Unraveling the Co-Morbidity between COVID-19 and Neurodegenerative Diseases Through Multi-scale Graph Analysis: A Systematic Investigation of Biological Databases and Text Mining

Negin Sadat Babaiha[1], Stefan Geissler[2], Vincent Nibart[2], Heval Atas Güvenilir[1], Vinay Srinivas Bharadhwaj[1], Alpha Tom Kodamullil[1,3], Juergen Klein[1], Marc Jacobs[1], Martin Hofmann-Apitius[1, 4, *]

1. Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven 53757 Sankt Augustin, Germany.
2. Kairntech SAS, 29 Chemin du Vieux Chêne 38240 Meylan, France
3. Causality Biomodels, Kinfra Hi-Tech Park, Kalamassery, Cochin 683503, Kerala, India.
4. Bonn-Aachen International Center for Information Technology (b-it), University of Bonn, 53115 Bonn, Germany

* Corresponding author(s):
Martin Hofmann-Apitius
martin.hofmann-apitius@scai.fraunhofer.de

Abstract

The COVID-19 pandemic has produced an overwhelming volume of research, yet much of it remains focused on individual diseases, largely ignoring the complex relationships between comorbidities. Although extensive literature exists on both neurodegenerative diseases (NDDs), namely Alzheimer's and Parkinson's, and COVID-19, their intersection remains underexplored. Co-morbidity modeling is essential, as patients, particularly those hospitalized, often present with multiple conditions. This study addresses this gap by investigating the crosstalk between COVID-19 and NDDs using a combination of knowledge graphs built from curated biomedical datasets and text mining tools. We conducted comprehensive graph analyses, including path analysis, phenotype coverage assessment, and mapping of cellular and genetic factors, to examine how various Knowledge Graphs (KGs)—such as PrimeKG,

DrugBank, OpenTargets, and those generated from natural language processing (NLP) methodologies—illuminate molecular and phenotypic relationships between these diseases. Our findings reveal significant variability in graph density and connectivity across datasets, each offering unique insights into the landscape of COVID-19 and NDD co-morbidities. By integrating structured biological data with unstructured textual data, this study aimed to optimize co-morbidity modeling, maximize recall for identifying potential co-morbidity mechanisms, and consolidate this information in a dedicated co-morbidity hypothesis database. Key genetic and inflammatory markers, particularly immune response genes, emerged as consistent features across multiple KGs, reinforcing their potential role in COVID-19–NDD interactions. This integrative approach advances our understanding of the underlying mechanisms linking these diseases and facilitates the identification of potential therapeutic targets.

All data, methodologies, and detailed instructions for accessing the co-morbidity hypothesis database are publicly available and thoroughly documented at: https://github.com/SCAI-BIO/covid-NDD-comorbidity-NLP .

Keywords: COVID-19, Neurodegenerative Diseases, Co-morbidity Analysis, Knowledge Graphs (KG), Natural Language Processing (NLP), Biological Databases

1 Introduction

The COVID-19 pandemic, which emerged in late 2019, was swiftly declared an international health emergency by the World Health Organization (WHO) due to its rapid global spread and severe impact on public health [1], [2], [3]. The disease, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), resulted in severe mortality and emergency cases worldwide. The COVID-19 pandemic has posed a complex and multifaceted challenge, overwhelming global healthcare systems with an unprecedented surge in patient numbers and severe resource shortages, while also triggering widespread social and economic disruptions across the globe [4], [5]. Besides, the pervasive uncertainty surrounding the virus and its effects has exacerbated mental health issues, with a marked increase in cases of depression, anxiety, and burnout among the global population [1], [6].

COVID-19 disease primarily targets the respiratory system, but it has also been associated with a wide range of symptoms and complications affecting multiple organ systems such as the cardiovascular and immune system [1], [7]. Recent studies have revealed complex immune dysregulation in COVID-19, particularly in severe cases [8]. Macrophages and monocytes, key components of the innate immune system, play critical roles in pathogen recognition and inflammation control [9]. Macrophages, as tissue-resident cells, initiate local immune responses, while monocytes are recruited from blood during infection. The disease progression follows a time-dependent model where initial interferon response is crucial for viral control [8]. However, delayed or prolonged responses can trigger hyperinflammation, characterized by excessive mononuclear phagocyte activation and dysregulated tissue repair. Severe cases exhibit distinct immunological changes, including elevated pro-inflammatory cytokines, neutrophilia, lymphopenia, and reduced HLA-DR expression on monocytes [8].

Beyond its well-documented respiratory manifestations and immune dysregulation, from the start of the pandemic, the disease has been linked to a range of neurological complications [10]. Early in the pandemic, reports began to surface detailing these complications, which varied from mild symptoms such as headaches and myalgia [11], to more severe and potentially life-threatening conditions, including seizures, strokes, and Guillain-Barré syndrome [12]. Subsequent in vitro studies indicated the potential for SARS-CoV-2 to infect neurons and astrocytes, raising concerns about direct viral involvement in neurological dysfunction [13]. However, findings from autopsy studies suggest that such dysfunction is unlikely to result from direct viral invasion in vivo [12], [14]. Instead, these studies propose that the virus affects the brain indirectly, potentially through mechanisms such as immune cell activation, the release of peripherally generated inflammatory mediators, or alterations in the blood-brain barrier [12], [13], [14], [15]. Nonetheless, preliminary evidence hints towards a direct aggregation potential for soluble amyloid beta of SARS-CoV-2, and HIV, as highlighted in a recent preprint [16]. In their study, they demonstrate that certain viruses, including SARS-CoV-2 and HSV-1, can induce the aggregation of proteins involved in neurodegenerative diseases (NDDs) through a physicochemical process known as heterogeneous nucleation (HEN). This mechanism allows viruses to catalyze protein clumping in cerebrospinal fluid (CSF) without requiring viral replication, suggesting that these viral infections may directly contribute to the development of NDDs by triggering protein aggregation pathways [16]. Recent comprehensive analyses have further elucidated the complex relationship between COVID-19 and NDDs [17]. More than two-thirds of hospitalized COVID-19 patients experience incomplete recovery even months

after infection, with specific implications for neurological health. As mentioned in their study, the virus' neurological impact appears to operate through five distinct pathways: the olfactory epithelial route, blood-brain barrier penetration, lateral ventricles and choroid plexus involvement, vagus nerve transmission, and corneal epithelial pathway [17]. These multiple entry routes may explain some of the virus´ diverse neurological effects. Particularly concerning is the virus's interaction with existing neurodegenerative conditions: Alzheimer's disease patients show increased COVID-19 susceptibility and mortality risk, while Parkinson's disease patients demonstrate a 58% higher hospitalization rate compared to healthy controls [17]. In multiple sclerosis patients, approximately 29% experience prolonged COVID-19 symptoms lasting over four weeks, predominantly manifesting as fatigue [17]. The emergence of sophisticated diagnostic approaches, including FDG-PET imaging and EEG monitoring, has enabled better assessment of these neurological impacts. While direct causal relationships between SARS-CoV-2 and NDDs remain under investigation, mounting evidence suggests that COVID-19 may act as both a trigger for new neurodegenerative conditions and an accelerator of existing ones, particularly through mechanisms of sustained inflammation and immune response dysregulation [17].

Given these findings, there is a growing interest in understanding the relationship between COVID-19 and the development of NDDs, such as Alzheimer’s disease (AD). To investigate the mechanisms by which COVID-19 may cause neurological complications, a research group conducted a study focused on the neurological aspects of COVID-19 under a project known as CIAO (https://www.ciao-covid.net). This project aimed to uncover the existing knowledge on COVID-19 pathogenesis by utilizing the Adverse Outcome Pathway (AOP) framework [18]. The AOP framework systematically maps disease progression from an initial molecular event, such as SARS-CoV-2 infection, through key biological processes, ultimately leading to adverse outcomes like respiratory distress, organ failure, insomnia, or cognitive impairments such as brain fog [19]. However, the rapidly evolving variants of COVID-19 and the broad spectrum of clinical symptoms present significant challenges to applying this framework effectively. Given the vast amount of research, clinical trials, and data generated in the years following the COVID-19 pandemic, there is a wealth of valuable information that can be leveraged to better understand the underlying mechanisms. The large volume of unstructured textual data makes natural language processing (NLP) and text mining particularly valuable tools in this endeavor. For instance, a recent study employed NLP and explored the corpora around neurological toxicity and COVID-19 by developing a language model named NeuroCORD [20]. This model

leverages Bidirectional Encoder Representations from Transformers (BERT) embeddings and has been trained on a corpus of literature abstracts to classify and distinguish COVID-19 articles that focus on neurological disorders from those covering other topics [20]. In the contentious debate over COVID-19 vaccinations, a widely cited study used text mining on Twitter (now “X”) data to assess public hesitancy, offering key insights into societal responses to the pandemic [21]. Building on this approach, other researchers have employed techniques like co-occurrence analysis, clustering, and topic modeling, to investigate the common manifestations of COVID-19 and identify potential therapeutic agents [22].

Despite the vast amount of research on NDDs such as AD and Parkinson's (PD), and the rapidly expanding body of literature on COVID-19, there remains a significant gap when it comes to studying the crosstalk between these two conditions. The dominance of "one-disease-centric" research presents a challenge for mining information on comorbidities, which are a critical aspect of real-world health dynamics. In clinical settings, patients rarely present with a single disease; comorbidities are the norm, particularly in those hospitalized. Early studies of hospitalized COVID-19 patients, conducted through the ISARIC4C protocol [23], revealed that over three-quarters had at least one co-morbidity, with conditions such as cardiac disease, pulmonary disease, chronic kidney disease, obesity, cancer, neurological disorders, dementia, and liver disease all associated with higher in-hospital mortality [24]. Analysis of primary care data from 40% of the English population further revealed hypertension (34.3%), asthma (15.9%), and diabetes (9.9%) as the most common comorbidities [24]. More recent research has shifted focus to multimorbidity and revealed that among 1,706 severe COVID-19 cases in the UK Biobank, 25.3% of patients had multiple conditions, with stroke and hypertension being the most prevalent combination, and chronic kidney disease paired with diabetes showing the highest associated risk (OR 4.93; 95% CI 3.36-7.22) [24]. In a separate study, Romagnolo et al. investigated the influence of pre-existing neurological conditions on COVID-19 outcomes in a cohort of 332 patients, of whom 22.6% were diagnosed with neurological disorders [25]. Their study found that patients with neurological diseases had a significantly higher case fatality rate (48.0% vs. 24.0%) and were independently associated with increased mortality. Specifically, NDDs were linked to the highest mortality (73.9% vs. 39.1%), while cerebrovascular diseases showed a higher, though non-significant, mortality rate. These patients were also older, had more comorbidities, and presented with more severe COVID-19 symptoms. The unprecedented surge in studies regarding COVID-19-related comorbidities, although offering critical insights, often includes speculation and frequently falls short of

delivering the robust, evidence-based data necessary for drawing definitive conclusions about COVID-19 comorbidities. Most of these insights are derived from large studies that focus only on hospitalized patients, which introduces bias into the findings [24]. Additionally, in general population studies, factors such as non-random sampling may distort the results, further complicating the interpretation of the data [24], [26]. Besides, robust insights into comorbidities between COVID-19 and NDDs primarily come from observational cohort studies, such as UK Biobank [27], or from experimental systems like organoid studies and blood-brain barrier (BBB) models, but these approaches can limit the generalizability of the findings.

Our current study, conducted within the framework of the COMMUTE project (www.commute-project.eu), aims to bridge existing gaps by integrating knowledge from both neurological diseases and COVID-19 research. Utilizing a graph-based approach, we model and mine co-morbidity relationships to gain a deeper understanding of their interactions. This methodology enables us to synthesize a vast and fragmented body of literature into a unified framework, facilitating a comprehensive analysis of existing knowledge stored in databases through the lens of Knowledge Graphs (KGs). By integrating diverse data sources, our approach not only enhances the understanding of the relationships between SARS-CoV-2 infection and NDDs but also offers novel insights into how COVID-19 may elevate the risk of NDDs at both the population and individual levels. The graph-based approach provides several advantages, such as visualizing complex relationships, integrating heterogeneous data sources, identifying hidden patterns, and allowing for predictive modeling of potential co-morbidities.
KGs are structured data formats designed to represent, organize, and analyze complex biological and medical information by capturing intricate relationships between diverse entities. Constructing disease-specific co-morbidity KGs can be approached from several perspectives, including manual curation, database integration, and the use of automated tools. Integrating curated databases provides a robust and reliable method for creating these KGs, especially for complex diseases like COVID-19. Curated databases offer high-quality, validated data, ensuring the accuracy and completeness of the resulting graphs, which enhances their effectiveness in analyzing and understanding disease mechanisms.

For instance, DisGeNET [28] is a widely used database that catalogs disease-gene and disease-variant associations, aggregating information from trusted sources like UniProt [29], ClinVar

[30], and the GWAS Catalog [31]. Integrating curated data from such resources facilitates the creation of detailed, precise disease-gene relationship maps. Databases such as OpenTargets [32], DrugBank [33], and PrimeKG [34] further enrich the depth and utility of knowledge graphs. However, there are limitations to this approach. Relying solely on curated resources can restrict the scope of the KG, as these databases may lag in capturing the latest research findings and emerging trends still under investigation. Moreover, the curation process is inherently subjective and may introduce bias, as it relies on the judgment of curators. Additionally, integrating data from multiple sources can be challenging due to variations in data formats, standards, and ontologies, and curated databases often contain redundant, inconsistent, or outdated records, complicating the integration process despite efforts to address these issues.

In contrast, constructing KGs through text mining of unstructured sources, such as scientific literature, offers a dynamic and complementary approach. Frameworks like the Integrated Network and Dynamical Reasoning Assembler (INDRA) (https://github.com/bgyori/indra) exemplify this method by automating the extraction and assembly of mechanistic models from textual data. Leveraging Natural Language Processing (NLP) techniques like REACH (https://indra.readthedocs.io/en/stable/modules/sources/reach/index.html) and TRIPS (https://indra.readthedocs.io/en/stable/modules/sources/trips/index.html), INDRA identifies entities such as genes and proteins and their interactions, structuring these interactions into standardized statements linked to database identifiers to ensure accuracy. This process allows for the discovery of novel associations that may not yet be captured by curated databases. While text mining offers flexibility and the potential to uncover new associations, it also presents challenges, including the need to continuously optimize and retrain NLP models to adapt to new datasets and terminologies. Furthermore, the inherent variability and ambiguity of natural language can introduce errors, highlighting the need for rigorous curation to maintain the quality and reliability of KGs.

Curated databases and text mining offer complementary strengths, and their integration holds great potential for advancing knowledge graph construction and biomedical discovery. In this study, we combine comprehensive database investigation and text mining to build mechanism-based co-morbidity KGs. By focusing on the mechanistic understanding of co-morbidity between COVID-19 and NDDs, we explore how different data sources and tools contribute to generating testable hypotheses. Our goal is to optimize co-morbidity modeling by leveraging

the strengths of these resources, ultimately creating a comprehensive database of hypotheses that not only reflects current knowledge on co-morbidity mechanisms but also uncovers new insights derived from graph mining across the integrated data.

2 Materials and Methods

Our workflow starts with collecting co-morbidity-specific information on COVID-19 and NDDs, specifically AD and PD, in curated databases. This initial step provides a foundational understanding of potential associations and interactions, ensuring a data-driven approach to subsequent analyses. After collecting the data, the interactions are systematically integrated into KG representations, enabling structured visualization and analysis of the relationships within each data source. These data sources are often stored in diverse formats, necessitating transformation and preprocessing to ensure compatibility with the task at hand. For instance, some databases primarily focus on single relationship types, such as disease-gene associations, while others encompass multiple relationship types across various biological entities. To address these disparities, custom scripts are developed to handle data extraction and preprocessing, manage various files, and construct graphs that accurately reflect the underlying data structure. Once prepared, the processed graphs are uploaded to Neo4j graph database (https://neo4j.com), enabling efficient storage, querying, and analysis of the integrated data. This workflow ensures a cohesive and scalable approach to handling heterogeneous datasets and supports the creation of comprehensive KGs for advanced analyses.

In parallel to the information retrieval from databases, we constructed a co-morbidity KG through information extraction from relevant literature. This began with a search of PubMed for publications addressing COVID-19 and NDD correlations in recent years, as detailed in section 2.1. Once the textual data was retrieved, filtered, and the relevance of the selected publications was validated, NLP techniques were employed to identify key entities and extract their relationships. These techniques enabled the systematic analysis of complex biomedical concepts, facilitating the mapping of interactions and associations critical to understanding the correlations between COVID-19 and NDDs. The resulting KG captures a wide spectrum of entities, and their interactions related to the COVID-19 and NDD co-morbidity.

A comparative analysis was then conducted between the KGs derived from databases and those constructed via text mining. This analysis leveraged graph-based path analysis and reasoning algorithms to uncover central nodes, biomarkers, drug targets, and identifiable disease pathways. An overview of the workflow is shown in Figure 1.

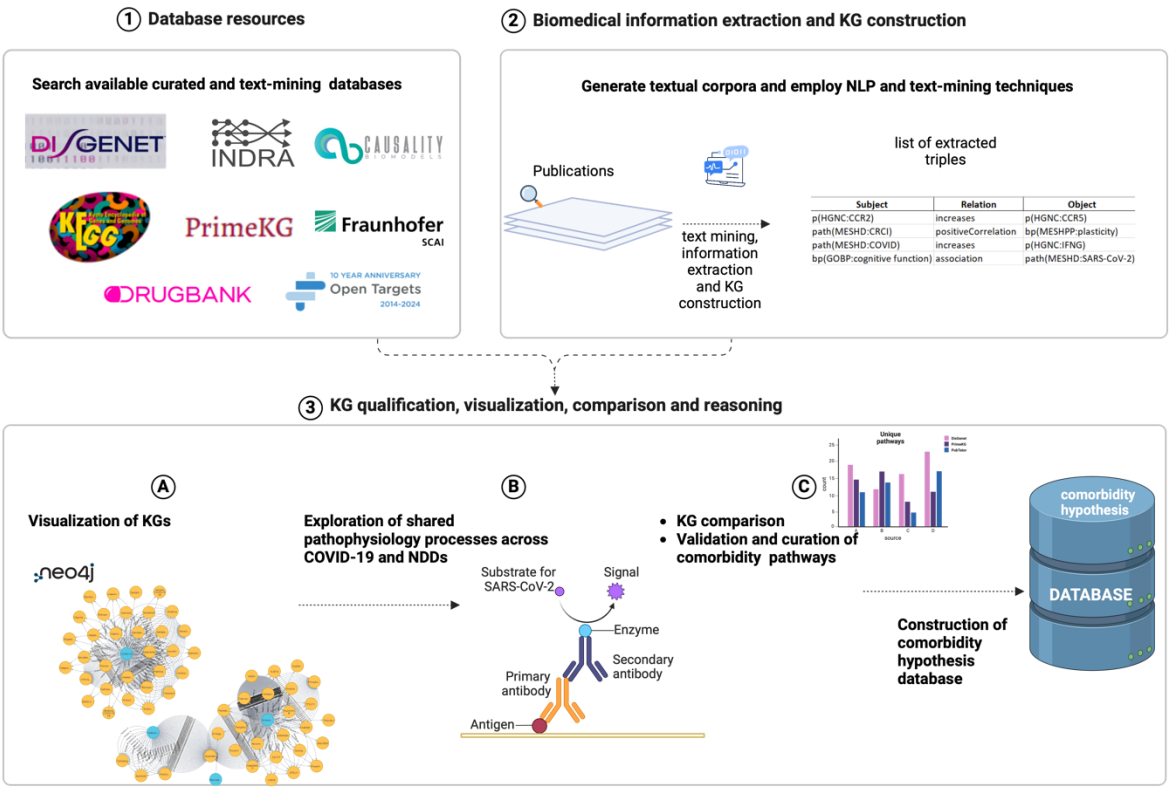

Figure 1. Overview of the co-morbidity KG construction and analysis pipeline. The workflow integrates data on COVID-19 and NDDs from curated databases (1) and PubMed publications to construct KGs (2). Using NLP techniques, relationships between key biomedical entities are extracted and visualized via Neo4j(2,3-A), facilitating the identification of critical nodes, biomarkers, and shared disease pathways (3-B). Comparative analysis of KGs from databases and text mining uncovers insights into the interactions between COVID-19 and NDDs (C). Finally, the candidate hypothesis reflecting potential co-morbidity between COVID-19 and NDDs is stored in a database, which will be further enriched and validated.

2.1 Textual Data

2.1.1 PubMed Search

The overview of textual corpora generation and processing is represented in Figure 2. We started our corpora generation with a focused search in the PubMed database, using a combination of keywords and Medical Subject Headings (MeSH) (https://www.ncbi.nlm.nih.gov/mesh/) terms. We thoroughly reviewed recent relevant literature to find key topics and emerging trends in research on COVID-19 and its effects on the nervous system. We paid special attention to highly relevant and frequently cited papers to identify common keywords and phrases used by experts. We also consulted with specialists in neurology and infectious diseases to make sure our search strategy was accurate and relevant. As a result, we created a specific list of search terms to look for such as combinations like "COVID-19 AND Blood-Brain Barrier Disruption," "COVID-19 AND Neuronal Infection," and "COVID-19 AND Cerebrovascular System," among others (a complete list of the search terms can be found in the Supplementary File, section 1.). We utilized the E-utility public API [35] to retrieve PMIDs for the relevant search term list. To ensure the inclusion of the most current research, we focused on articles published between 2021 and 2024.

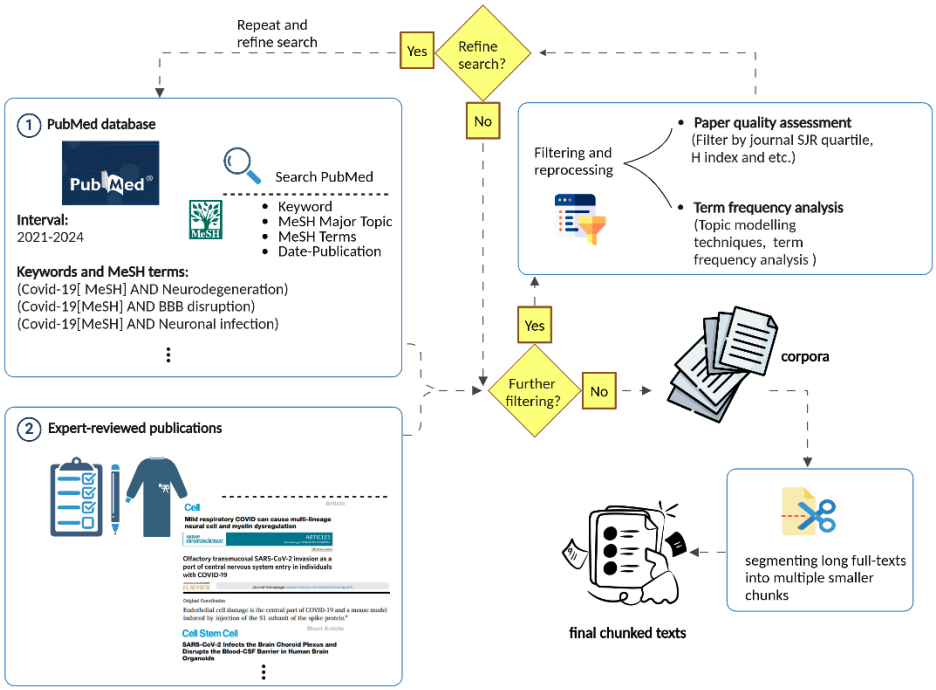

Figure 2. The strategies used to generate and refine a corpus of scientific literature focused on the co-morbidity between COVID-19 and NDDs. Initially, we conducted a search on PubMed using relevant keywords and MeSH terms related to COVID-19-NDD co-morbidity and defined appropriate time intervals. Consultation with domain experts allowed us to select a

highly relevant set of corpora. The extracted literature underwent a post-processing and assessment phase, utilizing topic modeling and journal evaluations. As the search terms were refined and updated, the process became more comprehensive, involving an expanded search on PubMed. Finally, the list of selected publications was divided into manageable chunks, preparing them for further processing by NLP models.

2.1.2 Expert Curation

In addition to conducting an extensive search using the E-utility API, we curated a targeted selection of papers based on expert recommendations. These recommendations were provided by specialists in Clinical Neurology and Infectious Diseases, who are part of the COMMUTE project consortium. This consortium represents leading research institutions from Germany, Spain, Luxembourg, and The Netherlands, ensuring that the selection process was informed by a diverse and highly qualified panel of experts. This combined approach allowed us to capture a comprehensive and focused set of publications relevant to the intersection of COVID-19 and NDDs. A complete list of the PMIDs of these publications is provided in our project GitHub repository data (https://github.com/SCAI-BIO/covid-NDD-comorbidity-NLP).

2.2 Textual Corpora Assessment

To enhance our PubMed search strategy and ensure the most relevant papers were included in our corpus, we reviewed term frequencies to identify any overlooked papers and important terms that were absent from our initial search. We analyzed how often our search terms appeared in the full text of the publications. We observed that some MeSH terms might have been missing because newly published articles had not yet been assigned these terms. We then incorporated these terms into our search, refining our strategy for a more comprehensive review. The corpora generated through keyword and MeSH term searches, however, may still contain irrelevant papers. To further filter the publications and identify the most related ones, we employed several strategies: 1) ranking publications and 2) employment of topic modeling. First, we filtered papers based on their publication ranking in high-quality journals within the domain. Additionally, we applied topic modeling and term frequency analysis to extract the

most pertinent topics for each set of corpora. Details of these approaches are outlined in the Supplementary File, section 2.

2.3 Pre-Processing of Textual Corpora

To process the textual corpus with NLP tools, we employed the BioC API for PubMed Central (PMC) [36] to extract full-text documents. Of the 100 documents targeted, 91 were successfully retrieved and freely accessible. Given the large size of the full-text documents, we optimized our processing by segmenting each document into paragraphs and treating them separately. Each document was typically divided into 100 to 300 paragraphs. This method ensured efficient processing and manageable data handling for our NLP pipeline.

2.4 Co-morbidity KG Construction

The comprehensive collection of co-morbidity KGs was constructed using two main methods, as previously shown in Figure 1. First, we integrated data from trusted biomedical databases and curated sources to gather reliable information. Second, we used NLP and text mining to extract useful data directly from scientific texts. The following sections will explain each method in more detail.

2.4.1 Available Databases and Sources

2.4.1.1 KEGG Pathway Database

We utilized the KEGG Pathway database [37] to gather detailed information on COVID-19 and NDDs. Using the KEGG API, we retrieved relevant data for each disease by first extracting KEGG IDs for each COVID-19, AD, PD, and NDDs (general), and then obtaining comprehensive disease-specific information, including associated pathways, genes, and drugs. This provided valuable insights into the genetic, molecular, and therapeutic aspects of these diseases. The technical details for data extraction from KEGG are added to the Supplementary File, section 3.

2.4.1.2 DisGeNET

DisGeNET, a database that collects over 380,000 gene-disease associations across more than 16,000 genes and 13,000 diseases, was used to gather key information on COVID-19, AD, PD,

and NDDs. By leveraging the DisGeNET R package, we extracted Unified Medical Language System (UMLS) IDs for each disease and then used DisGeNET API to identify relevant disease-gene associations to build a KG for further analysis and comparison.

2.4.1.3 DrugBank

DrugBank, a trusted resource offering comprehensive details on drugs, including chemical properties, mechanisms of action, and protein-gene interactions, was utilized in our study. We accessed DrugBank version 5.1.12, which includes 2,358 approved drugs and nearly 366,000 drug-drug interactions. This data was filtered to focus specifically on drugs associated with COVID-19, AD, PD, and NDDs for further analysis.

2.4.1.4 OpenTargets

OpenTargets is a comprehensive platform that aggregates and refines data from various public sources, scoring and ranking target-disease associations based on genomics, transcriptomics, and other biological data. We utilized OpenTargets (v24.06) to identify agents targeting COVID-19 and NDDs by accessing the platform's user interface and downloading targets for each disease individually for further analysis.

2.4.1.5 INDRA Database

INDRA is a source developed at Harvard Medical School using automated extraction of molecular mechanisms from both text and curated databases. It compiles mechanistic knowledge into a machine-readable format called "statements" and integrates these into the non-redundant knowledge base. Using the INDRA API (v1.0), we extracted disease-related statements for both COVID and NDD and only considered the statements where the confidence score was higher than 0.85.

2.4.1.6 PrimeKG

PrimeKG, introduced in 2023, is a comprehensive graph that integrates multi-dimensional disease-related data from 20 high-quality biomedical resources. It maps over 17,000 diseases with more than 4 million connections, encompassing disease-related proteins, biological processes, pathways, anatomical structures, and therapeutic compounds.

To investigate the relationship between COVID-19 and NDDs, we leveraged PrimeKG's extensive KG by searching for both COVID-19 (e.g., SARS-CoV-2, COVID-19, coronavirus) and NDDs (e.g., Alzheimer's, Parkinson's, ALS). In total, there are x nodes and y edges in the COVID-NDD subgraph extracted from PrimeKG.

2.4.1.7 Causality Biomodels (CBM)

CBM is a company (https://causalitybiomodels.com), specializing in bio-curation and knowledge extraction. [*For transparency reasons:* CBM is a spin-off started by and owned by members of Fraunhofer SCAI.] By extracting semantic information from published sources, CBM develops valuable knowledge models in the life sciences, a capability demonstrated in our previous two papers [38], [39]. The textual corpora discussed in Section 2.1, which were reviewed by domain experts, were manually curated by CBM, resulting in the extraction of over 3,000 mechanism triples encoded in Biological Expression Language (BEL) (https://bel.bio). This curated graph is also used as a gold standard for subsequent comparisons.

2.4.1.8 Manually Curated Disease Maps for COVID-19 and NDDs

We also utilized KGs developed by colleagues at the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI) to analyze the co-morbidity between COVID-19 and NDDs. The COVID-19 KG, developed by Domingo-Fernández et al. in 2021, is a comprehensive cause-and-effect model of COVID-19 pathophysiology, integrating data from over 160 research articles [40]. It contains more than 4,000 nodes and 10,000 relationships representing biological entities such as proteins, genes, chemicals, and biological processes. This graph primarily focuses on host-pathogen interactions, including viral invasion, immune response, comorbidities, and drug-target interactions, making it a vital resource for understanding the virus's biological impact and therapeutic strategies.

In contrast, the NeuroMMSig KG, first developed by the same group in 2017, focuses on NDDs like AD and PD [41]. This mechanism-enrichment graph is encoded in BEL [42] and represents causal relationships between genes, proteins, and biological processes, integrating multimodal data such as genetic, epigenetic, and imaging features. The graph enables the exploration of

disease mechanisms and the identification of potential drug targets by comparing molecular pathways involved in AD and PD.

These COVID-19, AD, and PD KGs, referred to as SCAI-DMaps, are used for further analysis in this study.

2.4.2 Text mining and Natural Language Processing Pipelines

2.4.2.1 Sherpa Kairntech

In this study, we utilized the Sherpa workflow developed by Kairntech [43], for text-based information extraction. Sherpa is a user-friendly, web-based machine learning platform that supports various NLP tasks, including entity recognition and relation extraction using entity fishing (https://github.com/kermitt2/entity-fishing) and openNRE package (https://github.com/thunlp/OpenNRE), respectively. It has been previously employed in our work to extract biological relationships and encode them in BEL [38], [39]. Sherpa was originally trained on a dataset of 39,099 triples, which were compiled by integrating multiple published KGs curated by experts and covering various diseases, including AD, PD, and epilepsy, as detailed in our previous work [38]. For this work, Sherpa was updated with a dataset focusing on COVID-19 and NDDs curated by experts in CBM. The platform achieved an accuracy of 93% after training on this dataset, which included more than 3,000 BEL triples, ensuring high-quality extraction of relationships between biomedical entities for further analysis. More details about the updated pipeline are documented in the Supplementary File, section 4.

2.4.2.2 PubTator3

We also leveraged PubTator 3.0 [44], an advanced AI-driven tool developed by the National Library of Medicine (NLM), to extract comprehensive biomedical information from scientific literature. PubTator processes PubMed abstracts and full-text articles, identifying key biological entities and their relationships across various entity types such as chemical–disease, gene-disease, and chemical–gene. Applying PubTator 3.0 API, we extracted annotations in .xml format, which were then parsed and processed into triples (entity-relationship-entity) and

saved in an Excel file for further analysis in the study. More complementary information about information extraction with PubTator is added to the Supplementary File, section 5.

2.5 Visualization of KGs and Graph Algorithms

To facilitate visualization, navigation, and querying of KGs, we employed an open-source version of the Neo4j platform. Its advanced visualization features enable clear representation and differentiation of various node and relationship types, such as diseases and genes. This capability provides an intuitive and comprehensive view of complex biological interconnections, enhancing both analysis and interpretability. Additionally, Neo4j's advanced capabilities, such as the Cypher query language and the integration of powerful libraries like the Graph Data Science (GDS) library and the Awesome Procedures on Cypher (APOC) library, facilitate efficient querying and manipulation of graph data. Since the Sherpa tool identifies entity IDs extracted from various databases, such as MeSH, we implemented a post-processing step within the Sherpa text mining pipeline to ensure uniformity and compatibility. This step standardizes and normalizes the extracted entities to align with predefined namespaces and ontologies, ensuring consistency and facilitating interoperability. Details of this post-processing workflow are provided in the Supplementary File, sections 6 and 7. Once processed, all the resulting triples from all KGs were systematically loaded into the Neo4j database for further analysis and exploration.

3 Results

3.1 General Overview of the Extracted Co-morbidity KGs

The visualization and analysis of KGs from various data sources, as depicted in Figure 3., offers valuable insights into both the structural characteristics of the KGs and the specific biological contexts they aim to represent. Table 1 and Figure 4. provide a comparative overview of several KGs, highlighting differences in the number of nodes, triples, and their densities.

Focusing on the subgraph of PrimeKG centered on both COVID-19 and NDDs—comprising 628 unique nodes and 41,868 distinct triples—it becomes evident that this versatile graph is designed for scalability and comprehensive analysis. Its relatively high density of 0.1063 reflects a much denser graph structure compared to many others, indicating a tightly connected

network of relationships across many entities. Similarly, other large-scale sources like OpenTargets (15,822 nodes, 30,008 triples) and SCAI-DMaps (6,838 nodes, 12,191 triples) present sparse structures, but OpenTargets has a slightly higher density (0.0002397) than SCAI-DMaps (0.0002607). These KGs are designed to cover broad biological domains, offering extensive information with less dense interconnections.

The CBM graph, with 665 unique nodes and 3,057 triples, provides a moderately sized but sparse KG, with a density of 0.000757. This suggests that while it covers a significant range of entities and relationships, the connections are spread thinly across the dataset. It strikes a balance between scale and specificity, offering a resource that is more expansive than highly specialized graphs like DrugBank, but less dense. CBM is well-suited for exploring broad patterns or integrating with other KGs to generate enriched networks.

In contrast, DrugBank provides a smaller but much denser graph. With only 64 nodes and 715 triples, DrugBank has the highest density (0.1773), indicating a tightly connected graph with a higher number of relationships per node. This focused structure likely represents highly curated information, especially about drug-disease relationships. DisGeNET, another smaller source (588 nodes, 1,172 triples), demonstrates a relatively higher density (0.0067911), highlighting its specialized focus, particularly in gene-disease associations. Both DrugBank and DisGeNET serve as dense, focused KGs that deeply explore disease-specific relationships.

The Sherpa text mining graph, with 758 unique nodes and 1,771 relationships, presents a moderately sized yet specific dataset. Its moderate density of 0.0030864 suggests that it balances meaningful relationships with coverage. Sherpa focuses on regulatory mechanisms and intervention strategies, bridging the gap between the sparsity of large graphs like PrimeKG and the dense connectivity of DrugBank.

When merging different sources into a combined KG, a broader and more comprehensive network emerges. For example, combining data from sources like DisGeNET, OpenTargets, DrugBank, and INDRA results in a graph with a significant number of unique nodes and relationships, especially with OpenTargets contributing 15,822 unique nodes and 30,008 triples. This merged KG offers an expansive view, integrating information about gene-disease associations, pharmacological pathways, and potential interventions, particularly in the context of COVID-19 and neurological complications.

In terms of specificity, the KEGG graph contains fewer unique nodes (100) and triples (88), but is highly curated, focusing on well-established metabolic and signaling pathways. PubTator3 text mining graph, as well, is a small KG derived from biomedical literature, with only 80 unique nodes and 68 relationships, derived mainly from term co-occurrence frequency in scientific research papers related to COVID-19 and NDDs.

The structural diversity across these KGs—ranging from the sparsity of large-scale graphs like PrimeKG and CBM to the density of DrugBank—illustrates the variation of the resources investigated in co-morbidity analysis. Whether offering a broad, exploratory framework (PrimeKG, OpenTargets) or a detailed, specialized network (DrugBank, DisGeNET), each KG provides unique insights into biological interactions in both COVID-19 and NDDs. The figures representing the top 10 nodes with the highest centrality degree for each data source are provided in the Supplementary File. As depicted in these figures, several key nodes reveal important insights into potential comorbidities between neurological disorders and other conditions. "Spinocerebellar ataxia" in PrimeKG emerges as a central node, underscoring its significant role in both motor dysfunction and neurodegenerative comorbidities, suggesting that its involvement might extend beyond movement disorders to other systemic conditions. Additionally, "Fatigue", a prominent symptom in PrimeKG, appears as a central node, potentially linking a wide range of neurological and systemic diseases, including those associated with viral infections like COVID-19, thus highlighting fatigue as a common underlying feature in co-morbidity networks. In SCAI-DMaps, "APP" and "Amyloid-beta" represent critical molecular mechanisms involved in AD, potentially connecting the pathophysiology of AD with other neurological disorders and systemic conditions, suggesting that these biomarkers may play a role in disease interactions. Furthermore, the centrality of "Inflammation" and "Neuronal loss" in PubTator3 and Sherpa highlights the crucial role of inflammation as a driver of comorbidities between COVID-19 and NDDs providing insights into how inflammatory pathways could exacerbate disease progression and worsen patient outcomes. Finally, Wolff-Parkinson-White Syndrome, identified in DisGeNET, suggests the possibility of an overlap between cardiovascular and NDDs, indicating the complex nature of comorbidities and their potential interdependence. This insight points to the need for further exploration into how cardiovascular dysfunction might exacerbate neurological conditions, providing a deeper understanding of disease interactions and informing clinical management strategies for patients with multiple comorbidities.

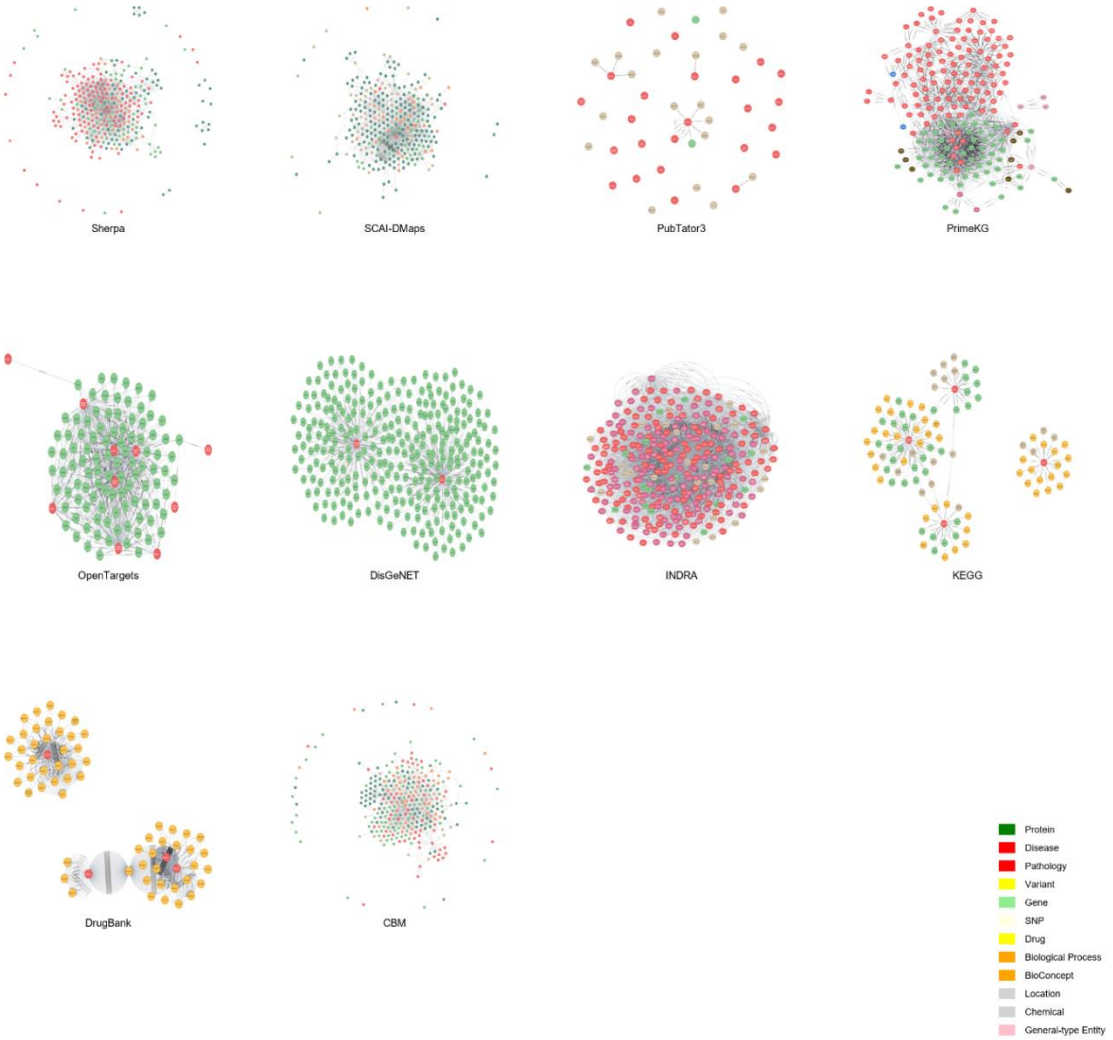

Figure 3. The graph representations in Neo4j provide a comprehensive snapshot of each extracted KG from individual sources. Various node types, such as genes, proteins, diseases, pathologies, and others, are distinguished by unique colors to enhance clarity.

Table 1. General overview of extracted KGs.

Source	Unique_Nodes	Unique_Triples	Density
PrimeKG	628	41868	0,10633
OpenTargets	15822	30008	0,00024
SCAI-DMaps	6838	12191	0,000261
INDRA	851	7137	0,019733
CBM	665	3057	0,000757
Sherpa	758	1771	0,003086

DisGeNET	588	1172	0,006791
DrugBank	64	715	0,177331
KEGG	100	88	0,017778
PubTator3	80	68	0,010759

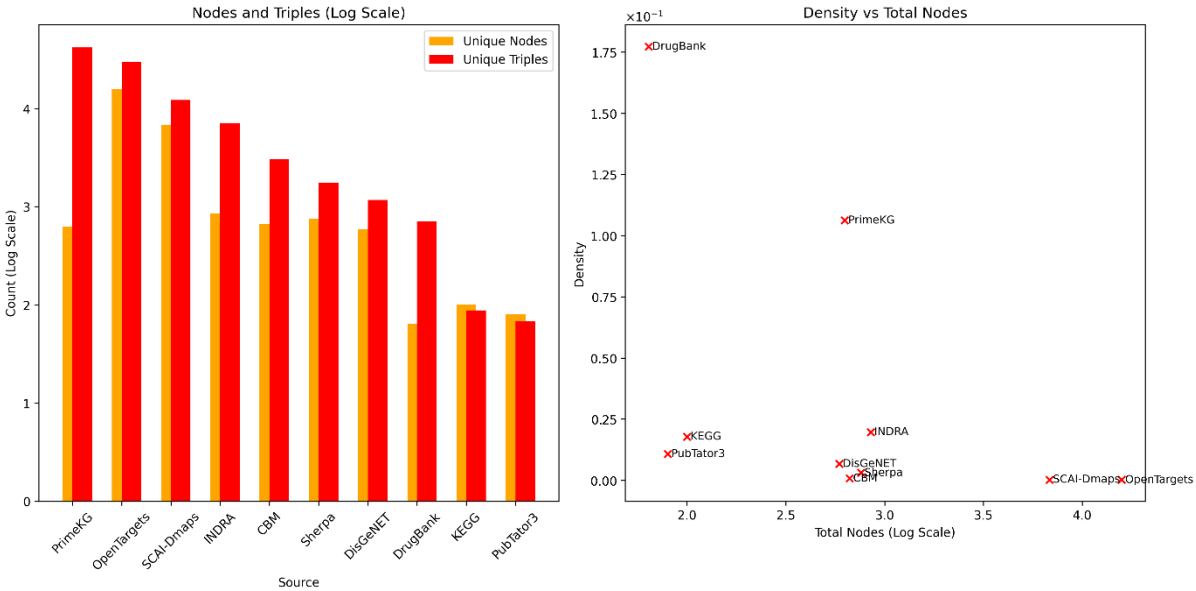

Figure 4. Comparative analysis of COVID-NDD knowledge across different sources. Bar charts (logarithmic scale) show the number of unique nodes and semantic triples from each source, while the scatter plot displays their graph densities relative to total node count, revealing the scale and connectivity patterns of COVID-NDD relationships in each knowledge base.

3.2 Shared Pathophysiology Pathways between COVID-19 and NDD using Shortest Path Analysis

The first step in investigating the interactions between COVID-19 and NDDs was to perform shortest path analysis, aimed at uncovering direct or indirect connections between the two conditions across various graphs. For this means, we developed a custom Cypher query to address the challenges posed by the variations in node names and namespaces across different graphs. This query is designed to uncover connections between COVID-19 and NDDs through

a series of structured and methodical steps, as outlined in the Supplementary File, section 8. A central aspect of our analysis was determining the shortest paths between nodes representing various forms of COVID-19 (e.g., COVID-19, SARS-CoV, and COVID) and NDDs (e.g., Alzheimer, Alzheimer's, Parkinson, Parkinson's, etc.), while accommodating variations in path lengths or hops. The outcomes of this analysis are summarized in Table 2. Initial findings showed that in the KEGG pathway database, no direct or indirect connections existed between COVID-19 and NDD-related genes, drugs, or pathways. Figure 5. visualizes this path analysis across multiple co-morbidity KGs. The combined dataset of DisGeNET, OpenTargets, DrugBank, and INDRA, demonstrated the highest levels of connectivity and the most diverse pathways linking two diseases. PrimeKG ranked second, still showcasing a strong degree of connectivity. These patterns highlight that different KGs vary in their connectivity pattern and require varying step lengths to establish links between COVID-19 and NDD-related entities. CBM represents a middle ground in terms of connectivity, exhibiting steady, moderate growth across hop levels, indicating a stable network with limited expansion potential. In contrast, graphs from SCAI-DMaps and Sherpa show constrained growth, with minimal increases in unique nodes and paths, making them more suitable for focused, targeted analyses rather than studies requiring extensive connectivity. PubTator3, however, has a highly limited co-morbidity network, showing no expansion beyond its initial connections, underscoring its constrained utility for broader co-morbidity analysis.

Table 2. Shortest path analysis between COVID-19 and NDDs nodes using various path lengths (hops).

Sources	Unique Nodes (Hop = 3)	Unique Paths (Hop = 3)	Total Paths (Hop = 3)	Unique Nodes (Hop = 5)	Unique Paths (Hop = 5)	Total Paths (Hop = 5)
DisGeNET-OpenTargets-DrugBank-INDRA	62	195	208	97	798	898
PrimeKG	57	64	64	169	413	415
CBM	25	38	39	32	66	67
SCAI-DMaps	17	14	14	22	44	44
Sherpa	3	6	7	6	10	11
PubTator3	1	1	1	1	1	1

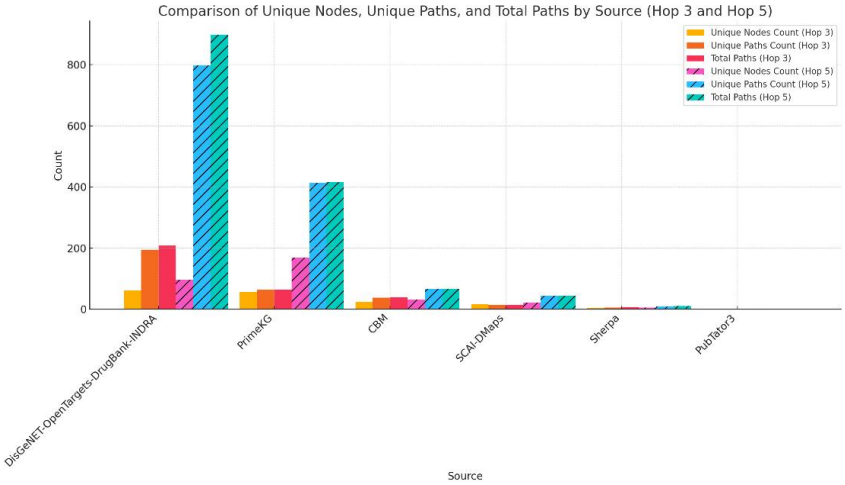

Figure 5. Bar chart summarizing pathway analyses across each co-morbidity KG. For a more comprehensive view, data from the following four sources were integrated for analysis: DisGeNET, OpenTargets, DrugBank, and INDRA. The KEGG pathways database did not reveal any pathways connecting COVID-19-related genes, drugs, or networks.

3.3 Exploring Phenotypic and Clinical Endpoint Coverage in Various Graphs Using Human Phenotype Ontology (HPO) Terms

To validate and evaluate the extent to which phenotypes are represented within the paths connecting COVID-19 and NDDs, we first looked at DisGeNET for genes linked to each COVID-19 and NDDs. Next, we leveraged the Human Phenotype Ontology (HPO) [45] to find phenotypic terms corresponding to each gene associated with these diseases. This process enabled us to extract a set of common phenotypes potentially shared between COVID-19 and various NDDs. Notable examples of these overlapping phenotypes include "Neurofibrillary tangles," "Neuroinflammation," and "Memory impairment," as detailed in Section 8 of the Supplementary File.

To further refine the phenotype information, we cross-referenced the phenotypic terms from the HPO with corresponding terms and synonyms from the MESH database. This comparison

enabled us to compile a comprehensive list of phenotype terms and their synonyms, significantly improving our ability to identify phenotype matches within our KGs. Subsequently, we applied fuzzy matching techniques to detect paths where node names loosely aligned with any phenotype or its synonyms. This approach facilitated a more flexible and thorough analysis of the relationships between COVID-19 and NDDs, capturing subtle variations and alternative representations of phenotype terms within KGs.

The bar chart visualizes this information by summarizing the distribution of shared phenotypes across different KGs, represented as the count of phenotypes in shortest paths connecting COVID-19 and NDDs (Figure 6.). The results highlight a significant variation in phenotype coverage among various KGs. PrimeKG stands out with its notable representation of phenotypes, including key terms such as “Focal Impaired Awareness Cognitive Seizure with Memory Impairment”, “Focal Impaired Awareness Sensory Seizure with Olfactory features”, “Status Epilepticus with Ictal Paresis”, “Neurofibrillary Tangles”, “Central Diabetes Insipidus”, indicating its comprehensive coverage of clinically relevant phenotypes linked to COVID -19 and neurodegenerative processes. Following PrimeKG, the combined dataset DisGeNET-OpenTargets-DrugBank-INDRA also exhibits substantial phenotypic diversity, with emphasis on terms or synonyms of “Progressive Leukoencephalopathy” and “Peripheral Demyelination”. In contrast, CBM displays a more limited set of phenotypic terms, with greater prominence given to phenotypes such as “Focal Impaired Awareness Seizure”, “Upper Motor Neuron Dysfunction”, suggesting a focus on specific neurocognitive impairments. Similarly, Sherpa and SCAI-DMaps have more constrained phenotype coverage reflecting a more focused emphasis on the molecular mechanisms underlying neurodegenerative conditions rather than a broad clinical phenotype spectrum.

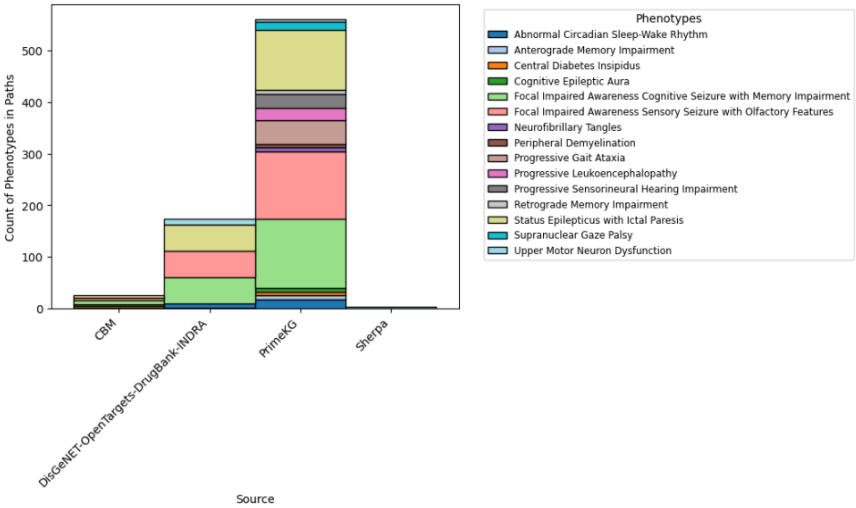

Figure 6. Distribution of shared phenotypes between COVID-19 and NDDs, extracted from path analysis in various KGs. The stacked bar chart displays the frequency of paths containing specific phenotypes associated with both NDD and COVID-19 across different data sources. Each color segment within a bar represents a distinct phenotype identified within the shared path between COVID and NDD using fuzzy matching, highlighting the coverage, diversity, and prevalence of shared phenotypic characteristics across different co-morbidity KGs.

3.4 Construction of the Co-morbidity Hypothesis Database

3.4.1 Co-morbidity Criteria

To investigate the comorbid relationship between COVID-19 and NDDs, we applied a comprehensive strategy based on several criteria to detect relevant triples and pathways from multiple databases and text mining approaches. The first criterion, direct co-occurrence evidence, was defined as the simultaneous presence of COVID-19 and an NDD within the same path within a maximum of 3 hops. The paths extracted this way were then curated and validated manually and added to the hypothesis space. In the second criterion, shared phenotypic outcomes, we identified clinical symptoms or complications common to both diseases, such as “cognitive decline” or “seizure”, with data sourced from the HPO, as formerly described. In fact, by constructing a curated list of common phenotypes, we aimed to identify biological pathways that reveal functional or mechanistic overlaps at the non-molecular level. This process was complemented by independent literature reviews to collect supporting evidence, including associated genes, protein pathways, cell types, genetic risk factors, and biological processes. To ensure accurate pathway identification, we utilized Neo4j’s Levenshtein fuzzy

matching tool, enabling precise recognition of nodes and pathways despite variations in terminology across datasets. The third criterion, common pathophysiological mechanisms, focused on overlapping biological pathways, such as “neuroinflammation”, “oxidative stress”, and “mitochondrial dysfunction”, found in both COVID-19 and NDDs. This was supported by looking through pathway databases like KEGG and Reactome (https://reactome.org), as well as reviewing literature and experimental studies, with examples like IL-6-mediated inflammation and its contribution to both COVID-19 and NDDs [46]. The fourth criterion, shared genetic susceptibility, involved identifying genetic markers linked to both diseases. For instance, the genes NPR3 and TLR7 have been associated with both neurodevelopmental disorders and COVID-19, as evidenced by data from DisGeNET and OpenTargets. The final criterion involved a thorough manual review of potential comorbid mechanisms conducted by domain experts. This step ensured a comprehensive evaluation of the proposed relationships and their biological plausibility. These criteria led to the creation of a co-morbidity database, which integrates these findings into a robust resource for understanding the interaction between COVID-19 and NDDs, offering insights into shared pathophysiological processes, genetic markers, and clinical outcomes. To ensure interoperability, candidate co-morbidity hypotheses were harmonized by standardizing entity names across multiple namespaces and ontologies, such as MeSH, ChEBI, Disease Ontology (DO) (https://disease-ontology.org) and Ontology Lookup Service (OLS) (https://www.ebi.ac.uk/ols4). The harmonized hypotheses were integrated into a publicly available Neo4j Aura database, enabling researchers to explore, curate, and experimentally validate the data. Figure 7. Represents a snapshot of this database.

3.4.2 Enrichment Analysis and Insights of the Co-morbidity Database

To enrich the hypothesis network with genetic data, we integrated risk variants and their documented associations with both COVID-19 and NDDs. This integration was based on data from genome-wide association studies (GWAS), accessible through GWAS Catalog (https://www.ebi.ac.uk/gwas/), which provides a comprehensive database of genetic loci linked to various diseases and traits. Additionally, we supplemented this data with evidence extracted from scientific literature, highlighting established and emerging connections between these genetic risk variants and the two conditions [47], [48], [49]. This approach ensured a more robust and comprehensive framework, enhancing the biological relevance of the hypothesis network by combining high-throughput genetic data with curated research findings. Notably, two specific risk variants, rs5117-C, and rs13107325-T, have been identified as significant

single nucleotide polymorphisms (SNPs) associated with PD, AD, and COVID-19, emphasizing shared genetic susceptibilities among the disease, as documented in GWAS studies [50], [51], [52], [53].

The hypothesis network is composed of more than 1,900 nodes and 4,900 edges, with a density of 0.002. The network is predominantly composed of nodes categorized as "GENE," "DISEASE," "BIOLOGICAL PROCESS," "CHEMICAL," and "PHENOTYPE." The most frequently observed relationship types within the database include "ASSOCIATION," "INCREASE," "INVOLVED_IN_PATHWAY," "POSITIVE CORRELATION," and "DECREASE." The total bar charts illustrating the distribution of relationship types, node types, and namespace types are provided in the Supplementary File.

To assess the relevance of the constructed hypothesis database and investigate the genetic intersections between COVID-19 and NDDs, we conducted a comprehensive manual review of multiple studies [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65]. This effort aimed to identify specific genetic variants, polymorphisms, and risk loci that may influence disease susceptibility and progression, offering deeper insights into potential shared biological mechanisms. Our analysis uncovered significant genetic associations between COVID-19 severity and NDDs, all integrated into the constructed hypothesis database. Notable genes associated with increased COVID-19 severity include TMPRSS2 and Furin [58] identified through Sherpa, and TLR7, identified through CBM. Among genes linked to NDDs, Sherpa identified SNCA, whose variants are associated with early-onset PD characterized by dementia and pyramidal signs [64]. Furthermore, PINK1, whose mutations represent a monogenic cause of PD, particularly in early-onset forms [64] was detected via SCAI-DMaps, DisGeNET, and CBM. The AD-associated gene ABCA7, extensively documented in the literature for its SNPs linked to Alzheimer's, was prominently highlighted by DisGeNET [57]. Several genes were specifically connected to the COVID-19–neurodegeneration link, including NLRP3 [65] as well as OAS1, and CXCL10 [63]. These genes were identified through multiple sources and databases, including CBM, SCAI-DMaps, DisGeNET, and OpenTargets.

Finally, we employed Over-Representation Analysis (ORA) to determine whether known biological pathways are significantly enriched in our gene list [66]. To achieve this, we utilized the DAVID bioinformatics tool (https://davidbioinformatics.nih.gov/), inputting our list of gene symbols, after which DAVID computed the Fisher’s Exact Test and p-value [67] ,

and generated an Enrichment Score. To ensure fair and accurate visualization, we removed statistical outliers and presented the results in a bar chart for clarity.

As illustrated in Figure 8., the pathway enrichment findings reveal key biological mechanisms linking COVID-19 and NDDs. The Nitric Oxide (NO) Signaling Pathway suggests that SARS-CoV-2 may disrupt vascular and neuronal function, contributing to cognitive impairment and neuroinflammation [68], [69]. The NF-κB Activation Pathway, a central regulator of inflammation [70], is over-activated in severe COVID-19 cases, paralleling the chronic neuroinflammatory processes observed in AD and PD [71], [72], [73] . Furthermore, the Chemokine Signaling Pathway highlights how persistent immune activation in COVID-19 could exacerbate neurodegenerative progression by enhancing microglial activation and synaptic dysfunction [74].

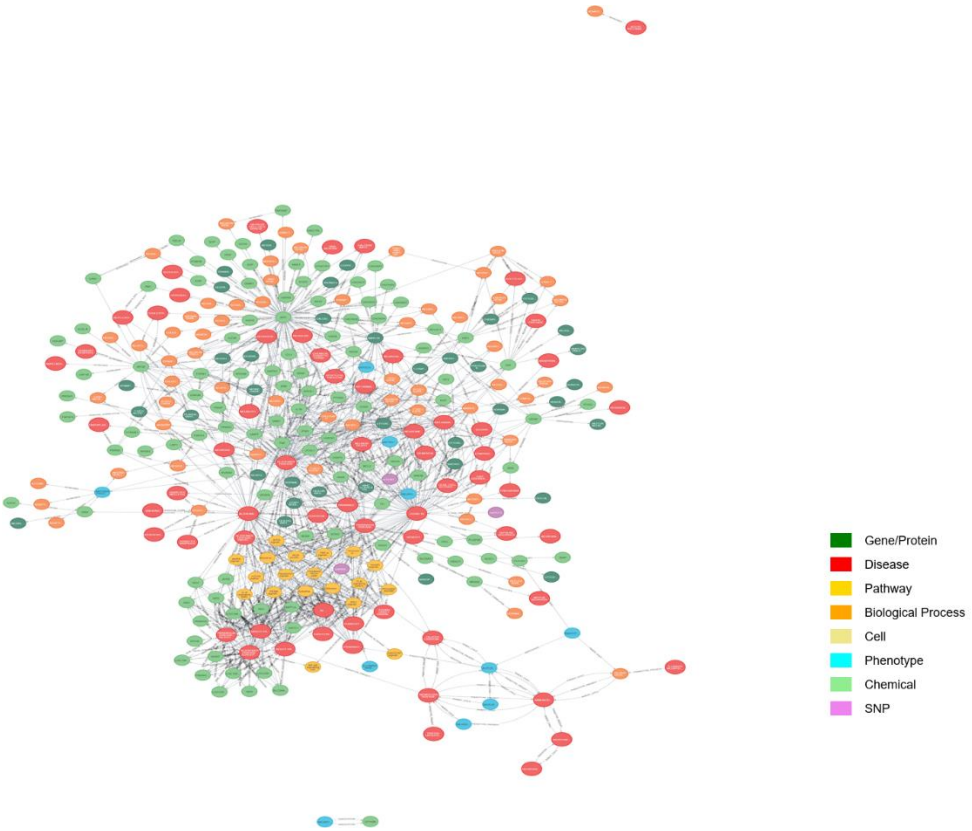

Figure 7. The snapshot of the COVID-NDD co-morbidity hypothesis database hosted in Neo4j AuraDB.

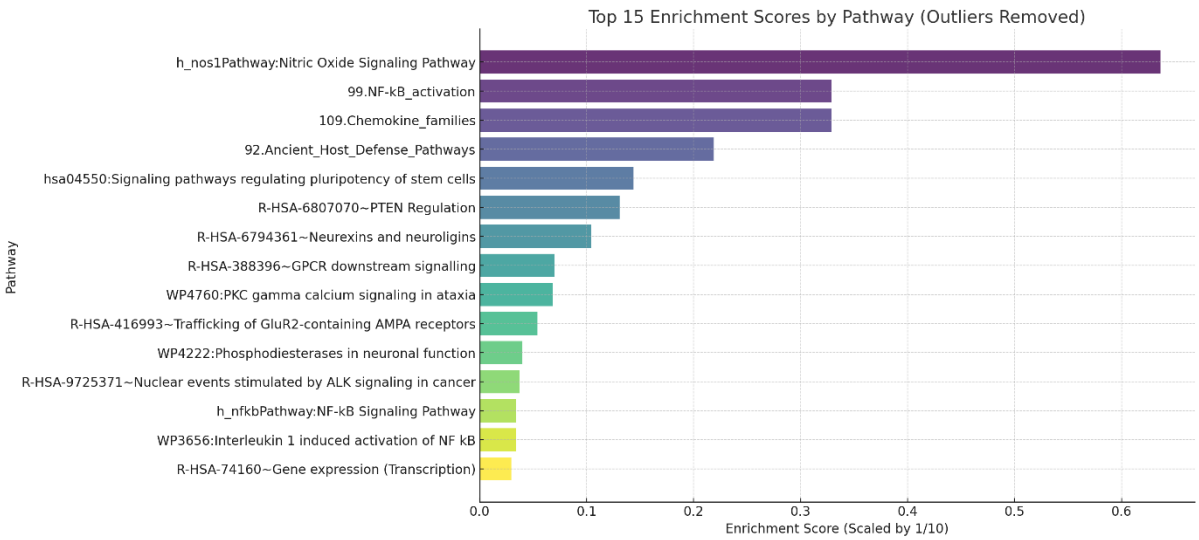

Figure 8. Over-Representation Analysis (ORA) showcasing the top enriched pathways derived from the hypothesis database gene list, analyzed using the DAVID bioinformatics tool.

4 Discussion

This study explored the comorbidity between COVID-19 and NDDs by analyzing their relationships within diverse KGs. Using insights from biomedical databases and text mining tools applied to relevant publications, we examined mechanisms and pathways linking COVID-19 to long-term neurological complications. Our approach, which combined graph algorithms with analyses of comorbid mechanisms, endpoints, and symptoms, allowed for a detailed exploration of how COVID-19-related processes intersect with NDD pathways. These findings contribute to the growing body of knowledge on the broader health implications of the virus, particularly its potential to exacerbate or initiate neurodegenerative processes.

4.1.1 Insights from various Co-Morbidity KGs

The structural diversity of the analyzed KGs, reflected in their distinct densities, connectivity patterns, and thematic focuses, highlights their complementary roles in biomedical research. DrugBank, for example, focuses on high-resolution drug-disease relationships, emphasizing molecular interactions relevant to precision pharmacological targeting. This specificity makes it particularly suitable for identifying therapeutic interventions. On the other hand, broader resources such as PrimeKG and OpenTargets take a systems-level approach, capturing the interdependencies between genes, proteins, and biological processes. This perspective is

particularly useful for studying diseases with overlapping mechanisms or shared phenotypic manifestations, such as the complex interplay between COVID-19 and NDDs.

Intermediate-sized KGs offer valuable, targeted molecular insights. SCAI-DMaps uncovers relationships involving neurotrophic factors, such as how GDNF and BDNF influence PD progression (with decreases observed in Parkinson’s). It also highlights immune system responses, revealing the upregulation of IL2 in the context of COVID-19. Sherpa maps the associations between COVID-19 and mood disorders, emphasizing the psychological impact of the virus KGs extracted through CBM and SCAI-DMaps contribute critical mechanistic insights, such as the role of mitochondrial metabolism in PD (as detailed in the study by Kannarkat et al. [75]), the processes governing blood-brain barrier development, and the dynamics of neuroinflammation and cytokine feedback loops.

DisGeNET and UNIPROT contribute to genetic associations (ND4 ASSOCIATION Alzheimer Disease, mitochondrial), while PrimeKG captures complex phenotypes (X-linked inheritance patterns). Recent COVID-specific insights from CBM graph also highlight Long COVID's effects on cognition and mental health.

These complementary KG strengths, from molecular mechanisms to systemic interactions, provide a comprehensive understanding of COVID-19-NDD relationships.

4.1.2 Key Genetic and Inflammatory Markers Included in the Hypothesis Database

Across analyses, several genetic and inflammatory markers consistently emerged as pivotal in understanding the intersection of COVID-19 and NDDs. Markers such as TLR7 (Toll-like Receptor 7), VEGFA (Vascular Endothelial Growth Factor A), APOE (Apolipoprotein E), and a variety of cytokines highlight the central roles of genetic susceptibility and immune responses. These markers are well-documented as contributors to inflammatory processes that can exacerbate both viral infections and neurodegenerative conditions [76], [77], [78].

APP's interaction with oxidative stress response pathways provides another mechanistic link, documented in SCAI-DMaps [79]. CBM data emphasizes neuroinflammation's role in modulating adaptive immune responses, while literature evidence confirms feedback loops between inflammatory cytokine production and neuroinflammation.

Neutrophil activation which is a critical mechanism involved in severe COVID-19 pathophysiology [80], was successfully captured by CBM KG. During infection, activated neutrophils release inflammatory mediators, form extracellular traps (NETs), and produce reactive oxygen species, leading to tissue damage and organ dysfunction [80]. This mechanism, , suggests a potential link between acute COVID-19 severity and long-term neurological complications, highlighting its importance as a therapeutic target. Neutrophil activation which is a critical mechanism involved in severe COVID-19 pathophysiology [80], was successfully captured by CBM KG. During infection, activated neutrophils release inflammatory mediators, form extracellular traps (NETs), and produce reactive oxygen species, leading to tissue damage and organ dysfunction [80]. This mechanism, suggests a potential link between acute COVID-19 severity and long-term neurological complications, highlighting its importance as a therapeutic target.

Sherpa further identified TRPV4, a calcium-permeable ion channel, as a notable marker involved in COVID-19 pathophysiology. TRPV4's association with inflammation, respiratory distress, and cellular stress underscores its potential role as a mechanistic link between acute COVID-19 symptoms and long-term neurological complications [81], [82]. The consistent identification of these markers across KGs reinforces their relevance for further research and highlights their potential as therapeutic targets.

4.1.3 Challenges of Data Heterogeneity

A key challenge encountered in this study was the heterogeneity of data across KGs. Each KG is derived from different sources and reflects unique strengths and limitations. For instance, some KGs focus on molecular interactions with high specificity but may lack phenotypic or clinical associations, while others provide broader, systems-level insights that might overlook molecular details. This diversity necessitates the integration of multiple KGs to create a more complete representation of the mechanisms underlying COVID-19 and NDD comorbidities.

Node harmonization is a critical step in integrating data from diverse KGs but is inherently challenging. Aligning entities to standardized ontologies ensures consistency and enables meaningful comparisons, yet discrepancies in terminologies, identifiers, and classifications across datasets complicate this process. For example, the same biological entity may appear under different names or identifiers across KGs, potentially leading to redundancies or

misrepresentations. Advanced methodologies are required to reconcile these differences and create unified datasets.

Another limitation involves the extraction of data via APIs, which, while efficient, often excludes some interactions and pathway visualizations available through user interfaces. This can result in the omission of critical insights necessary for understanding complex disease mechanisms. Additionally, some KGs, such as SCAI-DMaps, were developed over a year ago and may lack updates reflecting the latest research on COVID-19 and neurological complications. Maintaining the relevance and accuracy of KGs requires significant resources, including manual curation, which is both time-consuming and expensive.

4.1.4 Text Mining Limitations and Future Directions

Text mining tools like Sherpa demonstrated value in relation extraction but revealed some limitations in key areas. One primary challenge was incomplete entity recognition. For example, Sherpa occasionally extracted partial terms, such as identifying "HF" (heart failure) without capturing qualifiers like "acute." Such omissions can reduce the precision of downstream analyses by failing to account for important context. Another challenge was the handling of negations. Misinterpretation of relationships due to overlooked negation words sometimes resulted in incorrect correlations being identified. Additionally, the complexity of sentence structures and the presence of passive verb forms led to occasional inaccuracies in relationship extraction.

To address these limitations, advanced NLP techniques, such as contextual embeddings and large language models (LLMs), could enhance Sherpa’s performance. Expanding the scope of text mining to include non-molecular factors—such as behavioral symptoms, clinical manifestations, environmental influences, and lifestyle factors—would provide a more holistic understanding of COVID-19-NDD comorbidities. Incorporating additional datasets, such as clinical texts, epidemiological reports, and public health data, could further enrich the analysis and provide deeper insights into these complex relationships.

4.1.5 Toward a Comprehensive Co-Morbidity Database

The findings highlight the need for a dedicated co-morbidity database focused on COVID-19 and its neurological complications. While the current database integrates insights from

resources like the GWAS catalog, bridging the gap between molecular mechanisms and clinical manifestations requires additional data layers. These layers include protein-protein interactions, tissue-specific gene expression profiles, and clinical observations from electronic health records (EHRs). Integrating such diverse data would enable researchers to generate more robust hypotheses and facilitate the identification of actionable insights for therapeutic development.

However, integrating diverse datasets is not without challenges. Variability in diagnostic coding, inconsistencies in phenotypic definitions, and the lack of standardized approaches for linking molecular data to clinical endpoints complicate the process. Leveraging machine learning and network-based approaches, along with ontology-based mappings, is essential for establishing meaningful connections between these data types. Addressing these challenges would enable the creation of a comprehensive resource that bridges molecular, phenotypic, and clinical data layers.

4.1.6 Hypothesis Validation and Future Directions

Validation of the hypotheses generated in this study is critical to ensure their robustness and translational potential. Computational approaches, such as reverse causal reasoning (RCR) and network perturbation methods [83], candidate mechanism perturbation amplitude [84], or clinal embeddings for patients [85], provide valuable tools for assessing the plausibility of identified mechanisms. Additionally, experimental validation through molecular and cellular assays is essential for verifying key hypotheses. This process involves aligning computational findings with measurable entities in laboratory settings, enabling targeted testing of specific mechanisms.

To facilitate this, the proposed co-morbidity database should support expert annotation of hypotheses, enabling researchers to incorporate domain-specific knowledge. Automated workflows for hypothesis testing and data selection would further enhance the database's utility and ensure it remains dynamic and reflective of the latest scientific advances [86].

Given the rapidly evolving nature of biomedical knowledge, continuous updates to co-morbidity graphs are essential. Automated and adaptive procedures for updating and validating these graphs will ensure their ongoing relevance. These efforts will advance our understanding

of COVID-19-related neurological complications and inform the development of targeted therapeutic strategies.

By addressing the outlined challenges and expanding the scope of future research, this work lays the foundation for more comprehensive investigations into the comorbid mechanisms of COVID-19 and NDDs. The integration of diverse datasets and continuous validation of hypotheses will be pivotal in translating computational findings into actionable insights that can guide disease modeling and therapeutic development.

5 Conclusion

This study utilized a diverse collection of databases and resources, structured as KGs, alongside advanced text mining techniques, to investigate co-morbidity mechanisms between COVID-19 and NDDs. Through various graph-based analyses and phenotype coverage explorations, we highlighted how diverse KGs contribute unique insights into the molecular and clinical intersections between these conditions. Smaller KGs offer targeted pharmacological data, while larger and medium-sized KGs capture broader systemic interactions, particularly in inflammation and immune responses. The study underscores the need for an integrated approach to capture these multifaceted connections but also recognizes limitations such as data heterogeneity, the need for constant updating of existing KGs, and the lack of experimental validations. Future enhancements, including advancements in text mining tools and the integration of diverse data types—such as clinical data, EHRs, and real-world evidence—hold significant potential to deepen our understanding of comorbid pathways. By bridging molecular insights with clinical observations, these advancements could provide a more holistic view of disease mechanisms and facilitate the identification of actionable therapeutic targets. Such progress will greatly improve our hypothesis database, allowing better testing and validation of COVID-19 and NDD comorbidities. This will help develop targeted treatments for at-risk populations by addressing both the underlying molecular mechanisms and their clinical effects more effectively.

Data and Code Availability

The data and source codes used in this study are available at: https://github.com/SCAI-BIO/covid-NDD-comorbidity-NLP.

Funding

This work was supported by the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI) and the Bonn-Aachen International Center for Information Technology (b-it) foundation, Bonn, Germany. Additional funding came from the COMMUTE project, which is funded by the European Union under grant agreement number 101136957.

Acknowledgement

The authors would like to thank Prof. Laura Furlong, IMIM Barcelona, for her support with the DisGeNet analysis. The authors would also like to thank Dr. Benjamin Gyori, Harvard Medical School, for fruitful discussions and guidance with the INDRA system.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. Pan *et al.*, “Multivariate patterns of brain functional connectome associated with COVID-19-related negative affect symptoms,” *Transl Psychiatry*, vol. 14, no. 1, pp. 1–9, Jan. 2024, doi: 10.1038/s41398-024-02741-1.

[2] “The unequal effects of the health–economy trade-off during the COVID-19 pandemic | Nature Human Behaviour.” Accessed: Aug. 09, 2024. [Online]. Available: https://www.nature.com/articles/s41562-023-01747-x

[3] “The impact of the COVID-19 pandemic and associated disruptions in health-care provision on clinical outcomes in people with diabetes: a systematic review - The Lancet Diabetes & Endocrinology.” Accessed: Aug. 09, 2024. [Online]. Available: https://www.thelancet.com/journals/landia/article/PIIS2213-8587(23)00351-0/abstract

[4] M. L. Ranney, V. Griffeth, and A. K. Jha, “Critical Supply Shortages — The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic,” *New England Journal of Medicine*, vol. 382, no. 18, p. e41, Apr. 2020, doi: 10.1056/NEJMp2006141.

[5] M. Nicola *et al.*, “The socio-economic implications of the coronavirus pandemic (COVID-19): A review,” *Int J Surg*, vol. 78, pp. 185–193, Jun. 2020, doi: 10.1016/j.ijsu.2020.04.018.

[6] “Mental health in the COVID-19 pandemic | QJM: An International Journal of Medicine | Oxford Academic.” Accessed: Aug. 09, 2024. [Online]. Available: https://academic.oup.com/qjmed/article/113/5/311/5813733?login=false

[7] S. A. Narayanan *et al.*, “A comprehensive SARS-CoV-2 and COVID-19 review, Part 2: host extracellular to systemic effects of SARS-CoV-2 infection,” *Eur J Hum Genet*, vol. 32, no. 1, pp. 10–20, Jan. 2024, doi: 10.1038/s41431-023-01462-1.

[8] R. Knoll, J. L. Schultze, and J. Schulte-Schrepping, “Monocytes and Macrophages in COVID-19,” *Front. Immunol.*, vol. 12, Jul. 2021, doi: 10.3389/fimmu.2021.720109.

[9] J. L. Schultze, A. Schmieder, and S. Goerdt, “Macrophage activation in human diseases,” *Seminars in Immunology*, vol. 27, no. 4, pp. 249–256, Aug. 2015, doi: 10.1016/j.smim.2015.07.003.

[10] L. Dale, “Neurological Complications of COVID-19: A Review of the Literature,” *Cureus*, vol. 14, no. 8, p. e27633, doi: 10.7759/cureus.27633.
[11] S. L. Armijo-olivo and I. C. Gadotti, “Chapter 4 - Temporomandibular Disorders*The authors, editors, and publisher wish to acknowledge Martin Parfitt for his contributions on this topic in the previous edition.,” in *Pathology and Intervention in Musculoskeletal Rehabilitation (Second Edition)*, D. J. Magee, J. E. Zachazewski, W. S. Quillen, and R. C. Manske, Eds., W.B. Saunders, 2016, pp. 119–156. doi: 10.1016/B978-0-323-31072-7.00004-X.
[12] B. D. Michael *et al.*, “Para-infectious brain injury in COVID-19 persists at follow-up despite attenuated cytokine and autoantibody responses,” *Nat Commun*, vol. 14, no. 1, p. 8487, Dec. 2023, doi: 10.1038/s41467-023-42320-4.
[13] E. Normandin *et al.*, “Neuropathological features of SARS-CoV-2 delta and omicron variants,” *Journal of Neuropathology and Experimental Neurology*, vol. 82, no. 4, p. 283, Feb. 2023, doi: 10.1093/jnen/nlad015.
[14] R. B. Domingues, F. B. V. de M. Leite, and C. Senne, “Cerebrospinal fluid analysis in patients with COVID-19-associated central nervous system manifestations: a systematic review,” *Arquivos de Neuro-Psiquiatria*, vol. 80, no. 3, p. 296, Feb. 2022, doi: 10.1590/0004-282X-ANP-2021-0117.
[15] “SARS-CoV-2 and the brain: A review of the current knowledge on neuropathology in COVID-19 - PMC.” Accessed: Oct. 18, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8420197/
[16] W. Christ *et al.*, “SARS-CoV-2 and HSV-1 Induce Amyloid Aggregation in Human CSF Resulting in Drastic Soluble Protein Depletion,” Sep. 21, 2024, *bioRxiv*. doi: 10.1101/2022.09.15.508120.
[17] J. Zhao, F. Xia, X. Jiao, and X. Lyu, “Long COVID and its association with neurodegenerative diseases: pathogenesis, neuroimaging, and treatment,” *Front. Neurol.*, vol. 15, Apr. 2024, doi: 10.3389/fneur.2024.1367974.
[18] H. T. Hogberg *et al.*, “The Adverse Outcome Pathway Framework Applied to Neurological Symptoms of COVID-19,” *Cells*, vol. 11, no. 21, p. 3411, Oct. 2022, doi: 10.3390/cells11213411.
[19] “Project to establish a COVID-19 AOP,” CIAO Project. Accessed: Aug. 05, 2024. [Online]. Available: https://www.ciao-covid.net
[20] L. Wu, S. Ali, H. Ali, T. Brock, J. Xu, and W. Tong, “NeuroCORD: A Language Model to Facilitate COVID-19-Associated Neurological Disorder Studies,” *Int J Environ Res Public Health*, vol. 19, no. 16, p. 9974, Aug. 2022, doi: 10.3390/ijerph19169974.
[21] “Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset - ScienceDirect.” Accessed: Aug. 05, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422017407
[22] A. Karami, B. Bookstaver, M. Nolan, and P. Bozorgi, “Investigating diseases and chemicals in COVID-19 literature with text mining,” *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100016, Nov. 2021, doi: 10.1016/j.jjimei.2021.100016.
[23] S. R. Knight *et al.*, “Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score,” *BMJ*, vol. 370, p. m3339, Sep. 2020, doi: 10.1136/bmj.m3339.
[24] C. D. Russell, N. I. Lone, and J. K. Baillie, “Comorbidities, multimorbidity and COVID-19,” *Nat Med*, vol. 29, no. 2, pp. 334–343, Feb. 2023, doi: 10.1038/s41591-022-02156-9.
[25] A. Romagnolo *et al.*, “Neurological comorbidities and COVID-19-related case fatality: A cohort study,” *Journal of the Neurological Sciences*, vol. 428, p. 117610, Sep. 2021, doi: 10.1016/j.jns.2021.117610.
[26] M. J. Hua, G. Butera, O. Akinyemi, and D. Porterfield, “Biases and limitations in observational studies of Long COVID prevalence and risk factors: A rapid systematic

umbrella review," *PLoS One*, vol. 19, no. 5, p. e0302408, May 2024, doi: 10.1371/journal.pone.0302408.
[27] G. Douaud *et al.*, "SARS-CoV-2 is associated with changes in brain structure in UK Biobank," *Nature*, vol. 604, no. 7907, pp. 697–707, Apr. 2022, doi: 10.1038/s41586-022-04569-5.
[28] J. Piñero *et al.*, "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res*, vol. 45, no. Database issue, pp. D833–D839, Jan. 2017, doi: 10.1093/nar/gkw943.
[29] UniProt Consortium, "UniProt: a worldwide hub of protein knowledge," *Nucleic Acids Res*, vol. 47, no. D1, pp. D506–D515, Jan. 2019, doi: 10.1093/nar/gky1049.
[30] M. J. Landrum *et al.*, "ClinVar: improving access to variant interpretations and supporting evidence," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1062–D1067, Jan. 2018, doi: 10.1093/nar/gkx1153.
[31] E. Sollis *et al.*, "The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource," *Nucleic Acids Research*, vol. 51, no. D1, pp. D977–D985, Jan. 2023, doi: 10.1093/nar/gkac1010.
[32] G. Koscielny *et al.*, "Open Targets: a platform for therapeutic target identification and validation," *Nucleic Acids Research*, vol. 45, no. Database issue, p. D985, Dec. 2016, doi: 10.1093/nar/gkw1055.
[33] C. Knox *et al.*, "DrugBank 6.0: the DrugBank Knowledgebase for 2024," *Nucleic Acids Res*, vol. 52, no. D1, pp. D1265–D1275, Jan. 2024, doi: 10.1093/nar/gkad976.
[34] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," *Sci Data*, vol. 10, no. 1, p. 67, Feb. 2023, doi: 10.1038/s41597-023-01960-3.
[35] E. Sayers, "A General Introduction to the E-utilities," in *Entrez Programming Utilities Help [Internet]*, National Center for Biotechnology Information (US), 2022. Accessed: Jun. 14, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK25497/
[36] "PMC text mining subset in BioC: about three million full-text articles and growing | Bioinformatics | Oxford Academic." Accessed: Nov. 14, 2024. [Online]. Available: https://academic.oup.com/bioinformatics/article/35/18/3533/5305021
[37] M. Kanehisa, "The KEGG Database," in *'In Silico' Simulation of Biological Processes*, John Wiley & Sons, Ltd, 2002, pp. 91–103. doi: 10.1002/0470857897.ch8.
[38] N. S. Babaiha *et al.*, "A natural language processing system for the efficient updating of highly curated pathophysiology mechanism knowledge graphs," *Artificial Intelligence in the Life Sciences*, vol. 4, p. 100078, Dec. 2023, doi: 10.1016/j.ailsci.2023.100078.
[39] N. S. Babaiha, S. G. Rao, J. Klein, B. Schultz, M. Jacobs, and M. Hofmann-Apitius, "Rationalism in the face of GPT hypes: Benchmarking the output of large language models against human expert-curated biomedical knowledge graphs," *Artificial Intelligence in the Life Sciences*, vol. 5, p. 100095, Jun. 2024, doi: 10.1016/j.ailsci.2024.100095.
[40] D. Domingo-Fernández *et al.*, "COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology," *Bioinformatics*, vol. 37, no. 9, pp. 1332–1334, Jun. 2021, doi: 10.1093/bioinformatics/btaa834.
[41] D. Domingo-Fernández *et al.*, "Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment," *Bioinformatics*, vol. 33, no. 22, pp. 3679–3681, Nov. 2017, doi: 10.1093/bioinformatics/btx399.
[42] BEL.bio, "BEL.bio · BEL.bio," BEL.bio. Accessed: Aug. 30, 2023. [Online]. Available: https://bel.bio/
[43] S. Geißler, "The Kairntech Sherpa – An ML Platform and API for the Enrichment of (not only) Scientific Content," in *Proceedings of the 1st International Workshop on Language Technology Platforms*, Marseille, France: European Language Resources Association, May 2020, pp. 54–58. Accessed: Dec. 12, 2022. [Online]. Available: https://aclanthology.org/2020.iwltp-1.9

[44] “PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge | Nucleic Acids Research | Oxford Academic.” Accessed: Nov. 19, 2024. [Online]. Available: https://academic.oup.com/nar/article/52/W1/W540/7640526
[45] “Human Phenotype Ontology.” Accessed: Nov. 20, 2024. [Online]. Available: https://hpo.jax.org/
[46] S. Montazersaheb *et al.*, “COVID-19 infection: an overview on cytokine storm and related interventions,” *Virology Journal*, vol. 19, no. 1, p. 92, May 2022, doi: 10.1186/s12985-022-01814-1.
[47] N. Matveeva *et al.*, “Shared genetic architecture of COVID-19 and Alzheimer’s disease,” *Front. Aging Neurosci.*, vol. 15, Oct. 2023, doi: 10.3389/fnagi.2023.1287322.
[48] E. Kovalenko *et al.*, “GWAS and polygenic risk score of severe COVID-19 in Eastern Europe,” *Front. Med.*, vol. 11, Sep. 2024, doi: 10.3389/fmed.2024.1409714.
[49] L. Ibanez, F. H. G. Farias, U. Dube, K. A. Mihindukulasuriya, and O. Harari, “Polygenic Risk Scores in Neurodegenerative Diseases: a Review,” *Curr Genet Med Rep*, vol. 7, no. 1, pp. 22–29, Mar. 2019, doi: 10.1007/s40142-019-0158-0.
[50] M. Alcalde-Herraiz, M. Català, A. Prats-Uribe, R. Paredes, J. Xie, and D. Prieto-Alhambra, “Genome-wide association studies of COVID-19 vaccine seroconversion and breakthrough outcomes in UK Biobank,” *Nat Commun*, vol. 15, no. 1, p. 8739, Oct. 2024, doi: 10.1038/s41467-024-52890-6.
[51] Y. Chen, C. Fan, and J. Liu, “Investigating the shared genetic architecture between COVID-19 and obesity: a large-scale genome wide cross-trait analysis,” *Front Endocrinol (Lausanne)*, vol. 15, p. 1325939, Jan. 2024, doi: 10.3389/fendo.2024.1325939.
[52] O. B. Smeland *et al.*, “Genome-wide association analysis of Parkinson’s disease and schizophrenia reveals shared genetic architecture and identifies novel risk loci,” *Biol Psychiatry*, vol. 89, no. 3, pp. 227–235, Feb. 2021, doi: 10.1016/j.biopsych.2020.01.026.
[53] J. S. Reddy *et al.*, “Genome-wide analysis identifies a novel LINC-PINT splice variant associated with vascular amyloid pathology in Alzheimer’s disease,” *Acta Neuropathologica Communications*, vol. 9, no. 1, p. 93, May 2021, doi: 10.1186/s40478-021-01199-2.
[54] J. Schwartzentruber *et al.*, “Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer’s disease risk genes,” *Nat Genet*, vol. 53, no. 3, pp. 392–402, Mar. 2021, doi: 10.1038/s41588-020-00776-w.
[55] M. A. A. DeMichele-Sweet *et al.*, “Genome-wide association identifies the first risk loci for psychosis in Alzheimer disease,” *Mol Psychiatry*, vol. 26, no. 10, pp. 5797–5811, Oct. 2021, doi: 10.1038/s41380-021-01152-8.
[56] A. Rocchi, S. Pellegrini, G. Siciliano, and L. Murri, “Causative and susceptibility genes for Alzheimer’s disease: a review,” *Brain Res Bull*, vol. 61, no. 1, pp. 1–24, Jun. 2003, doi: 10.1016/s0361-9230(03)00067-4.
[57] L. Heath *et al.*, “Manifestations of Alzheimer’s disease genetic risk in the blood are evident in a multiomic analysis in healthy adults aged 18 to 90,” *Sci Rep*, vol. 12, no. 1, p. 6117, Apr. 2022, doi: 10.1038/s41598-022-09825-2.
[58] A. Ishak *et al.*, “The association of COVID-19 severity and susceptibility and genetic risk factors: A systematic review of the literature,” *Gene*, vol. 836, p. 146674, Aug. 2022, doi: 10.1016/j.gene.2022.146674.
[59] C. Dieter, L. de A. Brondani, C. B. Leitão, F. Gerchman, N. E. Lemos, and D. Crispim, “Genetic polymorphisms associated with susceptibility to COVID-19 disease and severity: A systematic review and meta-analysis,” *PLOS ONE*, vol. 17, no. 7, p. e0270627, Jun. 2022, doi: 10.1371/journal.pone.0270627.
[60] C. Cappadona, V. Rimoldi, E. M. Paraboschi, and R. Asselta, “Genetic susceptibility to severe COVID-19,” *Infection, Genetics and Evolution*, vol. 110, p. 105426, Jun. 2023, doi: 10.1016/j.meegid.2023.105426.

[61] “GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19 | Nature.” Accessed: Jan. 24, 2025. [Online]. Available: https://www.nature.com/articles/s41586-023-06034-3
[62] “Host genetic determinants of COVID-19 susceptibility and severity: A systematic review and meta-analysis - Eshetie - 2023 - Reviews in Medical Virology - Wiley Online Library.” Accessed: Jan. 24, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/rmv.2466
[63] “Emerging potential mechanisms and predispositions to the neurological manifestations of COVID-19 - Journal of the Neurological Sciences.” Accessed: Jan. 24, 2025. [Online]. Available: https://www.jns-journal.com/article/S0022-510X(21)00302-6/fulltext
[64] C. M. Lill, “Genetics of Parkinson’s disease,” *Molecular and Cellular Probes*, vol. 30, no. 6, pp. 386–396, Dec. 2016, doi: 10.1016/j.mcp.2016.11.001.
[65] C. Li, J. Liu, J. Lin, and H. Shang, “COVID-19 and risk of neurodegenerative disorders: A Mendelian randomization study,” *Transl Psychiatry*, vol. 12, no. 1, pp. 1–6, Jul. 2022, doi: 10.1038/s41398-022-02052-3.
[66] E. 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*, vol. 20, no. 18, pp. 3710–3715, Dec. 2004, doi: 10.1093/bioinformatics/bth456.
[67] G. J. G. Upton, “Fisher’s Exact Test,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 155, no. 3, pp. 395–402, 1992, doi: 10.2307/2982890.
[68] Z. S. Katusic, L. V. d’Uscio, and T. He, “Emerging Roles of Endothelial Nitric Oxide in Preservation of Cognitive Health,” *Stroke*, vol. 54, no. 3, pp. 686–696, Mar. 2023, doi: 10.1161/STROKEAHA.122.041444.
[69] P. M. Liy, N. N. A. Puzi, S. Jose, and S. Vidyadaran, “Nitric oxide modulation in neuroinflammation and the role of mesenchymal stem cells,” *Exp Biol Med (Maywood)*, vol. 246, no. 22, pp. 2399–2406, Nov. 2021, doi: 10.1177/1535370221997052.
[70] “NF-κB signaling in inflammation | Signal Transduction and Targeted Therapy.” Accessed: Feb. 02, 2025. [Online]. Available: https://www.nature.com/articles/sigtrans201723
[71] “The role of SARS-CoV-2-mediated NF-κB activation in COVID-19 patients | Hypertension Research.” Accessed: Feb. 02, 2025. [Online]. Available: https://www.nature.com/articles/s41440-023-01460-2
[72] E. Sun, A. Motolani, L. Campos, and T. Lu, “The Pivotal Role of NF-kB in the Pathogenesis and Therapeutics of Alzheimer’s Disease,” *Int J Mol Sci*, vol. 23, no. 16, p. 8972, Aug. 2022, doi: 10.3390/ijms23168972.
[73] B. Kaltschmidt, L. P. Helweg, J. F. W. Greiner, and C. Kaltschmidt, “NF-κB in neurodegenerative diseases: Recent evidence from human genetics,” *Front. Mol. Neurosci.*, vol. 15, Aug. 2022, doi: 10.3389/fnmol.2022.954541.
[74] B. P. Festa *et al.*, “Microglial-to-neuronal CCR5 signaling regulates autophagy in neurodegeneration,” *Neuron*, vol. 111, no. 13, pp. 2021-2037.e12, Jul. 2023, doi: 10.1016/j.neuron.2023.04.006.
[75] G. T. Kannarkat, J. M. Boss, and M. G. Tansey, “The role of innate and adaptive immunity in Parkinson’s disease,” *J Parkinsons Dis*, vol. 3, no. 4, pp. 493–514, 2013, doi: 10.3233/JPD-130250.
[76] C. Wang *et al.*, “Deficiency of Tlr7 and Irf7 in mice increases the severity of COVID-19 through the reduced interferon production,” *Commun Biol*, vol. 7, no. 1, pp. 1–16, Sep. 2024, doi: 10.1038/s42003-024-06872-5.
[77] A. Philippe *et al.*, “VEGF-A plasma levels are associated with impaired DLCO and radiological sequelae in long COVID patients,” *Angiogenesis*, vol. 27, no. 1, pp. 51–66, Feb. 2024, doi: 10.1007/s10456-023-09890-9.
[78] R. Talotta, “Impaired VEGF-A-Mediated Neurovascular Crosstalk Induced by SARS-CoV-2 Spike Protein: A Potential Hypothesis Explaining Long COVID-19 Symptoms and COVID-19 Vaccine Side Effects?,” *Microorganisms*, vol. 10, no. 12, p. 2452, Dec. 2022, doi: 10.3390/microorganisms10122452.

[79] M. E. D. Mingoti, A. G. Bertollo, J. L. B. Simões, G. R. Francisco, M. D. Bagatini, and Z. M. Ignácio, “COVID-19, Oxidative Stress, and Neuroinflammation in the Depression Route,” *J Mol Neurosci*, vol. 72, no. 6, pp. 1166–1181, 2022, doi: 10.1007/s12031-022-02004-y.
[80] N. Reusch *et al.*, “Neutrophils in COVID-19,” *Front. Immunol.*, vol. 12, Mar. 2021, doi: 10.3389/fimmu.2021.652470.
[81] F. Zhang, H. Mehta, H. H. Choudhary, R. Islam, and K. A. Hanafy, “TRPV4 Channel in Neurological Disease: from Molecular Mechanisms to Therapeutic Potential,” *Mol Neurobiol*, Sep. 2024, doi: 10.1007/s12035-024-04518-5.
[82] R. G. Scheraga, B. D. Southern, L. M. Grove, and M. A. Olman, “The Role of Transient Receptor Potential Vanilloid 4 in Pulmonary Inflammatory Diseases,” *Front Immunol*, vol. 8, p. 503, May 2017, doi: 10.3389/fimmu.2017.00503.
[83] F. Martin, A. Sewer, M. Talikka, Y. Xiang, J. Hoeng, and M. C. Peitsch, “Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models,” *BMC Bioinformatics*, vol. 15, no. 1, p. 238, Jul. 2014, doi: 10.1186/1471-2105-15-238.
[84] R. Karki, A. T. Kodamullil, C. T. Hoyt, and M. Hofmann-Apitius, “Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm,” *BMC Bioinformatics*, vol. 20, no. 1, p. 494, Oct. 2019, doi: 10.1186/s12859-019-3101-1.
[85] V. S. Bharadhwaj *et al.*, “CLEP: a hybrid data- and knowledge-driven framework for generating patient representations,” *Bioinformatics*, vol. 37, no. 19, p. 3311, May 2021, doi: 10.1093/bioinformatics/btab340.
[86] W. L. Chin and T. Lassmann, “SampleExplorer: Using language models to discover relevant transcriptome data,” *Bioinformatics*, p. btae759, Dec. 2024, doi: 10.1093/bioinformatics/btae759.