N. Jensen, "Targeted mass spectrometry analysis of the proteins IGF1, IGF2, IBP2, IBP3 and A2GL by blood protein precipitation," *Journal of Proteomics*, vol. 113, pp. 29–37, Jan. 2015, doi: https://doi.org/10.1016/j.jprot.2014.09.013.

[15] J. N. Adkins *et al.*, "Toward a Human Blood Serum Proteome," *Molecular & Cellular Proteomics*, vol. 1, no. 12, pp. 947–955, Dec. 2002, doi: https://doi.org/10.1074/mcp.m200066-mcp200.

[16] S. Khatib-Shahidi, M. Andersson, J. L. Herman, T. A. Gillespie, and R. M. Caprioli, "Direct Molecular Analysis of Whole-Body Animal Tissue Sections by Imaging MALDI Mass Spectrometry," *Analytical Chemistry*, vol. 78, no. 18, pp. 6448–6456, Aug. 2006, doi: https://doi.org/10.1021/ac060788p.

[17] G.-C. Yuan *et al.*, "Challenges and emerging directions in single-cell analysis," *Genome Biology*, vol. 18, no. 1, May 2017, doi: https://doi.org/10.1186/s13059-017-1218-y.

[18] A. S. Venteicher *et al.*, "Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq," *Science*, vol. 355, no. 6332, Mar. 2017, doi: https://doi.org/10.1126/science.aai8478.

[19] L. Pompella *et al.*, "Pancreatic Cancer Molecular Classifications: From Bulk Genomics to Single Cell Analysis," *International Journal of Molecular Sciences*, vol. 21, no. 8, p. 2814, Apr. 2020, doi: https://doi.org/10.3390/ijms21082814.

[20] L. Haghverdi, A. T. L. Lun, M. D. Morgan, and J. C. Marioni, "Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors," *Nature Biotechnology*, vol. 36, no. 5, pp. 421–427, May 2018, doi: https://doi.org/10.1038/nbt.4091.

[21] J. Eberwine, J.-Y. Sul, T. Bartfai, and J. Kim, "The promise of single-cell sequencing," *Nature Methods*, vol. 11, no. 1, pp. 25–27, Dec. 2013, doi: https://doi.org/10.1038/nmeth.2769.

[22] R. Sandberg, "Entering the era of single-cell transcriptomics in biology and medicine," *Nature Methods*, vol. 11, no. 1, pp. 22–24, Dec. 2013, doi: https://doi.org/10.1038/nmeth.2764.

[23] Matthew H. Spitzer and Garry P. Nolan, "Mass Cytometry: Single Cells, Many Features," *Cell*, vol. 165, no. 4, pp. 780–791, May 2016, doi: https://doi.org/10.1016/j.cell.2016.04.019.

[24] R. Zenobi, "Single-Cell Metabolomics: Analytical and Biological Perspectives," *Science*, vol. 342, no. 6163, Dec. 2013, doi: https://doi.org/10.1126/science.1243259.

[25] E. Marco *et al.*, "Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape," *Proceedings of the National Academy of Sciences*, vol. 111, no. 52, pp. E5643–E5650, Dec. 2014, doi: https://doi.org/10.1073/pnas.1408993111.

[26] I. Tirosh *et al.*, "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq," *Science*, vol. 352, no. 6282, pp. 189–196, Apr. 2016, doi: https://doi.org/10.1126/science.aad0501.

[27] Y. Wang *et al.*, "Clonal evolution in breast cancer revealed by single nucleus genome sequencing," *Nature*, vol. 512, no. 7513, pp. 155–160, Jul. 2014, doi: https://doi.org/10.1038/nature13600.

[28] F. Paul *et al.*, "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors," *Cell*, vol. 163, no. 7, pp. 1663–1677, Dec. 2015, doi: https://doi.org/10.1016/j.cell.2015.11.013.

[29] G. Guo *et al.*, "Mapping Cellular Hierarchy by Single-Cell Analysis of the Cell Surface Repertoire," *Cell Stem Cell*, vol. 13, no. 4, pp. 492–505, Oct. 2013, doi: https://doi.org/10.1016/j.stem.2013.07.017.

[30] I. C. Macaulay *et al.*, "Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells," *Cell Reports*, vol. 14, no. 4, pp. 966–977, Feb. 2016, doi: https://doi.org/10.1016/j.celrep.2015.12.082.

[31] A. Olsson *et al.*, "Single-cell analysis of mixed-lineage states leading to a binary cell fate choice," *Nature*, vol. 537, no. 7622, pp. 698–702, Sep. 2016, doi: https://doi.org/10.1038/nature19348.

[32] F. Buettner *et al.*, "Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells," *Nature Biotechnology*, vol. 33, no. 2, pp. 155–160, Jan. 2015, doi: https://doi.org/10.1038/nbt.3102.

[33] S. Skylaki, O. Hilsenbeck, and T. Schroeder, "Challenges in long-term imaging and quantification of single-cell dynamics," *Nature Biotechnology*, vol. 34, no. 11, pp. 1137–1144, Nov. 2016, doi: https://doi.org/10.1038/nbt.3713.

[34] R. Bacher and C. Kendziorski, "Design and computational analysis of single-cell RNA-sequencing experiments," *Genome Biology*, vol. 17, no. 1, Apr. 2016, doi: https://doi.org/10.1186/s13059-016-0927-y.

[35] Y. Lavin *et al.*, "Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses," *Cell*, vol. 169, no. 4, pp. 750-765.e17, May 2017, doi: https://doi.org/10.1016/j.cell.2017.04.014.

[36] C. Melzer, Y. Yang, and R. Hass, "Interaction of MSC with tumor cells," *Cell Communication and Signaling*, vol. 14, no. 1, Sep. 2016, doi: https://doi.org/10.1186/s12964-016-0143-0.

[37] K.-Y. Ma *et al.*, "Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response–related genes," *JCI Insight*, vol. 4, no. 4, Feb. 2019, doi: https://doi.org/10.1172/jci.insight.121387.

[38] M. Janiszewska *et al.*, "Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments," *Nature Cell Biology*, vol. 21, no. 7, pp. 879–888, Jul. 2019, doi: https://doi.org/10.1038/s41556-019-0346-x.

[39] Q. Wang *et al.*, "Single-cell profiling guided combinatorial immunotherapy for fast-evolving CDK4/6 inhibitor-resistant HER2-positive breast cancer," *Nature Communications*, vol. 10, no. 1, Aug. 2019, doi: https://doi.org/10.1038/s41467-019-11729-1.

[40] I. Xhangolli, B. Dura, G. Lee, D. Kim, Y. Xiao, and R. Fan, "Single-cell Analysis of CAR-T Cell Activation Reveals A Mixed TH1/TH2 Response Independent of Differentiation," *Genomics, Proteomics & Bioinformatics*, vol. 17, no. 2, pp. 129–139, Apr. 2019, doi: https://doi.org/10.1016/j.gpb.2019.03.002.

[41] K. L. Frieda *et al.*, "Synthetic recording and in situ readout of lineage information in single cells," *Nature*, vol. 541, no. 7635, pp. 107–111, Nov. 2016, doi: https://doi.org/10.1038/nature20777.

[42] S. Shah, E. Lubeck, W. Zhou, and L. Cai, "In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus," *Neuron*, vol. 92, no. 2, pp. 342–357, Oct. 2016, doi: https://doi.org/10.1016/j.neuron.2016.10.001.

[43] J. H. Lee *et al.*, "Highly multiplexed subcellular RNA sequencing in situ," *Science (New York, N.Y.)*, vol. 343, no. 6177, pp. 1360–1363, Mar. 2014, doi: https://doi.org/10.1126/science.1250212.

[44] T. Hashimshony, F. Wagner, N. Sher, and I. Yanai, “CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification,” *Cell Reports*, vol. 2, no. 3, pp. 666–673, Sep. 2012, doi: https://doi.org/10.1016/j.celrep.2012.08.003.

[45] D. Ramsköld *et al.*, “Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells,” *Nature Biotechnology*, vol. 30, no. 8, pp. 777–782, Jul. 2012, doi: https://doi.org/10.1038/nbt.2282.

[46] D. A. Jaitin *et al.*, “Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types,” *Science*, vol. 343, no. 6172, pp. 776–779, Feb. 2014, doi: https://doi.org/10.1126/science.1247651.

[47] T. Nagano *et al.*, “Single-cell Hi-C reveals cell-to-cell variability in chromosome structure,” *Nature*, vol. 502, no. 7469, pp. 59–64, Sep. 2013, doi: https://doi.org/10.1038/nature12593.

[48] V. Ramani *et al.*, “Massively multiplex single-cell Hi-C,” *Nature Methods*, vol. 14, no. 3, pp. 263–266, Jan. 2017, doi: https://doi.org/10.1038/nmeth.4155.

[49] R. Aebersold *et al.*, “How many human proteoforms are there?,” *Nature Chemical Biology*, vol. 14, no. 3, pp. 206–214, Mar. 2018, doi: https://doi.org/10.1038/nchembio.2576.

[50] V. Petrosius and E. M. Schoof, “Recent advances in the field of single-cell proteomics,” *Translational Oncology*, vol. 27, p. 101556, Jan. 2023, doi: https://doi.org/10.1016/j.tranon.2022.101556.

[51] D. B. Bekker-Jensen *et al.*, “An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes,” *Cell Systems*, vol. 4, no. 6, pp. 587-599.e4, Jun. 2017, doi: https://doi.org/10.1016/j.cels.2017.05.009.

[52] S. S. Boon *et al.*, “Human papillomavirus type 18 oncoproteins exert their oncogenicity in esophageal and tongue squamous cell carcinoma cell lines distinctly,” *BMC Cancer*, vol. 19, no. 1, Dec. 2019, doi: https://doi.org/10.1186/s12885-019-6413-7.

[53] J. Fan, K. Slowikowski, and F. Zhang, “Single-cell transcriptomics in cancer: computational challenges and opportunities,” *Experimental & Molecular Medicine*, vol. 52, no. 9, pp. 1452–1465, Sep. 2020, doi: https://doi.org/10.1038/s12276-020-0422-0.

[54] S. Müller and A. Diaz, “Single-Cell mRNA Sequencing in Cancer Research: Integrating the Genomic Fingerprint,” *Frontiers in Genetics*, vol. 8, May 2017, doi: https://doi.org/10.3389/fgene.2017.00073.

[55] P. J. Sapienza and A. L. Lee, “Using NMR to study fast dynamics in proteins: methods and applications,” *Current opinion in pharmacology*, vol. 10, no. 6, p. 723, Dec. 2010, doi: https://doi.org/10.1016/j.coph.2010.09.006.

[56] D. R. Bandura *et al.*, “Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry,” *Analytical Chemistry*, vol. 81, no. 16, pp. 6813–6822, Aug. 2009, doi: https://doi.org/10.1021/ac901049w.

[57] O. Ornatsky, D. Bandura, V. Baranov, M. Nitz, M. A. Winnik, and S. Tanner, “Highly multiparametric analysis by mass cytometry,” *Journal of Immunological Methods*, vol. 361, no. 1, pp. 1–20, Sep. 2010, doi: https://doi.org/10.1016/j.jim.2010.07.002.

[58] S. Han *et al.*, "Application value of CyTOF 2 mass cytometer technology at single-cell level in human gastric cancer cells," *Experimental Cell Research*, vol. 384, no. 1, p. 111568, Nov. 2019, doi: https://doi.org/10.1016/j.yexcr.2019.111568.

[59] A. J. Lamble, M. Dietz, T. Laderas, S. McWeeney, and E. F. Lind, "Integrated functional and mass spectrometry-based flow cytometric phenotyping to describe the immune microenvironment in acute myeloid leukemia," *Journal of Immunological Methods*, vol. 453, pp. 44–52, Feb. 2018, doi: https://doi.org/10.1016/j.jim.2017.11.010.

[60] S. C. Bendall *et al.*, "Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum," *Science (New York, N.Y.)*, vol. 332, no. 6030, pp. 687–696, May 2011, doi: https://doi.org/10.1126/science.1198704.

[61] Y. Lei *et al.*, "Applications of single-cell sequencing in cancer research: progress and perspectives," *Journal of Hematology & Oncology*, vol. 14, no. 1, Jun. 2021, doi: https://doi.org/10.1186/s13045-021-01105-2.

[62] A. Koussounadis, S. P. Langdon, I. H. Um, D. J. Harrison, and V. A. Smith, "Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system," *Scientific Reports*, vol. 5, no. 1, Jun. 2015, doi: https://doi.org/10.1038/srep10775.

[63] H. Boekweg and S. Payne, "Challenges and opportunities for single cell computational proteomics," *Molecular & Cellular Proteomics*, vol. 0, no. 0, Feb. 2023, doi: https://doi.org/10.1016/j.mcpro.2023.100518.

[64] N. Slavov, "Great Gains in Mass Spectrometry Data Interpretation," *Journal of Proteome Research*, vol. 22, no. 3, pp. 659–659, Mar. 2023, doi: https://doi.org/10.1021/acs.jproteome.3c00099.

[65] B. Budnik, E. Levy, G. Harmange, and N. Slavov, "SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation," *Genome Biology*, vol. 19, no. 1, Oct. 2018, doi: https://doi.org/10.1186/s13059-018-1547-5.

[66] H. Specht *et al.*, "Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2," *Genome Biology*, vol. 22, no. 1, Jan. 2021, doi: https://doi.org/10.1186/s13059-021-02267-5.

[67] J. Derks *et al.*, "Increasing the throughput of sensitive proteomics by plexDIA," *Nature Biotechnology*, pp. 1–10, Jul. 2022, doi: https://doi.org/10.1038/s41587-022-01389-w.

[68] R. G. Huffman *et al.*, "Prioritized single-cell proteomics reveals molecular and functional polarization across primary macrophages," Mar. 2022, doi: https://doi.org/10.1101/2022.03.16.484655.

[69] H. Chen and P. C. Boutros, "VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R," *BMC Bioinformatics*, vol. 12, no. 1, Jan. 2011, doi: https://doi.org/10.1186/1471-2105-12-35.

[70] A. Bateman *et al.*, "UniProt: the Universal Protein Knowledgebase in 2023," *Nucleic Acids Research*, Nov. 2022, doi: https://doi.org/10.1093/nar/gkac1052.

[71] F. Koopmans *et al.*, "SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse," *Neuron*, vol. 103, no. 2, pp. 217-234.e4, Jul. 2019, doi: https://doi.org/10.1016/j.neuron.2019.05.002.

[72] I. Rivals, L. Personnaz, L. Taing, and M.-C. . Potier, “Enrichment or depletion of a GO category within a class of genes: which test?,” *Bioinformatics*, vol. 23, no. 4, pp. 401–407, Dec. 2006, doi: https://doi.org/10.1093/bioinformatics/btl633.

[73] S. Lê, J. Josse, and F. Husson, “FactoMineR: AnRPackage for Multivariate Analysis,” *Journal of Statistical Software*, vol. 25, no. 1, 2008, doi: https://doi.org/10.18637/jss.v025.i01.

[74] L. McInnes, J. Healy, and J. Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,” *arXiv.org*, 2018. https://arxiv.org/abs/1802.03426

[75] A. Kassambara and F. Mundt, “factoextra: Extract and Visualize the Results of Multivariate Data Analyses,” *R-Packages*, Apr. 01, 2020. https://cran.r-project.org/package=factoextra

[76] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, “clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters,” *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, May 2012, doi: https://doi.org/10.1089/omi.2011.0118.

[77] G. Yu, *Introduction | Biomedical Knowledge Mining using GOSemSim and clusterProfiler*. Available: https://yulab-smu.top/biomedical-knowledge-mining-book/

[78] D. Szklarczyk *et al.*, “STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” *Nucleic Acids Research*, vol. 47, no. Database issue, pp. D607–D613, Jan. 2019, doi: https://doi.org/10.1093/nar/gky1131.

[79] Taiyun, “taiyun/corrplot,” *GitHub*, Dec. 07, 2022. https://github.com/taiyun/corrplot (accessed Jan. 06, 2023).

[80] C. Vaklavas, S. W. Blume, and W. E. Grizzle, “Translational Dysregulation in Cancer: Molecular Insights and Potential Clinical Applications in Biomarker Development,” *Frontiers in Oncology*, vol. 7, Jul. 2017, doi: https://doi.org/10.3389/fonc.2017.00158.

[81] L. Boussemart *et al.*, “eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies,” *Nature*, vol. 513, no. 7516, pp. 105–109, Sep. 2014, doi: https://doi.org/10.1038/nature13572.

[82] M. L. Truitt *et al.*, “Differential Requirements for eIF4E Dose in Normal Development and Cancer,” *Cell*, vol. 162, no. 1, pp. 59–71, Jul. 2015, doi: https://doi.org/10.1016/j.cell.2015.05.049.

[83] S. H. Reich *et al.*, “Structure-based Design of Pyridone–Aminal eFT508 Targeting Dysregulated Translation by Selective Mitogen-activated Protein Kinase Interacting Kinases 1 and 2 (MNK1/2) Inhibition,” *Journal of Medicinal Chemistry*, vol. 61, no. 8, pp. 3516–3540, Mar. 2018, doi: https://doi.org/10.1021/acs.jmedchem.7b01795.

[84] D. Silvera, S. C. Formenti, and R. J. Schneider, “Translational control in cancer,” *Nature Reviews. Cancer*, vol. 10, no. 4, pp. 254–266, Apr. 2010, doi: https://doi.org/10.1038/nrc2824.

[85] E. Martínez-Salas, G. Lozano, J. Fernandez-Chamorro, R. Francisco-Velilla, A. Galan, and R. Diaz, “RNA-Binding Proteins Impacting on Internal Initiation of Translation,” *International Journal of Molecular Sciences*, vol. 14, no. 11, pp. 21705–21726, Nov. 2013, doi: https://doi.org/10.3390/ijms141121705.

[86] B. Zhang *et al.*, "A comprehensive expression landscape of RNA-binding proteins (RBPs) across 16 human cancer types," *RNA Biology*, vol. 17, no. 2, pp. 211–226, Oct. 2019, doi: https://doi.org/10.1080/15476286.2019.1673657.

[87] A. E. Kudinov, J. Karanicolas, E. A. Golemis, and Y. Boumber, "Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets," *Clinical Cancer Research*, vol. 23, no. 9, pp. 2143–2153, Apr. 2017, doi: https://doi.org/10.1158/1078-0432.ccr-16-2728.

[88] L. P. Vu *et al.*, "Functional screen of MSI2 interactors identifies an essential role for SYNCRIP in myeloid leukemia stem cells," *Nature Genetics*, vol. 49, no. 6, pp. 866–875, Jun. 2017, doi: https://doi.org/10.1038/ng.3854.

[89] A. D'Ambrogio, K. Nagaoka, and J. D. Richter, "Translational control of cell growth and malignancy by the CPEBs," *Nature Reviews Cancer*, vol. 13, no. 4, pp. 283–290, Feb. 2013, doi: https://doi.org/10.1038/nrc3485.

[90] P. Song, F. Yang, H. Jin, and X. Wang, "The regulation of protein translation and its implications for cancer," *Signal Transduction and Targeted Therapy*, vol. 6, no. 1, pp. 1–9, Feb. 2021, doi: https://doi.org/10.1038/s41392-020-00444-9.

[91] N. Kondrashov *et al.*, "Ribosome-Mediated Specificity in Hox mRNA Translation and Vertebrate Tissue Patterning," *Cell*, vol. 145, no. 3, pp. 383–397, Apr. 2011, doi: https://doi.org/10.1016/j.cell.2011.03.028.

[92] G. R. Kardos, M.-S. Dai, and G. P. Robertson, "Growth inhibitory effects of large subunit ribosomal proteins in melanoma," *Pigment Cell & Melanoma Research*, vol. 27, no. 5, pp. 801–812, Jun. 2014, doi: https://doi.org/10.1111/pcmr.12259.

[93] Y. Zhu *et al.*, "RPL34-AS1-induced RPL34 inhibits cervical cancer cell tumorigenesis via the MDM2-P53 pathway," *Cancer Science*, vol. 112, no. 5, pp. 1811–1821, May 2021, doi: https://doi.org/10.1111/cas.14874.

[94] G. LI *et al.*, "shRNA-mediated RPS15A silencing inhibits U937 acute myeloid leukemia cell proliferation and enhances apoptosis," *Molecular Medicine Reports*, vol. 13, no. 5, pp. 4400–4406, Mar. 2016, doi: https://doi.org/10.3892/mmr.2016.5064.

[95] Y. Wu *et al.*, "ITGA6 and RPSA synergistically promote pancreatic cancer invasion and metastasis via PI3K and MAPK signaling pathways," *Experimental Cell Research*, vol. 379, no. 1, pp. 30–47, Jun. 2019, doi: https://doi.org/10.1016/j.yexcr.2019.03.022.

[96] T. Lefebvre *et al.*, "TRPM7/RPSA Complex Regulates Pancreatic Cancer Cell Migration," *Frontiers in Cell and Developmental Biology*, vol. 8, Jul. 2020, doi: https://doi.org/10.3389/fcell.2020.00549.

[97] F. Wei *et al.*, "Ribosomal protein L34 promotes the proliferation, invasion and metastasis of pancreatic cancer cells," *Oncotarget*, vol. 7, no. 51, pp. 85259–85272, Nov. 2016, doi: https://doi.org/10.18632/oncotarget.13269.

[98] J. Liang, Y. Liu, L. Zhang, J. Tan, E. Li, and F. Li, "Overexpression of microRNA-519d-3p suppressed the growth of pancreatic cancer cells by inhibiting ribosomal protein S15A-mediated Wnt/β-catenin signaling," *Chemico-Biological Interactions*, vol. 304, pp. 1–9, May 2019, doi: https://doi.org/10.1016/j.cbi.2019.02.026.

[99] J. R. Mills *et al.*, “RNAi screening uncovers Dhx9 as a modifier of ABT-737 resistance in an Eμ-myc/Bcl-2 mouse model,” *Blood*, vol. 121, no. 17, pp. 3402–3412, Apr. 2013, doi: https://doi.org/10.1182/blood-2012-06-434365.

[100] F. Robert *et al.*, “Translation initiation factor eIF4F modifies the dexamethasone response in multiple myeloma,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 37, pp. 13421–13426, Sep. 2014, doi: https://doi.org/10.1073/pnas.1402650111.

[101] T. Lee and J. Pelletier, “The biology of DHX9 and its potential as a therapeutic target,” *Oncotarget*, vol. 7, no. 27, Mar. 2016, doi: https://doi.org/10.18632/oncotarget.8446.

[102] Z. Sun *et al.*, “Conserved recurrent gene mutations correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers,” *BMC medical genomics*, vol. 7, p. 32, Jun. 2014, doi: https://doi.org/10.1186/1755-8794-7-32.

[103] Y-h. Taguchi and H. Wang, “Genetic Association between Amyotrophic Lateral Sclerosis and Cancer,” *Genes*, vol. 8, no. 10, p. 243, Sep. 2017, doi: https://doi.org/10.3390/genes8100243.

[104] M. Yamaguchi, Y. Azuma, and H. Yoshida, “ALS and Cancer,” *Journal of Carcinogenesis & Mutagenesis*, vol. 07, no. 06, 2016, doi: https://doi.org/10.4172/2157-2518.1000e122.

[105] U. Abramczyk, M. Nowaczyński, A. Słomczyński, P. Wojnicz, P. Zatyka, and A. Kuzan, “Consequences of COVID-19 for the Pancreas,” *International Journal of Molecular Sciences*, vol. 23, no. 2, p. 864, Jan. 2022, doi: https://doi.org/10.3390/ijms23020864.

[106] P. A. van Dam *et al.*, “SARS-CoV-2 and cancer: Are they really partners in crime?,” *Cancer Treatment Reviews*, vol. 89, p. 102068, Sep. 2020, doi: https://doi.org/10.1016/j.ctrv.2020.102068.

[107] S. K. Boddu, G. Aurangabadkar, and M. S. Kuchay, “New onset diabetes, type 1 diabetes and COVID-19,” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 6, pp. 2211–2217, Nov. 2020, doi: https://doi.org/10.1016/j.dsx.2020.11.012.

[108] P. Katopodis *et al.*, “Pan-cancer analysis of transmembrane protease serine 2 and cathepsin L that mediate cellular SARS-CoV-2 infection leading to COVID-19,” *International Journal of Oncology*, vol. 57, no. 2, pp. 533–539, Aug. 2020, doi: https://doi.org/10.3892/ijo.2020.5071.

[109] S. Becker *et al.*, “COVID-19 can lead to rapid progression of cervical intraepithelial neoplasia by dysregulating the immune system: A hypothesis,” *Journal of Reproductive Immunology*, vol. 154, p. 103763, Dec. 2022, doi: https://doi.org/10.1016/j.jri.2022.103763.

[110] E. Vavoulidis, C. Margioula-Siarkou, S. Petousis, and K. Dinas, “SARS-CoV-2 infection and impact on female genital tract: An untested hypothesis,” *Medical Hypotheses*, vol. 144, p. 110162, Nov. 2020, doi: https://doi.org/10.1016/j.mehy.2020.110162.

[111] O. Ondič *et al.*, “SARS-CoV-2 RNA may rarely be present in a uterine cervix LBC sample at the asymptomatic early stage of COVID 19 disease,” *Cytopathology: Official Journal of the British Society for Clinical Cytology*, vol. 32, no. 6, pp. 766–770, Nov. 2021, doi: https://doi.org/10.1111/cyt.12995.

[112] N. Ghosh, I. Saha, and Dariusz Plewczynski, “Unveiling the Biomarkers of Cancer and COVID-19 and Their Regulations in Different Organs by Integrating RNA-Seq Expression and Protein–Protein Interactions,” *ACS omega*, vol. 7, no. 48, pp. 43589–43602, Nov. 2022, doi: https://doi.org/10.1021/acsomega.2c04389.

[113] A. B. Leduc, R. Huffman, J. Cantlon, S. A. Khan, and Nikolai Slavov, “Exploring functional protein covariation across single cells using nPOP,” *Genome Biology*, vol. 23, no. 1, Dec. 2022, doi: https://doi.org/10.1186/s13059-022-02817-5.

[114] A. A. Petelski *et al.*, “Multiplexed single-cell proteomics using SCoPE2,” Mar. 2021, doi: https://doi.org/10.1101/2021.03.12.435034.

[115] M. Yang, R. Nelson, and A. Ros, “Toward Analysis of Proteins in Single Cells: A Quantitative Approach Employing Isobaric Tags with MALDI Mass Spectrometry Realized with a Microfluidic Platform,” *Analytical Chemistry*, vol. 88, no. 13, pp. 6672–6679, Jun. 2016, doi: https://doi.org/10.1021/acs.analchem.5b03419.

[116] P. Zhao, S. Bhowmick, J. Yu, and J. Wang, “Highly Multiplexed Single-Cell Protein Profiling with Large-Scale Convertible DNA-Antibody Barcoded Arrays,” *Advanced Science*, vol. 5, no. 9, p. 1800672, Aug. 2018, doi: https://doi.org/10.1002/advs.201800672.

[117] S. K. Saka *et al.*, “Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues,” *Nature Biotechnology*, vol. 37, no. 9, pp. 1080–1090, Sep. 2019, doi: https://doi.org/10.1038/s41587-019-0207-y.

[118] A. A. Petelski *et al.*, “Multiplexed single-cell proteomics using SCoPE2,” *Nature Protocols*, vol. 16, no. 12, pp. 5398–5425, Dec. 2021, doi: https://doi.org/10.1038/s41596-021-00616-z.

[119] R. Chen, L. Mao, R. Shi, W. Wang, and J. Cheng, “circRNA MYLK Accelerates Cervical Cancer via Up-Regulation of RHEB and Activation of mTOR Signaling,” *Cancer Management and Research*, vol. Volume 12, pp. 3611–3621, May 2020, doi: https://doi.org/10.2147/cmar.s238172.