UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

Guilherme Reis de Oliveira

OmicScope: from quantitative proteomics to Systems Biology

OmicScope: da proteômica quantitativa à biologia de sistemas

CAMPINAS
2024

Guilherme Reis de Oliveira

OmicScope: from quantitative proteomics to Systems Biology
OmicScope: da proteômica quantitativa à biologia de sistemas

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor of Genetics and Molecular Biology, specializing in Bioinformatics.

Tese Dissertação apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Doutor em Genética e Biologia Molecular, na área de Bioinformática.

Orientador: Daniel Martins de Souza

Co-Orientador: Alessandro dos Santos Farias

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO GUILHERME REIS DE OLIVEIRA E ORIENTADA PELO PROF. DANIEL MARTINS DE SOUZA.

CAMPINAS
2024

Informações Complementares

Título em outro idioma: OmicScope : da proteômica quantitativa à biologia de sistemas
Palavras-chave em inglês:
Proteomics
Systems biology
Data science
Área de concentração: Bioinformática
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
Daniel Martins de Souza [Orientador]
Helder Takashi Imoto Nakaya
Mariana Lima Boroni Martins
Paulo Costa Carvalho
Thiago Verano-Braga
Data de defesa: 18-03-2024
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-7696-0716
- Currículo Lattes do autor: http://lattes.cnpq.br/2234045116362023

Campinas, 18 de Março de 2024.

COMISSÃO EXAMINADORA

Prof. Dr. Daniel Martins de Souza

Prof. Dr. Helder Takashi Imoto Nakaya

Profa. Dra. Mariana Lima Boroni Martins

Prof. Dr. Paulo Costa Carvalho

Prof. Dr. Thiago Verano-Braga

Os membros da Comissão Examinadora acima assinaram a Ata de defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Tese e na Secretaria do Programa de Genética e Biologia Molecular do Instituto de Biologia.

You can avoid reality, but you cannot avoid the consequences of avoiding reality.

– Ayn Rand

Agradecimentos

Primeiramente, é necessário agradecer àqueles que apostam tudo, todos os dias, em um investimento sem retorno financeiro e alto grau de incerteza: meu pai e minha mãe. Ao meu pai, Adejair/Degano, agradeço seus milhares de quilômetros rodados, suas madrugadas de cuidado, e a cada centavo batalhado e investidos na sua família. À minha mãe, Regina/Guê, agradeço por cada abraço, risada, olhar e comida, ao seu infinito amor e a nossa amizade, que quase me faz acreditar em vidas passadas, almas e deuses.

Ainda no âmbito do laço sanguíneo, agradeço à minha irmã e ao meu irmão, Luana e Silas, por me ensinarem a lidar e aceitar o diferente e, obviamente, agradeço também a paciência dos dois em aceitar minhas diferenças. À Luana, ainda, agradeço por me ajudar a ver o lado bom das pessoas acima de tudo. Finalizando essa etapa, agradeço aos demais familiares (também conhecido como parentes, para aqueles que possuem menos afeto ou tem apego aos termos etimológicos), que estão comigo em corpo ou memória, por propiciarem um ambiente construtivo e carinhoso para que um ser humano se desenvolva em plenitude.

Avançando em tópicos, mas de forma alguma influenciando a ordem de importância, agradeço ao meu companheiro, Gilmar, por ser minha maior conquista. Especialmente, agradeço o seu amor de todos os dias, que se manifesta no cuidado, nas taças de vinho, nas nossas risadas diárias, na trilha sonora, na insistência com o exercício físico, nas nossas incríveis refeições, e nos finais de semana que dedicamos exclusivamente a nós.

No âmbito acadêmico, começo agradecendo ao Prof. Daniel, por abrir as portas do seu laboratório em 2015 a um aluno de graduação em biologia, cheio de sonhos, vontades e certezas. A ele sou grato pelas oportunidades, discussões, e por praticamente todas as pessoas que citarei e agradecerei abaixo, sem ele essa tese seria completamente diferente. Certa vez, ele me disse "meu laboratório é como um barco e eu sou o capitão. Mas nesse barco não cabe toda a tripulação, então alguns têm que ficar nadando e, quando eu vejo que estão para se afogar [em prazos ou projetos], eu resgato e outro tem que ir para a água! Você eu geralmente deixo nadando". Hoje, revisitando essa anedota,

vejo que nadei para locais onde as boias nem sempre chegavam. Agradeço ao Daniel pela paciência com o marujo enérgico e, por nesse momento, me oferecer um lugar no barco para chegarmos em terra firme.

Agradeço, também, aos meus amigos e colegas do Laboratório de Neuroproteômica, que são parte fundamental da minha história e desta tese. Especialmente, agradeço à primeira pessoa que me acolheu, minha amiga Verônica, que me ensinou os primeiros passos no mundo da proteômica (do cultivo de células à análise de dados), sobre relacionamentos no trabalho e sobre filosofia científica; também ao Paulinho pelos seus ensinamentos práticos, pelas piadas e por ser um exemplo de pessoa que trabalha pela família; à minha amiga Valéria por me ensinar sobre farmacologia, importância de "sentar e ler", e sobre o valor da resiliência.

Além disso, também gostaria de agradecer aos meus amigos Mariana e Fábio, pela sua generosidade em me doar seu tempo e imensurável conhecimento, eles sempre serão grandes exemplos de virtude e conhecimento. Ainda no âmbito da proteômica, e de forma especial, agradeço ao meu amigo Victor/Little, por dedicar seu tempo a discutir ideias implementadas nessa tese, ler meus textos, oferecer *feedbacks*, dar muitas risadas, fazer planos, e das nossas eternas discussões políticas de altíssimo nível.

A todos os amigos que vida me proporcionou, dentro e fora da Unicamp, pelo suporte intelectual e emocional, pelas milhares de gargalhadas, drinks, apoio e acolhimento.

A todos os colaboradores, cientistas, banca, revisores e assessores envolvidos neste trabalho e em outros que pude participar até este momento.

Ao Instituto de Biologia da Unicamp, pelo suporte estrutural e financeiro oferecido a mim e a todos os alunos, tendo sido praticamente minha casa durante 10 anos da minha vida. Especialmente, agradeço ao funcionário Marcos Akira, da Informática, por dar todo o suporte para que eu disponibilizasse o OmicScope no servidor do IB.

Resumo

A análise em larga escala de proteínas tem avançado significativamente a nossa compreensão sobre fenômenos naturais, da busca por biomarcadores à descoberta de novos mecanismos biológicos. Neste contexto, a proteômica baseada em espectrometria de massas se destaca como a principal abordagem para análise quantitativa de proteínas em sistemas biológicos. Embora apresente uma vantagem intrínseca na geração de informação, analisar essa extensa quantidade de dados implica diversos desafios que demandam grande investimento de tempo, como realizar análises de proteômica diferencial, análises de enriquecimento, busca por interação proteína-proteína, integração de resultados independentes, e outras aplicações. Nesta tese, apresentamos o OmicScope, um pacote desenvolvido para as análises de dados que sucedem os processos de identificação e quantificação de proteínas. O OmicScope é dividido em três componentes principais que realizam, respectivamente, proteômica diferencial, análise de enriquecimento e meta-análise. Cada módulo oferece seu próprio conjunto de funções, que incluem a geração de figuras e de redes, permitindo a análise dos dados a nível de sistemas. Além de sua distribuição como pacote Python, o OmicScope também pode ser acessado como um aplicativo web através do endereço https://omicscope.ib.unicamp.br/. Para demonstrar as melhorias trazidas pelo OmicScope ao nosso laboratório, dois artigos originais desenvolvidos pelo nosso grupo foram incluídos nessa tese, exemplificando os avanços no nosso fluxo de análise de dados e a apresentação dos resultados, tanto antes quanto durante o desenvolvimento da ferramenta.

Abstract

The large-scale analysis of proteins has significantly advanced our understanding of biological phenomena, ranging from the search for biomarkers to discovery of new biochemical mechanisms. In this context, mass-spectrometry-based proteomics stands out as the primary approach for quantitatively analyzing proteins in biological systems. Despite its intrinsic advantage in generating data, analyzing this data presents challenges, such as the expressive investment of time to carry out differential proteomics and enrichment analysis, exploring protein-protein interactions, integrating data with independent results, and other applications. This thesis introduces OmicScope, a package developed for downstream data analysis following the protein identification and quantitation process. OmicScope is divided into three major components to perform, respectively, differential proteomics, enrichment analysis, and meta-analysis. Each mode offers a range of functions, including the generation of publication quality figures and networks, facilitating systems-level data analysis. In addition to its distribution as a Python package, OmicScope is also available as a web app at https://omicscope.ib.unicamp.br/. To showcase the enhancements brought by OmicScope to our laboratory, two original articles developed by our group are included, exemplifying the advancements in our data analysis pipeline and the presentation of results both before and during the tool's development.

List of figures

Principal Figures

Supplementary Figures

List of tables

List of abbreviations

2D-PAGE	Two-Dimensional Polyacrylamide Gel Electrophoresis
ACN	Acetonitrile
AmBic	Ammonium Bicarbonate
ANOVA	Analysis of Variance
BBB	Blood-Brain-Barrier
CAU	Caudate Nucleus
CER	Cerebellum
CID	Collision Induced Energy
CNS	Central Nervous System
COVID-19	Coronavirus Disease
DAVID	Database for Annotation, Visualization and Integrated Discovery
DDA	Data-Dependent Acquisition
DIA	Data-Independent Acquisition
DRPs	Differentially Regulated Proteins
ESI	Electrospray Ionization
FA	Formic acid
FDR	False Discovery Rate
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GUI	Graphical User Interfaces
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid Chromatography coupled to MS
LFQ	Label-Free Quantitation
MALDI	Matrix-Assisted Laser Desorption Ionization
MS	Mass Spectrometry
NES	Normalized Enrichment Score
NMDA	N-Methyl-d-Aspartate
NSAF	Normalized Spectral Abundance Factor
NSC	Neural Stem Cell

ORA	Over-Representation Analysis
PCA	Principal Component Analysis
PCC	Posterior Cingulate Cortex
PPI	Protein-Protein Interaction
PSM	Peptide Spectral Match
Q	Quadrupole
RP-LC	Reversed Phase Liquid Chromatography
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SILAC	Stable Isotope Labeling by Amino Acids in Cell Culture
SVG	Scalable Vector Graphics
TOF	Time-Of-Flight
UPLC	Ultra High-Performance Liquid Chromatography
XIC	Extracted Ion Chromatogram

Contents

1   GENERAL INTRODUCTION

1.1   INTRODUCTION TO MS-BASED PROTEOMICS

The term "proteome" was coined by Marc Wilkins to describe the "total complement of proteins that can be encoded by a given genome"[1]. While the term was introduced in 1995, the large-scale analysis of proteins dates back decades to the use of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)[2]. This technique involves loading a complex protein mixture into a gel, allowing the separation of individual proteins based on their isoelectric point and molecular weight. Experimental outcomes comprise images of gels potentially containing thousands of spots, each one theoretically corresponding to a single protein. However, despite its utility, 2D-PAGE has limitations, including low reproducibility between gel replicates and experiments, as well as the inability to directly identify proteins within each spot[3]. The last limitation was addressed later by integrating 2D-PAGE with advanced technologies like Matrix-Assisted Laser Desorption Ionization (MALDI)[4] and Electrospray Ionization (ESI)[5], enabling the ionization of large molecules for subsequent analysis via Mass Spectrometry (MS). The introduction of MS into protein analysis has revolutionized our understanding about the proteome. Coupled with technological advancements and computational tools, this technique has enabled the large-scale identification of proteins in paths previously unexplored.

Presently, MS-based proteomics has evolved beyond the necessity for the 2D-PAGE technique, enabling in-depth proteome characterization and offering researchers versatility in addressing diverse inquiries. This technique facilitates large-scale

---

[1] Valerie C. Wasinger et al., "Progress with Gene-Product Mapping of the Mollicutes: Mycoplasma Genitalium," *ELECTROPHORESIS* 16, no. 1 (1995): 1090–94, https://doi.org/10.1002/elps.11501601185.

[2] Patricia Kahn, "From Genome to Proteome: Looking at a Cell's Proteins," *Science* 270, no. 5235 (October 20, 1995): 369–70, https://doi.org/10.1126/science.270.5235.369.

[3] Shao-En Ong and Matthias Mann, "Mass Spectrometry–Based Proteomics Turns Quantitative," *Nature Chemical Biology* 1, no. 5 (October 2005): 252–62, https://doi.org/10.1038/nchembio736.

[4] Koichi Tanaka et al., "Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry," *Rapid Communications in Mass Spectrometry* 2, no. 8 (1988): 151–53, https://doi.org/10.1002/rcm.1290020802.

[5] John B. Fenn et al., "Electrospray Ionization for Mass Spectrometry of Large Biomolecules," *Science* 246, no. 4926 (October 6, 1989): 64–71, https://doi.org/10.1126/science.2675315.

exploration of protein-protein interactions[6], protein structures[7], post-translational modifications (PTMs)[8], drug-protein interactions[9], protein stability[10], and more. In proteomics, these applications are achieved through two primary approaches: top-down and bottom-up proteomics[11].

In top-down approach, proteins are analyzed under native conditions, providing valuable insights into PTMs and proteoforms. However, studying intact proteins introduces some technical challenges, elevated costs, limitations in proteome depth, and complexities in spectra data analysis[12]. In contrast, bottom-up proteomics infers protein information through peptide analysis (Figure 1). Peptides, being more easily fractionated, ionized, and fragmented, enhance sensitivity and the proteome coverage. For these reasons, bottom-up approach is established as the predominant pipeline for protein large-scale analysis.

Briefly, the backbone of the bottom-up approach involves[13]: 1) extracting proteins from the target biological system (e.g., plasma, cell extract, tissue), 2) enzymatic digestion of proteins, 3) analyzing peptides using Liquid Chromatography coupled to MS (LC-MS), 4) identifying and quantifying proteins, and 5) conducting differential proteomics analysis. Additionally, enrichment analysis is commonly integrated as a sixth step in most biological investigations. Despite each step in the workflow being essential to guarantee high quality results, LC-MS analysis plays a central role as it yields the raw results.

[6] Fan Liu et al., "The Interactome of Intact Mitochondria by Cross-Linking Mass Spectrometry Provides Evidence for Coexisting Respiratory Supercomplexes," *Molecular & Cellular Proteomics: MCP* 17, no. 2 (February 2018): 216–32, https://doi.org/10.1074/mcp.RA117.000470.

[7] Aline M. Santos et al., "FERM Domain Interaction with Myosin Negatively Regulates FAK in Cardiomyocyte Hypertrophy," *Nature Chemical Biology* 8, no. 1 (January 2012): 102–10, https://doi.org/10.1038/nchembio.717.

[8] Elise J. Needham et al., "Personalized Phosphoproteomics Identifies Functional Signaling," *Nature Biotechnology* 40, no. 4 (April 2022): 576–84, https://doi.org/10.1038/s41587-021-01099-9.

[9] Severin Lechner et al., "Chemoproteomic Target Deconvolution Reveals Histone Deacetylases as Targets of (R)-Lipoic Acid," *Nature Communications* 14, no. 1 (June 15, 2023): 3548, https://doi.org/10.1038/s41467-023-39151-8.

[10] Mikhail M. Savitski et al., "Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome," *Science* 346, no. 6205 (October 3, 2014): 1255784, https://doi.org/10.1126/science.1255784.

[11] Yaoyang Zhang et al., "Protein Analysis by Shotgun/Bottom-up Proteomics," *Chemical Reviews* 113, no. 4 (April 10, 2013): 2343–94, https://doi.org/10.1021/cr3003533.

[12] Zhang et al.; Timothy K. Toby et al., "A Comprehensive Pipeline for Translational Top-down Proteomics from a Single Blood Draw," *Nature Protocols* 14, no. 1 (January 2019): 119–52, https://doi.org/10.1038/s41596-018-0085-7.

[13] Steven R. Shuken, "An Introduction to Mass Spectrometry-Based Proteomics," *Journal of Proteome Research* 22, no. 7 (July 7, 2023): 2151–71, https://doi.org/10.1021/acs.jproteome.2c00838.
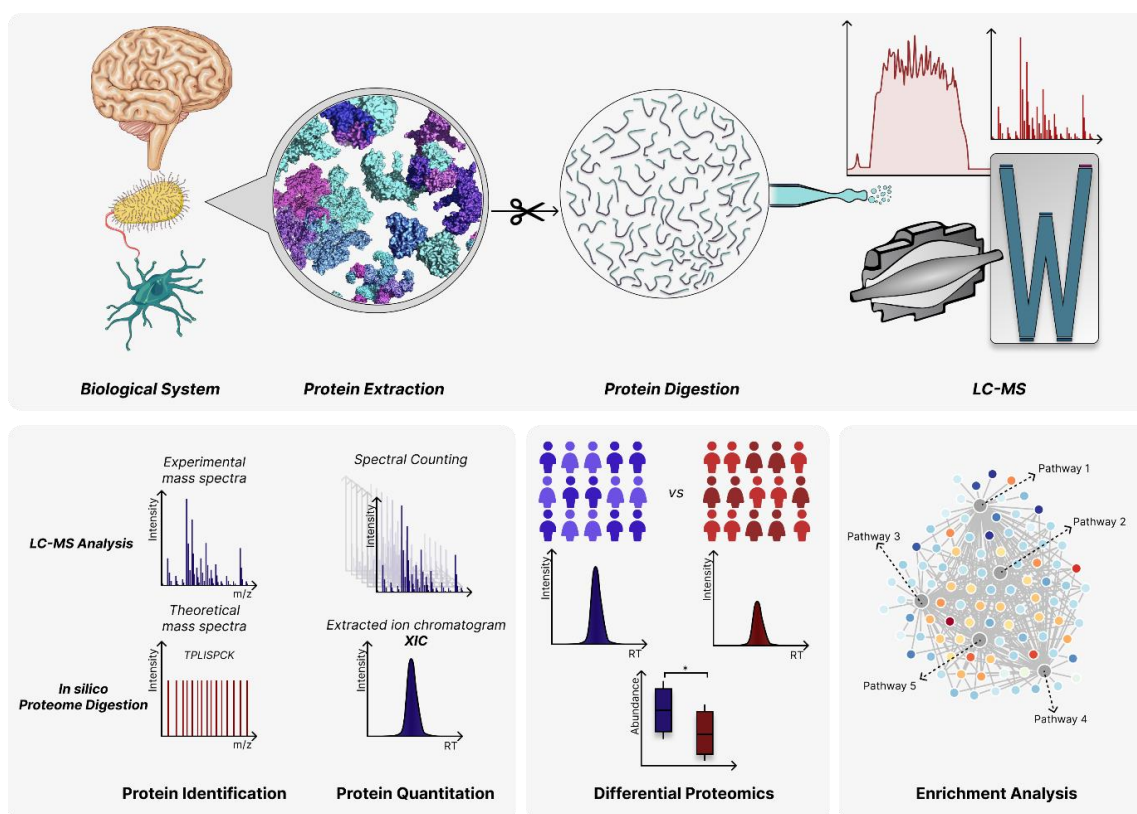
Figure 1-1 General bottom-up proteomics pipeline. The proteomics pipeline initiates with the protein extraction process, followed by protein digestion and peptide analysis using LC-MS platforms. The computational step involves conducting protein identification and quantitation. Subsequent analyses encompass differential proteomics to identify differentially regulated proteins, and, when applied, enrichment analysis.

Due to the elevated sample complexity in proteomics research[14], MS setups commonly incorporate nano Ultra High-Performance Liquid Chromatography (UPLC) for the prior separation of peptides via reversed-phase chromatography (RP-LC). This setup decreases sample complexity, avoiding ion suppression and increasing technical sensitivity[15]. The peptides loaded onto the UPLC are subsequently analyzed in the MS equipment, which comprises three main components: the ion source, mass analyzer(s), and detector. In the first component, often the ESI source, molecules are ionized and transitioned to gaseous state, enabling the ion flighting and subsequent analysis on mass analyzers. Mass analyzers work discriminating ions according to their mass-to-charge ratios, thereby determining the advantages and limitations of MS equipment. In

---

[14] Shuken.

[15] Matthias Wilm and Matthias Mann, "Analytical Properties of the Nanoelectrospray Ion Source," *Analytical Chemistry* 68, no. 1 (January 1, 1996): 1–8, https://doi.org/10.1021/ac9509519; Yangyang Bian et al., "Robust, Reproducible and Quantitative Analysis of Thousands of Proteomes by Micro-Flow LC–MS/MS," *Nature Communications* 11, no. 1 (January 9, 2020): 157, https://doi.org/10.1038/s41467-019-13973-x.

large-scale proteomics applications, MS systems typically incorporate a quadrupole (Q) followed by a collision chamber and a high-resolution mass analyzer, such as time-of-flight (TOF) or Orbitrap. Additionally, modern MS configurations, known as tribrids, feature three mass analyzers, such as Q-LTQ-Orbitrap or Q-Orbitrap-TOF[16]. This modular nature of MS provides a highly versatile tool for conducting proteomics experiments, enabling the equipment to acquire mass spectra using diverse acquisition modes.

In large-scale proteomics, acquisition modes segregate into two categories: Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) (Figure 2). Data-Dependent Acquisition allows the direct association between precursors and their transitions by initially acquiring precursor ions mass spectrum (MS, full scan), followed by the selection of specific precursors for fragmentation, and subsequent transition ions acquisition (MS/MS or tandem-MS)[17]. On the other hand, DIA is a class of several MS/MS experiments that do not require the selection of particular precursor ions but rather simultaneous fragmentation of multiple precursors[18]. Classical examples within this category include SWATH[19], where precursor ions are selected based on a window mass range, and MS[E20], which alternates between low and high energy without utilizing any window cutoff. In addition to inherent differences related to manufacturers, data generated by DIA and DDA exhibit marked dissimilarities concerning precursor and transition assignments, requiring the utilization of distinct computational tools for subsequent peptide and protein identification.

---

[16] Hamish I. Stewart et al., "Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis," *Analytical Chemistry* 95, no. 42 (October 24, 2023): 15656−64, https://doi.org/10.1021/acs.analchem.3c02856.

[17] Reta Birhanu Kitata, Jhih-Ci Yang, and Yu-Ju Chen, "Advances in Data-Independent Acquisition Mass Spectrometry towards Comprehensive Digital Proteome Landscape," *Mass Spectrometry Reviews* 42, no. 6 (2023): 2324−48, https://doi.org/10.1002/mas.21781.

[18] John D. Venable et al., "Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra," *Nature Methods* 1, no. 1 (October 2004): 39–45, https://doi.org/10.1038/nmeth705; Kitata, Yang, and Chen, "Advances in Data-Independent Acquisition Mass Spectrometry towards Comprehensive Digital Proteome Landscape."

[19] Ludovic C. Gillet et al., "Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis," *Molecular & Cellular Proteomics : MCP* 11, no. 6 (June 2012): O111.016717, https://doi.org/10.1074/mcp.O111.016717.

[20] Jeffrey C. Silva et al., "Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs," *Analytical Chemistry* 77, no. 7 (April 1, 2005): 2187−2200, https://doi.org/10.1021/ac048455k.
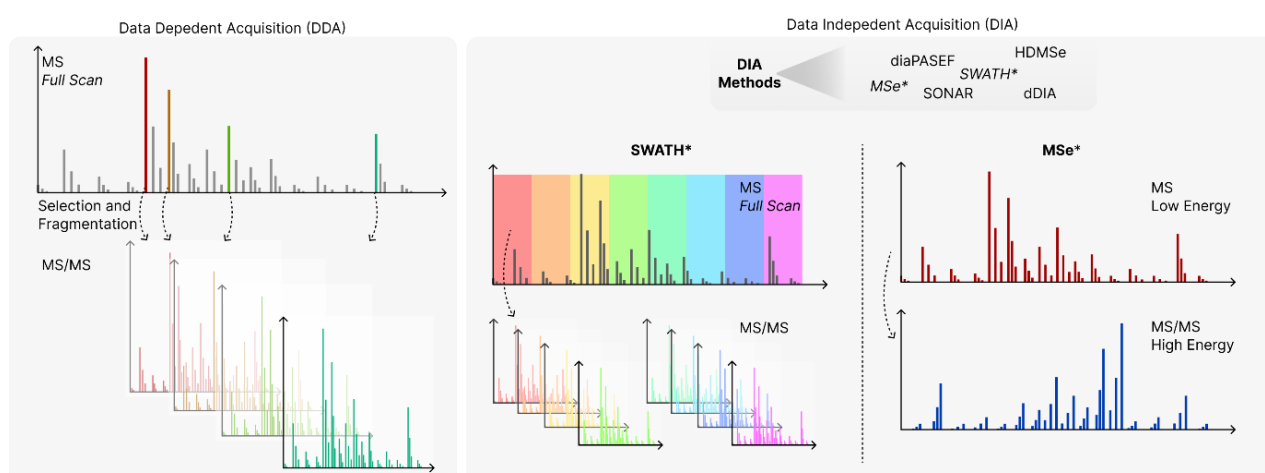
Figure 1-2 Acquisition methods in proteomics. Data-dependent acquisition (DDA) involves an initial full scan to determine precursor ions and their intensities, followed by the selection and fragmentation of the most intense precursors. Data-independent acquisition encompasses various MS/MS experiments without prior selection of precursors. Examples include SWATH and MS^E, which select based on a mass window and alternate between low and high energies, respectively.

In bottom-up proteomics, mass spectra stem from peptides obtained through sample digestion, allowing this approach to infer the presence of proteins in a sample via peptide identification. To carry out peptide assignment, the search engine requires a proteome database containing protein sequences from the target organisms in a FASTA format. This reference proteome undergoes *in silico* digestion considering experimental conditions, such as the use of trypsin and reducing reagents. Trypsin cleaves proteins at arginine (R) and lysine (K) residues, leading search engines to calculate precursor monoisotopic mass. Considering fragments potentially generated in collision chamber, such as –y and –b fragments for Collision Induced Dissociation (CID) experiments, search engines also generate theoretical mass spectra based on peptide sequence. Then, computational tools match theoretical and experimental mass spectra result and generate a peptide spectral match (PSM) associated with a similarity score (Figure 3).

Since score varies according to run and search engine, quality control and statistical analysis for peptide identification often rely on the target-decoy approach[21]. This method involves reversing or scrambling protein sequences, creating nonexistent sequences (decoy sequences) that are also searched against experimental mass spectra.

---

[21] Joshua E. Elias and Steven P. Gygi, "Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry," *Nature Methods* 4, no. 3 (March 2007): 207–14, https://doi.org/10.1038/nmeth1019.

The distribution of PSM scores from both target and decoy sequences helps estimate and filter peptides based on false discovery rates.
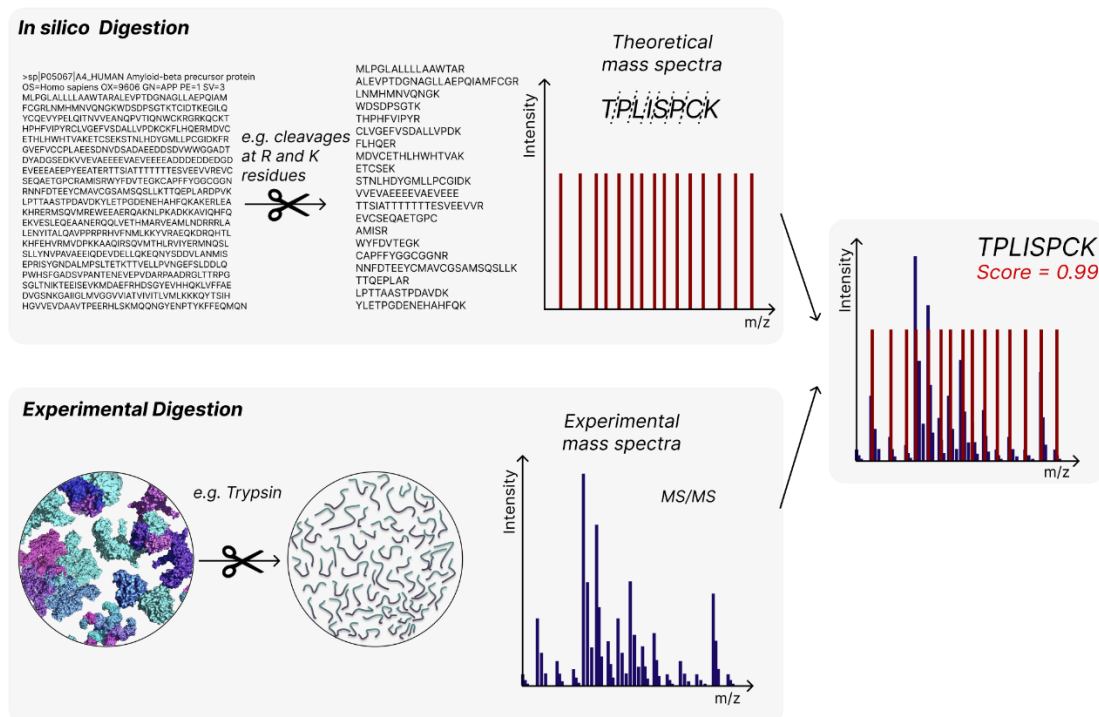
Figure 1-3 General search engine workflow. Comparable to experimental procedures, search engines conduct in silico digestions of FASTA files, considering protease cleavage sites along with potential variable and fixed modifications. Utilizing CID-derived peptide fragmentats, these computational tools generate theoretical spectra to compare against experimental data, yielding a confident score assessing the spectral match.

Although this workflow is widely adopted among proteomics search engines, each tool employs distinct methods to match theoretical and experimental spectra, calculate scores, implement thresholds, and report results. These differences become more pronounced when considering that some software performing protein identification is also capable of quantitative proteomics.

In quantitative proteomics, the goal is to extract quantitative data on peptides in a sample, thereby inferring protein abundance. Despite this information being independent from protein identification process, the quantitative values are also derived from the same LC-MS raw data. Intra-sample protein quantitation can be accessed using normalized spectral abundance factor (NSAF), which considers the spectral counts and protein lengths to measure relative protein amounts [22]. While comparing multiple samples, protein quantitation becomes more laborious and involves several methods,

[22] Boris Zybailov et al., "Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces c Erevisiae*," *Journal of Proteome Research* 5, no. 9 (September 1, 2006): 2339–47, https://doi.org/10.1021/pr060161n.

categorized into four main classes: metabolic labeling, stable isotope labeling, internal standards, and label-free approaches (Figure 4). While each method has its strengths and weaknesses (extensively reviewed in [23] and [24]), label-free proteomics has prevailed due to cost-effectiveness and broad dynamic range. In label-free quantitation (LFQ), a primary approach involves assuming a peptide quantity is proportional to the area of respective precursor extracted ion chromatogram (XIC), enabling peptide quantitation across multiple samples. Post XIC measurements, quantitative proteomics tools normalize data for inter-sample comparison, commonly using total ion current as normalization factor. Regarding protein quantitation, software tools vary in protein normalization methods, handling missing values, and performance, which profoundly impact downstream analysis[25].
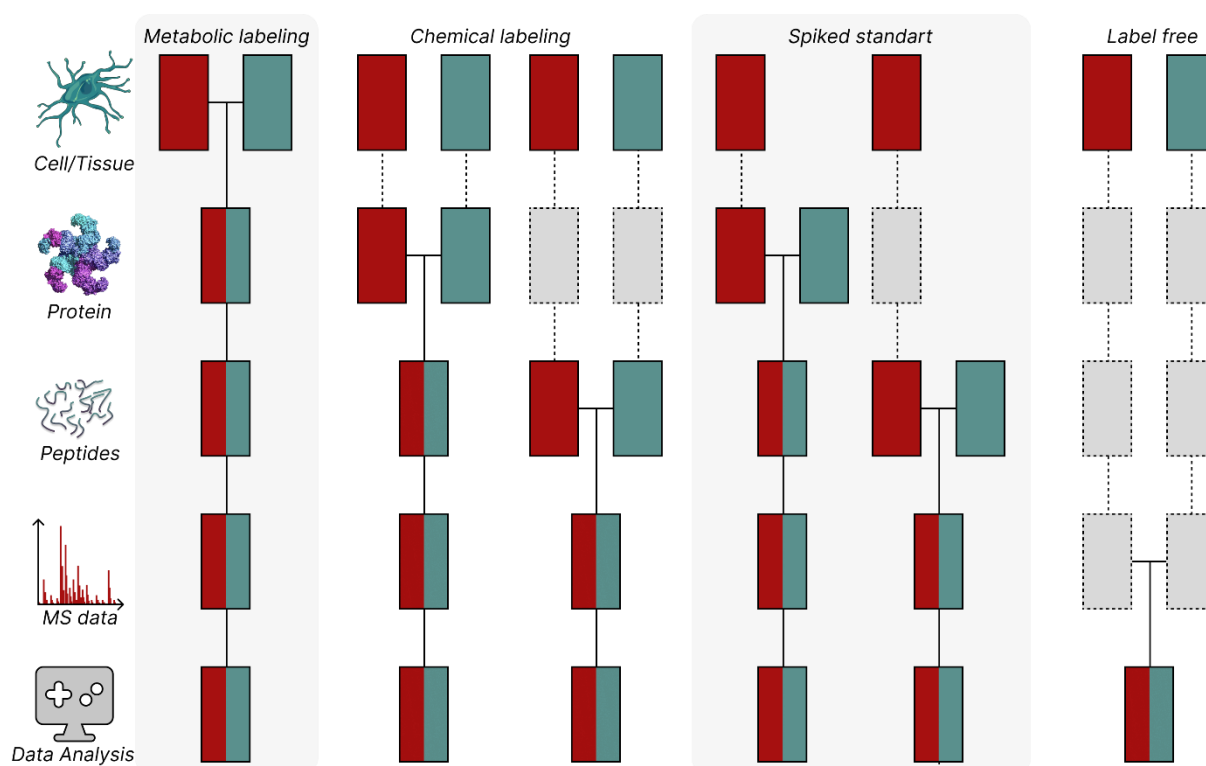
Figure 1-4 General categories of quantitative proteomics workflows. Each class possesses distinct strengths and limitations, particularly concerning cost-effectiveness, quantitative variation, and protein dynamic

---

[23] Marcus Bantscheff et al., "Quantitative Chemical Proteomics Reveals Mechanisms of Action of Clinical ABL Kinase Inhibitors," *Nature Biotechnology* 25, no. 9 (September 2007): 1035–44, https://doi.org/10.1038/nbt1328.

[24] Marcus Bantscheff et al., "Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present," *Analytical and Bioanalytical Chemistry* 404, no. 4 (September 1, 2012): 939–65, https://doi.org/10.1007/s00216-012-6203-4.

[25] Tommi Välikangas, Tomi Suomi, and Laura L Elo, "A Comprehensive Evaluation of Popular Proteomics Software Workflows for Label-Free Proteome Quantification and Imputation," *Briefings in Bioinformatics* 19, no. 6 (November 27, 2018): 1344–55, https://doi.org/10.1093/bib/bbx054.

range. Red and green boxes denote experimental conditions, while horizontal lines indicate sample combination points and steps contributing to experimental variation (dashed lines). Adapted from [26].

As previously mentioned, proteomics offers multiple software options for both protein identification and quantitation from MS-raw data. Since DDA method is the most used approach, several computational tools are designed to process DDA-derived raw data, such as MaxQuant[27] and PatternLabV[28]. On the other hand, DIA experiments utilize a different set of tools, like DIA-NN[29] and Progenesis Qi for Proteomics (Waters Co), to process chimerical MS2 spectra for identification and quantitation. Due to this diversity of tools, establishing a straightforward bioinformatics protocol for downstream proteomics analysis, such as differential proteomics, becomes challenging.

1.2   DIFFERENTIAL PROTEOMICS ANALYSIS

In genomics and transcriptomics, researchers conduct differential expression analysis to compare different groups and identify distinct gene patterns between populations. Considering definition and experimental approaches, in proteomics, the workflow focuses not on evaluating changes in expression patterns but rather on differences in protein abundance across multiple groups. Protein abundance can be regulated by several processes and may not always correlate directly with the expression of a specific gene[30]. Therefore, the term used to assess changes across groups when working with proteins is "differential proteomics analysis".

The differential proteomics workflow encompasses a series of computational steps designed to comparing protein abundance in a large scale manner, employing quantitative parameters to distinguish groups. This pipeline has evolved in parallel with

[26] Bantscheff et al., "Quantitative Mass Spectrometry in Proteomics."

[27] Jürgen Cox and Matthias Mann, "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification," *Nature Biotechnology* 26, no. 12 (December 2008): 1367–72, https://doi.org/10.1038/nbt.1511; Stefka Tyanova, Tikira Temu, and Juergen Cox, "The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics," *Nature Protocols* 11, no. 12 (December 2016): 2301–19, https://doi.org/10.1038/nprot.2016.136.

[28] Marlon D. M. Santos et al., "Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V," *Nature Protocols* 17, no. 7 (July 2022): 1553–78, https://doi.org/10.1038/s41596-022-00690-x.

[29] Vadim Demichev et al., "DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput," *Nature Methods* 17, no. 1 (January 2020): 41–44, https://doi.org/10.1038/s41592-019-0638-x.

[30] Fredrik Edfors et al., "Gene-specific Correlation of RNA and Protein Levels in Human Cells and Tissues," *Molecular Systems Biology* 12, no. 10 (October 2016): 883, https://doi.org/10.15252/msb.20167144; Yansheng Liu, Andreas Beyer, and Ruedi Aebersold, "On the Dependency of Cellular Protein Levels on mRNA Abundance," *Cell* 165, no. 3 (April 21, 2016): 535–50, https://doi.org/10.1016/j.cell.2016.03.014.

other computational proteomics strategies, enabling the exploration of biomarkers and biological mechanisms within increasingly diverse and complex proteomes [31]. Consequently, its exploratory nature should encompass a range of biological inquiries, considering both the technical and biological variables inherent in experimental designs.

In differential proteomics, statistical analysis forms the core, involving the management of technical and biological replicates, data filtering, normalization, and application of suitable statistical tests[32]. While each step is crucial for ensuring the most reliable and confident outcomes, selecting the right statistical test poses a challenge. This challenge is particularly complex due to various factors associated with experimental design, including variable dependencies, the number of groups being compared, and the necessity for multiple hypothesis correction.

Proteomics studies typically employ a static experimental design, aiming to uncover molecular signatures for disease stratification or identify biological mechanisms related to diseases or treatments (Figure 5). This experimental design assumes independence between variables, meaning each group features distinct characteristics unrelated to protein levels in other groups[33]. As protein abundance is frequently log-transformed to approximate a normal distribution in the data, comparisons between two groups can be performed using an independent t-test. For studies involving more than two groups, under similar principles of independence, Analysis of Variance (ANOVA) is employed to assess differentially regulated proteins (DRPs) among these groups. Both the t-test and ANOVA assume data with a normal distribution and equal variances, being statistical tests suitable for comparing groups with at least three replicates[34].

---

[31] Ruedi Aebersold and Matthias Mann, "Mass-Spectrometric Exploration of Proteome Structure and Function," *Nature* 537, no. 7620 (September 2016): 347–55, https://doi.org/10.1038/nature19949.

[32] Bantscheff et al., "Quantitative Mass Spectrometry in Proteomics."

[33] Bantscheff et al.; Bantscheff et al., "Quantitative Chemical Proteomics Reveals Mechanisms of Action of Clinical ABL Kinase Inhibitors."

[34] Bantscheff et al., "Quantitative Mass Spectrometry in Proteomics"; Bantscheff et al., "Quantitative Chemical Proteomics Reveals Mechanisms of Action of Clinical ABL Kinase Inhibitors."
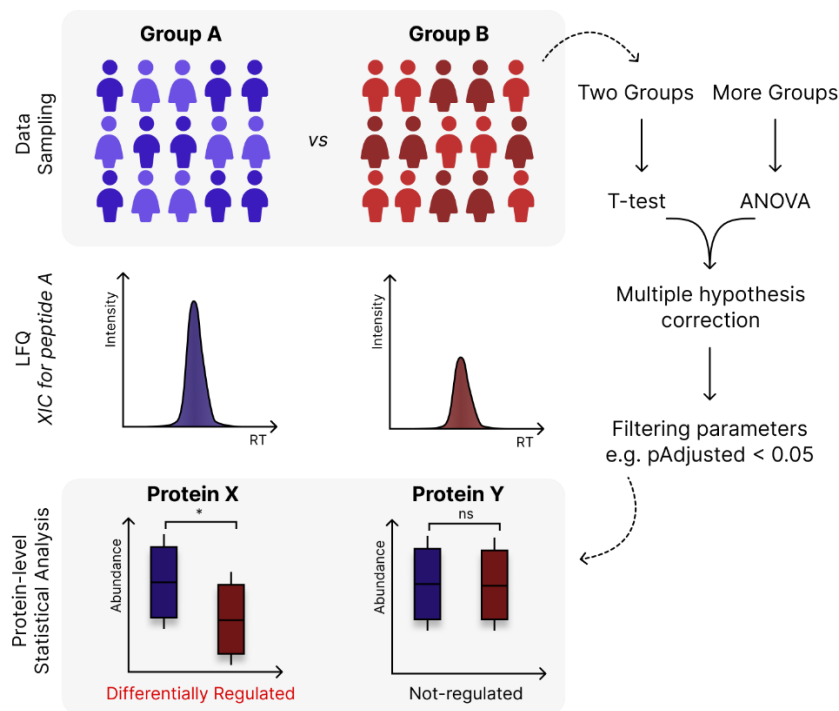
Figure 1-5 Static experimental design. In the LFQ approach, computational tools analyze the extracted ion chromatograms (XIC) from individual peptides to assess protein abundance and compare various groups. Within static experimental designs, protein levels across groups remain independent, enabling statistical analyses such as t-tests or ANOVA to distinguish between groups. Subsequently, multiple hypothesis corrections and filtering are employed to identify differentially regulated proteins.

In addition to the static experimental designs, the incorporation of time as a variable in longitudinal analysis significantly broadens the spectrum of inquiries and challenges (Figure 6). In time-course experiments, it is possible to evaluate the progression of a disease over time, the variation of protein abundance between different treatments, or take into account data sampling details – such as the independence of samples or time ranges in data sampling. While seeking differentially regulated proteins, one approach involves predefining thresholds and counting instances where protein levels surpass specific fold-change cutoffs[35]. Critically, this ad hoc methodology relies on arbitrary thresholds and assumes uniform turnover processes for all proteins, which inherently limits its applicability[36]. Alternatively, specialized statistical methods tailored for this analysis have emerged, integrating automated workflows and considering protein abundance for longitudinal proteomics analysis. Storey (2005) proposed a method initially developed for microarrays and embedded in the EDGE R

---

[35] Ziv Bar-Joseph, Anthony Gitter, and Itamar Simon, "Studying and Modelling Dynamic Biological Processes Using Time-Series Gene Expression Data," *Nature Reviews Genetics* 13, no. 8 (August 2012): 552–64, https://doi.org/10.1038/nrg3244.

[36] Bar-Joseph, Gitter, and Simon.

package, assuming gene expressions vary over time according to a natural cubic spline[37]. This approach fits null and full models—no differential expression and differential expression, respectively—employing an F-test to compute p-values. Storey's methodology accommodates many longitudinal experimental features, spanning multiple range sampling, related or independent sampling, and comparisons within or between groups (Figure 6). Although initially devised for microarrays, researchers have applied Storey's approach to evaluate protein abundance in longitudinal proteomics studies, showcasing its adaptability across different experimental contexts[38].
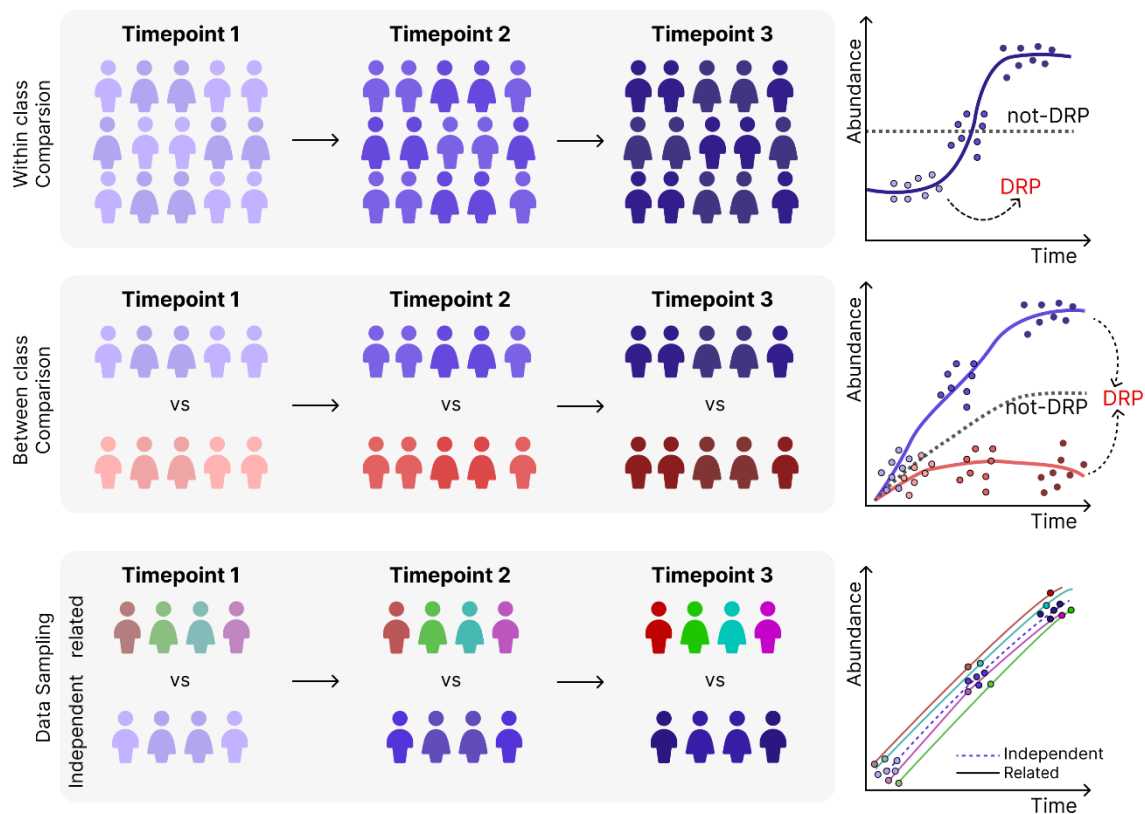
Figure 1-6 Longitudinal experimental design. Incorporating time into experimental designs allows various inquiries, including comparisons within or between groups and considering related or independent data

[37] John D. Storey et al., "Significance Analysis of Time Course Microarray Experiments," *Proceedings of the National Academy of Sciences* 102, no. 36 (September 6, 2005): 12837–42, https://doi.org/10.1073/pnas.0504609102.

[38] Nils Kurzawa et al., "A Computational Method for Detection of Ligand-Binding Proteins from Dose Range Thermal Proteome Profiles," *Nature Communications* 11, no. 1 (November 13, 2020): 5783, https://doi.org/10.1038/s41467-020-19529-8; Brian C. Searle et al., "Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry," *Nature Communications* 9, no. 1 (December 3, 2018): 5128, https://doi.org/10.1038/s41467-018-07454-w.

sampling. In Storey methodology[39], protein levels exhibit variation based on a spline, and differentially regulated proteins (DRPs) are identified by comparing full (DRPs) or null (not DRPs) models.

After executing both static and longitudinal pipelines, the tools generate nominal p-values for each protein, indicating the likelihood of observing DRPs by random chance. Given the simultaneous evaluation of thousand proteins, multiple hypothesis testing increases the risk of identifying false positives, requiring additional steps to mitigate false discoveries. In this context, the widely adopted approach in differential expression and proteomics analyses is the Benjamini-Hochberg correction for controlling false discovery rates (FDR)[40]. This correction involves sorting, ranking, and recalculating p-values, assigning critical values based on pre-specified FDR[41]. The resulting adjusted p-values, also referred to as p-adjusted, identify DRPs, providing a more reliable discrimination between the studied groups.

In quantitative proteomics, alongside statistical analysis, a crucial aspect involves measuring how much a condition alters protein abundance relative to others. A common method is averaging individual protein abundances across analyzed conditions and calculating ratios, termed fold change, between experimental groups versus control groups[42]. The fold change also serves as an additional filter to enhance bolster confidence in defining DRPs. However, this approach assumes uniform variances across all proteins, which may not hold true and potentially leads to the loss of valuable information. Noteworthy, within 'omics approaches, log2-transformation of fold-change values finds widespread application[43], aiding data interpretation and visualization by normalizing the alterations between conditions.

While performing differential proteomics analysis, a variety of computational tools are available, offering distinct strengths and limitations. The first category encompasses tools like PatternLab V and Progenesis QI for Proteomics, combining search engines with differential proteomics capabilities, providing comprehensive platforms within a unified environment. Despite their rich feature sets, these tools may

---

[39] Storey et al., "Significance Analysis of Time Course Microarray Experiments."

[40] Yoav Benjamini and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)* 57, no. 1 (1995): 289–300.

[41] Benjamini and Hochberg.

[42] Bantscheff et al., "Quantitative Chemical Proteomics Reveals Mechanisms of Action of Clinical ABL Kinase Inhibitors."

[43] Lily Ting et al., "Normalization and Statistical Analysis of Quantitative Proteomics Data Generated by Metabolic Labeling," *Molecular & Cellular Proteomics: MCP* 8, no. 10 (October 2009): 2227–42, https://doi.org/10.1074/mcp.M800462-MCP200.

lack extensive visualization functions or dedicated statistical workflows to effectively address experimental questions. The second category includes tools like Perseus[44] and AlphaPeptStats[45], downstream to search engines and suited specifically for differential proteomics analysis. While these tools often offer enhanced visualization and statistical analysis features, their applicability might be constrained to a limited set of accepted software inputs, often built around templates like MaxQuant. Lastly, the third category comprises R packages and Python libraries, enabling a comprehensive proteomics analysis via command line or notebooks. However, these tools demand a steep learning curve in programming skills, posing a time-consuming challenge and an initial barrier for users seeking to conduct thorough statistical analyses in proteomics.

After delineating DRPs, researchers encounter a vast list, often comprising hundreds or thousands of proteins affected in a biological condition. To further characterize these proteins based on the study's objectives, various approaches come into play. These include selecting potential biomarker candidates, unraveling co-regulated proteins via clustering algorithms, exploring protein-protein interactions, and comprehending the collective impact of proteins within a system. While these pursuits are crucial for a comprehensive understanding of proteome dynamics, enrichment analysis is a common method used to examine the role of proteins at the system level.

1.3   ENRICHMENT ANALYSIS

Enrichment analysis comprises a range of methods used to compare experimental results, such as DRPs, with predefined/theoretical datasets, typically cataloged in databases. The term 'enrichment' signifies the primary aim of this analysis: whether a specific set of genes/proteins appears more frequently than expected by chance within a list of genes/proteins, often obtained through high-throughput

---

[44] Stefka Tyanova et al., "The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data," *Nature Methods* 13, no. 9 (September 2016): 731–40, https://doi.org/10.1038/nmeth.3901; Stefka Tyanova and Juergen Cox, "Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research," in *Cancer Systems Biology: Methods and Protocols*, ed. Louise von Stechow, Methods in Molecular Biology (New York, NY: Springer, 2018), 133–48, https://doi.org/10.1007/978-1-4939-7493-1_7.

[45] Elena Krismer et al., "AlphaPeptStats: An Open-Source Python Package for Automated and Scalable Statistical Analysis of Mass Spectrometry-Based Proteomics," *Bioinformatics* 39, no. 8 (August 1, 2023): btad461, https://doi.org/10.1093/bioinformatics/btad461.

techniques[46]. In proteomics, the list of proteins is used to explore potential biological pathways linked to phenotypes, to investigate transcription factors influencing protein abundance, and to pinpoint cell markers for further in-depth examination of cellular specificity.

The most prevalent method for conducting enrichment analysis is Over-Representation Analysis (ORA)[47]. ORA treats the experimental data and the target library as categorical variables, allowing the creation of a contingency matrix that considers the count of experimental proteins, proteins in the target library, and the total possible entities – often termed background, like the number of proteins in a database, or proteins identified in a study (Figure 7). The association between these variables can be assessed using a Fisher's exact test, determining the likelihood of the variables being correlated by random chance[48]. Enrichment analysis commonly involves databases (e.g., KEGG[49]) comprising various gene sets (e.g. 'Glycolysis/Gluconeogenesis' or 'MAPK signaling'). To comprehensively explore these databases, experimental data are tested against multiple gene sets, requiring multiple hypothesis tests for p-value correction[50]. Once again, the Benjamini-Hochberg method can correct p-values to control FDR. Terms with adjusted p-values below a predefined threshold are considered enriched terms, often selected for subsequent biological investigations and validations.

---

[46] Elizabeth I. Boyle et al., "GO::TermFinder—Open Source Software for Accessing Gene Ontology Information and Finding Significantly Enriched Gene Ontology Terms Associated with a List of Genes," *Bioinformatics* 20, no. 18 (December 12, 2004): 3710–15, https://doi.org/10.1093/bioinformatics/bth456.

[47] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki, "Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists," *Nucleic Acids Research* 37, no. 1 (January 2009): 1–13, https://doi.org/10.1093/nar/gkn923.

[48] Huang, Sherman, and Lempicki.

[49] Minoru Kanehisa et al., "KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs," *Nucleic Acids Research* 45, no. D1 (January 4, 2017): D353–61, https://doi.org/10.1093/nar/gkw1092.

[50] Huang, Sherman, and Lempicki, "Bioinformatics Enrichment Tools."
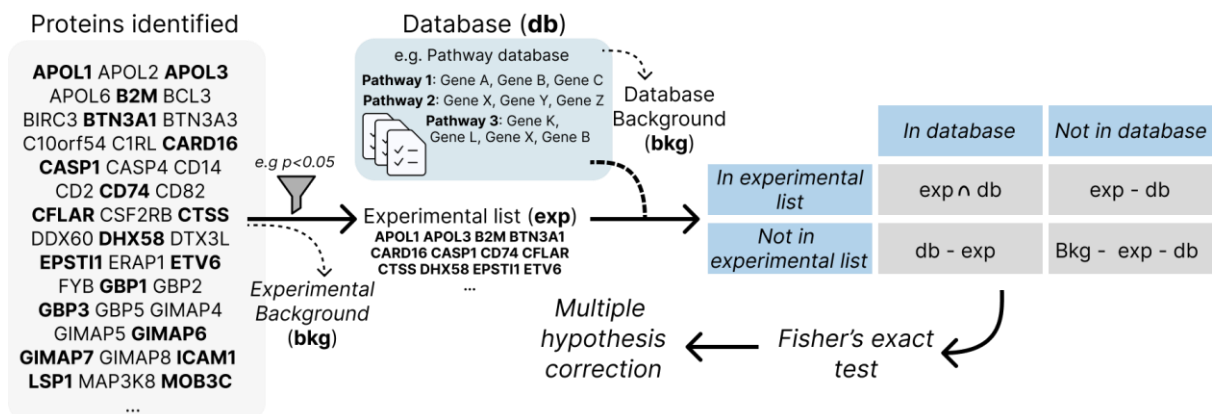
Figure 1-7 Over-representation analysis workflow. In traditional enrichment (or over-representation analysis), differentially regulated proteins are compared against databases containing numerous terms associated with genes. Utilizing the target database, a list of proteins, and background information (defined as either all genes in databases or all experimentally identified proteins), a contingency matrix is constructed. Fisher's test is then applied to each term, followed by multiple hypothesis correction.

In addition to ORA, functional class scoring approach incorporates gene expression levels within statistical framework to perform enrichment analysis, being Gene-Set Enrichment Analysis (GSEA) the most popular method[51]. GSEA involves ranking the experimental gene list based on fold-change between two conditions, where the algorithm determines whether members of a gene set library tend to cluster towards the top (up-regulated) or bottom (down-regulated) of this ranked list (Figure 8). In essence, this statistical approach evaluates whether a specific term (e.g., Glycolysis pathway) exhibits a higher occurrence of up or down-regulated genes than expected. Furthermore, GSEA provides not only p-values but also the Normalized Enrichment Score (NES), indicating the extent to which a library gene set is overrepresented at the extremes (top or bottom) of the entire ranked list. Despite biological mechanisms being regulated by intricate networks of interactions, GSEA is often employed as a tool to anticipate regulation within biological pathways, allowing researchers to design experiments that corroborate the regulatory processes associated with their findings[52].

---

[51] Aravind Subramanian et al., "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles," *Proceedings of the National Academy of Sciences* 102, no. 43 (October 25, 2005): 15545–50, https://doi.org/10.1073/pnas.0506580102.

[52] Aravind Subramanian et al., "GSEA-P: A Desktop Application for Gene Set Enrichment Analysis," *Bioinformatics* 23, no. 23 (December 1, 2007): 3251–53, https://doi.org/10.1093/bioinformatics/btm369.
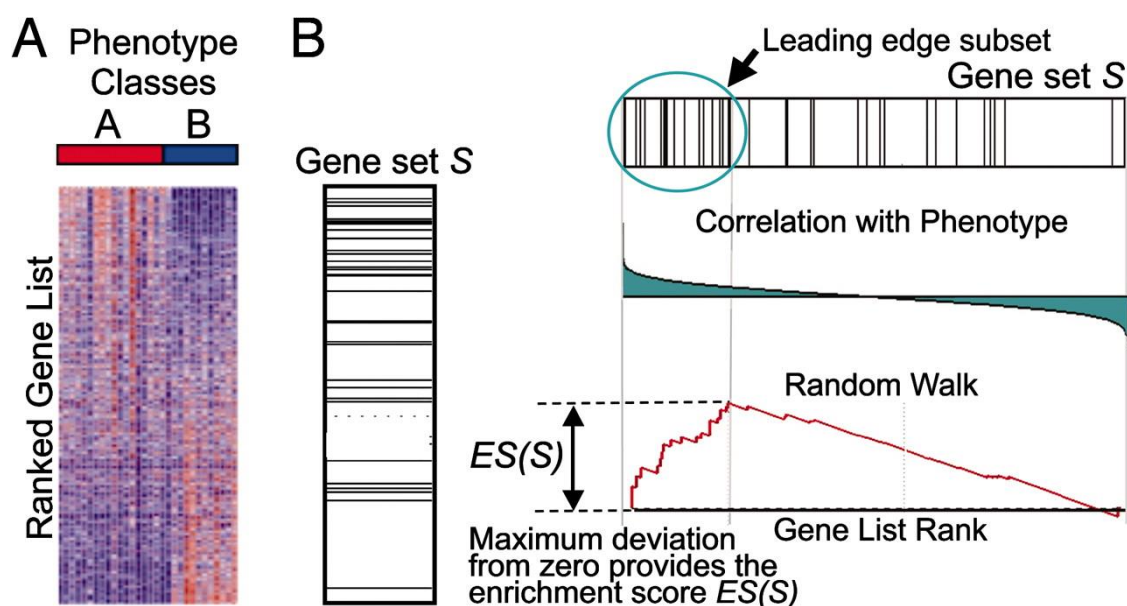
Figure 1-8 Gene Set Enrichment Analysis (GSEA) workflow. GSEA utilizes quantitative information to rank and assess whether an experimental gene list shows over-representation at the top or bottom of a gene set (Gene Set S), typically derived from a database, which may correlate with a determined phenotype. The maximum deviation from zero yields the enrichment score, which can be positive or negative when genes are over-represented at the top or bottom of the list, respectively. Figure from Subramanian [53].

In conventional proteomics computational tools, the integration of enrichment analysis remains relatively scarce. While certain tools like Perseus requires the download of specific databases for Over-Representation Analysis (ORA), others like AlphaPeptStats utilize a limited selection of databases for enrichment analysis, such as Gene Ontology libraries. These limitations prompt users to turn to third-party tools prevalent in genomics and transcriptomics fields, such as DAVID [54], Enrichr [55], Metascape [56], and Reactome [57]. Despite offering streamlined pipelines, these tools introduce additional steps into the proteomics workflow. They often require the conversion of protein accessions into alternative entries, might not perform both ORA and Gene-Set Enrichment Analysis (GSEA), and typically provide either a limited database selection or a constrained toolset for analysis and visualization.

[53] Subramanian et al., "Gene Set Enrichment Analysis," October 25, 2005.

[54] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki, "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources," *Nature Protocols* 4, no. 1 (January 2009): 44–57, https://doi.org/10.1038/nprot.2008.211.

[55] Edward Y. Chen et al., "Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool," *BMC Bioinformatics* 14 (April 15, 2013): 128, https://doi.org/10.1186/1471-2105-14-128.

[56] Yingyao Zhou et al., "Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets," *Nature Communications* 10, no. 1 (April 3, 2019): 1523, https://doi.org/10.1038/s41467-019-09234-6.

[57] Marc Gillespie et al., "The Reactome Pathway Knowledgebase 2022," *Nucleic Acids Research* 50, no. D1 (January 7, 2022): D687–92, https://doi.org/10.1093/nar/gkab1028.

While enrichment analysis adds an extra layer of information to researchers' inquiries, deciphering this massive data presents a notable challenge in proteomics. Extracting biological relevance from differential analysis and enrichment results demands considerable time, especially for non-programmers who often depend on various computational tools or engage in manual data curation. To address this limitation, a widely employed strategy involves the adoption of systems biology methodologies to manage these extensive datasets.

1.4 SYSTEMS BIOLOGY

Systems biology emerged as a scientific field that evolved in parallel with molecular biology, providing a holistic approach to comprehending the intricate complexities inherent to biological systems[58]. While integrating into omics pipelines, this methodology aids researchers in mining and understanding the roles played by thousands of molecular entities implicated in diseases and biological phenomena. Much like other disciplines that aim to study phenomena at the system-level, systems biology applies network theory to extract information about nature, delving into its intricate details by analyzing the interactions within biological systems [59].

In network theory, a network is a graph characterized by nodes and edges, each possessing distinct attributes[60]. Typically, nodes in proteomics represent proteins, while edges denote protein-protein interactions (PPI). The term PPI encompasses various forms of interaction in the literature and databases—ranging from physical contact between proteins (physical interactions) to proteins co-expressed or engaged in shared biological mechanisms (functional interactions)[61]. These nodes and their interactions combine to build the network structure, and alterations within this structure constitute the network dynamics [62]. Over recent decades, studies have highlighted the close relationship between disturbances in both node (such as proteins, genes, and metabolites) edge properties and the onset of diseases and disorders. Notably, recent

---

[58] Hiroaki Kitano, "Systems Biology: A Brief Overview," *Science* 295, no. 5560 (March 2002): 1662–64, https://doi.org/10.1126/science.1069492.
[59] Kitano.
[60] Albert-László Barabási and Zoltán N. Oltvai, "Network Biology: Understanding the Cell's Functional Organization," *Nature Reviews Genetics* 5, no. 2 (February 2004): 101–13, https://doi.org/10.1038/nrg1272.
[61] Damian Szklarczyk et al., "The STRING Database in 2021: Customizable Protein–Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets," *Nucleic Acids Research* 49, no. D1 (January 8, 2021): D605–12, https://doi.org/10.1093/nar/gkaa1074.
[62] Barabási and Oltvai, "Network Biology."

research by Krogan's group underscored substantial variations in PPIs within breast, and head and neck cancer cell lines. Their findings suggest that drugs targeting dysfunctional interactions could serve as potential treatments for respective diseases[63].

When exploring biological and molecular networks, the distribution of connections among nodes (known as degree) follows a non-random pattern within the network. Typically, most studies consider natural networks presenting a scale-free topology, wherein the degree distribution follows a power law[64]. In essence, this means that a few nodes exhibit a higher number of connections, while the majority of nodes have fewer links (Figure 9). These nodes with an elevated degree are termed "hubs" and significantly influence the network structure. Biologically, hub proteins tend to exhibit several characteristics: they may correspond to essential genes, display slower evolution, demonstrate a tendency towards higher abundance, and have profound phenotypic outcomes stemming from their modification or deletion[65].

---

[63] Minkyu Kim et al., "A Protein Interaction Landscape of Breast Cancer," *Science* 374, no. 6563 (October 2021): eabf3066, https://doi.org/10.1126/science.abf3066; Danielle L. Swaney et al., "A Protein Network Map of Head and Neck Cancer Reveals PIK3CA Mutant Drug Sensitivity," *Science* 374, no. 6563 (October 2021): eabf2911, https://doi.org/10.1126/science.abf2911.

[64] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo, "Network Medicine: A Network-Based Approach to Human Disease," *Nature Reviews Genetics* 12, no. 1 (January 2011): 56–68, https://doi.org/10.1038/nrg2918.

[65] Barabási, Gulbahce, and Loscalzo.
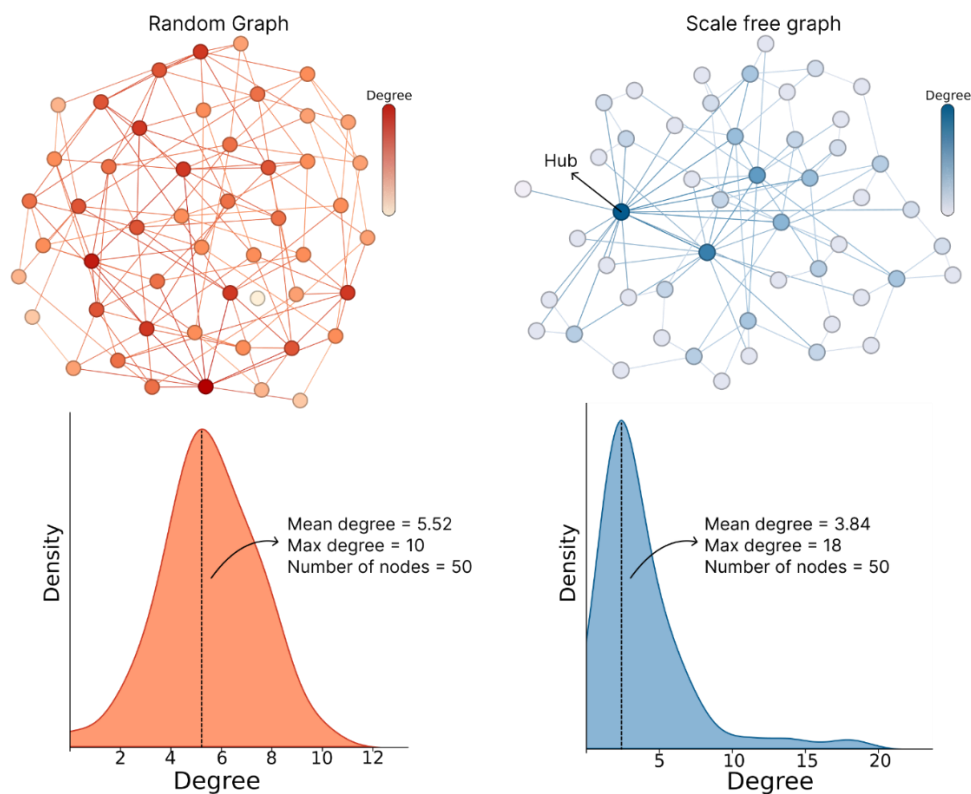
Figure 1-9 Random topology vs. Scale-Free Topology. Two graphs were generated comprising 50 nodes, with degrees following either a random distribution (left) or a scale-free distribution (right). The random topology exhibits a uniform edge distribution, whereas the scale-free topology displays a deviation in degree distribution, characterized by numerous nodes with few links and a few nodes with notably high degrees (hubs).

Additionally, biological networks encompass numerous nodes forming densely interconnected local neighborhoods, referred to as communities or topological modules (Figure 10)[66]. Within these regions, nodes exhibit a stronger propensity to connect with others within the same community rather than with nodes outside it. Detecting these communities within a complex network involves leveraging clustering algorithms, such as the widely applied Louvain algorithm [67]. This method perform a hierarchical clustering approach considering both the amount and strength of pairwise connections within the network. Consequently, as communities are delineated based on link properties, proteins allocated to a specific module often exhibit more shared characteristics than would be anticipated by random chance [68]. In proteomics, this

---

[66] Barabási, Gulbahce, and Loscalzo.

[67] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: Guaranteeing Well-Connected Communities," *Scientific Reports* 9, no. 1 (March 26, 2019): 5233, https://doi.org/10.1038/s41598-019-41695-z.

[68] Traag, Waltman, and van Eck.

phenomenon results in the clustering of proteins involved in the same complex, similar biological pathways, or shared cellular compartments.
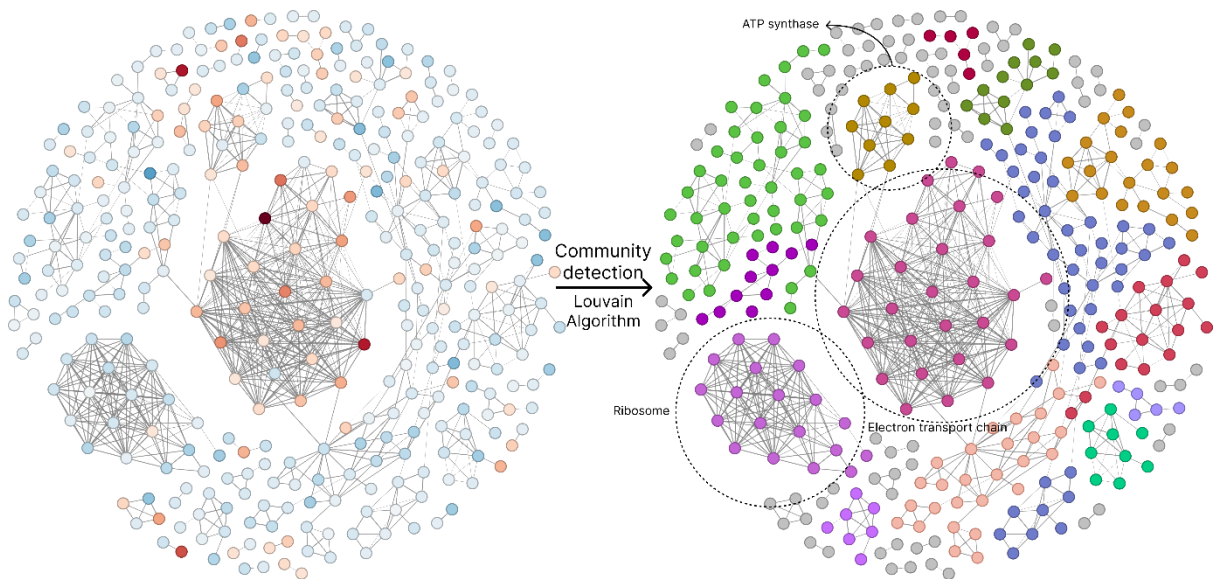
Figure 1-10 Community detection in PPI networks. Complex networks present intricate information, and the search for communities, facilitated by methods like the Louvain algorithm, aids in dissecting network structures. The graph was generated using differentially regulated proteins (up-regulated in red and down-regulated in blue) and includes physical interactions from STRING database.

Enrichment analysis also takes advantage of the power of systems biology approach. When exploring databases, such as KEGG, the enriched terms often exhibit common genes or proteins, which connect distinct biological pathways. However, when employing enrichment using hierarchical databases, such as Gene Ontology and Reactome, the results often exhibit a substantial redundancy[69]. This redundancy arises as child terms entirely overlap with parent terms, leading to repetitive outcomes that highlight the same pathways. To address the integration of distinct terms and mitigate data redundancy, graph representation proves valuable in visualization. In this network, referred to as an enrichment map, nodes symbolize terms while links are established based on pairwise correlation metrics, such as the Jaccard similarity index (Figure 11)[70]. This method generates a weighted network, which can be further analyzed through modularity analysis to identify communities at the enrichment level.

---

[69] Zhou et al., "Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets."

[70] Daniele Merico et al., "Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation," ed. Timothy Ravasi, *PLoS ONE* 5, no. 11 (November 15, 2010): e13984, https://doi.org/10.1371/journal.pone.0013984.
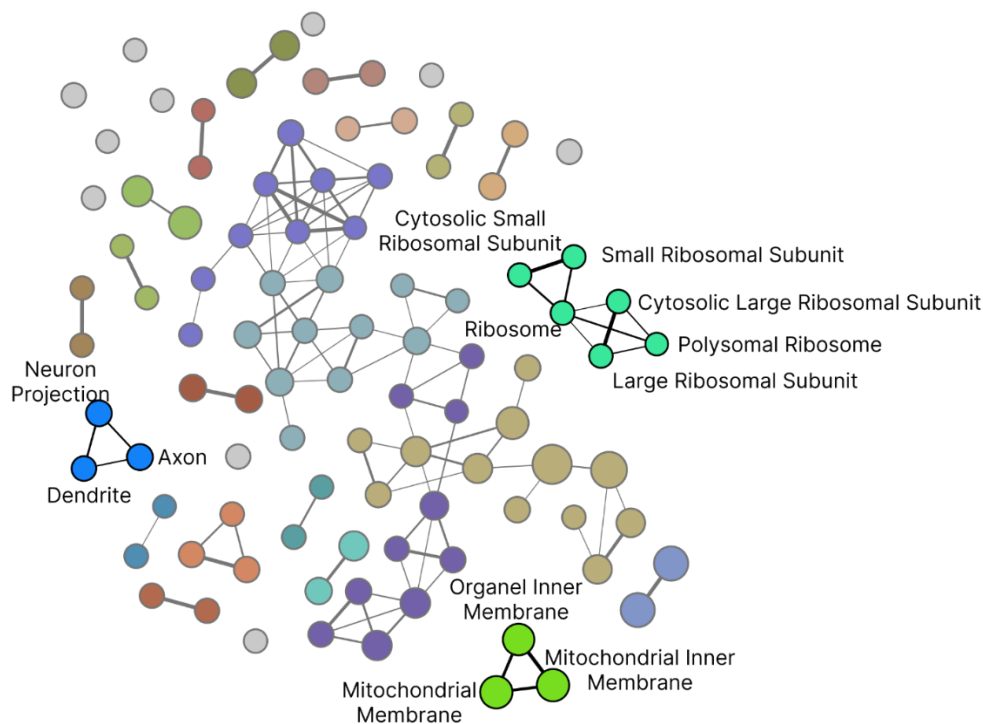
Figure 1-11 Enrichment Map. An enrichment map connects enriched terms that share common genes/proteins, effectively mitigating data redundancy within hierarchical databases. This map was generated using the same input from Figure 11 and compared against the Gene Ontology Cellular Compartment 2023.

Additionally, systems biology's integrative nature allows meta-analysis studies, aiding in the comparison of independent data and mitigating false discovery rates associated with individual studies [71]. Similar to the strategy employed in enrichment map, each study encompassed within a meta-analysis can be represented as a node, with linkages indicative of the proportional sharing of proteins/genes between studies. Alternatively, a more intricate network can be generated depicting studies and respective proteins as nodes, connected by unweighted edges[72]. This last visualization facilitates the identification of protein clusters linking two or more studies. These clusters can undergo deeper investigation, exploring shared characteristics among proteins, such as cellular compartments, offering valuable insights into biological inquiries related to the studies involved in the meta-analysis.

Noteworthy, the practical application of the systems biology approach often requires tools capable of analyzing complex networks. While programming languages

---

[71] Shashank Tripathi et al., "Meta- and Orthogonal Integration of Influenza 'OMICs' Data Defines a Role for UBR4 in Virus Budding," *Cell Host & Microbe* 18, no. 6 (December 9, 2015): 723–35, https://doi.org/10.1016/j.chom.2015.11.002.

[72] Tianzhi Wu et al., "clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data," *The Innovation* 2, no. 3 (August 28, 2021): 100141, https://doi.org/10.1016/j.xinn.2021.100141.

offer libraries for network analysis, software tools equipped with graphical user interfaces (GUI) offer a dynamic network analysis and visualization, crucial for enhancing performance and gaining insights regarding biological systems. Cytoscape stands out by providing a comprehensive environment for biological network analysis, featuring specific plugins for enrichment analysis, PPI exploration, database comparisons, and data visualization[73]. However, it does have drawbacks, including loading times and a steep learning curve, which can be further impacted using plugins. On the other hand, Gephi also serves as a well-established tool for network analysis, offering various plugins to enhance network insights and visualization, although its learning curve is comparable to Cytoscape, and it lacks plugins tailored for biological data[74]. Despite both tools accepting universal network file formats like graphml, creating a file containing all necessary network information can pose challenges, consuming valuable time and acting as a barrier to the wider adoption of the systems biology approach in proteomics workflows.

In the specific context of proteomics, only a few tools incorporate network analysis directly into the workflow, often necessitating the use of third-party software. The Cox group has recently integrated a network module into Perseus, offering proteomics researchers a unified environment for both differential proteomics and systems biology analyses[75]. Despite the developers' efforts to provide a tutorial for configuring the network module, implementing this type of analysis remains a laborious task and demands some programming skills for thorough execution. While the learning curve for Perseus can be time-consuming, its workflow lacks meta-analysis capabilities for comparing results across independent groups.

Taking the above information together, proteomics stands as a field in constant evolution, from technical advances to data analysis workflows. In the last case, researchers often need multiple computational tools to conduct thorough data analysis, encompassing search engines for protein identification, quantitative tools, pipelines for

---

[73] Paul Shannon et al., "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Research* 13, no. 11 (November 2003): 2498–2504, https://doi.org/10.1101/gr.1239303.

[74] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," *Proceedings of the International AAAI Conference on Web and Social Media* 3, no. 1 (March 19, 2009): 361–62, https://doi.org/10.1609/icwsm.v3i1.13937.

[75] Jan Daniel Rudolph and Jürgen Cox, "A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis," *Journal of Proteome Research* 18, no. 5 (May 3, 2019): 2052–64, https://doi.org/10.1021/acs.jproteome.8b00927.

differential proteomics, algorithms for enrichment analysis considering multiple databases, and systems-level tools. Despite notable advancements in concatenating search engine and quantitative proteomics steps, the downstream analysis lacks a unified computational environment, especially given the diversity of experimental proteomics workflows available.

In this context, this thesis aimed to develop a novel computational proteomics tool capable of executing 1) differential proteomics analysis, 2) enrichment analysis, and 3) comparison of multiple independent studies (meta-analysis) within a unified environment. Given the diversity in proteomics workflows utilizing multiple search engines for protein identification and quantitation, this tool needs to handle data files from various sources. Furthermore, the tool must incorporate network analysis to facilitate a system-level assessment of biological questions, spanning from differential proteomics outcomes to comparisons across multiple studies. Crucially, this computational tool aims for inclusivity, catering to both experienced programmers and proteomics beginners, offering accessibility through programming languages and a user-friendly graphical interface.

As a result, this thesis is structured into three chapters, showcasing the progression of data analysis within our group and the development of our computational tool. The chapters are organized as follows:

- **Chapter 1: Digging deeper in the proteome of different regions from schizophrenia brains**. This chapter focuses on conducting differential proteomics across three post-mortem brain regions from patients with schizophrenia. The findings, **published in the Journal of Proteomics in 2020**, exemplify the efforts in our group's approach to proteomics data analysis. However, despite the substantial time and effort invested in this analysis, a retrospective analysis shows that the analyses and figures generated exhibited inefficiencies. They suffered from issues such as inefficient use of space, redundancy, limited information, confusing network representations, and inadequate depiction of group overlaps. More information about this critical review can be found in "*Thesis Considerations*" section on this chapter.

- **Chapter 2: OmicScope: from differential proteomics to systems biology.** Positioned as the thesis's core, this chapter introduces and elucidates OmicScope, our solution tailored for handling quantitative proteomics data. We

comprehensively tested the OmicScope pipeline using real-world and benchmark datasets, showcasing its capabilities. **This chapter has been submitted for publication.**

- **Chapter 3: Diving into the proteomic atlas of SARS-CoV-2 infected cells: insights of viral infection in different cell types.** Here, OmicScope was employed to evaluate nine SARS-CoV-2 infected cell lines, culminating in the construction of an in vitro infectome based on proteomic findings. The utilization of the OmicScope package markedly improved paper performance compared to the 2020 publication. It facilitated quicker work development. OmicScope was pivotal in conducting enrichment analysis, meta-analyses, and generating most figures for the nine in vitro models. This chapter is an example of the progression observed in our laboratory regarding proteomics data analysis, highlighting the advancements realized through the development and application of OmicScope. More information can be found in "*Thesis Considerations*" section on this chapter.

2 CHAPTER 1: DIGGING DEEPER IN THE PROTEOME OF DIFFERENT REGIONS FROM SCHIZOPHRENIA BRAINS

2.1 THESIS CONSIDERATIONS

This chapter is the full article published in the Journal of Proteomics in 2020 – DOI: *10.1016/j.jprot.2020.103814*. In this article, we analyzed the proteome of three different post-mortem brain regions from patients with schizophrenia and compared them with mentally healthy controls. To increase the proteome coverage, we performed subcellular fractionation of mitochondria and nucleus for each brain region, resulting in 9 independent differential proteomics analyses that were concatenated. In experimental terms, we used the bottom-up approach with whole proteome and subcellular fractions, and performed MS analysis in a Q-TOF with HDMS^E acquisition mode (DIA). We used Progenesis Qi for Proteomics (Non-Linear, Waters) for protein identification, quantitation and differential proteomics, followed by downstream data analysis using **four distinct** tools: R language, Cytoscape, DAVID, and Ingenuity Pathway Analysis (IPA, QIAGEN).

Upon critical review of the data analysis process related to this article, we identified several issues. The first one concerns the time invested to analyze and concatenate the 9 datasets, which involved learning R language, Cytoscape and its plugins, DAVID and IPA. The process to concatenate all data took approximately 2 years, starting in 2018 and ending with the publication date in 2020. Most of this time was dedicated to analyzing data and choosing appropriate formats to present the information.

Another notable issue was the use of a licensed software, IPA, to conduct enrichment and network analysis. Although IPA provides a comprehensive and curated database, using licensed software for scientific purposes limits the reproducibility of results for groups that do not have financial support to pay for the license. Currently, even our group does not have the IPA license, which prevents us from reproducing and re-analyzing the data if needed.

Regarding visualization features, figures 1-3 present the same template: one circular diagram, three horizontal bar plots, and one network. The circular diagram was

an attempt to design an alternative to Venn-Diagram, but it became a confusing representation of overlaps, especially for more than 2 conditions. Moreover, despite bar plot being a classical plot in scientific papers, here it was used repeatedly throughout the article, without information gain, highlight for specific terms or pivotal outcomes. The last category, network plots, present a complex structure, mainly in the link distribution, which makes data visualization and interpretation difficult.

Finally, the last significant data analysis issue in this work relates to the integration of quantitative and enrichment results. In general, the work focused on enrichment results to analyze data, due to the challenges in concatenating the multiple data. However, differential proteomics has the ability to indicate protein regulation, and associating this outcome with enrichment result could provide more precise information about the dysregulation in the brain of patients with schizophrenia.

This chapter summarizes and showcases the challenges in analyzing multiple proteomics datasets and is an example of the need for the development of an integrative tool to accelerate the data analysis process and improve its quality.

2.2 ABSTRACT

Schizophrenia is a psychiatric disorder that affects 21 million people worldwide. Despite several studies having been shown that some brain regions may play a critical role in the pathophysiology of schizophrenia, the molecular basis to explain this diversity is still lacking. The cerebellum (CER), caudate nucleus (CAU), and posterior cingulate cortex (PCC) are areas associated with negative and cognitive symptoms in schizophrenia. In this study, we performed shotgun proteomics of the aforementioned brain regions, collected postmortem from patients with schizophrenia and compared with the mentally healthy group. In addition, we performed a proteomic analysis of nuclear and mitochondrial fractions of these same regions. Our results presented 106, 727 and 135 differentially regulated proteins in the CAU, PCC, and CER, respectively. Pathway enrichment analysis revealed dysfunctions associated with synaptic processes in the CAU, transport in the CER, and in energy metabolism in the PCC. In all brain areas, we found that proteins related to oligodendrocytes and the metabolic processes were dysregulated in schizophrenia.

2.3 SIGNIFICANCE

Schizophrenia is a complex and heterogeneous psychiatric disorder. Despite much research having been done to increase the acknowledgment about the role of each region in the pathophysiology of this disorder, the molecular mechanisms underlying it are still lacking. We performed shotgun proteomics in the postmortem cerebellum (CER), caudate nucleus (CAU) and posterior cingulate cortex (PCC) from patients with schizophrenia and compared with healthy controls. Our findings suggest that each aforementioned region presents dysregulations in specific molecular pathways, such as energy metabolism in the PCC, transport in the CER, and synaptic process in the CAU. Additionally, these areas presented dysfunctions in oligodendrocytes and metabolic processes. Our results may highlight future directions for the development of novel clinical approaches for specific therapeutic targets.

2.4 INTRODUCTION

Schizophrenia is a severe psychiatric disorder that affects 21 million people around the world[76]. It is characterized by the presence of positive (e.g. hallucinations, delusions) and negative symptoms (disorganized speech, social withdraw, anhedonia), and cognitive deficits[77]. Schizophrenia is a neurodevelopmental disorder, in which N-methyl-D-aspartate receptor (NMDAr) hypofunction and dysfunction in dopaminergic activity in mesolimbic and mesocortical pathways are associated with symptoms development[78]. In addition, several brain imaging studies have described the role of each brain area in the pathophysiology of schizophrenia.

The cerebellum (CER) is a well-known brain region related to motor coordination, as well as functions in language, emotions, sleep, and visceral responses[79].

---

[76] WHO, "Schizophrenia," accessed May 1, 2019, https://www.who.int/news-room/fact-sheets/detail/schizophrenia.

[77] Robert Freedman, "Schizophrenia," *N. Engl. J. Med.* 349, no. 18 (October 30, 2003): 1738–49, https://doi.org/10.1056/NEJMra035458.

[78] Joshua T Kantrowitz and Daniel C Javitt, "N-Methyl-d-Aspartate (NMDA) Receptor Dysfunction or Dysregulation: The Final Common Pathway on the Road to Schizophrenia?," *Brain Res. Bull.* 83, no. 3–4 (September 2010): 108–21, https://doi.org/10.1016/j.brainresbull.2010.04.006; Oliver D Howes et al., "The Role of Genes, Stress, and Dopamine in the Development of Schizophrenia," *Biol. Psychiatry* 81, no. 1 (January 1, 2017): 9–20, https://doi.org/10.1016/j.biopsych.2016.07.014.

[79] Stacey L Reeber, Tom S Otis, and Roy V Sillitoe, "New Roles for the Cerebellum in Health and Disease," *Front. Syst. Neurosci.* 7 (November 14, 2013): 83, https://doi.org/10.3389/fnsys.2013.00083; Oliver Baumann et al., "Consensus Paper: The Role of the Cerebellum in Perceptual Processes," *Cerebellum* 14, no. 2 (April 2015): 197–220, https://doi.org/10.1007/s12311-014-0627-7.

Increasing recognition of the importance of the cerebellum has pointed to a potential role of this area in the pathophysiology of schizophrenia[80]. A recent mega-analysis showed that patients with schizophrenia present a decrease in gray matter in posterior regions of the cerebellum, which has functional connectivity with associative areas, such as the cortex[81]. In addition, evidence suggests a correlation between the positive symptoms and cerebello-thalamo-cortical disconnectivity[82].

The posterior cingulate cortex (PCC) plays a pivotal role in cognitive functions and in the default mode network [83]. Patients with schizophrenia show alterations in the default mode network when presented to passive tasks[84]. Despite the PCC exhibiting a high metabolic rate, imaging studies have described disturbances in glucose metabolic rate in the PCC from patients with schizophrenia[85]. Recently, Kirino and colleagues have found that patients with schizophrenia present an enhanced functional connectivity between cortical areas, such as the PCC and caudate nucleus, suggesting an additional disruption in cortico-striatal networks[86].

---

[80] Nancy C Andreasen and Ronald Pierson, "The Role of the Cerebellum in Schizophrenia," *Biol. Psychiatry* 64, no. 2 (July 15, 2008): 81–88, https://doi.org/10.1016/j.biopsych.2008.01.003.

[81] T Moberget et al., "Cerebellar Volume and Cerebellocerebral Structural Covariance in Schizophrenia: A Multisite Mega-Analysis of 983 Patients and 1349 Healthy Controls," *Mol. Psychiatry* 23, no. 6 (June 2018): 1512–20, https://doi.org/10.1038/mp.2017.106.

[82] Jessica A Bernard, Joseph M Orr, and Vijay A Mittal, "Cerebello-Thalamo-Cortical Networks Predict Positive Symptom Progression in Individuals at Ultra-High Risk for Psychosis," *Neuroimage Clin* 14 (March 6, 2017): 622–28, https://doi.org/10.1016/j.nicl.2017.03.001; Hengyi Cao et al., "Cerebello-Thalamo-Cortical Hyperconnectivity as a State-Independent Functional Neural Signature for Psychosis Prediction and Characterization," *Nat. Commun.* 9, no. 1 (September 21, 2018): 3836, https://doi.org/10.1038/s41467-018-06350-7.

[83] Robert Leech and David J Sharp, "The Role of the Posterior Cingulate Cortex in Cognition and Disease," *Brain* 137, no. 1 (2014): 12–32, https://doi.org/10.1093/brain/awt162.

[84] Susan Whitfield-Gabrieli et al., "Hyperactivity and Hyperconnectivity of the Default Network in Schizophrenia and in First-Degree Relatives of Persons with Schizophrenia," *Proc. Natl. Acad. Sci. U. S. A.* 106, no. 4 (January 27, 2009): 1279–84, https://doi.org/10.1073/pnas.0809141106; Randy L Buckner, "The Brain's Default Network: Origins and Implications for the Study of Psychosis," *Dialogues Clin. Neurosci.* 15, no. 3 (September 2013): 351–58.

[85] M Mehmet Haznedar et al., "Cingulate Gyrus Volume and Metabolism in the Schizophrenia Spectrum," *Schizophr. Res.* 71, no. 2–3 (December 1, 2004): 249–62, https://doi.org/10.1016/j.schres.2004.02.025; Serge A Mitelman et al., "Metabolic Disconnection Between the Mediodorsal Nucleus of the Thalamus and Cortical Brodmann's Areas of the Left Hemisphere in Schizophrenia," *American Journal of Psychiatry* 162, no. 9 (2005): 1733–35, https://doi.org/10.1176/appi.ajp.162.9.1733; Jeong-Hee Kim et al., "Altered Interregional Correlations between Serotonin Transporter Availability and Cerebral Glucose Metabolism in Schizophrenia: A High-Resolution PET Study Using [C]DASB and [F]FDG," *Schizophr. Res.* 182 (April 2017): 55–65, https://doi.org/10.1016/j.schres.2016.10.020.

[86] Eiji Kirino et al., "Functional Connectivity of the Caudate in Schizophrenia Evaluated with Simultaneous Resting-State Functional MRI and Electroencephalography Recordings," *Neuropsychobiology* 77, no. 4 (2019): 165–75, https://doi.org/10.1159/000490429.

The caudate nucleus (CAU), which plays a pivotal role in learning and reward processes, is a basal nucleus and a part of the striatum[87]. Dysfunctions in the CAU are related to negative symptoms, such as deficits in reward processes, which is hypothesized to be due to a dysregulation of dopamine pathways in the striatum[88]. In addition, the CAU was also associated with cognitive symptoms, since this region and cortical areas connected to the CAU present abnormal hemispheric specialization, which can affect memory and language functions[89].

Despite all these efforts to understand schizophrenia, a molecular basis related to the brain regions mentioned above is still lacking. Proteomics is a powerful approach to investigate complex and multifactorial diseases such as schizophrenia and other psychiatric disorders[90]. Several investigations have been done using postmortem brain tissue from patients with schizophrenia, mainly with the prefrontal cortex[91], anterior

---

[87] Brian Knutson and Jeffrey C Cooper, "Functional Magnetic Resonance Imaging of Reward Prediction," *Curr. Opin. Neurol.* 18, no. 4 (August 2005): 411–17, https://doi.org/10.1097/01.wco.0000173463.24758.f6.

[88] James M Gold et al., "Negative Symptoms of Schizophrenia Are Associated with Abnormal Effort-Cost Computations," *Biol. Psychiatry* 74, no. 2 (July 15, 2013): 130–36, https://doi.org/10.1016/j.biopsych.2012.12.022; Alexis E Whitton, Michael T Treadway, and Diego A Pizzagalli, "Reward Processing Dysfunction in Major Depression, Bipolar Disorder and Schizophrenia," *Curr. Opin. Psychiatry* 28, no. 1 (January 2015): 7–12, https://doi.org/10.1097/YCO.0000000000000122; Mette Ødegaard Nielsen et al., "Negative Symptoms and Reward Disturbances in Schizophrenia Before and After Antipsychotic Monotherapy," *Clin. EEG Neurosci.* 49, no. 1 (January 2018): 36–45, https://doi.org/10.1177/1550059417744120; Sarah Saperia et al., "Reward-Driven Decision-Making Impairments in Schizophrenia," *Schizophr. Res.*, November 12, 2018, https://doi.org/10.1016/j.schres.2018.11.004.

[89] Sophia Mueller et al., "Abnormalities in Hemispheric Specialization of Caudate Nucleus Connectivity in Schizophrenia," *JAMA Psychiatry* 72, no. 6 (June 2015): 552–60, https://doi.org/10.1001/jamapsychiatry.2014.3176.

[90] Juliana M Nascimento and Daniel Martins-de-Souza, "The Proteome of Schizophrenia," *NPJ Schizophr* 1 (March 4, 2015): 14003, https://doi.org/10.1038/npjschz.2014.3; Juliana M Nascimento et al., "Proteomics and Molecular Tools for Unveiling Missing Links in the Biochemical Understanding of Schizophrenia," *Proteomics Clin. Appl.* 10, no. 12 (December 2016): 1148–58, https://doi.org/10.1002/prca.201600021; Verônica M Saia-Cereda et al., "Psychiatric Disorders Biochemical Pathways Unraveled by Human Brain Proteomics," *Eur. Arch. Psychiatry Clin. Neurosci.* 267, no. 1 (February 2017): 3–17, https://doi.org/10.1007/s00406-016-0709-2.

[91] S Prabakaran et al., "Mitochondrial Dysfunction in Schizophrenia: Evidence for Compromised Brain Metabolism and Oxidative Stress," *Mol. Psychiatry* 9, no. 7 (July 2004): 684–97, 643, https://doi.org/10.1038/sj.mp.4001511; Daniel Martins-de-Souza et al., "Proteomic Analysis of Dorsolateral Prefrontal Cortex Indicates the Involvement of Cytoskeleton, Oligodendrocyte, Energy Metabolism and New Potential Markers in Schizophrenia," *J. Psychiatr. Res.* 43, no. 11 (July 2009): 978–86, https://doi.org/10.1016/j.jpsychires.2008.11.006.

temporal lobe[92], orbitofrontal cortex [93], corpus callosum[94], and thalamus[95]. Proteomic approaches even allow for the investigation of subcellular fractions, such as the nucleus and mitochondria. Studies have shown that nuclei from white and gray matter are differentially regulated in patients with schizophrenia [96], as well as mitochondria presenting aberrant size, location, and number among the brain areas[97].

Here, we performed shotgun proteomics to characterize the proteome of postmortem Posterior Cingulate Cortices, Caudate Nuclei, and Cerebella from patients with schizophrenia and compared them with a mentally healthy control group. Additionally, we carried out a subcellular fractionation of mitochondria and nuclei from the same brain regions, performing each respective proteomic analysis. Therefore, this approach allows a whole, mitochondrial, and nuclear analysis to hypothesize about novel molecular features associated with schizophrenia spectrum disorder.

2.5 MATERIALS AND METHODS

2.5.1 *Brain extraction and storage*

Brain samples were collected postmortem and provided by the Brain Net Europe Consortium. Patient samples (n = 5, for each region) came from the State Mental Hospital, Wiesloch, Germany. The individuals were chronic patients with residual symptoms, diagnosed antemortem according to the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) criteria and were medicated. The mentally healthy group (n = 5, for each region) did not present any brain disorder or somatic disease, neither were they medicated with antidepressants or antipsychotics. Control samples

---

[92] Daniel Martins-de-Souza et al., "Alterations in Oligodendrocyte Proteins, Calcium Homeostasis and New Potential Markers in Schizophrenia Anterior Temporal Lobe Are Revealed by Shotgun Proteome Analysis," *J. Neural Transm.* 116, no. 3 (March 2009): 275–89, https://doi.org/10.1007/s00702-008-0156-y.

[93] Erika Velásquez et al., "Synaptosomal Proteome of the Orbitofrontal Cortex from Schizophrenia Patients Using Quantitative Label-Free and iTRAQ-Based Shotgun Proteomics," *J. Proteome Res.* 16, no. 12 (December 1, 2017): 4481–94, https://doi.org/10.1021/acs.jproteome.7b00422.

[94] Verônica M Saia-Cereda et al., "Differential Proteome and Phosphoproteome May Impact Cell Signaling in the Corpus Callosum of Schizophrenia Patients," *Schizophr. Res.* 177, no. 1–3 (November 2016): 70–77, https://doi.org/10.1016/j.schres.2016.03.022.

[95] Daniel Martins-de-Souza et al., "Proteome Analysis of the Thalamus and Cerebrospinal Fluid Reveals Glycolysis Dysfunction and Potential Biomarkers Candidates for Schizophrenia," *J. Psychiatr. Res.* 44, no. 16 (December 2010): 1176–89, https://doi.org/10.1016/j.jpsychires.2010.04.014.

[96] Verônica M Saia-Cereda et al., "The Nuclear Proteome of White and Gray Matter from Schizophrenia Postmortem Brains," *Mol Neuropsychiatry* 3, no. 1 (July 2017): 37–52, https://doi.org/10.1159/000477299.

[97] Rosalinda C Roberts, "Postmortem Studies on Mitochondria in Schizophrenia," *Schizophr. Res.* 187 (September 2017): 17–25, https://doi.org/10.1016/j.schres.2017.01.056.

came from the Institute of Neuropathology, Heidelberg University, Heidelberg, Germany. Both groups were represented by German Caucasians with no history of alcohol or drug abuse. Supplementary Table 1 shows individual aspects of patients and healthy subjects, as well as the respective brain regions collected. Additional information about patients can be found in Saia-Cereda, 2015[98].

2.5.2   Subcellular fractionation, protein extraction, and digestion

Subcellular fractionation, protein extraction, and digestion were carried out following the protocol described by Reis-de-Oliveira[99]. Approximately 45 mg of brain tissue was weighed out in duplicate; the first sample was used for subcellular fractionation and the second was for whole-proteome analysis. The portion used for subcellular fractionation was submitted to cell lysis and a sucrose gradient, followed by differential centrifugation to isolate high membranous organelles (e.g. nuclei) and mitochondria. Both enrichment and whole tissue samples were added in denaturing buffer (Tris-SDS, 2-mercaptoethanol, glycerol, and bromophenol blue), followed by mechanical lysis with pestle and ultrasonication. The samples were then heated at 95ºC and centrifuged at 21,000 x g. Samples were desalted using electrophoresis on a 12% polyacrylamide gel. This experiment was carried out on a short electrophoresis run and a unique gel spot was generated and digest. Each gel lane was trimmed and submitted to in-gel trypsin digestion. The peptides were dried and stored at -80ºC[100].

2.5.3   LC-MS/MS Analyses

Peptides were resuspended in ammonium formate (50 mM, pH 10), quantified by a spectrophotometer (DS-11, DeNovix), and analyzed by liquid chromatography tandem-mass spectrometry (LC-MS/MS). Samples were injected into an ACQUITY UPLC M-Class (Waters, Co) coupled to a Synapt G2-Si mass spectrometer (Waters, Co). The peptides were separated using an HSS T3 1.8 µm x 75 µm x 150 mm column, carrying out a binary gradient from 3 to 40% acetonitrile (ACN) with 0.1% formic acid,

---

[98] Verônica M Saia-Cereda et al., "Proteomics of the Corpus Callosum Unravel Pivotal Players in the Dysfunction of Cell Signaling, Structure, and Myelination in Schizophrenia Brains," *Eur. Arch. Psychiatry Clin. Neurosci.* 265, no. 7 (October 2015): 601–12, https://doi.org/10.1007/s00406-015-0621-1.

[99] Guilherme Reis-de-Oliveira, Mariana Fioramonte, and Daniel Martins-de-Souza, "A Complete Proteomic Workflow to Study Brain-Related Disorders via Postmortem Tissue," *Methods Mol. Biol.* 1916 (2019): 319–28, https://doi.org/10.1007/978-1-4939-8994-2_31.

[100] Reis-de-Oliveira, Fioramonte, and Martins-de-Souza.

followed by a binary gradient from 40 to 85% ACN at a flow rate of 400 nL/min. When performing two-dimensional liquid chromatography (2D), the peptides were first loaded on an XBridge BEH130 C18 5µm x 300 µm x 50 mm column and fractionated in five elutions before the stepwise gradient across the analytical column, using 11.4%, 14.7%, 17.4%, 20.7%, and 50.0% ACN. The chromatographic conditions were based on previous work from our group[101]. Details about each experiment are described in **Table 1**.

NanoElectrospray ionization was set to positive mode and Data-Independent Acquisition was performed using ion mobility separation and fragmented using collision-induced dissociation (CID) (HDMSe, Waters Co.). The acquisition range utilized was 50 to 2000 *m/z*, using a collision energy fragmentation ramp of 19 to 53 eV. Glu-fibrinopeptide B was used as reference lock mass compound, at a flow rate of 500 nL/min and 100 fmol/uL. CER analyses were enhanced by optimizing precursor fragmentation efficiency based on drift time-specific collision energy profiles (UDMSe).

Table 2-1 Chromatographic and mass spectrometry methods for each experiment performed.

Brain Region	Sample type	UPLC system	Fractions/ sample	Gradient time (min)	MS acquisition mode
CAU	WP	1D	-	95	HDMSE
CAU	MIT	1D	-	95	HDMSE
CAU	NUC	1D	-	95	HDMSE
PCC	WP	1D	-	120	HDMSE
PCC	MIT	1D	-	54	HDMSE
PCC	NUC	1D	-	54	HDMSE
CER	WP	2D	5	54	UDMSE
CER	MIT	2D	5	54	UDMSE
CER	NUC	2D	5	54	UDMSE

CAU = Caudate Nucleus; PCC = Posterior Cingulate Cortex; CER = Cerebellum; WP = Whole-proteomic; MIT = Mitochondrial-enrichment; NUC = Nucleus-enrichment.

[101] Saia-Cereda et al., "Differential Proteome and Phosphoproteome May Impact Cell Signaling in the Corpus Callosum of Schizophrenia Patients"; Reis-de-Oliveira, Fioramonte, and Martins-de-Souza, "A Complete Proteomic Workflow to Study Brain-Related Disorders via Postmortem Tissue."

2.5.4 Data Processing and Statistical Analysis

Progenesis QI for Proteomics® (version 3.0) was used to analyze MS raw files. MS/MS spectra were searched against the *Homo sapiens* database (Uniprot Reviewed database, March 2019), using the Ion Accounting algorithm (Version 4.0)[102]. Cysteine carbamidomethylation was applied as fixed modification and methionine oxidation as variable one. Trypsin was set as the digestion enzyme with a maximum of 1 missed cleavage and 600 kDa as the maximum peptide mass. The ion set requirements were 2 or more fragments/peptide, 5 or more fragments/protein, 1 or more peptides/protein, and a false discovery rate (FDR) of 1% for peptide and protein identification. Quantitation was performed using the 3 most abundant peptides for each protein[103]. Proteins with absolute mass error of 20 ppm were deleted.

Progenesis QI for proteomics performs a variation of the one-factor ANOVA calculation, which assumes that the conditions are independent, and the means of the conditions are all equal. Therefore, the tests return a p-value that takes into account the mean difference and the variance and also the sample size [104]. Proteins differentially regulated (ANOVA<0.05) were used to perform the *in silico* analyses in Database for Annotation, Visualization and Integrated Discovery (DAVID)[105] and Ingenuity Pathway Analysis® (IPA, QIAGEN). Data visualization was performed using R environment and Cytoscape[106].

2.6 RESULTS

2.6.1 Proteomic analysis

Th*e postmortem* caudate nucleus (CAU), cerebellum (CER), and posterior cingulate cortex (PCC) samples from patients with schizophrenia and the control group were submitted to differential centrifugation in order to perform subcellular

---

[102] Guo-Zhong Li et al., "Database Searching and Accounting of Multiplexed Precursor and Product Ion Spectra from the Data Independent Analysis of Simple and Complex Peptide Mixtures," *Proteomics* 9, no. 6 (March 2009): 1696–1719, https://doi.org/10.1002/pmic.200800564.

[103] Jeffrey C Silva et al., "Absolute Quantification of Proteins by LCMSE," *Molecular & Cellular Proteomics* 5, no. 1 (2006): 144–56, https://doi.org/10.1074/mcp.m500230-mcp200.

[104] Timothy Clough et al., "Protein Quantification in Label-Free LC-MS Experiments," *J. Proteome Res.* 8, no. 11 (November 2009): 5275–84, https://doi.org/10.1021/pr900610q.

[105] Huang, Sherman, and Lempicki, "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources," January 2009.

[106] Shannon et al., "Cytoscape."

fractionation of high-density organelles (e.g. nuclei) and low-density mitochondria. The
proteins from all experiments, whole tissue and respective enrichments, were extracted
and digested in-gel with trypsin. The peptides were loaded in the liquid-
chromatographic system coupled to a Q-TOF mass spectrometer and the data were
analyzed in Progenesis® QI for Proteomics (Figure 1A).
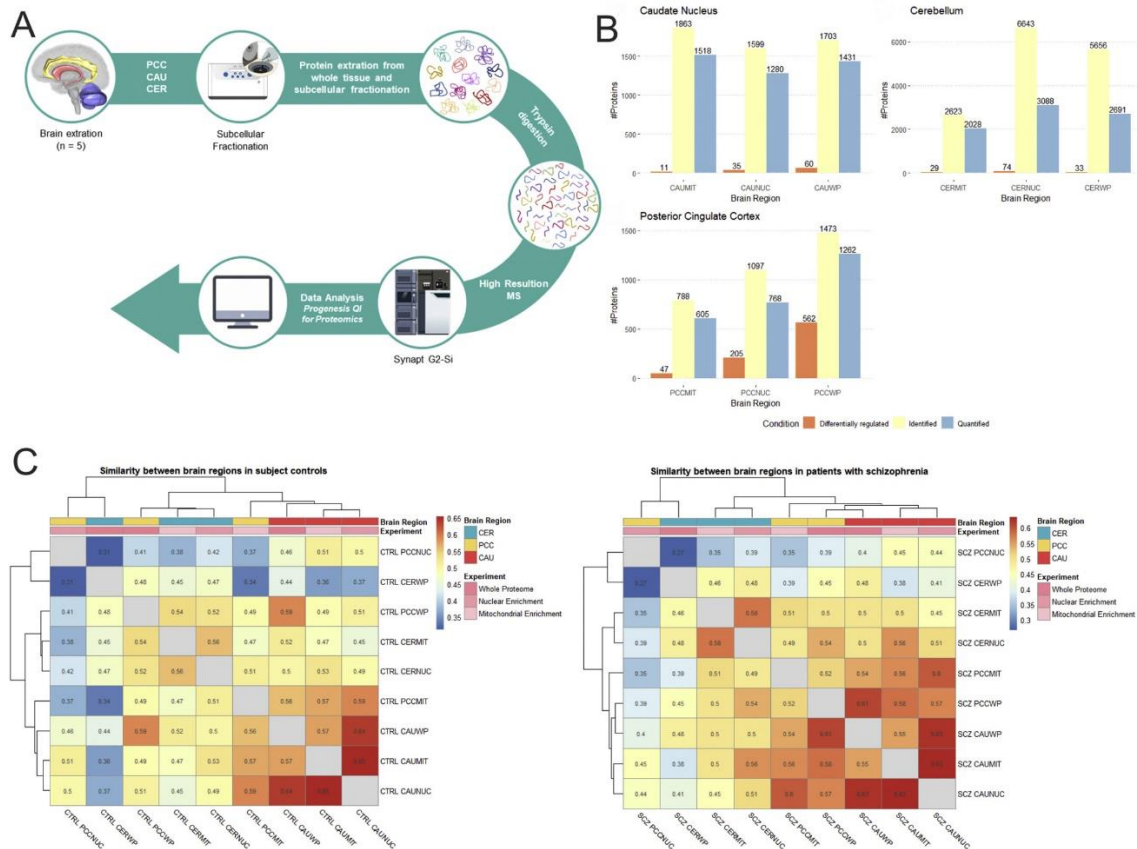
Figure 2-1 Proteomic analysis of postmortem brain tissue from patients with schizophrenia and controls. **A)**
Experimental workflow for sample preparation and quantitative mass spectrometry. **B)** Number of
proteins identified, quantified, and differentially regulated for the three brain regions and subcellular
fractions. **C)** Pearson correlation between analyses of control subjects and patients with schizophrenia.

The CER analyses represented the highest number of proteins quantified, with
1518, 1599, and 1703 proteins from mitochondrial enrichment (CERMIT), nuclear
enrichment (CERNUC), and the whole proteome (CERWP), respectively. For the CAU
samples, 1518 (CAUMIT), 1599 (CAUNUC), and 1703 (CAUWP) proteins were quantified
among the experiments. Finally, the PCC was the region with the lowest number of
quantitated proteins: 788 proteins in mitochondrial enrichment (PCCMIT), 1097 in
nuclear enrichment (PCCNUC), and 1473 in the whole proteome (PCCWP) analysis
(Figure 1B).

We also calculated Pearson's correlation coefficient between the patient and control datasets (Figure 1C). The hierarchical clustering suggests a switching in the proteomic similarities among the brain areas of schizophrenia patients when compared with control subjects. This data suggests that patients with schizophrenia share more similarities in proteome profiling between the PCC and CAU than the control group, in which the PCC is closer to the CER.

2.6.2   Cerebellum

Proteomic analysis of the CER revealed 135 proteins to be differently regulated in patients with schizophrenia compared to controls (ANOVA <0.05, Figure 2A and Supplementary Table 2). Functional enrichment analyses carried out using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) suggest these proteins are mainly related to cellular junctions and cytoskeleton processes (Figure 2B).

The network analysis performed in IPA (Figure 2C) presented dysfunction in cellular organization, which highlighted the role of these proteins in the over-migration of neurons (p = 1.40e-06) and Purkinje cells (p = 3.14E-05), and other cellular movements (p = 1.97e-06). Additionally, deregulated proteins in these networks can lead to dysfunctions in the proteasome, NfkB complex, caspase, and ERK signaling pathways (Figure 2D).
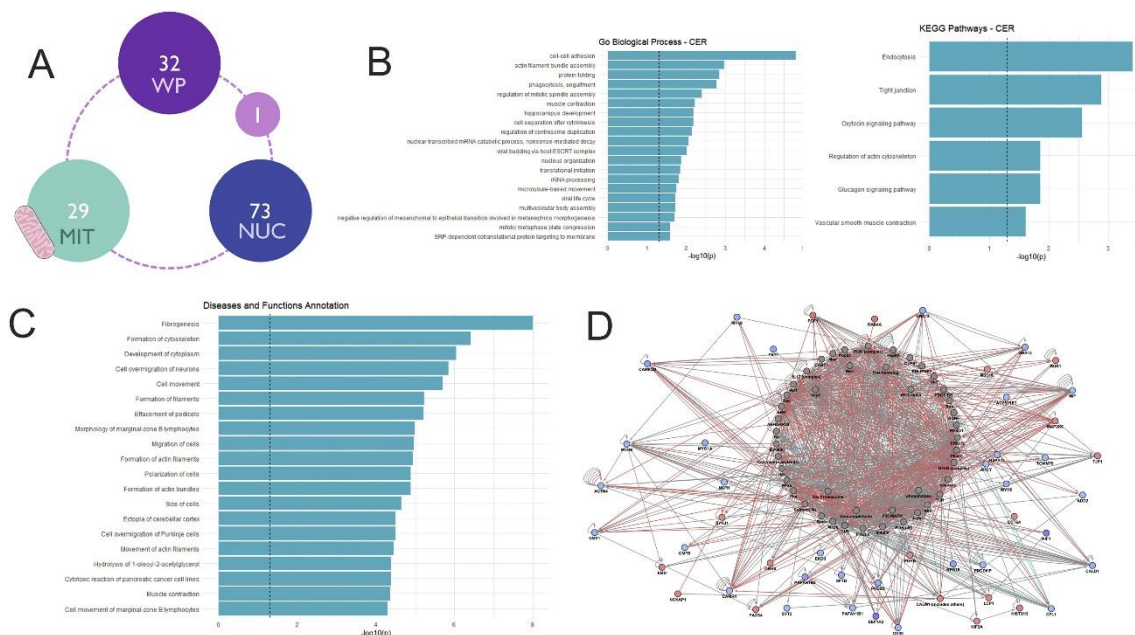
Figure 2-2 Proteomic analysis of the cerebellum. **A)** Diagram with numbers of proteins found differentially regulated in different fractions. **B)** Pathway enrichment analysis performed in KEGG pathway database under Gene Ontology (GO) Biological Process. **C)** Diseases and functional annotation performed using IPA. **D)** Merging of network analysis from each experiment, performed with IPA. Red arrow: activation; blue arrow: inhibition; red circles: proteins found upregulated in the proteomic analysis; blue circles: proteins found downregulated in the proteomic analysis. WP = Whole proteome; MIT = Mitochondrial enrichment; NUC = Nuclear enrichment.

2.6.3   Caudate Nucleus

The proteomic analysis of the CAU showed 106 proteins differentially regulated in patients with schizophrenia when compared with control subjects (ANOVA <0.05, Figure 3A, Supplementary Table 3). The in silico functional analysis carried out using GO and KEGG revealed that the CAU region from patients presented dysregulations in vesicle-mediated transport (p=2.0e-04), and the synaptic vesicle cycle (p=3.11e-04, Figure 3B).

The network analysis performed in IPA reinforced the dysregulation of transport synaptic vesicles (p=2.55e-08) in the top scoring networks. Additionally, IPA revealed that patients with schizophrenia presented disturbances in the apoptotic process in neurons (p = 4.80e-05, Figure 3C). Proteins related to these networks can also affect key molecules of the immunological process (e.g. TCR, NFkB complex), signalling (ERK1/2), and hormonal pathways (LSH, growth hormone, Figure 3D).
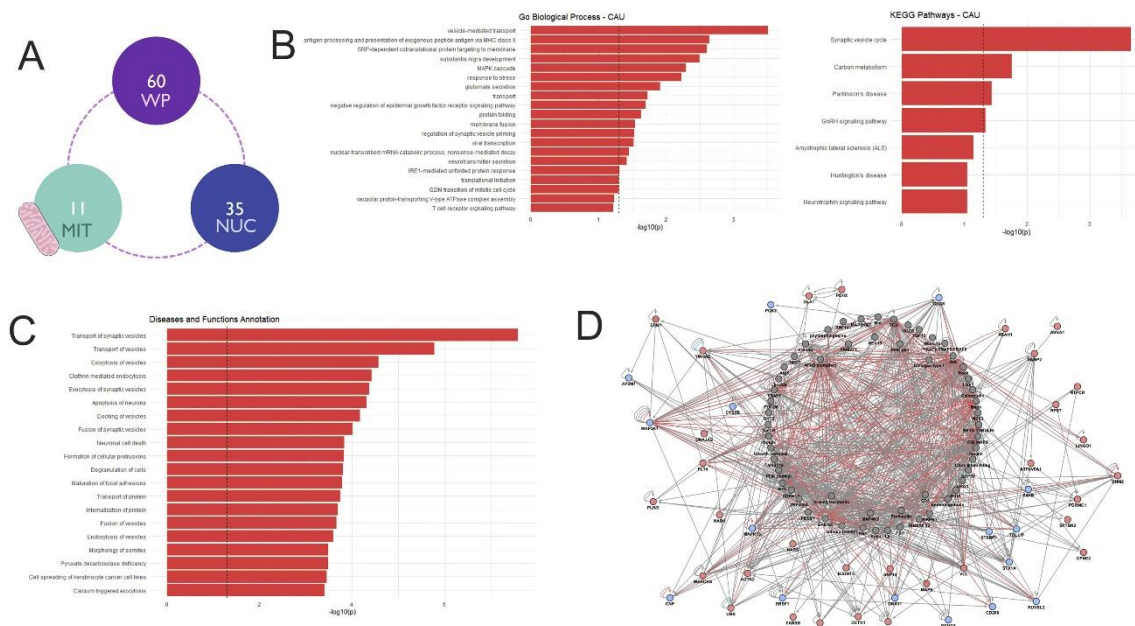
Figure 2-3 Proteomic analysis of the caudate nucleus. **A)** Diagram with numbers of proteins found differentially regulated in different fractions. **B)** Pathway enrichment analysis performed in KEGG pathway database under GO Biological Process. **C)** Diseases and functional annotation performed using IPA. **D)** Merging of network analysis from each experiment, performed with IPA. Red arrow: activation; blue arrow: inhibition; red circles: proteins found upregulated in the proteomic analysis; blue circles: proteins found downregulated in the proteomic analysis. WP = Whole proteome; MIT = Mitochondrial enrichment; NUC = Nuclear enrichment.

2.6.4   Posterior Cingulate Cortex

The posterior cingulate cortex showed a total of 727 proteins differentially regulated in patients compared to controls (ANOVA <0.05, Figure 4A-B, Supplementary Table 4). The in silico analysis revealed these proteins are related to the tricarboxylic acid cycle (p = 2.96e-15), carbon metabolism (p = 9.24e-25), oxidative phosphorylation (p = 2.92e-22), and neuronal processes, such as synaptic vesicle cycle (p = 2.61e-11) and dopaminergic synapses (p = 6.55e-07).

The further network analyses carried out in IPA uncovered that proteins deregulated in the merged networks were related to transport of proteins (p = 1.20e-07), vesicles (p = 6.38e-06), and molecules (p = 1.16e-05); mitochondrial disorders (p = 3.73e-05); and neuritogenesis (p = 1.18e-04, Figure 4C). In addition, this merged network presented proteins related to energy metabolism (e.g. phosphatases, ATPases), as well as signaling pathways (e.g. ERK1/2, 14-3-3) and the immunological process (e.g. NFkB, TCR).
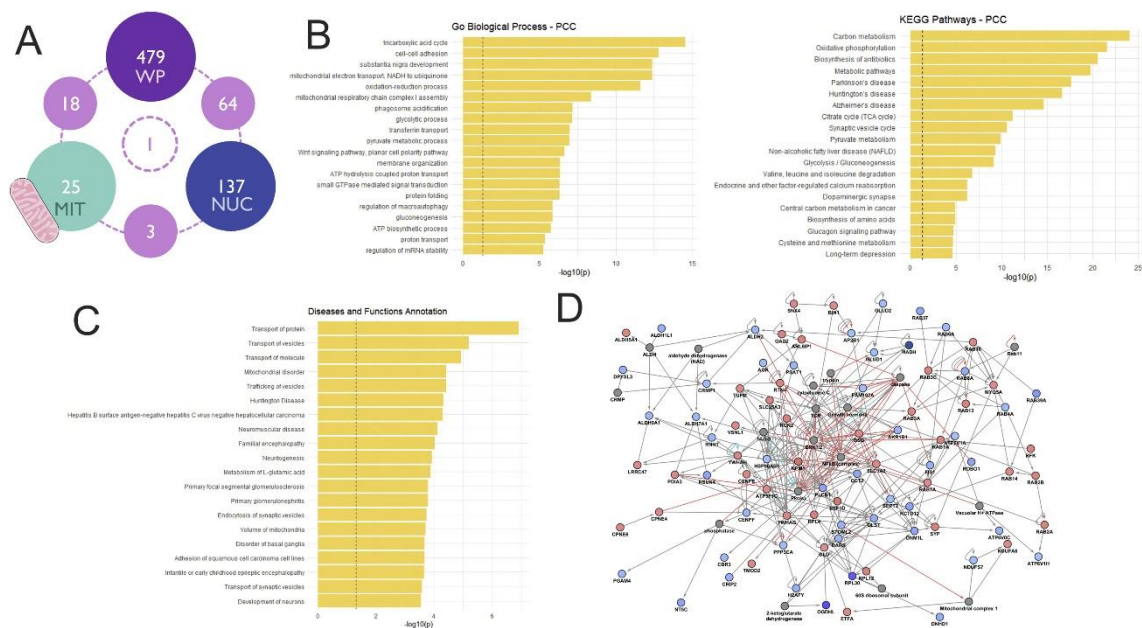
Figure 2-4 Proteomic analysis of the posterior cingulate cortex. **A)** Diagram with numbers of proteins found differentially regulated in different fractions. **B)** Pathway enrichment analysis performed in KEGG pathway database under Gene Ontology Biological Process. **C)** Diseases and functional annotation performed using IPA. **D)** Merging of network analysis from each experiment, performed with IPA. Red arrow: activation; blue arrow: inhibition; red circles: proteins found upregulated in the proteomic analysis; blue circles: proteins found downregulated in the proteomic analysis. WP = Whole proteome; MIT = Mitochondrial enrichment; NUC = Nuclear enrichment.

2.6.5   All brain regions

According to the proteomic analyses, patients with schizophrenia presented 897 differentially regulated proteins. Of the three regions, the PCC was the region that was most affected, while the CAU was the least affected (Figure 5A). Although there is no shared dysregulated protein among all three brain regions, the PCC shared 35 proteins with both the CAU and CER, while the CAU and CER just presented 1 protein in common (VPS35).

The enriched pathways generated from differentially regulated proteins were submitted to the Enrichment Map[107] and MCODE[108] plugins in Cytoscape in order to select the regions with higher connections in the pathway networks (Figure 5C). This analysis showed a relationship between the brain areas and their main affected pathways: the PCC was highly represented by energy metabolism and signaling pathways, the CAU was most deregulated in neurotransmission processes, and the CER

---

[107] Merico et al., "Enrichment Map."

[108] Gary D. Bader and Christopher WV Hogue, "An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks," *BMC Bioinformatics* 4, no. 1 (January 13, 2003): 2, https://doi.org/10.1186/1471-2105-4-2.

presented a dysregulation in transport-related proteins. In addition, translational processes were affected in all the brain regions.
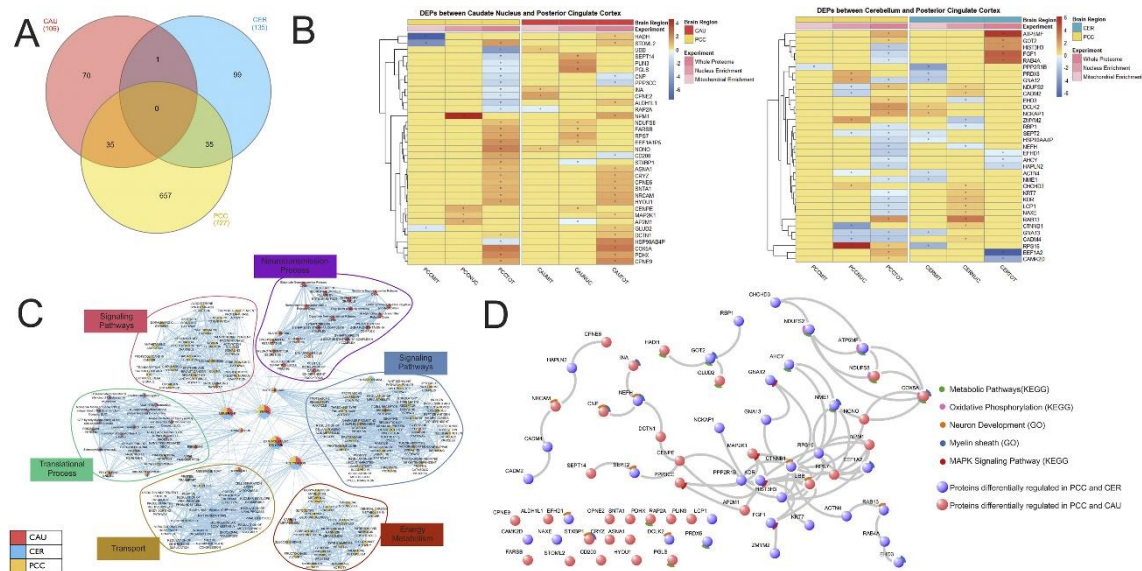
Figure 2-5 Association between brain areas. **A)** Venn diagram with proteins found differentially regulated in each cerebral region. **B)** Heatmaps of proteins found differentially regulated in more than one brain region. **C)** Network of pathways deregulated according to proteomic analysis, showing clusters associated with individual cerebral regions. **D)** Protein-protein interactions of proteins deregulated in the posterior cingulate cortex and other brain regions. WP = Whole proteome; MIT = Mitochondrial enrichment; NUC = Nuclear enrichment; CAU = Caudate nucleus; PCC = Posterior Cingulate Cortex; CER = Cerebellum.

As mentioned above, 35 differentially regulated proteins were identified in the CAU and PCC areas, as well as in the CER and PCC (Figure 5B). We performed a pathway enrichment analysis of just these proteins that were deregulated in the PCC and other regions (Figure 5D). Despite the absence of shared proteins among all three regions, the proteins deregulated in more than one region were related to metabolic pathways, oxidative phosphorylation, neuron development, the myelin sheath, and the MAPK signaling pathway. Taken together, despite our data suggesting compartmentalization of cellular dysfunctions throughout the brain of patients with schizophrenia, these regions also share some dysregulations in their biological processes.

2.7 DISCUSSION

As far as we are aware, this is the first study using shotgun proteomics to investigate multiple postmortem brain areas from patients with schizophrenia, areas

which are known to be associated with different functions, such as movement coordination, default mode network, and reward processes. These divergent functions were reflected in our proteomic results, in which each brain region presented specific dysregulated pathways, such as the transport process in the CER, synaptic processes in the CAU, and energy metabolism in the PCC.

Proteomic analyses of all brain areas suggested that the proteome of patients with schizophrenia share different profiles in each brain area compared to healthy subjects. These data suggest that functional dysconnectivity observed in imaging studies with schizophrenia patients[109] may be associated with the molecular changes observed in our study.

2.7.1 *Cerebella from patients present dysregulation in transport-related processes*

The cerebellum has emerged as a pivotal brain region in schizophrenia development, such as the cortico-cerebellar-thalamic-cortical circuits highlighted in gene expression analyses [110]. Our findings showed that patients with schizophrenia present a dysfunction in proteins related to cytoskeleton processes, such as cell over-migration of Purkinje cells (Figure 2C). Purkinje cells are GABAergic interneurons present in the cerebellum and they were found reduced in patients with schizophrenia and bipolar disorder[111]. G Protein Subunit Alpha 12 (GNA12) and G Protein Subunit Alpha 13 (GNA13) are transducer proteins associated with migration[112] (Gan *et al.*, 2012) and cell-adhesion processes[113]. Studies have shown that these genes are affected in

---

[109] Garry D Honey et al., "Functional Dysconnectivity in Schizophrenia Associated with Attentional Modulation of Motor Function," *Brain* 128, no. Pt 11 (November 2005): 200, https://doi.org/10.1093/brain/awh632; Tumbwene E Mwansisya et al., "Task and Resting-State fMRI Studies in First-Episode Schizophrenia: A Systematic Review," *Schizophr. Res.* 189 (November 2017): 9–18, https://doi.org/10.1016/j.schres.2017.02.026.

[110] Andreasen and Pierson, "The Role of the Cerebellum in Schizophrenia"; Tharani Sundararajan, Ann M Manzardo, and Merlin G Butler, "Functional Analysis of Schizophrenia Genes Using GeneAnalytics Program and Integrated Databases," *Gene* 641 (January 30, 2018): 25–34, https://doi.org/10.1016/j.gene.2017.10.035.

[111] Ekrem Maloku et al., "Lower Number of Cerebellar Purkinje Neurons in Psychosis Is Associated with Reduced Reelin Expression," *Proc. Natl. Acad. Sci. U. S. A.* 107, no. 9 (March 2, 2010): 4407–11, https://doi.org/10.1073/pnas.0914483107.

[112] Xiaoqing Gan et al., "PRR5L Degradation Promotes mTORC2-Mediated PKC-δ Phosphorylation and Cell Migration Downstream of Gα12," *Nat. Cell Biol.* 14, no. 7 (May 20, 2012): 686–96, https://doi.org/10.1038/ncb2507.

[113] Thomas E Meigs et al., "Galpha12 and Galpha13 Negatively Regulate the Adhesive Functions of Cadherin," *J. Biol. Chem.* 277, no. 27 (July 5, 2002): 24594–600, https://doi.org/10.1074/jbc.M201984200.

schizophrenia[114] and their partial ablation in a mouse model affects Purkinje cell migration and the development of the cerebellum[115]. Disruption in neuronal migration was reported in patients with schizophrenia, playing a crucial role in the neurodevelopment of cerebellum and other brain regions[116]. In addition, our data also showed upregulation of the AIF1 protein, which is an activation marker for microglial cells, suggesting a potential inflammatory status in the brains of patients. The inflammation has critical relevance to schizophrenia[117]. A recent postmortem study has shown that the cerebellum of patients with schizophrenia presented an over-activation of the TLR4 pathway, which is directly associated with inflammatory conditions[118].

2.7.2   Caudate Nuclei from patients present dysregulation in synaptic pathways

The CAU is a region rich in D2 receptors[119], which are pivotal receptors in schizophrenia pathophysiology[120]. Our data showed that proteins dysregulated in the CAU are related to transport in synaptic vesicles (AP2M1, DNM2, EPN1, SPTBN2, STX1A, VAMP2). Accordingly, dysfunctions in synaptic organization have been reported in the CAU from patients with schizophrenia, which present an increase in synaptic density in striatal areas[121]. In addition, proteomic analyses also showed a

[114] Peilin Jia et al., "Network-Assisted Investigation of Combined Causal Signals from Genome-Wide Association Studies in Schizophrenia," *PLoS Comput. Biol.* 8, no. 7 (July 5, 2012): e1002587, https://doi.org/10.1371/journal.pcbi.1002587.

[115] Alexandra Moers et al., "Galpha12/Galpha13 Deficiency Causes Localized Overmigration of Neurons in the Developing Cerebral and Cerebellar Cortices," *Mol. Cell. Biol.* 28, no. 5 (March 2008): 1480–88, https://doi.org/10.1128/MCB.00651-07.

[116] Hanna Jaaro-Peled et al., "Neurodevelopmental Mechanisms of Schizophrenia: Understanding Disturbed Postnatal Brain Maturation through Neuregulin-1-ErbB4 and DISC1," *Trends Neurosci.* 32, no. 9 (September 2009): 485–95, https://doi.org/10.1016/j.tins.2009.05.007; Jessica A Bernard and Vijay A Mittal, "Cerebellar-Motor Dysfunction in Schizophrenia and Psychosis-Risk: The Importance of Regional Cerebellar Analysis Approaches," *Front. Psychiatry* 5 (November 25, 2014): 160, https://doi.org/10.3389/fpsyt.2014.00160.

[117] Norbert Müller et al., "The Role of Inflammation in Schizophrenia," *Front. Neurosci.* 9 (October 21, 2015): 372, https://doi.org/10.3389/fnins.2015.00372.

[118] Karina S MacDowell et al., "Differential Regulation of the TLR4 Signalling Pathway in Post-Mortem Prefrontal Cortex and Cerebellum in Chronic Schizophrenia: Relationship with SP Transcription Factors," *Prog. Neuropsychopharmacol. Biol. Psychiatry* 79, no. Pt B (October 3, 2017): 481–92, https://doi.org/10.1016/j.pnpbp.2017.08.005.

[119] R A Lahti, R C Roberts, and C A Tamminga, "D2-Family Receptor Distribution in Human Postmortem Tissue: An Autoradiographic Study," *Neuroreport* 6, no. 18 (December 15, 1995): 2505–12, https://doi.org/10.1097/00001756-199512150-00015.

[120] Oliver D Howes and Shitij Kapur, "The Dopamine Hypothesis of Schizophrenia: Version III--the Final Common Pathway," *Schizophr. Bull.* 35, no. 3 (May 2009): 549–62, https://doi.org/10.1093/schbul/sbp006.

[121] Emma Perez-Costas, Miguel Melendez-Ferro, and Rosalinda C Roberts, "Basal Ganglia Pathology in Schizophrenia: Dopamine Connections and Anomalies," *J. Neurochem.* 113, no. 2 (April 2010): 20, https://doi.org/10.1111/j.1471-4159.2010.06604.x.

dysregulation in apoptosis of neurons (ASAH1, CD200, CYCS, LINGO1, MAP3K1, MAPK13, P4HB, STXBP1), which could explain the reduced number of neurons found in the CAU in patients with schizophrenia[122]. Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 1 (LINGO-1) was also found to be upregulated in schizophrenia in the dorsolateral prefrontal cortex and hippocampus[123]. LINGO-1 plays a role in brain development, regulating the maturation of oligodendrocytes and neuronal outgrowth, and whose upregulation could lead to a reduction of myelin-related processes (reviewed in [124]). Our data also showed that proteins associated with carbon metabolism and pyruvate decarboxylase deficiency (PDHX, PGLS, DLAT, PGK2, GLUD2, ALDOB) were deregulated, showing a disturbance in energy metabolism of the CAU in schizophrenia. These results are in line with alterations in mitochondrial density, size, and distribution already described in neurons and glial cells from the CAU of patients[125].

2.7.3 *Posterior Cingulate Cortices from patients present dysregulation in energy metabolism*

The PCC was the most affected brain region in patients with schizophrenia. Although the PCC presents a high metabolic rate at default states [126], some studies have shown a reduction in its metabolic rate in patients with schizophrenia during verbal working memory tasks[127].This is in line with our proteomic analysis, since it revealed a

---

[122] Pawel Kreczmanski et al., "Volume, Neuron Density and Total Neuron Number in Five Subcortical Regions in Schizophrenia," *Brain* 130, no. Pt 3 (March 2007): 678–92, https://doi.org/10.1093/brain/awl386.

[123] F Fernandez-Enright et al., "Novel Implications of Lingo-1 and Its Signaling Partners in Schizophrenia," *Transl. Psychiatry* 4 (January 21, 2014): e348, https://doi.org/10.1038/tp.2013.121.

[124] Jessica L Andrews and Francesca Fernandez-Enright, "A Decade from Discovery to Therapy: Lingo-1, the Dark Horse in Neurological and Psychiatric Disorders," *Neurosci. Biobehav. Rev.* 56 (September 2015): 97–114, https://doi.org/10.1016/j.neubiorev.2015.06.009.

[125] Natalya A Uranova et al., "Ultrastructural Alterations of Synaptic Contacts and Astrocytes in Postmortem Caudate Nucleus of Schizophrenic Patients," *Schizophr. Res.* 22, no. 1 (October 1996): 81–83, https://doi.org/10.1016/0920-9964(96)00059-X; Natalya Uranova et al., "Electron Microscopy of Oligodendroglia in Severe Mental Illness," *Brain Res. Bull.* 55, no. 5 (July 2001): 597–610, https://doi.org/10.1016/S0361-9230(01)00528-7; Shahza M Somerville, Robert R Conley, and Rosalinda C Roberts, "Mitochondria in the Striatum of Subjects with Schizophrenia," *World J. Biol. Psychiatry* 12, no. 1 (February 2011): 48–56, https://doi.org/10.3109/15622975.2010.505662; Roberts, "Postmortem Studies on Mitochondria in Schizophrenia."

[126] M E Raichle et al., "A Default Mode of Brain Function," *Proc. Natl. Acad. Sci. U. S. A.* 98, no. 2 (January 16, 2001): 676–82, https://doi.org/10.1073/pnas.98.2.676.

[127] Haznedar et al., "Cingulate Gyrus Volume and Metabolism in the Schizophrenia Spectrum"; Kim et al., "Altered Interregional Correlations between Serotonin Transporter Availability and Cerebral Glucose Metabolism in Schizophrenia: A High-Resolution PET Study Using [C]DASB and [F]FDG."

disturbance in carbon and energy metabolism, such as glycolysis, the tricarboxylic acid cycle, and oxidative phosphorylation. These processes were already associated with other brain areas in patients with schizophrenia, with dysregulations in protein and metabolite levels[128]. Proteomic analyses of the anterior cingulate cortex revealed a dysfunction in energy metabolism proteins in schizophrenia and bipolar disorder[129]. It was already described that patients with schizophrenia present anterior cingulate cortex decreased mitochondrial and synaptic density, which could suggest a correlation between energy process and synaptic function[130]. In addition, our data also showed that PCC presented a downregulation in superoxide dismutase 1 (SOD1). SOD1 is antioxidant metalloprotein already described to be downregulated in cerebrospinal fluid[131] and plasma[132] from patients with schizophrenia. Positive symptoms presented in patients with schizophrenia has been inversely associated with SOD levels[133], as well as cognitive deficits related to antioxidant status[134].

Proteins related to synaptic vesicles and the transport of proteins were also found differentially regulated in the PCC. Synaptophysin (SYP) is a presynaptic protein

---

[128] Martins-de-Souza et al., "Proteomic Analysis of Dorsolateral Prefrontal Cortex Indicates the Involvement of Cytoskeleton, Oligodendrocyte, Energy Metabolism and New Potential Markers in Schizophrenia"; Daniel Martins-de-Souza et al., "Proteome Analysis of Schizophrenia Patients Wernicke's Area Reveals an Energy Metabolism Dysregulation," *BMC Psychiatry* 9 (April 30, 2009): 17, https://doi.org/10.1186/1471-244X-9-17; B Dean et al., "Evidence for Impaired Glucose Metabolism in the Striatum, Obtained Postmortem, from Some Subjects with Schizophrenia," *Transl. Psychiatry* 6, no. 11 (November 15, 2016): e949, https://doi.org/10.1038/tp.2016.226.

[129] D Clark et al., "A Proteome Analysis of the Anterior Cingulate Cortex Gray Matter in Schizophrenia," *Mol. Psychiatry* 11, no. 5 (May 2006): 459–70, 423, https://doi.org/10.1038/sj.mp.4001806; M Föcking et al., "Proteomic Analysis of the Postsynaptic Density Implicates Synaptic Function and Energy Pathways in Bipolar Disorder," *Transl. Psychiatry* 6, no. 11 (November 29, 2016): e959, https://doi.org/10.1038/tp.2016.224.

[130] R C Roberts et al., "Decreased Synaptic and Mitochondrial Density in the Postmortem Anterior Cingulate Cortex in Schizophrenia," *Schizophr. Res.* 168, no. 1–2 (October 2015): 543–53, https://doi.org/10.1016/j.schres.2015.07.016.

[131] Jennifer M Coughlin et al., "Reduced Superoxide Dismutase-1 (SOD1) in Cerebrospinal Fluid of Patients with Early Psychosis in Association with Clinical Features," *Schizophr. Res.* 183 (May 2017): 64–69, https://doi.org/10.1016/j.schres.2016.10.040.

[132] Ömer Akyol et al., "The Indices of Endogenous Oxidative and Antioxidative Processes in Plasma from Schizophrenic Patients," *Prog. Neuropsychopharmacol. Biol. Psychiatry* 26, no. 5 (June 2002): 995–1005, https://doi.org/10.1016/S0278-5846(02)00220-8; Cristina Gonzalez-Liencres et al., "Oxidative Stress in Schizophrenia: A Case-Control Study on the Effects on Social Cognition and Neurocognition," *BMC Psychiatry* 14 (September 24, 2014): 268, https://doi.org/10.1186/s12888-014-0268-x.

[133] Zhiwei Wu et al., "Elevated Plasma Superoxide Dismutase in First-Episode and Drug Naive Patients with Schizophrenia: Inverse Association with Positive Symptoms," *Prog. Neuropsychopharmacol. Biol. Psychiatry* 36, no. 1 (January 10, 2012): 34–38, https://doi.org/10.1016/j.pnpbp.2011.08.018.

[134] Xiang Yang Zhang et al., "Plasma Total Antioxidant Status and Cognitive Impairments in Schizophrenia," *Schizophr. Res.* 139, no. 1–3 (August 2012): 66–72, https://doi.org/10.1016/j.schres.2012.04.009.

directly associated to synaptic density[135] (Calhoun *et al.*, 1996) and was linked to intellectual disabilities[136]. Although our results have pointed to an upregulation of SYP, a recent meta-analysis found this protein to be downregulated in patients with schizophrenia in the anterior cingulate cortex and hippocampus[137]. Dysregulation in the synaptic process was already described in metabolomic analysis of anterior cingulate cortex, which presents lower GABA levels in patients with schizophrenia than healthy subjects[138]. Taken together, these data could suggest that synaptic deficits have regional specificities, affecting each area in different ways compared to others.

2.7.4   Immunological system and signalling pathways are dysfunctioned among all brain regions

In all brain regions, the Nuclear Factor-κB Transcriptional Complex (NFκB) were enriched in IPA network analysis. NFκB plays a key role in the immune system, controlling the regulation of inflammatory molecules, such as cytokines[139]. Recently, Volk and colleagues have shown that the prefrontal cortex from patients with schizophrenia presents higher levels of NFκB transcripts, suggesting an immune activation in these patients[140]. In addition, several reports have shown that patients with schizophrenia present dysfunctions in immune systems, including high levels of

---

[135] M E Calhoun et al., "Comparative Evaluation of Synaptophysin-Based Methods for Quantification of Synapses," *J. Neurocytol.* 25, no. 12 (December 1996): 821–28, https://doi.org/10.1007/bf02284844.

[136] Patrick S Tarpey et al., "A Systematic, Large-Scale Resequencing Screen of X-Chromosome Coding Exons in Mental Retardation," *Nat. Genet.* 41, no. 5 (May 2009): 535–43, https://doi.org/10.1038/ng.367.

[137] Emanuele Felice Osimo et al., "Synaptic Loss in Schizophrenia: A Meta-Analysis and Systematic Review of Synaptic Protein and mRNA Measures," *Mol. Psychiatry* 24, no. 4 (April 2019): 20, https://doi.org/10.1038/s41380-018-0041-5.

[138] Ryne C Ramaker et al., "Post-Mortem Molecular Profiling of Three Psychiatric Disorders," *Genome Med.* 9, no. 1 (July 28, 2017): 72, https://doi.org/10.1186/s13073-017-0458-5.

[139] Bastian Hoesel and Johannes A Schmid, "The Complexity of NF-κB Signaling in Inflammation and Cancer," *Mol. Cancer* 12 (August 2, 2013): 86, https://doi.org/10.1186/1476-4598-12-86.

[140] David W Volk et al., "The Role of the Nuclear Factor-κB Transcriptional Complex in Cortical Immune Activation in Schizophrenia," *Biol. Psychiatry* 85, no. 1 (January 1, 2019): 20, https://doi.org/10.1016/j.biopsych.2018.06.015.

cytokines cortical areas [141], overactivation of microglial cells[142], and genomic variants associated with Major Histocompatibility Complex (MHC)[143].

The extracellular-regulated protein kinase (ERK) was also enriched in IPA networking analysis for all the brain regions. This pathway was already found dysregulated in schizophrenia, being associated with a reduction in synaptic plasticity in these patients[144]. In addition, a recent proteomic analysis using the prefrontal cortex from patients with schizophrenia has shown that the suppression in GNA13-ERK signaling affects the synaptic plasticity in the patients[145]. Since synaptic plasticity is a biological process directly associated with learning and memory, dysfunction in this process are associated with the negative and cognitive symptoms found in patients with schizophrenia[146].

2.7.5 Schizophrenia patients present a brain compartmentalization of protein dysfunctions

Taking these dysregulations together, it can be concluded that our results showed differences among the brain regions encompassing proteins and molecular pathways. The changes in the CER were related to dysfunctions in transport processes, while the CAU had associations with synaptic processes and the PCC with energy metabolism and signaling pathways. These results agree with studies that showed divergences of patterns among brain regions in schizophrenia[147]. On the other hand,

---

[141] S G Fillman et al., "Increased Inflammatory Markers Identified in the Dorsolateral Prefrontal Cortex of Individuals with Schizophrenia," *Mol. Psychiatry* 18, no. 2 (February 2013): 206–14, https://doi.org/10.1038/mp.2012.110.

[142] Peter S Bloomfield et al., "Microglial Activity in People at Ultra High Risk of Psychosis and in Schizophrenia: An [(11)C]PBR28 PET Brain Imaging Study," *Am. J. Psychiatry* 173, no. 1 (January 2016): 44–52, https://doi.org/10.1176/appi.ajp.2015.14101358.

[143] Schizophrenia Working Group of the Psychiatric Genomics Consortium, "Biological Insights from 108 Schizophrenia-Associated Genetic Loci," *Nature* 511, no. 7510 (July 24, 2014): 421–27, https://doi.org/10.1038/nature13595.

[144] Peixiong Yuan et al., "Altered Levels of Extracellular Signal-Regulated Kinase Signaling Proteins in Postmortem Frontal Cortex of Individuals with Mood Disorders and Schizophrenia," *J. Affect. Disord.* 124, no. 1–2 (July 2010): 164–69, https://doi.org/10.1016/j.jad.2009.10.017.

[145] Mio Hirayama-Kurogi et al., "Downregulation of GNA13-ERK Network in Prefrontal Cortex of Schizophrenia Brain Identified by Combined Focused and Targeted Quantitative Proteomics," *J. Proteomics* 158 (March 31, 2017): 31–42, https://doi.org/10.1016/j.jprot.2017.02.009.

[146] Lin Mei and Wen-Cheng Xiong, "Neuregulin 1 in Neural Development, Synaptic Plasticity and Schizophrenia," *Nat. Rev. Neurosci.* 9, no. 6 (June 2008): 437–52, https://doi.org/10.1038/nrn2392.

[147] Paolo Fusar-Poli et al., "Neurofunctional Correlates of Vulnerability to Psychosis: A Systematic Review and Meta-Analysis," *Neurosci. Biobehav. Rev.* 31, no. 4 (January 12, 2007): 465–84, https://doi.org/10.1016/j.neubiorev.2006.11.006; Jim van Os and Shitij Kapur, "Schizophrenia," *Lancet* 374, no. 9690 (August 2009): 635–45, https://doi.org/10.1016/S0140-6736(09)60995-8.

our data also showed proteins that were differentially regulated in more than one
region, which were related to the myelin sheath (STXBP1, EHD3, CNP, NEFH, INA,
EEF1A2, NME1, COX5A, GOT2), metabolic pathways (PDHX, PRDX6, PGLS, GLUD2,
GOT2, NME1, COX5A, NDUFS8, ATP5MF, HADH, AHCY), and neuron development
(CNP, NRCAM, NEFH, SERT2, EFHD1, STXBP1, DCLK2, RAB13, UBB, CTNNB1, Figure
5 C-D). 2',3'-Cyclic-nucleotide 3'-phosphodiesterase (CNP) is an oligodendrocyte marker
associated with the size of the oligodendrocyte population in tissue. Our results showed
lower levels of CNP in the CAU and PCC in schizophrenia, suggesting a decrease in the
oligodendrocyte population in schizophrenia. Studies have shown decreased
oligodendrocyte density in the CAU[148] and prefrontal cortex[149], and in the anterior
frontal cortex, anterior cingulate cortex, and hippocampus [150] of patients with
schizophrenia. Since oligodendrocytes play a pivotal role in myelination and energy
support to neuron axons [151], we suggest that dysfunctions in both processes are
intrinsically related to the whole-brain dysfunctions seen in schizophrenia.

2.7.6   Limitations

       When working with postmortem brain tissues in psychiatric disorders, the
main limitations are associated with the inherent characteristics of this model, such as
the sample size, methods of collection, and storage conditions. In this study, the major
limitation is indeed the modest sample size used in our analysis. In addition, the samples
used in this study were collected from patients that were treated with antipsychotics.
Although the treatment may affect the identification of novel proteins associated with
schizophrenia, Chan and colleagues showed that antipsychotics normalize the effects of

---

[148] V M Vostrikov and N A Uranova, "Reduced Density of Oligodendrocytes and Oligodendrocyte Clusters in the Caudate Nucleus in Major Psychiatric Illnesses," *Schizophr. Res.*, October 22, 2019, https://doi.org/10.1016/j.schres.2019.10.027.

[149] Natalya A Uranova et al., "The Role of Oligodendrocyte Pathology in Schizophrenia," *Int. J. Neuropsychopharmacol.* 10, no. 4 (August 2007): 537–45, https://doi.org/10.1017/S1461145707007626.

[150] S W Flynn et al., "Abnormalities of Myelination in Schizophrenia Detected in Vivo with MRI, and Post-Mortem with Analysis of Oligodendrocyte Proteins," *Mol. Psychiatry* 8, no. 9 (September 2003): 811–20, https://doi.org/10.1038/sj.mp.4001337; Stella Dracheva et al., "Myelin-Associated mRNA and Protein Expression Deficits in the Anterior Cingulate Cortex and Hippocampus in Elderly Schizophrenia Patients," *Neurobiol. Dis.* 21, no. 3 (March 2006): 531–40, https://doi.org/10.1016/j.nbd.2005.08.012.

[151] Aiman S Saab, Iva D Tzvetanova, and Klaus-Armin Nave, "The Role of Myelin and Oligodendrocytes in Axonal Energy Metabolism," *Curr. Opin. Neurobiol.* 23, no. 6 (December 2013): 1065–72, https://doi.org/10.1016/j.conb.2013.09.008.

disease according to lifetime medication[152]. Finally, each brain region was analyzed with different chromatographic and MS acquisition methods, which affect protein identification and comparison between the regions. Despite these limitations, postmortem brain tissue is a unique model to perform molecular analysis directly in the brains of patients with schizophrenia. In addition, our findings are supported by several other results in the literature and for all experiments, we applied the same statistical parameters to obtain the highest proteome coverage and trust in the data as possible.

2.8 CONCLUSION

This study provides a differential pattern of proteins in the CER, CAU and PCC from patients with schizophrenia compared with healthy controls. Our results pointed to 135, 106 and 727 deregulated proteins in the aforementioned regions, respectively. In addition, pathway enrichment analysis showed dysfunctions in synaptic processes in the CAU, transport in the CER, and energy metabolism in the PCC. Our study also reinforces the role of oligodendrocytes and metabolic processes in these areas in schizophrenia. Further analyses using other Omic approaches, such as lipidomics and metabolomics, are needed to complement our results and show if these changes can be seen among the brain regions also in metabolic levels.

---

[152] M K Chan et al., "Evidence for Disease and Antipsychotic Medication Effects in Post-Mortem Brain from Schizophrenia Patients," *Mol. Psychiatry* 16, no. 12 (December 2011): 1189–1202, https://doi.org/10.1038/mp.2010.100.

3   CHAPTER 2: OMICSCOPE: FROM DIFFERENTIAL PROTEOMICS TO SYSTEMS BIOLOGY

3.1 ABSTRACT

Shotgun proteomics analysis presents multifaceted challenges, demanding diverse tool integration for insights. Addressing this complexity, OmicScope emerges as a comprehensive solution for quantitative proteomics data analysis. Engineered to handle various data formats, it conducts differential proteomics analysis for both static and longitudinal designs. Empowered with Enrichr, OmicScope allows user to perform Over Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) with over 224 databases. Additionally, its Nebula module facilitates multi-omics integrative analysis, providing a systems biology approach for enriched insights. Complete with a data visualization toolkit and accessible as both a Python package and a web application (https://omicscope.ib.unicamp.br/), OmicScope democratizes proteomics analysis, offering an efficient, high-quality pipeline for researchers.

3.2 INTRODUCTION

Mass spectrometry-based proteomics has emerged as an indispensable tool for unraveling the intricate molecular mechanisms underlying complex diseases and biological phenomena. This technique enables simultaneous interrogation of thousands of proteins, allowing for the discovery of novel protein candidates without prior knowledge or defined targets. The flexibility provided by shotgun proteomics enables large-scale exploration of protein-protein interaction (PPI) networks[153], subcellular processes[154], protein thermal stability[155], protein-drug interactions [156], and signaling

---

[153] Liu et al., "The Interactome of Intact Mitochondria by Cross-Linking Mass Spectrometry Provides Evidence for Coexisting Respiratory Supercomplexes."

[154] Claire M. Mulvey et al., "Spatiotemporal Proteomic Profiling of the Pro-Inflammatory Response to Lipopolysaccharide in the THP-1 Human Leukaemia Cell Line," *Nature Communications* 12, no. 1 (October 1, 2021): 5773, https://doi.org/10.1038/s41467-021-26000-9; Peter J. Thul et al., "A Subcellular Map of the Human Proteome," *Science* 356, no. 6340 (May 26, 2017): eaal3321, https://doi.org/10.1126/science.aal3321.

[155] Nils Kurzawa et al., "Deep Thermal Profiling for Detection of Functional Proteoform Groups," *Nature Chemical Biology* 19, no. 8 (August 2023): 962–71, https://doi.org/10.1038/s41589-023-01284-8; Savitski et al., "Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome."

[156] Bantscheff et al., "Quantitative Chemical Proteomics Reveals Mechanisms of Action of Clinical ABL Kinase Inhibitors"; Lechner et al., "Chemoproteomic Target Deconvolution Reveals Histone Deacetylases as Targets of (R)-Lipoic Acid."

pathways[157]. Furthermore, due to the pivotal role of proteins in bridging genotypic information to phenotypic outcomes, proteomics complements other 'omics disciplines, including genomics, transcriptomics, and metabolomics. Despite its widespread adoption and continuous technical advancements, proteomics continues to present substantial challenges, with data analysis complexity remaining a prominent issue.

To achieve comprehensive information from proteomics data, the current approach requires the use of an array of software tools. This encompasses raw data processing, protein identification, quantitation, differential proteomics, and enrichment analysis (e.g., MaxQuant[158], PatternLab V[159], DIA-NN[160], Perseus[161], PatternLab V, Progenesis QI for Proteomics, MSstats [162]; DAVID [163], Enrichr [164]). While this conventional pipeline forms the core of proteomics data analysis, additional steps like exploring protein-protein interactions, conducting network analyses, and customizing data visualization are often necessary.

The complexity and number of computational tools present an initial barrier, particularly for non-programmers and newcomers to proteomics, as mastering the functions, capabilities, and limitations of each software tool demands a steep learning curve. Furthermore, the challenge with tools for differential proteomics analysis involves accommodating the wide spectrum of data formats generated by search engines and quantitative proteomics software (Supplementary Table 1). Considering the distinctive features and necessities of each tool, encompassing data normalization, structure, and treatment of missing values, any new software tool must account for these varied input options. Additionally, a versatile and generic format is crucial to facilitate data importation from emerging proteomics tools, incorporate innovative

---

[157] Needham et al., "Personalized Phosphoproteomics Identifies Functional Signaling."

[158] Cox and Mann, "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification."

[159] Santos et al., "Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V."

[160] Demichev et al., "DIA-NN."

[161] Tyanova et al., "The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data."

[162] "MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale | Journal of Proteome Research," accessed September 30, 2023, https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00834.

[163] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki, "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources," *Nature Protocols* 4, no. 1 (January 2009): 44–57, https://doi.org/10.1038/nprot.2008.211.

[164] Chen et al., "Enrichr."

statistical analyses, and integrate data from other 'omics' platforms, such as genomics and transcriptomics.

Statistical analysis for differential proteomics can exhibit substantial variations contingent upon the experimental design adopted by researchers (Supplementary Table 1). Typically, experimental designs can be broadly categorized into two main types: static and longitudinal [165] . Despite the common occurrence of both static and longitudinal analyses, the majority of proteomics tools often lack comprehensive coverage of the wide spectrum of experimental designs.

Within the proteomics workflow, enrichment analysis assumes a pivotal role in aiding researchers to uncover system-level biological insights. One widely adopted approach is over-representation analysis (ORA), wherein experimentally derived entities are compared against annotated databases to ascertain whether biologically relevant properties are overrepresented in the experimental gene list[166]. Alternatively, gene-set enrichment analysis (GSEA) compares two conditions to find coordinated changes in gene expression in biologically relevant databases[167]. The inherent data requirements associated with enrichment analysis compel researchers to seek web tools that provide extensive datasets to perform such analyses. Enrichr, for instance, offers access to over 224 distinct and regularly updated libraries[168]. Enrichment analysis typically plays a pivotal role in the proteomics workflow, highlighting potential biological pathways for targeted validation through orthogonal approaches.

Despite the importance of enrichment analysis within the broader proteomics workflow, its power can be significantly enhanced by comparing individual experiments with independent results or data published by third-party research groups. This meta-analysis approach reduces false discovery rates and enables a more reliable assessment of molecular features associated with biological phenomena, as numerous sources contribute to a systems-level investigation[169]. While specific enrichment tools

---

[165] Storey et al., "Significance Analysis of Time Course Microarray Experiments."

[166] Boyle et al., "GO."

[167] Aravind Subramanian et al., "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles," *Proceedings of the National Academy of Sciences* 102, no. 43 (October 25, 2005): 15545–50, https://doi.org/10.1073/pnas.0506580102.

[168] Chen et al., "Enrichr."

[169] Zhou et al., "Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets"; Tripathi et al., "Meta- and Orthogonal Integration of Influenza 'OMICs' Data Defines a Role for UBR4 in Virus Budding."

support the analysis of multiple gene lists, many existing proteomics tools frequently lack the capacity to conduct a broad analysis across multiple experiments and integrate data with other omics technologies (Supplementary Table 1). For researchers aiming to analyze multiple gene lists, several web-based tools are accessible, including Metascape[170] and DAVID[171]. Nevertheless, these tools often overlook crucial protein statistics, such as protein fold changes, and offer a restricted array of figures that can integrate enrichment outcomes with respective protein attributes.

In light of the complexities inherent in the proteomics workflow, we introduce OmicScope—an integrated solution designed to streamline proteomics data analysis from differential expression to system-level integration. Here, we elucidate the features of OmicScope, which is available both as a Python package and a web application (https://omicscope.ib.unicamp.br). OmicScope empowers users to perform differential expression analysis across diverse platforms, conduct enrichment analysis with access to over 224 libraries offered by Enrichr, and seamlessly integrate independent studies to enhance biological insights. Additionally, we offer a comprehensive suite of graphical outputs, encompassing gold-standard proteomics figures and a unique set of plots tailored to enhance comprehension in individual experiments and systems-level analyses. OmicScope also provides a versatile array of output files for integration with third-party software, including tables, vectorized images, network files, and its output format. Together, OmicScope stands as a user-friendly tool engineered to facilitate proteomics data analysis, integration, and interpretation for the entire research community.

3.3 METHODS

3.3.1 OmicScope architecture and User Interface

OmicScope is developed in Python (v. 3.11) and distributed in Pypi repository (https://pypi.org/project/omicscope/) under MIT license. The comprehensive

---

[170] Zhou et al., "Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets."

[171] Huang, Sherman, and Lempicki, "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources," January 2009.

OmicScope pipeline comprises three modules: OmicScope, EnrichmentScope and
Nebula (Figure 1), designed to handle differential expression analysis, enrichment
analysis, and multi-study comparisons, respectively.

The web application (https://omicscope.ib.unicamp.br) was developed using
the Streamlit framework, leveraging the OmicScope package in the background.
Interactive figures were developed separately using Altair v.4 [172] and the Vega
visualization grammar. Step-by-step instructions for using the web application are
provided in the Supplementary material, covering both OmicScope and Nebula
workflows.

3.3.2 *Study design and data collection*

For our study, we utilized previously published COVID-19 datasets to run the
OmicScope pipeline. Crunfli et al. [173] (PXD023781) performed label-free quantitative
proteomics on Progenesis QI for Proteomics, evaluating 2,278 proteins in post-mortem
brain tissue from COVID-19 patients. They identified 735 proteins as differentially
regulated (pAdjusted < 0.05, Supplementary Table 2). This biological relevant dataset
served as a demonstration for a single analysis in OmicScope, showcasing its ability to
handle quantitative proteomics data, generate publication quality figures, and perform
enrichment analysis. Moreover, we conducted the complete workflow within
OmicScope using the benchmark dataset supplied by Demichev, and the corresponding
results can be found in Supplementary Figure 2. Demichev applied Frag-Pipe and DIA-
NN methodologies to analyze data initially published by Meier. Meier's study involved
spiking 15 ng and 45 ng of Yeast digest into Hela digest (PXD017703).

To showcase the capabilities of the Nebula module, we combined Crunfli's
dataset with differentially regulated proteins/genes found by Nie et al. 2021[174] and
Wang et al. 2021[175]. Nie and colleagues evaluated autopsied tissues from seven organs
of COVID-19 patients, of which we focused on the liver, the most affected organ

---

[172] Jacob VanderPlas et al., "Altair: Interactive Statistical Visualizations for Python," *Journal of Open Source Software* 3, no. 32 (December 10, 2018): 1057, https://doi.org/10.21105/joss.01057.

[173] Fernanda Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients," *Proceedings of the National Academy of Sciences* 119, no. 35 (August 30, 2022): e2200960119, https://doi.org/10.1073/pnas.2200960119.

[174] Xiu Nie et al., "Multi-Organ Proteomic Landscape of COVID-19 Autopsies," *Cell* 184, no. 3 (February 4, 2021): 775-791.e14, https://doi.org/10.1016/j.cell.2021.01.004.

[175] Si Wang et al., "A Single-Cell Transcriptomic Landscape of the Lungs of Patients with COVID-19," *Nature Cell Biology* 23, no. 12 (December 2021): 1314–28, https://doi.org/10.1038/s41556-021-00796-6.

according to their findings. Additionally, Wang and colleagues assessed protein and RNA levels in lungs from COVID-19 patients, providing differentially regulated genes and proteins. Files related to single-study analyses or Nebula workflows are available as supplementary files.

3.3.3 Input and data structure

OmicScope offers six distinct modules for integrating external data into its pipeline, four of which rely on widely used proteomic software for protein identification and quantitation: Progenesis Qi for Proteomics, PatternLab V[176], MaxQuant[177], and DIA-NN [178]. For alternative sources, the General and Snapshot methods support data importing, with Snapshot offering a more concise file format with limited information. Additionally, our import methods categorize raw data into three primary matrices: assay (protein abundances), phenotype data (metadata or pdata), and row data (protein information). Users also have the option to independently import pdata into the workflow, replacing original pdata when necessary, to tailor statistical analysis to specific needs. Detailed information on input data, method specifications, and data structure can be found in the appendix and supplementary table 3.

3.3.4 Differential Proteomics Analysis – OmicScope Module

Differential proteomics analysis within the OmicScope workflow is optional, contingent on input data. In cases where p-value information is absent, OmicScope prior joins technical replicates calculating average protein abundance, followed by filtering stage of proteins detected in all conditions. By default, OmicScope applies log2-transformation for statistical tests.

OmicScope conducts two data analysis pipelines based on experimental design: static and longitudinal. In the static approach, it performs t-tests or Analysis of Variance (ANOVA) for two or more conditions, respectively. For longitudinal analysis, OmicScope adapted the workflow proposed by Storey in 2005, employing a natural cubic spline in a generalized linear model to model gene expression over time[179]. After

---

[176] Santos et al., "Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V."
[177] Cox and Mann, "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification."
[178] Demichev et al., "DIA-NN."
[179] Storey et al., "Significance Analysis of Time Course Microarray Experiments."

obtaining nominal protein p-values, OmicScope corrects for multiple hypotheses using the Benjamini-Hochberg method[180]. Further details on the statistical workflow are provided in the appendix.

The result of differential proteomics analysis, even if previously imported into OmicScope, is the quantitative data (quant_data). Quantitative data is a comprehensive table containing all protein information available, including gene names, log2-transformed fold changes, log10-transformed p-values, mean protein abundances across conditions, and more. This data serves as the basis for generating figures in the OmicScope module and is also employed for enrichment analysis and figures in the EnrichmentScope module.

3.3.5   *Enrichment Analysis – EnrichmentScope module*

Enrichment analysis is an integral part of the OmicScope workflow, handled within the EnrichmentScope module. This module employs the GSEApy package[181] to conduct two distinct methods of enrichment analysis: over-representation analysis (ORA) and Gene-Set Enrichment Analysis (GSEA). EnrichmentScope leverages data from the OmicScope object to query databases provided by the Enrichr API. OmicScope currently supports Human, Mouse, Yeast, Fly, Fish, and Worm as target organisms due to its dependency on Enrichr[182].

For ORA, EnrichmentScope uses the proteins differentially regulated in the OmicScope object as input, employing a Fisher's exact test against target databases. In contrast, GSEA utilizes assay data and phenotype data to determine whether a database term is significantly enriched at the top or bottom of a ranked list of genes based on their differential regulation. In both cases, EnrichmentScope generates a table containing evaluated terms, Benjamini-Hochberg-adjusted p-values, Combined Score (ORA) or Normalized Enrichment Score (GSEA), log10-transformed p-values, proteins related to each term, and the respective regulation of each protein.

---

[180] Benjamini and Hochberg, "Controlling the False Discovery Rate."

[181] Zhuoqing Fang, Xinyuan Liu, and Gary Peltz, "GSEApy: A Comprehensive Package for Performing Gene Set Enrichment Analysis in Python," *Bioinformatics (Oxford, England)* 39, no. 1 (January 1, 2023): btac757, https://doi.org/10.1093/bioinformatics/btac757.

[182] Chen et al., "Enrichr."

3.3.6 Multi-Studies Integration – Nebula Module

The Nebula module within OmicScope facilitates the integration of data from independent studies, enabling multi-omics analysis. Input data for Nebula is provided by the OmicScope and EnrichmentScope modules. While OmicScope exports quantitative analysis, EnrichmentScope exports both quantitative and enrichment results. The output is a text file with omics extension, including conditions and statistical parameters applied in the respective study. For each analysis conducted in OmicScope, one file can be exported and later imported into Nebula.

To import data into the Nebula workflow, users must place all exports in the same directory (or zip file, in web application) and import them collectively. Once imported, Nebula divides each file into quantitative and, where applicable, enrichment data. Differentially regulated proteins are defined based on a user-defined p-value cutoff, consistently applied to all imported data. To ensure reproducibility and consistency between OmicScope and EnrichmentScope outputs, Nebula remaps genes reported in EnrichmentScope to match the naming and format used in OmicScope data.

3.3.7 Outputs

Each module described above offers a unique set of visualization tools tailored to specific analyses (see appendix for details). In the OmicScope package, figures are primarily generated using matplotlib, seaborn, networkx, and pycirclize[183] packages. In the web application, plots are generated based on Altair and Vega libraries. This setup allows for the export of figures in vectorized formats (e.g., SVG) and high-resolution images. Additionally, data used to construct networks can be exported as graphml files, compatible with network-specific tools like Cytoscape[184] and Gephi[185].

3.4 RESULTS

3.4.1 Overview

The OmicScope pipeline includes three primary components: OmicScope, EnrichmentScope, and Nebula (Figure 1). Briefly, once quantitative data is inserted into

---

[183] Yuki Shimoyama, "pyCirclize: Circular Visualization in Python," Python, December 2022, https://github.com/moshi4/pyCirclize.
[184] Shannon et al., "Cytoscape."
[185] Bastian, Heymann, and Jacomy, "Gephi."

the workflow, the OmicScope determines differentially regulated proteins (DRPs). These DRPs are then subjected to enrichment analysis using the EnrichmentScope algorithm, aiming to elucidate key biological features. Additionally, individual studies analyzed using the OmicScope and/or EnrichmentScope algorithms can be exported and used as input for Nebula. In Nebula, users can analyze independent studies collectively, establishing correlations and identifying shared features across independent results. Each component, when activated, generates a set of figures and tables, streamlining user interactions for both the package and web application.
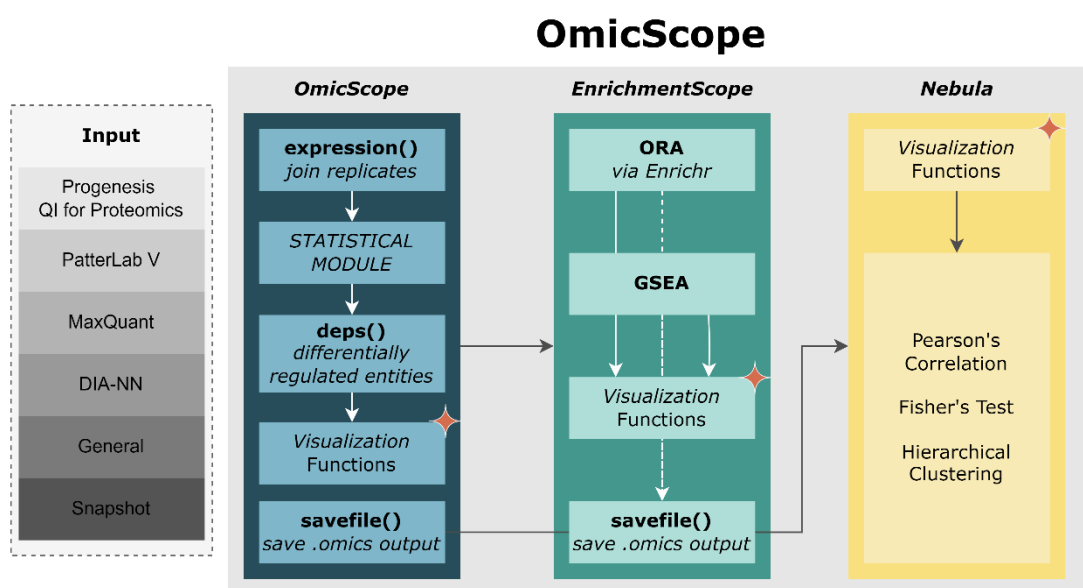
Figure 3-1 OmicScope workflow. The OmicScope workflow begins with the import of data from various sources, including outputs from proteomics tools and generic formats. Once imported, the OmicScope module determines differentially regulated proteins. These proteins are then directed to the EnrichmentScope module, which facilitates over-representation and gene-set enrichment analyses. Data derived from both OmicScope and EnrichmentScope can be seamlessly used as input for Nebula, a module that integrates results from multiple studies using a systems biology approach. Each module within OmicScope is equipped with its own visualization toolset and allows for the export of tables, vectorized images, and graphml files.

Proteomics research exhibits substantial diversity in experimental workflows, including mass spectrometer selection, acquisition modes, fragmentation methods, and quantitative approaches. This inherent diversity requests a wide array of software tools for protein identification and quantitation, each with its strengths and limitations, leading to interoperability challenges[186].

---

[186] Chen Chen et al., "Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis," *International Journal of Molecular Sciences* 21, no. 8 (April 20, 2020): 2873, https://doi.org/10.3390/ijms21082873; Bantscheff et al., "Quantitative Mass Spectrometry in Proteomics."

To address these challenges, OmicScope offers six data import methods (See Methods, Appendix and Figure 1), including four tailored to widely adopted proteomic software: MaxQuant, PatternLab V, DIA-NN, and Progenesis QI for Proteomics. These methods seamlessly import outputs from respective software considering their unique characteristics.

For software not yet integrated into OmicScope, the "General" method allows users to create custom spreadsheets for input into the OmicScope pipeline. This method accepts generic expression files, making OmicScope compatible with data from various omics platforms, such as genomics and transcriptomics. "General" method is able to perform differential proteomics analysis or import existing statistical analyses based on imported spreadsheet.

Aiming to provide an import method that joins succinctness, simplicity, and speed, we implemented "Snapshot" method, in which the users can import proteomics results containing assessed proteins, along with their associated fold changes and statistical outcomes. While Snapshot presents certain limitations concerning the number of plots that can be generated (Supplementary Table 3), this method substantially improves interoperability across studies, especially given that many studies typically provide restricted information from their analyses, as demonstrated in the cases of Nie 2021 and Wang 2021.

3.4.2   OmicScope: the core module

The central module of OmicScope shares the same name as the tool described herein. This module plays a pivotal role in data organization and filtration, as well as in conducting differential proteomics analysis, defining differentially regulated entities, and generating ready-to-publish figures (Figure 2A).

To provide maximum versatility, pre-filtering and statistical steps are optional within the OmicScope pipeline. When no statistical results are provided, OmicScope autonomously conducts statistical analysis, filtering data based on pre-specified parameters and selecting the most suitable statistical tests based on the data architecture (see Appendix section for details). This flexible architecture accommodates various experimental designs, including static and longitudinal approaches. In static cases, comparisons between independent groups are typically made using t-tests for binary comparisons or One-way ANOVA for more than two independent conditions. In

longitudinal analyses, OmicScope employs the Storey approach[187], considering that differentially regulated genes vary over time based on natural cubic splines. In this longitudinal approach, statistical evaluations consider both within-group and between-group comparisons. Once nominal p-value is calculated, OmicScope performs Benjamini-Hochberg multiple hypothesis correction [188] . By default, OmicScope designates proteins as differentially regulated if their adjusted p-value is below 0.05, although users can define other parameters, such as fold-change and nominal p-value cutoffs (Figure 2A).

OmicScope module offers a visualization toolkit for data overviewing, clustering, and protein-specific features (Supplementary figure 1). In the overview category, users can generate bar plots, volcano plots, MA-plots, and dynamic range plots, facilitating the visualization of data distribution and normalization, providing initial insights into the dataset. The clustering category includes functions for hierarchical clustering, principal component analysis (PCA), and K-means clustering, allowing users to compare samples based on protein abundances and assess sample clustering. In this category, users can select various metrics and calculation methods to perform the data clustering. Lastly, protein-specific category aims to extract deeper insights about selected proteins, using bar plots and box plots. In this category, OmicScope also includes an integration with STRING API, providing a protein-protein interaction network of differentially regulated proteins.

To demonstrate the capabilities of OmicScope, we employed previously published COVID-19 studies as illustrative examples (refer to the Methods section for details). These studies employed quantitative proteomics and transcriptomics to investigate SARS-CoV-2's effects on various tissues. Specifically, we conducted a single analysis example, showcasing both differential proteomics and enrichment analysis, using proteins quantified by Crunfli[189] in the brain tissue of patients who succumbed to SARS-CoV-2 complications. In this study, the authors meticulously detailed the processing parameters and furnished quantitative outputs from the analysis, ensuring

---

[187] Storey et al., "Significance Analysis of Time Course Microarray Experiments."
[188] Benjamini and Hochberg, "Controlling the False Discovery Rate."
[189] Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients."

reproducibility, and enabling result comparisons. For multi-study integration, we incorporated data from Crunfli, Nie[190], and Wang[191] studies.

Crunfli's dataset was imported into OmicScope with default parameters, filtering out contaminants[192], and resulting in the identification of 721 differentially regulated proteins (Figure 2B). After OmicScope defines the differentially regulated proteins, proteomics figures can be generated using a dedicated function for each plot type. For scatter plots and heatmaps, users can specify gene names as arguments to highlight specific target proteins (as demonstrated in Figure 2C). Additionally, for clustering analyses, users optionally can set a p-value cutoff to filter proteins and conduct analyses based on statistical significance (Figure 2D).

In Crunfli's dataset, for instance, we selected the MAPK family, including MAPK1, MAPK14, and MAPK9, all of which showed upregulation in SARS-CoV-2 infection compared to the control group (Figure 2E). Moreover, the protein-specific category includes a function for exploring protein-protein interactions (PPIs) by querying the STRING database [193]. In this network analysis, users can identify communities based on the Louvain algorithm[194] and filter data based on protein p-values and/or specific proteins. In our analyzes, we filtered proteins based on a p-value threshold (pAdjusted < 0.005), applied the Louvain algorithm to conduct modularity analysis, and exported the data to facilitate data visualization (Figure 2F).

While the Crunfli dataset offers advantages for our pivotal analysis, it does pose a limitation due to the relatively small number of evaluated proteins in the study. To address this, we expanded our analysis by utilizing the OmicScope workflow with a quantitative proteomics dataset provided by Meier [195] and Demichev [196]. Meier introduced two distinct concentrations of Yeast digest into 200 ng of HeLa digest, while

---

[190] Nie et al., "Multi-Organ Proteomic Landscape of COVID-19 Autopsies."

[191] Wang et al., "A Single-Cell Transcriptomic Landscape of the Lungs of Patients with COVID-19."

[192] Ashley M. Frankenfield et al., "Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics," *Journal of Proteome Research* 21, no. 9 (September 2, 2022): 2104–13, https://doi.org/10.1021/acs.jproteome.2c00145.

[193] Szklarczyk et al., "The STRING Database in 2021."

[194] Traag, Waltman, and van Eck, "From Louvain to Leiden."

[195] Florian Meier et al., "diaPASEF: Parallel Accumulation–Serial Fragmentation Combined with Data-Independent Acquisition," *Nature Methods* 17, no. 12 (December 2020): 1229–36, https://doi.org/10.1038/s41592-020-00998-0.

[196] Vadim Demichev et al., "Dia-PASEF Data Analysis Using FragPipe and DIA-NN for Deep Proteomics of Low Sample Amounts," *Nature Communications* 13, no. 1 (July 8, 2022): 3944, https://doi.org/10.1038/s41467-022-31492-0.

Demichev employed Frag-Pipe and DIA-NN workflows, resulting in the evaluation of over 12,000 proteins, specifically identifying differentially regulated proteins from the yeast digest. Through OmicScope, we pinpointed proteins highlighted by the authors exhibiting two distinct expression profiles, emphasizing differential abundance among yeast protein concentrations (Supplementary figure 2). These outcomes highlight OmicScope's capacity to handle varying data formats and sizes, performing a robust analysis of differential proteomics.
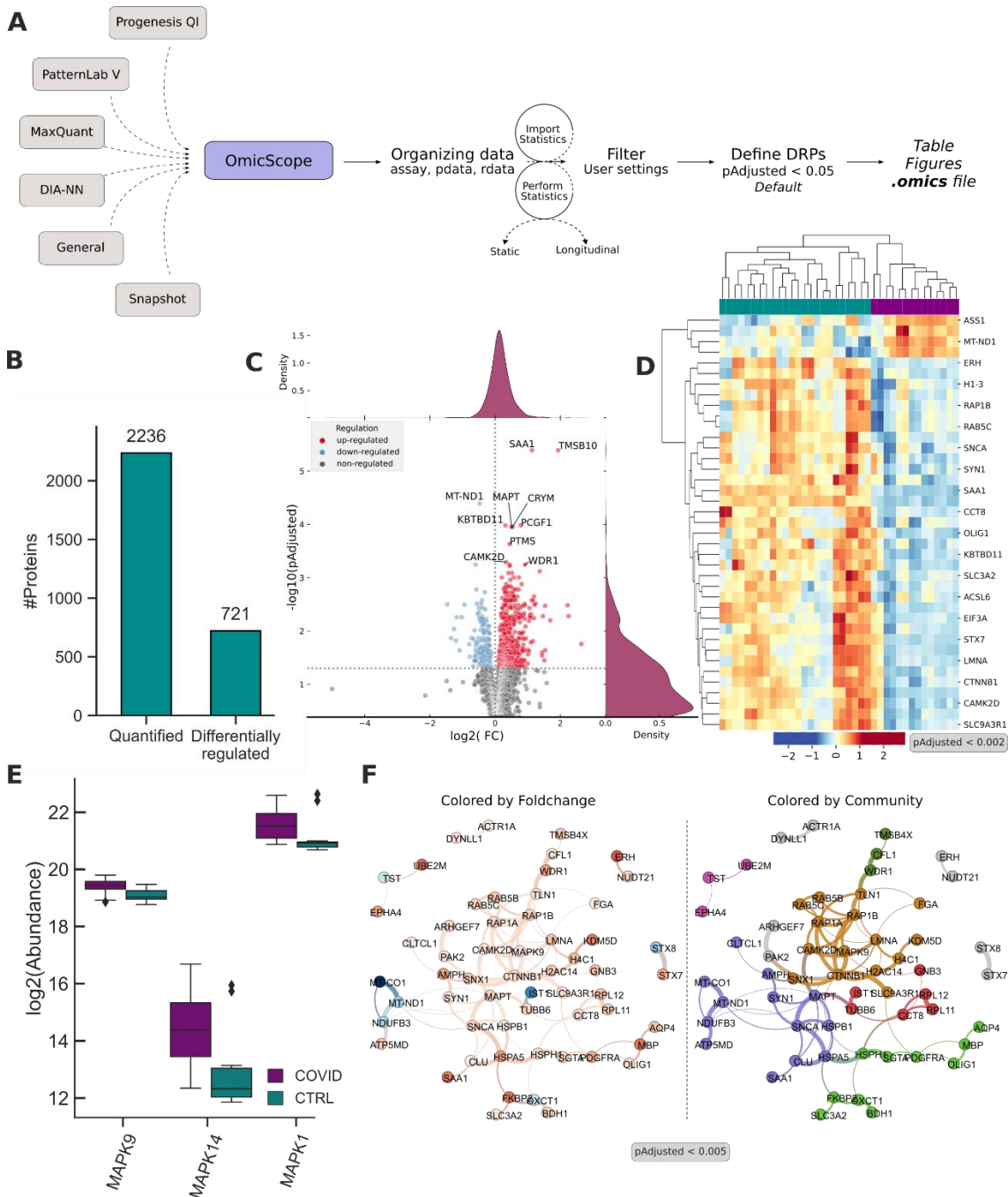
*Figure 3-2 OmicScope provides a robust platform for differential proteomics analysis and comprehensive data
visualization.* **A)** OmicScope offers various data import methods, including established software and
generic approaches. Once data is successfully imported, OmicScope defines data architecture, performs
differential proteomics analysis, filters data, identifies differentially regulated proteins (DRPs), and
generates tables, figures, and exports. **B-E)** Illustrative figures generated by OmicScope: **B)** Bar plot
displaying the count of identified proteins and DRPs. **(C)** Volcano plot with accompanying density plot
highlighting the top 10 DRPs based on Adjusted p-value. **D)** Heatmap of DRPs with Adjusted p-value less
than 0.002, with colors representing log2(fold-change). COVID-19 patients and controls are denoted as
dark cyan and purple, respectively. **E)** Boxplot depicting the abundance of proteins identified from the
MAPK family. **F)** Protein-protein interaction network generated by OmicScope with DRPs having an
Adjusted p-value less than 0.005. In the left graph, proteins are colored based on log2(fold change), while
the right graph represents proteins colored according to their communities identified using Louvain
algorithms.

3.4.3   EnrichmentScope: enhancing biological insights

One of the critical and challenging aspects of omics studies is extracting meaningful biological insights from hundreds or even thousands of differentially regulated entities. A commonly applied method for this purpose is enrichment analysis, wherein experimental gene or protein sets are compared against pre-established datasets, which may encompass biological pathways, molecular functions, kinase-associated genes, and other relevant categories. EnrichmentScope addresses this challenge by furnishing specialized enrichment analysis capabilities.

After executing the OmicScope module, users can proceed to perform enrichment analysis on EnrichmentScope module, specifying between two approaches: Over-Representation Analysis (ORA, conventional enrichment) or Gene-set Enrichment Analysis (GSEA). Then, users must select specific databases, choosing between the 224 libraries offered by Enrichr[197]. Noteworthy, EnrichmentScope can optionally consider more stringent background, using all proteins identified in the study for enrichment analysis. Once the analysis is performed, the module provides a result table and a toolkit of visualization functions, including the ability to export quantitative and enrichment data (Figure 3A, Supplementary Figure 3).

EnrichmentScope offers effective visualization tools like dot plots, facilitating the assessment of enrichment statistics and the number of proteins considered for enrichment (Figures 3B–C). Users can select top enriched terms based on adjusted p-values to identify relevant biological processes (Figure 3B). Another dot plot option allows users to explore protein regulation in depth, illustrating the number of differentially regulated proteins and splitting in up- and down-regulation in each enriched term (Figure 3C). In Crunfli's study, for instance, the top 10 enriched terms using KEGG Database were filtered, and pathways related to neurodegenerative diseases were selected, showing the ratio of up- and down-regulated proteins (Figures 3B–C).

EnrichmentScope also generates heatmaps and network graphs, both linking enriched terms to respective proteins (Figure 3D-E). These visualizations reveal protein fold changes and proteins overlap among groups, shedding light on key factors in biological events. In the previously chosen pathways, proteins related to processes such

---

[197] Chen et al., "Enrichr."

as the proteasome, electron transport chain, and cytoskeleton were shared across all neurodegenerative processes, offering insights into the effects of SARS-CoV-2 on COVID-19 patients (Figure 3E). Following this analysis, users can further investigate proteins of interest within the OmicScope module using functions like box plots, protein-protein interaction networks, and more.

A challenge encountered in enrichment analysis is dealing with data redundancy, particularly prevalent in hierarchical databases such as Reactome[198] and Gene Ontology[199], which can lead to an overwhelming amount of information, as many pathways indicate a similar biological function (Supplementary Figure 4). To address this limitation, EnrichmentScope apply systems biology approach similar to what is proposed by EnrichmentMap, wherein enrichment terms are represented as nodes within a network[200] (Figure 3F). The enrichment map is created by calculating pairwise Jaccard similarity indices, considering genes shared between enriched terms (See appendix). By default, EnrichmentScope establishes links between terms when the Jaccard Similarity Index exceeds 0.25, enabling graph construction. Additionally, EnrichmentScope automatically searches for communities within the enrichment map, labeling nodes (terms) that present highest intra-module degree. Besides providing a cleaner representation of the network, this strategy also simplifies information extraction, reduces data redundancy without omitting any data, and aids in selecting targets for further experimental validation (Figure 3F).

[198] Gillespie et al., "The Reactome Pathway Knowledgebase 2022."

[199] Michael Ashburner et al., "Gene Ontology: Tool for the Unification of Biology," *Nature Genetics* 25, no. 1 (May 2000): 25–29, https://doi.org/10.1038/75556.

[200] "Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation | PLOS ONE," accessed October 1, 2023, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013984.
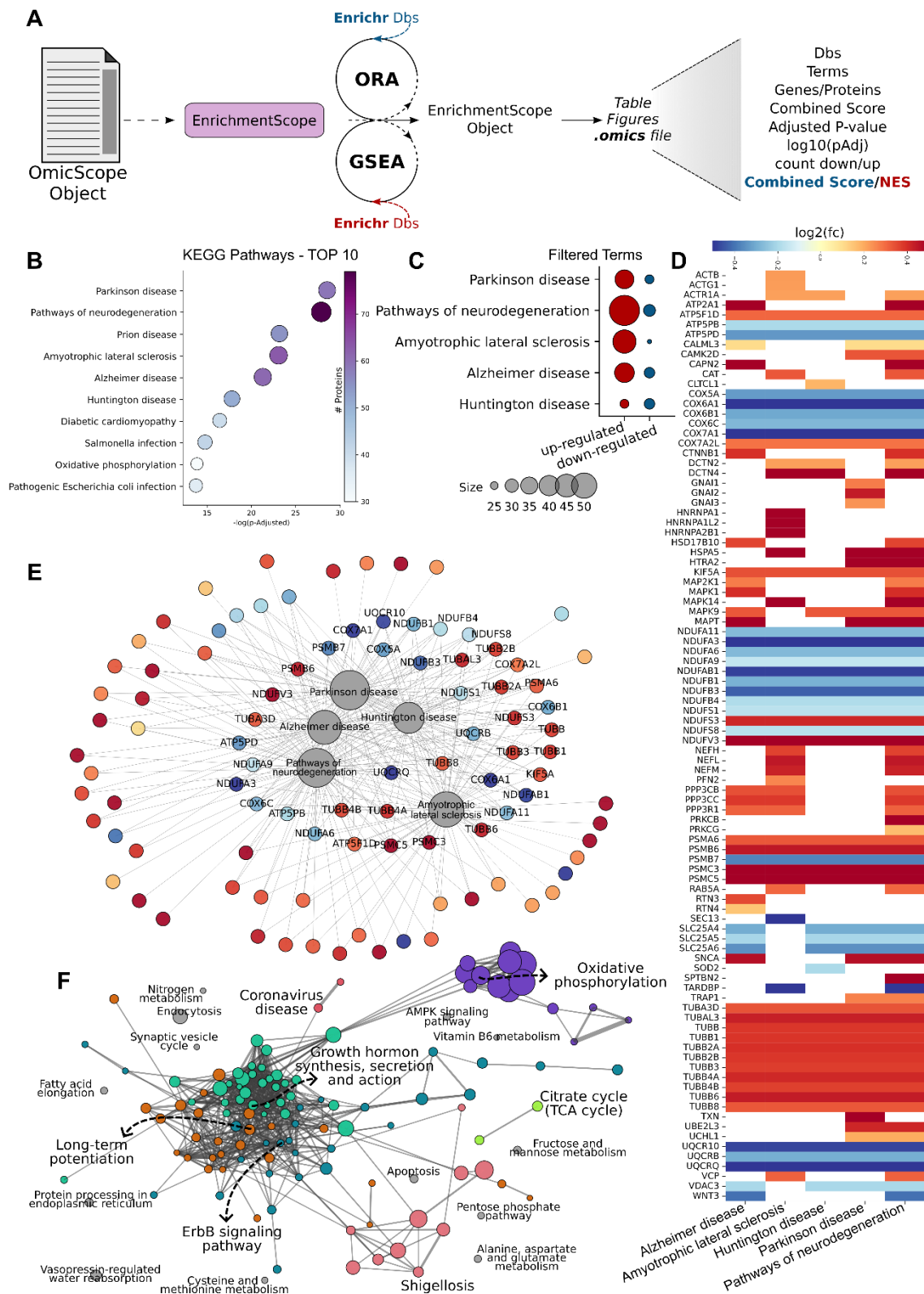
Figure 3-3 EnrichmentScope employs a systems biology approach for enrichment analysis based on data provided by OmicScope. **A)** EnrichmentScope performs Over-represented analysis (ORA) or Gene Set Enrichment Analysis (GSEA) using Enrichr libraries. **B-F)** Depiction of figures generated using the EnrichmentScope module: **B)** Dot plot illustrating the top 10 enriched terms in the analysis. **C)** Dot plot

showcasing the count of up- and down- regulated proteins in terms related to neurodegenerative diseases. **D)** Heatmap of differentially regulated proteins associated with terms from C. **E)** Network connecting enriched terms with their respective proteins, colored based on foldchange. The labeled proteins are shared among all processes. **F)** Enrichment map displaying all enriched pathways, colored by modules defined using the Louvain algorithm. Term labels were determined based on intra-module connectivity and p-value.

3.4.4   *Nebula: from singular studies to meta-analysis*

The advent of Omics platforms has exponentially increased the accumulation of data over the years, driving scientists to develop tools capable of comparing independent studies or even integrating experiments in a multi-omics fashion. Therefore, OmicScope introduces the Nebula module, designed to enhance data integration, interpretability, and comparison between studies.

The Nebula workflow utilizes the outputs of OmicScope/EnrichmentScope for data integration and visualization. These outputs have the extension ".omics" and can be generated by running the OmicScope module, which returns quantitative data, or the EnrichmentScope module, which provides both quantitative and enrichment results. For each independent analysis, one of these previously described modules must be executed, and Nebula will read each output file to compile them into a unified object. Once the files are imported into Nebula, a set of visualization functions becomes available for conducting studies comparisons at the protein and/or enrichment levels (Figure 4A, Supplementary figure 5).

To demonstrate Nebula's capabilities, we used data from Crunfli 2022, Nie 2021, and Wang 2021. These selected studies assessed the effects of SARS-CoV-2 on patients' tissues, with Crunfli examining proteomic signatures in the brain, Wang evaluating proteomics and transcriptomics effects in the lungs, and Nie reporting the liver as the most affected organ in proteomics terms. In Nie's and Wang's studies, the authors just provided differentially regulated proteins and genes, enabling the application of the Snapshot method for over-representation analysis (Figure 4B).
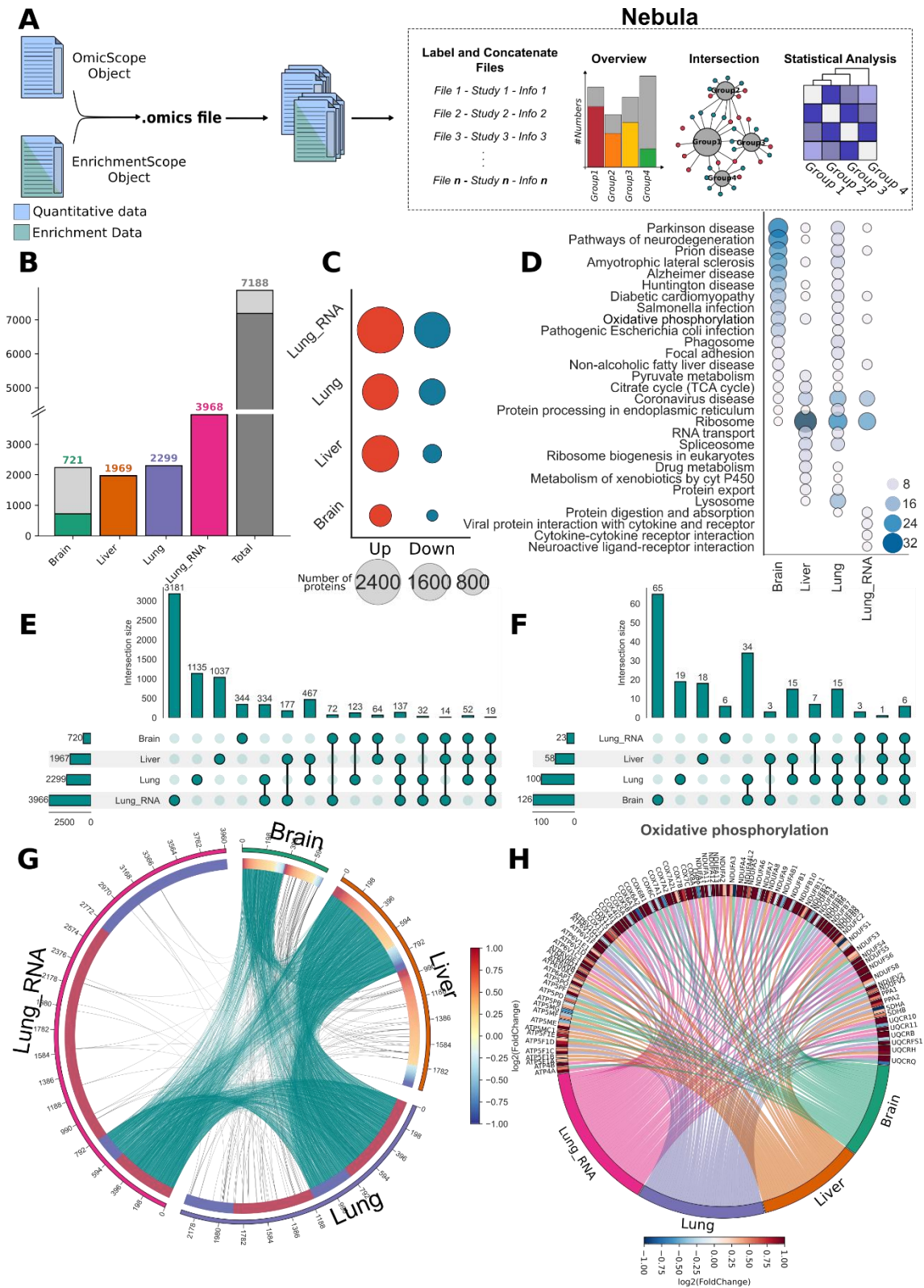
Figure 3-4 Nebula, the multi-omics integrative module, compares independent studies utilizing data outputs from OmicScope and EnrichmentScope. **A)** Nebula facilitates comparative analysis of independent studies based on OmicScope and EnrichmentScope outputs. **B-H)** Figures generated using Nebula: **B)** Bar plot depicting the count of whole (gray) and differentially regulated proteins/genes (colored) across various studies, as well as the combined count. **C)** Dot plot showing the count of up-regulated and down-

regulated entities. **D)** Top 10 enriched pathways according to the KEGG database for all organs. **E-F)** Upset plots for **E)** proteins and **F)** enrichment terms, illustrating overlapping sizes among conditions. **G)** Circular plot displaying all differentially regulated proteins and their shared relationships among evaluated groups (cyan links), along with shared enrichment terms among groups (black links). Each protein is annotated with its respective foldchange. **H)** Circular plot depicting proteins differentially regulated in Oxidative phosphorylation among studies, with accompanying foldchange values.

Nebula's pipeline supports various plots that facilitate the simultaneous comparison of all target groups. Bar plots and dot plots offer an initial overview of the groups by comparing the number of proteins and pathways evaluated in each condition, serving as initial steps in establishing associations between studies (Figures 4B-D). In the selected datasets, the lungs exhibited the highest number of differentially regulated proteins and genes, followed by the liver and brain (Figure 4B). Utilizing the Nebula integrative analysis approach, we noteworthy all examined tissues presents a consistent elevated number of up-regulated entities when compared to down-regulated counterparts (Figure 4C). When filtering enrichment terms to highlight the top 10 pathways identified in each condition, Nebula can pinpoint several potential pathways worthy of further investigation (Figure 4D).

To delve deeper into comparisons, Nebula offers tools for examining overlaps at both protein and enrichment levels. While Venn diagrams are commonly used for visualizing overlaps, they have limitations when comparing more than four conditions, producing illegible plots (Supplementary Figure 6). To overcome this limitation, Nebula includes circus plots and upset plots in its pipeline (Figure 4E-F). In the Upset plot[201], each condition is depicted in a row, while columns illustrate non-zero intersections exclusively among the labeled groups specified in the frame (Figure 4E-F). The advantage of the Upset plot lies in its readability and the absence of limitations regarding the number of groups analyzed. In the example datasets, only 19 proteins and genes exhibited dysregulation in all tissues, whereas the largest overlap encompassed 467 differentially regulated proteins between the lung and liver proteomes (Figure 4E). On the other hand, when examining overlaps in enrichment terms, the highest overlap was found between brain and lung proteomes, with 34 terms exclusively shared between these two tissues.

---

[201] Alexander Lex et al., "UpSet: Visualization of Intersecting Sets," *IEEE Transactions on Visualization and Computer Graphics* 20, no. 12 (December 2014): 1983–92, https://doi.org/10.1109/TVCG.2014.2346248.

In addition to the Upset plot, Nebula can also perform comparisons across groups using circular plots. In this plot, Nebula links each group with lines, with each link representing a protein that overlaps between those conditions. Each protein also displays its respective fold change in the respective study, generating a circular heatmap. This circular plot complements the Upset plot by providing a view of the proportion of up- and downregulated proteins shared among groups. As expected, in the studies under evaluation, the major shared proteins were up-regulated (Figure 4-G).

Nebula also offers a three-dimensional interpretation of data, considering groups, proteins, and enrichment terms simultaneously. In the "circular_term" function, the user specifies an enrichment term to be searched in all datasets, followed by the filtering of proteins associated with those terms in each study. Nebula then generates a circular plot that connects study and proteins, color-coding proteins based on their respective fold changes per group. In the example datasets, "oxidative phosphorylation," enriched in all studies, was chosen to demonstrate that major proteins in this pathway were indeed up-regulated in all organs (Figure 4H).

Nebula's array of visual representations also comprises network and statistical analyses. Similar to the methodology employed in EnrichmentScope, Nebula generates a graphical representation that establishes connections between groups and their corresponding differentially regulated proteins, which also can be exported to third-party software tools (Figure 5A).

Two other systems biology strategies employed by Nebula to assess the similarity between studies in a pairwise fashion are similarity analysis and Fisher's exact test. In pairwise similarity analysis, Nebula computes similarity indices using the Jaccard algorithm by default across the target studies[202]. Nebula is also capable of using different correlational metrics, like Pearson, Euclidean, and others, to calculate the similarity index using protein fold change. On the other hand, in pairwise Fisher's exact tests Nebula compares the overlap between studies by considering the entire set of imported proteins as the background, which results in pairwise p-values. Similar to conventional enrichment analysis, users can optionally specify alternative background sizes, such as the number of reviewed proteins in a specific organism according to the

---

[202] "Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation | PLOS ONE."

Uniprot database. The results from both the similarity and Fisher's analysis can be visualized using heatmaps and graphs. In the network representation, each node represents a group, while links are depicted as either similarity indices or p-values, filtered based on pre-defined thresholds (Figure 5 B-C).

In the example discussed here, differentially regulated proteins from the four groups were compared using the Fisher's Exact Test approach, utilizing the reviewed human proteome database as the background (proteome size: 20,423 proteins). The heatmap showcases all pairwise p-values generated in this analysis (Figure 5B), whereas the network representation filters p-values below 0.05 and connects each group accordingly, using log10(p-value) to link groups (Figure 5C). This analysis effectively illustrates the relationship between the Liver and Lung proteomes, as previously observed in the Upset plot, while also highlighting associations between the other proteome datasets.
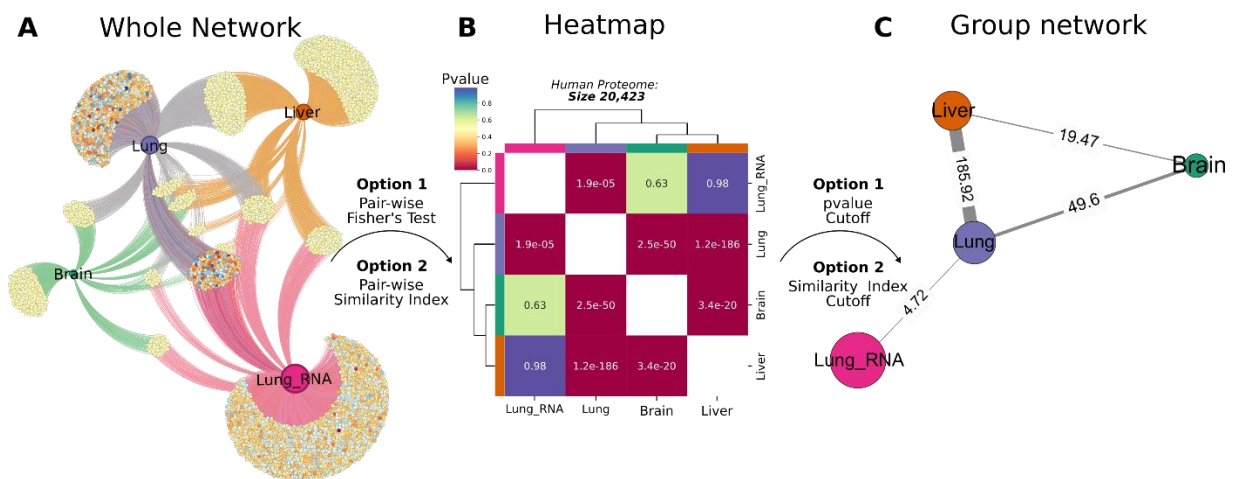
Figure 3-5 Systems Biology Approach with Nebula. **A)** Nebula employs a systems biology approach, presenting proteins differentially regulated for each study as networks, enabling a detailed exploration of shared proteins among groups. **B)** Differentially regulated proteins can be used to compare the independent studies in a pairwise manner using Fisher's Test or similarity indices, with a heatmap displaying each pairwise result. Users have the option to define background size or utilize the sum of all identified genes in among conditions as background for Fisher's Test. **C)** Based on results from similarity indices or Fisher's Test, users can generate a network, linking studies that share similar under specified thresholds.

3.4.5   User Interface and output flexibility

OmicScope is designed to cater to the diverse needs of the proteomics community, offering both code-based and user-friendly options for data analysis. The code is freely available on the PyPI repository under the MIT License, allowing experienced programmers and bioinformaticians to contribute and enhance its

capabilities. For non-programmer users, OmicScope is accessible as a Streamlit app (https://omicscope.ib.unicamp.br), providing a user-friendly interface with a high level of interactivity, making it easy for users to interact with the generated plots.

Furthermore, all OmicScope platforms prioritize the reporting of proteomics results to the scientific community. Figures can be exported in a vectorized manner, data used for all plots can be easily obtained, and networks can be exported to third-party software using the universal graphml file format.

3.5 DISCUSSION

Alongside the maturation of the field of MS-based Proteomics the field has witnessed the evolution of mass spectrometers, leading to enhanced resolution, sensitivity, and consequently, the capability to identify thousands of proteins. In response to these advancements, the field of computational analysis has progressed in parallel with technological advancements, offering robust methods for protein identification and quantification. The ability to assess thousands of proteins in a single experiment generates extensive datasets that necessitate a systems-level approach for meaningful interpretation. In this context, OmicScope was not designed to replace existing specialized tools but to function as a versatile pipeline capable of harnessing the powerful features of various analysis platforms. Its primary objective is to facilitate effective, integrative, and snapshot downstream analysis of proteomics datasets.

Here, we evaluated nine pipelines commonly used for downstream proteomics analysis (Supplementary Table 1). Our findings underscore the need for more integrative tools equipped to handle the diversity and complexity of proteomics data. While PatternLabV [203] and Progenesis QI for Proteomics were included in our comparison due to their capacity for differential expression analysis and related visualizations, their primary function revolves around protein identification and quantitation from raw MS files. On the other hand, MSstats[204] stands out by including a wide array of proteomics search engines, encompassing nine distinct tools that can be seamlessly integrated into its pipeline. Another noteworthy candidate,

---

[203] Santos et al., "Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V."

[204] "MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale | Journal of Proteome Research."

AlphaPeptStats[205], offers compatibility with inputs from four different tools and even accommodates the generic mzTab format, ensuring versatility for users. Similarly, OmicScope also provides four import methods based on quantitative proteomics engines. However, OmicScope distinguishes itself by offering two generic methods, the General and Snapshot approaches. This flexibility allows integration and analysis of data from other omics platforms, including genomics and transcriptomics, and facilitates the straightforward incorporation of previously reported data, which is often presented in a more concise format.

OmicScope offers distinct advantages compared to the other software tools mentioned in this study. While DEP, MSPypeline, and MSstats primarily rely on R architecture and offer diverse linear models for differential proteomics analysis, and InfernoRDN, Pattern Lab V, Perseus, Progenesis, and AlphaPeptStat utilize t-tests and ANOVA for statistical analysis, OmicScope combines frequentist and probabilistic approaches to address both static and longitudinal experimental designs.

In static analyses, OmicScope employs t-tests or one-way ANOVA to effectively identify differentially regulated proteins among two or more conditions, ensuring robust results in various scenarios. On the other hand, while performing longitudinal analysis, OmicScope employs the Storey methodology, incorporating a generalized linear model and considering gene variability over time through natural cubic splines. This approach allows OmicScope to investigate within- and between-group variations comprehensively, providing a more nuanced understanding of the data. Additionally, OmicScope employs the Benjamini-Hochberg method to correct p-values, effectively controlling false discovery rates. This ensures a higher level of statistical rigor, minimizing the risk of false positives in the results. Overall, OmicScope's ability to adapt to both static and longitudinal designs while offering a powerful and accurate statistical framework makes it a versatile choice for differential proteomics analysis.

In our survey of existing tools, we observed that only four applications incorporate enrichment analysis within their workflows. Perseus, MsPypeline, and AlphaPeptStat support over-representation analysis, though Perseus requests the download of target databases for to perform ORA. In contrast, MSPypeline and

---

[205] Krismer et al., "AlphaPeptStats."

AlphaPeptStat are limited in their capacity to enrich data with a restricted number of databases. Moreover, the DEP package primarily focuses on Gene Set Enrichment Analysis, employing the Enrichr API to encompass an extensive array of databases for the enrichment analysis. In line with DEP, OmicScope harnesses the vast resource of 224 databases available in Enrichr. Notably, OmicScope goes a step further by implementing both Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) algorithms. Beyond the analysis itself, OmicScope offers a rich set of innovative visualization features, including network analysis, heatmaps, and dot plots. These visualization tools serve a dual purpose: they enhance the presentation of enrichment analysis statistics and incorporate protein fold-change data, facilitating a deeper integration between proteomics and enrichment results, providing a broader overview of the entire landscape of changes within the system.

In addition to the conventional proteomics workflow, the integration of results with independent and orthogonal datasets has emerged as a valuable strategy to mitigate false discovery rates[206]. While the Perseus workflow stands out for its capacity to facilitate multi-study and multi-omics comparisons, other evaluated tools often require the use of third-party software and packages to perform additional analyses. OmicScope introduces the Nebula module, specifically designed for multi-study comparisons. Nebula empowers users with innovative visualization strategies that facilitate both an overall view and pairwise comparisons of target studies. While Venn diagrams are conventionally employed to distinguish sets of proteins, they are constrained by the limited number of sets that can be simultaneously analyzed[207]. Nebula effectively addresses this limitation by introducing Upset plots and circular plots that provide insights into overlapping protein sets and the corresponding fold changes observed in each study. Additionally, the Nebula workflow extends to network analysis for group comparisons, featuring Fisher's exact test and similarity indices that highlight relationships between target studies.

To cater to the diverse community of proteomics researchers, several software tools incorporate a Graphical User Interface (GUI) alongside their primary packages.

---

[206] Tripathi et al., "Meta- and Orthogonal Integration of Influenza 'OMICs' Data Defines a Role for UBR4 in Virus Budding."

[207] Lex et al., "UpSet."

DEP, MSPypeline, and AlphaPeptStats offer both command-line and GUI, ensuring accessibility for users across a range of technical backgrounds. Similarly, OmicScope provides both a Python package and a web application (https://omicscope.ib.unicamp.br) for running OmicScope, EnrichmentScope, and Nebula modules. While the web application delivers the interactivity necessary for data analysis and interpretation, the Python package enhances flexibility by facilitating integration with other pipelines and allowing developers to contribute updates to input formats and algorithms. Furthermore, both the web application and Python package support data export in various formats, including tables containing all information required for generating figures and graphml files that enable network visualization in third-party software, such as Cytoscape[208].

In conclusion, OmicScope has been meticulously crafted to empower experimentalists for the comprehensive analysis and interpretation of proteomics datasets. Our tool guides proteomics researchers in intricate biological questions, unraveling drug mechanisms of action, and elucidating molecular pathways of disorders. OmicScope stands as an open-access resource, readily available at https://omicscope.ib.unicamp.br and can be conveniently downloaded from the PyPI repository.

3.6 APPENDIX

3.6.1 Data organization and input methods.

To ensure versatile compatibility with various data formats, OmicScope adopts the data organization approach outlined by Morgan et al. in 2023[209], which divides data into three primary components: assay, pdata (metadata and phenotype data), and rdata (comprising information regarding proteins), as visually depicted in Supplementary Figure 7.

The assay represents the abundance matrix A, with dimensions i x j, where i denotes the number of proteins, and j represents the number of samples. Meanwhile, the rdata is a compound matrix R, with dimensions i x r, where r indicates the number

---

[208] Shannon et al., "Cytoscape."
[209] Martin Morgan et al., "SummarizedExperiment: SummarizedExperiment Container" (Bioconductor version: Release (3.17), 2023), https://doi.org/10.18129/B9.bioc.SummarizedExperiment.

of compound features, such as Accession, gene name, and p-value. Pdata is the sample matrix P, with dimensions j x p, where p refers to the number of columns describing samples.

In order to ensure seamless integration within OmicScope, data terminologies have been standardized among the import methods. Additionally, each import method incorporates suggestions previously reported by the respective authors in their publications[210].

- **Progenesis Qi for proteomics:** Progenesis exports a .csv file containing all the necessary information for the OmicScope pipeline. The assay data corresponds to normalized abundance levels, while rdata includes all available information about proteins. Pdata is extracted from columns located below the "Normalized Abundance" label. Due to the simplicity of data exported by Progenesis, OmicScope also accommodates Excel files (with extensions .xls or .xlsx) containing unique sheets.

- **MaxQuant**: MaxQuant exports the "proteinGroups.txt" file, which provides a comprehensive description of the assay and rdata. While importing these data, OmicScope filters out reverse proteins and contaminants, and selects the abundance based on the 'LFQ intensity' columns. As pdata is missing in the "proteinGroups" file, OmicScope necessitates additional pdata to define biological conditions and execute the statistical workflow.

- **DIA-NN**: DIA-NN exports the "main output," which contains a comprehensive description of the assay and rdata. During the import process, OmicScope filters reverse proteins and contaminants and selects the abundance based on 'MaxLFQ.' Similar to MaxQuant, additional pdata is required to define biological conditions and perform the statistical workflow.

- **PatternLab V**: PatternLab exports an Excel file that contains assay, pdata, and rdata. The assay is extracted from the "Proteins" sheet, which includes XIC-based protein quantitation. OmicScope normalizes protein abundance based on information provided in the PatternLab output and filters out identified reverses

---

[210] Cox and Mann, "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification"; Tyanova and Cox, "Perseus"; Demichev et al., "DIA-NN"; Santos et al., "Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V."

and contaminants. Finally, rdata consists of additional information about proteins presented in the "Proteins" sheet, and pdata is constructed based on the "Class Description" sheet.

- **Snapshot**: The Snapshot method is a simplified version of other methods and comprises an Excel spreadsheet containing information about the studied conditions and four additional columns (accession, gene name, log2-transformed fold-change, and p-value), as shown in Supplementary Figure 8. Optionally, users can also add a "TotalMean" column containing the abundance mean for each protein.

- **General**: General is an OmicScope method that enables the analysis of data generated from other sources. Users are required to construct an Excel file containing three sheets: assay, rdata, and pdata (Supplementary Figure 9).

1. *Assay*: This represents the abundance matrix. The assay columns must be named according to the samples described in pdata, and the number of rows must match the number in rdata.

2. *Rdata*: Rdata contains information about proteins/genes. Users must ensure the existence of two columns: "Accession" and "Description." "Accession" contains the protein identifier, while "Description" contains protein FASTA header. Optionally, users can add other protein features, including differential proteomics results, which must be labeled naming columns with "pvalue" or "pAdjusted". The number of rows in rdata must match the rows in the assay.

3. *Pdata*: Pdata contains information about the samples evaluated in the study. Users must ensure the existence of three columns: "Sample", "Condition", and "Biological". The "Sample" column should contain names that match with assay data, "Condition" specifies the conditions to be compared (e.g. "Control" and "Treatment"), and "Biological" refers to the number labeling biological replicates among samples.

For longitudinal experimental design, additional columns can be included to facilitate statistical analysis (Supplementary Figure 10). It is mandatory to insert the "TimeCourse" column to annotate the respective sample time-point. In cases where related sampling is performed, such as when the same

individual is sampled over time, an "Individual" column must be added to
pdata. Noteworthy, in longitudinal experiments, "Biological" considers an
individual at a specific time-point.

During data import, users can specify various parameters to fine-tune the
OmicScope functions, such as control group selection, experimental design, p-value and
fold-change cutoffs, log2-transformation, nominal or adjusted p-values to define
differentially regulated proteins, and degrees of freedom (exclusive for longitudinal
analysis). Furthermore, OmicScope allows users to filter out contaminants from the
analysis using the Frankenfield 2022 list of the most commonly found protein
contaminants[211].

3.6.2   Differential Proteomics Analysis

In the OmicScope differential proteomics workflow, users have the option to
import previous statistical results or execute the OmicScope statistical pipeline. When
importing previous results, OmicScope searches in rdata for columns that may represent
statistical analysis, such as "pvalue," "qvalue," and "p-Adjusted" columns. If OmicScope
identifies any of these terms, the algorithm utilizes the previous statistical analysis to
define result data, known as quantitative data, or "quant_data." Subsequently,
OmicScope determines differentially regulated proteins based on the filtering
parameters specified by users.

While performing statistical workflow, OmicScope considers the pdata matrix
and/or user-defined parameters to perform the appropriate experimental design.
Initially, the algorithm calculates the mean abundance level among biological replicates
for each protein. This is followed by a filtering stage where proteins are selected if they
are detected in at least one sample for each analyzed condition. Users also can disable
log2-transformation to perform statistical analysis.

OmicScope provides two workflows for statistical analysis: static and
longitudinal (Supplementary Figure 11). For static analysis, OmicScope assumes a
normal distribution between groups, homogeneity of variances, and group
independence. Based on these assumptions, OmicScope performs an independent T-test
or Analysis of Variance (ANOVA) for two or more groups, respectively. Alternatively,

---

based on user-input parameters, users can perform a paired t-test, assuming related observations, independence of differences, and a normal distribution. Additionally, for ANOVA analysis, post-hoc tests have been implemented, in which proteins with pAdjusted values less than a specified threshold undergo a Tukey-HSD test for pair-wise comparison between groups. It is important to note that due to the pair-wise comparison, the Tukey-HSD test may take some time to complete the analysis.

In the longitudinal workflow, OmicScope assesses whether protein abundance varies over time. To achieve this, the workflow adapts the method suggested by Storey in 2005[212], in which gene expression is modulated according to a natural cubic spline in a generalized linear model. The Storey method takes into account differential proteomics considering within- and between-group analysis, defining differentially related proteins based on expression over time or through a comparison between groups.

Once nominal p-values have been calculated for either longitudinal or static approaches, OmicScope performs a multiple hypothesis correction according to the Benjamini-Hochberg method[213]. OmicScope then calculates the fold change for each protein among groups, performs log2-transformation on fold change and p-value, and finally generates quantitative data results for plotting figures and performing enrichment analysis.

3.6.3   Figures toolset

OmicScope offers a comprehensive set of data visualization tools that have been specifically designed to emphasize key data results. Moreover, OmicScope, EnrichmentScope, and Nebula provide unique setups for figures that can be plotted, and all available figure options can be saved in scalable vector graphics (SVG) or PNG formats. For functions involving graphs, OmicScope exports a graphML file that can be imported into other specialized software, such as Cytoscape[214] or Gephi[215].

---

[212] Storey et al., "Significance Analysis of Time Course Microarray Experiments."
[213] Benjamini and Hochberg, "Controlling the False Discovery Rate."
[214] Shannon et al., "Cytoscape."
[215] Bastian, Heymann, and Jacomy, "Gephi."

3.6.3.1 OmicScope-class figures

Within the OmicScope class, figures can be categorized into three distinct types: overview, clustering, and protein-specific.

In the overview category, data can be visualized in terms of protein abundances, fold changes, and statistical significance. OmicScope offers volcano, MA, and dynamic range plots to visualize data normalization and distribution. The volcano plot presents both fold changes (x-axis) and statistical significance (y-axis) in log-scale. The MA plot displays protein fold change (y-axis) against its average (x-axis), and the dynamic range plot focuses on proteome coverage, showing ranked proteins (x-axis) and their respective abundance (y-axis). When conducting multiple group comparisons, the volcano plot and MA plot present only positive axes for fold changes and include a legend showing the respective comparisons, addressing the data's multidimensionality.

In the clustering category, OmicScope implements three clustering algorithms to demonstrate how proteins can be used to group conditions: hierarchical clustering, Principal Component Analysis (PCA), and K-means. Hierarchical clustering is employed alongside a heatmap and can be used to visualize pair-wise correlations between samples or protein regulation throughout the samples. PCA is used to illustrate how samples from different conditions can be grouped based on protein abundance. The K-means algorithm is also implemented to depict sample clustering, which is particularly useful for longitudinal statistical analysis. As the K-means algorithm requires a specific K-value for clustering analysis, OmicScope automatically applies the Kneedle algorithm to determine the optimal K-value[216]. Alternatively, users can pre-specify the best K-value based on data characteristics.

The last category of OmicScope plots is protein-specific. This category allows users to evaluate target proteins using boxplots, barplots, and protein-protein interaction (PPI) networks. Boxplots and barplots compare the abundance of target proteins across groups, while the PPI function uses the STRING API to retrieve known protein-protein interactions, including functional or physical interactions[217]. The

---

[216] Ville Satopaa et al., "Finding a 'Kneedle' in a Haystack: Detecting Knee Points in System Behavior," in *2011 31st International Conference on Distributed Computing Systems Workshops* (2011 31st International Conference on Distributed Computing Systems Workshops (ICDCS Workshops), Minneapolis, MN, USA: IEEE, 2011), 166–71, https://doi.org/10.1109/ICDCSW.2011.20.
[217] Szklarczyk et al., "The STRING Database in 2021."

PPInteractions function enables users to set the evidence score to consider protein-protein interactions (default to 0.6), search for communities based on the Louvain algorithm, and choose between physical or functional interactions. Notably, the STRING API can search up to 2000 proteins to retrieve PPIs. In cases data presents more than 2000 differentially regulated proteins, OmicScope will filter the top proteins based on p-values.

3.6.3.2   EnrichmentScope figures

In the EnrichmentScope class, figures have been designed to emphasize enrichment results and include quantitative values reported by OmicScope. These plots can be categorized into three main types: dot plots, heatmaps, and graphs.

Dot plots serve two primary purposes. They are used to evaluate enrichment statistical results and depict overall protein deregulation. In the first case, the dot plot associates enriched terms (y-axis) with statistical significance (x-axis). In the latter case, for each enriched term (y-axis), EnrichmentScope splits and counts the number of up-regulated and down-regulated proteins and plots the data with dot size proportional to the number of proteins.

The heatmap category showcases proteins associated with each enriched term. During Over-Representation Analysis (ORA) or Gene Set Enrichment Analysis (GSEA)[218] users can generate heatmaps in which proteins are colored based on the enrichment-adjusted p-value or protein fold change. For GSEA, users also have the option to color the proteins based on the Normalized Enrichment Score (NES) associated with the respective term.

For graph visualization, EnrichmentScope provides two distinct functions: enrichment network and enrichment map. In the enrichment network, enriched terms are linked to the proteins, making it easier to visualize proteins associated with multiple terms. On the other hand, the enrichment map is implemented like the approach proposed by Merico in 2010[219], in which terms are connected and weighted according to the Jaccard similarity index. This index calculates the similarity between two enriched terms as the ratio between overlap and the union among both datasets. Since

---

[218] Subramanian et al., "Gene Set Enrichment Analysis," October 25, 2005.
[219] "Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation | PLOS ONE."

the overlap of genes in the pathways evaluated increases as the index increases, EnrichmentScope considers links when the similarity index is higher than 0.25 (by default). Additionally, EnrichmentScope performs community detection within the enrichment map using the Louvain algorithm to define modules and label the central node in each community. This labeling is done by selecting the node with the highest degree within the target community. In cases where more than one node shares the maximum degree, the node with the highest adjusted p-value is selected.

3.6.3.3 Nebula figures

Nebula figures have been implemented to facilitate the visual comparison of independent studies or groups. These figures include bar plots, dot plots, upset plots, circular plots, and graphs (Supplementary Figure 5). Some of these plots also perform statistical analysis to identify similarities between groups.

To initially assess the differences between groups, Nebula offers bar plots and dot plots for protein- and enrichment-level data. In the protein approach, the bar plot displays the number of entities quantified and differentially regulated across all studies, while the dot plot counts the number of up- and down-regulated entities, providing insights into the differential regulation of data. On the enrichment level, Nebula offers a dot plot that sorts enriched terms according to adjusted p-values in each study, followed by filtering stage based on the N top terms. All filtered terms are then combined into a unified list, which serves as a template for filtering terms across all studies, offering information about terms that are highly relevant to one group and potentially relevant to others.

Additionally, upset plot has been implemented to visualize intersections between groups[220]. It takes into consideration entities that are differentially regulated and enriched terms. The upset plot displays the intersection of groups in a dot frame, highlighting the studies being compared. On the left/right side of the dot frame, a bar plot indicates the number of proteins related to each study. Above the dot frame, a second bar plot shows the number of proteins that are shared only in the studies highlighted in the dot frame.

---

[220] Lex et al., "UpSet."

To consolidate features derived from differential proteomics and enrichment analysis, Nebula includes two circular plots inspired by Circos [221] and circlize [222] approaches. In the first circular plot, differentially regulated proteins in each group are displayed alongside a heatmap, showing the respective protein fold changes. Proteins that are shared between studies connect the respective studies, providing a visual representation of the regulation of overlapping proteins. Users can also add enrichment links to the plot, suggesting an amount of shared enrichment terms among groups. In the circlize approach, users select a term to filter all enriched terms that contain the searched word. Nebula retrieves all proteins associated with those terms and links them with the respective groups.

The last set of figures in Nebula includes network analysis to compare the independent studies. Firstly, users can plot all groups linked to their respective proteins, providing a similar figure proposed by EnrichmentScope. Alternatively, users can also perform more quantitative comparison using two approaches: similarity analysis and Fisher's exact test.

The similarity analysis calculates a similarity index between groups. By default, Nebula employs a pairwise Jaccard similarity algorithm to provide the similarity index. Users can optionally apply other similarity algorithms, such as Pearson's and Euclidean measures, taking into consideration the fold change of proteins to provide the similarity index.

On the other hand, Fisher's exact test is offered as an "Enrichment-like" approach. Nebula assesses the chance of pair-wise overlap occurring randomly compared to the whole set of proteins imported as background, considering all files. Optionally, users can specify the background size to be compared, such as the entire Human proteome, which comprises over 20,000 reviewed and annotated proteins in Uniprot database.

After performing similarity or Fisher's exact test, the data is available in a visual format as a heatmap and can also be represented as a network, in which nodes represent groups and links display the similarity index or log10(p-value).

---

[221] Martin Krzywinski et al., "Circos: An Information Aesthetic for Comparative Genomics," *Genome Research* 19, no. 9 (January 9, 2009): 1639–45, https://doi.org/10.1101/gr.092759.109.

[222] Zuguang Gu et al., "Circlize Implements and Enhances Circular Visualization in R," *Bioinformatics* 30, no. 19 (October 1, 2014): 2811–12, https://doi.org/10.1093/bioinformatics/btu393.

3.6.4   OmicScope App

The OmicScope App is a user-friendly interface developed using the Streamlit framework. The application consists of three main pages: Home, OmicScope, and Nebula. Each page serves a specific purpose and provides an easy-to-navigate environment for users to interact with the OmicScope platform. Below is a detailed description of each page within the OmicScope App:

Home Page (Supplementary Figure 12): The Home page serves as an introduction to the OmicScope platform. It provides a brief description of the OmicScope architecture and functionalities, along with figures that are also presented in this manuscript. Users can get an overview of the capabilities of OmicScope by exploring this page.

OmicScope Page (Supplementary Figure 13): The OmicScope page is where users can access the core OmicScope and EnrichmentScope modules. This page is divided into different sections and present several features:

1. ***Sidebar***: The sidebar is where users can upload their quantitative data file, select the appropriate import method, and define statistical parameters and filtering stages to run the OmicScope pipeline. Here are some key features in the sidebar:

    File Upload: upload quantitative data files.

    Input Method Selection: Users can choose the appropriate input method that matches the uploaded file. This step is crucial for accurate data processing.

    Additional Parameters: Users can modify various parameters related to data analysis, including defining a control group, customizing statistical parameters, and enabling protein-protein interaction (PPI) searching through the STRING API.

    Enrichment Analysis: Users can opt to run the EnrichmentScope module by selecting a checkbox. This allows for additional fine-tuning of enrichment parameters, such as selecting the target database, type of enrichment analysis, organism, or pAdjusted cutoff related to enrichment analysis.

2. **Main Page**: The main page displays tables and interactive figures generated based on the uploaded data and user-defined parameters. These figures can be easily exported as PNG or SVG files. Each figure also comes with its set of

adjustable parameters, allowing users to fine-tune various aspects of the visualization. Additionally, at the end of the page, users can download all the raw data used to generate the figures and GraphML file, which contains the information required to plot graphs in third-party software. This approach allows the integration of OmicScope data with other tools and workflows.

Nebula Page (Supplementary Figure 14): The Nebula page is dedicated to our integrative module, which allows users to compare multiple independent studies or groups. This page is organized as follows:

1. **Sidebar**: In the sidebar, users are required to upload a zip file containing all the omics files to be analyzed together. It's important to note that in OmicScope package, Nebula imports omics files from a conventional folder structure. After successful data import, Nebula reports the number of studies, groups, and the quantity of enrichment results imported.

2. **Main Page**: The main page displays the generated figures. While most figures are interactive, the circular plots are static images. To enhance interactivity, each figure comes with parameters that allow users to fine-tune specific features, such as colors and sizes. Finally, at the end of the page, users can find a downloadable button that provides access to files that can be used in third-party software and workflows, enabling seamless integration with other analysis tools.

The OmicScope App, with its user-friendly interface and easy access to OmicScope, EnrichmentScope, and Nebula modules, simplifies the process of importing data, performing analyses, and visualizing results. It offers a comprehensive set of features for bioinformatic analysis in a user-friendly and accessible manner.

3.7 SUPPLEMENTARY FIGURES
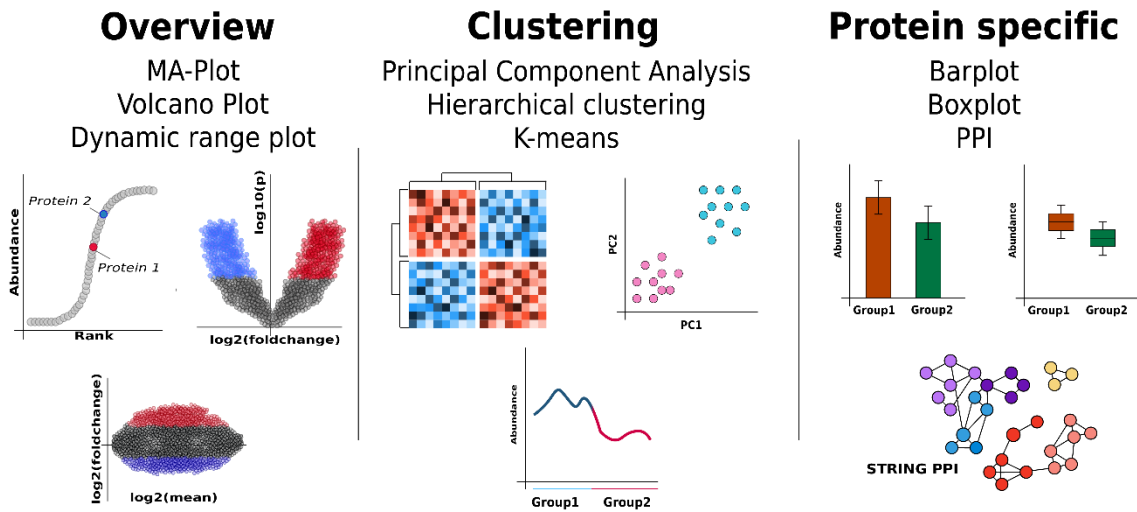
Supplementary Figure 3-1 OmicScope Figure Toolset. The OmicScope figure toolset comprises three subcategories of plots: overview, clustering, and protein-specific. Overview figures include volcano plots, MA plots, and dynamic range plots. The clustering category features Principal Component Analysis (PCA), hierarchical clustering, and K-means. The protein-specific set encompasses bar plots, boxplots, and protein-protein interaction networks.
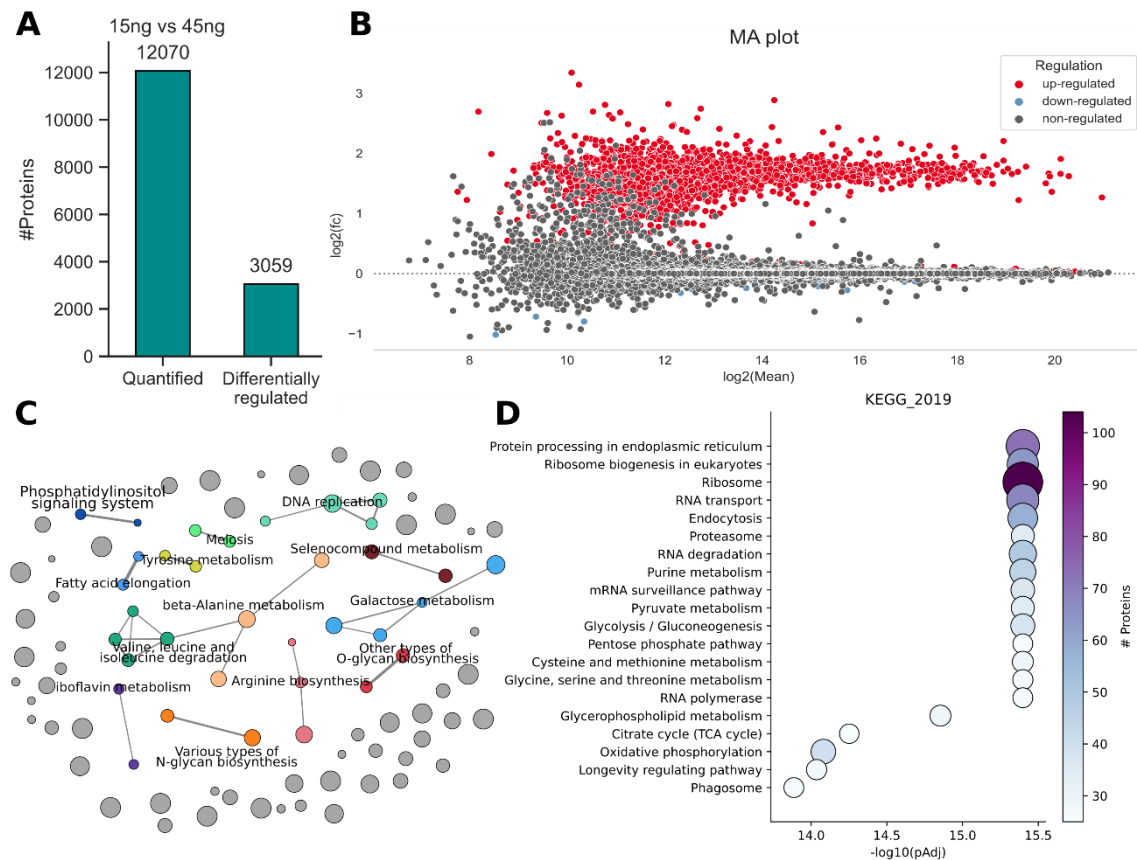
Supplementary Figure 3-2 Benchmark Dataset. This dataset comprises two concentrations of Yeast digest spiked into Hela digest and subsequently analyzed using DIA-NN and OmicScope. A) The study identified a total of 12,000 proteins, among which 3059 exhibit differential abundance between 45ng and 15ng Yeast digest concentrations. B) The MA plot illustrates the two distinct patterns of expression, wherein the red proteins indicate the up-regulation

of yeast proteins log2(45ng/15ng). C - D) Enrichment analysis was conducted utilizing the KEGG Yeast database. The results are presented using network and dot plot functions to visualize enriched pathways and their associated proteins.
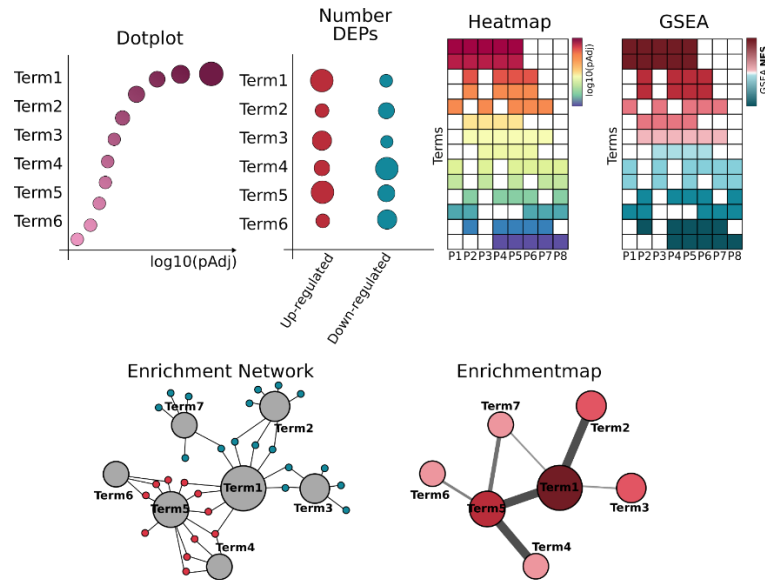
Supplementary Figure 3-3 EnrichmentScope Figure Toolset. The EnrichmentScope toolset offers dotplots, heatmaps, and graphs to visualize enrichment results and relationships between enrichment results and target proteins.
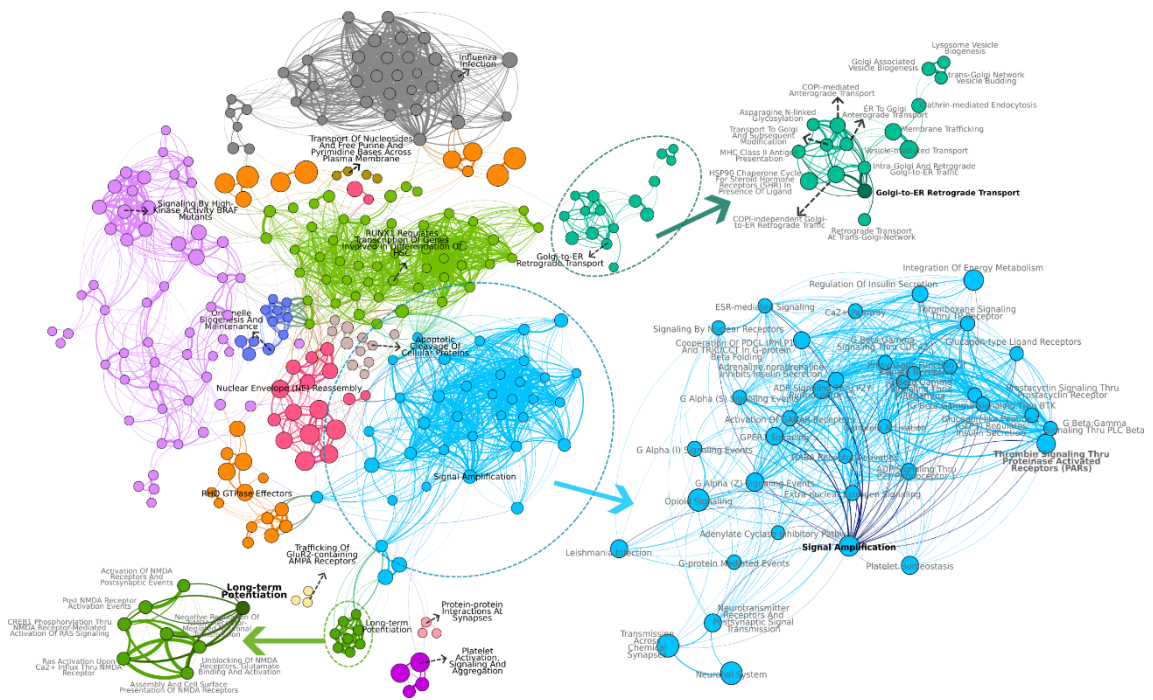
Supplementary Figure 3-4 Reactome Enrichment Map. Hierarchical databases, such as Reactome and Gene Ontology, often contain redundant terms, making data representation and interpretation complex during data analysis. EnrichmentScope performs modularity analysis to identify highly connected regions, followed by the selection of nodes that present higher degree and Adjusted P-value, respectively. These steps help reduce redundancy without omitting data.
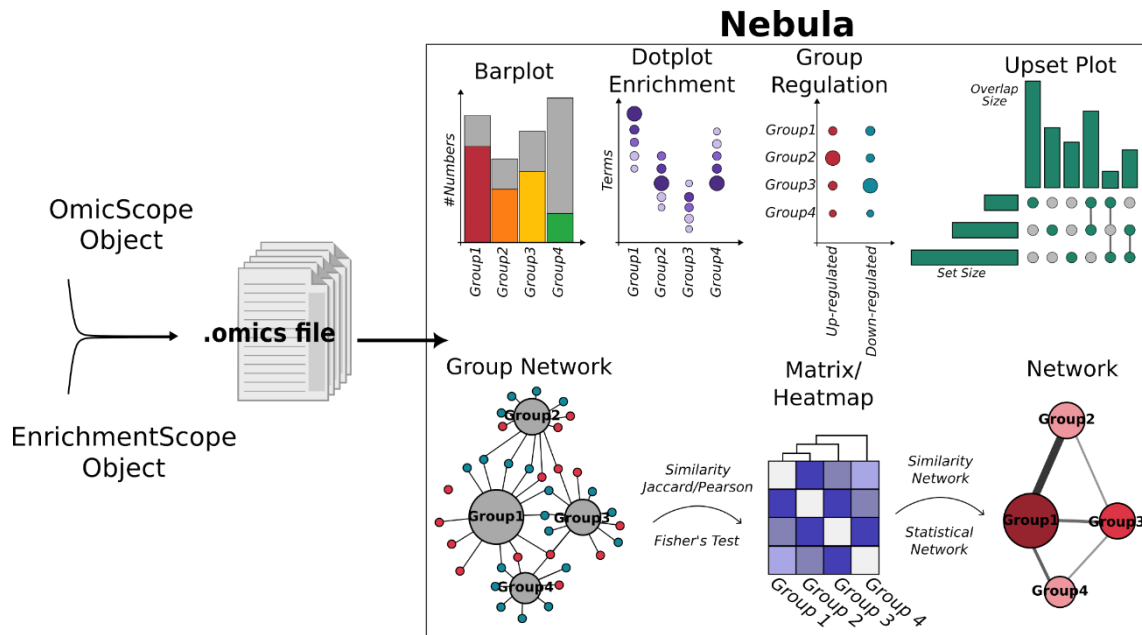
Supplementary Figure 3-5 Nebula Figure Toolset. The Nebula workflow allows the import of multiple omics files for simultaneous analysis. Once imported into Nebula, the algorithm facilitates the comparison of groups using bar plots, dot plots, heatmaps, circular plots, and graphs.
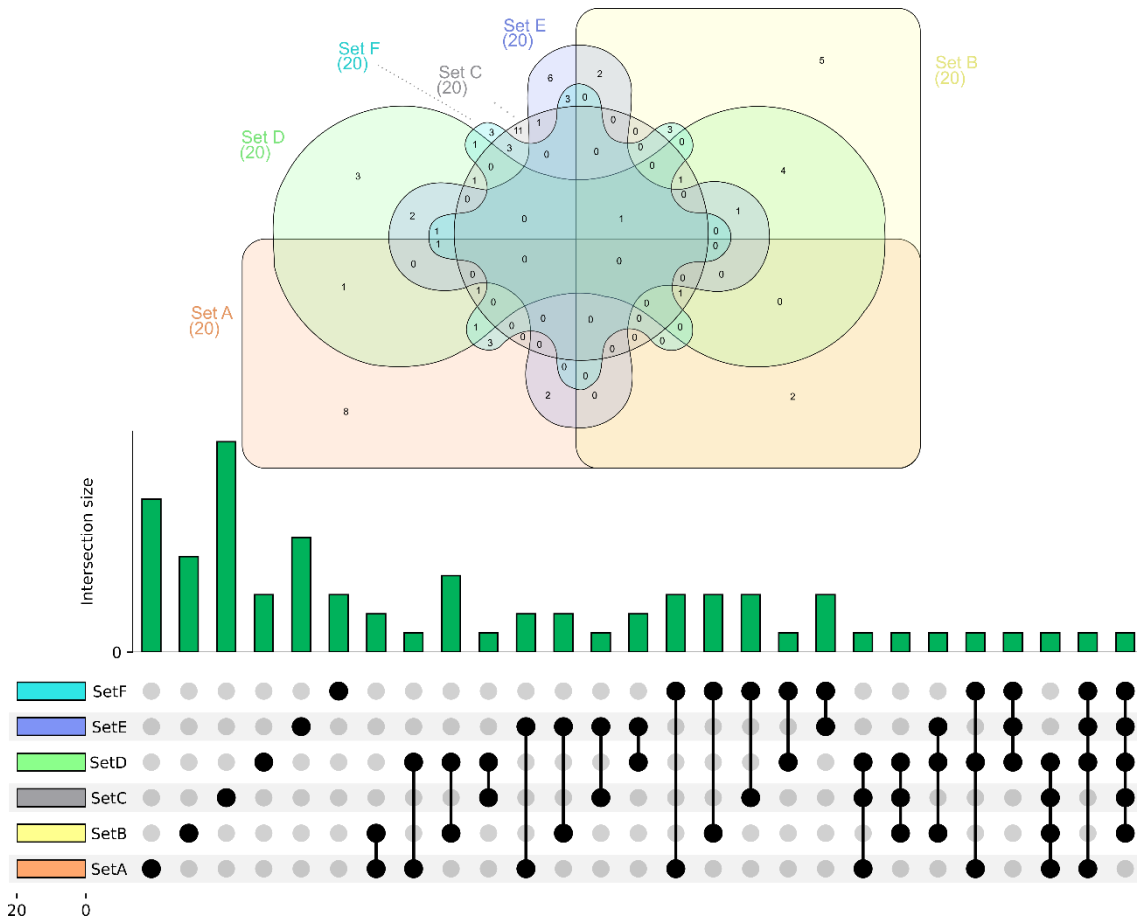
Supplementary Figure 3-6 Overlap Between 6 Sets Using Venn Diagram and Upset Plot. This figure displays the overlap between six sets using both Venn Diagram (top) and Upset Plot (bottom) representations.
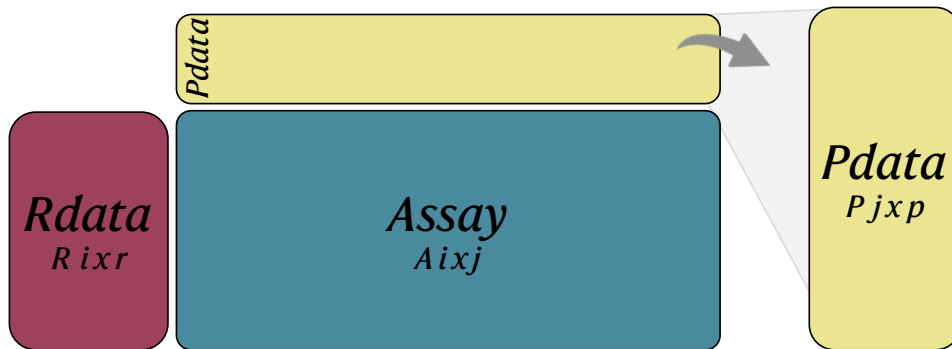
Supplementary Figure 3-7 OmicScope Structure. In OmicScope, data is organized into three categories: assay, rdata, and pdata. Assay corresponds to protein abundance, rdata contains protein features, and pdata includes phenotype data associated with each sample.

Supplementary Figure 3-8 "Snapshot" Method. This figure illustrates the data structure used to import data into OmicScope using the "Snapshot" method.

1. assay

	A	B	C	D	E	F
1	VCC_HB_1	VCC_HB_1	VCC_HB_2	VCC_HB_2	VCC_HB_3	VCC_HB_3
2	29388.5	31109.3	25218.1	30907	23835	22672.4
3	70813.1	64469.5	58254.9	59316.1	63091	59335.3
4	100754	101200	73013.3	73493.9	97668.4	99522
5	25880.3	37691	29926.9	34600.9	25963.2	25789.6
6	1019192	1109406	1060396	1078239	1003426	985367
7	565028	616371	570302	609880	655115	661696
8	1628191	1878949	1396965	1490581	1584860	1624611
9	45790.6	50332.6	53062.7	61519.8	43293.6	42383.5
10	1445203	1616672	1464048	1535869	1410360	1435973
11	971137	1131267	996292	1061151	830755	826155

3. pdata

	A	B	C
1	Sample	Condition	Biological
2	VCC_HB_1_1_2020	COVID	1
3	VCC_HB_1_2	COVID	1
4	VCC_HB_2_1	COVID	2
5	VCC_HB_2_1_2	COVID	2
6	VCC_HB_3_1	COVID	3
7	VCC_HB_3_1_2	COVID	3
8	VCC_HB_4_1	COVID	4
9	VCC_HB_4_1_2	COVID	4
10	VCC_HB_5_1	COVID	5

2. rdata

	Accession	Peptide co	Unique pe	Confidence	Anova (p)	q Value	Max fold c	Power	Highest m	Lowest me	Description
2	P0DJI8	1	1	6.8809	0	0	2.192654	1	COVID	CTRL	Serum amyloid A-1 protein OS=Homo sapiens OX=9606 GN=SAA1 PE=1 SV=1
3	P63313	2	0	24.1939	0	0	3.823799	1	COVID	CTRL	Thymosin beta-10 OS=Homo sapiens OX=9606 GN=TMSB10 PE=1 SV=2
4	P03886	3	0	24.0213	1.30E-07	4.07E-05	1.386199	0.999998	CTRL	COVID	NADH-ubiquinone oxidoreductase chain 1 OS=Homo sapiens OX=9606 GN=MT-ND1 PE=1 SV=1
5	Q9B5M1	2	2	12.267	5.52E-07	0.000105	1.726615	0.999984	COVID	CTRL	Polycomb group RING finger protein 1 OS=Homo sapiens OX=9606 GN=PCGF1 PE=1 SV=2
6	O94819	32	16	190.5708	5.58E-07	0.000105	1.245223	0.999984	CTRL	Kelch repeat and BTB domain-containing protein 11 OS=Homo sapiens OX=9606 GN=KBTBD11 PE=1 SV=1
7	Q14894	17	8	146.9671	7.83E-07	0.000111	1.45195	0.999974	COVID	CTRL	Ketimine reductase mu-crystallin OS=Homo sapiens OX=9606 GN=CRYM PE=1 SV=1
8	P10636	36	15	288.0723	8.29E-07	0.000111	1.416853	0.999972	COVID	CTRL	Microtubule-associated protein tau OS=Homo sapiens OX=9606 GN=MAPT PE=1 SV=5
9	P20962	1	0	6.3974	1.99E-06	0.000233	1.364264	0.999909	COVID	CTRL	Parathymosin OS=Homo sapiens OX=9606 GN=PTMS PE=1 SV=2

Supplementary Figure 3-9 Data Structure to Import Data into OmicScope Using the "General" Method. For the "General" method, OmicScope uses Excel spreadsheet, which should contain three sheets: assay, rdata, and pdata.

Minimum information — Time point — Related sampling

	A	B	C	D	E
1	Sample	Condition	Biological	TimeCourse	Individual
2	Sample1_Day1_Bio1_1	Control	1	1	1
3	Sample1_Day1_Bio1_2	Control	1	1	1
4	Sample2_Day1_Bio2_1	Control	2	1	2
5	Sample2_Day1_Bio2_2	Control	2	1	2
6	Sample3_Day1_Bio3_1	Control	3	1	3
7	Sample3_Day1_Bio3_2	Control	3	1	3
8	Sample4_Day2_Bio1_1	Control	4	3	1
9	Sample4_Day2_Bio1_2	Control	4	3	1
10	Sample5_Day2_Bio2_1	Control	5	3	2
11	Sample5_Day2_Bio2_2	Control	5	3	2
12	Sample6_Day2_Bio3_1	Control	6	3	3
13	Sample6_Day2_Bio3_2	Control	6	3	3
14	Sample7_Day3_Bio1_1	Control	7	5	1
15	Sample7_Day3_Bio1_2	Control	7	5	1
16	Sample8_Day3_Bio2_1	Control	8	5	2
17	Sample8_Day3_Bio2_2	Control	8	5	2
18	Sample9_Day3_Bio3_1	Control	9	5	3
19	Sample9_Day3_Bio3_2	Control	9	5	3
20	Sample10_Day4_Bio1_1	Control	10	7	1
21	Sample10_Day4_Bio1_2	Control	10	7	1

Supplementary Figure 3-10 Example of pdata structure. To import pdata into OmicScope, the Excel spreadsheet must include "Sample," "Condition," and "Biological" columns. When performing longitudinal analysis, pdata must also contain a "TimeCourse" column to specify the sample time points. Users can optionally add an "Individual" column to assign whether related data sampling was performed.
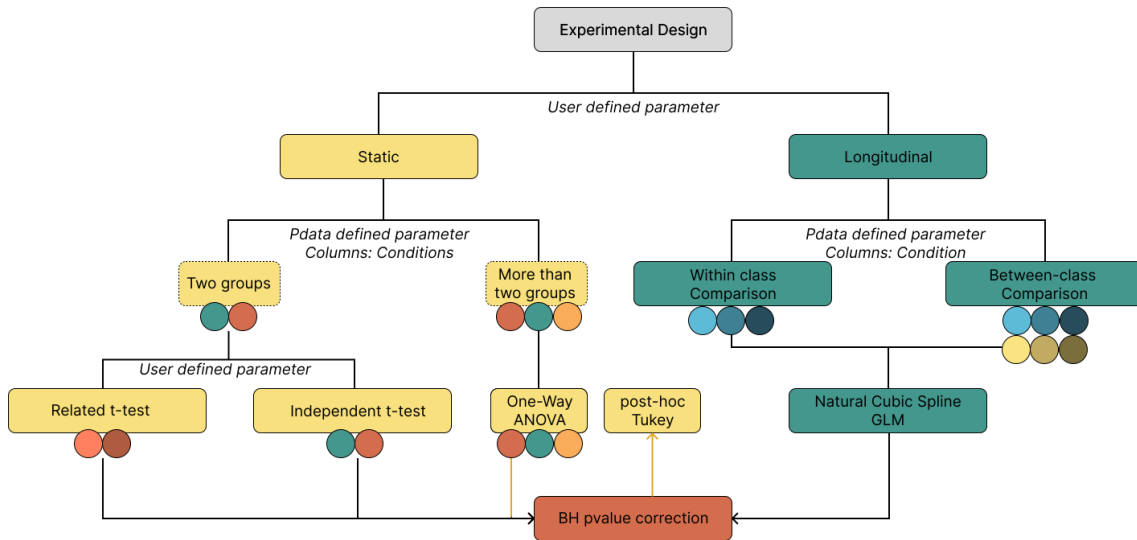
Supplementary Figure 3-11 OmicScope statistical pipeline. The OmicScope statistical workflow includes both static and longitudinal experiments. Depending on the number of groups analyzed, the algorithm selects appropriate statistical tests and multiple hypothesis corrections.

Supplementary Figure 3-12 OmicScope App Home Page. The home page provides a concise description of the OmicScope architecture and features.

Supplementary Figure 3-13 OmicScope and EnrichmentScope modules on the Web Application. The OmicScope and EnrichmentScope modules operate on the same page, with Enrichment being an optional analysis. Users can access global parameters and customize figures based on individual parameters. In the end of page, users can download all the files used to generate the figures.
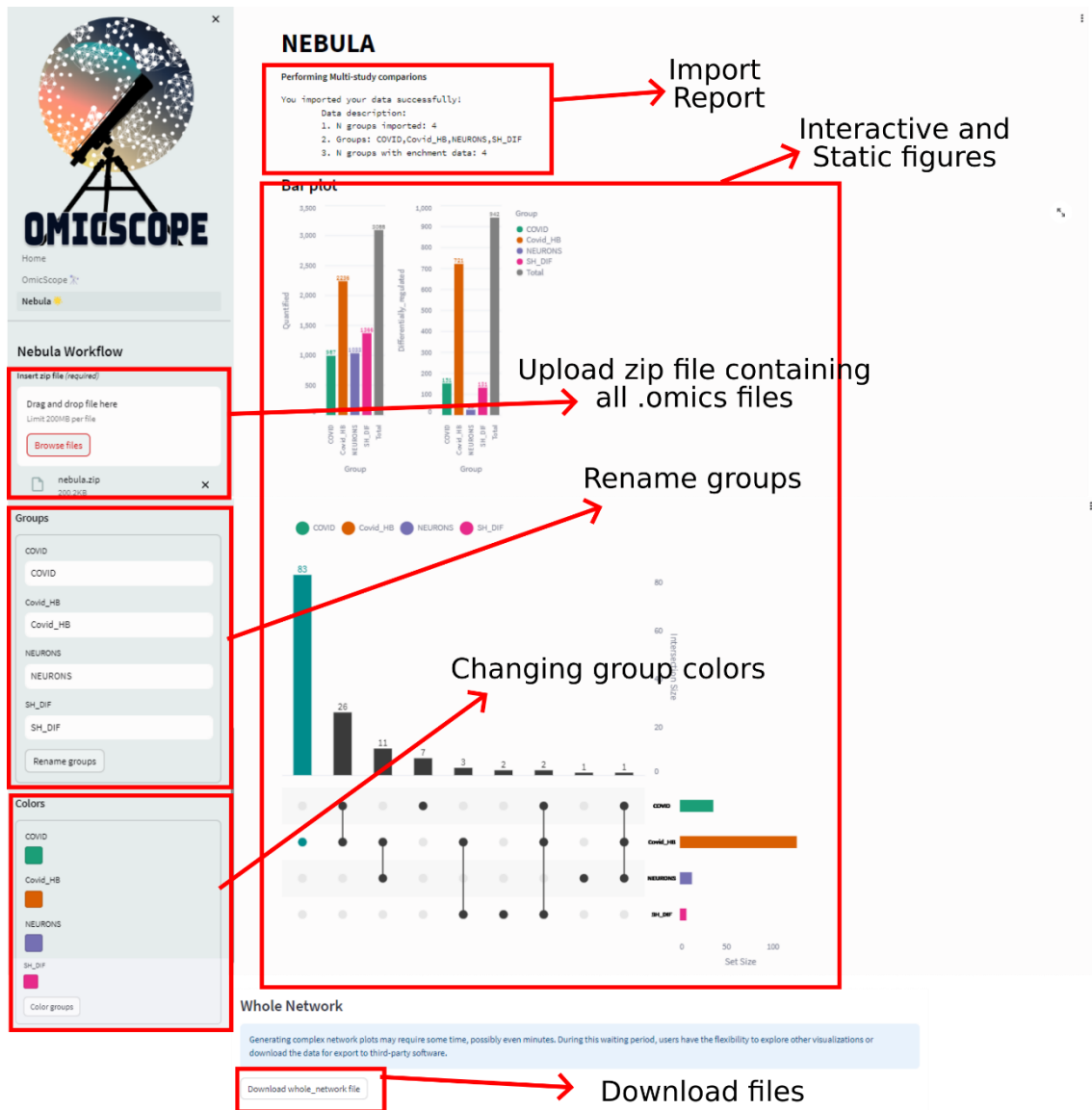
Supplementary Figure 3-14 Nebula module on the Web Application. To import data into Nebula, users are required to place all omics files into a zip file and then import it via the web application. Subsequently, several interactive figures are generated, and users can download raw data for use in further analyses.

Supplementary Table 3-1 Comparison between software

Software	User Interface	Input		Differential Expression					Enrichment Analysis			Meta analysis	Export	Source Code	Observations
				Static		Longitudinal			ORA	GSEA	Figures				
		ProteomicsPlatforms	Generic Formats	Two-Groups	Multiple-Groups	Within-Group	Between-Groups	Figures							
DEP	R package with Shiny App	MaxQuant, IsobarQuant		Use R package Limma to perform differential expression - empirical Bayes statistics				Barplots, volcano, heatmap, boxplots, density plot, cummulative sum plot, PCA, correlation, coverage, pvalue histogram,		Perform using EnrichR API	barplot			open source	
InfernoRDN (and DAnTE)	Desktop Application		.csv files	ANOVA				Histograms, Boxplot, Correlation Diagrams,MA Plot, PCA, PLS, Heatmaps, protein rollup						open source	InfernoRDN supersedes all previous DAnTE (Data Analysis Tool Extension), DanteR, and Inferno versions.
MSstats (also Included in Skyline)	R package	DIA-Umpire, MaxQuant, OpenMS, OpenSwath, Proteome discoverer, Skyline, Progenesis, DIA-NN, Spectronaut		linear mixed-effects model				Profile, QC, Condition Plots, Heatmap, Volcano, Comparison Plot						open source	
PatternLab V	Desktop Application	MS raw files		T-TEST	pair-wise t-test	Longitudinal analysis is performed using Kmeans algorithm in TrendQuest module		volcano (tfold), heatmap, venn Diagrams, histogram, k-means (TrendQuest), PCA (Buzios)					Export several file format for quantitative analysis. Exports png/svg format for plots and the raw data for each plot.	not public (open access)	PatternLab V presents algorithms for protein identification and quantitation and some downstream analysis.
Perseus	Desktop Application	MaxQuant	.txt file	T-TEST	ANOVA	Periodicity Analysis (cosine fitting)		User can dynamically change de parameters and axis of basic plots: 3D plot, Histogram, Multi scatter plot, Profile plot, Scatter plot	Fisher Exact text (user must download target databases)		User can dynamically change de parameters and axis of basic plots: 3D plot, Histogram, Multi scatter plot, Profile plot, Scatter plot	Coupled to R enviroment, Perseus perform DE analysis and can compare pair-wise omics experiments using '2D annotation enrichment'	Perseus export all matrix as text files and figures in several formats.	not public (open access)	Perseus is the most comprehensive desktop application for proteomcis experiments. However, the tool requires a lot of clicks to perform data handling. Due to its complexity, the developers suggests courses and book chapters to facilitate the learning process.
Progenesis Qi for Proteomics	Desktop Application	MS raw files		T-TEST	ANOVA	ANOVA		PCA, barplots, boxplots (for general analysis and protein abundance in each sample)				-	.csv files and png figures	not public (requires license)	Progenesis is a platform to perform protein quantitation. While coupled to PLGS (Waters), it also performs protein identification.
MSPypeline	Python package with GUI	MaxQuant		Use R package Limma to perform differential expression - empirical Bayes statistics				volcano, venn diagram, boxplots, heatmap, barplots, pca, histogram, Dynamic range	Enrichment Analysiss using a summarized version of GO, Hallmarks, and Biocarta		barplots and dotplots			open source	
AlphaPeptStat	Python package, web application, desktop application	DIA-NN, MaxQuant, Spectronaut,FragPipe	mzTab	T-TEST	ANOVA, ANCOVA			PCA, UMAP, tSNE, Volcano, Clustermap, dendogram	Enrichment Analysis with Gene Ontology				figures can be saved in svg and data can be manually copied and paste on excel	open source	
OmicScope (This paper)	Python package and web application	Progenesis, DIA-NN, PatternLabV, and MaxQuant	General, Snapshot (excel files)	T-TEST	ANOVA	GLM (natural cubic spline, Storey 2005)	GLM (natural cubic spline, Storey 2005)	volcano plot with density plot, heatmap, MA-plot, Dynamic Range plot, bar plot, box plot, PPI networks, K-means, and PCA	Perform using EnrichR API GSEAPY	Perform using EnrichR API via GSEAPY	dot plots, heatmaps, enrichment map, term-protein network	Import methods include generical methods that allow transcriptomics evaluation. Nebula module totally design to meta-analysis. Including specific plots and network analysis to explore multiple studies at once	graphml, dataframe	open source	

Supplementary Table 3-3 OmicScope input method features

	Feature	General	PatternLab	Progenesis	MaxQuant	DIA-NN	Snapshot
Input	Extension	.xlsx/.xls	.xlsx/.xls	.csv/.xls/.xlsx	.txt	.txt	.xlsx/.xls
	Require pdata	No	No	No	Yes	Yes	No
	Accept pdata	Yes	Yes	Yes	Yes	Yes	No
	Contaminant Filtering	Yes	Yes	Yes	Yes	Yes	No
	Reverse Filtering	No	Yes	No	Yes	Yes	No
	NA based filtering	Yes	Yes	Yes	Yes	Yes	No
Statistical Analysis	Import statistics	Optional	No	Yes	No	No	Yes
	Perform statistics	Optional	Yes	Optional	Yes	Yes	No
OmicScope	Identification	Yes	Yes	Yes	Yes	Yes	Yes
	Dynamic range	Yes	Yes	Yes	Yes	Yes	**No**
	Volcano	Yes	Yes	Yes	Yes	Yes	Yes
	MA plot	Yes	Yes	Yes	Yes	Yes	Yes
	Barplot	Yes	Yes	Yes	Yes	Yes	**No**
	Boxplot	Yes	Yes	Yes	Yes	Yes	**No**
	Heatmap	Yes	Yes	Yes	Yes	Yes	**No**
	Correlation	Yes	Yes	Yes	Yes	Yes	**No**
	PCA	Yes	Yes	Yes	Yes	Yes	**No**
	Bigtrend	Yes	Yes	Yes	Yes	Yes	**No**
	PPI	Yes	Yes	Yes	Yes	Yes	Yes
EnrichmentScope	ORA algorithm	Yes	Yes	Yes	Yes	Yes	Yes
	GSEA algorithm	Yes	Yes	Yes	Yes	Yes	**No**
	Dotplot	Yes	Yes	Yes	Yes	Yes	Yes
	Heatmap	Yes	Yes	Yes	Yes	Yes	Yes
	number_deps	Yes	Yes	Yes	Yes	Yes	Yes
	enrichment_network	Yes	Yes	Yes	Yes	Yes	Yes
	enrichment_map	Yes	Yes	Yes	Yes	Yes	Yes
Nebula	Barplot	Yes	Yes	Yes	Yes	Yes	Yes
	dotplot_enrichment	Yes	Yes	Yes	Yes	Yes	Yes
	diff_reg	Yes	Yes	Yes	Yes	Yes	Yes
	protein_overlap	Yes	Yes	Yes	Yes	Yes	Yes
	enrichment_overlap	Yes	Yes	Yes	Yes	Yes	Yes
	similarity_heatmap	Yes	Yes	Yes	Yes	Yes	Yes
	similarity_network	Yes	Yes	Yes	Yes	Yes	Yes
	fisher_heatmap	Yes	Yes	Yes	Yes	Yes	Yes
	fisher_network	Yes	Yes	Yes	Yes	Yes	Yes
	whole_network	Yes	Yes	Yes	Yes	Yes	Yes
	circular_path	Yes	Yes	Yes	Yes	Yes	Yes
	circos_plot	Yes	Yes	Yes	Yes	Yes	Yes

4 CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

4.1 THESIS CONSIDERATIONS

This chapter is a full article currently in revision process. Here, we analyzed the proteome of nine in vitro cellular models infected with SARS-CoV-2 and compared them with respective Mock controls. These models can be divided into four categories, according to their ontologies: central nervous systems cells, immune cells, gastroenterological cells and adipocytes. As occurred in Chapter 1, we used the bottom-up approach for all cell types and performed MS analysis in a Q-TOF with HDMS^E acquisition mode (DIA), followed by protein identification, quantitation and differential proteomics in Progenesis Qi for Proteomics (Non-Linear, Waters). However, in this article, the OmicScope was in development, and it was responsible for more than 70% of figures generated and used to discuss our results. Moreover, all the tools used for data analysis were freely available as a python package or a web application, which did not pose any issue for data analysis reproducibility.

Compared to Chapter 1, the whole data analysis process took some months to be carried out and submitted for publication. Indeed, most of the time between data generation and submission was dedicated to writing the manuscript and discussing the results with the literature.

Regarding visualization features, figures 2-5 also present a template, which comprises: Venn Diagram, heatmap (or boxplot in figure 2), dot plot of enrichment results, and circular diagram. Heatmaps and boxplots were used to visualize protein fold-changes of proteins across the evaluated cell types, showcasing the comparison of groups in protein abundance level. The dot plot figure was used to compare the cellular response to SARS-CoV-2 in an enrichment level, highlighting unique or overlapped enriched terms. Joining fold-change attribute and enrichment outcomes, we also compared the cellular response using the circular diagram strategy. These strategies adopted allowed us to discuss and evaluate the changes triggered by SARS-CoV-2 in a

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

112

more comprehensive manner, integrating data from differential proteomics and enrichment analysis.

Finally, this chapter shows how OmicScope can be used to improve the data analysis process, joining quantitative proteomics with enrichment results using meta-analysis approach.

4.2 ABSTRACT

The COVID-19 pandemic was initiated by the rapid spread of a SARS-CoV-2 strain. Though mainly classified as a respiratory disease, SARS-CoV-2 infects multiple tissues throughout the human body, leading to a wide range of symptoms in patients. To better understand how SARS-CoV-2 affects the proteome from cells with different ontologies, this work generated an infectome atlas of 9 cell models, including cells from brain, blood, digestive system, and adipocyte tissue. Our data shows that SARS-CoV-2 infection mainly triggers dysregulations on proteins related to cellular structure and energy metabolism. Despite these pivotal processes, heterogeneity of infection was also observed, highlighting many proteins and pathways uniquely dysregulated in one cell type or ontological group. These data have been made searchable online via a tool that will permit future submissions of proteomic data (https://reisdeoliveira.shinyapps.io/Infectome_App/) to enrich and expand this knowledgebase.

4.3 INTRODUCTION

The COVID-19 pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been responsible for more than 649·8 million confirmed cases and 6·6 million deaths as of December 2022[223], drastically affecting the global economy and social behavior. COVID-19 is manifested by pulmonary infection and a cytokine storm in more severe stages of the disease. However, a growing number of studies have also proven the presence of SARS-CoV-2 in various other tissues and

---

[223] "COVID-19 Cases | WHO COVID-19 Dashboard," datadot, accessed January 11, 2024, https://data.who.int/dashboards/covid19/cases.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

113

biofluids, such as cardiac muscle tissue, the central nervous system (CNS), the kidneys, the gastrointestinal system, and adipose tissue[224].

SARS-CoV-2 is able to infect human cells and tissues mainly through their expression of angiotensin-converting enzyme 2 (ACE2), the main port of entry for the virus[225], however, other receptors have been described as secondary for viral infection, such as TPMRS22, NPR1, and BSG/CD147[226]. Once inside the cell, SARS-CoV-2 then alters cell function at the proteomic level at minimum, often resulting in tissue dysfunction and damage.

Xiu *et al.*, 2021 performed a broad proteomic analysis in distinct post-mortem tissues from several organs obtained from patients who died of COVID-19[227], in which they discovered distinct patterns of protein expression for each infected organ. To complement this study, an easily searchable compilation of general and tissue-specific molecular mechanisms that are triggered by the virus in different cell types is relevant. With this in mind, we analyzed proteomic data from several distinct SARS-CoV-2-infected cell types: neural stem cell (NSC)-derived astrocytes and neurons[228]; SH-SY5Y (human neuroblastoma cells) and CACO-2 (intestinal epithelial cells) cultured cells; T-

---

[224] Xin Zou et al., "Single-Cell RNA-Seq Data Analysis on the Receptor ACE2 Expression Reveals the Potential Risk of Different Human Organs Vulnerable to 2019-nCoV Infection," *Frontiers of Medicine* 14, no. 2 (April 2020): 185–92, https://doi.org/10.1007/s11684-020-0754-0; Aravinthan Varatharaj et al., "Neurological and Neuropsychiatric Complications of COVID-19 in 153 Patients: A UK-Wide Surveillance Study," *The Lancet. Psychiatry* 7, no. 10 (October 2020): 875–82, https://doi.org/10.1016/S2215-0366(20)30287-X; Lu Lin et al., "Gastrointestinal Symptoms of 95 Cases with SARS-CoV-2 Infection," *Gut* 69, no. 6 (June 2020): 997–1001, https://doi.org/10.1136/gutjnl-2020-321013; Yassine Yachou et al., "Neuroinvasion, Neurotropic, and Neuroinflammatory Events of SARS-CoV-2: Understanding the Neurological Manifestations in COVID-19 Patients," *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 41, no. 10 (October 2020): 2657–69, https://doi.org/10.1007/s10072-020-04575-3; Aakriti Gupta et al., "Extrapulmonary Manifestations of COVID-19," *Nature Medicine* 26, no. 7 (July 2020): 1017–32, https://doi.org/10.1038/s41591-020-0968-3.
[225] Zou et al., "Single-Cell RNA-Seq Data Analysis on the Receptor ACE2 Expression Reveals the Potential Risk of Different Human Organs Vulnerable to 2019-nCoV Infection"; Li Yang et al., "COVID-19: Immunopathogenesis and Immunotherapeutics," *Signal Transduction and Targeted Therapy* 5, no. 1 (July 25, 2020): 128, https://doi.org/10.1038/s41392-020-00243-2; Nie et al., "Multi-Organ Proteomic Landscape of COVID-19 Autopsies."
[226] Titilola D. Kalejaiye et al., "SARS-CoV-2 Employ BSG/CD147 and ACE2 Receptors to Directly Infect Human Induced Pluripotent Stem Cell-Derived Kidney Podocytes," *Frontiers in Cell and Developmental Biology* 10 (2022): 855340, https://doi.org/10.3389/fcell.2022.855340; Markus Hoffmann et al., "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor," *Cell* 181, no. 2 (April 16, 2020): 271-280.e8, https://doi.org/10.1016/j.cell.2020.02.052.
[227] Nie et al., "Multi-Organ Proteomic Landscape of COVID-19 Autopsies."
[228] Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients."

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

114

cells and monocytes isolated from human blood[229], hepatocytes; and adipocytes isolated from visceral and subcutaneous tissue[230].

This information brings new insight into how the virus disturbs biochemical processes and can therefore be used to generate new integrated therapeutic targets, which may be useful to treat COVID-19 systemically and in a tissue-specific manner. Moreover, with the creation of the online SARS-Cov-2 infectome, a new tool to research proteins of interest will become available for public browsing, also including how SARS-CoV-2 can affect these proteins in each cell type and which biological mechanisms are altered by the infection.

4.4 RESULTS

Before any subsequent analyses, a viral kinetics analysis at three time points indicated that all cells analyzed here were indeed infected by SARS-CoV-2. Data regarding T-cells, monocytes, adipocytes, and astrocytes were included in their originally published articles[231]. Despite being able to replicate within CNS cells, the virus was not able to secrete viral particles into the medium; all other cell types showed signs of both viral replication and secretion (Supplementary Figure 1).

Proteomic analyses of SARS-CoV-2-infected cells were used to build what we will hereby refer to as the SARS-CoV-2 infectome, a compilation of differentially regulated proteins in cells in response to SARS-CoV-2 infection (Figure 1A, and Supplementary Table 1). Among all analyses, a total of 3098 proteins were quantified, of which 1652 were found differentially regulated (p-value < 0·05, Figure 1B). Pair-wise Pearson's correlation analysis, followed by hierarchical clustering, showed that almost all cell types clustered as expected by their ontology at the whole-proteome level, a

---

[229] Natalia S. Brunetti et al., "SARS-CoV-2 Uses CD4 to Infect T Helper Lymphocytes," *eLife* 12 (July 31, 2023): e84790, https://doi.org/10.7554/eLife.84790.

[230] Tatiana Dandolini Saccon et al., "SARS-CoV-2 Infects Adipose Tissue in a Fat Depot- and Viral Lineage-Dependent Manner," *Nature Communications* 13, no. 1 (September 29, 2022): 5722, https://doi.org/10.1038/s41467-022-33218-8.

[231] Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients"; Brunetti et al., "SARS-CoV-2 Uses CD4 to Infect T Helper Lymphocytes"; Saccon et al., "SARS-CoV-2 Infects Adipose Tissue in a Fat Depot- and Viral Lineage-Dependent Manner"; Ana Campos Codo et al., "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis," *Cell Metabolism* 32, no. 3 (September 1, 2020): 437-446.e5, https://doi.org/10.1016/j.cmet.2020.07.007.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

115

pattern that was markedly distinct while considering only differentially regulated proteins (Supplementary Figure 2). These differences may be explained by the high diversity effects elicited by SARS-CoV-2 in each cell type, which can vary in the number of differentially regulated proteins as well as their regulation direction (Figure 1C). Interestingly, neuronal cell models were unique in that they presented more upregulated proteins in their infectome. Additionally, we observed that 551 proteins were dysregulated in more than one cell type; however, none were found differentially regulated in all cell types (Figure 1D, Supplementary Figure 3).

A pathway enrichment analysis for each infectome identified 151 biological pathways associated with SARS-CoV-2 infection, with the neuronal infectome being the largest contributor (Supplementary Figure 4). When filtering for pathways enriched in at least 50% of infectomes, SARS-CoV-2 modulated the protein related to energy metabolism (mainly glycolysis), infectious diseases, protein metabolism (both synthesis and degradation processes), signaling/homeostatic pathways, and potential comorbidities (Figure 1E) in nearly all cells, in line with expected responses to viral infection and viral particle production. Six cellular infectomes (neurons, astrocytes, monocytes, hepatocytes, and both types of adipocytes) also enriched for Coronavirus disease, highlighting the canonical infection pathways in several cell types outside of the respiratory system. The absence of these results in SH-SY5Y cells and T-cells suggests other avenues for viral infection and proliferation.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

116
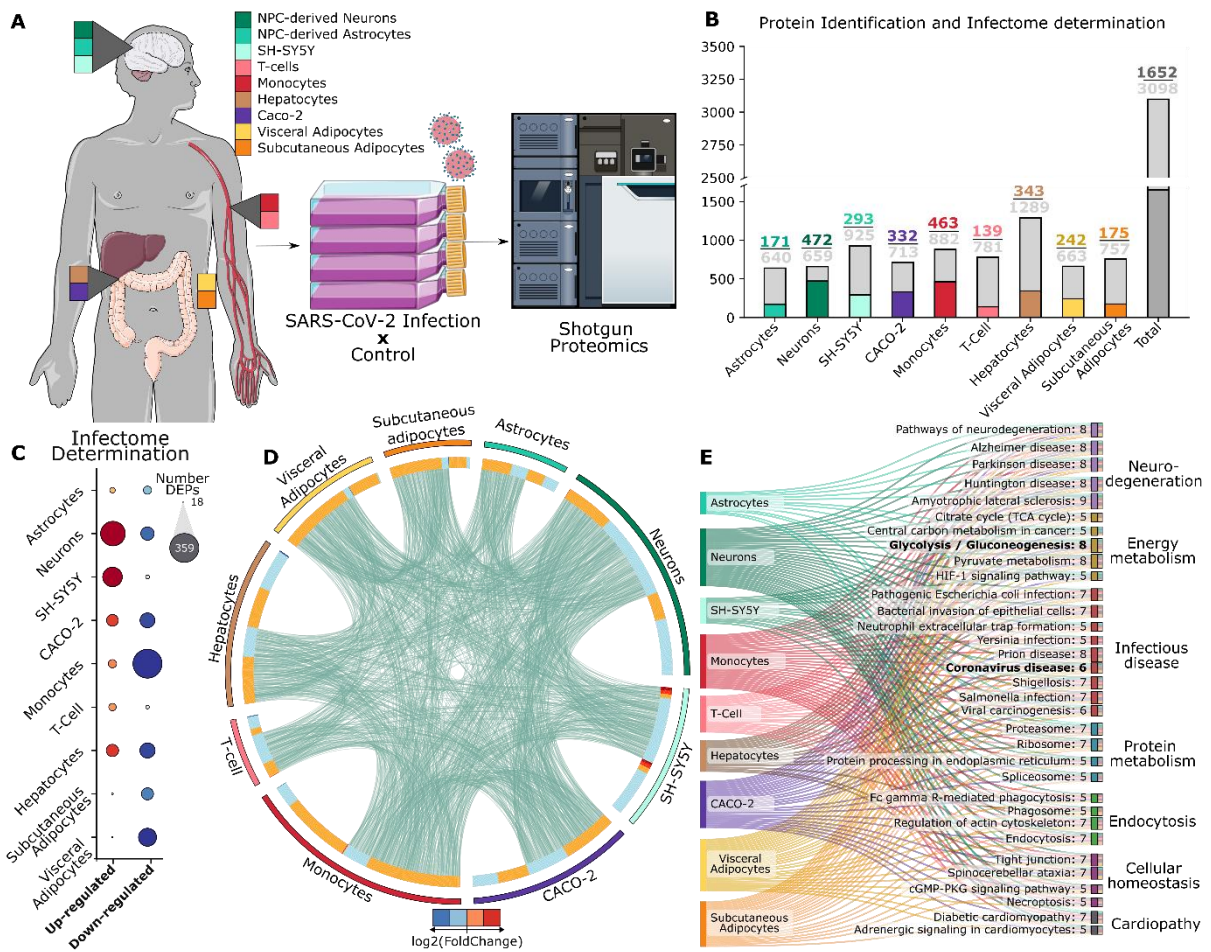
Figure 4-1 SARS-CoV-2 infects cells from different tissues and elicits changes in protein abundance. **A)** Cell lines from distinct tissues and biological systems were infected with SARS-CoV-2 in vitro and the infectomes (proteins differentially regulated due to viral response) were obtained by shotgun proteomics. **B)** Total number of identified proteins (gray) and differentially regulated proteins (colored) in each cell type. **C)** Proteins differentially regulated in the respective infectomes separated by direction of regulation. Bubble size indicates the number of differentially regulated proteins (DRPs). **D)** Overview of SARS-CoV-2 infectome in all cell types (outer circle), showing proteins that are shared among groups as well as their respective regulation in the heatmap (inner circle). **E)** Sankey plot of pathways enriched in at least 50% of infectomes.

Given the variety of cellular functions and complexity of the respective infectomes of each cell type, we categorized differential protein expression by ontology: CNS cells (neurons, differentiated SH-SY5Y neuron-like cells, and astrocytes), gastroenterological cells (CACO-2, hepatocytes), white blood cells (monocytes and T-cells), and adipose tissue (differentiated visceral and subcutaneous adipocytes). This allowed an evaluation of similarity among cellular infectomes within ontological groups, useful in increasing our understanding about changes triggered by SARS-CoV-2 to better predict body response to infection.

All results and findings derived from these proteomic analyses were compiled and published in a unique database, forming the SARS-CoV-2 Infectome Atlas. This atlas, being the first of its kind, details the proteomic dysregulations elicited by SARS-CoV-2 in nine cellular infectomes, groupable by their tissue ontology. Individual proteins are searchable, providing information about normalized fold change differences, as well as altered pathways and their constituent proteins. This allows for targeted searching of proteins and pathways of interest within a SARS-CoV-2 context (https://reisdeoliveira.shinyapps.io/Infectome_App/).

4.5 SARS-CoV-2 INFECTION MODULATES CNS CELL PROTEOMES

The infectome of CNS cells was composed of 810 differentially regulated proteins, with astrocytes presenting the fewest dysregulations (Figure 2A). Nine differentially regulated proteins were observed in all cell types, primarily associated with translation machinery (RPL29, RPL22, PABPC1, RPL12) and energy metabolism (TKT, PHGDH). Interestingly, however, these proteins presented a distinct regulation profile in each CNS cell, despite similarities in function and role.

In energy metabolism, for instance, PHGDH, a pivotal protein throughout neurodevelopment, was upregulated in all infected cell types; whereas TKT was upregulated in both neuronal cell types but downregulated in astrocytes (Figure 2B). Astrocytic marker vimentin (VIM), also present in excitatory neurons, was downregulated in infected astrocytes, and upregulated in both infected neuronal cell types (Figure 2B). This divergent regulation pattern indicates that, even when cells share proximal ontological characteristics, SARS-CoV-2 can induce different and even divergent protein-level responses.

To understand how SARS-CoV-2 can potentially affect these *in vitro* CNS models at the functional level, the Kyoto Encyclopedia for Genes and Genomes (KEGG) was used for *in silico* enrichment analyses[232]. Performing a pair-wise comparison of the top ten pathways enriched for each cellular model (Figure 2C), the NPC-derived neuron infectome was the only CNS cell type seen to incorporate all the top ten terms found in

---

[232] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research* 28, no. 1 (January 1, 2000): 27–30, https://doi.org/10.1093/nar/28.1.27.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

118

the other CNS cell types. When both neuronal infectomes were compared, differentiated
SH-SY5Y cells enriched 67% fewer terms than NPC-derived neurons, highlighting that
SARS-CoV-2 response can differ by cell model, even of the same cell type
(Supplementary Figure 4). Nevertheless, at the pathway level, these cellular infectomes
still shared several enriched pathways, such as glycolysis, pentose phosphate pathway,
and proteasomes (Figure 2C).

Figure 4-2 SARS-CoV-2 induces changes in the proteomes of CNS cell types. A) Venn diagram showing
differentially regulated proteins among NPC-derived astrocytes and neurons and differentiated SH-SY5Y
neurons. B) Normalized abundance of the differentially regulated proteins associated with energy
metabolism (PHGDH, TKT) and cellular marker (VIM) for CNS cells. Gray bars represent the abundance

Chapter 3: Diving into the proteomic atlas of SARS-CoV-2 infected cells: insights of viral infection in different cell types

119

in uninfected cells. C) Pathway enrichment analyses for each CNS infectome. The top 10 pathways were selected for each cellular model, and then a pair-wise comparison was performed to visualize enrichment in other cell types. Bubble size indicates adjusted p-value on a -log10 scale. D) Upregulated (red) and downregulated (blue) proteins associated with glycolysis in different CNS cell types. Proteins in bold and italic are pivotal in glycolysis and lactate generation, respectively.

Since glycolysis was found enriched among the top ten enriched terms of all CNS infectomes (p-adj. < 1·0e-6), we highlighted the proteins differentially regulated in this pathway in each cell type (Figure 2D). Most of the proteins (66%) associated with glycolysis were upregulated. In neuronal models, the proteins that play pivotal roles in glycolytic regulation (hexokinase [HK1], phosphofructokinase [PFKP/PFKL], and pyruvate kinase [PKM]) were upregulated after infection (Figure 2D, bold); while in NSC-derived cells, lactate dehydrogenase-associated LDHAL6B was down-regulated. Additionally, PGM2 and PGM4 were upregulated in SH-SY5Y cells and NSC-derived astrocytes, both of which are essential in 5'-phosphopentose metabolism and nucleotide synthesis.

4.6 SARS-CoV-2 infection modulates gastroenterological cell proteomes

Proteome changes were evaluated in intestinal epithelial cells (CACO-2 cell line) and hepatoma cells (HepG2 cell line), both of which derive from cancer patients and present epithelial-like morphology. Infected CACO-2 cells presented 332 deregulated proteins while HepG2 cells presented 343 deregulated proteins; 38 proteins were deregulated in both cell lines (Figure 3A). These overlapping proteins are related to vesicle-mediated transport (import into the nucleus, endocytosis, and exocytosis) and cellular metabolic processes (Figure 3B).

Of the terms obtained from KEGG enrichment, 11 pathways were shared between both gastroenterological cell lines (Figure 3C). Among the enriched terms, we found bacterial invasion of epithelial cells, Salmonella infection, and pathogenic *E. coli* infection. These processes could be associated with the diarrhea induced by SARS-CoV-2 infection through a mechanism similar to previously documented bacterial invasion processes in the gut. Among the proteins associated with these pathways, only CDC42, ANXA2 and ARF1 were differentially regulated in both cell types; nevertheless, protein

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

120

families (such as tubulin or MAPK), are shared in both cell types (Figure 3D), which could lead to similar effects on epithelial homeostasis.

Figure 4-3 SARS-CoV-2 induces changes in the proteomes of gastroenterological cell types. A) Venn diagram showing differentially regulated proteins in HepG2 and CACO-2 cells. B) Heatmap of fold changes for each protein found in both gastroenterological cell types. C) Pathway enrichment analyses for hepatocytes and CACO-2 cells. Bubble size indicates adjusted p-value on a -log10 scale. D) Upregulated (red) and downregulated (blue) proteins related to bacterial invasion of epithelial cells. Proteins highlighted in bold are from the MAPK and tubulin families.

4.7 SARS-CoV-2 INFECTION MODULATES IMMUNE CELL PROTEOMES

The immunological cells analyzed in our current datasets consist of SARS-CoV-2-infected primary monocytes and T lymphocytes. By testing these two cell types, both myeloid and lymphoid lineages can be tested, covering the innate immune response with monocytes and the adaptive response with the lymphocytes. Our quantitative analysis found 139 proteins dysregulated in T-cells and 463 in monocytes, with 35 proteins dysregulated in both cell types (Figure 4A). These proteins are associated with cellular metabolic processes; though the heatmap shows that fold change – and therefore likely the cellular response as well – can vary between myeloid and lymphoid cells (Figure 4B).

Pathways enriched by the monocyte and lymphocyte infectomes (Figure 4C) presented a large overlap between cell types, of which several terms were associated with immune system responses (such as infections and diseases). Among these terms, we highlight FcγR-mediated phagocytosis, a signaling pathway that interplays adaptive and innate immune responses mediated by antibodies. We also noticed that, within this pathway, lymphocytes presented only upregulated proteins, while monocytes were more diverse in their response (Figure 4D). Only the monocyte infectome was enriched for COVID-19 disease, suggesting that these myeloid cells may play a pivotal role in the canonically activated pathways during the course of SARS-CoV-2 infection.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-COV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

122

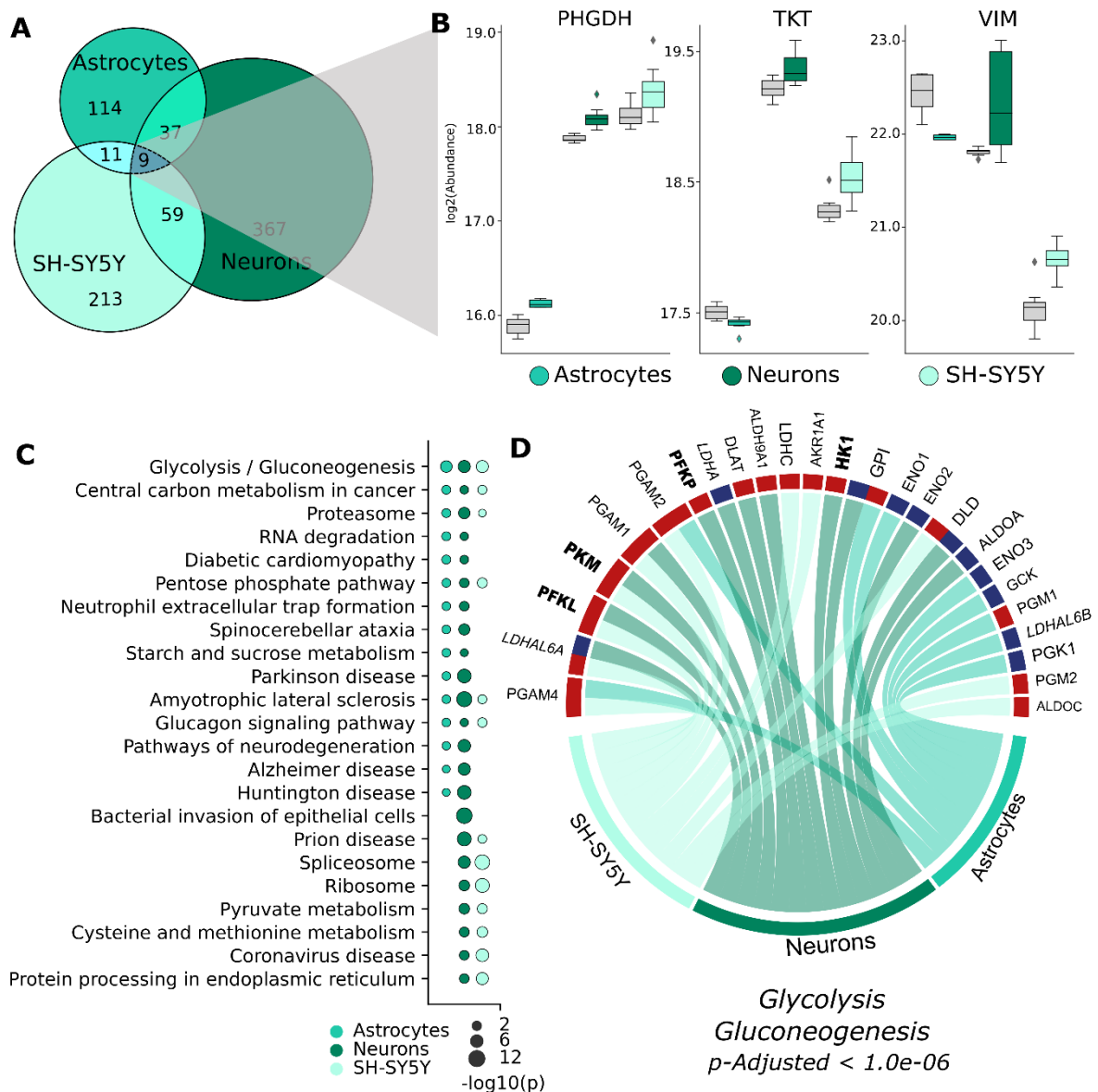*Figure 4-4 SARS-CoV-2 induces changes in the proteomes of immunological cells.* A) Venn diagram showing differentially regulated proteins in monocytes and lymphocytes. B) Heatmap of fold changes for proteins found in both immune cells. C) Pair-wise comparison of top pathways enriched against the KEGG database. Bubble size indicates adjusted p-value on a -log10 scale. D) Upregulated (red) and downregulated (blue) proteins involved in Fc gamma R-mediated phagocytosis.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

123

4.8 SARS-CoV-2 INFECTION MODULATES VISCERAL AND SUBCUTANEOUS ADIPOCYTE PROTEOMES

Since visceral and subcutaneous adipocytes are known targets for SARS-CoV-2 infection and storage[233] we investigated the proteomic modulations that occur in these cell types. 242 (visceral) and 175 (subcutaneous) proteins were found to be deregulated after SARS-CoV-2 infection. While nearly 60% of identified proteins were observed in both cell types, 44 proteins were dysregulated in both (Figure 5A), most of which were in the same direction. No pathways were found to be enriched, despite the presence of many glucose metabolism proteins (Figure 5B).

When comparing the top enriched pathways from both adipocyte types, only proteasome processes were unique to visceral adipocytes; subcutaneous adipocytes presented more unique processes, namely protein processing in the endoplasmic reticulum, central carbon metabolism in cancer, glucagon signaling pathway, and necroptosis (Figure 5C). Despite these differences, TCA cycle proteins were found downregulated in both types, indicating dysregulated NAD^+ and FAD reduction, culminating in lowered ATP production (Figure 5D).

---

[233] Paul MacDaragh Ryan and Noel M. Caplice, "Is Adipose Tissue a Reservoir for Viral Spread, Immune Activation, and Cytokine Amplification in Coronavirus Disease 2019?," *Obesity (Silver Spring, Md.)* 28, no. 7 (July 2020): 1191–94, https://doi.org/10.1002/oby.22843; Ilja L. Kruglikov and Philipp E. Scherer, "The Role of Adipocytes and Adipocyte-Like Cells in the Severity of COVID-19 Infections," *Obesity (Silver Spring, Md.)* 28, no. 7 (July 2020): 1187–90, https://doi.org/10.1002/oby.22856.
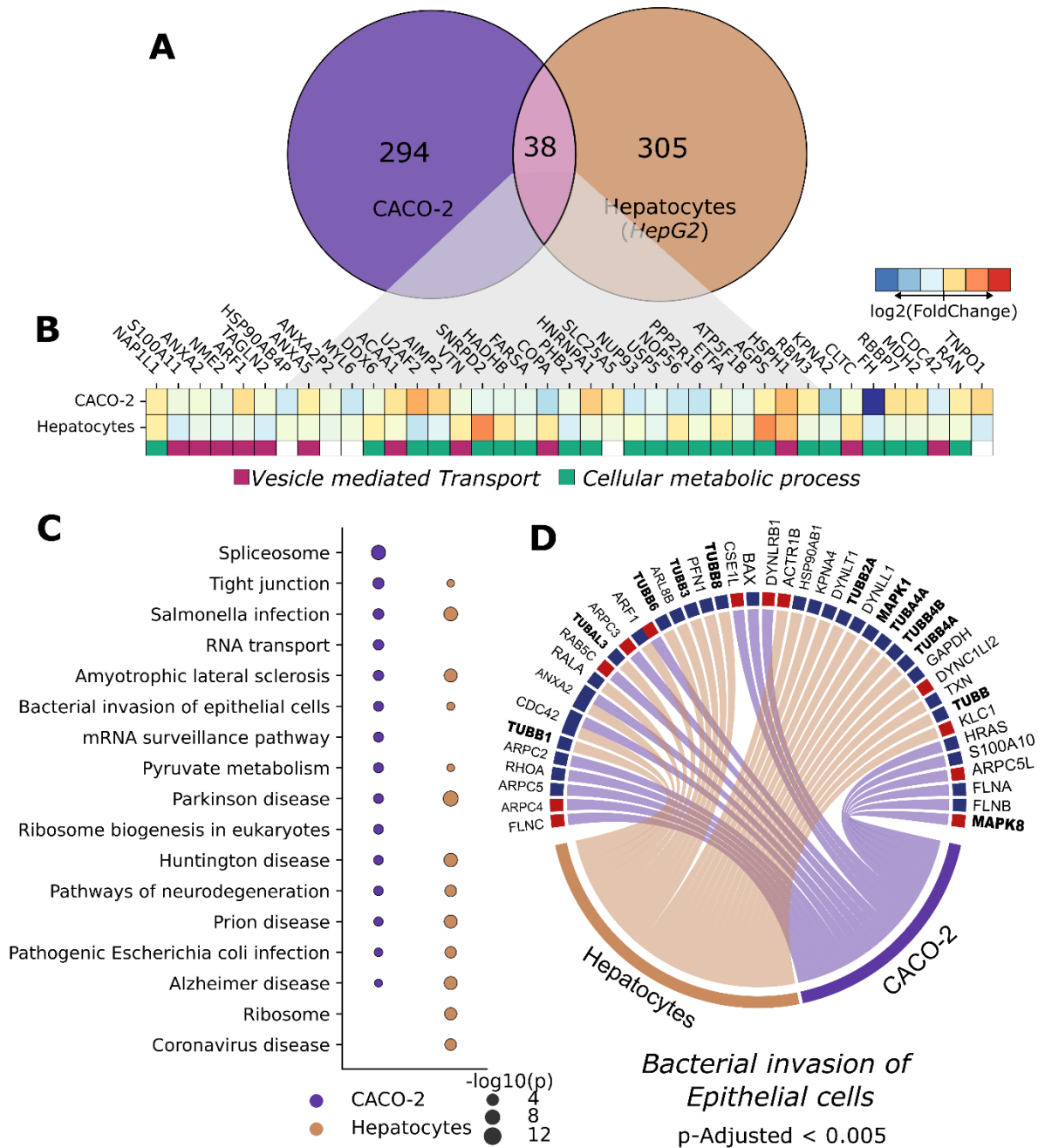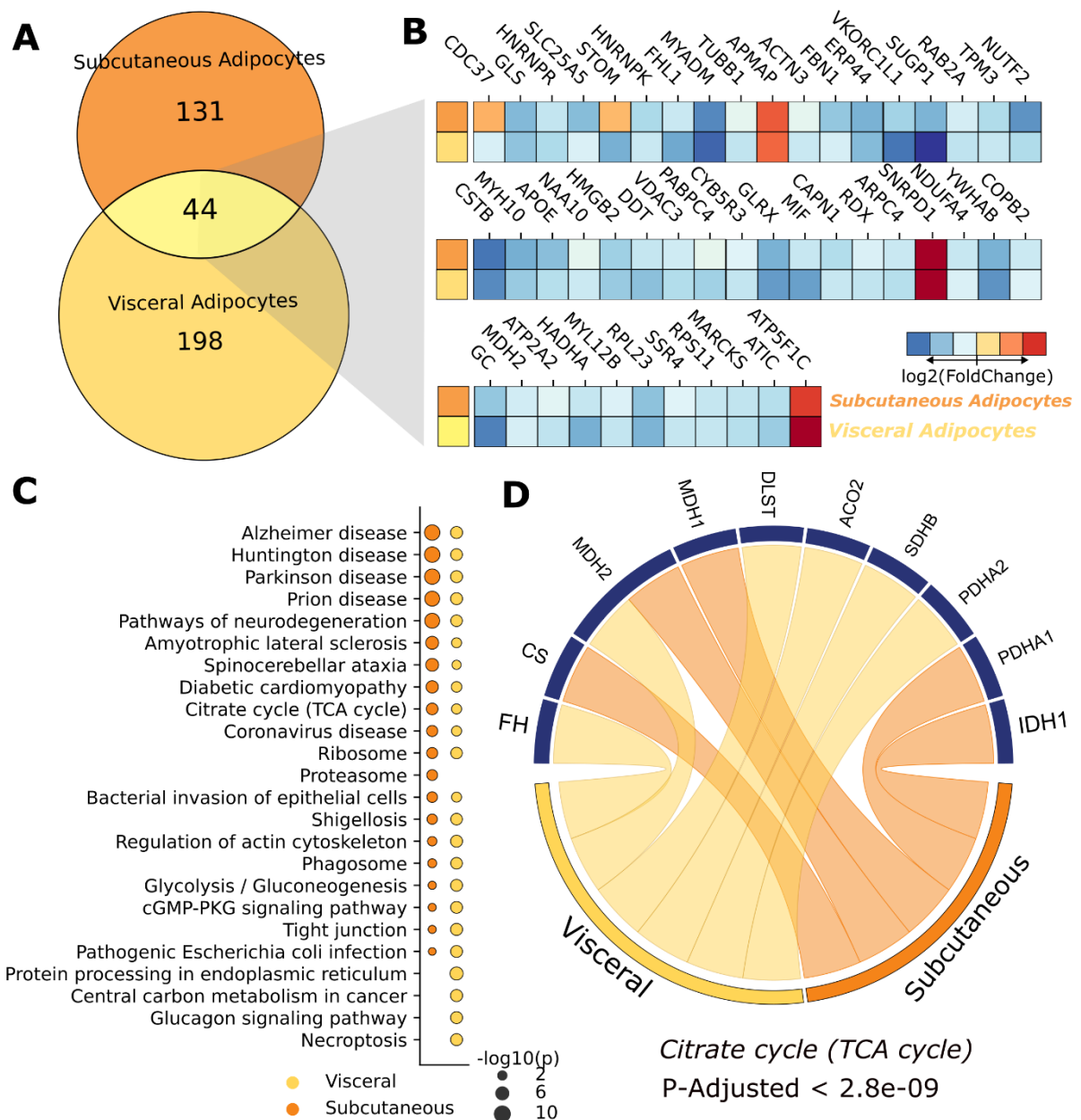
Figure 4-5 SARS-CoV-2 induces changes in the proteome of adipocytes. A) Venn diagram showing differentially regulated proteins in visceral and subcutaneous adipocytes. B) Heatmap of fold changes for proteins found in both adipocyte types. C) Pair-wise comparison of top pathways enriched against KEGG database. Bubble size indicates adjusted p-value on a -log10 scale. D) Upregulated (red) and downregulated (blue) proteins involved in the citrate (TCA) cycle.

4.9 DISCUSSION

COVID-19 was initially classified as a respiratory disease; however, many studies have since proven that SARS-CoV-2 has the capacity to infect several other cell types and tissues. It has since been classified as a systemic disease with effects including

Chapter 3: Diving into the proteomic atlas of SARS-CoV-2 infected cells: insights of viral infection in different cell types

125

kidney failure, permanent liver damage, diarrhea, and neurological symptoms and sequelae. An increasing necessity to better understand the molecular and biochemical mechanisms triggered by SARS-CoV-2 infection throughout the human body led us to create the SARS-CoV-2 Infectome Atlas, in which several human cell types infected by SARS-CoV-2 were analyzed.

SARS-CoV-2 has been found in human brain tissue collected *postmortem* and has also been shown to infect CNS models such as organoids and hamster brains[234]. Nonetheless, there is still no consensus as to if the brain alterations are directly related to the viral infection or if it is a secondary effect. Moreover, the mechanisms utilized by the virus to infect these cells and which changes are triggered or provoked by its replication are still a matter of discussion. This work reinforces the ability of SARS-CoV-2 to infect different CNS cell types and subsequently trigger proteomic changes. NPC-derived neurons exhibited the most affected model by the size of its infectome. Around 25% of the total identified proteomes from NSC-derived astrocytes, neurons, and differentiated SH-SY5Y neuroblastoma cells were deregulated after infection. Neurons or astrocytes have been hypothesized to be some of the first affected cells within the CNS due to their proximity to the blood-brain barrier (BBB), which would act as the port of entry of the virus into the CNS. Another noncompeting hypothesis is that axonal projections that reach the olfactory bulb may carry the virus into the CNS, thereby bypassing the BBB[235].

Furthermore, infection of the CNS increases IL-6 levels, primarily as a response by astrocytes and microglia, potentially reducing the integrity of the BBB and promoting further viral invasion into the CNS[236]. This increase in IL-6 levels has been

---

[234] Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients"; Fadi Jacob et al., "Human Pluripotent Stem Cell-Derived Neural Cells and Brain Organoids Reveal SARS-CoV-2 Neurotropism Predominates in Choroid Plexus Epithelium," *Cell Stem Cell* 27, no. 6 (December 3, 2020): 937-950.e9, https://doi.org/10.1016/j.stem.2020.09.016.

[235] Yachou et al., "Neuroinvasion, Neurotropic, and Neuroinflammatory Events of SARS-CoV-2"; Iván Alquisiras-Burgos et al., "Neurological Complications Associated with the Blood-Brain Barrier Damage Induced by the Inflammatory Response During SARS-CoV-2 Infection," *Molecular Neurobiology* 58, no. 2 (February 2021): 520–35, https://doi.org/10.1007/s12035-020-02134-7.

[236] Orkide O. Koyuncu, Ian B. Hogue, and Lynn W. Enquist, "Virus Infections in the Nervous System," *Cell Host & Microbe* 13, no. 4 (April 17, 2013): 379–93, https://doi.org/10.1016/j.chom.2013.03.010; María Erta, Albert Quintana, and Juan Hidalgo, "Interleukin-6, a Major Cytokine in the Central Nervous System," *International Journal of Biological Sciences* 8, no. 9 (2012): 1254–66, https://doi.org/10.7150/ijbs.4679.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

126

observed in the cerebrospinal fluid of patients with COVID-19[237]. In this study, the astrocyte infectome enriched for biological pathways involved in other chronic inflammatory diseases and responses to infectious agents, suggesting that these cells may be involved in the primary response to SARS-CoV-2 infection.

Within brain tissue, SARS-CoV-2 presence is concentrated in astrocytes despite a more accentuated proteomic change in neurons[238]. While comparing results from the enrichment analysis, NPC-derived neurons and differentiated SH-5YSY cells had a higher overlap of enriched pathways than with astrocytes, likely due to ontological similarity. Many of the enriched pathways were associated with neurodegenerative disease, glucose energy metabolism, COVID-19 disease, and ribosomal translational functions. The enrichment for neurodegenerative diseases could potentially be explained by increases in cell death and apoptosis.

Energy metabolism was affected in all CNS cells analyzed in this study, making it a pathway of interest for further study. Proteins belonging to glycolytic pathways were generally upregulated in neurons and SH-5YSY cells, while in astrocytes they were downregulated. Calmodulin-dependent protein kinase (CAMK2D) was found upregulated in astrocytes, a protein that plays an important role in glutamatergic synapses through the post NMDA receptor activation events, possibly representing a higher uptake of glutamate from the extracellular medium.

In all cell types, PHGDH was found upregulated (Figure 2B), a protein which catalyzes the synthesis of L-serine, the precursor to D-serine, which is essential for NMDA receptor function throughout neurodevelopment. The upregulation of PHGDH in these cells could be related to an increased consumption of glutamate by the infected CNS cells to offset the increased energy demands resulting from the replication of viral particles. Crunfli *et al.* showed an imbalance in the glutamate/glutamine interplay

---

[237] Keith D. Rochfort et al., "Downregulation of Blood-Brain Barrier Phenotype by Proinflammatory Cytokines Involves NADPH Oxidase-Dependent ROS Generation: Consequences for Interendothelial Adherens and Tight Junctions," *PloS One* 9, no. 7 (2014): e101815, https://doi.org/10.1371/journal.pone.0101815; Jiyong Zhang et al., "Anti-IL-6 Neutralizing Antibody Modulates Blood-Brain Barrier Function in the Ovine Fetus," *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 29, no. 5 (May 2015): 1739–53, https://doi.org/10.1096/fj.14-258822.
[238] Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients."

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

127

between SARS-CoV-2 infected astrocytes and neurons[239] and proposed that the higher mortality observed in neurons and SH-5YSY cells after being exposed to medium conditioned with infected astrocytes is due to a disruption of this feedback system between astrocytes and neurons.

More than half of COVID-19 patients report gastrointestinal symptoms, most commonly diarrhea, but also including anorexia, nausea, vomiting, abdominal pain and gastrointestinal bleeding[240]. ACE2 is abundantly expressed in the intestinal epithelial cell cytoplasm [241], enabling SARS-CoV-2 to directly target gastrointestinal cells, specifically gastric and intestinal epithelial cells[242]. Based on this evidence and the various gastroenterological symptoms, we chose to investigate the cellular response to infection in epithelial intestinal cells (CACO-2) and hepatocytes when creating our SARS-CoV-2 infection proteome map. The pathway analysis from the 332 proteins in the CACO-2 infectome mainly enriched for pathways related to chronic inflammatory diseases, energy metabolism, and biosynthesis of amino acids, highlighting the necessity of SARS-CoV-2 to induce changes in energy and protein metabolism to quickly replicate and create new viral particles.

Bojkova *et al.* used the same cell line labelled with stable isotopic amino acids (SILAC) to study the proteomic profile after SARS-CoV-2 infection[243], discovering 459 proteins dysregulated by viral infection, enriching for glycolysis and nucleotide metabolism pathways. They also tested inhibitors of protein translation, RNA splicing, glycolysis, and nucleotide synthesis at non-toxic concentrations, which, in line with our findings, inhibited viral replication. Another study raised the hypothesis that, upon the ACE2-mediated entry of SARS-CoV-2 into the gastrointestinal tract via the small

[239] Crunfli et al.

[240] Han-Yu Lei et al., "Potential Effects of SARS-CoV-2 on the Gastrointestinal Tract and Liver," *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie* 133 (January 2021): 111064, https://doi.org/10.1016/j.biopha.2020.111064.

[241] Lulin Zhou et al., "SARS-CoV-2 Targets by the pscRNA Profiling of ACE2, TMPRSS2 and Furin Proteases," *iScience* 23, no. 11 (November 20, 2020): 101744, https://doi.org/10.1016/j.isci.2020.101744.

[242] Fei Xiao et al., "Evidence for Gastrointestinal Infection of SARS-CoV-2," *Gastroenterology* 158, no. 6 (May 2020): 1831-1833.e3, https://doi.org/10.1053/j.gastro.2020.02.055.

[243] Denisa Bojkova et al., "Proteomics of SARS-CoV-2-Infected Host Cells Reveals Therapy Targets," *Nature* 583, no. 7816 (July 2020): 469–72, https://doi.org/10.1038/s41586-020-2332-7.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

128

intestine, viral invasion and expansion occur, triggering gastrointestinal inflammation and potentially evolving into a strong cytokine response[244].

Within the gastroenterological system, the liver is the organ that is considered to be most impacted by COVID-19, with approximately 58-75% of patients presenting hepatic injury and/or increased levels of transaminases[245]. While patients with COVID-19 generally present only moderate liver damage, more severe cases have been reported[246]. Wang and colleagues were the first to provide evidence of cytopathy in hepatocytes caused by SARS-CoV-2 infection that could cause liver impairment in the infected host[247]. Our analysis indeed shows that SARS-CoV-2 elicits proteomic changes in the HepG2 cell line, affecting proteins related to energy metabolism, the proteasome, amino acid biosynthesis, and pyruvate metabolism. When filtering for proteins associated with carbon metabolism in hepatocytes, there was an increase in proteins involved in pyruvate production, indicating once again that glucose metabolism is crucial during viral replication.

Adipose tissue has also been associated with SARS-CoV-2 infection and the subsequent inflammatory response[248]. Since the entry of SARS-CoV-2 into the digestive

---

[244] Hui Zhang et al., "Specific ACE2 Expression in Small Intestinal Enterocytes May Cause Gastrointestinal Symptoms and Injury after 2019-nCoV Infection," *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 96 (July 2020): 19–24, https://doi.org/10.1016/j.ijid.2020.04.027.

[245] Chaolin Huang et al., "Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China," *Lancet (London, England)* 395, no. 10223 (February 15, 2020): 497–506, https://doi.org/10.1016/S0140-6736(20)30183-5; Nanshan Chen et al., "Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study," *Lancet (London, England)* 395, no. 10223 (February 15, 2020): 507–13, https://doi.org/10.1016/S0140-6736(20)30211-7; Mansoor N. Bangash, Jaimin Patel, and Dhruv Parekh, "COVID-19 and the Liver: Little Cause for Concern," *The Lancet. Gastroenterology & Hepatology* 5, no. 6 (June 2020): 529–30, https://doi.org/10.1016/S2468-1253(20)30084-4; Yijin Wang et al., "SARS-CoV-2 Infection of the Liver Directly Contributes to Hepatic Impairment in Patients with COVID-19," *Journal of Hepatology* 73, no. 4 (October 2020): 807–16, https://doi.org/10.1016/j.jhep.2020.05.002.

[246] Chen et al., "Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China"; Bangash, Patel, and Parekh, "COVID-19 and the Liver."

[247] Wang et al., "SARS-CoV-2 Infection of the Liver Directly Contributes to Hepatic Impairment in Patients with COVID-19."

[248] Kruglikov and Scherer, "The Role of Adipocytes and Adipocyte-Like Cells in the Severity of COVID-19 Infections"; Aaron J. Wilk et al., "Multi-Omic Profiling Reveals Widespread Dysregulation of Innate Immunity and Hematopoiesis in COVID-19," *The Journal of Experimental Medicine* 218, no. 8 (August 2, 2021): e20210582, https://doi.org/10.1084/jem.20210582; Martin Zickler et al., "Replication of SARS-CoV-2 in Adipose Tissue Determines Organ and Systemic Lipid Metabolism in Hamsters and Humans," *Cell Metabolism* 34, no. 1 (January 4, 2022): 1–2, https://doi.org/10.1016/j.cmet.2021.12.002; A. Basolo et al., "Adipose Tissue in COVID-19: Detection of SARS-CoV-2 in Adipocytes and Activation of the Interferon-

Chapter 3: Diving into the proteomic atlas of SARS-CoV-2 infected cells: insights of viral infection in different cell types

129

system, obesity, and high quantities of adipose tissue are risk factors and indicators of poor prognosis for severe COVID-19 and its evolution[249], we evaluated infectomes of primary stromal-vascular cells isolated from subcutaneous (abdomen) or visceral (omentum) adipose tissues. Subcutaneous and visceral adipose tissue cells presented different responses to infection, with only 44 proteins in common. A pathway enrichment analysis highlighted changes in pyruvate metabolism, carbon metabolism, glycolysis/gluconeogenesis, and the TCA cycle in both cell types.

These dysregulations could be associated with the hyperglycemia that is often reported in acute COVID-19 patients and other metabolic dysfunctions that is seen in this disease[250]. Hospitalized COVID-19 patients with hyperglycemia also presented a higher prevalence of insulin resistance when compared to ARDS (acute respiratory distress syndrome)-positive and ARDS-negative controls, due to changes in adipokine levels (adiponectin and leptin), instead of beta-cell function[251]. This was also observed in other studies that analyzed adipokine levels in infected hamsters[252]. The energy metabolism dysfunctions seen in both adipocyte infectomes could be disrupting the proper functioning of these cells, inducing changes in adipokine secretion. When epithelial intestinal cells (CACO-2), hepatocytes (HepG2), and both subcutaneous and visceral adipose tissue were compared, only MDH2 (malate dehydrogenase 2) was found in all infectomes, being upregulated in CACO-2 cells, and downregulated in all other cell types (Supplementary Figure 5).

The immune response that is mounted upon SARS-CoV-2 infection is against the spike and nucleocapsid viral components and seeks to eliminate the virus from the

Alpha Response," *Journal of Endocrinological Investigation* 45, no. 5 (May 2022): 1021–29, https://doi.org/10.1007/s40618-022-01742-5.

[249] Norbert Stefan et al., "Obesity and Impaired Metabolic Health in Patients with COVID-19," *Nature Reviews. Endocrinology* 16, no. 7 (July 2020): 341–42, https://doi.org/10.1038/s41574-020-0364-6.

[250] Moritz Reiterer et al., "Hyperglycemia in Acute COVID-19 Is Characterized by Insulin Resistance and Adipose Tissue Infectivity by SARS-CoV-2," *Cell Metabolism* 33, no. 11 (November 2, 2021): 2174-2188.e5, https://doi.org/10.1016/j.cmet.2021.09.009.

[251] Reiterer et al.

[252] Jasper Fuk-Woo Chan et al., "Simulation of the Clinical and Pathological Manifestations of Coronavirus Disease 2019 (COVID-19) in a Golden Syrian Hamster Model: Implications for Disease Pathogenesis and Transmissibility," *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 71, no. 9 (December 3, 2020): 2428–46, https://doi.org/10.1093/cid/ciaa325; Sin Fun Sia et al., "Pathogenesis and Transmission of SARS-CoV-2 in Golden Hamsters," *Nature* 583, no. 7818 (July 2020): 834–38, https://doi.org/10.1038/s41586-020-2342-5.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

130

host[253]. Lower levels of neutralizing antibodies in circulation and a progression to severe COVID-19 are strongly associated with an exacerbated inflammatory response, alterations in the immune system response characterized by protein synthesis reduction, a "cytokine storm", lymphocytopenia, and T-cell exhaustion[254]. Immune responses occur through two different mechanisms: the innate immune response, mediated by macrophages, dendritic cells, and natural killer cells, and/or the adaptive immune response, mediated by T-cells and B-cells.

To cover as many of the proteomic changes brought about by SARS-CoV-2 infection as possible, we used human primary T-cells and monocytes as infection models, thereby screening both innate and adaptive responses[255]. The resulting pathway analyses showed a high similarity between the T-cells and monocytes, enriching for several pathways related to response to infection, phagocytosis, neutrophil cellular trap formation, and, once again, energy metabolism. Ribosomal changes and coronavirus disease were identified only in monocytes, likely since monocytes are responsible for the activation of macrophages and produce the initial immune response. In agreement with this hypothesis, chemokine signaling activation and leukocyte trans-endothelial migration were both enriched exclusively in monocytes. The T-cell infectome also exclusively enriched for certain biological pathways, which included apoptotic and necroptotic pathways, supporting a study that saw a decrease in CD4$^+$ T-cell viability after SARS-CoV-2 infection[256]. As SARS-CoV-2 replication is dependent on the presence of glucose, this highlights the many pathways involved in glucose metabolism that were enriched in both cell types. Moreover, this suggests that energy production

---

[253] Xiaonan Zhang et al., "Viral and Host Factors Related to the Clinical Outcome of COVID-19," *Nature* 583, no. 7816 (July 2020): 437–40, https://doi.org/10.1038/s41586-020-2355-0.

[254] Prabhu S. Arunachalam et al., "Systems Biological Assessment of Immunity to Mild versus Severe COVID-19 Infection in Humans," *Science (New York, N.Y.)* 369, no. 6508 (September 4, 2020): 1210–20, https://doi.org/10.1126/science.abc6261; Davide F. Robbiani et al., "Convergent Antibody Responses to SARS-CoV-2 in Convalescent Individuals," *Nature* 584, no. 7821 (August 2020): 437–42, https://doi.org/10.1038/s41586-020-2456-9.

[255] Nie et al., "Multi-Organ Proteomic Landscape of COVID-19 Autopsies"; Brunetti et al., "SARS-CoV-2 Uses CD4 to Infect T Helper Lymphocytes"; Codo et al., "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis."

[256] Brunetti et al., "SARS-CoV-2 Uses CD4 to Infect T Helper Lymphocytes."

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

131

dysfunctions could be directly related to an exacerbated immune response and resulting cytokine storm[257].

When all cell types were compared together, a large overlap was observed in identified proteins; however, when only dysregulated proteins were considered, this overlap vanished (Supplementary Figure 3). Nie *et al.* observed a similar trend in their proteomic analysis of seven tissue types from patients who died from COVID-19[258]. They showed that only 39 proteins were commonly dysregulated of a total of 11394 identified proteins, representing less than 0.004% of identifications. Despite the lack of overlap of individual proteins, energy metabolism was affected in every infectome (Supplementary Figure 6), which strongly supports the hypothesis that glucose metabolism is a key factor for SARS-CoV-2 replication inside the host cell. This hypothesis is further supported by several studies that have reported disturbances in energy production in tissues and cells infected by SARS-CoV-2[259].

To facilitate the future investigation of all the data collected and analyzed in this work, we have constructed an open-source tool to facilitate the search for proteins found to be dysregulated in each cell type and tissue type. Other data include their relative abundance in comparison to controls and the biological pathways they enrich, all of which is freely available at https://reisdeoliveira.shinyapps.io/Infectome_App/. Future proteomic analyses can then be submitted to this database to expand this SARS-CoV-2 Infectome Atlas.

4.10 DATA SHARING STATEMENT

The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[60,61] with dataset identifiers 10·6019/PXD023781, PXD030910, PXD0269521, PXD020967. The unique primary cells and biopsies used in this study can only be shared upon approval by the local ethics committee and donor consent.

---

[257] Codo et al., "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis."

[258] Nie et al., "Multi-Organ Proteomic Landscape of COVID-19 Autopsies."

[259] Nie et al.; Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients"; Codo et al., "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis."

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

132

4.11 METHODS

4.11.1 Generation of human astrocytes (hES-derived)

Differentiation of glial progenitor cells was performed from neural stem cells (NSC) derived from pluripotent human embryonic stem cells (hES, BR-1 cell line, RRID:CVCL_C062),[260] according to the method published by Trindade *et al.*[261] NSCs were cultured on plates coated with Geltrex Matrix (Thermo Fisher Scientific, MA, USA) using 1:1 Neurobasal™:advanced DMEM/F12 medium and 2% neural induction supplement. Upon reaching 50% confluence, the medium was changed to DMEM/F12 (Dulbecco's modified Eagle medium/F12), 1% N2 supplement, 1% fetal bovine serum (FBS), and 1% penicillin-streptomycin and cells were maintained at 37°C in humidified air with 5% atmospheric CO_2 for 21 days. At this stage, cells were considered glial progenitor cells (GPC).

Subsequently, GPCs were plated at low density (30-40% confluence) on Geltrex-coated plates with DMEM/F12 medium, 1% GlutaMAX Supplement, 10% FBS and 1% penicillin-streptomycin. The differentiation medium was replaced every 2-3 days. After 4 weeks of differentiation, the cells were considered mature astrocytes. These cells were plated on Geltrex-coated coverslips at a density of $4x10^4$ cells for immunostaining assays (24-well plates) or $2 \cdot 5x10^5$ cells for viral load, proteomic, and metabolomic analysis (6-well plates).

All products used for cell culture are from Thermo Fisher Scientific, MA, USA. The characterization of the BR-1 lineage as astrocyte cells has been previously described elsewhere[262]. We generated eight batches of human astrocytes from BR-1-derived NSCs.

---

[260] Ana M. Fraga et al., "Establishment of a Brazilian Line of Human Embryonic Stem Cells in Defined Medium: Implications for Cell Therapy in an Ethnically Diverse Population," *Cell Transplantation* 20, no. 3 (2011): 431–40, https://doi.org/10.3727/096368910X522261.

[261] Pablo Trindade et al., "Short and Long TNF-Alpha Exposure Recapitulates Canonical Astrogliosis Events in Human-Induced Pluripotent Stem Cells-Derived Astrocytes," *Glia* 68, no. 7 (July 2020): 1396–1409, https://doi.org/10.1002/glia.23786.

[262] Fraga et al., "Establishment of a Brazilian Line of Human Embryonic Stem Cells in Defined Medium"; Trindade et al., "Short and Long TNF-Alpha Exposure Recapitulates Canonical Astrogliosis Events in Human-Induced Pluripotent Stem Cells-Derived Astrocytes"; Pítia Flores Ledur et al., "Zika Virus Infection Leads to Mitochondrial Failure, Oxidative Stress and DNA Damage in Human iPSC-Derived Astrocytes," *Scientific Reports* 10, no. 1 (January 27, 2020): 1218, https://doi.org/10.1038/s41598-020-57914-x; Yiping Yan et al., "Efficient and Rapid Derivation of Primitive Neural Stem Cells and Generation of Brain Subtype Neurons from Human Pluripotent Stem Cells," *Stem Cells Translational Medicine* 2, no. 11 (November 2013): 862–70, https://doi.org/10.5966/sctm.2013-0080.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

133

The NSCs were of different passages and were used as biological replicates in independent experiments. Our internal control showed that approximately 97% of the neural stem cell-derived astrocytes in culture expressed GFAP, 80·4% expressed vimentin, and 12·9% expressed SOX-2 (markers of progenitor cells). The neural stem cell-derived astrocyte culture expressed more astrocytic markers than progenitor cell markers, indicating an excellent effectiveness of human astrocyte generation. Method described in our previous work[263].

4.11.2 Differentiation of the SH-SY5Y human neuroblastoma cell line

The SH-SY5Y cell line (SH-SY5Y ATCC-CRL-2266, RRID:CVCL_0019), kindly donated by Prof. Dr. Gustavo J. S. Pereira (Federal University of São Paulo, UNIFESP), was cultivated using a previously documented neuronal differentiation protocol[264] using DMEM/F12 medium, 10% FBS and 1% penicillin-streptomycin at 37°C in humidified air with 5% atmospheric CO_2. The SH-SY5Y cells were plated and, upon reaching 25-30% confluency, the medium was changed to neuronal differentiation medium consisting of DMEM/F12 with 1% FBS and 10 µM retinoic acid (Sigma Aldrich). The differentiation medium was replaced every 2-3 days over the course of 2 weeks. These differentiated SH-SY5Y cells are known to more closely relate to adrenergic neurons, but they also express dopaminergic markers[265].

4.11.3 NSC differentiation into neurons

Human NSC-derived neurons were cultivated following the protocol described by Thermo Fisher Scientific[266]. NSCs were plated on Geltrex-coated plates and

---

[263] Crunfli et al., "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients."

[264] Jane Kovalevich and Dianne Langford, "Considerations for the Use of SH-SY5Y Neuroblastoma Cells in Neurobiology," *Methods in Molecular Biology (Clifton, N.J.)* 1078 (2013): 9–21, https://doi.org/10.1007/978-1-62703-640-5_2; Mackenzie M. Shipley, Colleen A. Mangold, and Moriah L. Szpara, "Differentiation of the SH-SY5Y Human Neuroblastoma Cell Line," *Journal of Visualized Experiments: JoVE*, no. 108 (February 17, 2016): 53193, https://doi.org/10.3791/53193; Helena Xicoy, Bé Wieringa, and Gerard J. M. Martens, "The SH-SY5Y Cell Line in Parkinson's Disease Research: A Systematic Review," *Molecular Neurodegeneration* 12, no. 1 (January 24, 2017): 10, https://doi.org/10.1186/s13024-017-0149-0.

[265] Kovalevich and Langford, "Considerations for the Use of SH-SY5Y Neuroblastoma Cells in Neurobiology."

[266] G. J. Brewer, "Serum-Free B27/Neurobasal Medium Supports Differentiated Growth of Neurons from the Striatum, Substantia Nigra, Septum, Cerebral Cortex, Cerebellum, and Dentate Gyrus," *Journal of Neuroscience Research* 42, no. 5 (December 1995): 674–83, https://doi.org/10.1002/jnr.490420510; Y.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

134

maintained with NEM medium at 37°C in humidified air with 5% atmospheric CO_2. Upon reaching 40% confluency, the medium was changed to neuronal differentiation medium consisting of 1:1 DMEM/F12:neurobasal medium with 1% B27 supplement (Thermo Fisher Scientific, Carlsbad, CA, USA) and 1% GlutaMAX (Thermo Fisher Scientific, Carlsbad, CA, USA). The medium was renewed every four days over the course of 20 days by removing half of the volume and replacing the same volume with fresh medium. Medium renewal was performed in this manner since factors secreted by the differentiating cells are important for successful differentiation. Two control cell lines were used: GM23279A, obtained from a female subject (available at Coriell; RRID:CVCL_F178) and BR-1 (RRID:CVCL_C062)[267]. Both cell lines were cultivated following previously published protocols[268] and have been extensively characterized elsewhere[269]. We also used FACS to analyze cellular markers and found a *bona fide* neuronal phenotype with the expression of the neuronal markers synaptophysin (75·9% of cells), MAP2 (99·9%), and β-tubulin (99%), as well as astrocytic marker GFAP (8·1%).

4.11.4 Human adipose tissue mesenchymal stem cell isolation, culture, and adipocyte differentiation

Patients, samples collection and adipocyte differentiation were previous described in our work[270]. For this work it was used the same samples but with

---

Elkabetz and L. Studer, "Human ESC-Derived Neural Rosettes and Neural Stem Cell Progression," *Cold Spring Harbor Symposia on Quantitative Biology* 73 (2008): 377–87, https://doi.org/10.1101/sqb.2008.73.052; Cleber A. Trujillo et al., "Novel Perspectives of Neural Stem Cell Differentiation: From Neurotransmitters to Therapeutics," *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* 75, no. 1 (January 2009): 38–53, https://doi.org/10.1002/cyto.a.20666.

[267] Fraga et al., "Establishment of a Brazilian Line of Human Embryonic Stem Cells in Defined Medium."

[268] Ledur et al., "Zika Virus Infection Leads to Mitochondrial Failure, Oxidative Stress and DNA Damage in Human iPSC-Derived Astrocytes"; Bárbara S. Casas et al., "hiPSC-Derived Neural Stem Cells from Patients with Schizophrenia Induce an Impaired Angiogenesis," *Translational Psychiatry* 8, no. 1 (February 22, 2018): 48, https://doi.org/10.1038/s41398-018-0095-9; Livia Goto-Silva et al., "Computational Fluid Dynamic Analysis of Physical Forces Playing a Role in Brain Organoid Cultures in Two Different Multiplex Platforms," *BMC Developmental Biology* 19, no. 1 (March 7, 2019): 3, https://doi.org/10.1186/s12861-019-0183-y; Joseph A. White et al., "Excess Rab4 Rescues Synaptic and Behavioral Dysfunction Caused by Defective HTT-Rab4 Axonal Transport in Huntington's Disease," *Acta Neuropathologica Communications* 8, no. 1 (July 1, 2020): 97, https://doi.org/10.1186/s40478-020-00964-z.

[269] Brewer, "Serum-Free B27/Neurobasal Medium Supports Differentiated Growth of Neurons from the Striatum, Substantia Nigra, Septum, Cerebral Cortex, Cerebellum, and Dentate Gyrus"; Elkabetz and Studer, "Human ESC-Derived Neural Rosettes and Neural Stem Cell Progression"; Trujillo et al., "Novel Perspectives of Neural Stem Cell Differentiation."

[270] Saccon et al., "SARS-CoV-2 Infects Adipose Tissue in a Fat Depot- and Viral Lineage-Dependent Manner."

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

135

independent runs in the LC-MS/MS system. Briefly, human adipose tissue-derived
mesenchymal stem cells (hADSCs) were isolated from abdominal subcutaneous adipose
tissue and visceral omental adipose tissue of three individuals who underwent
abdominal surgery (i.e., bariatric surgery or cholecystectomy) at the Clinical Hospital
of the University of Campinas, Campinas, SP, Brazil. These cells were isolated prior to
the COVID-19 pandemic and hence the donors were not infected with SARS-CoV-2. The
subjects received written and oral information before providing written informed
consent for the collection of the biopsy and use of the tissues. All material used in the
procedure was sterile. Biopsies were collected during surgery and transported to the
laboratory within 30 minutes in sealed, sterile falcon tubes for initiation of the
procedure of stromal-vascular fraction isolation. The tissue was weighed, cut into small
pieces on a petri dish, and digested with 25-30 mL of lysis buffer (1 mg/mL collagenase
in HBSS containing 2% BSA, filtered through a 0·22 μm filter) at 37°C for 30-50 min with
slight agitation, until homogeneous. The homogenate was filtered through a 250 μm
filter and collected in a sterile falcon tube. After a brief rest period (approximately 5
minutes), the infranatant containing the stromal-vascular fraction was collected using
a Pasteur pipette and centrifuged for 5 min at 200 x g and 4°C. The supernatant was
discarded, and the cell pellet was washed with HBSS. The centrifugation and washing
steps were repeated twice. The pellet was then resuspended in BM-1 medium (ZenBio)
with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (penicillin 10,000
units/mL, streptomycin 10 mg/mL).

Cells were cultured at 37°C and 5% atmospheric CO_2 until semi-confluent (80-
90%). They were then trypsinized and frozen in freezing medium (10% DMSO, 50% FBS,
and 40% basal culture medium) and stored in a liquid nitrogen biorepository for further
assays (e.g., adipocyte differentiation followed by exposure with agents, lipid
quantification, determination of cell viability, gene expression analysis, proteomics,
immunodetection of proteins, quantification of cell secretion products, and metabolic
analysis).

hADSCs were thawed in BM-1 medium supplemented with 10% FBS, 1%
penicillin/streptomycin, 17 ng/mL bFGF, and 15 ng/mL BMP4, and cultured at 37°C and
5% atmospheric CO_2. Cells were expanded (i.e., split 1-3 times) and then seeded onto 24-

Chapter 3: Diving into the proteomic atlas of SARS-CoV-2 infected cells: insights of viral infection in different cell types

136

well plates at a density of 40,000 cells/cm^2 for adipocyte differentiation. When cells reached 100% confluency (day 0), the medium was replaced with BM-1 supplemented with 3% FBS, 1% penicillin/streptomycin, 0·1 µM dexamethasone, 500 µM 3-isobutyl-1 methylxanthine (IBMX), 20 nM insulin, 5 nM triiodothyronine (T3) and 10 ng/mL BMP4. Cells were cultured for 7 days, and the medium was changed every 2-3 days. On day 7, the medium was replaced with BM-1 supplemented with 3% FBS, 1% penicillin/streptomycin, 0·1 µM dexamethasone, 20 nM insulin, 5 nM T3 and 10 ng/mL BMP4 until the cells were fully differentiated (day 10). Cells were then cultured for an additional 3 days with BM-1 supplemented with 3% FBS and 1% penicillin/streptomycin before being subjected to the assays described below.

4.11.5 Virus strain

The HIAE-02-SARS-CoV-2/SP02/human/2020/BRA (GenBank accession number MT126808·1) virus strain was used for all *in vitro* experiments. The virus was isolated from the first confirmed case of COVID-19 in Brazil and kindly donated by Prof. Dr. Edison Durigon (ICB-USP). VSV-eGFP-SARS-CoV-2 was engineered and donated by Prof. Dr. Sean P.J. Whelan (Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA) for SARS-CoV-2 entry experiments[271]. Viral stock was propagated in Vero CCL-81 cells (ATCC, RRID:CVCL_0059), cultivated in DMEM supplemented with 10% heat-inactivated FBS and 1% penicillin and streptomycin (Gibco, Walthmam, MA, USA), and incubated at 37°C with 5% atmospheric CO$_2$. Viral titer was determined by the plaque-forming assay using Vero cells. The Brazilian ZIKV strain (BeH823339, 589; GenBank accession KU729217) was used for proteomic experiments.

4.11.6 Human lymphocyte isolation and mixed lymphocyte reaction

PBMCs isolated from buffy coats from healthy volunteers were incubated at 37°C and 5% atmospheric CO$_2$ for 2 h to allow monocyte adherence to the plate surface. Non-adherent cells in suspension were collected and stained with carboxyfluorescein succinimidyl ester (Cell-Trace CFSE). Cells were incubated with CFSE diluted in pre-warmed PBS to the desired concentrations at 37°C for 15 min, resuspended in fresh, pre-

---

[271] James Brett Case et al., "Neutralizing Antibody and Soluble ACE2 Inhibition of a Replication-Competent VSV-SARS-CoV-2 and a Clinical Isolate of SARS-CoV-2," *Cell Host & Microbe* 28, no. 3 (September 9, 2020): 475-485.e5, https://doi.org/10.1016/j.chom.2020.06.021.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

137

warmed medium for 30 min, and then washed with PBS, according to the
manufacturer's recommendations. Is this work it was reanalalyzed the proteomic data
obtained by our previous work. Patients steatments and ethical issues are also described
at Codo *et al*[272].

4.11.7   *Human colon cancer cell line (CACO-2) culture*

Human colon cancer cell line CACO-2 (RRID:CVCL_0025) was obtained from
the American Type Culture Collection (ATCC). Cells were maintained in Dulbecco's
modified Eagle medium (Gibco) supplemented with 20% fetal bovine serum (FBS) and
1% penicillin-streptomycin at 37°C and 5% atmospheric CO_2. CACO-2 cells were
exposed to SARS-CoV-2 for 1 h at an MOI of 0·1 (P3 21/05/20 - 10^4 UFF/mL) under gentle
agitation at room temperature. After viral adsorption, cells were washed twice with
phosphate-buffered saline (PBS) and incubated with DMEM supplemented with 20% FBS
and 1% penicillin and streptomycin for 24 h at 37°C and 5% atmospheric CO_2.

4.11.8   *Human liver carcinoma cell line (HepG2) culture*

HepG2 cells (human liver carcinoma cell line HepG2/CD81++) were cultured
for 3 days in Dulbecco's modified Eagle medium (DMEM; Vitrocell, Campinas, SP,
Brazil), enriched with 10% (v/v) fetal bovine serum, under a humidified conditions with
5% atmospheric CO2 at 37°C. After reaching 80% confluence, the cells underwent viral
infection of 0·1 MOI, prior to preparation for proteomic analyses, as described below.

4.11.9   *Blood sample collection and lymphocyte separation*

Each COVID-19 patient had heparin and plain blood tubes collected. Whole
blood, serum, and plasma samples were separated. Peripheral blood mononuclear cells
(PBMCs) from patients and buffy coats were obtained via Histopaque-1077 density
gradient (Sigma-Aldrich). Samples were diluted in Hanks balanced salt solution (1:1)
and gently poured into 15- or 50-mL conical tubes containing 3 or 10 mL of Histopaque,
respectively. Samples were centrifuged at 1400 rpm for 30 min at 4°C without
acceleration or braking. After the PBMC layer was transferred to new tube, lymphocytes
were sorted and incubated overnight with RPMI 1640 (Gibco) containing 10% fetal

---

272 Codo et al., "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a
HIF-1α/Glycolysis-Dependent Axis."

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

138

bovine serum (FBS) and 1% penicillin-streptomycin (P/S), at 37°C with 5% atmospheric CO_2. Is this work it was reanalysed the proteomic data obtained by our previous work. Patients statements and ethical issues are also described at Davanzo GG, *et al* 2020[12]

4.11.10 In vitro astrocyte infection

Astrocytes were infected with SARS-CoV-2 for 1 h (MOI of 0·1 for proteomic, viral load, and bioenergetic assays) under gentle agitation at room temperature. After viral adsorption, cells were washed twice with phosphate-buffered saline (PBS) and incubated with DMEM/F12 supplemented with 10% FBS, 1% GlutaMAX, and 1% penicillin and streptomycin for 24 h at 37°C and 5% atmospheric CO2.

4.11.11 LC-MS/MS sample preparation, analyses, and data processing

All cellular experiments underwent the same preparation process, as follows. Cells were lysed chemically with lysis buffer (100 mM Tris-HCl, 1 mM EDTA, 150 mM NaCl, 1% Triton-X, and protease and phosphatase inhibitors) then mechanically with an ultrasonication probe during three cycles of 20 s with 90% frequency, on ice. The total protein extract was quantified by Pierce BCA protein assay (Thermo Fisher Scientific, MA, USA) according to the manufacturer's instructions. 30 µg of total protein extract from each sample was transferred to a Microcon-10 centrifugal filter with a 10 kDa cutoff for FASP protein digestion[273]. In brief, proteins were reduced (10 mM dithiothreitol) at 56°C for 40 min, alkylated (50 mM iodoacetamide) at room temperature for 20 min in the dark, and digested overnight with trypsin at 37°C in 50 mM ammonium bicarbonate (AmBic), pH 8·0. Peptides were recovered from the filter in 50 mM AmBic and trypsin activity was quenched by adding formic acid (FA) to a final concentration of 1% (v/v), whereupon the peptides were desiccated in a SpeedVac vacuum concentrator and stored at -80°C until analysis.

Digested peptides were resuspended in 0·1% FA. LC-MS/MS analyses were performed in an ACQUITY UPLC M-Class System (Waters Corporation, Milford, MA) coupled online to a Synapt G2-Si mass spectrometer (Waters Corporation, Milford, MA). 1 µg of peptides were loaded onto a trapping column (Symmetry C18 5 µm, 180 µm × 20

---

[273] Ute Distler et al., "Label-Free Quantification in Ion Mobility-Enhanced Data-Independent Acquisition Proteomics," *Nature Protocols* 11, no. 4 (April 2016): 795–812, https://doi.org/10.1038/nprot.2016.042.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

139

mm, Waters Corporation, Milford, MA) and subsequently separated in the analytical column (HSS T3 C18 1·8 μm, 75 μm × 150 mm; Waters Corporation, Milford, MA). For gradient elution, 0·1% FA was used as eluent A and acetonitrile-FA (99·9% ACN:0·1% FA) as eluent B. A reversed phase gradient was carried out over a 120-minute method, with a linear gradient running from 3 - 60% eluent B over 90 min at 300 nL/min. In the Synapt G2-Si, the peptide spectra were acquired by ion mobility-enhanced data-independent acquisition (HDMS^E). Mass spectrometry analysis was performed in Resolution Mode, switching between low (4 eV) and high (25 to 60 eV) collision energies, using a scan time of 1·0 s per function over 50 to 2000 m/z. The wave velocity for ion mobility separation was 1000 m/s and the transfer wave velocity was 175 m/s. A human [Glu1]-Fibrinopeptide B standard (Waters Corporation, Milford, MA) was used as the reference lock-mass compound. Each sample was acquired in at least duplicate.

The raw data from each experiment were processed in Progenesis QI for Proteomics, version 4·0·x (Waters Corporation, Milford, MA). Tandem mass spectra were searched against a reviewed *Homo sapiens* proteome (UNIPROT, release 2020-04), using tolerance parameters of 20 ppm for precursor ions and 10 ppm for product ions. For peptide identification, carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine as a variable modification, up to two missed cleavages were permitted and the false discovery rate (FDR) was limited to 1%. Protein identification was performed using a minimum of one fragment ion per peptide, three fragment ions per protein, and one peptide per protein. Sample acquisition until this point was performed by experimenter blind to group assignment and outcome assessment. Analyses after this point were performed with predefined algorithms and cutoffs, reducing the possibility of experimenter-induced bias.

The label-free quantitative analysis was carried out using the relative abundance intensity of all peptides of a protein with at least one unique peptide after normalisation by all peptide intensities. The expression analysis was performed considering the technical replicates for each experimental condition, following the hypothesis that each group is independent. Proteins with ANOVA (p) ≤ 0·05 between groups were considered differentially regulated. Data integration and visualization was

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

140

carried out in OmicScope package[62] (v.1.2.2), using Progenesis Method and default parameters.

4.11.12 RNA extraction and viral load

Total RNA extraction was performed using TRI Reagent (Sigma, St Louis, USA) according to the manufacturer's instructions. RNA concentration was determined by a DeNovix spectrophotometer and RNA integrity was assessed by visualisation of 28S and 18S ribosomal RNA on a 1% agarose gel. Reverse transcription was performed with 0·5 µg of RNA using a GoScript reverse transcriptase kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. qPCR was performed using astrocyte cDNA diluted 1:10 and the qPCR SybrGreen Supermix (Qiagen, Valencia, CA, USA) containing forward and reverse primers in RNAse-free water. All reactions were performed in a CFX384 Touch real-time PCR detection system (Biorad, Hercules, CA, USA) and cycling conditions were set as follows: 50°C for 2 min; 95°C for 10 min; (95°C for 15 s; 60°C for 1 min) x 40 cycles. To evaluate primer specificity, a melting curve analysis was performed by heating samples from 65 to 99°C (1°C increment changes at 5 s intervals). All sample measurements were performed in duplicate. Primers were designed with PrimerBlast and used at a concentration of 200 nM. Data were normalised to the expression of 18S (Fwd 5' CCCAACTTCTTAGAGGGACAAG 3'; Rev 5' CATCTAAGGGCATCACAGACC 3') and the relative quantification value of each target gene was determined using a comparative CT method[274]. For virus detection, SARS-CoV-2 nucleocapsid N1 primers (Fwd 5' CAATGCTGCAATCGTGCTAC 3'; Rev 5' GTTGCGACTACGTGATGAGG 3') were used as previously described[275]. Data were expressed as mean ± SEM. Statistical significance was calculated by a two-tailed unpaired Student's t-test. All analyses were performed using GraphPad Prism 8·0 (San Diego, CA, USA) and a significance level of $p \leq 0.05$ was adopted.

---

[274] Thomas D. Schmittgen and Kenneth J. Livak, "Analyzing Real-Time PCR Data by the Comparative C(T) Method," *Nature Protocols* 3, no. 6 (2008): 1101–8, https://doi.org/10.1038/nprot.2008.73.

[275] Codo et al., "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis"; Joungha Won et al., "Development of a Laboratory-Safe and Low-Cost Detection Protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19)," *Experimental Neurobiology* 29, no. 2 (April 30, 2020): 107–19, https://doi.org/10.5607/en20009.

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES
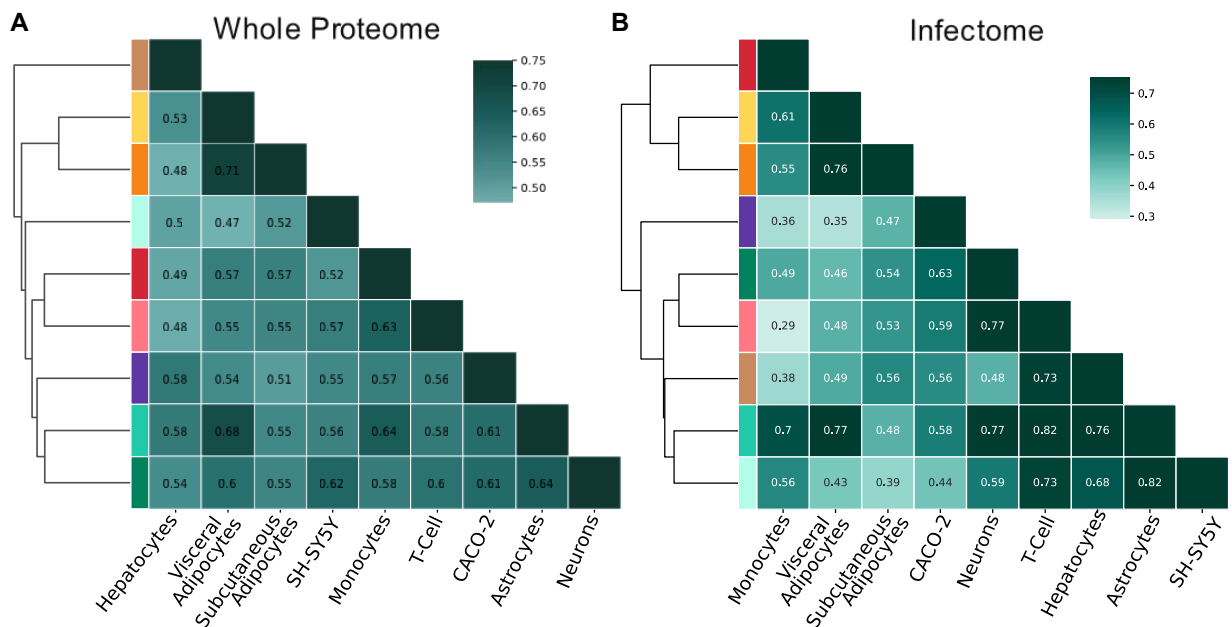
141

4.11.13 Role of the funding source

Funding sources had no influence on the study design; collection, analysis or interpretation of data; writing of the report; or decision to submit.
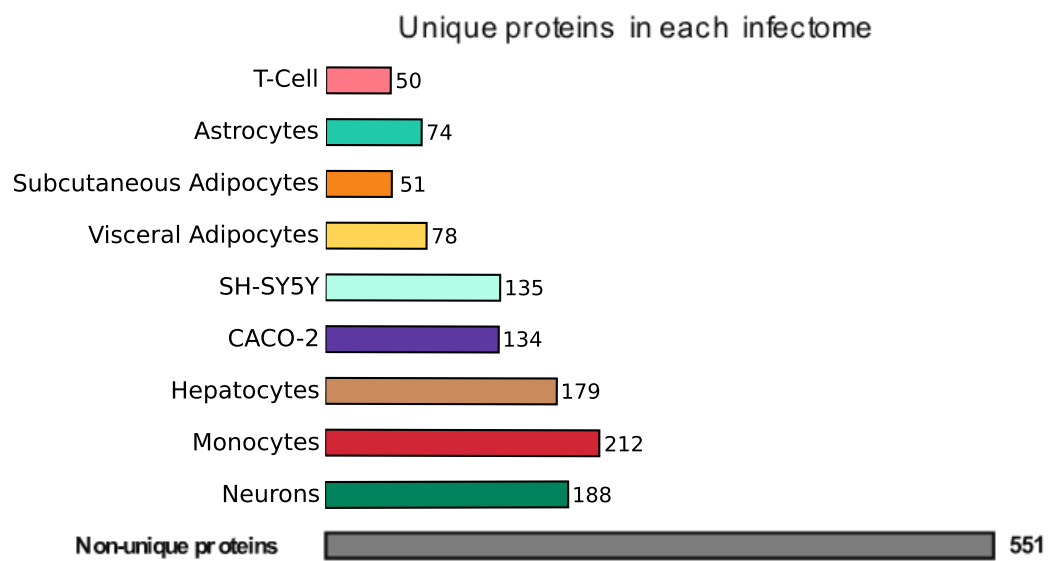
4.12 SUPPLEMENTARY INFORMATION
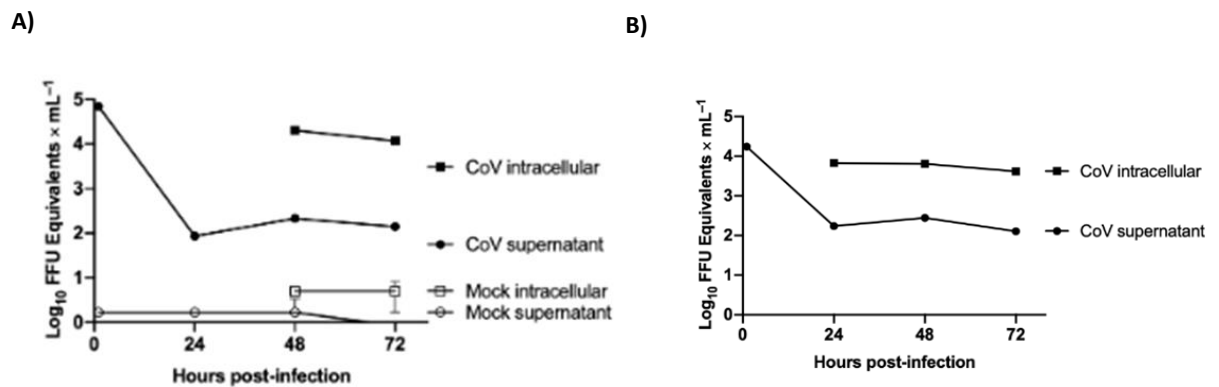
Supplementary Figure 4-1 A) Viral kinetic analysis in SH-SY5Y cell line. B) Viral kinetic analysis in NSC-derived neurons.

Supplementary Figure 4-2 Hierarchical clustering for pair-wise Pearson correlation among A) whole proteomes and B) infectomes.

Unique proteins in each infectome

T-Cell	50
Astrocytes	74
Subcutaneous Adipocytes	51
Visceral Adipocytes	78
SH-SY5Y	135
CACO-2	134
Hepatocytes	179
Monocytes	212
Neurons	188
Non-unique proteins	551

Supplementary Figure 4-3 Number of unique proteins differentially regulated in each infectome. Although no protein was found in all infectomes, 551 proteins were differentially regulated in more than one cell type.
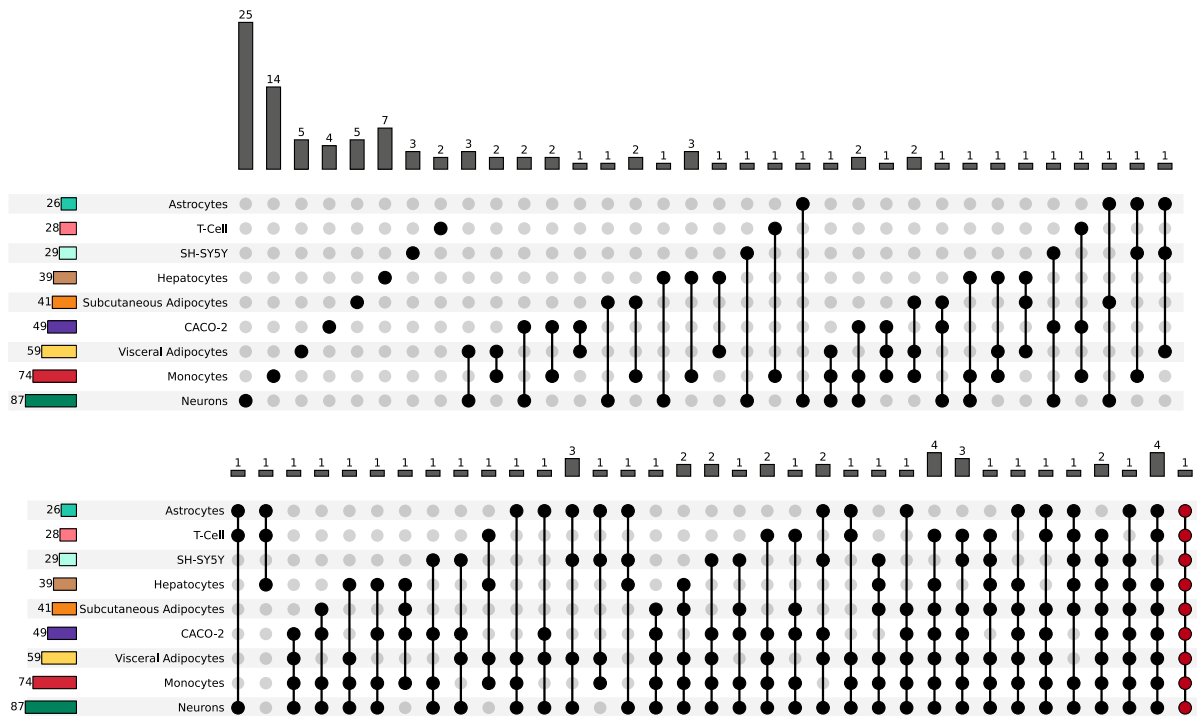
Supplementary Figure 4-4 Upset plot showing the number of overlapping pathways among infectomes. Vertical bars indicate the number of proteins found exclusively in groups marked by the respective connected dots. Horizontal bars indicate the total number of pathways found in each infectome.
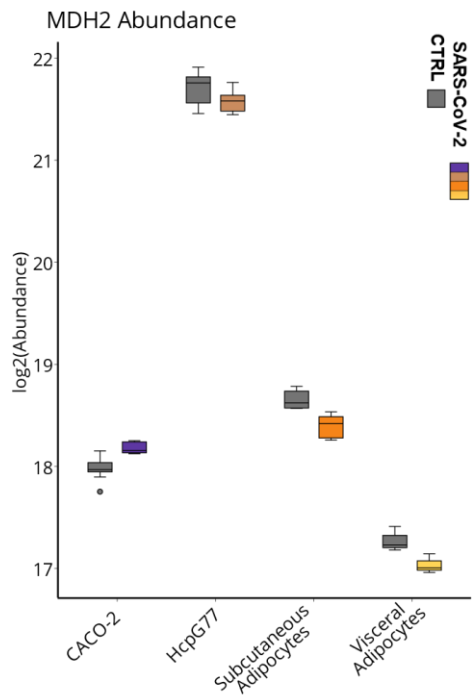
Supplementary Figure 4-5 Boxplot of MDH2 abundance for control (CTRL) and SARS-CoV-2-infected cells

CHAPTER 3: DIVING INTO THE PROTEOMIC ATLAS OF SARS-CoV-2 INFECTED
CELLS: INSIGHTS OF VIRAL INFECTION IN DIFFERENT CELL TYPES

144
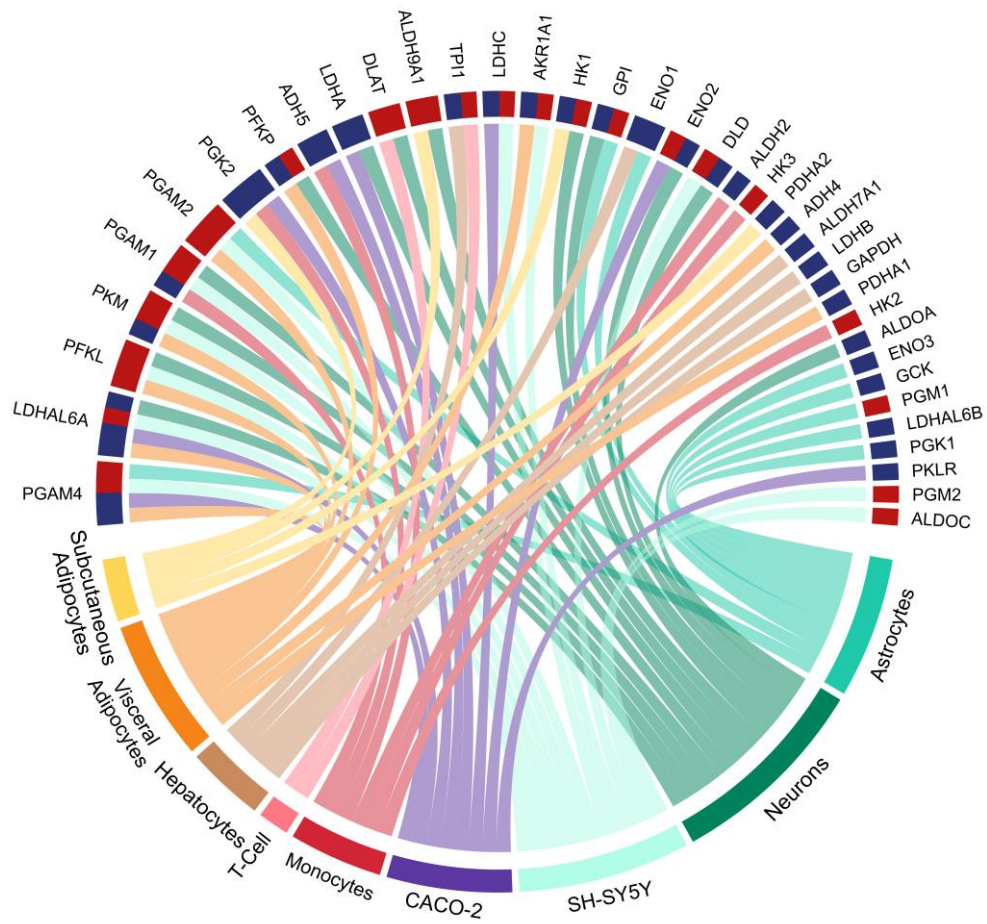
Supplementary Figure 4-6 Proteins found in all SARS-CoV-2 infectomes that are related to energy metabolism. Red indicates upregulation; blue indicates downregulation.

5   CONCLUSION

This thesis presented OmicScope, a novel and integrative package for proteomics data analysis, focusing on processes downstream to protein identification and quantitation. OmicScope is designed to interface with existing proteomics software and to bridge proteomics with other omics technologies and systems biology. Our tool is available as a Python package and a web application (https://omicscope.ib.unicamp.br) and enables differential proteomics analysis, enrichment analysis, and meta-analysis. OmicScope's integrative workflow allows users to perform differential proteomics analyses with complex experimental designs and to prompt link them with enrichment analyses, integrating quantitative and enrichment results. Moreover, the Nebula module compares independent studies using data from differential proteomics and enrichment analysis, even from third-party sources.

Critically, OmicScope has some limitations, such as restricted input format options, limited flexibility in some normalization and filtering steps, and reliance on the Enrichr API for enrichment analysis. However, these limitations are expected to be overcome with our ongoing efforts and our open-source approach. This approach also encourages community developers to add new features and improvements to the tool.

Our group, known for applying shotgun proteomics to reveal biochemical mechanisms related to brain disorders and SARS-CoV-2 infection, has incorporated OmicScope into our data analysis workflow. Chapters one and three of this thesis illustrate the benefits triggered by the development and usage of our computational tool. Additionally, our software is currently supporting several projects in our lab, helping users to efficiently analyse data and generate valuable biological insights. OmicScope was developed based on the experimental designs used by our group and has been applied to projects that assess proteome changes caused by drugs and diseases, including those involving time-course treatments or neurodevelopmental processes. The web application, which enables real-time discussion and brainstorming of results during lab meetings, has been instrumental in enhancing decision-making and guiding hypothesis validation. As a future outlook, we expect OmicScope to be adopted by the scientific community with similar acceptance as it has been in our group, contributing to the widespread use of proteomics and facilitating novel discoveries.

6   REFERENCES

Aebersold, Ruedi, and Matthias Mann. "Mass-Spectrometric Exploration of Proteome Structure and Function." *Nature* 537, no. 7620 (September 2016): 347–55. https://doi.org/10.1038/nature19949.

Akyol, Ömer, Hasan Herken, Efkan Uz, Ersin Fadıllıoğlu, Süheyla Ünal, Sadık Söğüt, Hüseyin Özyurt, and Haluk Asuman Savaş. "The Indices of Endogenous Oxidative and Antioxidative Processes in Plasma from Schizophrenic Patients." *Prog. Neuropsychopharmacol. Biol. Psychiatry* 26, no. 5 (June 2002): 995–1005. https://doi.org/10.1016/S0278-5846(02)00220-8.

Alquisiras-Burgos, Iván, Irlanda Peralta-Arrieta, Luis Antonio Alonso-Palomares, Ana Elvira Zacapala-Gómez, Eric Genaro Salmerón-Bárcenas, and Penélope Aguilera. "Neurological Complications Associated with the Blood-Brain Barrier Damage Induced by the Inflammatory Response During SARS-CoV-2 Infection." *Molecular Neurobiology* 58, no. 2 (February 2021): 520–35. https://doi.org/10.1007/s12035-020-02134-7.

Andreasen, Nancy C, and Ronald Pierson. "The Role of the Cerebellum in Schizophrenia." *Biol. Psychiatry* 64, no. 2 (July 15, 2008): 81–88. https://doi.org/10.1016/j.biopsych.2008.01.003.

Andrews, Jessica L, and Francesca Fernandez-Enright. "A Decade from Discovery to Therapy: Lingo-1, the Dark Horse in Neurological and Psychiatric Disorders." *Neurosci. Biobehav. Rev.* 56 (September 2015): 97–114. https://doi.org/10.1016/j.neubiorev.2015.06.009.

Arunachalam, Prabhu S., Florian Wimmers, Chris Ka Pun Mok, Ranawaka A. P. M. Perera, Madeleine Scott, Thomas Hagan, Natalia Sigal, et al. "Systems Biological Assessment of Immunity to Mild versus Severe COVID-19 Infection in Humans." *Science (New York, N.Y.)* 369, no. 6508 (September 4, 2020): 1210–20. https://doi.org/10.1126/science.abc6261.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25, no. 1 (May 2000): 25–29. https://doi.org/10.1038/75556.

Bader, Gary D., and Christopher WV Hogue. "An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks." *BMC Bioinformatics* 4, no. 1 (January 13, 2003): 2. https://doi.org/10.1186/1471-2105-4-2.

Bangash, Mansoor N., Jaimin Patel, and Dhruv Parekh. "COVID-19 and the Liver: Little Cause for Concern." *The Lancet. Gastroenterology & Hepatology* 5, no. 6 (June 2020): 529–30. https://doi.org/10.1016/S2468-1253(20)30084-4.

Bantscheff, Marcus, Dirk Eberhard, Yann Abraham, Sonja Bastuck, Markus Boesche, Scott Hobson, Toby Mathieson, et al. "Quantitative Chemical Proteomics Reveals Mechanisms of Action of Clinical ABL Kinase Inhibitors." *Nature Biotechnology* 25, no. 9 (September 2007): 1035–44. https://doi.org/10.1038/nbt1328.

Bantscheff, Marcus, Simone Lemeer, Mikhail M. Savitski, and Bernhard Kuster. "Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present." *Analytical and Bioanalytical Chemistry* 404, no. 4 (September 1, 2012): 939–65. https://doi.org/10.1007/s00216-012-6203-4.

Barabási, Albert-László, Natali Gulbahce, and Joseph Loscalzo. "Network Medicine: A Network-Based Approach to Human Disease." *Nature Reviews Genetics* 12, no. 1 (January 2011): 56–68. https://doi.org/10.1038/nrg2918.

Barabási, Albert-László, and Zoltán N. Oltvai. "Network Biology: Understanding the Cell's Functional Organization." *Nature Reviews Genetics* 5, no. 2 (February 2004): 101–13. https://doi.org/10.1038/nrg1272.

Bar-Joseph, Ziv, Anthony Gitter, and Itamar Simon. "Studying and Modelling Dynamic Biological Processes Using Time-Series Gene Expression Data." *Nature Reviews Genetics* 13, no. 8 (August 2012): 552–64. https://doi.org/10.1038/nrg3244.

Basolo, A., A. M. Poma, D. Bonuccelli, A. Proietti, E. Macerola, C. Ugolini, L. Torregrossa, et al. "Adipose Tissue in COVID-19: Detection of SARS-CoV-2 in Adipocytes and Activation of the Interferon-Alpha Response." *Journal of Endocrinological Investigation* 45, no. 5 (May 2022): 1021–29. https://doi.org/10.1007/s40618-022-01742-5.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: An Open Source Software for Exploring and Manipulating Networks." *Proceedings of the International AAAI Conference on Web and Social Media* 3, no. 1 (March 19, 2009): 361–62. https://doi.org/10.1609/icwsm.v3i1.13937.

Baumann, Oliver, Ronald J Borra, James M Bower, Kathleen E Cullen, Christophe Habas, Richard B Ivry, Maria Leggio, et al. "Consensus Paper: The Role of the Cerebellum in Perceptual Processes." *Cerebellum* 14, no. 2 (April 2015): 197–220. https://doi.org/10.1007/s12311-014-0627-7.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57, no. 1 (1995): 289–300.

Bernard, Jessica A, and Vijay A Mittal. "Cerebellar-Motor Dysfunction in Schizophrenia and Psychosis-Risk: The Importance of Regional Cerebellar Analysis Approaches." *Front. Psychiatry* 5 (November 25, 2014): 160. https://doi.org/10.3389/fpsyt.2014.00160.

Bernard, Jessica A, Joseph M Orr, and Vijay A Mittal. "Cerebello-Thalamo-Cortical Networks Predict Positive Symptom Progression in Individuals at Ultra-High Risk for Psychosis." *Neuroimage Clin* 14 (March 6, 2017): 622–28. https://doi.org/10.1016/j.nicl.2017.03.001.

Bian, Yangyang, Runsheng Zheng, Florian P. Bayer, Cassandra Wong, Yun-Chien Chang, Chen Meng, Daniel P. Zolg, et al. "Robust, Reproducible and Quantitative Analysis of Thousands of Proteomes by Micro-Flow LC–MS/MS." *Nature Communications* 11, no. 1 (January 9, 2020): 157. https://doi.org/10.1038/s41467-019-13973-x.

Bloomfield, Peter S, Sudhakar Selvaraj, Mattia Veronese, Gaia Rizzo, Alessandra Bertoldo, David R Owen, Michael Ap Bloomfield, et al. "Microglial Activity in People at Ultra High Risk of Psychosis and in Schizophrenia: An [(11)C]PBR28 PET Brain Imaging Study." *Am. J. Psychiatry* 173, no. 1 (January 2016): 44–52. https://doi.org/10.1176/appi.ajp.2015.14101358.

Bojkova, Denisa, Kevin Klann, Benjamin Koch, Marek Widera, David Krause, Sandra Ciesek, Jindrich Cinatl, and Christian Münch. "Proteomics of SARS-CoV-2-Infected Host Cells Reveals Therapy Targets." *Nature* 583, no. 7816 (July 2020): 469–72. https://doi.org/10.1038/s41586-020-2332-7.

Boyle, Elizabeth I, Shuai Weng, Jeremy Gollub, Heng Jin, David Botstein, J. Michael Cherry, and Gavin Sherlock. "GO::TermFinder—Open Source Software for

Accessing Gene Ontology Information and Finding Significantly Enriched Gene Ontology Terms Associated with a List of Genes." *Bioinformatics* 20, no. 18 (December 12, 2004): 3710–15. https://doi.org/10.1093/bioinformatics/bth456.

Brewer, G. J. "Serum-Free B27/Neurobasal Medium Supports Differentiated Growth of Neurons from the Striatum, Substantia Nigra, Septum, Cerebral Cortex, Cerebellum, and Dentate Gyrus." *Journal of Neuroscience Research* 42, no. 5 (December 1995): 674–83. https://doi.org/10.1002/jnr.490420510.

Brunetti, Natalia S., Gustavo G. Davanzo, Diogo de Moraes, Allan J. R. Ferrari, Gabriela F. Souza, Stéfanie Primon Muraro, Thiago L. Knittel, et al. "SARS-CoV-2 Uses CD4 to Infect T Helper Lymphocytes." *eLife* 12 (July 31, 2023): e84790. https://doi.org/10.7554/eLife.84790.

Buckner, Randy L. "The Brain's Default Network: Origins and Implications for the Study of Psychosis." *Dialogues Clin. Neurosci.* 15, no. 3 (September 2013): 351–58.

Calhoun, M E, M Jucker, L J Martin, G Thinakaran, D L Price, and P R Mouton. "Comparative Evaluation of Synaptophysin-Based Methods for Quantification of Synapses." *J. Neurocytol.* 25, no. 12 (December 1996): 821–28. https://doi.org/10.1007/bf02284844.

Cao, Hengyi, Oliver Y Chén, Yoonho Chung, Jennifer K Forsyth, Sarah C McEwen, Dylan G Gee, Carrie E Bearden, et al. "Cerebello-Thalamo-Cortical Hyperconnectivity as a State-Independent Functional Neural Signature for Psychosis Prediction and Characterization." *Nat. Commun.* 9, no. 1 (September 21, 2018): 3836. https://doi.org/10.1038/s41467-018-06350-7.

Casas, Bárbara S., Gabriela Vitória, Marcelo N. do Costa, Rodrigo Madeiro da Costa, Pablo Trindade, Renata Maciel, Nelson Navarrete, Stevens K. Rehen, and Verónica Palma. "hiPSC-Derived Neural Stem Cells from Patients with Schizophrenia Induce an Impaired Angiogenesis." *Translational Psychiatry* 8, no. 1 (February 22, 2018): 48. https://doi.org/10.1038/s41398-018-0095-9.

Case, James Brett, Paul W. Rothlauf, Rita E. Chen, Zhuoming Liu, Haiyan Zhao, Arthur S. Kim, Louis-Marie Bloyet, et al. "Neutralizing Antibody and Soluble ACE2 Inhibition of a Replication-Competent VSV-SARS-CoV-2 and a Clinical Isolate of SARS-CoV-2." *Cell Host & Microbe* 28, no. 3 (September 9, 2020): 475-485.e5. https://doi.org/10.1016/j.chom.2020.06.021.

Chan, Jasper Fuk-Woo, Anna Jinxia Zhang, Shuofeng Yuan, Vincent Kwok-Man Poon, Chris Chung-Sing Chan, Andrew Chak-Yiu Lee, Wan-Mui Chan, et al. "Simulation of the Clinical and Pathological Manifestations of Coronavirus Disease 2019 (COVID-19) in a Golden Syrian Hamster Model: Implications for Disease Pathogenesis and Transmissibility." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 71, no. 9 (December 3, 2020): 2428–46. https://doi.org/10.1093/cid/ciaa325.

Chan, M K, T M Tsang, L W Harris, P C Guest, E Holmes, and S Bahn. "Evidence for Disease and Antipsychotic Medication Effects in Post-Mortem Brain from Schizophrenia Patients." *Mol. Psychiatry* 16, no. 12 (December 2011): 1189–1202. https://doi.org/10.1038/mp.2010.100.

Chen, Chen, Jie Hou, John J. Tanner, and Jianlin Cheng. "Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis." *International Journal of Molecular Sciences* 21, no. 8 (April 20, 2020): 2873. https://doi.org/10.3390/ijms21082873.

Chen, Edward Y., Christopher M. Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R. Clark, and Avi Ma'ayan. "Enrichr: Interactive and

Collaborative HTML5 Gene List Enrichment Analysis Tool." *BMC Bioinformatics* 14 (April 15, 2013): 128. https://doi.org/10.1186/1471-2105-14-128.

Chen, Nanshan, Min Zhou, Xuan Dong, Jieming Qu, Fengyun Gong, Yang Han, Yang Qiu, et al. "Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study." *Lancet (London, England)* 395, no. 10223 (February 15, 2020): 507–13. https://doi.org/10.1016/S0140-6736(20)30211-7.

Clark, D, I Dedova, S Cordwell, and I Matsumoto. "A Proteome Analysis of the Anterior Cingulate Cortex Gray Matter in Schizophrenia." *Mol. Psychiatry* 11, no. 5 (May 2006): 459–70, 423. https://doi.org/10.1038/sj.mp.4001806.

Clough, Timothy, Melissa Key, Ilka Ott, Susanne Ragg, Gunther Schadow, and Olga Vitek. "Protein Quantification in Label-Free LC-MS Experiments." *J. Proteome Res.* 8, no. 11 (November 2009): 5275–84. https://doi.org/10.1021/pr900610q.

Codo, Ana Campos, Gustavo Gastão Davanzo, Lauar de Brito Monteiro, Gabriela Fabiano de Souza, Stéfanie Primon Muraro, João Victor Virgilio-da-Silva, Juliana Silveira Prodonoff, et al. "Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis." *Cell Metabolism* 32, no. 3 (September 1, 2020): 437-446.e5. https://doi.org/10.1016/j.cmet.2020.07.007.

Coughlin, Jennifer M, Lindsay N Hayes, Teppei Tanaka, Meifang Xiao, Robert H Yolken, Paul Worley, F Markus Leweke, and Akira Sawa. "Reduced Superoxide Dismutase-1 (SOD1) in Cerebrospinal Fluid of Patients with Early Psychosis in Association with Clinical Features." *Schizophr. Res.* 183 (May 2017): 64–69. https://doi.org/10.1016/j.schres.2016.10.040.

Cox, Jürgen, and Matthias Mann. "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." *Nature Biotechnology* 26, no. 12 (December 2008): 1367–72. https://doi.org/10.1038/nbt.1511.

Crunfli, Fernanda, Victor C. Carregari, Flavio P. Veras, Lucas S. Silva, Mateus Henrique Nogueira, André Saraiva Leão Marcelo Antunes, Pedro Henrique Vendramini, et al. "Morphological, Cellular, and Molecular Basis of Brain Infection in COVID-19 Patients." *Proceedings of the National Academy of Sciences* 119, no. 35 (August 30, 2022): e2200960119. https://doi.org/10.1073/pnas.2200960119.

datadot. "COVID-19 Cases | WHO COVID-19 Dashboard." Accessed January 11, 2024. https://data.who.int/dashboards/covid19/cases.

Dean, B, N Thomas, E Scarr, and M Udawela. "Evidence for Impaired Glucose Metabolism in the Striatum, Obtained Postmortem, from Some Subjects with Schizophrenia." *Transl. Psychiatry* 6, no. 11 (November 15, 2016): e949. https://doi.org/10.1038/tp.2016.226.

Demichev, Vadim, Christoph B. Messner, Spyros I. Vernardis, Kathryn S. Lilley, and Markus Ralser. "DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput." *Nature Methods* 17, no. 1 (January 2020): 41–44. https://doi.org/10.1038/s41592-019-0638-x.

Demichev, Vadim, Lukasz Szyrwiel, Fengchao Yu, Guo Ci Teo, George Rosenberger, Agathe Niewienda, Daniela Ludwig, et al. "Dia-PASEF Data Analysis Using FragPipe and DIA-NN for Deep Proteomics of Low Sample Amounts." *Nature Communications* 13, no. 1 (July 8, 2022): 3944. https://doi.org/10.1038/s41467-022-31492-0.

Distler, Ute, Jörg Kuharev, Pedro Navarro, and Stefan Tenzer. "Label-Free Quantification in Ion Mobility-Enhanced Data-Independent Acquisition Proteomics." *Nature Protocols* 11, no. 4 (April 2016): 795–812. https://doi.org/10.1038/nprot.2016.042.

Dracheva, Stella, Kenneth L Davis, Benjamin Chin, Derek A Woo, James Schmeidler, and Vahram Haroutunian. "Myelin-Associated mRNA and Protein Expression Deficits in the Anterior Cingulate Cortex and Hippocampus in Elderly Schizophrenia Patients." *Neurobiol. Dis.* 21, no. 3 (March 2006): 531–40. https://doi.org/10.1016/j.nbd.2005.08.012.

Edfors, Fredrik, Frida Danielsson, Björn M Hallström, Lukas Käll, Emma Lundberg, Fredrik Pontén, Björn Forsström, and Mathias Uhlén. "Gene-specific Correlation of RNA and Protein Levels in Human Cells and Tissues." *Molecular Systems Biology* 12, no. 10 (October 2016): 883. https://doi.org/10.15252/msb.20167144.

Elias, Joshua E., and Steven P. Gygi. "Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry." *Nature Methods* 4, no. 3 (March 2007): 207–14. https://doi.org/10.1038/nmeth1019.

Elkabetz, Y., and L. Studer. "Human ESC-Derived Neural Rosettes and Neural Stem Cell Progression." *Cold Spring Harbor Symposia on Quantitative Biology* 73 (2008): 377–87. https://doi.org/10.1101/sqb.2008.73.052.

"Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation | PLOS ONE." Accessed October 1, 2023. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013984.

Erta, María, Albert Quintana, and Juan Hidalgo. "Interleukin-6, a Major Cytokine in the Central Nervous System." *International Journal of Biological Sciences* 8, no. 9 (2012): 1254–66. https://doi.org/10.7150/ijbs.4679.

Fang, Zhuoqing, Xinyuan Liu, and Gary Peltz. "GSEApy: A Comprehensive Package for Performing Gene Set Enrichment Analysis in Python." *Bioinformatics (Oxford, England)* 39, no. 1 (January 1, 2023): btac757. https://doi.org/10.1093/bioinformatics/btac757.

Fenn, John B., Matthias Mann, Chin Kai Meng, Shek Fu Wong, and Craig M. Whitehouse. "Electrospray Ionization for Mass Spectrometry of Large Biomolecules." *Science* 246, no. 4926 (October 6, 1989): 64–71. https://doi.org/10.1126/science.2675315.

Fernandez-Enright, F, J L Andrews, K A Newell, C Pantelis, and X F Huang. "Novel Implications of Lingo-1 and Its Signaling Partners in Schizophrenia." *Transl. Psychiatry* 4 (January 21, 2014): e348. https://doi.org/10.1038/tp.2013.121.

Fillman, S G, N Cloonan, V S Catts, L C Miller, J Wong, T McCrossin, M Cairns, and C S Weickert. "Increased Inflammatory Markers Identified in the Dorsolateral Prefrontal Cortex of Individuals with Schizophrenia." *Mol. Psychiatry* 18, no. 2 (February 2013): 206–14. https://doi.org/10.1038/mp.2012.110.

Flynn, S W, D J Lang, A L Mackay, V Goghari, I M Vavasour, K P Whittall, G N Smith, et al. "Abnormalities of Myelination in Schizophrenia Detected in Vivo with MRI, and Post-Mortem with Analysis of Oligodendrocyte Proteins." *Mol. Psychiatry* 8, no. 9 (September 2003): 811–20. https://doi.org/10.1038/sj.mp.4001337.

Föcking, M, P Dicker, L M Lopez, M Hryniewiecka, K Wynne, J A English, G Cagney, and D R Cotter. "Proteomic Analysis of the Postsynaptic Density Implicates

Synaptic Function and Energy Pathways in Bipolar Disorder." *Transl. Psychiatry* 6, no. 11 (November 29, 2016): e959. https://doi.org/10.1038/tp.2016.224.

Fraga, Ana M., Marina Sukoyan, Prithi Rajan, Daniela Paes de Almeida Ferreira Braga, Assumpto Iaconelli, José Gonçalves Franco, Edson Borges, and Lygia V. Pereira. "Establishment of a Brazilian Line of Human Embryonic Stem Cells in Defined Medium: Implications for Cell Therapy in an Ethnically Diverse Population." *Cell Transplantation* 20, no. 3 (2011): 431–40. https://doi.org/10.3727/096368910X522261.

Frankenfield, Ashley M., Jiawei Ni, Mustafa Ahmed, and Ling Hao. "Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics." *Journal of Proteome Research* 21, no. 9 (September 2, 2022): 2104–13. https://doi.org/10.1021/acs.jproteome.2c00145.

Freedman, Robert. "Schizophrenia." *N. Engl. J. Med.* 349, no. 18 (October 30, 2003): 1738–49. https://doi.org/10.1056/NEJMra035458.

Fusar-Poli, Paolo, Jorge Perez, Matthew Broome, Stefan Borgwardt, Anna Placentino, Eduardo Caverzasi, Mariachiara Cortesi, et al. "Neurofunctional Correlates of Vulnerability to Psychosis: A Systematic Review and Meta-Analysis." *Neurosci. Biobehav. Rev.* 31, no. 4 (January 12, 2007): 465–84. https://doi.org/10.1016/j.neubiorev.2006.11.006.

Gan, Xiaoqing, Jiyong Wang, Chen Wang, Eeva Sommer, Tohru Kozasa, Srinivasa Srinivasula, Dario Alessi, Stefan Offermanns, Melvin I Simon, and Dianqing Wu. "PRR5L Degradation Promotes mTORC2-Mediated PKC-δ Phosphorylation and Cell Migration Downstream of Gα12." *Nat. Cell Biol.* 14, no. 7 (May 20, 2012): 686–96. https://doi.org/10.1038/ncb2507.

Gillespie, Marc, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, et al. "The Reactome Pathway Knowledgebase 2022." *Nucleic Acids Research* 50, no. D1 (January 7, 2022): D687–92. https://doi.org/10.1093/nar/gkab1028.

Gillet, Ludovic C., Pedro Navarro, Stephen Tate, Hannes Röst, Nathalie Selevsek, Lukas Reiter, Ron Bonner, and Ruedi Aebersold. "Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis." *Molecular & Cellular Proteomics : MCP* 11, no. 6 (June 2012): O111.016717. https://doi.org/10.1074/mcp.O111.016717.

Gold, James M, Gregory P Strauss, James A Waltz, Benjamin M Robinson, Jamie K Brown, and Michael J Frank. "Negative Symptoms of Schizophrenia Are Associated with Abnormal Effort-Cost Computations." *Biol. Psychiatry* 74, no. 2 (July 15, 2013): 130–36. https://doi.org/10.1016/j.biopsych.2012.12.022.

Gonzalez-Liencres, Cristina, Cumhur Tas, Elliot C Brown, Soner Erdin, Ece Onur, Zeynep Cubukcoglu, Omer Aydemir, Aysen Esen-Danaci, and Martin Brüne. "Oxidative Stress in Schizophrenia: A Case-Control Study on the Effects on Social Cognition and Neurocognition." *BMC Psychiatry* 14 (September 24, 2014): 268. https://doi.org/10.1186/s12888-014-0268-x.

Goto-Silva, Livia, Nadia M. E. Ayad, Iasmin L. Herzog, Nilton P. Silva, Bernard Lamien, Helcio R. B. Orlande, Annie da Costa Souza, et al. "Computational Fluid Dynamic Analysis of Physical Forces Playing a Role in Brain Organoid Cultures in Two Different Multiplex Platforms." *BMC Developmental Biology* 19, no. 1 (March 7, 2019): 3. https://doi.org/10.1186/s12861-019-0183-y.

Gu, Zuguang, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors. "Circlize Implements and Enhances Circular Visualization in R." *Bioinformatics* 30, no. 19 (October 1, 2014): 2811–12. https://doi.org/10.1093/bioinformatics/btu393.

Gupta, Aakriti, Mahesh V. Madhavan, Kartik Sehgal, Nandini Nair, Shiwani Mahajan, Tejasav S. Sehrawat, Behnood Bikdeli, et al. "Extrapulmonary Manifestations of COVID-19." *Nature Medicine* 26, no. 7 (July 2020): 1017–32. https://doi.org/10.1038/s41591-020-0968-3.

Haznedar, M Mehmet, Monte S Buchsbaum, Erin A Hazlett, Lina Shihabuddin, Antonia New, and Larry J Siever. "Cingulate Gyrus Volume and Metabolism in the Schizophrenia Spectrum." *Schizophr. Res.* 71, no. 2–3 (December 1, 2004): 249–62. https://doi.org/10.1016/j.schres.2004.02.025.

Hirayama-Kurogi, Mio, Yohei Takizawa, Yasuto Kunii, Junya Matsumoto, Akira Wada, Mizuki Hino, Hiroyasu Akatsu, et al. "Downregulation of GNA13-ERK Network in Prefrontal Cortex of Schizophrenia Brain Identified by Combined Focused and Targeted Quantitative Proteomics." *J. Proteomics* 158 (March 31, 2017): 31–42. https://doi.org/10.1016/j.jprot.2017.02.009.

Hoesel, Bastian, and Johannes A Schmid. "The Complexity of NF-κB Signaling in Inflammation and Cancer." *Mol. Cancer* 12 (August 2, 2013): 86. https://doi.org/10.1186/1476-4598-12-86.

Hoffmann, Markus, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, et al. "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor." *Cell* 181, no. 2 (April 16, 2020): 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052.

Honey, Garry D, Edith Pomarol-Clotet, Philip R Corlett, Rebekah A E Honey, Peter J McKenna, Edward T Bullmore, and Paul C Fletcher. "Functional Dysconnectivity in Schizophrenia Associated with Attentional Modulation of Motor Function." *Brain* 128, no. Pt 11 (November 2005): 2597–2611. https://doi.org/10.1093/brain/awh632.

Howes, Oliver D, and Shitij Kapur. "The Dopamine Hypothesis of Schizophrenia: Version III--the Final Common Pathway." *Schizophr. Bull.* 35, no. 3 (May 2009): 549–62. https://doi.org/10.1093/schbul/sbp006.

Howes, Oliver D, Robert McCutcheon, Michael J Owen, and Robin M Murray. "The Role of Genes, Stress, and Dopamine in the Development of Schizophrenia." *Biol. Psychiatry* 81, no. 1 (January 1, 2017): 9–20. https://doi.org/10.1016/j.biopsych.2016.07.014.

Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, et al. "Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China." *Lancet (London, England)* 395, no. 10223 (February 15, 2020): 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. "Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists." *Nucleic Acids Research* 37, no. 1 (January 2009): 1–13. https://doi.org/10.1093/nar/gkn923.

———. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." *Nature Protocols* 4, no. 1 (January 2009): 44–57. https://doi.org/10.1038/nprot.2008.211.

———. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." *Nature Protocols* 4, no. 1 (January 2009): 44–57. https://doi.org/10.1038/nprot.2008.211.

Jaaro-Peled, Hanna, Akiko Hayashi-Takagi, Saurav Seshadri, Atsushi Kamiya, Nicholas J Brandon, and Akira Sawa. "Neurodevelopmental Mechanisms of Schizophrenia: Understanding Disturbed Postnatal Brain Maturation through Neuregulin-1-ErbB4 and DISC1." *Trends Neurosci.* 32, no. 9 (September 2009): 485–95. https://doi.org/10.1016/j.tins.2009.05.007.

Jacob, Fadi, Sarshan R. Pather, Wei-Kai Huang, Feng Zhang, Samuel Zheng Hao Wong, Haowen Zhou, Beatrice Cubitt, et al. "Human Pluripotent Stem Cell-Derived Neural Cells and Brain Organoids Reveal SARS-CoV-2 Neurotropism Predominates in Choroid Plexus Epithelium." *Cell Stem Cell* 27, no. 6 (December 3, 2020): 937-950.e9. https://doi.org/10.1016/j.stem.2020.09.016.

Jia, Peilin, Lily Wang, Ayman H Fanous, Carlos N Pato, Todd L Edwards, International Schizophrenia Consortium, and Zhongming Zhao. "Network-Assisted Investigation of Combined Causal Signals from Genome-Wide Association Studies in Schizophrenia." *PLoS Comput. Biol.* 8, no. 7 (July 5, 2012): e1002587. https://doi.org/10.1371/journal.pcbi.1002587.

Kahn, Patricia. "From Genome to Proteome: Looking at a Cell's Proteins." *Science* 270, no. 5235 (October 20, 1995): 369–70. https://doi.org/10.1126/science.270.5235.369.

Kalejaiye, Titilola D., Rohan Bhattacharya, Morgan A. Burt, Tatianna Travieso, Arinze E. Okafor, Xingrui Mou, Maria Blasi, and Samira Musah. "SARS-CoV-2 Employ BSG/CD147 and ACE2 Receptors to Directly Infect Human Induced Pluripotent Stem Cell-Derived Kidney Podocytes." *Frontiers in Cell and Developmental Biology* 10 (2022): 855340. https://doi.org/10.3389/fcell.2022.855340.

Kanehisa, M., and S. Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28, no. 1 (January 1, 2000): 27–30. https://doi.org/10.1093/nar/28.1.27.

Kanehisa, Minoru, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. "KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs." *Nucleic Acids Research* 45, no. D1 (January 4, 2017): D353–61. https://doi.org/10.1093/nar/gkw1092.

Kantrowitz, Joshua T, and Daniel C Javitt. "N-Methyl-d-Aspartate (NMDA) Receptor Dysfunction or Dysregulation: The Final Common Pathway on the Road to Schizophrenia?" *Brain Res. Bull.* 83, no. 3–4 (September 2010): 108–21. https://doi.org/10.1016/j.brainresbull.2010.04.006.

Kim, Jeong-Hee, Jong-Hoon Kim, Young-Don Son, Yo-Han Joo, Sang-Yoon Lee, Hang-Keun Kim, and Myung-Kyun Woo. "Altered Interregional Correlations between Serotonin Transporter Availability and Cerebral Glucose Metabolism in Schizophrenia: A High-Resolution PET Study Using [C]DASB and [F]FDG." *Schizophr. Res.* 182 (April 2017): 55–65. https://doi.org/10.1016/j.schres.2016.10.020.

Kim, Minkyu, Jisoo Park, Mehdi Bouhaddou, Kyumin Kim, Ajda Rojc, Maya Modak, Margaret Soucheray, et al. "A Protein Interaction Landscape of Breast Cancer." *Science* 374, no. 6563 (October 2021): eabf3066. https://doi.org/10.1126/science.abf3066.

Kirino, Eiji, Shoji Tanaka, Mayuko Fukuta, Rie Inami, Reiichi Inoue, and Shigeki Aoki. "Functional Connectivity of the Caudate in Schizophrenia Evaluated with Simultaneous Resting-State Functional MRI and Electroencephalography

Recordings." *Neuropsychobiology* 77, no. 4 (2019): 165–75. https://doi.org/10.1159/000490429.

Kitano, Hiroaki. "Systems Biology: A Brief Overview." *Science* 295, no. 5560 (March 2002): 1662–64. https://doi.org/10.1126/science.1069492.

Kitata, Reta Birhanu, Jhih-Ci Yang, and Yu-Ju Chen. "Advances in Data-Independent Acquisition Mass Spectrometry towards Comprehensive Digital Proteome Landscape." *Mass Spectrometry Reviews* 42, no. 6 (2023): 2324–48. https://doi.org/10.1002/mas.21781.

Knutson, Brian, and Jeffrey C Cooper. "Functional Magnetic Resonance Imaging of Reward Prediction." *Curr. Opin. Neurol.* 18, no. 4 (August 2005): 411–17. https://doi.org/10.1097/01.wco.0000173463.24758.f6.

Kovalevich, Jane, and Dianne Langford. "Considerations for the Use of SH-SY5Y Neuroblastoma Cells in Neurobiology." *Methods in Molecular Biology (Clifton, N.J.)* 1078 (2013): 9–21. https://doi.org/10.1007/978-1-62703-640-5_2.

Koyuncu, Orkide O., Ian B. Hogue, and Lynn W. Enquist. "Virus Infections in the Nervous System." *Cell Host & Microbe* 13, no. 4 (April 17, 2013): 379–93. https://doi.org/10.1016/j.chom.2013.03.010.

Kreczmanski, Pawel, Helmut Heinsen, Valentina Mantua, Fritz Woltersdorf, Thorsten Masson, Norbert Ulfig, Rainald Schmidt-Kastner, et al. "Volume, Neuron Density and Total Neuron Number in Five Subcortical Regions in Schizophrenia." *Brain* 130, no. Pt 3 (March 2007): 678–92. https://doi.org/10.1093/brain/awl386.

Krismer, Elena, Isabell Bludau, Maximilian T Strauss, and Matthias Mann. "AlphaPeptStats: An Open-Source Python Package for Automated and Scalable Statistical Analysis of Mass Spectrometry-Based Proteomics." *Bioinformatics* 39, no. 8 (August 1, 2023): btad461. https://doi.org/10.1093/bioinformatics/btad461.

Kruglikov, Ilja L., and Philipp E. Scherer. "The Role of Adipocytes and Adipocyte-Like Cells in the Severity of COVID-19 Infections." *Obesity (Silver Spring, Md.)* 28, no. 7 (July 2020): 1187–90. https://doi.org/10.1002/oby.22856.

Krzywinski, Martin, Jacqueline Schein, İnanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. "Circos: An Information Aesthetic for Comparative Genomics." *Genome Research* 19, no. 9 (January 9, 2009): 1639–45. https://doi.org/10.1101/gr.092759.109.

Kurzawa, Nils, Isabelle Becher, Sindhuja Sridharan, Holger Franken, André Mateus, Simon Anders, Marcus Bantscheff, Wolfgang Huber, and Mikhail M. Savitski. "A Computational Method for Detection of Ligand-Binding Proteins from Dose Range Thermal Proteome Profiles." *Nature Communications* 11, no. 1 (November 13, 2020): 5783. https://doi.org/10.1038/s41467-020-19529-8.

Kurzawa, Nils, Isabelle Rose Leo, Matthias Stahl, Elena Kunold, Isabelle Becher, Anastasia Audrey, Georgios Mermelekas, et al. "Deep Thermal Profiling for Detection of Functional Proteoform Groups." *Nature Chemical Biology* 19, no. 8 (August 2023): 962–71. https://doi.org/10.1038/s41589-023-01284-8.

Lahti, R A, R C Roberts, and C A Tamminga. "D2-Family Receptor Distribution in Human Postmortem Tissue: An Autoradiographic Study." *Neuroreport* 6, no. 18 (December 15, 1995): 2505–12. https://doi.org/10.1097/00001756-199512150-00015.

Lechner, Severin, Raphael R. Steimbach, Longlong Wang, Marshall L. Deline, Yun-Chien Chang, Tobias Fromme, Martin Klingenspor, et al. "Chemoproteomic Target Deconvolution Reveals Histone Deacetylases as Targets of (R)-Lipoic Acid."

Nature Communications 14, no. 1 (June 15, 2023): 3548. https://doi.org/10.1038/s41467-023-39151-8.

Ledur, Pítia Flores, Karina Karmirian, Carolina da Silva Gouveia Pedrosa, Leticia Rocha Quintino Souza, Gabriela Assis-de-Lemos, Thiago Martino Martins, Jéssica de Cassia Cavalheiro Gomes Ferreira, et al. "Zika Virus Infection Leads to Mitochondrial Failure, Oxidative Stress and DNA Damage in Human iPSC-Derived Astrocytes." *Scientific Reports* 10, no. 1 (January 27, 2020): 1218. https://doi.org/10.1038/s41598-020-57914-x.

Leech, Robert, and David J Sharp. "The Role of the Posterior Cingulate Cortex in Cognition and Disease." *Brain* 137, no. 1 (2014): 12–32. https://doi.org/10.1093/brain/awt162.

Lei, Han-Yu, Ying-He Ding, Kai Nie, Yin-Miao Dong, Jia-Hao Xu, Meng-Ling Yang, Meng-Qi Liu, et al. "Potential Effects of SARS-CoV-2 on the Gastrointestinal Tract and Liver." *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie* 133 (January 2021): 111064. https://doi.org/10.1016/j.biopha.2020.111064.

Lex, Alexander, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. "UpSet: Visualization of Intersecting Sets." *IEEE Transactions on Visualization and Computer Graphics* 20, no. 12 (December 2014): 1983–92. https://doi.org/10.1109/TVCG.2014.2346248.

Li, Guo-Zhong, Johannes P C Vissers, Jeffrey C Silva, Dan Golick, Marc V Gorenstein, and Scott J Geromanos. "Database Searching and Accounting of Multiplexed Precursor and Product Ion Spectra from the Data Independent Analysis of Simple and Complex Peptide Mixtures." *Proteomics* 9, no. 6 (March 2009): 1696–1719. https://doi.org/10.1002/pmic.200800564.

Lin, Lu, Xiayang Jiang, Zhenling Zhang, Siwen Huang, Zhenyi Zhang, Zhaoxiong Fang, Zhiqiang Gu, et al. "Gastrointestinal Symptoms of 95 Cases with SARS-CoV-2 Infection." *Gut* 69, no. 6 (June 2020): 997–1001. https://doi.org/10.1136/gutjnl-2020-321013.

Liu, Fan, Philip Lössl, Beverley M. Rabbitts, Robert S. Balaban, and Albert J. R. Heck. "The Interactome of Intact Mitochondria by Cross-Linking Mass Spectrometry Provides Evidence for Coexisting Respiratory Supercomplexes." *Molecular & Cellular Proteomics: MCP* 17, no. 2 (February 2018): 216–32. https://doi.org/10.1074/mcp.RA117.000470.

Liu, Yansheng, Andreas Beyer, and Ruedi Aebersold. "On the Dependency of Cellular Protein Levels on mRNA Abundance." *Cell* 165, no. 3 (April 21, 2016): 535–50. https://doi.org/10.1016/j.cell.2016.03.014.

MacDowell, Karina S, Raquel Pinacho, Juan C Leza, Joan Costa, Belén Ramos, and Borja García-Bueno. "Differential Regulation of the TLR4 Signalling Pathway in Post-Mortem Prefrontal Cortex and Cerebellum in Chronic Schizophrenia: Relationship with SP Transcription Factors." *Prog. Neuropsychopharmacol. Biol. Psychiatry* 79, no. Pt B (October 3, 2017): 481–92. https://doi.org/10.1016/j.pnpbp.2017.08.005.

Maloku, Ekrem, Ignacio R Covelo, Ingeborg Hanbauer, Alessandro Guidotti, Bashkim Kadriu, Qiaoyan Hu, John M Davis, and Erminio Costa. "Lower Number of Cerebellar Purkinje Neurons in Psychosis Is Associated with Reduced Reelin Expression." *Proc. Natl. Acad. Sci. U. S. A.* 107, no. 9 (March 2, 2010): 4407–11. https://doi.org/10.1073/pnas.0914483107.

Martins-de-Souza, Daniel, Wagner F Gattaz, Andrea Schmitt, Giuseppina Maccarrone, Eva Hunyadi-Gulyás, Marcos N Eberlin, Gustavo H M F Souza, et al. "Proteomic Analysis of Dorsolateral Prefrontal Cortex Indicates the Involvement of Cytoskeleton, Oligodendrocyte, Energy Metabolism and New Potential Markers in Schizophrenia." *J. Psychiatr. Res.* 43, no. 11 (July 2009): 978–86. https://doi.org/10.1016/j.jpsychires.2008.11.006.

Martins-de-Souza, Daniel, Wagner F Gattaz, Andrea Schmitt, José C Novello, Sérgio Marangoni, Christoph W Turck, and Emmanuel Dias-Neto. "Proteome Analysis of Schizophrenia Patients Wernicke's Area Reveals an Energy Metabolism Dysregulation." *BMC Psychiatry* 9 (April 30, 2009): 17. https://doi.org/10.1186/1471-244X-9-17.

Martins-de-Souza, Daniel, Wagner F Gattaz, Andrea Schmitt, Christiane Rewerts, Sérgio Marangoni, José C Novello, Giuseppina Maccarrone, Christoph W Turck, and Emmanuel Dias-Neto. "Alterations in Oligodendrocyte Proteins, Calcium Homeostasis and New Potential Markers in Schizophrenia Anterior Temporal Lobe Are Revealed by Shotgun Proteome Analysis." *J. Neural Transm.* 116, no. 3 (March 2009): 275–89. https://doi.org/10.1007/s00702-008-0156-y.

Martins-de-Souza, Daniel, Giuseppina Maccarrone, Thomas Wobrock, Inga Zerr, Philipp Gormanns, Stefan Reckow, Peter Falkai, Andrea Schmitt, and Christoph W Turck. "Proteome Analysis of the Thalamus and Cerebrospinal Fluid Reveals Glycolysis Dysfunction and Potential Biomarkers Candidates for Schizophrenia." *J. Psychiatr. Res.* 44, no. 16 (December 2010): 1176–89. https://doi.org/10.1016/j.jpsychires.2010.04.014.

Mei, Lin, and Wen-Cheng Xiong. "Neuregulin 1 in Neural Development, Synaptic Plasticity and Schizophrenia." *Nat. Rev. Neurosci.* 9, no. 6 (June 2008): 437–52. https://doi.org/10.1038/nrn2392.

Meier, Florian, Andreas-David Brunner, Max Frank, Annie Ha, Isabell Bludau, Eugenia Voytik, Stephanie Kaspar-Schoenefeld, et al. "diaPASEF: Parallel Accumulation–Serial Fragmentation Combined with Data-Independent Acquisition." *Nature Methods* 17, no. 12 (December 2020): 1229–36. https://doi.org/10.1038/s41592-020-00998-0.

Meigs, Thomas E, Mary Fedor-Chaiken, Daniel D Kaplan, Robert Brackenbury, and Patrick J Casey. "Galpha12 and Galpha13 Negatively Regulate the Adhesive Functions of Cadherin." *J. Biol. Chem.* 277, no. 27 (July 5, 2002): 24594–600. https://doi.org/10.1074/jbc.M201984200.

Merico, Daniele, Ruth Isserlin, Oliver Stueker, Andrew Emili, and Gary D. Bader. "Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation." Edited by Timothy Ravasi. *PLoS ONE* 5, no. 11 (November 15, 2010): e13984. https://doi.org/10.1371/journal.pone.0013984.

Mitelman, Serge A, William Byne, Eileen M Kemether, Erin A Hazlett, and Monte S Buchsbaum. "Metabolic Disconnection Between the Mediodorsal Nucleus of the Thalamus and Cortical Brodmann's Areas of the Left Hemisphere in Schizophrenia." *American Journal of Psychiatry* 162, no. 9 (2005): 1733–35. https://doi.org/10.1176/appi.ajp.162.9.1733.

Moberget, T, N T Doan, D Alnæs, T Kaufmann, A Córdova-Palomera, T V Lagerberg, J Diedrichsen, et al. "Cerebellar Volume and Cerebellocerebral Structural Covariance in Schizophrenia: A Multisite Mega-Analysis of 983 Patients and 1349 Healthy Controls." *Mol. Psychiatry* 23, no. 6 (June 2018): 1512–20. https://doi.org/10.1038/mp.2017.106.

Moers, Alexandra, Alexander Nürnberg, Sandra Goebbels, Nina Wettschureck, and Stefan Offermanns. "Galpha12/Galpha13 Deficiency Causes Localized Overmigration of Neurons in the Developing Cerebral and Cerebellar Cortices." *Mol. Cell. Biol.* 28, no. 5 (March 2008): 1480–88. https://doi.org/10.1128/MCB.00651-07.

Morgan, Martin, Valerie Obenchain, Jim Hester, and Hervé Pagès. "SummarizedExperiment: SummarizedExperiment Container." Bioconductor version: Release (3.17), 2023. https://doi.org/10.18129/B9.bioc.SummarizedExperiment.

"MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale | Journal of Proteome Research." Accessed September 30, 2023. https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00834.

Mueller, Sophia, Danhong Wang, Ruiqi Pan, Daphne J Holt, and Hesheng Liu. "Abnormalities in Hemispheric Specialization of Caudate Nucleus Connectivity in Schizophrenia." *JAMA Psychiatry* 72, no. 6 (June 2015): 552–60. https://doi.org/10.1001/jamapsychiatry.2014.3176.

Müller, Norbert, Elif Weidinger, Bianka Leitner, and Markus J Schwarz. "The Role of Inflammation in Schizophrenia." *Front. Neurosci.* 9 (October 21, 2015): 372. https://doi.org/10.3389/fnins.2015.00372.

Mulvey, Claire M., Lisa M. Breckels, Oliver M. Crook, David J. Sanders, Andre L. R. Ribeiro, Aikaterini Geladaki, Andy Christoforou, et al. "Spatiotemporal Proteomic Profiling of the Pro-Inflammatory Response to Lipopolysaccharide in the THP-1 Human Leukaemia Cell Line." *Nature Communications* 12, no. 1 (October 1, 2021): 5773. https://doi.org/10.1038/s41467-021-26000-9.

Mwansisya, Tumbwene E, Aimin Hu, Yihui Li, Xudong Chen, Guowei Wu, Xiaojun Huang, Dongsheng Lv, et al. "Task and Resting-State fMRI Studies in First-Episode Schizophrenia: A Systematic Review." *Schizophr. Res.* 189 (November 2017): 9–18. https://doi.org/10.1016/j.schres.2017.02.026.

Nascimento, Juliana M, Sheila Garcia, Verônica M Saia-Cereda, Aline G Santana, Caroline Brandao-Teles, Giuliana S Zuccoli, Danielle G Junqueira, et al. "Proteomics and Molecular Tools for Unveiling Missing Links in the Biochemical Understanding of Schizophrenia." *Proteomics Clin. Appl.* 10, no. 12 (December 2016): 1148–58. https://doi.org/10.1002/prca.201600021.

Nascimento, Juliana M, and Daniel Martins-de-Souza. "The Proteome of Schizophrenia." *NPJ Schizophr* 1 (March 4, 2015): 14003. https://doi.org/10.1038/npjschz.2014.3.

Needham, Elise J., Janne R. Hingst, Benjamin L. Parker, Kaitlin R. Morrison, Guang Yang, Johan Onslev, Jonas M. Kristensen, et al. "Personalized Phosphoproteomics Identifies Functional Signaling." *Nature Biotechnology* 40, no. 4 (April 2022): 576–84. https://doi.org/10.1038/s41587-021-01099-9.

Nie, Xiu, Liujia Qian, Rui Sun, Bo Huang, Xiaochuan Dong, Qi Xiao, Qiushi Zhang, et al. "Multi-Organ Proteomic Landscape of COVID-19 Autopsies." *Cell* 184, no. 3 (February 4, 2021): 775-791.e14. https://doi.org/10.1016/j.cell.2021.01.004.

Nielsen, Mette Ødegaard, Egill Rostrup, Brian Villumsen Broberg, Sanne Wulff, and Birte Glenthøj. "Negative Symptoms and Reward Disturbances in Schizophrenia Before and After Antipsychotic Monotherapy." *Clin. EEG Neurosci.* 49, no. 1 (January 2018): 36–45. https://doi.org/10.1177/1550059417744120.

Ong, Shao-En, and Matthias Mann. "Mass Spectrometry–Based Proteomics Turns Quantitative." *Nature Chemical Biology* 1, no. 5 (October 2005): 252–62. https://doi.org/10.1038/nchembio736.

Os, Jim van, and Shitij Kapur. "Schizophrenia." *Lancet* 374, no. 9690 (August 2009): 635–45. https://doi.org/10.1016/S0140-6736(09)60995-8.

Osimo, Emanuele Felice, Katherine Beck, Tiago Reis Marques, and Oliver D Howes. "Synaptic Loss in Schizophrenia: A Meta-Analysis and Systematic Review of Synaptic Protein and mRNA Measures." *Mol. Psychiatry* 24, no. 4 (April 2019): 549–61. https://doi.org/10.1038/s41380-018-0041-5.

Perez-Costas, Emma, Miguel Melendez-Ferro, and Rosalinda C Roberts. "Basal Ganglia Pathology in Schizophrenia: Dopamine Connections and Anomalies." *J. Neurochem.* 113, no. 2 (April 2010): 287–302. https://doi.org/10.1111/j.1471-4159.2010.06604.x.

Prabakaran, S, J E Swatton, M M Ryan, S J Huffaker, J T-J Huang, J L Griffin, M Wayland, et al. "Mitochondrial Dysfunction in Schizophrenia: Evidence for Compromised Brain Metabolism and Oxidative Stress." *Mol. Psychiatry* 9, no. 7 (July 2004): 684–97, 643. https://doi.org/10.1038/sj.mp.4001511.

Raichle, M E, A M MacLeod, A Z Snyder, W J Powers, D A Gusnard, and G L Shulman. "A Default Mode of Brain Function." *Proc. Natl. Acad. Sci. U. S. A.* 98, no. 2 (January 16, 2001): 676–82. https://doi.org/10.1073/pnas.98.2.676.

Ramaker, Ryne C, Kevin M Bowling, Brittany N Lasseigne, Megan H Hagenauer, Andrew A Hardigan, Nicholas S Davis, Jason Gertz, et al. "Post-Mortem Molecular Profiling of Three Psychiatric Disorders." *Genome Med.* 9, no. 1 (July 28, 2017): 72. https://doi.org/10.1186/s13073-017-0458-5.

Reeber, Stacey L, Tom S Otis, and Roy V Sillitoe. "New Roles for the Cerebellum in Health and Disease." *Front. Syst. Neurosci.* 7 (November 14, 2013): 83. https://doi.org/10.3389/fnsys.2013.00083.

Reis-de-Oliveira, Guilherme, Mariana Fioramonte, and Daniel Martins-de-Souza. "A Complete Proteomic Workflow to Study Brain-Related Disorders via Postmortem Tissue." *Methods Mol. Biol.* 1916 (2019): 319–28. https://doi.org/10.1007/978-1-4939-8994-2_31.

Reiterer, Moritz, Mangala Rajan, Nicolás Gómez-Banoy, Jennifer D. Lau, Luis G. Gomez-Escobar, Lunkun Ma, Ankit Gilani, et al. "Hyperglycemia in Acute COVID-19 Is Characterized by Insulin Resistance and Adipose Tissue Infectivity by SARS-CoV-2." *Cell Metabolism* 33, no. 11 (November 2, 2021): 2174-2188.e5. https://doi.org/10.1016/j.cmet.2021.09.009.

Robbiani, Davide F., Christian Gaebler, Frauke Muecksch, Julio C. C. Lorenzi, Zijun Wang, Alice Cho, Marianna Agudelo, et al. "Convergent Antibody Responses to SARS-CoV-2 in Convalescent Individuals." *Nature* 584, no. 7821 (August 2020): 437–42. https://doi.org/10.1038/s41586-020-2456-9.

Roberts, R C, K A Barksdale, J K Roche, and A C Lahti. "Decreased Synaptic and Mitochondrial Density in the Postmortem Anterior Cingulate Cortex in Schizophrenia." *Schizophr. Res.* 168, no. 1–2 (October 2015): 543–53. https://doi.org/10.1016/j.schres.2015.07.016.

Roberts, Rosalinda C. "Postmortem Studies on Mitochondria in Schizophrenia." *Schizophr. Res.* 187 (September 2017): 17–25. https://doi.org/10.1016/j.schres.2017.01.056.

Rochfort, Keith D., Laura E. Collins, Ronan P. Murphy, and Philip M. Cummins. "Downregulation of Blood-Brain Barrier Phenotype by Proinflammatory

Cytokines Involves NADPH Oxidase-Dependent ROS Generation: Consequences for Interendothelial Adherens and Tight Junctions." *PloS One* 9, no. 7 (2014): e101815. https://doi.org/10.1371/journal.pone.0101815.

Rudolph, Jan Daniel, and Jürgen Cox. "A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis." *Journal of Proteome Research* 18, no. 5 (May 3, 2019): 2052–64. https://doi.org/10.1021/acs.jproteome.8b00927.

Ryan, Paul MacDaragh, and Noel M. Caplice. "Is Adipose Tissue a Reservoir for Viral Spread, Immune Activation, and Cytokine Amplification in Coronavirus Disease 2019?" *Obesity (Silver Spring, Md.)* 28, no. 7 (July 2020): 1191–94. https://doi.org/10.1002/oby.22843.

Saab, Aiman S, Iva D Tzvetanova, and Klaus-Armin Nave. "The Role of Myelin and Oligodendrocytes in Axonal Energy Metabolism." *Curr. Opin. Neurobiol.* 23, no. 6 (December 2013): 1065–72. https://doi.org/10.1016/j.conb.2013.09.008.

Saccon, Tatiana Dandolini, Felippe Mousovich-Neto, Raissa Guimarães Ludwig, Victor Corasolla Carregari, Ana Beatriz Dos Anjos Souza, Amanda Stephane Cruz Dos Passos, Matheus Cavalheiro Martini, et al. "SARS-CoV-2 Infects Adipose Tissue in a Fat Depot- and Viral Lineage-Dependent Manner." *Nature Communications* 13, no. 1 (September 29, 2022): 5722. https://doi.org/10.1038/s41467-022-33218-8.

Saia-Cereda, Verônica M, Juliana S Cassoli, Daniel Martins-de-Souza, and Juliana M Nascimento. "Psychiatric Disorders Biochemical Pathways Unraveled by Human Brain Proteomics." *Eur. Arch. Psychiatry Clin. Neurosci.* 267, no. 1 (February 2017): 3–17. https://doi.org/10.1007/s00406-016-0709-2.

Saia-Cereda, Verônica M, Juliana S Cassoli, Andrea Schmitt, Peter Falkai, and Daniel Martins-de-Souza. "Differential Proteome and Phosphoproteome May Impact Cell Signaling in the Corpus Callosum of Schizophrenia Patients." *Schizophr. Res.* 177, no. 1–3 (November 2016): 70–77. https://doi.org/10.1016/j.schres.2016.03.022.

Saia-Cereda, Verônica M, Juliana S Cassoli, Andrea Schmitt, Peter Falkai, Juliana M Nascimento, and Daniel Martins-de-Souza. "Proteomics of the Corpus Callosum Unravel Pivotal Players in the Dysfunction of Cell Signaling, Structure, and Myelination in Schizophrenia Brains." *Eur. Arch. Psychiatry Clin. Neurosci.* 265, no. 7 (October 2015): 601–12. https://doi.org/10.1007/s00406-015-0621-1.

Saia-Cereda, Verônica M, Aline G Santana, Andrea Schmitt, Peter Falkai, and Daniel Martins-de-Souza. "The Nuclear Proteome of White and Gray Matter from Schizophrenia Postmortem Brains." *Mol Neuropsychiatry* 3, no. 1 (July 2017): 37–52. https://doi.org/10.1159/000477299.

Santos, Aline M., Deborah Schechtman, Alisson C. Cardoso, Carolina F. M. Z. Clemente, Júlio C. Silva, Mariana Fioramonte, Michelle B. M. Pereira, et al. "FERM Domain Interaction with Myosin Negatively Regulates FAK in Cardiomyocyte Hypertrophy." *Nature Chemical Biology* 8, no. 1 (January 2012): 102–10. https://doi.org/10.1038/nchembio.717.

Santos, Marlon D. M., Diogo B. Lima, Juliana S. G. Fischer, Milan A. Clasen, Louise U. Kurt, Amanda Caroline Camillo-Andrade, Leandro C. Monteiro, et al. "Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V." *Nature Protocols* 17, no. 7 (July 2022): 1553–78. https://doi.org/10.1038/s41596-022-00690-x.

Saperia, Sarah, Susana Da Silva, Ishraq Siddiqui, Ofer Agid, Z Jeff Daskalakis, Arun Ravindran, Aristotle N Voineskos, Konstantine K Zakzanis, Gary Remington,

and George Foussias. "Reward-Driven Decision-Making Impairments in Schizophrenia." *Schizophr. Res.*, November 12, 2018. https://doi.org/10.1016/j.schres.2018.11.004.

Satopaa, Ville, Jeannie Albrecht, David Irwin, and Barath Raghavan. "Finding a 'Kneedle' in a Haystack: Detecting Knee Points in System Behavior." In *2011 31st International Conference on Distributed Computing Systems Workshops*, 166–71. Minneapolis, MN, USA: IEEE, 2011. https://doi.org/10.1109/ICDCSW.2011.20.

Savitski, Mikhail M., Friedrich B. M. Reinhard, Holger Franken, Thilo Werner, Maria Fälth Savitski, Dirk Eberhard, Daniel Martinez Molina, et al. "Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome." *Science* 346, no. 6205 (October 3, 2014): 1255784. https://doi.org/10.1126/science.1255784.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. "Biological Insights from 108 Schizophrenia-Associated Genetic Loci." *Nature* 511, no. 7510 (July 24, 2014): 421–27. https://doi.org/10.1038/nature13595.

Schmittgen, Thomas D., and Kenneth J. Livak. "Analyzing Real-Time PCR Data by the Comparative C(T) Method." *Nature Protocols* 3, no. 6 (2008): 1101–8. https://doi.org/10.1038/nprot.2008.73.

Searle, Brian C., Lindsay K. Pino, Jarrett D. Egertson, Ying S. Ting, Robert T. Lawrence, Brendan X. MacLean, Judit Villén, and Michael J. MacCoss. "Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry." *Nature Communications* 9, no. 1 (December 3, 2018): 5128. https://doi.org/10.1038/s41467-018-07454-w.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13, no. 11 (November 2003): 2498–2504. https://doi.org/10.1101/gr.1239303.

Shimoyama, Yuki. "pyCirclize: Circular Visualization in Python." Python, December 2022. https://github.com/moshi4/pyCirclize.

Shipley, Mackenzie M., Colleen A. Mangold, and Moriah L. Szpara. "Differentiation of the SH-SY5Y Human Neuroblastoma Cell Line." *Journal of Visualized Experiments: JoVE*, no. 108 (February 17, 2016): 53193. https://doi.org/10.3791/53193.

Shuken, Steven R. "An Introduction to Mass Spectrometry-Based Proteomics." *Journal of Proteome Research* 22, no. 7 (July 7, 2023): 2151–71. https://doi.org/10.1021/acs.jproteome.2c00838.

Sia, Sin Fun, Li-Meng Yan, Alex W. H. Chin, Kevin Fung, Ka-Tim Choy, Alvina Y. L. Wong, Prathanporn Kaewpreedee, et al. "Pathogenesis and Transmission of SARS-CoV-2 in Golden Hamsters." *Nature* 583, no. 7818 (July 2020): 834–38. https://doi.org/10.1038/s41586-020-2342-5.

Silva, Jeffrey C., Richard Denny, Craig A. Dorschel, Marc Gorenstein, Ignatius J. Kass, Guo-Zhong Li, Therese McKenna, et al. "Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs." *Analytical Chemistry* 77, no. 7 (April 1, 2005): 2187–2200. https://doi.org/10.1021/ac048455k.

Silva, Jeffrey C, Marc V Gorenstein, Guo-Zhong Li, Johannes P C Vissers, and Scott J Geromanos. "Absolute Quantification of Proteins by LCMSE." *Molecular & Cellular Proteomics* 5, no. 1 (2006): 144–56. https://doi.org/10.1074/mcp.m500230-mcp200.

Somerville, Shahza M, Robert R Conley, and Rosalinda C Roberts. "Mitochondria in the Striatum of Subjects with Schizophrenia." *World J. Biol. Psychiatry* 12, no. 1 (February 2011): 48–56. https://doi.org/10.3109/15622975.2010.505662.

Stefan, Norbert, Andreas L. Birkenfeld, Matthias B. Schulze, and David S. Ludwig. "Obesity and Impaired Metabolic Health in Patients with COVID-19." *Nature Reviews. Endocrinology* 16, no. 7 (July 2020): 341–42. https://doi.org/10.1038/s41574-020-0364-6.

Stewart, Hamish I., Dmitry Grinfeld, Anastassios Giannakopulos, Johannes Petzoldt, Toby Shanley, Matthew Garland, Eduard Denisov, et al. "Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis." *Analytical Chemistry* 95, no. 42 (October 24, 2023): 15656–64. https://doi.org/10.1021/acs.analchem.3c02856.

Storey, John D., Wenzhong Xiao, Jeffrey T. Leek, Ronald G. Tompkins, and Ronald W. Davis. "Significance Analysis of Time Course Microarray Experiments." *Proceedings of the National Academy of Sciences* 102, no. 36 (September 6, 2005): 12837–42. https://doi.org/10.1073/pnas.0504609102.

Subramanian, Aravind, Heidi Kuehn, Joshua Gould, Pablo Tamayo, and Jill P. Mesirov. "GSEA-P: A Desktop Application for Gene Set Enrichment Analysis." *Bioinformatics* 23, no. 23 (December 1, 2007): 3251–53. https://doi.org/10.1093/bioinformatics/btm369.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102, no. 43 (October 25, 2005): 15545–50. https://doi.org/10.1073/pnas.0506580102.

———. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102, no. 43 (October 25, 2005): 15545–50. https://doi.org/10.1073/pnas.0506580102.

Sundararajan, Tharani, Ann M Manzardo, and Merlin G Butler. "Functional Analysis of Schizophrenia Genes Using GeneAnalytics Program and Integrated Databases." *Gene* 641 (January 30, 2018): 25–34. https://doi.org/10.1016/j.gene.2017.10.035.

Swaney, Danielle L., Dana J. Ramms, Zhiyong Wang, Jisoo Park, Yusuke Goto, Margaret Soucheray, Neil Bhola, et al. "A Protein Network Map of Head and Neck Cancer Reveals PIK3CA Mutant Drug Sensitivity." *Science* 374, no. 6563 (October 2021): eabf2911. https://doi.org/10.1126/science.abf2911.

Szklarczyk, Damian, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, et al. "The STRING Database in 2021: Customizable Protein–Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets." *Nucleic Acids Research* 49, no. D1 (January 8, 2021): D605–12. https://doi.org/10.1093/nar/gkaa1074.

Tanaka, Koichi, Hiroaki Waki, Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida, Tamio Yoshida, and T. Matsuo. "Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry." *Rapid Communications in Mass Spectrometry* 2, no. 8 (1988): 151–53. https://doi.org/10.1002/rcm.1290020802.

Tarpey, Patrick S, Raffaella Smith, Erin Pleasance, Annabel Whibley, Sarah Edkins, Claire Hardy, Sarah O'Meara, et al. "A Systematic, Large-Scale Resequencing

Screen of X-Chromosome Coding Exons in Mental Retardation." *Nat. Genet.* 41, no. 5 (May 2009): 535–43. https://doi.org/10.1038/ng.367.

Thul, Peter J., Lovisa Åkesson, Mikaela Wiking, Diana Mahdessian, Aikaterini Geladaki, Hammou Ait Blal, Tove Alm, et al. "A Subcellular Map of the Human Proteome." *Science* 356, no. 6340 (May 26, 2017): eaal3321. https://doi.org/10.1126/science.aal3321.

Ting, Lily, Mark J. Cowley, Seah Lay Hoon, Michael Guilhaus, Mark J. Raftery, and Ricardo Cavicchioli. "Normalization and Statistical Analysis of Quantitative Proteomics Data Generated by Metabolic Labeling." *Molecular & Cellular Proteomics: MCP* 8, no. 10 (October 2009): 2227–42. https://doi.org/10.1074/mcp.M800462-MCP200.

Toby, Timothy K., Luca Fornelli, Kristina Srzentić, Caroline J. DeHart, Josh Levitsky, John Friedewald, and Neil L. Kelleher. "A Comprehensive Pipeline for Translational Top-down Proteomics from a Single Blood Draw." *Nature Protocols* 14, no. 1 (January 2019): 119–52. https://doi.org/10.1038/s41596-018-0085-7.

Traag, V. A., L. Waltman, and N. J. van Eck. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports* 9, no. 1 (March 26, 2019): 5233. https://doi.org/10.1038/s41598-019-41695-z.

Trindade, Pablo, Erick Correia Loiola, Juciano Gasparotto, Camila Tiefensee Ribeiro, Pablo Leal Cardozo, Sylvie Devalle, José Alexandre Salerno, et al. "Short and Long TNF-Alpha Exposure Recapitulates Canonical Astrogliosis Events in Human-Induced Pluripotent Stem Cells-Derived Astrocytes." *Glia* 68, no. 7 (July 2020): 1396–1409. https://doi.org/10.1002/glia.23786.

Tripathi, Shashank, Marie O. Pohl, Yingyao Zhou, Ariel Rodriguez-Frandsen, Guojun Wang, David A. Stein, Hong M. Moulton, et al. "Meta- and Orthogonal Integration of Influenza 'OMICs' Data Defines a Role for UBR4 in Virus Budding." *Cell Host & Microbe* 18, no. 6 (December 9, 2015): 723–35. https://doi.org/10.1016/j.chom.2015.11.002.

Trujillo, Cleber A., Telma T. Schwindt, Antonio H. Martins, Janaína M. Alves, Luiz Eugênio Mello, and Henning Ulrich. "Novel Perspectives of Neural Stem Cell Differentiation: From Neurotransmitters to Therapeutics." *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* 75, no. 1 (January 2009): 38–53. https://doi.org/10.1002/cyto.a.20666.

Tyanova, Stefka, and Juergen Cox. "Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research." In *Cancer Systems Biology: Methods and Protocols*, edited by Louise von Stechow, 133–48. Methods in Molecular Biology. New York, NY: Springer, 2018. https://doi.org/10.1007/978-1-4939-7493-1_7.

Tyanova, Stefka, Tikira Temu, and Juergen Cox. "The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics." *Nature Protocols* 11, no. 12 (December 2016): 2301–19. https://doi.org/10.1038/nprot.2016.136.

Tyanova, Stefka, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. "The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data." *Nature Methods* 13, no. 9 (September 2016): 731–40. https://doi.org/10.1038/nmeth.3901.

Uranova, Natalya A, Manuel F Casanova, Nathan M DeVaughn, Diana D Orlovskaya, and Dmitry V Denisov. "Ultrastructural Alterations of Synaptic Contacts and Astrocytes in Postmortem Caudate Nucleus of Schizophrenic Patients."

Schizophr. Res. 22, no. 1 (October 1996): 81–83. https://doi.org/10.1016/0920-9964(96)00059-X.

Uranova, Natalya A, Victor M Vostrikov, Olga V Vikhreva, Ivetta S Zimina, Natalya S Kolomeets, and Diana D Orlovskaya. "The Role of Oligodendrocyte Pathology in Schizophrenia." *Int. J. Neuropsychopharmacol.* 10, no. 4 (August 2007): 537–45. https://doi.org/10.1017/S1461145707007626.

Uranova, Natalya, Diana Orlovskaya, Olga Vikhreva, Ivetta Zimina, Natalya Kolomeets, Victor Vostrikov, and Valentina Rachmanova. "Electron Microscopy of Oligodendroglia in Severe Mental Illness." *Brain Res. Bull.* 55, no. 5 (July 2001): 597–610. https://doi.org/10.1016/S0361-9230(01)00528-7.

Välikangas, Tommi, Tomi Suomi, and Laura L Elo. "A Comprehensive Evaluation of Popular Proteomics Software Workflows for Label-Free Proteome Quantification and Imputation." *Briefings in Bioinformatics* 19, no. 6 (November 27, 2018): 1344–55. https://doi.org/10.1093/bib/bbx054.

VanderPlas, Jacob, Brian E. Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. "Altair: Interactive Statistical Visualizations for Python." *Journal of Open Source Software* 3, no. 32 (December 10, 2018): 1057. https://doi.org/10.21105/joss.01057.

Varatharaj, Aravinthan, Naomi Thomas, Mark A. Ellul, Nicholas W. S. Davies, Thomas A. Pollak, Elizabeth L. Tenorio, Mustafa Sultan, et al. "Neurological and Neuropsychiatric Complications of COVID-19 in 153 Patients: A UK-Wide Surveillance Study." *The Lancet. Psychiatry* 7, no. 10 (October 2020): 875–82. https://doi.org/10.1016/S2215-0366(20)30287-X.

Velásquez, Erika, Fabio C S Nogueira, Ingrid Velásquez, Andrea Schmitt, Peter Falkai, Gilberto B Domont, and Daniel Martins-de-Souza. "Synaptosomal Proteome of the Orbitofrontal Cortex from Schizophrenia Patients Using Quantitative Label-Free and iTRAQ-Based Shotgun Proteomics." *J. Proteome Res.* 16, no. 12 (December 1, 2017): 4481–94. https://doi.org/10.1021/acs.jproteome.7b00422.

Venable, John D., Meng-Qiu Dong, James Wohlschlegel, Andrew Dillin, and John R. Yates. "Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra." *Nature Methods* 1, no. 1 (October 2004): 39–45. https://doi.org/10.1038/nmeth705.

Volk, David W, Annie E Moroco, Kaitlyn M Roman, Jessica R Edelson, and David A Lewis. "The Role of the Nuclear Factor-κB Transcriptional Complex in Cortical Immune Activation in Schizophrenia." *Biol. Psychiatry* 85, no. 1 (January 1, 2019): 25–34. https://doi.org/10.1016/j.biopsych.2018.06.015.

Vostrikov, V M, and N A Uranova. "Reduced Density of Oligodendrocytes and Oligodendrocyte Clusters in the Caudate Nucleus in Major Psychiatric Illnesses." *Schizophr. Res.*, October 22, 2019. https://doi.org/10.1016/j.schres.2019.10.027.

Wang, Si, Xiaohong Yao, Shuai Ma, Yifang Ping, Yanling Fan, Shuhui Sun, Zhicheng He, et al. "A Single-Cell Transcriptomic Landscape of the Lungs of Patients with COVID-19." *Nature Cell Biology* 23, no. 12 (December 2021): 1314–28. https://doi.org/10.1038/s41556-021-00796-6.

Wang, Yijin, Shuhong Liu, Hongyang Liu, Wei Li, Fang Lin, Lina Jiang, Xi Li, et al. "SARS-CoV-2 Infection of the Liver Directly Contributes to Hepatic Impairment in Patients with COVID-19." *Journal of Hepatology* 73, no. 4 (October 2020): 807–16. https://doi.org/10.1016/j.jhep.2020.05.002.

Wasinger, Valerie C., Stuart J. Cordwell, Anne Cerpa-Poljak, Jun X. Yan, Andrew A. Gooley, Marc R. Wilkins, Mark W. Duncan, Ray Harris, Keith L. Williams, and Ian Humphery-Smith. "Progress with Gene-Product Mapping of the Mollicutes: Mycoplasma Genitalium." *ELECTROPHORESIS* 16, no. 1 (1995): 1090–94. https://doi.org/10.1002/elps.11501601185.

White, Joseph A., Thomas J. Krzystek, Hayley Hoffmar-Glennon, Claire Thant, Katherine Zimmerman, Gary Iacobucci, Julia Vail, Layne Thurston, Saad Rahman, and Shermali Gunawardena. "Excess Rab4 Rescues Synaptic and Behavioral Dysfunction Caused by Defective HTT-Rab4 Axonal Transport in Huntington's Disease." *Acta Neuropathologica Communications* 8, no. 1 (July 1, 2020): 97. https://doi.org/10.1186/s40478-020-00964-z.

Whitfield-Gabrieli, Susan, Heidi W Thermenos, Snezana Milanovic, Ming T Tsuang, Stephen V Faraone, Robert W McCarley, Martha E Shenton, et al. "Hyperactivity and Hyperconnectivity of the Default Network in Schizophrenia and in First-Degree Relatives of Persons with Schizophrenia." *Proc. Natl. Acad. Sci. U. S. A.* 106, no. 4 (January 27, 2009): 1279–84. https://doi.org/10.1073/pnas.0809141106.

Whitton, Alexis E, Michael T Treadway, and Diego A Pizzagalli. "Reward Processing Dysfunction in Major Depression, Bipolar Disorder and Schizophrenia." *Curr. Opin. Psychiatry* 28, no. 1 (January 2015): 7–12. https://doi.org/10.1097/YCO.0000000000000122.

WHO. "Schizophrenia." Accessed May 1, 2019. https://www.who.int/news-room/fact-sheets/detail/schizophrenia.

Wilk, Aaron J., Madeline J. Lee, Bei Wei, Benjamin Parks, Ruoxi Pi, Giovanny J. Martínez-Colón, Thanmayi Ranganath, et al. "Multi-Omic Profiling Reveals Widespread Dysregulation of Innate Immunity and Hematopoiesis in COVID-19." *The Journal of Experimental Medicine* 218, no. 8 (August 2, 2021): e20210582. https://doi.org/10.1084/jem.20210582.

Wilm, Matthias, and Matthias Mann. "Analytical Properties of the Nanoelectrospray Ion Source." *Analytical Chemistry* 68, no. 1 (January 1, 1996): 1–8. https://doi.org/10.1021/ac9509519.

Won, Joungha, Solji Lee, Myungsun Park, Tai Young Kim, Mingu Gordon Park, Byung Yoon Choi, Dongwan Kim, Hyeshik Chang, V. Narry Kim, and C. Justin Lee. "Development of a Laboratory-Safe and Low-Cost Detection Protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19)." *Experimental Neurobiology* 29, no. 2 (April 30, 2020): 107–19. https://doi.org/10.5607/en20009.

Wu, Tianzhi, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, et al. "clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data." *The Innovation* 2, no. 3 (August 28, 2021): 100141. https://doi.org/10.1016/j.xinn.2021.100141.

Wu, Zhiwei, Xiang Yang Zhang, Huanhuan Wang, Wei Tang, Yu Xia, Feixue Zhang, Jiahong Liu, et al. "Elevated Plasma Superoxide Dismutase in First-Episode and Drug Naive Patients with Schizophrenia: Inverse Association with Positive Symptoms." *Prog. Neuropsychopharmacol. Biol. Psychiatry* 36, no. 1 (January 10, 2012): 34–38. https://doi.org/10.1016/j.pnpbp.2011.08.018.

Xiao, Fei, Meiwen Tang, Xiaobin Zheng, Ye Liu, Xiaofeng Li, and Hong Shan. "Evidence for Gastrointestinal Infection of SARS-CoV-2." *Gastroenterology* 158, no. 6 (May 2020): 1831-1833.e3. https://doi.org/10.1053/j.gastro.2020.02.055.

Xicoy, Helena, Bé Wieringa, and Gerard J. M. Martens. "The SH-SY5Y Cell Line in Parkinson's Disease Research: A Systematic Review." *Molecular*

Neurodegeneration 12, no. 1 (January 24, 2017): 10. https://doi.org/10.1186/s13024-017-0149-0.

Yachou, Yassine, Abdeslem El Idrissi, Vladimir Belapasov, and Said Ait Benali. "Neuroinvasion, Neurotropic, and Neuroinflammatory Events of SARS-CoV-2: Understanding the Neurological Manifestations in COVID-19 Patients." *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 41, no. 10 (October 2020): 2657–69. https://doi.org/10.1007/s10072-020-04575-3.

Yan, Yiping, Soojung Shin, Balendu Shekhar Jha, Qiuyue Liu, Jianting Sheng, Fuhai Li, Ming Zhan, et al. "Efficient and Rapid Derivation of Primitive Neural Stem Cells and Generation of Brain Subtype Neurons from Human Pluripotent Stem Cells." *Stem Cells Translational Medicine* 2, no. 11 (November 2013): 862–70. https://doi.org/10.5966/sctm.2013-0080.

Yang, Li, Shasha Liu, Jinyan Liu, Zhixin Zhang, Xiaochun Wan, Bo Huang, Youhai Chen, and Yi Zhang. "COVID-19: Immunopathogenesis and Immunotherapeutics." *Signal Transduction and Targeted Therapy* 5, no. 1 (July 25, 2020): 128. https://doi.org/10.1038/s41392-020-00243-2.

Yuan, Peixiong, Rulun Zhou, Yun Wang, Xiaoxia Li, Jianling Li, Guang Chen, Xavier Guitart, and Husseini K Manji. "Altered Levels of Extracellular Signal-Regulated Kinase Signaling Proteins in Postmortem Frontal Cortex of Individuals with Mood Disorders and Schizophrenia." *J. Affect. Disord.* 124, no. 1–2 (July 2010): 164–69. https://doi.org/10.1016/j.jad.2009.10.017.

Zhang, Hui, Hong-Bao Li, Jian-Rui Lyu, Xiao-Ming Lei, Wei Li, Gang Wu, Jun Lyu, and Zhi-Ming Dai. "Specific ACE2 Expression in Small Intestinal Enterocytes May Cause Gastrointestinal Symptoms and Injury after 2019-nCoV Infection." *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 96 (July 2020): 19–24. https://doi.org/10.1016/j.ijid.2020.04.027.

Zhang, Jiyong, Grazyna B. Sadowska, Xiaodi Chen, Seon Yeong Park, Jeong-Eun Kim, Courtney A. Bodge, Erin Cummings, et al. "Anti-IL-6 Neutralizing Antibody Modulates Blood-Brain Barrier Function in the Ovine Fetus." *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 29, no. 5 (May 2015): 1739–53. https://doi.org/10.1096/fj.14-258822.

Zhang, Xiang Yang, Da Chun Chen, Mei Hong Xiu, Wei Tang, Feixue Zhang, Lianjing Liu, Yuanling Chen, et al. "Plasma Total Antioxidant Status and Cognitive Impairments in Schizophrenia." *Schizophr. Res.* 139, no. 1–3 (August 2012): 66–72. https://doi.org/10.1016/j.schres.2012.04.009.

Zhang, Xiaonan, Yun Tan, Yun Ling, Gang Lu, Feng Liu, Zhigang Yi, Xiaofang Jia, et al. "Viral and Host Factors Related to the Clinical Outcome of COVID-19." *Nature* 583, no. 7816 (July 2020): 437–40. https://doi.org/10.1038/s41586-020-2355-0.

Zhang, Yaoyang, Bryan R. Fonslow, Bing Shan, Moon-Chang Baek, and John R. Yates. "Protein Analysis by Shotgun/Bottom-up Proteomics." *Chemical Reviews* 113, no. 4 (April 10, 2013): 2343–94. https://doi.org/10.1021/cr3003533.

Zhou, Lulin, Zubiao Niu, Xiaoyi Jiang, Zhengrong Zhang, You Zheng, Zhongyi Wang, Yichao Zhu, et al. "SARS-CoV-2 Targets by the pscRNA Profiling of ACE2, TMPRSS2 and Furin Proteases." *iScience* 23, no. 11 (November 20, 2020): 101744. https://doi.org/10.1016/j.isci.2020.101744.

Zhou, Yingyao, Bin Zhou, Lars Pache, Max Chang, Alireza Hadj Khodabakhshi, Olga Tanaseichuk, Christopher Benner, and Sumit K. Chanda. "Metascape Provides a

Biologist-Oriented Resource for the Analysis of Systems-Level Datasets." *Nature Communications* 10, no. 1 (April 3, 2019): 1523. https://doi.org/10.1038/s41467-019-09234-6.

Zickler, Martin, Stephanie Stanelle-Bertram, Sandra Ehret, Fabian Heinrich, Philine Lange, Berfin Schaumburg, Nancy Mounogou Kouassi, et al. "Replication of SARS-CoV-2 in Adipose Tissue Determines Organ and Systemic Lipid Metabolism in Hamsters and Humans." *Cell Metabolism* 34, no. 1 (January 4, 2022): 1–2. https://doi.org/10.1016/j.cmet.2021.12.002.

Zou, Xin, Ke Chen, Jiawei Zou, Peiyi Han, Jie Hao, and Zeguang Han. "Single-Cell RNA-Seq Data Analysis on the Receptor ACE2 Expression Reveals the Potential Risk of Different Human Organs Vulnerable to 2019-nCoV Infection." *Frontiers of Medicine* 14, no. 2 (April 2020): 185–92. https://doi.org/10.1007/s11684-020-0754-0.

Zybailov, Boris, Amber L. Mosley, Mihaela E. Sardiu, Michael K. Coleman, Laurence Florens, and Michael P. Washburn. "Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces c Erevisiae*." *Journal of Proteome Research* 5, no. 9 (September 1, 2006): 2339–47. https://doi.org/10.1021/pr060161n.
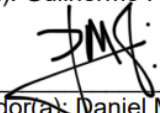
7 ETHICS COMMITTEE STATEMENT

COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br

UNICAMP

IB

DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada "*OmicScope: from quantitative proteomics to systems biology*", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: _____
Nome do(a) aluno(a): Guilherme Reis-de-Oliveira

Assinatura: _____
Nome do(a) orientador(a): Daniel Martins-de-Souza

Data: 18/03/2024

8 COPYRIGHT STATEMENT

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **OmicScope: from quantitative proteomics to systems biology**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.
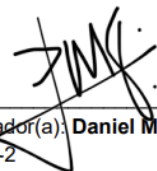
Campinas, 18 de Março de 2024

Assinatura : _____

Nome do(a) autor(a): **Guilherme Reis-de-Oliveira**

RG n.° 48.713.685-8

Assinatura : _____

Nome do(a) orientador(a): **Daniel Martins-de-Souza**

RG n.° 32.431.379-2