



eBRAIN-Health

Public report D4.3

Open access pathway knowledge graphs

Project number	101058516
Project title	eBRAIN-Health - Actionable Multilevel Health Data
Submission date	July 2025
Authors	Charite, Fraunhofer
Dissemination level	Public (PU)
Public project website	https://ebrain-health.eu



Funded by
the European Union

Table of Contents

1.	eBRAIN-Health.....	3
2.	eBRAIN-Health consortium.....	3
3.	Introduction	4
4.	Partners involved	4
5.	Description of work performed	4
5.1.	Ontology models	4
5.2.	Causal Biological Mechanism Knowledge Graphs (CBM KGs).....	5
5.3.	Alzheimer Knowledge Graph with EEG Features	7
5.4.	Neuro-PsychoMMSig (Neuro-Psychiatric Multimodal Signature).....	7
5.4.1.	Explorative Analysis	8
5.4.2.	Clinical Analysis.....	8
5.4.3.	Case Scenario	11
6.	Results.....	11
6.1.	Alzheimer Knowledge Graph with EEG Features	11
6.2.	Neuro-PsychoMMSig (Neuro-Psychiatric Multimodal Signature).....	12
6.2.1.	Explorative Analysis	13
6.2.2.	Clinical Analysis.....	14
6.2.3.	AD Case Study	14
6.2.4.	AD sample stratification by age	15
6.2.5.	Schizophrenia Case Study	15
7.	Conclusion, next steps.....	15

1. eBRAIN-Health

Project eBRAIN-Health will deliver a distributed research platform for modeling and simulating complex neurobiological phenomena of human brain function and dysfunction in a data protection compliant environment. It will provide thousands of multilevel virtual brains from patients and healthy human controls for research and innovation. Brain data from multiple sources will be pre-processed. Solving the societal grand challenge of dementia is a big task. Yet it appears feasible in a collective approach. Therefore, we will build an interdisciplinary digital twin for dementia for modeling and simulating complex phenomena at the service of research infrastructure communities. eBRAIN-Health-Cloud will offer end-to-end services for personalized complex brain modeling and simulations in distributed e-infrastructures with data protection by design and by default and simulation-ready human multiscale brain data that range from molecular (genomics, proteomics, metabolomics) and cellular to electrophysiology and imaging to behavioral, clinical, lifestyle and environmental data as well as data from wearables. Brain data are pre-processed and annotated such that they all relate to a common reference 3D brain space.

eBRAIN-Health-Cloud constitutes a blend of three large-scale research programs: the FET Flagship Human Brain Project with its EBRAINS Research Infrastructure, the EOSC project Virtual Brain Cloud with its Virtual Research Environment for sensitive data and the H2020 project AI-MIND with intelligent tools for dementia risk estimation. The project will have synergies to topics of the Digital Europe Program, such as artificial intelligence, cybersecurity and supercomputing and the Health Data Space.

eBRAIN-Health-Cloud offers a next generation clinical research infrastructure and creates an open yet protected space for groundbreaking digital health innovation by the research infrastructure communities comprising academia and the private sector.

2. eBRAIN-Health consortium

- CHARITE – Universitaetsmedizin Berlin, Germany
- EBRAINS, Belgium
- Forschungszentrum Juelich GmbH, Germany
- Stichting Radboud Universiteit, Netherlands
- Universidad Pompeu Fabra, Spain
- OSLO Universitetssykehus, Norway
- tp21 GMBH, Germany
- Fraunhofer Gesellschaft zur Foerderung der Angewandten Forschung eV, Germany
- INDOC RESEARCH EUROPE gGmbH, Germany
- Universitaet Wien, Austria
- Universidad Complutense de Madrid, Spain
- EODYNE Systems SL, Spain
- ATHENA – Research and Innovation Center, Greece
- University of Oslo, Norway
- Universita degli Studi di Roma la Sapienza, Italy
- Alzheimer Europe, Luxembourg
- Institute National de Recherche en Informatique et Automatique, France
- Centre Hospitalier Universitaire Vaudois, Switzerland
- The University of Edinburgh, United Kingdom

[Find the partners on our website](#)

3. Introduction

The main aim of this task is to develop a neurodegenerative disease (NDD) knowledge graph (KG) utilizing ontologies, including BRCTO, NIFO, PTS, and ADO. These ontologies function as structured dictionaries to extract entities from unstructured text, which are then transformed into KG using the ProMiner tool, a proprietary technology developed by Fraunhofer SCAI. In addition to that, it is aimed to integrate the resulting KGs into a pathway enrichment Server, enabling them to serve as prior knowledge for multimodal datasets. By combining structured information from KGs with diverse biological datasets, the platform facilitates the application of advanced analytical algorithms.

4. Partners involved

CHARITE (lead), FRAUNHOFER

5. Description of work performed

For establishing the task 4.3, we have developed the following resources and tools:

5.1. Ontology models

As part of the eBRAINS-Health project, a semantic framework has been developed to enable unified metadata annotation. This framework is built upon six ontologies specifically designed to support neurodegenerative and psychiatric research:

- Alzheimer Disease Ontology (ADO)
- EEG/MEG and Feature Terminology
- Pathway Terminology System (PTS)
- Neuroimaging Feature Ontology (NIFT)
- Brain Region and Cell Type Ontology (BRCTO)
- Human Physiology Simulation Ontology (HUPSON)

The semantic framework is managed through the eBRAINS-Health Ontology Look-up Service (OLS), which is based on the OLS platform from the European Bioinformatics Institute (EBI). As of March 2025, the framework includes 11,790 classes and 642 properties/relations derived from the six project-specific ontologies (**Figure 1** and **Figure 2**). This system provides a robust foundation for metadata annotation and semantic integration across diverse datasets. The semantic framework is accessible via the following link: <https://ols.ebrain.bio.scai.fraunhofer.de/index>.

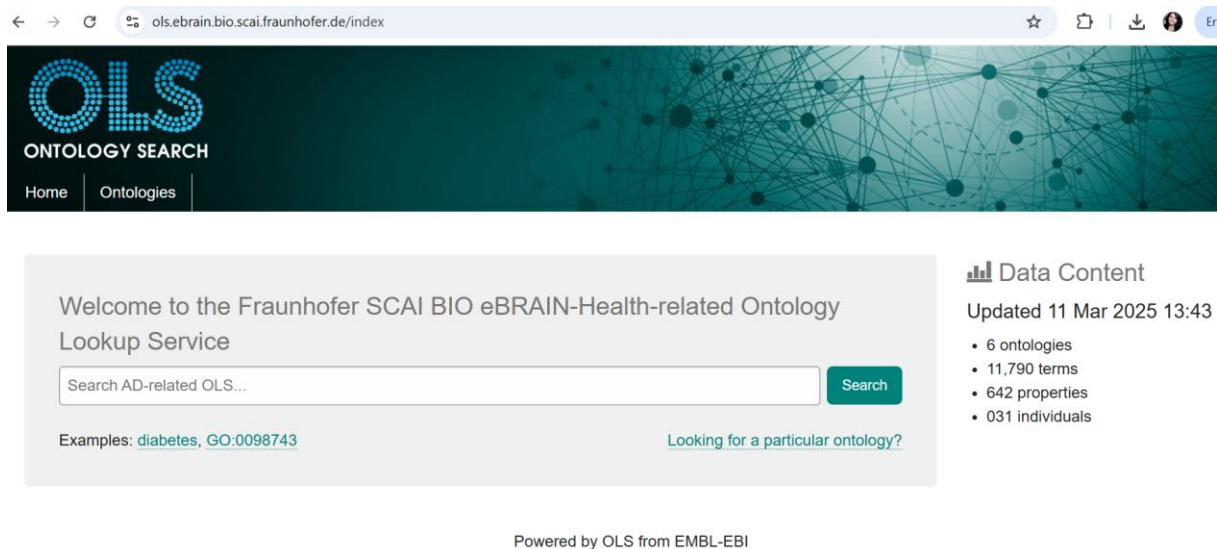


Figure 1: the eBRAINS-Health Ontology Look-up Service (OLS)

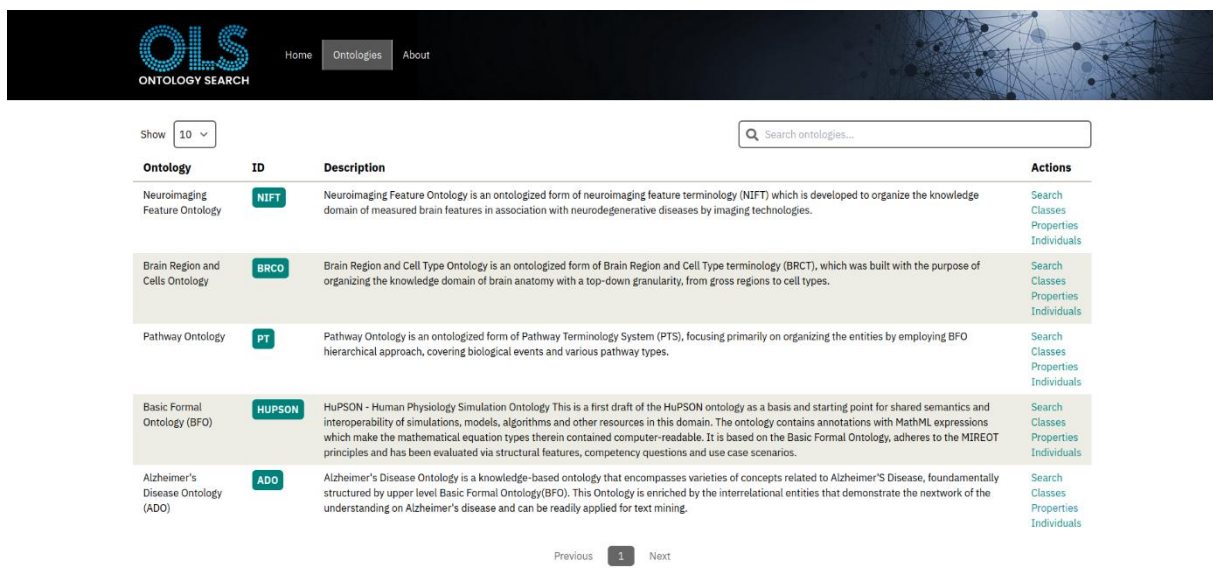


Figure 2: The available ontologies in the eBRAINS-Health Ontology Look-up Service (OLS)

5.2. Causal Biological Mechanism Knowledge Graphs (CBM KGs)

Fraunhofer has developed KGs by utilizing six ontologies, including BRCTO, NIFT, PTS, and ADO. These ontologies function as structured dictionaries to extract entities from unstructured text, which are then transformed into a KG using the ProMiner tool, a proprietary technology developed by Fraunhofer SCAI. Causal Biological Mechanism Knowledge Graphs (CBM KGs) have been developed using these KGs to enable in-depth, mechanism-driven exploration and analysis of neurological disorders such as Alzheimer's disease (AD), Parkinson's disease, and epilepsy, as well as psychiatric conditions like schizophrenia and bipolar disorder. Additionally, Type-2 diabetes mellitus (T2DM) is included due to its notable comorbidity and shared mechanistic pathways with schizophrenia and bipolar disorder. These CBM KGs were meticulously curated and encoded using the Biological Expression Language (BEL), which provides a robust semantic framework for representing complex causal interactions among diverse biological entities, including genes, proteins, metabolites, biological processes, phenotypes, and chemicals.

The KGs exhibit a high degree of granularity, encompassing thousands of causal relationships (edges) and biological entities (nodes), thereby supporting detailed mechanistic reasoning. A schema diagram (Figure 3) illustrates the multimodal and granular structure of these KGs, highlighting the diversity of nodes and relationships, while Table 1 summarizes key characteristics of the KGs.

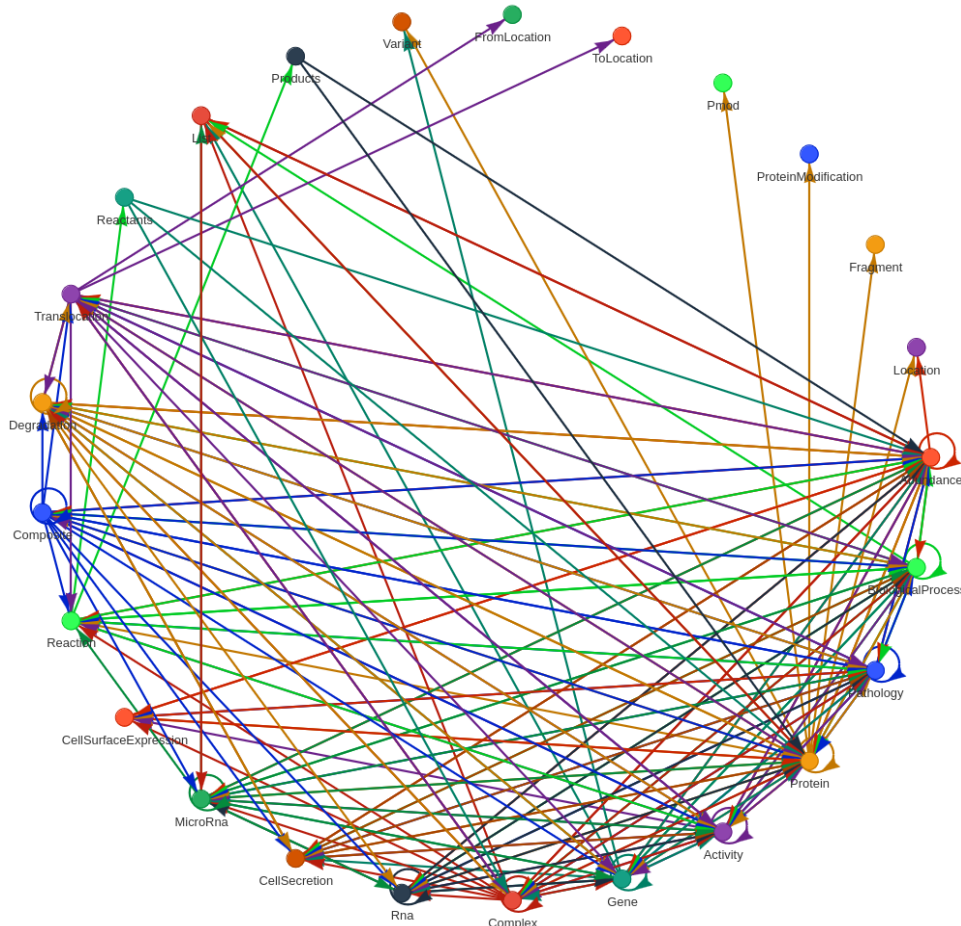


Figure 3: A schema diagram depicting the various node types (e.g., genes, proteins, biological processes, metabolites) and the causal relationships connecting them within the CBM knowledge graph.

Table 1: Statistical information about the CBM KGs.

Disease KG	Number of nodes	Number of edges	Number of pathways
Alzheimer's Disease	5,019	12,455	1,213
Parkinson's Disease	2,282	4,451	494
Epilepsy	2,390	7,614	642
Schizophrenia	2,836	5,760	1,934

Bipolar disorder	2,009	4,586	1,567
Diabetes Mellitus	425	617	182

5.3. Alzheimer Knowledge Graph with EEG Features

To facilitate extraction of knowledge from unstructured text and scientific publications, an essential EEG/MEG ontology was constructed and extended. Using this ontology, an EEG/MEG KG (KG) was developed by curating a corpus of research findings related to EEG biomarkers in AD. The corpus was initially compiled through a PubMed search targeting studies with titles or abstracts containing combinations of "Drug_name and EEG," "Protein_name and EEG," or "Drug_name and MEG." After filtering for studies on human subjects, the following datasets were obtained:

- 423 studies on Drug-EEG,
- 103 studies on Protein-EEG,
- 452 studies on Drug-MEG.

From these, 173 studies focusing on resting-state EEG with various drugs and proteins were manually selected. Additionally, the corpus was enriched with 17 review articles on AD-EEG and 10 reviews on AD mechanisms, resulting in a curated corpus of 200 papers. The entire corpus was annotated in Biological Expression Language (BEL) format and compiled using the ebel package to extract relevant entities from biological databases. In collaboration with our domain expert partners at Sapienza University of Rome (UNIROMA1), EEG terms were standardized, and an EEG curation guideline was developed and validated.

5.4. Neuro-PsychoMMSig (Neuro-Psychiatric Multimodal Signature)

The resulting KGs are integrated into the Mechanism Enrichment Server – Neuro-PsychoMMSig (Neuro-Psychiatric Multimodal Signature). This platform enables KG to serve as prior knowledge for multimodal datasets. By combining structured information from KGs with diverse biological datasets, the platform facilitates the application of advanced analytical algorithms.

Neuro-PsychoMMSig integrates six curated, disease-specific KGs, covering AD, Parkinson's disease, epilepsy, schizophrenia, bipolar disorder, and type 2 diabetes mellitus. By employing a KG-based framework, the platform organizes and integrates multifactorial and heterogeneous information associated with neurodegenerative and psychiatric disorders. The platform incorporates four advanced multi-omics analysis algorithms: CLinical Embedding of Patients (CLEP), Candidate Mechanism Perturbation Amplitude (CMPA), Gene Set Enrichment Analysis (GSEA), and Causal Robust Mapping in Meta-Analysis (CARMA). These algorithms enable researchers to uncover complex biological mechanisms, identify biomarkers, generate hypotheses, and stratify patients based on their unique data profiles. Additionally, the platform supports the identification of enriched causal pathways tailored to specific biological contexts or patient subgroups.

Neuro-PsychoMMSig consists of two main analysis methods: an explorative and clinical analysis, allowing the user to carry out network-based algorithms on -omics datasets using KGs as prior knowledge.

5.4.1. Explorative Analysis

Neuro-PsychoMMSig provides advanced qualitative exploration tools (**Figure 4**), enabling users to interactively query the CBM KGs using biomarkers. This interactive visualization facilitates the examination of subgraphs, the identification of relationship provenance, and the analysis of the structural and connectivity features underlying disease mechanisms. The visualization framework leverages D3.js for dynamic and responsive graph rendering. Available graph-based exploration functionalities include subgraph extraction and the analysis of local graph neighborhoods, allowing users to investigate specific biological contexts in detail.

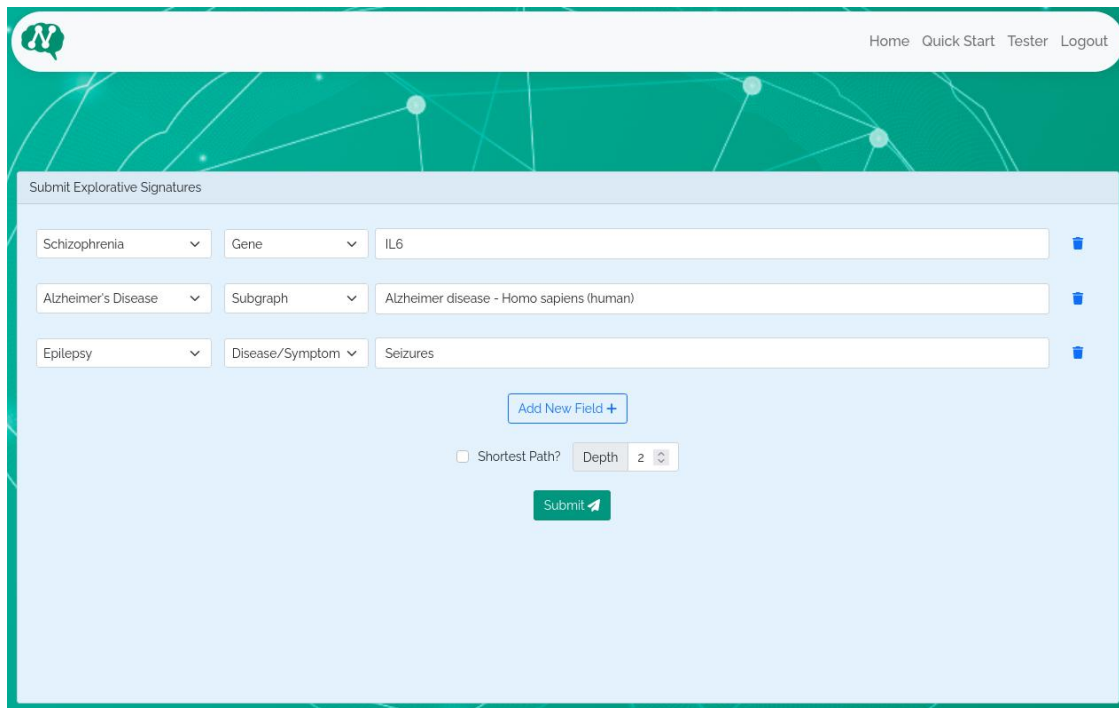


Figure 4: Explorative analysis interface in Neuro-PsychoMMSig.

5.4.2. Clinical Analysis

Neuro-PsychoMMSig offers robust data analysis capabilities, enabling users to conduct mechanistic analyses of multi-omics datasets integrated with disease-specific KGs. Each network-based algorithm is specifically designed to address distinct biological data modalities, such as gene expression and genome-wide association studies (GWAS).

By leveraging curated, domain-specific KGs as prior knowledge, these algorithms enhance the biological interpretability and contextual relevance of the analytical outcomes. Users can upload their omics datasets for analysis, which is performed asynchronously (**Figure 5**). The platform provides interactive visualizations of the results, complete with detailed provenance and biological context, enabling deeper insights into disease mechanisms.

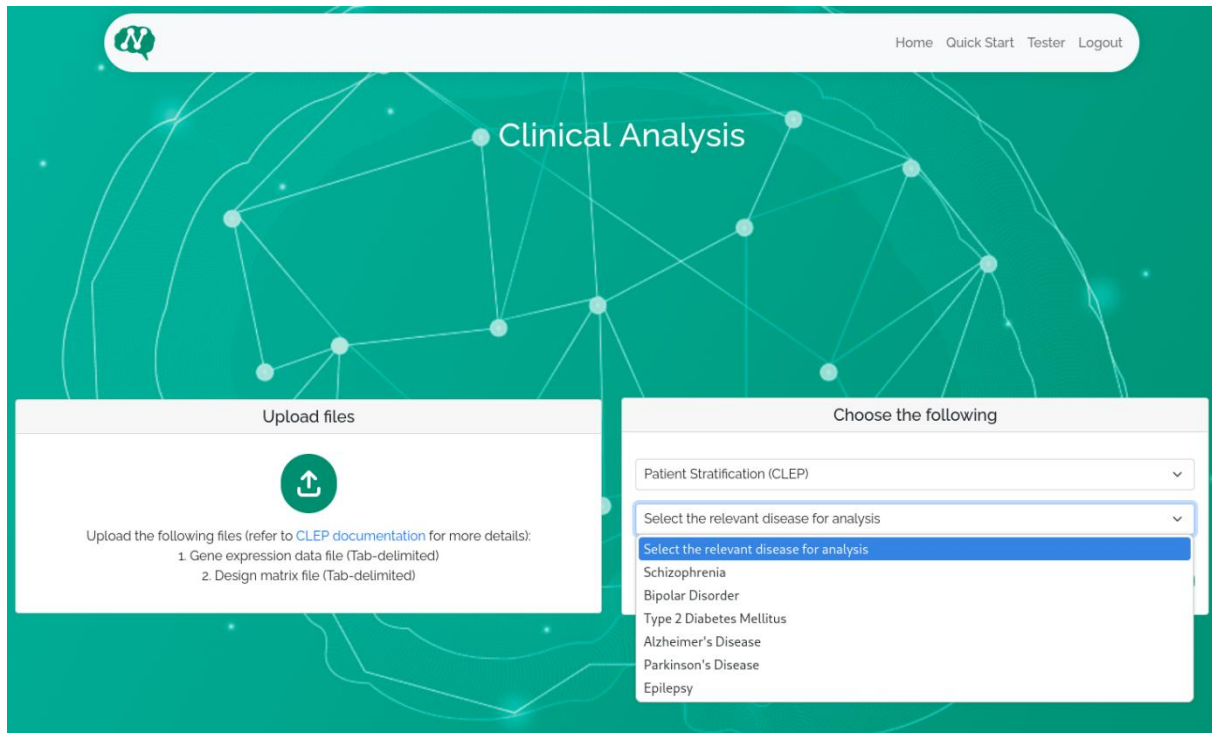


Figure 5: Pre-loaded disease KGs that can be used as prior knowledge for carrying out various data analysis algorithms.

The clinical analysis module incorporates four state-of-the-art computational algorithms:

a. Clinical Embeddings with Patient data (CLEP):

The CLEP algorithm integrates prior knowledge from domain-specific KGs into clinical datasets, enriching the interpretability of the data. For example, gene expression data can be analyzed in conjunction with the AD-KG. The workflow involves the following steps:

- Reference Distribution Calculation: For each gene in the dataset, an empirical cumulative distribution function (eCDF) was computed using the control group to establish a reference distribution.
- Patient Identification: Patients were identified as outliers if their gene expression values fell within the extreme ends of the eCDF, based on a predefined threshold (e.g., the top or bottom 5% of the distribution).
- Graph Integration: These outlier patients were linked to the corresponding gene in the KG with an edge, labeled as either +1 or -1, depending on whether the patient's expression value was in the upper or lower extreme of the distribution.
- Knowledge Graph Embedding: The augmented KG, now incorporating patient-specific data, was used to train a KG embedding model, RotatE. The embeddings generated for each patient were subsequently utilized for a classification task.

This approach enables the integration of patient-specific data into KG, facilitating downstream tasks such as patient stratification or disease classification.

b. Candidate Mechanism Perturbation Amplitude (CMPA)

CMPA algorithm is designed to assess the impact of genetic dysregulation on KGs and their subgraphs, such as pathways and biological processes. The workflow of CMPA is as follows:

- Identification of Dysregulated Genes: A gene expression dataset is analyzed to identify all dysregulated genes for each phenotype or subset of interest.

- Contextual Application in Knowledge Graph: The identified gene dysregulation is mapped onto the KG, allowing the algorithm to contextualize the dysregulation within the broader biological network.
- Effect Propagation: The impact of gene dysregulation is propagated throughout the network, enabling the calculation of the total effect on interconnected nodes and edges.
- CMPA Scoring: The propagated effects are quantified as CMPA scores for various subgraphs (e.g., pathways or biological processes). These scores are then compared to identify the most significantly impacted pathways or biological processes.

This approach provides a systematic way to link gene expression changes to higher-order biological mechanisms, offering insights into the pathways most affected by genetic dysregulation.

c. Gene Set Enrichment Analysis (GSEA)

GSEA is a computational approach used to evaluate whether predefined gene sets—such as pathways, functional categories, or other biologically relevant groups—exhibit statistically significant, coordinated expression changes under specific experimental conditions. The key steps in GSEA are as follows:

- Gene Ranking: All genes in the dataset are ranked based on their correlation with a particular phenotype or experimental condition (e.g., patients with schizophrenia versus controls).
- Gene Set Evaluation: GSEA assesses whether the genes within a predefined set are disproportionately represented at the extremes (top or bottom) of the ranked list. This indicates whether the gene set is enriched for genes strongly associated with the phenotype.
- Statistical Significance: The enrichment score is calculated for each gene set, and statistical significance is determined through permutation testing, which evaluates whether the observed enrichment is greater than expected by chance.
- Biological Insights: By identifying enriched gene sets, GSEA highlights dysregulated pathways or biological processes associated with the phenotype, such as pathways uniquely altered in patients diagnosed with schizophrenia compared to those without the diagnosis.

This method is particularly powerful for detecting subtle but coordinated changes in gene expression that might be missed when analyzing individual genes in isolation. It provides a pathway-level perspective, offering insights into the underlying biological mechanisms of complex phenotypes.

d. CAusal Robust Mapping method in meta-Analysis (CARMA)

We have also integrated CARMA, a Bayesian model designed for fine-mapping genome-wide association study (GWAS) datasets to identify putative causal single nucleotide polymorphisms (SNPs) within a sample. Fine-mapping approaches, such as CARMA, utilize probabilistic models to calculate posterior probabilities for each variant being causal. These models often incorporate prior knowledge from diverse sources, including epigenomic datasets, transcription factor binding sites, and chromatin accessibility profiles. This methodology has proven critical in bridging the gap between GWAS signals and functional biology, enabling researchers to prioritize specific variants for downstream experimental validation.

5.4.3. Case Scenario

a. Alzheimer's Disease

For benchmarking, we utilized the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, a leading resource for dementia research. The preprocessed dataset includes 260 cognitively healthy controls, 215 early mild cognitive impairment (MCI) patients, 225 late MCI patients, and 44 AD patients. For binary classification, all cognitively impaired patients ($n = 484$) were grouped into a single class. Preprocessed gene expression data [RMA normalized] was used as a baseline for benchmarking CLEP, with individual transcripts analyzed for genes with multiple transcripts. Blood plasma transcriptomic and GWAS data from ADNI were incorporated into the KG. Transcriptomic data processing for CMPA and GSEA was performed. Additionally, the ADNI GWAS dataset was analyzed using CARMA, leveraging summary statistics and functional annotations.

b. Age-Stratified Alzheimer's Disease Analysis

We analyzed the GSE33000 dataset, which includes postmortem prefrontal cortex samples from 624 individuals (demented and non-demented controls). Patients were stratified into four age groups: 50–69, 70–79, 80–89, and 90–100 years. Differential expression analysis was performed using Geo2R to compute log fold change (logFC) values for each group. The resulting profiles were analyzed with CMPA, and GSEA, identifying co-expressed gene pairs, disrupted pathways, and regulatory networks linked to neurodegenerative pathology across age groups.

c. Schizophrenia

For schizophrenia benchmarking, we used two gene expression datasets: GSE53987 and GSE21138. GSE53987 includes microarray profiles from schizophrenia, bipolar disorder, and major depressive disorder, with samples from the hippocampus, prefrontal cortex, and striatum of postmortem brains. We analyzed 48 schizophrenia and 55 unaffected samples using CMPA, GSEA, and CLEP. GSE21138 contains microarray profiles from the prefrontal cortex of 30 schizophrenia brains and 29 matched controls. Like GSE53987, it was analyzed using CMPA and GSEA to identify dysregulated pathways and gene expression changes.

6. Results

6.1. Alzheimer Knowledge Graph with EEG Features

The initial KG generated from this curation contained 4,603 nodes and 12,848 edges. Subsequent integration of gene-protein relations expanded the graph to 5,649 nodes and 13,935 edges. Further enrichment with drug-protein and protein-protein relations from databases such as DrugBank, IUPHAR, ChEBI, and IntAct resulted in a comprehensive graph with 25,674 nodes and 105,327 edges. This intricate network captures gene-protein-disease-drug-EEG relationships and serves as a foundational resource for analyzing AD pathways, comorbidities, and brain region-pathway mappings (**Figure 6**).



Figure 6: Alzheimer knowledge graph with EEG features


The KG supports several downstream applications, including:

- **Patient-Specific Embeddings:** Mapping EEG features from datasets to corresponding nodes in the KG to generate embeddings for patient stratification and precision medicine.
- **Drug Repurposing Benchmarking:** Evaluating the impact of EEG data integration by comparing drug repurposing performance on the AD-EEG KG versus a version of the KG without EEG nodes.

This KG represents a significant step toward leveraging EEG-informed embeddings for advancing AD research and personalized therapeutic strategies.

6.2. Neuro-PsychoMMSig (Neuro-Psychiatric Multimodal Signature)

We developed Neuro-PsychoMMSig, a web platform designed for intuitive exploration and analysis of its graph database. Users can generate subgraphs based on biomarkers such as genes, proteins, and biological processes across all diseases included in the platform. The platform also supports advanced data analysis using KG algorithms, including CLEP, CARMA, GSEA, and CMPA, enabling the identification of dataset-specific insights and the comparison of overlapping results across algorithms. A quick start guide is available at the website for easy operation (**Figure 7**).


Home Quick Start Tester Logout

How to use Neuro-PsychoMMSig

About
Explorative Analysis
Clinical Analysis

Neuro-PsychoMMSig is a web-based tool that allows users to carry out network based algorithms on -omics datasets. The tool comes pre-installed with knowledge graphs for schizophrenia, bipolar disorder, Alzheimer's disease, Parkinson's disease and type-2 diabetes mellitus. These knowledge graphs are either used to identify enriched mechanisms, or pathways using algorithms like the Mechanism Enrichment Analysis (CMPA) or the Pathway Enrichment Analysis (GSEA). They can also be used to identify putatively causal SNPs using finemapping approach and an embedding of the samples in the patient data can be generated using the CLEP algorithm. All the algorithms run on the server and are **NOT** uploaded to the cloud to avoid data privacy issues. The results of the algorithms are displayed in the form of interactive visualizations or downloadable files. This method of analysis is called the **Clinical Analysis**.

The **Explorative Analysis** on the other hand, is another way to interact with the knowledge graphs. It allows users to query the knowledge graphs using biomarkers and visualize the results in the form of interactive graphical visualizations. This visualization can be interacted with to identify the provenance of the relationships and the publication source for every interaction. Furthermore subgraphs/biological processes from gene ontology can be isolated for a visualization.

Figure 7: The quick start guide for Neuro-PsychoMMSig

6.2.1. Explorative Analysis

The explorative analysis feature of Neuro-PsychoMMSig enables researchers to uncover both shared and unique biomarkers across different diseases, shedding light on inter-disease relationships. For instance, users can query multiple KG), such as those for Alzheimer's and Parkinson's disease, to identify overlapping interactions, like those linked to mitochondrial dysfunction. Conversely, querying disease-specific biomarkers, such as those for schizophrenia, may reveal unique metabolic pathways. Targeted exploration within a single disease KG provides deeper insights into disease-specific mechanisms, such as protein and biological process interactions relevant to AD. The KG visualization page further enhances this functionality by allowing users to inspect individual edges within the graph in detail, as shown in **Figure 8**. By clicking on an edge, users can access provenance information, including supporting publications, ensuring transparency and enabling validation of the biological interactions identified. This feature facilitates a deeper understanding of the mechanisms underlying complex diseases.

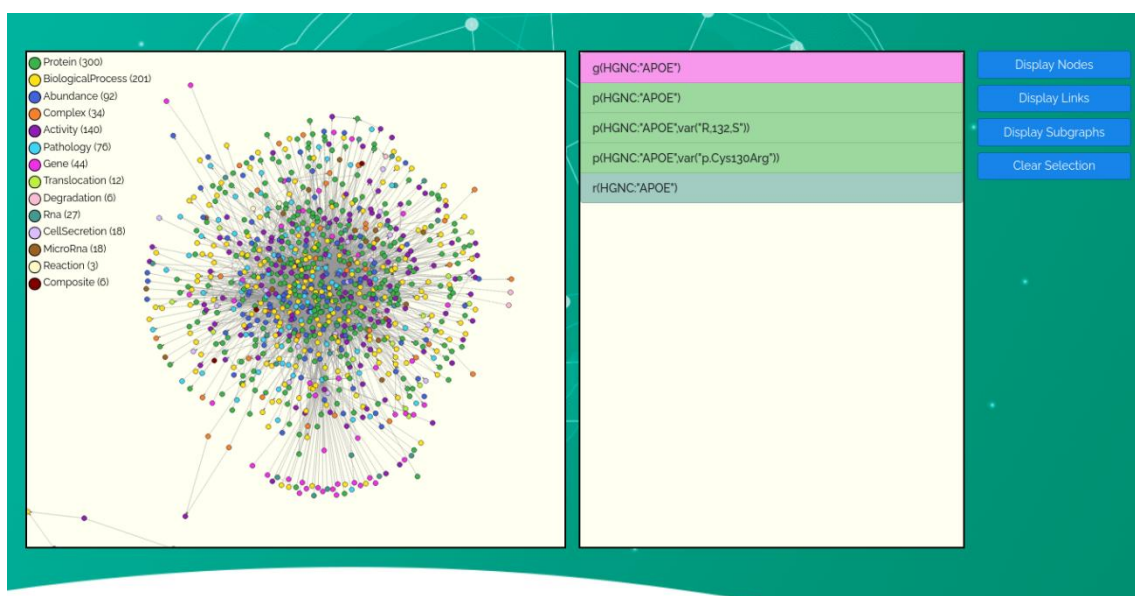
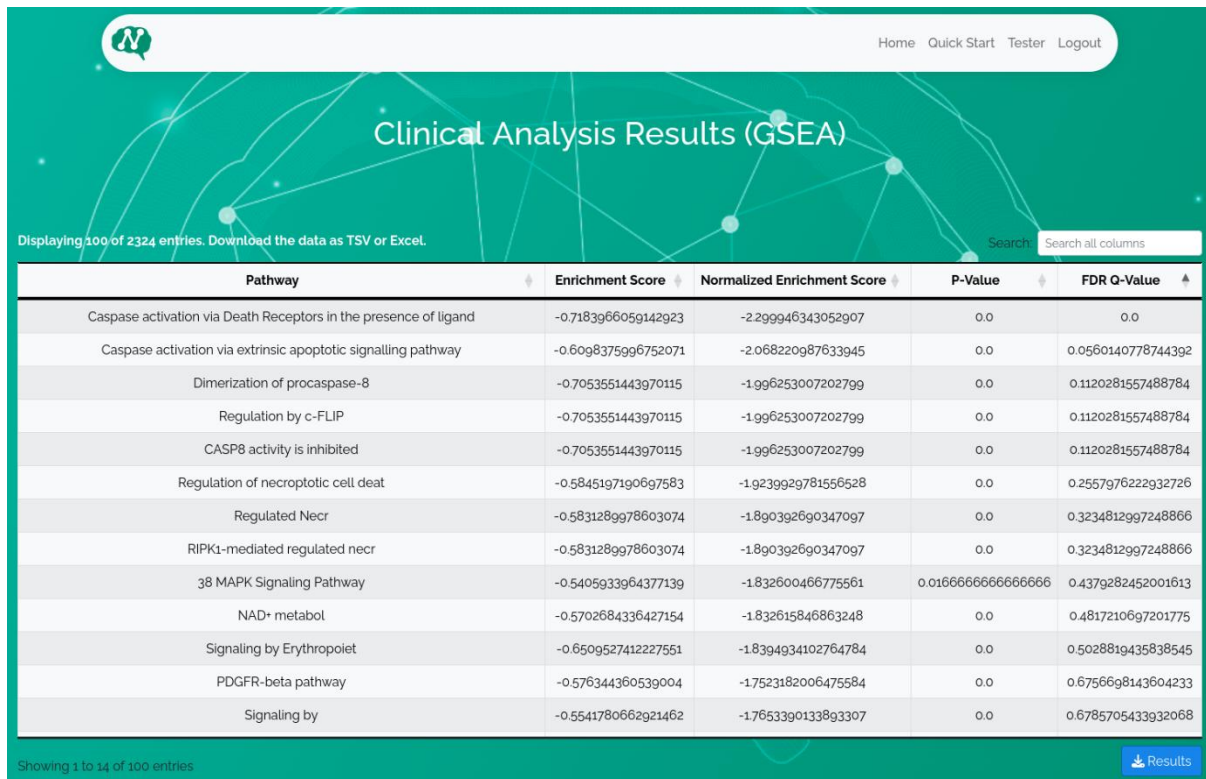


Figure 8: The visualization page for explorative analysis of KGs

6.2.2. Clinical Analysis

After selecting the analysis type and uploading the relevant dataset, users can submit their job, which is processed asynchronously using RabbitMQ workers via the Celery Django Python library. The processing time varies depending on the selected analysis, ranging from 1 minute to 24 hours.

Upon job completion, the platform automatically sends an email notification to the user, including a URL to the results page (**Figure 9**). Additionally, users can access their submitted jobs through the Accounts tab, which provides a comprehensive list of all jobs associated with the account. This ensures easy tracking and retrieval of analysis results.



Pathway	Enrichment Score	Normalized Enrichment Score	P-Value	FDR Q-Value
Caspase activation via Death Receptors in the presence of ligand	-0.7183966059142923	-2.299946343052907	0.0	0.0
Caspase activation via extrinsic apoptotic signalling pathway	-0.6098375996752071	-2.068220987633945	0.0	0.0560140778744392
Dimerization of procaspase-8	-0.7053551443970115	-1.996253007202799	0.0	0.1120281557488784
Regulation by c-FLIP	-0.7053551443970115	-1.996253007202799	0.0	0.1120281557488784
CASP8 activity is inhibited	-0.7053551443970115	-1.996253007202799	0.0	0.1120281557488784
Regulation of necroptotic cell death	-0.5845197190697583	-1.9239929781556528	0.0	0.2557976222932726
Regulated Necr	-0.5831289978603074	-1.890392690347097	0.0	0.3234812997248866
RIPK1-mediated regulated necr	-0.5831289978603074	-1.890392690347097	0.0	0.3234812997248866
38 MAPK Signaling Pathway	-0.5405933964377139	-1.832600466775561	0.016666666666666666	0.4379282452001613
NAD ⁺ metabol	-0.5702684336427154	-1.832615846863248	0.0	0.4817210697201775
Signaling by Erythropoiet	-0.6509527412227551	-1.8394934102764784	0.0	0.5028819435838545
PDGFR-beta pathway	-0.576344360539004	-1.7523182006475584	0.0	0.6756698143604233
Signaling by	-0.5541780662921462	-1.7653390133893307	0.0	0.6785705433932068

Figure 9: The result page for GSEA analysis

6.2.3. AD Case Study

In this study, we utilized the CMPA algorithm to assess perturbation amplitudes of Gene Ontology (GO) biological processes within disease-specific networks. Genetic and biological data were sourced from ADNI dataset. The CMPA algorithm was employed to quantify the impact of AD-associated genes on the AD-KG, enabling the evaluation of how these genes disrupted key biological processes.

Analysis of the AD-specific network revealed dysregulation in several GO biological processes, including cell growth (GO:0016049), G1/S phase cell cycle transition (GO:0044843), stress response to copper ions (GO:1990169), and neurofibrillary tangle assembly (GO:1902988). These findings align with prior research that highlighted the critical role of cell cycle dysregulation in AD, noting that cyclins and cyclin-dependent kinases (CDKs), which regulate the cell cycle, are significantly elevated in neurons of AD patients. Similarly, copper ion dysregulation has been implicated in AD pathogenesis, with elevated copper levels observed in AD patients and animal models, where chronic copper exposure has been linked to neurodegeneration.

GSEA further corroborated these findings, revealing dysregulation in pathways such as “Metallothionein bind metals” and “Response to metal ions.” Additionally, the association between neurofibrillary tangle assembly and AD is well-established, further supporting the observed results.

6.2.4. AD sample stratification by age

The CMPA analysis of the GSE33000 dataset was conducted to examine patterns of dysregulation across different age groups of AD patients. Although the dataset is not a time-series dataset, patients were categorized into age bins (50–69, 70–79, 80–89, and 90–100 years). The analysis revealed a positive trend in the dysregulation of biological processes such as neuron differentiation (GO:0030182), amyloid fibril formation (GO:1990000), and cell cycle G1/S phase transition (GO:0044843), indicating an increase in dysregulation with age. Conversely, a negative trend was observed in processes such as regulation of the MAPK cascade (GO:0043408), neutrophil aggregation (GO:0070488), and positive regulation of the ERAD pathway (GO:1904294).

The upward trend in neuron differentiation and amyloid fibril formation dysregulation aligns with expected AD pathology, while the dysregulation of the G1/S phase transition is consistent with findings from the ADNI dataset. The role of neutrophil adhesion in memory impairment has been demonstrated in mouse models and the downward trend in the dysregulation of the ERAD pathway has been linked to neuronal degeneration.

GSEA across all age groups identified the upregulation of genes involved in cristae formation (R-HSA-8949613), consistent with prior studies suggesting that amyloid oligomers may induce ion channel formation in mitochondria. Additionally, GSEA revealed the downregulation of genes associated with Interleukin-27 signaling (R-HSA-9020956), a pathway implicated in neuroprotective and neuroinflammatory processes. These findings provide further insights into age-related dysregulation in AD.

6.2.5. Schizophrenia Case Study

The CMPA analysis of the GSE53987 and GSE21138 datasets did not identify any relevant mechanisms of interest. In GSE53987, only two biological processes—cell death (GO:0008219) and autophagy (GO:0006914)—were notably downregulated. In contrast, GSE21138 exhibited an unexpected upregulation of Gamma Rhythm (MESH: D065818). While autophagy and cell death have been linked to schizophrenia, with evidence suggesting that impaired autophagy predisposes individuals to neurodegeneration and mood disorders, gamma rhythm is typically negatively correlated with schizophrenia. This negative correlation aligns with our schizophrenia KG, leading us to hypothesize that the observed gamma rhythm upregulation is a dataset-specific anomaly.

GSEA for both datasets failed to reveal any significantly dysregulated pathways ($FDR \leq 0.05$). However, we observed the downregulation of pathways such as the “IL17 signaling pathway” (WP2112) and “Caspase-mediated cleavage of cytoskeletal proteins” (R-HSA-264870). The IL17 signaling pathway, an immune response pathway, includes IL17, a protein considered a predictive biomarker for schizophrenia. Notably, the IL17 gene is typically upregulated in schizophrenia patients. Similarly, the caspase-mediated pathway involves caspase family proteins (e.g., CASP3, CASP7, CASP8), which are upregulated in schizophrenia and associated with apoptosis. We hypothesize that these discrepancies arise due to the lack of significant GSEA results, limiting the interpretability of these findings.

7. Conclusion, next steps

We have developed Neuro-PsychoMMSig, a web-based research platform designed to facilitate the exploration and analysis of neuropsychiatric and neurodegenerative disease mechanisms. By integrating disease-specific KGs with multi-omics data, the platform provides a flexible framework for researchers to perform tasks such as mechanism enrichment, patient stratification, and pathway analysis. It leverages established algorithms, including CMPA, CLEP, GSEA, and CARMA, to support comprehensive data analysis.

Rather than replacing existing workflows, Neuro-PsychoMMSig complements them by offering a customizable and extensible environment. A key feature of the platform is its support for self-hosted deployment, allowing researchers to securely analyze sensitive or proprietary datasets offline. This capability also enables users to incorporate custom KGs or extend the platform with new algorithms tailored to their specific research objectives.

Future work will focus on enhancing the platform's usability and interoperability. Planned improvements include integrating additional KGs and algorithms to support diverse data modalities, such as neuroimaging and longitudinal clinical records, thereby expanding the platform's applicability across a broader range of research domains.

Disclaimer

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101058516. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or other granting authorities. Neither the European Union nor other granting authorities can be held responsible for them.