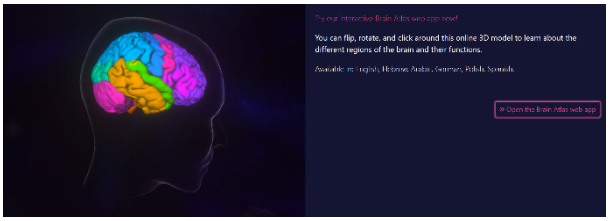
VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease



Public deliverable report

D5.2: Digital atlas that maps signaling pathways on corresponding brain regions for different neurodegenerative states in humans

Date	April 2021
Authors	Petra Ritter (CHARITE) Alpha Tom Kodamullil, Marc Jacobs, Aliaksandr Masny, Sepehr Golriz Khatami, Stephan Springstubbe (Fraunhofer SCAI)
Dissemination level	**public**
Website	https://VirtualBrainCloud-2020.eu

This project has received funding from the **European Union´s Horizon 2020** research and innovation programme under **grant agreement No 826421**

Table of content

1. Introduction

1.1. Mapping of clinically relevant signaling pathways on the human brain

In the present report we describe an outcome of the Virtual Brain Cloud (TVB-Cloud) project. We have developed a methodology to map information from knowledge bases to a parcellated brain model – as we build and use it with The Virtual Brain (TVB) simulation platform. By bringing the brain model and atlas information in the same space, features from atlases can inform and constrain model parameters and hence increase biological plausibility of simulations or adjust them for alterations due to disease pathogenesis.

One aim of this task is to identify the pathways which are specific for various brain regions and that alter in neurodegenerative disease. This is to interpret the high-level functional consequences of molecular interactions specific to certain brain regions. This is done by linking the signaling pathways, which is in this case the 'cause-effect mechanisms' to their underlying anatomical locations.

Central to this work is The Virtual Brain (TVB, www.thevirtualbrain.org). TVB offers an open-source neuroinformatic platform for brain simulation based on individual structural connectivity[1,2]. It has been widely used in the exploration of dynamics in the healthy brain as well as in research on AD as well as for stroke, epilepsy and brain tumors. In TVB-Cloud, we provide a proof-of-concept for refined brain simulation by using additional knowledge sources for informing the model parameters.

The central outcomes of this work are

1) A framework and software that maps features form knowledge bases obtained through text mining to the 3D brain space. A detailed description can be found in pur publication "Integrating biological knowledge with multi-scale brain simulation – The Virtual Brain knowledge base adapter (TVBase)" (Stefannovski et al. 2021).

2) An interactive web atlas viewer of selected pathways as extracted with the tool described in (1) – developed by CHARITE:

   https://www.brainsimulation.org/atlasweb_multiscale/

In addition, Fraunhofer SCAI has built a set of ontologies and terminologies which are integrated into an **o**ntology **l**ookup **s**ervice. This OLS act as the basis for our text mining software ProMiner to retrieve the specific corpus as well as to retrieve the information of relevant pathways in a specific region of brain. Fraunhofer SCAI also have developed NeuroMMSig, which is an inventory of multimodal molecular signatures of cause-effect subgraphs in the

[1] Ritter, P., M. Schirner, A. R. McIntosh and V. K. Jirsa (2013). "The virtual brain integrates computational modeling and multimodal neuroimaging." Brain Connect **3**(2): 121-145.
[2] Sanz Leon, P., S. A. Knock, M. M. Woodman, L. Domide, J. Mersmann, A. R. McIntosh and V. Jirsa (2013). "The Virtual Brain: a simulator of primate brain network dynamics." Frontiers in Neuroinformatics **7**(10).

context of Alzheimer's disease (AD). In the context of this deliverable, we have established the mapping among specific brain regions to the pathways based on the evidence from literature and then via pathways to the cause-effect subgraphs from NeuroMMSig.

2. Partners involved

LEAD: CHARITE, Partners involved: FRAUNHOFER, AMU, CODEBOX, CODEMART

3. Description of work performed

3.1 Tools and Resources

For establishing the tasks 5.2, we have developed the following resources and tools:

3.1.1 The Virtual Brain Knowledge Base Adapater TVBase

See Stefanovski et al. 2021 for a detailed description.

Our knowledge about the brain and neurological diseases is growing fast and covers findings from various research fields, e.g., genetics, neuroimaging, and computational neuroscience. The accessibility of this knowledge is increasing through publicly available databases derived from automated large-scale literature mining tools, e.g. SCAIView. Yet there remain integration gaps between the different scales of brain organization. To overcome these gaps and to enable the integration of various knowledge sources in self-consistent whole-brain models, we have developed a methodology to map information from automated literature-mining derived knowledge bases to 3D parcellated brain models – as we build them with The Virtual Brain simulation tool, i.e. a TVB knowledge base adapter (TVBase). By bringing the brain model and information from knowledge bases in the same brain coordinate space, features from these knowledge bases can inform and constrain model parameters and hence increase biological plausibility of simulations or adjust them for alterations due to disease pathogenesis. A transformation matrix links the comprehensive anatomical ontology Uberon to standard brain atlas parcellations. With our tool, user defined spatial feature maps can be extracted from the knowledge base SCAIView that can inform multi-scale brain simulation with The Virtual Brain. We demonstrate how these feature maps can serve as a proxy for biological features obtained from neuroimaging and neuropathological measurements or provide anatomical representation of biochemical pathway alterations in neurodegenerative diseases. We further validate our new approach by comparing simulation inferred predictions from an established brain imaging model to a brain model derived from a corresponding feature map. In summary our novel approach allows to map – in real time – information from the literature-mining knowledge base derived from SCAIView to the 3D standard brain space for effortless integration of biological information in multi-scale whole brain models.

Sources. We use existing knowledge from various sources: Empirical receptor densities from EBRAINS (Amunts, Mohlberg et al. 2020), PET data from ADNI (Frisoni and Weiner 2010, Weiner, Veitch et al. 2017, Stefanovski, Triebkorn et al. 2019), the neuropathological Braak classification (Braak and Braak 1991, Braak and Braak 1997, Braak, Alafuzoff et al. 2006) and the Knowledge Graph database NeuroMMSig (Domingo-Fernández, Kodamullil et al. 2017).

Receptor data from EBRAINS, ADNI, and Braak classification directly provide data suitable for the pipeline's later steps. NeuroMMSig provides 125 "subgraphs" associated with AD, e.g., referring to the glutamatergic system, amyloidogenesis, Tau, etc. Exemplarily, we show in this figure the glutamatergic subgraph, which consists of proteins, genes, imaging factors, MeSH terms, and other classes of knowledge. We extract the MeSH terms for further processing. Besides, we add MeSH terms referring to receptors where empirical data in EBRAINS exists. These MeSH terms are used for a query in SCAIView, a database of highly annotated scientific literature (Hanisch, Fundel et al. 2005). As one possible result of SCAIView, the user can extract statistics about the mentioning of anatomical terms, annotated by Uberon (Mungall, Torniai et al. 2012).

Raw Data. The Uberon statistics provide one value of relative entropy, a relevance measure per Uberon term, which is calculated by the appearance of this term in the query's selected subset. This already gives an anatomical assignment of the search MeSH term, but it is not mapped on a 3D brain. Instead, SCAIView provides a lexical mapping. This can be seen as raw data, comparable to autoradiography of receptor tracers from EBRAINS or the unprocessed PET from ADNI.

Topological assignment. Autoradiography data and PET are already 3D data, so the assignment can happen by co-registration of the data array with a brain parcellation and averaging data intensities per region. For Uberon statistics, a further method is necessary to transfer the lexical mapping into a 3D mapping. Therefore, we created a transfer matrix that assigns each Uberon term to the corresponding region of the Glasser parcellation (Glasser, Coalson et al. 2016). Consecutively, we can create an array that lists the relative entropy for a SCAIView search for each parcel of the Glasser atlas, which we call semantic association map.

Atlases. ADNI PET data is also co-registered to Glasser, and the Braak disease stages were manually assigned to Glasser areas. Julich-Brain receptor data is given in the Julich-Brain parcellation (Amunts, Mohlberg et al. 2020). The volumetric Julich-Brain data were projected onto the cortical surface and then parcellated by Glasser.

Brain Maps. Finally, for each MeSH term, BRAAK stage, receptor autoradiography, and disease stage in ADNI, we get one unique Glasser-parcellated brain map. By registering this map to the fsaverage brain of freesurfer, it is possible to re-sample it vertex by vertex and transfer it to other parcellations. As an example, we provide it here also for Julich-Brain, Desikan-Killiany (Desikan, Ségonne et al. 2006), and Destrieux (Destrieux, Fischl et al. 2010).

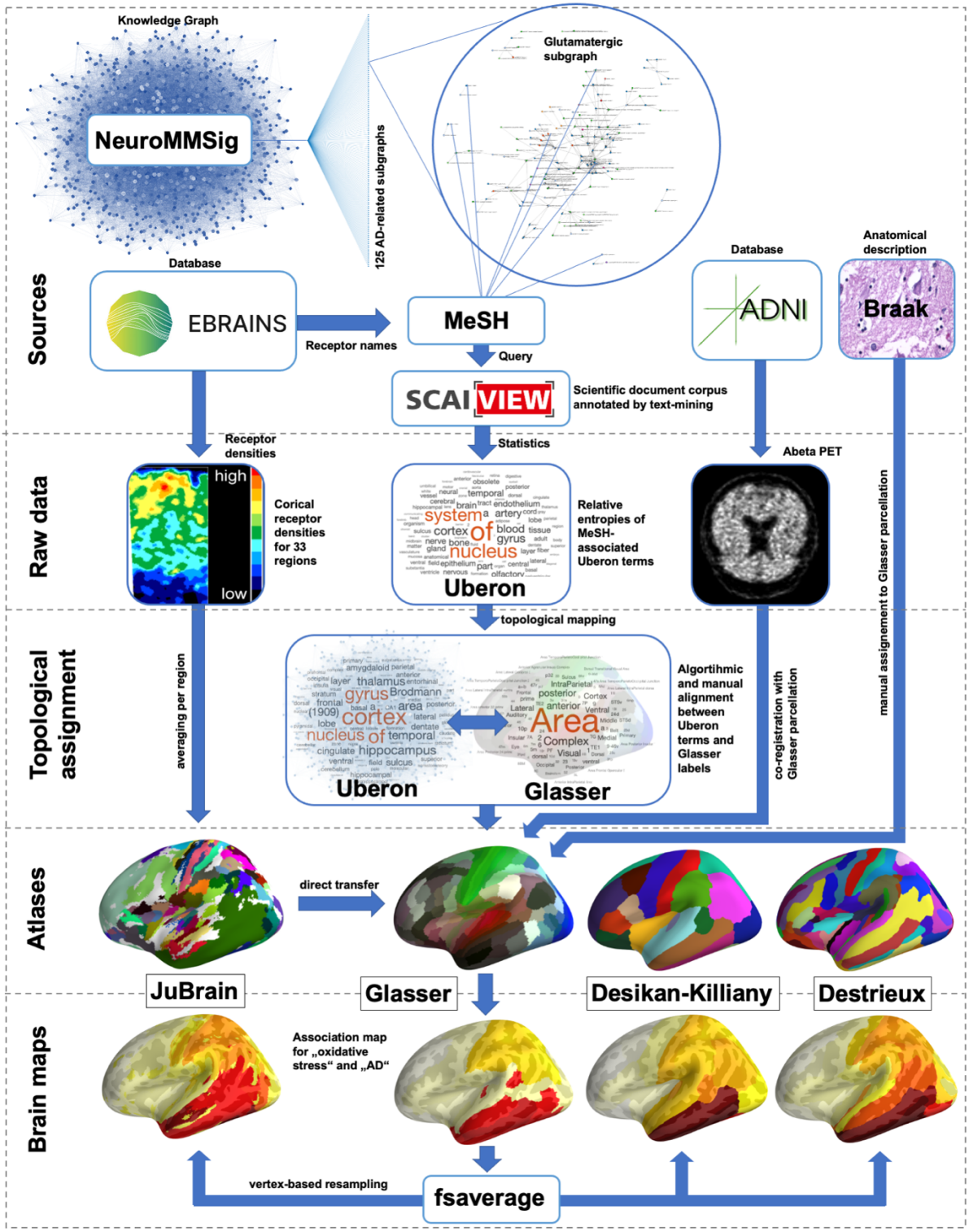


***Figure 1**. Workflow for the multi-source mapping methodology (Stefanovski et al. 2021)*

3.1.2 Interactive viewer of brain maps (receptor densities, proteinopathies, signaling pathways)

The Brain Atlas is an interactive software made in the cross-platform game engine Unity, where one can touch or click on a 3D brain model to learn more about different brain regions and brain functions. One can rotate and scale the brain for better views of brain regions, or

click a brain region or function from the side menu to automatically highlight them and rotate the model to an optimal viewing angle.

- The software is organised to facilitate translation and localisation via JSON files that can be added to the built applications resources folder. This means anyone can supply new translation files to the built application, with no need to access the source code and build again. Currently the application supports right-to-left and left-to-right languages, such as English, Polish, Arabic, Hebrew, and German.

The source project implements a Tree Node system for automatic generation of the concept of regions (and sub regions), and collections of regions into functions to facilitate modifications to the current Atlas, and the addition of new Atlases. To construct a new atlas, one defines regions (and associated sub regions) in arrays, and collections of regions into arrays for corresponding functions. The Tree Node system will search the scene for segmented brain mesh fragments with the same name, and assign the associated descriptive texts from the corresponding JSON file keys in the current language selected at run time.

The simple and flexible software was designed this way to quickly, accessibly and attractively display brain atlases for educational purposes for people of all ages and education levels.

URL: https://www.brainsimulation.org/atlasweb_multiscale/

3.1.3 Integration of Ontologies for TVB Cloud as the semantic layer

TVB Cloud OLS

This is the specific Ontology Look up service built for TVB-Cloud project. This can be accessed via https://ols.neuro.scaiview.com/index The TVB-Cloud OLS is based on the 'Ontology Lookup Service' (OLS) that is developed by European Bioinformatics Institute (EBI) for exploring ontologies. TVB-Cloud OLS is a repository for biomedical resources that aims to provide a single point of access to the latest ontology and terminology versions for translational neurodegeneration disease research. This version of OLS is used as the semantic layer for annotating the relevant entities for retrieving the specific literature that could be used for extracting the cause-effect mechanisms as well as to retrieve the information about which pathways are co-associated with which brain regions.

In order to achieve these outcomes, we have integrated the following seven ontologies in the TVB Cloud OLS:

- 7 ontologies (the brain region cell ontology, the FMA curated brain regions ontology, the neuro names ontology, the pathway tool, the PTS pathway dictionary, the virtual brain ontology, Uber-anatomy ontology)
- 34,331 terms
- 539 properties

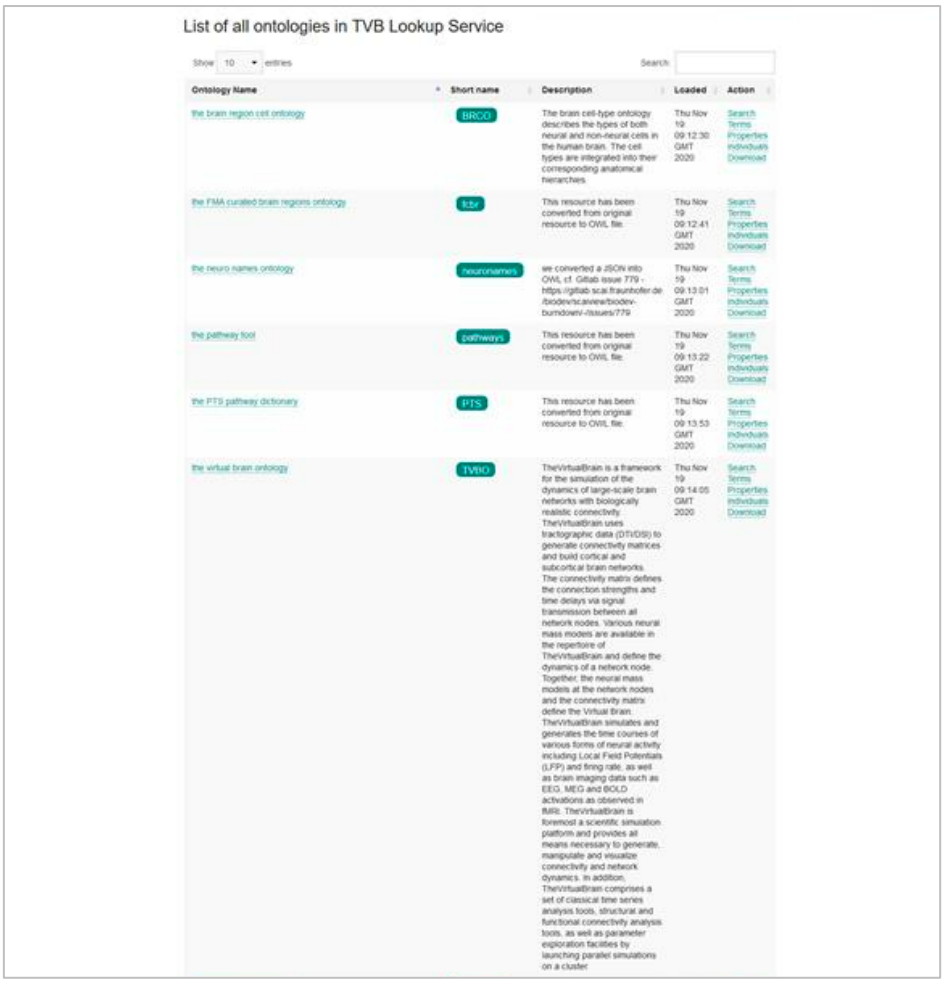


Figure 1: Instance of existing ontologies in TVB OLS

3.1.4. List of the ontologies integrated in TVB Cloud OLS:

- BRCT – Brain Region Cell Type ontology

The BRCT stands for Brain Region and Cell Type terminology, which was built with the purpose of organizing the knowledge domain of brain anatomy with a top-down granularity, from gross regions to cell types. BRCT describes the types of both neural and non-neural cells in the human brain and the cell types were integrated into their corresponding anatomical hierarchies. BRCT is dedicated terminology derived from Brain Region and Cell type Ontology (BRCO). BRCT contains more than 1,600 classes with average number of 5 children which showed a satisfactory F-score (0.80) in a named entity recognition task on an independent testing corpus composed of 100 manually annotated MEDLINE abstracts. Annotating unstructured data (e.g., from large cohort data sets like ADNI or the UK biobank or literature) using BRCT allows to apply automated workflows to link imaging features with other interested biological entities such as genes, pathways.

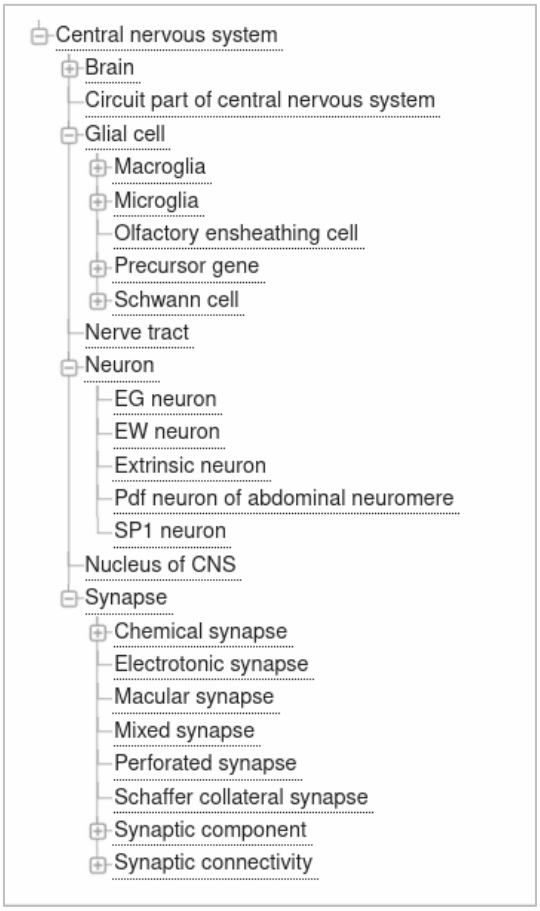


Figure 2: Instance of existing brain regions in BRCT

- PTS – Pathway Terminology System

The PTS stands for Pathway Terminology System consists of 6,596 classes, organized in a hierarchical structure. Each class has been annotated, whenever available, with additional information such as definition, source identifiers, and synonyms PTS was created by integrating terms from the Integrating Network Objects with Hierarchies (INOH) and Pathway Ontology (PW) and was populated with pathways from four popular public biomedical pathway databases, namely KEGG, Reactome, BioCarta, and Pathway Interaction Database. Further enrichment of the PTS was achieved by analysis of relevant scientific text. For this purpose, the phrases (with frequency of occurrence) from MEDLINE abstracts containing the word "pathway" and three words preceding it were extracted using a sub-corpus. If the pathway name inside the four-word phrase was either absent from the pathway reference name list or absent from the pathway synonym list, it was added to the pathway reference name dictionary. Integrating the PTS into the literature-mining environment such as SCAIView (the environment that enabled us to perform very context-specific literature searches based on combined semantics from multiple ontologies and terminologies), allows us to link pathways with other interested biological concepts such as imaging features. For example, we would be able to query the literature for cellular pathways relevant to the brain anatomy in both healthy and AD conditions.

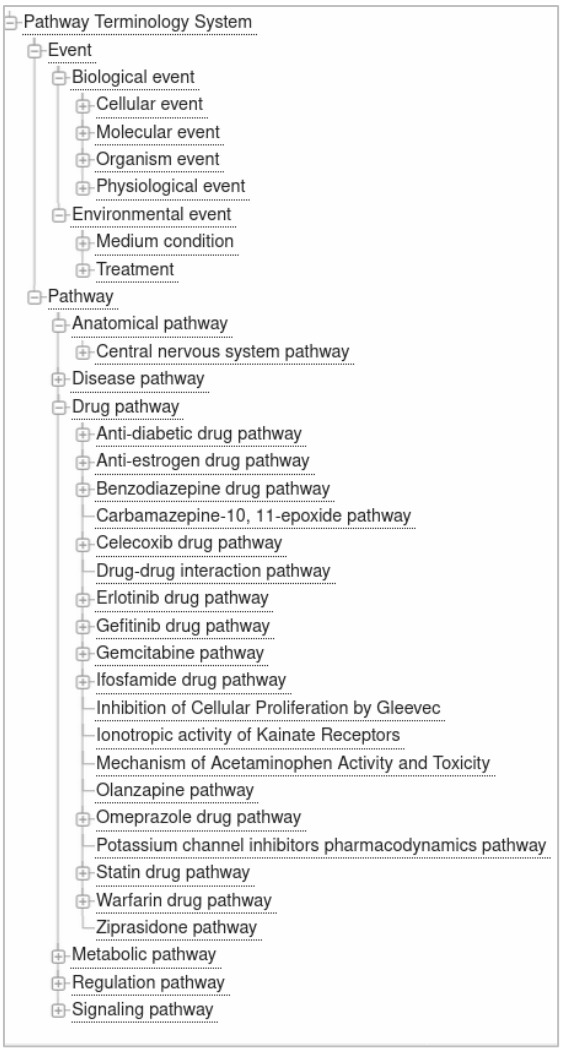


Figure 3: Instance of existing pathway names in PTS

- NeuroFMA - Neuro Foundational Model of Anatomy

The Foundational Model of Anatomy (FMA) Ontology provides a semantic framework for representing the anatomical entities and relationships that constitute the phenotypic organization of the human body. Similarly, the neuroanatomical content of the FMA models cytoarchitectural and morphological regions of the cerebral cortex, as well as white matter structure and connectivity. This modeling effort is driven by the need to correlate and reconcile the terms used in neuroanatomical labeling protocols. By providing an ontological framework that harmonizes multiple views of the neuroanatomical organization, the FMA provides developers with reusable and computable knowledge for a range of biomedical applications. The neuroanatomical content of the FMA was enhanced with detailed modeling for cerebral hemisphere brain labeling schemes, cerebral sulci, white matter structures, and neural connectivity relationships. NeuroFMA contains 2300 classes which by their incorporation we can rich dataset annotations and provides a means to correlate and integrate the findings with other external data and studies as discussed in the following section.

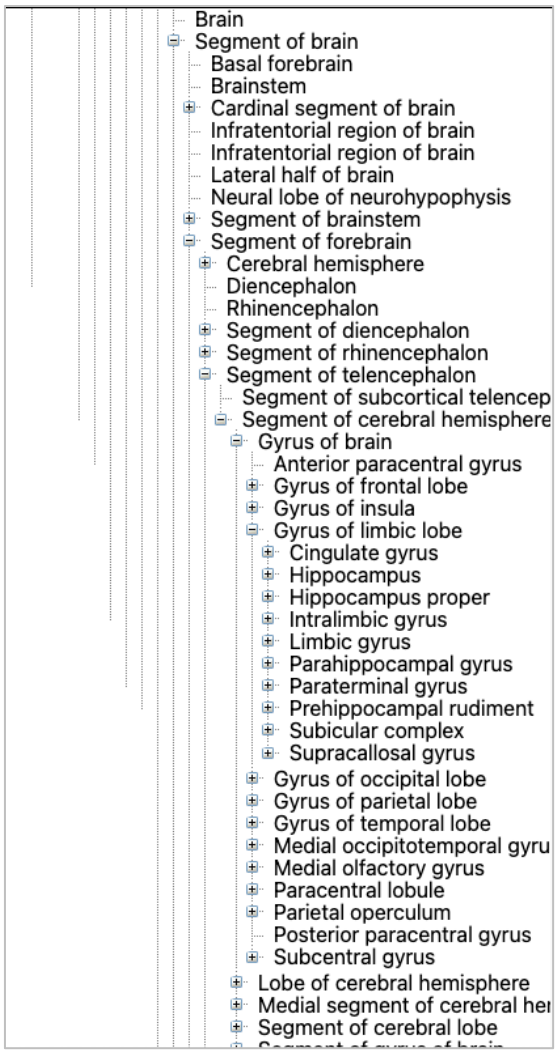


Figure 4: Instance of existing neuroanatomical terms in Neuro FMA

- NeuroNames Standard Nomenclature and Ontology

NeuroNames Ontology and Standard Nomenclature contains of some 3000 central nervous system structures in the human, macaque, rat and mouse. This version includes: 1) a set of unique Standard Terms for controlled vocabularies that enable unambiguous neuroanatomical tagging and indexing of data in websites and databases; 2) Synonyms for interpretation of search queries in multiple terminologies, including a comprehensive set of English and Latin terms found in major textbooks, research publications and brain atlases as well as terms for the most common brain structures in six other languages: French, German, Indonesian, Italian, Russian and Spanish; 3) operational Definitions of all structures (the feature of NeuroNames that makes it an ontology as well as a nomenclature); 4) Parents of structures in several different hierarchical models of neuroanatomical organization; 5) URLs to BrainInfo pages with links to further information on structures.

- ADO - Alzheimer's disease ontology:

Alzheimer's disease ontology (ADO) was developed with the purpose of containing information relevant to four main biological views—preclinical, clinical, etiological, and molecular/cellular mechanisms—and was enriched by adding synonyms and references. A first collection of terms and concepts related to AD was generated by scanning various knowledge sources, including review articles, content of online books, standard knowledgebases, encyclopedias, glossaries, and informative online sources and websites. Most of the disease-specific knowledge was acquired by performing Web-based searches on re-sources that focus on various aspects of the disease, such as treatment aspects, research overviews, neurological perspectives, diagnostic criteria, research proceedings, and so forth. Concepts were also extracted from compendiums such as the Encyclopedia of Alzheimer's Disease and the Atlas of Alzheimer's Disease. Another glossary of terms representing AD was collected from the Bibliography and Dictionary of Alzheimer' Disease. Whenever possible, any available hierarchical organization (structure) of the concepts was also extracted along with the concepts themselves. Corresponding definitions and synonyms were also taken into consideration as additional annotation of the concepts. ADO contains 1,565 classes and provides a first draft version of a semantic framework that is supposed to support research on Alzheimer´s Disease. ADO can be used to add meta-information to existing AD data, it provides the basis for efficient information retrieval and knowledge representation. It can automatically extract domain-specific knowledge related to AD, which can be used to gain further insights into disease mechanisms.

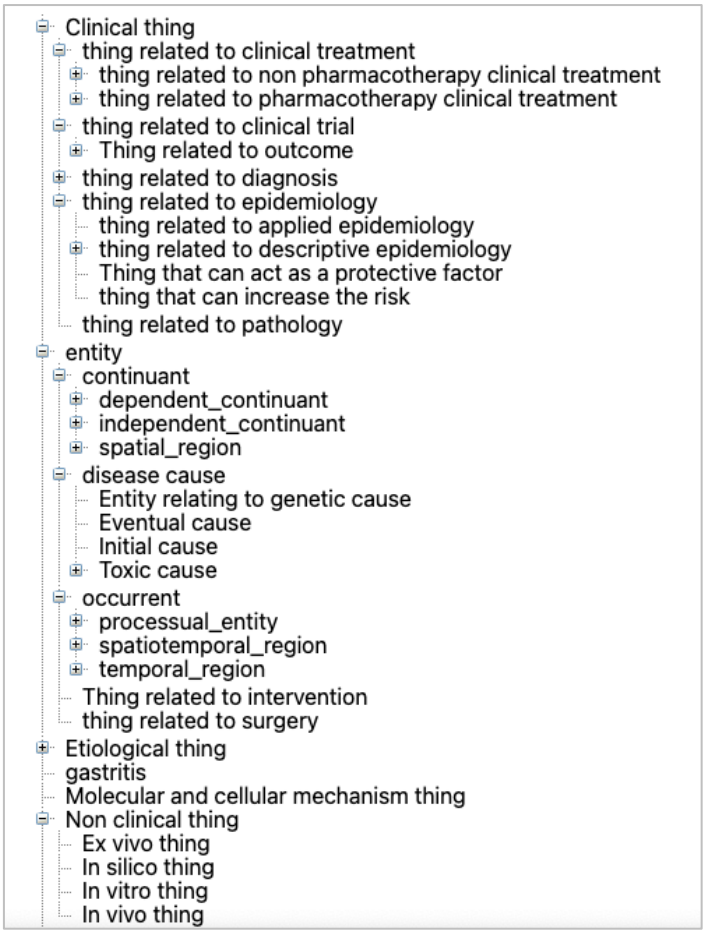


Figure 5: Instance of existing Alzheimer's disease related term in ADO

- NIFT - Neuro-Imaging Feature Terminology

Neuro-Imaging Feature Terminology (NIFT) was developed to organize the knowledge domain of measured brain features in association with neurodegenerative diseases by imaging technologies. The purpose is to identify quantitative imaging biomarkers that can be extracted from multi-modal brain imaging data. This terminology attempts to cover measured features and parameters in brain scans relevant to disease progression. The NIFT terminology concepts were gathered by collecting and reading relevant publications, e-books, websites, and medical blogs related to imaging in neurodegeneration. Following the initial literature search, we also adapted some concepts from already published, highly relevant ontologies such as QIBO and Radlex. Ontologies such as QIBO and Radlex had well-structured concepts such as Imaging Techniques and Imaging Agents, which were contextually relevant for the development of NIFT. Essential entities used in the ADNI were also included in our terminology system. Consequently, we enriched the NIFT with measured biomarkers obtained from the Biomedical Imaging Group Rotterdam (BIGR) pipeline, UMC Rotterdam, and neuGRID platform. NIFT comprises 7 major classes namely Algorithms, Brain Region, Clinical indices, Clinical trial information, Imaging technique, Measured Feature, and finally Radiopharmaceutical compound. There are in total of 1,221 terms in NIFT. NIFT can be used to correlate clinical diagnosis with imaging features for staging AD, retrieval and mining figure captions and full text from publications, and annotation of image scans.

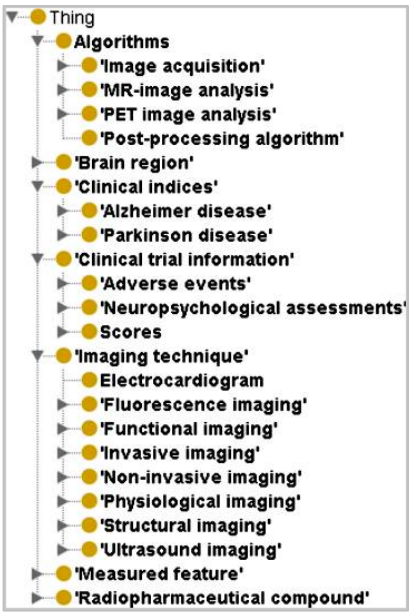


Figure 6: Instance of existing neuroimaging terms in NIFT

- Uberon

Uberon is an integrated multi-species anatomy ontology that integrates different anatomical ontologies and the datasets annotated using these ontologies. The ontology has been integral in a number of computational analyses that interpret human data using model organism phenotypes for translational research. Uberon contaions over 6,500 classes representing a variety of anatomical entities, organized according to traditional anatomical classification criteria. The ontology represents structures in a species-neutral way and includes extensive

associations to existing species-centric anatomical ontologies, allowing integration of model organism and human data. Uberon provides a necessary bridge between anatomical structures in different taxa for cross-species inference. It uses novel methods for representing taxonomic variation and has proved to be essential for translational phenotype analyses. Further, Uberon serves as a nexus for connecting multiple other biological ontologies such as the Cell Ontology (CL) and the Gene Ontology (GO), and in the modular construction of other multi-species anatomy ontologies. Uberon meets the current need for an integrative cross-species anatomy ontology amongst the Open Biomedical Ontologies (OBO) Foundry suite.

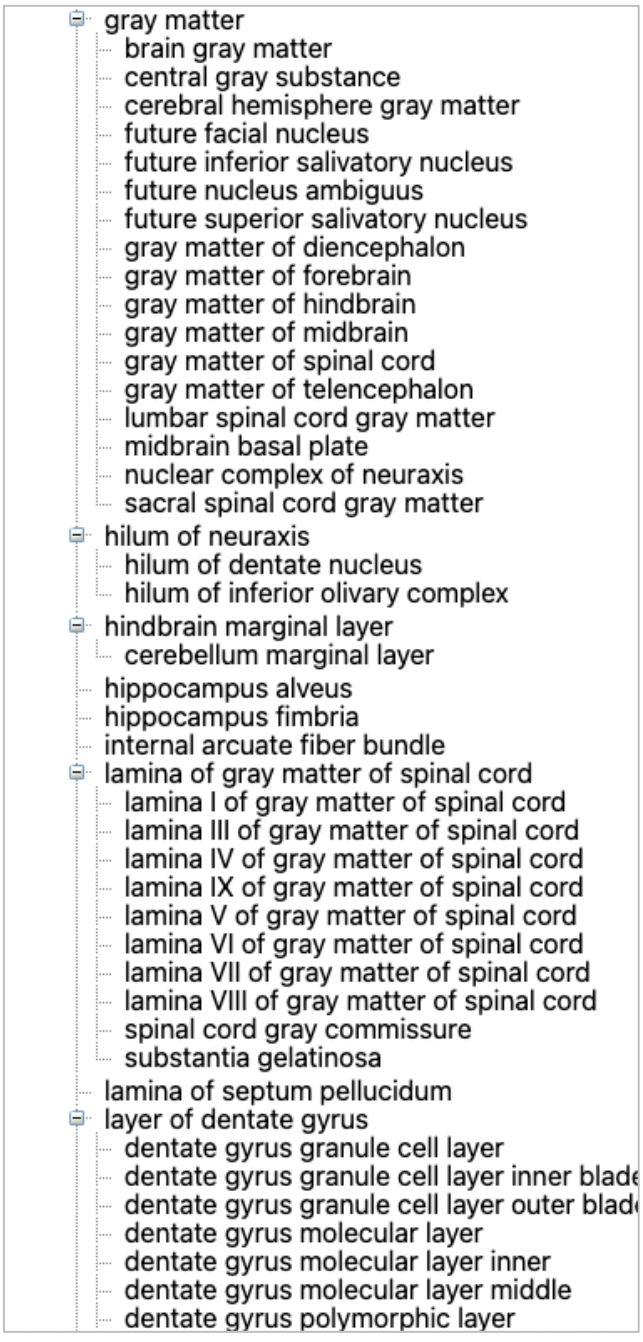


Figure 7: Instance of existing brain's anatomy terms in Uberon

- HGNC - HUGO Gene Nomenclature Committee

The HUGO Gene Nomenclature Committee (HGNC) based at EMBL's European Bioinformatics Institute (EMBL-EBI) assigns unique symbols and names to human genes. There are over 42,000 approved gene symbols in our current database of which over 19 000 are for protein-coding genes The HGNC is responsible for approving unique symbols and names for human loci, including protein-coding genes, ncRNA genes, and pseudogenes, to allow unambiguous scientific communication. Each symbol is unique, and each gene is only given one approved gene symbol. It is necessary to provide a unique symbol for each gene so that researchers can talk about them, and this also facilitates electronic data retrieval from publications and databases. In preference, each symbol maintains parallel construction in different members of a gene family and can also be used in other species. HGNC symbols are displayed in all major databases containing human gene and protein data including Ensembl, NCBI Gene, UniProt, GeneCards, and the UCSC genome browser, as well as resources focused on human disease and phenotypes such as Decipher, OMIM, Locus Reference Genomics, ClinVar, and GeneTests.

3.1.5. Text-mining Software, Tools and Resources

ProMiner

Scientific publications found in abstract data bases, full text journals or patents are the main and most up-to-date information source, but the amount of text is overwhelming for most life science areas. Recognition of life science terminology is a key prerequisite for performing automatic information retrieval and information extraction. Huge and complex terminologies with high numbers of synonymous expressions, ambiguous terminology and numerous generations of new names and classes present named entity recognition with a real challenge. ProMiner is a tool for specific terminology recognition and addresses several fundamental issues in named entity recognition in the field of life sciences. ProMiner can handle voluminous dictionaries, complex thesauri and large controlled vocabularies derived from ontologies. It allows the mapping of synonyms to reference names and data sources. ProMiner has been designed for high-speed tagging and execution of parallel workflows for multiple dictionaries. In the context of Task 5.2, we have generated text mining dictionaries for all the ontologies indexed in the ontology service OLS. We have tagged all available open access full text articles (PMC corpus), tagged them with a dedicated Alzheimer model in order to filter down to relevant documents – the TVB collection. We have used the Alzheimer ontology to filter the PMC corpus for relevant neurodegenrative articles. In total 1,543,765 documents have been selected. In the next step the TVB collection has been processed with all dictionaries. The specific annotations made by ProMiner then have been uploaded into a dedictated Neuro instance of SCAIView: https://neuro.scaiview.com/ .

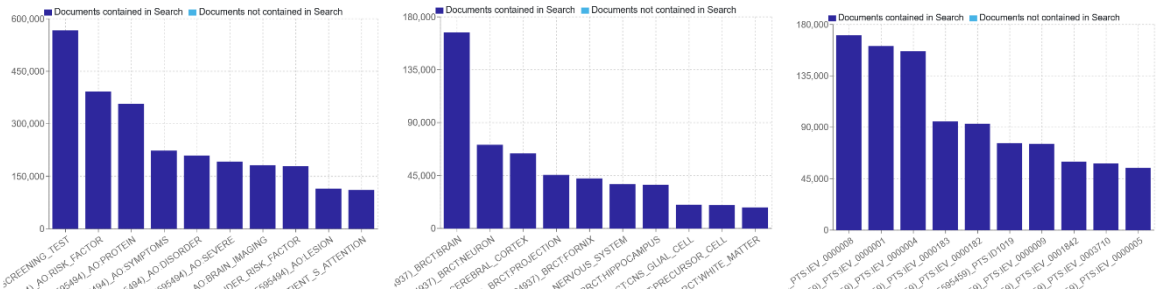


Figure 8: Distribution of most frequent terms from ADO, BRCO and PTS ontologies in the total corpus.

SCAIView Neuro

SCAIView is an advanced text mining system and semantic search engine that addresses questions of interest to general biomedical and life science researchers. Most of the current knowledge exists as unstructured text (publications, text fields in databases). SCAIView provides users with full-text and biomedical concept searches, which are supported by large biomedical terminologies and outstanding text mining technologies. Using the SCAI text mining pipeline which uses machine learning and rule-based named entity recognition, SCAIView extracts information about genes, drugs, SNPs and other life science entities based on the ontological concepts that occur in the documents. In the context of Task 5.2, we have used PTS, BRCT, TVB Cloud ontology, to find the literature corpus as well as to extract the pathways related with brain regions in the context of Alzheimer's disease.

Some example queries to drill down on a certain pathway – brain region - disease association. Searching for documents containing:

- Any pathway and any brain region: Found 62,683 documents
- Hippocampus: Found 24,110 documents
- Notch pathway: Found 4,336 documents
- Notch pathway and brain regions: Found 3,288 documents
- Notch pathway and Hippocampus: Found 344 documents
- Notch pathway and Hippocampus and Apoptosis: Found 157 documents

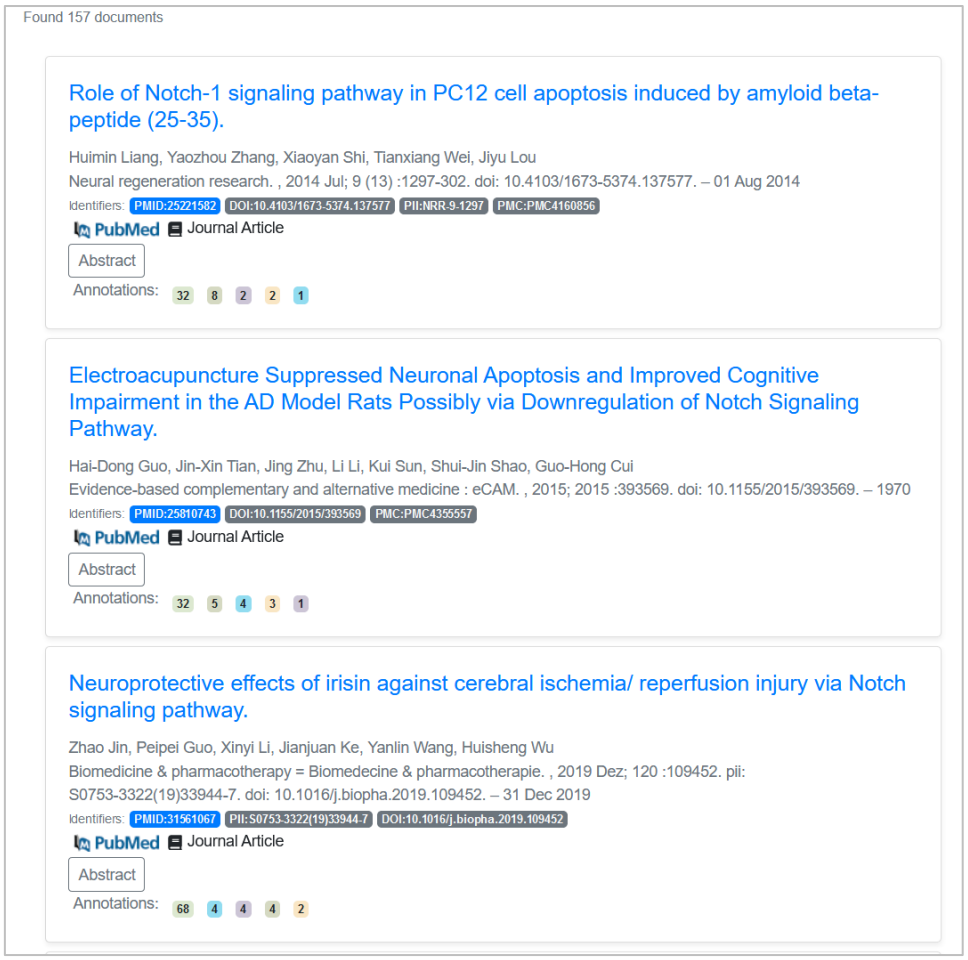


Figure 9: Instance of retrieved articles in the context of TVB

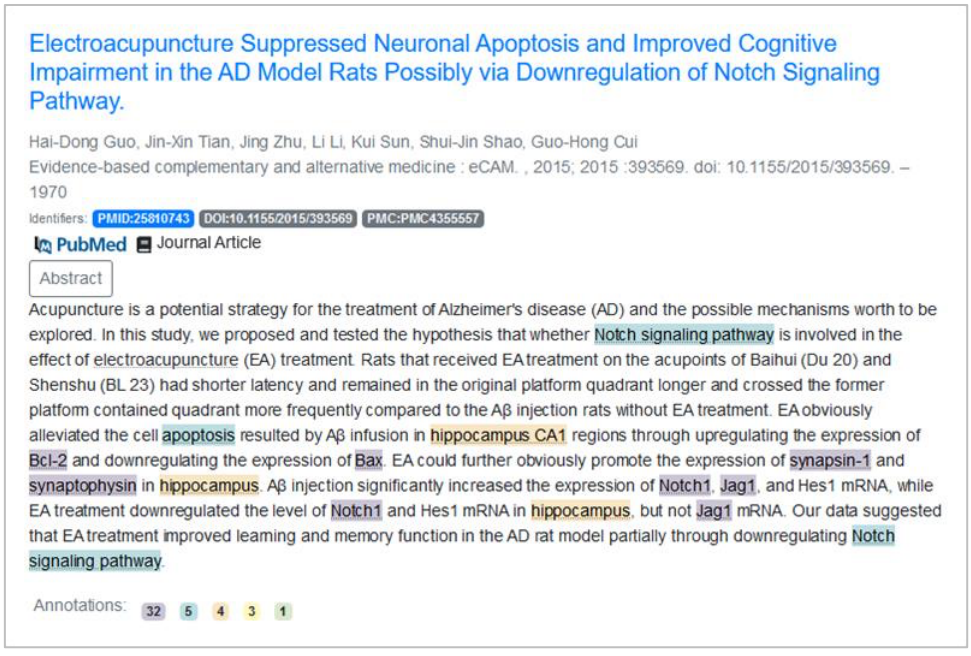


Figure 10: Instance of brain region – pathway association in sentence level
https://neuro.scaiview.com/document/PMID:25810743

BELIEF

BELIEF as a workflow that focuses on knowledge extraction from literature in the biomedical domain while integrating various available state-of-the-art solutions. The workflow uses a text mining pipeline to extract relationships from literature. In the current state, we focus on sentence-based extraction of protein–protein interactions and relations between genes/proteins, chemicals, diseases and biological processes. As a network modeling language, we use Biological Expression Language (BEL) to represent the extracted knowledge. The information extraction system translates the extracted relations directly into BEL statements (BEL encoded triples). The web-based curation interface visualizes the causal and correlative BEL statements and facilitates the biocuration.

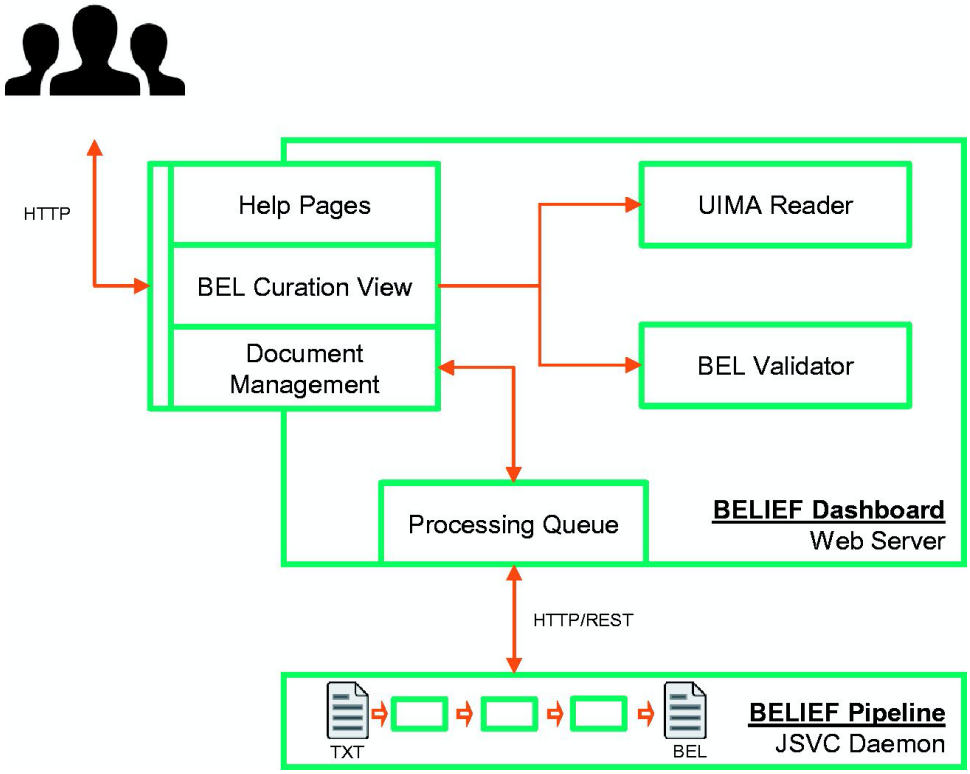


Figure 11: Architecture of semiautomatic information extraction workflow BELIEF. The workflow consists of a text mining pipeline (BELIEF Pipeline) and a web-based biocuration tool (BELIEF Dashboard). (Note: UIMA: Unstructured Information Management Architecture. UIMA Reader: A reader component to parse and extract information from UIMA XCAS documents. JSVC Daemon: A Java library that allows applications to run as daemons.)

- BELIEF Pipeline

The BELIEF Pipeline consists of several components from the fields of natural language processing (NLP), NER and relationship extraction (RE) implemented in the Unstructured Information Management Architecture (UIMA) framework 3. The full workflow is constructed as a non-interactive server application (based on commons daemon library 4) to support text processing on demand. The workflow also includes a REST 5 -based BELIEF service that periodically contacts the BELIEF Dashboard via a defined API to pull new unprocessed documents into the RE workflow. The input data for the text mining workflow is the document text itself and the respective output data are the identified named entities and relationships.

Hence, the runtime of the workflow strongly depends on the number of identified named entities and the document length.
Overall, the pipeline consists of several sequential minor and major steps (Figure 12). Owing to the flexible underlying UIMA architecture and the Common Analysis Structure (CAS) exchange format, the integration of new tools is simplified. A high level of modularity is required for the integration of new RE applications since several pre- and post-processing steps are necessary. Some of the pre-processing steps, such as the tagging of sentences or NER, can be shared by different tools. In this way, the duplication of work is prevented and runtime is optimized. The following sections describe the pipeline (Figure 12) in detail.

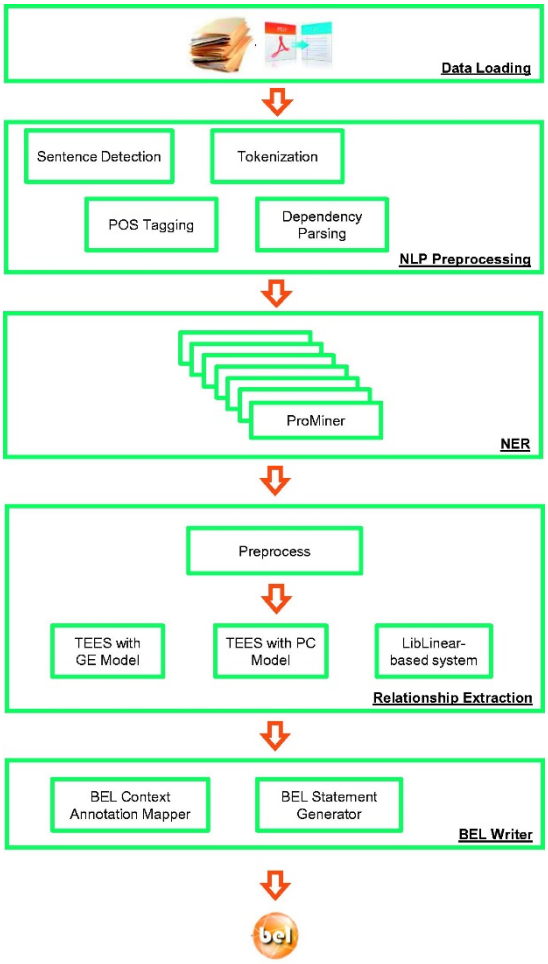


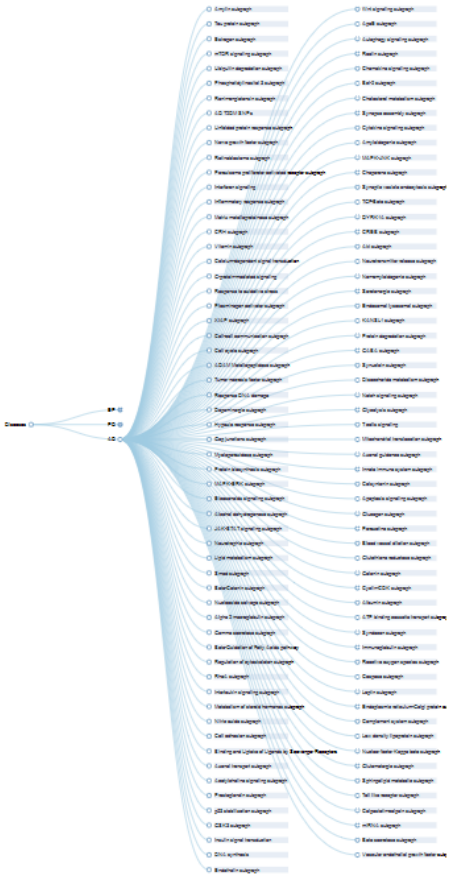
Figure 12: Architecture of the BELIEF text mining pipeline. (Note: POS Tagging: Part of Speech Tagging. NLP: Natural Language Processing. TEES: Turku Event Extraction System, a state-of-the-art relation extraction system. GE: Genia Event Extraction for NFkB knowledgebase. PC: Pathway Curation.)
BEL

NeuroMMSig

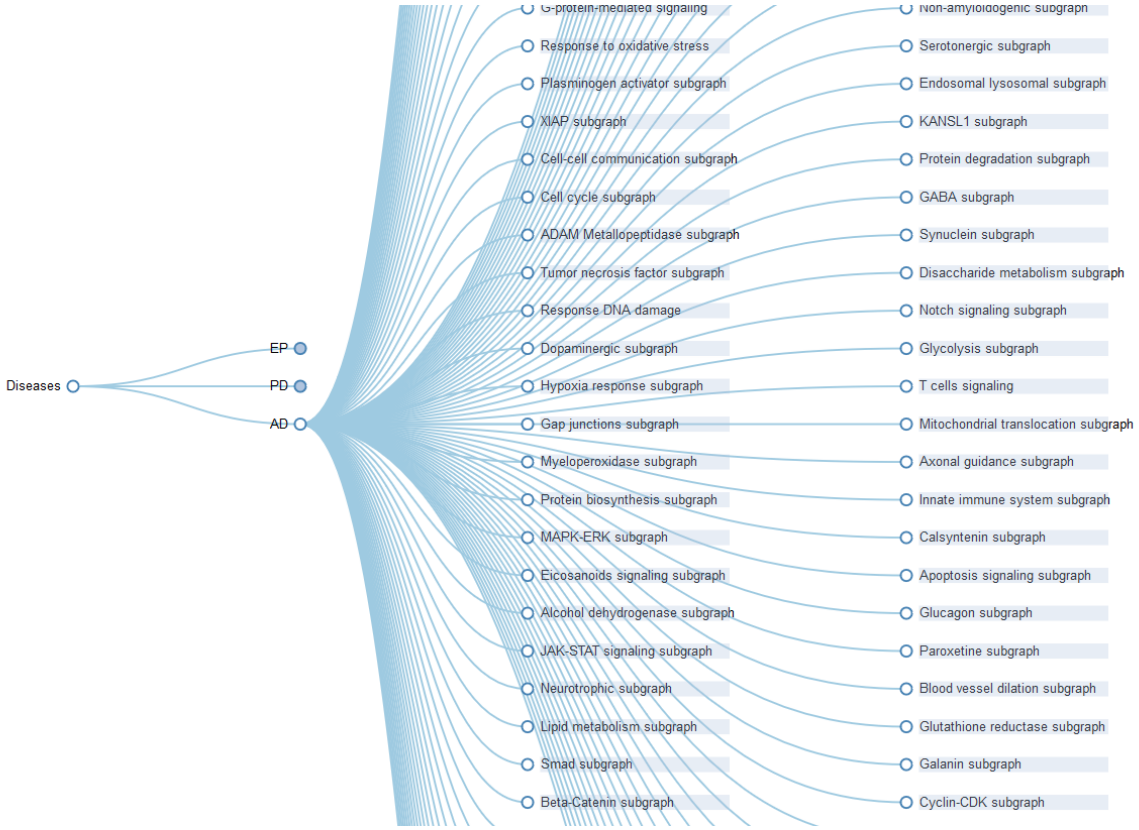
NeuroMMSig, a web server for mechanism enrichment that allows submission of multiscale data from molecular to clinical level to return mechanisms that fit best the data. We have focused on neurodegenerative diseases, as we try to establish a 'mechanism-based taxonomy of Alzheimer's Disease (AD), Parkinson's Disease (PD)' and Epilepsy. Disease knowledge assembly models were built using Biological Expression Language (BEL) which integrate

literature-derived 'cause and effect' relationships in the form of triples (Kodamullil et al., 2015). We have captured a representative subsample of the scientific knowledge on existing canonical pathways in AD and PD (Iyappan et al., 2016) which have been grouped into subgraphs. NeuroMMSig's subgraphs have been enriched with multimodal data (e.g. imaging features, variant information and drugs). In the context of AD, we have 124 sub graphs that describes the cause-effect mechanisms in the context of AD.

A)



B)



C)

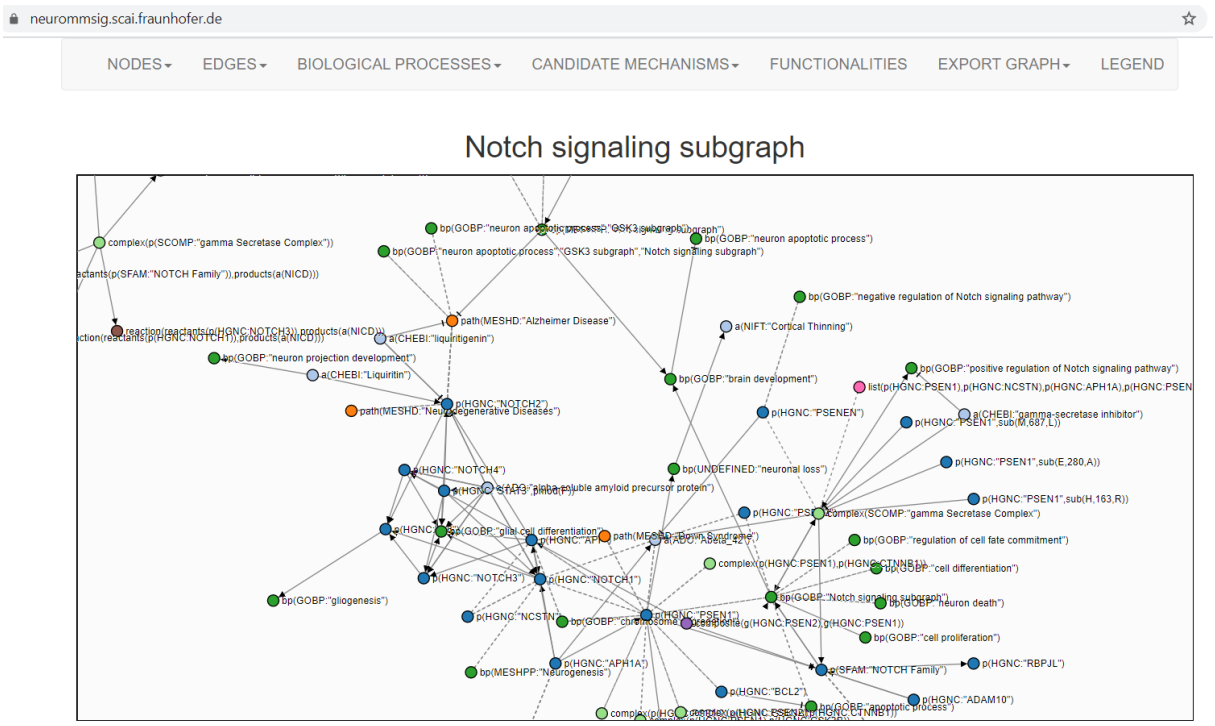


Figure 13: Instance of NeuroMMsig content representing the mechanism-based taxonomy of AD, PD, EP (A and B) and an interested subgraph(C).

3.1.6. Mapping of Brain Regions to Pathways to Cause-Effect mechanisms

In the context of Task 5.2, we have generated ontologies that are necessary for retrieving, indexing and extracting information related with Brain Regions, Pathways in the context of AD. The ontologies BRCT, Neuro FMA are used to detect the brain regions. PTS is used to detect the pathway mention in the text. The ADO and Neuronames are used to search for the specific articles that mention about Alzheimer's disease. NIFT, HGNC and TVB Cloud ontology was also integrated into TVB cloud OLS, so that, we could also annotate the associated imaging features, genes and TVB Cloud specific terms in the text, if needed. Specific dictionaries are made from these TVB specific ontologies and are used by ProMiner for tagging the Alzheimer specific literature. ProMiner have tagged all available open access full text articles (PMC corpus), tagged them with a dedicated Alzheimer model in order to filter down to relevant documents – the TVB collection. As the next step. the TVB collection has been processed with all dictionaries. The specific annotations made by ProMiner then have been uploaded into SCAIView.
Using ProMiner and SCAIView, we are able to detect the Brain Region-Pathway mentions/co-occurrences in the TVB Collection. This list is being used to map the Pathway mention to subgraphs/pathway names of NeuroMMSIG. In this way, we could identify the cause-effect mechanistic subgraph related with each pathway and brain regions. The cause-effect mechanisms from the literature are semi-automatically extracted by the BELIEF workflow and are then manually curated before loading that into the NeuroMMSiG subgraphs.

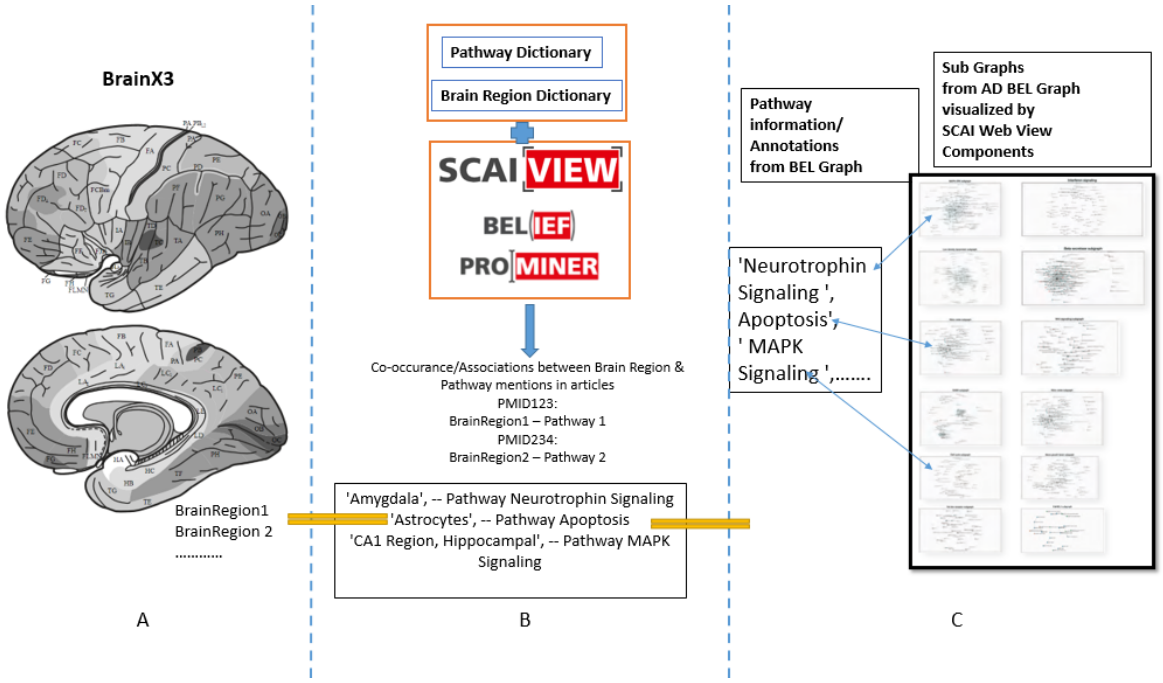


Figure 14: Sketch of Fraunhofer SCAI's tool set to map brain regions to pathways to cause-effect mechanisms

4. Results

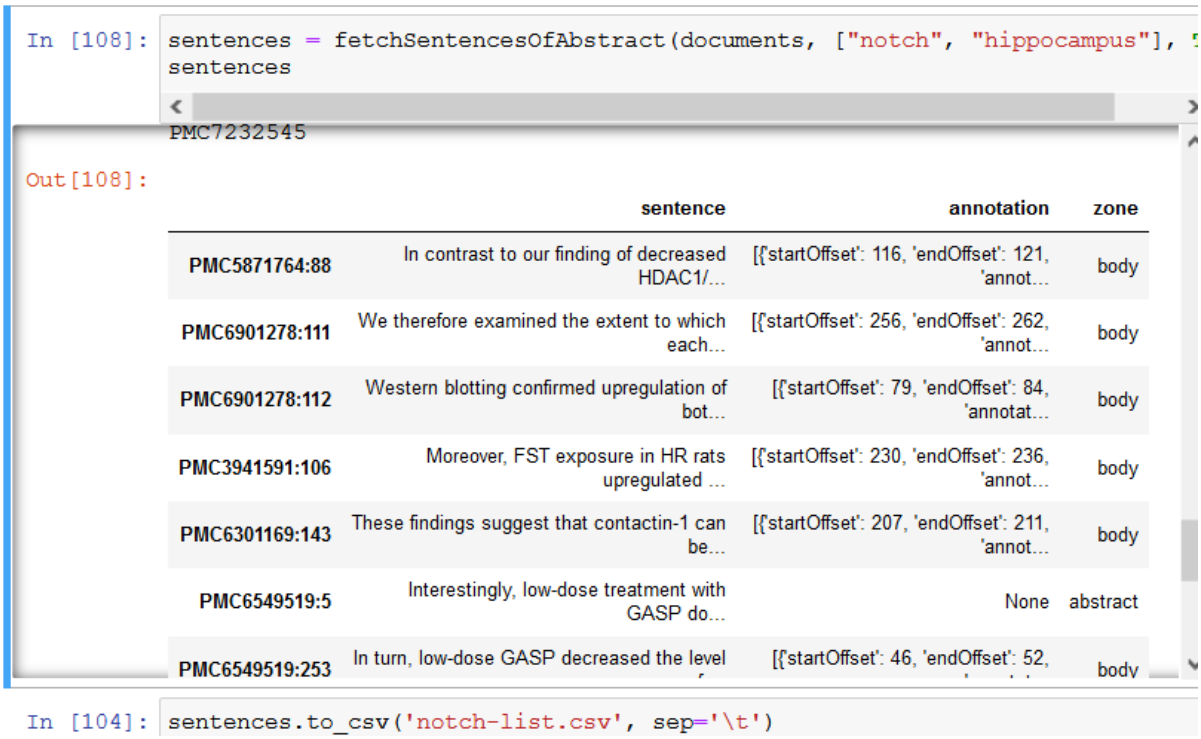
See Annex 1 for our main result: A software that maps features from knowledge bases to the 3D brain coordinate space.

In addition to this main deliverable, CHARITE produced a public interactive web atlas that displays feature maps from knowledge bases in a 3D brain model:

https://www.brainsimulation.org/atlasweb_multiscale/

4.1. Brain Region Pathway Mapping

As discussed above, ProMiner and SCAIView have used to identify the co-occurring pathway and brain region mentions at the sentence level in the context of AD. We have created a jupyter notebook for querying the SCAIView API https://api.neuro.scaiview.com/swagger-ui.html#/ . This notebook allows to query a collection of documents and then analyze the content of the documents in detail. In the list below information have been on the document, the number of the sentence, the sentence itself, the zone the sentence appears (title, abstract, body) and all ProMiner annotations. The mappings are included as the .csv file



```
In [108]: sentences = fetchSentencesOfAbstract(documents, ["notch", "hippocampus"], '
          sentences
```

PMC7232545

Out[108]:

	sentence	annotation	zone
PMC5871764:88	In contrast to our finding of decreased HDAC1/...	[{'startOffset': 116, 'endOffset': 121, 'annot...	body
PMC6901278:111	We therefore examined the extent to which each...	[{'startOffset': 256, 'endOffset': 262, 'annot...	body
PMC6901278:112	Western blotting confirmed upregulation of bot...	[{'startOffset': 79, 'endOffset': 84, 'annotat...	body
PMC3941591:106	Moreover, FST exposure in HR rats upregulated ...	[{'startOffset': 230, 'endOffset': 236, 'annot...	body
PMC6301169:143	These findings suggest that contactin-1 can be...	[{'startOffset': 207, 'endOffset': 211, 'annot...	body
PMC6549519:5	Interestingly, low-dose treatment with GASP do...	None	abstract
PMC6549519:253	In turn, low-dose GASP decreased the level	[{'startOffset': 46, 'endOffset': 52,	body

```
In [104]: sentences.to_csv('notch-list.csv', sep='\t')
```

Figure 15: Example of Notch Pathway – Hippocampus Brain Region detected in articles (PMCID) along with the supporting sentence.

4.2. Brain Region – Pathway – NeuroMMSiG subgraphs

The mappings of bran region and pathway are then linked with NeuroMMSig subgraphs. In figure 16, you find the mappings between Pathways (from PTS) and the mappings to NeuroMMSiG Ad subgraphs. In Figure 17, we have included the publication ID (PMC ID), sentence from which co-occurrance of Brain Region (BRCT) and Pathway (PTS) are detected and then the list of corresponding NeuroMMSiG subgraphs.

1	code	path	neurommsig_id	neurommsig_name
2	IEV_0003872	T cell proliferation	NEUROMMSIG100	T cells signaling
3	IEV_0000462	Toll-like receptor signaling pat	NEUROMMSIG104	Toll like receptor subgraph
4	IEV_0000027	Wnt signaling pathway	NEUROMMSIG109	Wnt signaling subgraph
5	IEV_0001552	Energy Metabolism	NEUROMMSIG111	Energy metabolic pathway
6	ID1005	Ubiquitin mediated proteolysis	NEUROMMSIG114	Ubiquitin degradation subgraph
7	IEV_0000012	Ubiquitination	NEUROMMSIG114	Ubiquitin degradation subgraph
8	IEV_0001699	AKT signaling	NEUROMMSIG13	Akt subgraph
9	ID2126	PI3K/AKT signaling	NEUROMMSIG13	Akt subgraph
10	ID1019	Apoptosis	NEUROMMSIG19	Apoptosis signaling subgraph
11	IEV_0000070	Apoptosis (Inhibition of cell su	NEUROMMSIG19	Apoptosis signaling subgraph
12	IEV_0000097	Cell death	NEUROMMSIG19	Apoptosis signaling subgraph
13	ID0278	cell death pathway	NEUROMMSIG19	Apoptosis signaling subgraph
14	ID0281	autophagy pathway	NEUROMMSIG21	Autophagy signaling subgraph
15	IEV_0000105	Cell adhesion	NEUROMMSIG32	Cell adhesion subgraph
16	IEV_0000077	Cell cycle	NEUROMMSIG34	Cell cycle subgraph
17	ID2568	cytokine pathway	NEUROMMSIG40	Cytokine signaling subgraph
18	IEV_0000098	Immune response	NEUROMMSIG61	Innate immune system subgrap
19	ID0146	insulin signaling pathway	NEUROMMSIG62	Insulin signal transduction
20	IEV_0001553	Lipid Metabolism	NEUROMMSIG67	Lipid metabolism subgraph

Figure 16: Example of Pathway as well as the corresponding subgraph from NeuroMMSiG

PMC5362350:177	We further showed that autophagy inhibitors exacerbated gp120/r	['Cell death', 'autophagy pathway', 'a	['Astrocyte']	['Apoptosis signaling subgraph', 'Autophagy signaling subgraph', 'Autophag
PMC2727358:101	In many other studies, cell death associated with excessive autoph	['Cell death', 'autophagy pathway', 'A	['Neuron', 'Brain']	['Apoptosis signaling subgraph', 'Autophagy signaling subgraph', 'NA', 'Apo
PMC4592743:296	For example, during HSV-1 infection, DRG neurons resist cytotoxic	['Cell death', 'autophagy pathway']	['Neuron']	['Apoptosis signaling subgraph', 'Autophagy signaling subgraph']
PMC6913523:253	Studies in a model of brain inflammation where cell death does no	['Cell death', 'Immune response']	['Brain']	['Apoptosis signaling subgraph', 'Innate immune system subgraph']
PMC6391192:165	Furthermore downregulation of a-syn in dopaminergic neurons lea	['Cell death', 'Immune response']	['Neuron']	['Apoptosis signaling subgraph', 'Innate immune system subgraph']
PMC2672307:174	Our work in adult focal stroke models demonstrates that the amou	['Cell death', 'Pathway', 'Cell death', 'T	['Brain']	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph', 'NA']
PMC4485545:91	For example, IPA identified â€œincreased cell death of brainâ€, â€	['Cell death', 'Gene expression', 'Cell d	['Brain', 'Thalamus', 'Dentate_gy	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC3460212:132	In contrast, intracellular Fas signaling involving traditional FADD re	['Apoptosis', 'Activation', 'Cell death']	['Brain', 'Brain']	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC6120585:21	Both TSP-1 and its highly homologous family member, TSP-2, can p	['Apoptosis', 'Death Receptor signalin	['Brain']	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC6612317:93	In contrast, our hiNC-derived Schwann cells are likely a better mod	['Cell death', 'Cell-cell signaling', 'cell c	['Schwann_Cell', 'Schwann_Cell'	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC4765324:164	As such, AEA induces apoptosis of human neuroblastoma cells thro	['Apoptosis', 'Pathway', 'Cell death']	['Neuron', 'CNS_glial_cell']	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC4948590:95	Methylene blue (MB) and Melatonin, show potentials for hypoxic-i	['Apoptosis', 'Translocation', 'Apoptos	['Brain', 'Pineal_body']	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC3364609:95	In microglia, selective activation of mGlu2 and mGlu3 receptors by	['Apoptosis', 'Activation', 'Cell death']	['Microglia']	['Apoptosis signaling subgraph', 'NA', 'Apoptosis signaling subgraph']
PMC5316312:195	In contrast to the early school of thought that non-cycling cells suc	['Cell death', 'Activation', 'Cell cycle', '	['Neuron', 'Neuron', 'Brain']	['Apoptosis signaling subgraph', 'NA', 'Cell cycle subgraph', 'Cell cycle subg
PMC3728697:89	This association of neuroserpin with social behavior is intriguing in	['Cell death', 'Cell proliferation', 'Cell c	['Brain']	['Apoptosis signaling subgraph', 'NA', 'Cell cycle subgraph']
PMC5959054:68	In 2014, Xie et al. demonstrated that the protective effects of ischa	['Apoptosis', 'Activation', 'neuronal de	['Brain']	['Apoptosis signaling subgraph', 'NA', 'NA', 'mTOR signaling subgraph', 'NA
PMC5559303:70	Pathological mechanisms leading to synaptic loss and inflammatio	['Apoptosis', 'Activation', 'Pathway', 'in	['Brain', 'Brain', 'Neuron', 'Hippo	['Apoptosis signaling subgraph', 'NA', 'NA', 'NA', 'Apoptosis signaling subgr
PMC5681369:44	The extrinsic/death receptor pathway can crosstalk to the intrinsic	['Apoptosis', 'mitochondrial apoptotic	['Microglia', 'Astrocyte', 'CNS_gl	['Apoptosis signaling subgraph', 'NA', 'NA', 'NA', 'NA']

Figure 17: Example of Pathway and Brain Region co-occurrences with corresponding PMC ID as well as the sentence. These are then mapped to corresponding subgraph from NeuroMMSiG

References

Amunts, K., H. Mohlberg, S. Bludau and K. Zilles (2020). "Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture." Science **369**(6506): 988-992.

Braak, H. and E. Braak (1991). "Neuropathological staging of Alzheimer-related changes." Acta neuropathologica **82**(4): 239-259.

Braak, H. and E. Braak (1997). "Frequency of stages of Alzheimer-related lesions in different age categories." Neurobiol Aging **18**(4): 351-357.

Braak, H., I. Alafuzoff, T. Arzberger, H. Kretzschmar and K. Del Tredici (2006). "Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry." Acta Neuropathol **112**(4): 389-404.

Domingo-Fernández, D., A. T. Kodamullil, A. Iyappan, M. Naz, M. A. Emon, T. Raschka, R. Karki, S. Springstubbe, C. Ebeling and M. Hofmann-Apitius (2017). "Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment." Bioinformatics **33**(22): 3679-3681.

A. M. Dale, R. P. Maguire and B. T. Hyman (2006). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." Neuroimage **31**(3): 968-980.

Destrieux, C., B. Fischl, A. Dale and E. Halgren (2010). "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature." Neuroimage **53**(1): 1-15.

Frisoni, G. B. and M. W. Weiner (2010). "Alzheimer's Disease Neuroimaging Initiative special issue." Neurobiol Aging **31**(8): 1259-1262

Hanisch, D., K. Fundel, H.-T. Mevissen, R. Zimmer and J. Fluck (2005). "ProMiner: rule-based protein and gene entity recognition." BMC Bioinformatics **6**(1): S14.

Mungall, C. J., C. Torniai, G. V. Gkoutos, S. E. Lewis and M. A. Haendel (2012). "Uberon, an integrative multi-species anatomy ontology." Genome Biology **13**(1): R5.

Stefanovski, L., P. Triebkorn, A. Spiegler, M. A. Diaz-Cortes, A. Solodkin, V. Jirsa, A. R. McIntosh and P. Ritter (2019). "Linking Molecular Pathways and Large-Scale Computational Modeling to Assess Candidate Disease Mechanisms and Pharmacodynamics in Alzheimer's Disease." Frontiers in Computational Neuroscience.

Weiner, M. W., D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, Jr., W. Jagust, J. C. Morris, R. C. Petersen, J. Salazar, A. J. Saykin, L. M. Shaw, A. W. Toga and J. Q. Trojanowski (2017). "The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement." Alzheimer's & Dementia: The Journal of the Alzheimer's Association **13**(5): 561-571