DALSO: domain ALS ontology

Teresa Podsiadły-Marczykowska[1], Peter Andersen[2], Marta Gromicho[3], Julian Grosskreutz[4], Magdalena Kuźma-Kozakiewicz[5], Susanne Petri[6], Katarzyna Szacka[5], Hilmi Uysal[7], Mamede de Carvalho[3], Maria Piotrkiewicz[1]

[1]Nalecz Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland, [2]Umeå University, Sweden, [3]Faculty of Medicine, University of Lisbon, Portugal, [4]Dept of Neurology, University of Lübeck, Germany, [5]Medical University of Warsaw, Poland, [6]Hannover Medical School, Germany, [7]Akdeniz University Faculty of Medicine, Antalya, Turkey.

Corresponding author: Maria Piotrkiewicz
Nalecz Institute of Biocybernetics and Biomedical Engineering PAS
4 Trojdena Str., 02-109 Warsaw, Poland
E-mail: mpiotrkiewicz@ibib.waw.pl
tel. +48664293172, ORCID No 0000-0001-9340-1724

Abstract

Amyotrophic lateral sclerosis (ALS) is an incurable, rapidly progressive neurodegenerative disease. During the course of ALS, virtually all skeletal muscles are gradually affected, including the respiratory muscles, and the disease is usually fatal within 2−5 years of symptom onset. Unequivocal and conclusive tests for ALS do not exist, its disease etiology is still unknown, and therapeutic options are limited. This paper presents the ALS domain ontology (DALSO), model containing formalized, semantic descriptions of a wide range of modeled disease related notions such as patient demographics, clinical findings and history, disease clinical features and diagnostic classifications, risk factors, genetics and pathophysiological mechanisms of motor neuron degeneration. The DALSO′s aim and information scope, design assumption, structure and implementation details are also described. DALSO covers the broad range of significant biomedical concepts ranging from clinical to molecular feature of the modeled disease, it represents a comprehensive, structured knowledge source for the ALS disease domain. To the best of authors′ knowledge, the DALSO is the first attempt to develop a formal, computational model representing knowledge of this fatal motor neuron disease. It provides the means for integrating and annotating clinical and research data, not only at the generic domain knowledge level, but also at the level of individual patient case studies. The DALSO is expressed in OWL2 language, contains 910 classes, is consistent and free of logical errors. Its syntactic correctness was validated by the Fact++ reasoner.

Key words: Amyotrophic Lateral Sclerosis, domain disease ontology, motor neuron

1. Introduction

1.1 ALS disease features

The etiology of amyotrophic lateral sclerosis (ALS) is still unknown, although this disease was first reported in 1869 by the French neurologist Jean-Martin Charcot (Charcot and Joffroy 1869). ALS is a fatal, rapidly progressive neurodegenerative disorder characterized by degeneration of both the upper motor neurons (MNs) of the motor cortex and lower MNs of the brainstem and spinal cord. The initial symptoms consist of progressive muscle weakness, wasting and spasticity. During the course of ALS, all voluntary muscles are eventually affected, including the respiratory muscles, and the disease is usually fatal within 2–5 years of symptom onset. There is a relatively small subgroup of patients that demonstrate slower disease progression, surviving even beyond 10 years (Witzel et al. 2022). Recently, the possibility was suggested that disease course becomes slower over time (Czaplinski et al. 2006). Respiratory insufficiency always occurs in patients with ALS, generally a the end stage, and is a major cause of mortality.

The mean age of onset is 60 years, but symptoms can start at any adult age. The male to female ratio is approximately 1.4:1, with hormones possibly playing a protective role in premenopausal women (Manjaly et al. 2010). The incidence of ALS is 2–4/100,000 per year, with heterogeneous distribution worldwide (Marin et al. 2017), and a prevalence of about 2–5/100,000 (Chiò et al., 2013), possibly increasing over the years, which can partly be explained by lengthening lifespan (Korner et al. 2013). The life-time risk is in the order of 1/400 for males and 1/550 for females (Ryan et al. 2019).

Approximately 5% of ALS patients have a first-degree relative with ALS. When more distant relatives are taken into account, about 10% of patients are found to have a familial case of ALS (fALS), and the remaining 90% of ALS cases are apparently sporadic (sALS) (Mejzini et al. 2019). It has been shown that genetic factors are involved in the pathogenesis of around 70% of fALS cases. Moreover, mutations in the same genes were found in evidently sporadic cases, although at lower frequencies, which resulted in oligogenic hypothesis suggesting that sALS cases also may be caused by yet undiscovered genetic mutations (Kim et al. 2020; van Blitterswijk et al. 2012; Alsultan et al. 2016). This hypothesis was tested in recently published study aimed towards characterization of genetic variability in genes related to sALS (Van Daele et al. 2023). The authors studied genetic variation in 90 ALS-associated genes in a large cohort of 6529 sALS patients. The study confirmed genetic involvement only in 28.17% of patients, but concluded that further research to test oligogenic hypothesis is needed.

Therapeutic options for ALS are limited and no preventive treatment exists. The most important part of current patient care consists of nutritional and respiratory management and alleviation of symptoms. The only drugs currently approved for ALS treatment are riluzole and edaravone, both providing relatively modest effects. Moreover, positive effects of edaravone are limited to a specific subpopulation of patients.

Nevertheless, the search for effective ALS therapies continues around the world. First of all, we should mention the Institute for the Development of ALS Therapy, whose research is financed by the ALS community. Several other clinical trials of new treatments are also ongoing (e.g. (Miller et al. 2022; Wiesenfarth et al. 2024), see also https://www.eventalways.com/2nd-als-drug-development-summit and https://www.biopharmadive.com/news/als-drug-research-progress-genetic-clinical-trials/611274/ Hopefully these activities will soon result in the emergence of effective therapies for ALS, and also bring us closer to understanding the etiology of this disease.

A fact that can aggravate prognosis of ALS is diagnostic delay (Campos et al. 2021, 2023) averaging 5–27 months, depending on the disease onset region (Sennfält et al. 2023), which urge to be changed (Andersen et al. 2012; Segura et al. 2023). This can be primarily attributed to the clinical heterogeneity of ALS, recognized as a consistent disease feature (e.g. Goutman et al. 2022) due to the existence of many conditions mimicking ALS (Traynor et al. 2000) and to the very commonly subtle symptoms of ALS onset.

An unequivocal and conclusive test for ALS does not exist (Turner and Swash 2015), despite many years of research and undeniable progress in understanding the disease mechanisms. Currently, the ALS diagnostic process is primarily based on clinical assessment by an experienced neurologist, relying heavily on individual knowledge and experience, although diagnostic support through electrophysiological investigations, electromyography (Fukushima et al. 2022; Goutman et al. 2022) and neuroimaging (Cengiz and Kuruoğlu 2020; Verstraete and Foerster 2015) is also recommended.

Despite the huge body of research into genetics, epidemiology, diet, lifestyle, and the many mechanisms of cellular pathogenesis observable in ALS, a single unifying theory for the disease etiology is still lacking.

1.2 Motivation for the ALS domain ontology construction

Domain ontological models have two interrelated aspects. The first is a specialized domain terminology, the second represents semantics inherent to particular terms. Both model aspects are crucial to the realization of domain model applications such as unequivocal data capture, database schema construction, data integration (Calvanese et al. 2017b; Zhang et al. 2017) or, relevant to the genetic data analysis, extensions of enrichment analysis (Shah et al. 2012). Other important features of the ontological domain model are explicit taxonomy and modular structure, semantic explicitness, explicit linkage of concepts and relations, and the final expression of concept definitions using machine readable formal language. These features indicate that an ontological model allowing for the semantic encoding of domain terminology is the method of choice for the integration and annotation of heterogeneous data sources and it can also be helpful in the formulation of unambiguous diagnostic criteria (Abdollahi et al. 2021; Steiner et al. 2017).

These features of ontological models are reasons for which they have been proven as practical solutions in biomedical areas such as: (i) knowledge formalization and sharing (Sánchez and Moreno 2008; Staab and Studer 2010), (ii) logical, unequivocal concept definition in biomedicine (Bug et al. 2008; Silachan and Tantatsanawong 2011; Tao et al. 2013) (iii) integration, comparison and reuse of data and information (Ruttenberg et al. 2009; Ivanović and Budimac 2014; Nielsen 2014), and eventually (iv) new methods for genetic data analyses (Shah et al. 2012; Bean et al. 2020; Segura et al. 2023).

Besides the lack of a definitive test for diagnosing ALS, sound clinical assessment of ALS patients is additionally complicated by the inherent deficiency of the precision of terms describing the symptoms, diagnostics and phenotypic criteria. (Ruttenberg et al. 2009) state that there is no clear distinction between diagnostic and phenotypic ALS classifications because the vocabulary of terms used for clinical patient features, descriptions and classification of ALS phenotypes has often been inconsistent. The first three of the above-mentioned features of ontological models indicate that they have the potential to provide a solution to ALS diagnostic problems.

The genetic causes of ALS remain unclear today and a better understanding of patients' genetic data could result in deeper insights into the mechanisms of the disease. This issue can also be addressed using ontological models.

Gene ontology (GO) (Tao et al. 2013), a widely accepted reference ontology in biomedicine, is today a valuable tool for describing gene functions, although it has known limitations, such as the lack of disease specific context. Therefore, GO cannot be used to perform specific disease-related enrichment analysis considering e.g. the problem of disease subtypes, syndromes or phenotypes, under- or over-represented in the given gene set in question (Osborne et al. 2009).

There are reference ontologies containing human disease-related concepts/terminology, but these models are characterized by high-level, broad coverage knowledge, which limits their use for specific disease problems. Shah et al. (2012) proposed the novel method of enrichment analyses consisting of referencing specific disease domain ontologies with GO.

In addition, further research on the cell-specific biochemistry and physiology of MNs and the cellular pathways disturbed during MN degeneration are likely to lead to the development of more effective neuroprotective treatments for patients.

In order to find possible created and published ALS domain ontology or ontologies related to ALS, we have searched internet resources, including PubMed, ScienceDirect, IEEE Xplore, Google Scholar, Research Gate, and Springer Link, along with the special purpose semantic search engine Watson and classical key word search using Google. Moreover, we searched also repositories such as the NCBO BioPortal (http://bioportal.bioontology.org/), currently the most comprehensive biomedical ontology repository, Ontology Lookup Service from the European Bioinformatics Institute (http://www.ebi.ac.uk/ontologylookup) and the Open Biomedical Ontology (OBO) Foundry (http://www.obofoundry.org/). The searches returned only one ALS-related ontology (OntoPaRON), which was used to describe the care pathway for patients from the Ile-de-France ALS network (Cardoso et al. 2021). OntoPaRON concentrates on the notions relating to area of medical care and social help for ALS patients, but the representation of medical concepts is poor and confined to basic anatomical concepts.

Therefore we concluded that to the best of our knowledge, the ontology of the ALS domain presented in this article is the first attempt to develop a semantic, computational model of ALS knowledge.

Notably, the search found two domain ontologies related to other neurodegenerative diseases, namely Parkinson's (Younesi et al. 2015) and Alzheimer's (Malhotra et al. 2014). In this way, the new ALS model will complement the spectrum of existing models related to neurodegenerative pathology.

The domain ontology of ALS disease should enable the integration of clinical and genetic data, as well as extended analysis of these data, which could ultimately translate into solving many problems both in ALS research and in clinical practice.

The paper is organized as follows: second section presents ontology construction methodology as well as knowledge acquisition and reuse problem; third section provides general model description, giving also in subsections the details of the ontology modules; section four concludes the paper.

2. DALSO construction – methodology, knowledge acquisition and reuse

DALSO development is closely related to the OnWebDUALS research project.

2.1. Ontology-based Web Database for Understanding ALS (OnWebDUALS)

The OnWebDUALS project was carried out under the EU Joint Programme – Neurodegenerative Disease Research (JPND) in 2015-2018. The project consortium consisted of 6 neurology departments from 5 countries, genetic laboratory from Sweden, and one research team from Nalecz Institute of Biocybernetics and Biomedical Engineering, Warsaw.

The main objective of this project was to investigate the interplay between the characteristics of ALS patients, their genetic profiles, and environmental influence. In addition, other aspects were investigated, such as the pattern of disease spreading from the first affected region to other regions of the body, family history, related diseases, injuries and surgeries, exercise, and diagnostic pathway in different countries.

The central to this project was to build an ALS domain ontology comprehensively representing a body of related medical knowledge and serving as the formal basis for building a standardized E-health record for anonymous ALS was patients, implemented in the European ALS Web-database.

ALS patients were interviewed during clinical practice using a standard questionnaire developed by the OnWebDUALS consortium at five European University Hospital centers: Akdeniz University, Antalya, Turkey; Hannover University Medical School, Germany; Jena University Hospital, Germany; Faculty of Medicine, University Hospital of Lisbon, Portugal; Medical University of Warsaw, Poland. Clinical evaluation at each center was performed by a designated senior neurologist with extensive clinical experience in the treatment of ALS. Ethical consent to implement the project was granted by the relevant Clinical Ethics Committees of the participating centers, and each patient and control subject gave written informed consent to the collection and analysis of his/hers medical data.

The recruitment process permitted to interview a total of 1376 patients from the following centers: University Hospital Jena: 148 patients; Hannover Medical School: 236 patients; Medical University of Warsaw: 286 patients; Faculty of Medicine, University of Lisbon: 471 patients; and Akdeniz University Faculty of Medicine, Antalya: 235 patients. For comparisons, about 1000 controls from the above referred centers were included.

Based on this large dataset, several papers were published so far, describing the influence of environment and lifestyle factors on ALS onset and progression (Kuraszkiewicz et al. 2018; Korner et al. 2019), risk factors for polyneuropathy in ALS (de Carvalho et al. 2020), the patterns of disease presentation (Prell et al. 2020; Gromicho et al. 2020; Gromicho et al. 2021; Matos et al. 2020), diagnostic pathway of ALS patients (Campos et al. 2021, 2023), the rate of disease progression (Prell, 2020; Barc et al. 2020), the possible biomarkers (Hertel et al. 2022), the impact of comorbidities (Diekmann et al. 2020; Pereira et al. 2021), the predictive impact of C9Orf72 mutation in respiratory decline (Miltenberger-Miltenyi et al. 2019), family history of neurodegenerative diseases (Campos et al., 2020), influence of ALS on fertility (Uysal et al. 2021), association of the contact sports with ALS (Henriques et al. 2023), as well as potential predictive measures (Kuraszkiewicz et al. 2020).

2.2 Applied methodology, DALSO aim and range

METHONTOLOGY (Fernández-López et al. 1997) is a method of ontology building that is grounded in software engineering methodologies. It is well suited for building ontologies from scratch, but also recommends and supports ontological knowledge reuse.

The main, general aim of DALSO is to formalize medical terminology in the domain of ALS disease, embedding it in the broad context of knowledge involved in ALS diagnosis, clinical finding, ALS disease features, risk factors, genetics and pathophysiological mechanisms of MN degeneration. The basic scope of ALS domain ontology has been defined during interviews with neurologists involved in the OnWebDUALS research project. These meetings aimed to clarify the purpose, significance and mutual relations of the ALS patient characteristics contained in the questionnaire prepared for data collection during the OnWebDUALS project. It should be noted that the initial questionnaire content was also advised by a group of 40 expert neurologists familiar with ALS-related problems, but not

involved in the project. The result of this assessment of ALS relevant notions has been described in (Mamede De Carvalho et al. 2017). It identified a consensual set of clinical data with 42 specific items that can be used as a minimal data set for patient registers and for clinical trials.

Compared to the above-mentioned minimal data set, the patient's questionnaire ultimately applied in the OnWebDUALS project has been subsequently significantly expanded to fulfill project needs. It contains finally 156 concepts, and is presented as the supplement file1. Analogous questionnaire for control subjects contains 85 concepts and is presented as the supplement file2.

Final questionnaire version contains notions from areas such as: (1) demography, patient environment, life style and job history; (2) ALS clinical symptoms, dates of clinical events needed for describing disease course, selected ALS laboratory diagnostic tests, disease features, clinical phenotypes, and ALS diagnostic criteria; (3) ALS clinical signs and selected neurological manifestation not linked directly with ALS disease; (4) clinical history, including a wide range of comorbidities and clinical events which could potentially constitute ALS risk factors, such as serious traumas and surgeries, and (5) patient gene mutations in SOD1 and C9orf72 genes studied in the Project. In addition, other factors were explored, such as the pattern of disease spreading from the first affected region to other regions of the body, family history, exercise, and diagnostic track in different countries.

The core DALSO range has been defined by the final OnWebDUALS project questionnaire content and expanded by supplementing concepts existing in OnWebDUALS questionnaire and by adding new items not included in the questionnaire but necessary for the completeness and logical correctness of the modeled domain.

The final scope of DALSO is presented in details in section 3 of the paper.

2.3 Knowledge acquisition

The goal of the knowledge acquisition and knowledge-reuse in the ontology development process is to gather knowledge resources necessary for reliable and complete description of the model domain. Sources of knowledge can generally be classified, by reason of their form, as non-ontological and ontological. Knowledge acquisition performed during DALSO construction can be divided into three interrelated stages:

1. Establishing the possible existence of ALS domain ontology with an aim and scope at least partially corresponding to those described above; in the case of the existence of an appropriate ontology, its usefulness in the current task should be assessed, otherwise the construction of ontologies should be continued, i.e. by performing subsequent steps 2 and 3. This stage was described in section 1.2. In the absence of reusable ontological knowledge resources related to ALS, we decided to build our ontology from scratch using non-ontological knowledge and classical knowledge acquisition techniques.
2. The primary set of relevant ontology concepts has been constructed on the basis of an ALS patient questionnaire used in the OnWebDUALS Project Additional terms were collected during structured and non-structured interviews with our medical experts as well as informal or semi-formal analysis of medical texts in the field of ALS disease (risk factors—112 articles, ALS clinical knowledge—22 articles; genetic and molecular mechanisms—18 articles). All the collected terms were analyzed not only with the aim of identifying individual concepts, but also to find relationships connecting them, in order to model complex concept features and hierarchical organization of the model.
3. Searching for top-level and reference ontologies enabling knowledge reuse (5 articles).

2.4. Knowledge reuse – top-level and reference ontologies in DALSO

The effective implementation of knowledge reuse is based on two factors, the appropriate ontology construction i.e. normalization (described in next paper section), and application to its construction reference top-level ontologies.

Top-level and referential formal ontologies play an important role in the construction and integration of domain ontologies, providing a well-founded reference model that can be shared across domains. They form reference terminologies with a common set of abstract classes and properties that facilitate interoperability between domain ontologies and support consistent modeling decisions. To ensure knowledge reuse and interoperability with DALSO, five top-level reference ontologies have been selected.

The first upper-level ontology SCTO (El-Sappagh et al. 2018), built specifically for SNOMED-CT ontology (described below), is playing the role of general top-level reference ontology and was imported as a whole. SCTO defines the concepts of general medical science, to serve as a frame of reference for fundamental aspects of medicine such as 'disorder', 'diagnosis', 'disease', 'disease course', 'symptom'. DALSO top-level classes are rooted in SCTO in order to avoid unclear interpretation and to enable compatibility between DALSO and other ontologies in biomedicine.

Next four biomedical reference ontologies have been used to externally reference general medical concepts in DALSO: (i) Foundational Model of Anatomy (FMA) ontology (Rosse and Mejino 2003) for human anatomy notions, (ii) Human Phenotype (HP) ontology (Robinson et al. 2008) for patient symptoms and findings, (iii) Disease Ontology (DO) (Schriml et al. 2012) for ALS patient comordibities and pathological conditions, and (iv) SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) ontology (Spackman 2000) was used to enhance granularity of symptoms, drugs, laboratory tests and epidemiology concepts.

The use of selected external concepts of top-level reference ontologies in DALSO meets the requirements of the minimum information to reference an external ontology term guidelines (MIREOT, Courtot et al. 2011).

3. DALSO general structure and information content

Medical data needed for reliable description or ALS diagnosis contain the results from many different medical knowledge areas: neurological symptoms such as upper and lower motor neuron damage, cerebellar, sensory and extrapyramidal involvement, etc., results of blood and genetic tests, electrophysiological measurements, ALS disease type classifications and clinimetric scales providing means for quantitative assessment of patient functional or mental state, and, finally, patient and his/her family medical history. These data are complemented with information relating to disease etiology and potential environmental, social and lifestyle risk factors.

To enable the construction of an ontology model, this complex and broad ALS-related medical information set should be adequately structured and modeled. The overall structure of the DALSO was based on guidelines formulated for construction of modular, normalized ontologies with explicit structure, being transparent for human readers, suitable for machine inference and supporting ontological model reuse and maintainability. Normalized ontology is an ontology organized into orthogonal modules, formed by sets of homogenous class hierarchies, joined with relations and class definitions used for restrictions and complex notions logical descriptions.

The DALSO contains 910 classes, structured in eight modules. The resulting, first level of DALSO is presented in Fig.1, together with the content of the modules and the basic relations

between them. The DALSO syntactic correctness was validated by Fact++ reasoner (Tsarkov and Horrocks 2006) embedded in ontology editor applied, Protégé-2000 ver. 4.3.

ALS Disease Description Concept module (no. 2 in Fig. 1) is the central part of the model. To be precise, the class **ALS Disease Description** in this module can be considered as the fundamental concept in DALSO. The ALS description is expressed in categories such as disease phenotype and genotype, genetic and non-genetic risk factors and demography. Each of these general characteristics is defined as one of the remaining classes in the **ALS Disease Description Concept** module.

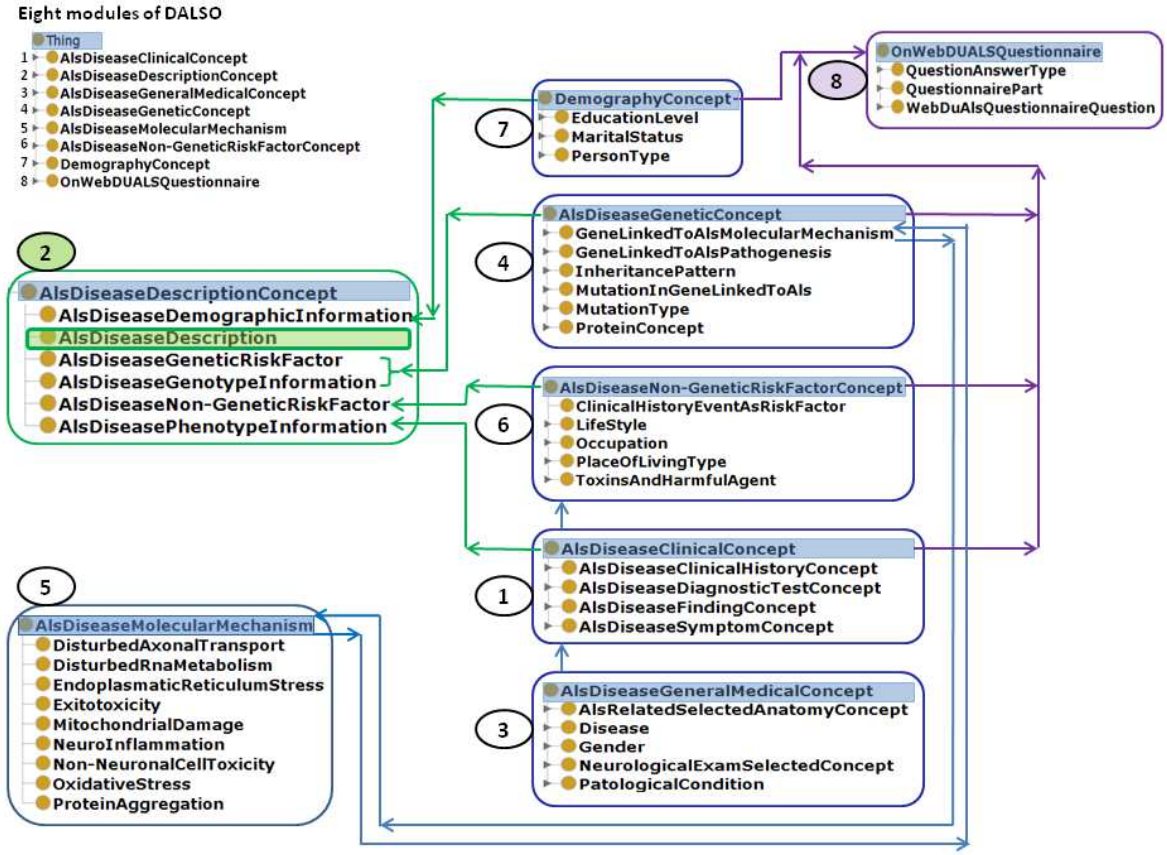

Fig. 1. The first two levels of DALSO: modules (1st level) with relations between them as shown by arrows and main classes (2nd level). Left upper corner, a list of eight DALSO modules (note that all the lists are sorted alphabetically). Middle left, central DALSO module, **ALS Disease Description Concept** (2), with the class **ALS Disease Description** highlighted in green. Central part of the figure, five DALSO modules (from top to bottom: 7, 4, 6, 1, 3) providing classes defining ALS disease characteristics. Left lower corner, module of ALS Disease Molecular Mechanisms (5), right upper corner, **OnWebDUALS** Project Questionnaire modeled as ontology module 8

3.1 Module ALS Disease Clinical Concept

The main aim of **ALS Disease Clinical Concept** module (1) is to provide classes necessary to define the disease phenotype. These classes represent clinical history, symptoms, diagnostics tests and clinical finding. All those classes described below allow to build the definition of the class **ALS Disease Phenotype Information** in DALSO module **ALS Disease Description Concept**.

The class **ALS Disease Clinical History Concept** models important facts from the medical history of the patient, which can act as ALS risk factors encompassing comorbidities (cancer, immune disease, viral infections, cardiovascular or metabolic diseases), blood transfusion or

donation, severe surgeries and traumas. All of these medical events that can be considered as ALS risk factors are used to build classes in the module **ALS Disease Non-Genetic Risk Factor Concept**.

The Class **ALS Disease Symptom Concept** contains notions defining the most common ALS patient complaints.

The class **ALS Disease Diagnostic Test Concep**t models the wide range of laboratory investigations, including blood, urine and cerebro-spinal fluid tests. Their results are used along with neurophysiology tests and neuroimaging to exclude ALS mimic syndromes, but also as ALS-specific prognostic biomarkers (Ghasemi et al. 2012; Sanderson et al. 2015; Turner et al. 2013a).

Clinical heterogeneity is a consistent feature of ALS (Al-Chalabi et al. 2016; Chiò et al. 2011; Ravits and La Spada 2009; Sabatelli et al. 2013). Major factors contributing to variable clinical phenotypes are age of onset, proportion of upper MN to lower MN deficits and unreliable prediction of disease progression (Chiò et al. 2011; Ravits and La Spada 2009; Sabatelli et al. 2013). Difficult, often delayed, diagnosis or inaccurate patient selection criteria for clinical trials, problems often encountered in the literature devoted to ALS, are partially related to ALS heterogeneity.

The content of the class **Als Disease Finding Concept** has been mainly dictated by the need of representation in DALSO the above-described ALS clinical heterogeneity. Its subclasses model concepts such as: (i) ALS clinical phenotypes based on age of onset, site of onset, disease progression rate and motor involvement; (ii) ALS diagnostic criteria, El Escorial (De Carvalho et al. 2008) and Awaji (De Carvalho and Swash 2009); (iii) ALS staging systems, Milano-Torino (MiToS) functional staging (Chiò et al. 2015) and King's clinical staging systems (Roche et al. 2012); (iv) scales assessing ALS patient psychic and cognitive state, including ALSFRS scale (Cedarbaum et al. 1999) a validated rating tool for monitoring the progression of disability in patients with ALS; (v) ALS type (fALS, sALS) based on disease extent in patient family.

3.2 Module ALS Disease General Medical Concept

ALS Disease General Medical Concept module (3) groups five classes, used to complement the definitions of classes modeling ALS-specific clinical concepts. The class **ALS Related Selected Anatomy Concept** represents concepts necessary to define neurological clinical findings related to specific regions of the body, and concepts from clinical history. The classes **Disease** and **Pathological Condition** are used to define comordibities or events from clinical history. The class **Neurological Exam Selected Concept** gather subclasses representing concepts from the neurological examination such as motor and sensory systems, coordination, gait, and reflexes. These notions are subsequently used in the module **ALS Disease Clinical Concept (1)** to build the class **ALS Disease Finding Concept**, which finally is used to define the class **ALS Disease Phenotype Information** in the module (2).

The authors of (Gkoutos et al. 2009) stated that HP reference ontology, standardized vocabulary of phenotypic abnormalities related to human diseases, usually presents as a whole the intrinsic complexity of the defined terms. To formalize precise, fine-grained and multi-aspect knowledge implicitly inherent with clinical phenotype, it is necessary to decompose it into basic information units.

We will explain this with an example of HP ontology derived phenotype trait, characteristic for bulbar onset of ALS, *Tongue Fasciculation* (id HP:0001308), which can be broken down into the anatomical concept *Tongue Muscle* and a motor system finding *Fasciculation*. Here, the elementary features are: (i) localization in two anatomic structures, tongue muscle and

head; (ii) localization in bulbar region, an anatomic area specific for ALS; (iii) the relation of the modeled phenotypic trait to specific MN type, in this case the lower MN. They have been modeled as subclasses of the root class **ALS Disease Selected Anatomy Concept**. The relation of the class Tongue Muscle Fasciculation with its features is modeled via four object type properties. The resulting *Tongue Fasciculation* composite concept is defined at a higher level of detail, allowing for more precise phenotype description. The definition of the class Tongue Muscle Fasciculation is presented in Fig. 2.

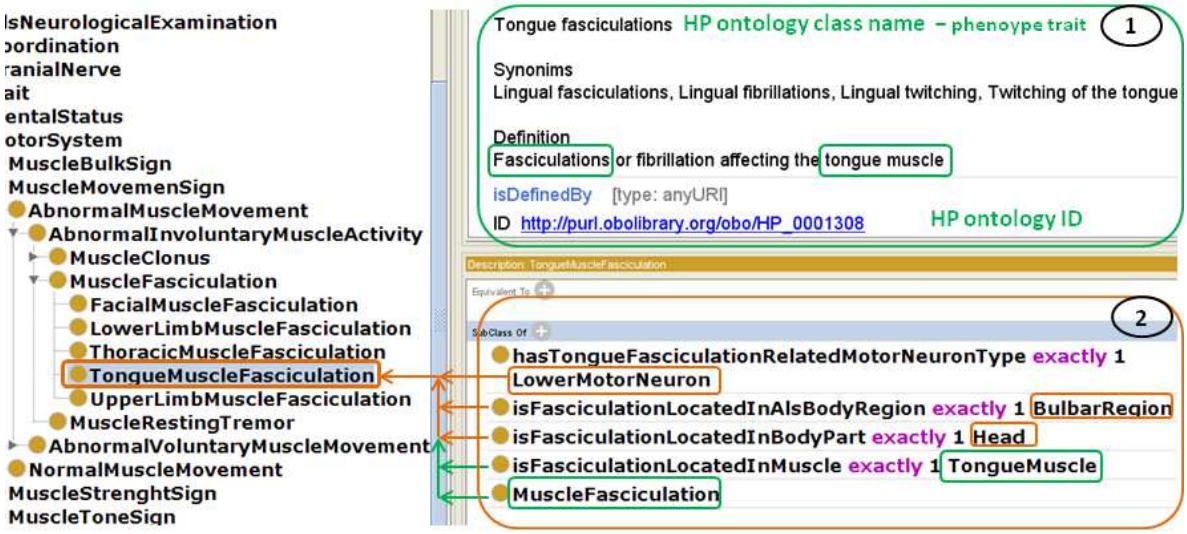

Fig. 2. The definition of the class Tongue Muscle Fasciculation in DALSO (screenshot from ontology editor Protege-2000). On the left, a fragment of class hierarchy, defined class highlighted in orange. Green panel (1), references from HP ontology to the class Tongue Muscle Fasciculation: HP term for a phenotype trait, HP identifier and textual definition. Orange panel (2), details of the class Tongue Muscle Fasciculation definition. Notions implicitly present in HP definition, Tongue Muscle and Fasciculation (marked with green) are made explicit, additional notions *Lower Motor Neuron*, *Bulbar Region* and *Head* are also made explicit (marked with orange). The resulting definition of *Tongue Fasciculation* in DALSO allows for deeper understanding of modeled phenotype trait.

3.3 Module ALS Disease Non-Genetic Risk Factor

Although many risk factors have been proposed and studied for ALS, little progress has been made in identifying those with a high degree of causality (Wang et al. 2017). Current research suggests that the only established risk factors for ALS are family history of ALS (fALS), male gender and older age. Identification of non-genetic ALS risk factors, i.e. environmental, toxic, infectious or metabolic etiologies of ALS, is highly relevant for detecting mechanisms of susceptibility to the sporadic form of ALS.

The knowledge regarding risk factors has been collected as a result of discussions among OnWebDUALS experts and a review on the literature on ALS risk factors was prepared in frames of the ONWebDUALS Project (eventually published as an opinion article, Kuraszkiewicz et al. 2018). Most of the authors point to possible ALS non-genetic risk factors that can be divided into the following four main groups: related to lifestyle, occupation, environment or medical history. However, the research in this area is generally inconclusive.

The class **LifeStyle** represents lifestyle-related ALS risk factors, including smoking (Armon 2003; Das et al. 2012; Turner 2013), dietary factors, nutritional status, vitamin deficiency (Rosenfeld and Ellis 2008; Verburgh 2014), physical fitness and body-mass index (Armon 2003; Turner 2013).

The class **Occupation** models in DALSO occupational risk-factors, as e.g., exposure of industry and construction workers to heavy metals, electromagnetic fields and toxic

chemicals, such as solvents and dyes or exposure of soldiers during military service to chemicals used in battlefields, severe injuries and stress. The class **Place Of Living** contains environmental risk factors, such as residential exposure to electromagnetic fields or to chemicals, including pesticides (Beard et al. 2016; Cragg et al. 2017; Das et al. 2012; Park et al. 2005; Su et al. 2016; Tai et al. 2017; Weisskopf et al. 2009).

The class **Clinical History Event As Risk Factor** represents selected medical history facts that can act as ALS risk factors, including comorbidities, e.g. viral infections (Alfahad and Nath 2013; Oluwole et al. 2007), autoimmune diseases (Turner et al. 2013a) or cancer (Freedman et al. 2013; Gibson et al. 2016), medical conditions such as leaky intestine (De Marchi et al. 2018; Kuraszkiewicz et al. 2018; Rowin et al. 2017; Wright et al. 2018), head and spine traumas or surgeries, electrical injuries (Berger et al. 2000; Chen et al. 2007; Das et al. 2012; Turner et al. 2013b) or stress (Okamoto et al. 2009).

Four above described classes representing ALS non-genetic risk factor serve in turn to build the definition of the class **ALS Disease Phenotype Information** in the module (2). Class **Toxins And Harmful Agent** is an auxiliary class that models concepts such as chemical, physical or emotional stressors, used to complement classes representing four main ALS disease risk factors.

3.4 Module ALS Disease Genetic Concept

The module **ALS Disease Genetic Concept (4)** gathers classes representing genetic concepts included in DALSO. The class **Gene Linked To ALS Pathogenesis** groups as subclasses genes related to ALS pathogenesis, listed in Amyotropic Lateral Sclerosis Online Genetics Database (ALSoD, http://alsod.iop.kcl.ac.uk/) (Abel et al. 2012). The authors of (Fundel and Zimmer 2006) revealed the existence of significant ambiguity within gene names and showed that the degree of overlap between different public data sources is generally relatively modest.

Thus, to ensure explicitness in gene identity, all subclasses of the class **Gene Linked To ALS Pathogenesis** have three datatype properties ensuring their unambiguous identification. These are: (1) gene name from ALSoD, (2) HUGO Nomenclature Committee (HGNC) (Wain et al. 2002) identifier and (3) identifier from the online catalogue of human genes and genetic disorders (Online Mendelian Inheritance in Man, OMIM, Hamosh et al. 2002), https://www.ncbi.nlm.nih.gov/omim). Fourth data type property of the class representing genes involved in ALS pathogenesis indicates the gene location within the chromosome. The class definition is completed by relations with two important model concepts, the classes **Mutations In Gene Linked To ALS** and **Protein Concept**.

The class **Mutation In Gene Linked To ALS** gathers sub-classes representing mutations in genes related to ALS pathogenesis. Their definitions are based on the guidelines of the Mutation Nomenclature of the Human Genome Variation Society (den Dunnen and Antonarakis 2000) and also on dependencies defined in the DALSO. According to mutation nomenclature, the gene mutation is defined using four datatype properties representing mutation, its original and mutated content, offset and count. Type of the mutation is modeled as the object type relation linking class representing mutation with the class **Mutation Type**. The association of the mutation and the gene is represented via the object type relation between the classes representing gene mutation and gene.

The class **Protein Concept** is defined by its name, function and relation to the gene by which the protein is encoded. It is linked with the module **ALS Disease Molecular Mechanism** (5).

3.5 Module ALS Disease Description Concept

As it was stated in section 3.1, **ALS Disease Description Concept** module with its class **ALS Disease Description** constitutes the most important part of the DALSO. The class **ALS**

Disease Description is characterized along five axes, phenotype and genotype information, genetic and non-genetic risk factors and demography (Fig. 3). Each axis is represented as one class in the module, with a class name indicating modeled content (see Fig.1). Each class modeling specified axis relates to a specific ontology module, gathering classes modeling its structure and content.

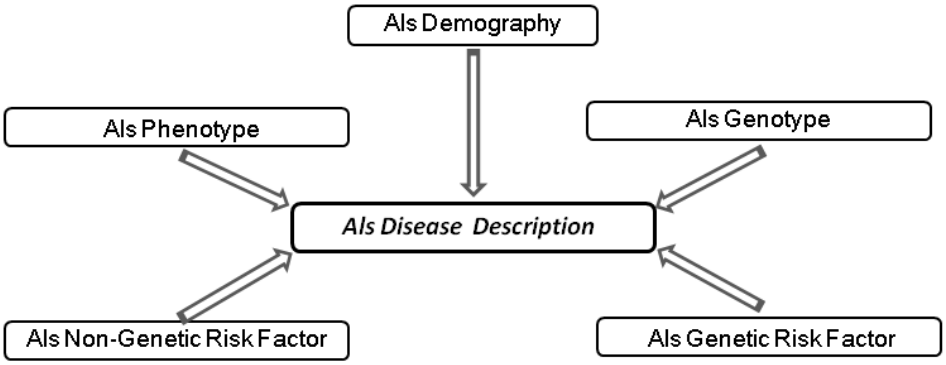

Fig. 3. General overview of the class **ALS Disease Description** structure. ALS is characterized along five information axes, phenotype and genotype information, genetic and non-genetic risk factors and demography.

3.6 Module Als Disease Molecular Mechanism

The module **ALS Disease Molecular Mechanism** (5) contains classes modeling mechanisms involved in MN degeneration. As many works indicate (Ferraiuolo et al. 2011; Shaw 2005; Taylor et al. 2016; Morgan and Orrell 2016; S. Chen et al. 2013; Souza et al. 2015; Pasinelli and Brown 2006; Zhou et al. 2017), mechanisms underlying MN death are multifactorial and MN degeneration probably results from a complex interplay of multiple mechanisms. The analysis of the cited above literature allowed distinguishing nine basic mechanisms involved in the pathway leading to the MN death:

1. Oxidative stress,
2. Excitotoxicity,
3. Mitochondrial damage,
4. Neuroinflammation,
5. Disturbed axonal transport,
6. Endoplasmic reticulum stress,
7. Protein aggregation,
8. Disturbances in RNA metabolism,
9. Non-neuronal cell dysfunction.

The complete picture of the pathogenic mechanisms of ALS is still unclear, thus true participation and significance of the above-listed putative mechanisms may in fact effectively vary in particular patients or within their subgroups. However, a deep understanding of the complex interplay between multiple processes is critical for development of new therapeutic approaches. This consideration motivated us to include in DALSO the module representing MN degeneration pathways. Many recent studies state that dysfunctions of vital molecular pathways underlying ALS pathogenesis are closely related to genetic factors, namely to mutations in specific genes (Pasinelli and Brown 2006; Chen et al. 2013; Souza et al. 2015; Morgan and Orrell 2016; Taylor et al. 2016). Each class in the module **ALS Disease Molecular Mechanism** is related to the class **Gene Linked to ALS Molecular Mechanism** in the module **(4)** representing specific genes involved in specific mechanism of motor neuron degeneration.

3.7 Modules OnWebDUALS Questionnaire and Demography Concept

The module **OnWebDUALS Questionnaire (8)** represents the content of questionnaire items used in the OnWebDUALS project (see supplement file1), defined using ontology concepts, except for concepts modeling the time periods and temporal sequence of events.

The module **Demography Concept (7),** the smallest DALSO module, gathers rudimentary demographic concepts, education level, marital status and family members. Its classes are used to model concepts in OnWebDUALS questionnaire and demographic information in the **ALS Disease Description** module (**2**, see Fig.1).

4. Conclusions

The DALSO is so far the only model representing a comprehensive, formal, structured knowledge source for the ALS domain. The ontology includes a wide range of ALS-related aspects, ranging from clinical information and clinical history data, to disease mechanisms and genetics. DALSO is the first ALS disease ontological model and it complements the spectrum of existing domain disease ontologies referring to neurodegenerative diseases (Malhotra et al. 2014; Younesi et al. 2015). The model shortcoming is a lack of concepts that allow modeling of the temporal sequence of events.

The DALSO is expressed in OWL2 language, contains 910 classes, is consistent and free of logical errors. Its syntactic correctness was validated by the Fact++ reasoner (Tsarkov and Horrocks 2006) embedded in Protégé-2000 ver 4.3 ontology editor, used to develop the model.

The DALSO provides the means for defining schemes of ALS patient resources and also for defining schemes for clinical trials. DALSO ontology used in combination with Ontop, an Ontology Based Data Access system (Calvanese et al. 2017a), can be used to access a relational database through ontological conceptual layers. The efficiency of this approach was proven by Zhang et al. (2018) for the integration of cancer data. DALSO can also be used for creation of a practicable health informatics service system (Djedidi and Aufaure 2007; Silachan and Tantatsanawong 2011).

Another potential application of DALSO is in refined versions of enrichment analysis, proposed by (Shah et al. 2012), who suggests the next step of genetic data analysis could be e.g. replacement of GO ontology with specific disease ontology.

Summing up, the DALSO can be used as a starting point for the development of systems for the diagnosis support, education of young doctors in this field, and for the integration and analysis of clinical, genetic and other data collected in various databases (Ivanović and Budimac 2014). This could help speed up the research on the ALS etiology and development of effective treatments for this lethal disease.

The final version of DALSO was created using the Protégé-2000 ver 4.3 editor. It can be uploaded at the link https://als.ibib.waw.pl/DAlsO.zip. Use of the DALSO ontology is free of charge, provided that any future publication based on its use cites this article.

Acknowledgment

This paper was supported by the grant JPND/01/2015 funded by the Polish National Center of Research and Development in frames of EU Joint Program of Neurodegenerative Research (JPND).

References

Abdollahi, M., Gao, X., Mei, Y., Ghosh, S., Li, J., & Narag, M. (2021). Substituting clinical features using synthetic medical phrases: Medical text data augmentation techniques. *Artif Intell Med, 120*, 102167, doi:10.1016/j.artmed.2021.102167.

Abel, O., Powell, J. F., Andersen, P. M., & Al-Chalabi, A. (2012). ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum Mutat, 33*(9), 1345-1351, doi:10.1002/humu.22157.

Al-Chalabi, A., Hardiman, O., Kiernan, M. C., Chiò, A., Rix-Brooks, B., & van den Berg, L. H. (2016). Amyotrophic lateral sclerosis: moving towards a new classification system. *Lancet Neurol, 15*(11), 1182-1194, doi:10.1016/s1474-4422(16)30199-5.

Alfahad, T., & Nath, A. (2013). Retroviruses and amyotrophic lateral sclerosis. *Antiviral Res, 99*(2), 180-187, doi:10.1016/j.antiviral.2013.05.006.

Alsultan, A., Waller, R., Heath, P., & Kirby, J. (2016). The genetics of amyotrophic lateral sclerosis: current insights. *Degenerative Neurological and Neuromuscular Disease, 6*, 49-64 doi:10.2147/DNND.S84956.

Andersen, P. M., Abrahams, S., Borasio, G. D., de Carvalho, M., Chiò, A., Van Damme, P., et al. (2012). EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)--revised report of an EFNS task force. *Eur J Neurol, 19*(3), 360-375, doi:10.1111/j.1468-1331.2011.03501.x.

Armon, C. (2003). An evidence-based medicine approach to the evaluation of the role of exogenous risk factors in sporadic amyotrophic lateral sclerosis. *Neuroepidemiology, 22*(4), 217-228, doi:10.1159/000070562.

Barc, K., Szacka, K., Nieporecki, K., de Carvalho, M., Gromicho, M., Grosskreutz, J., et al. (2020). Emotional Lability at Disease Onset Is an Independent Prognostic Factor of Faster Disease Progression in Amyotrophic Lateral Sclerosis. *Aging Dis, 11*(5), 1021-1028, doi:10.14336/AD.2019.1120.

Bean, D. M., Al-Chalabi, A., Dobson, R. J. B., & Iacoangeli, A. (2020). A Knowledge-Based Machine Learning Approach to Gene Prioritisation in Amyotrophic Lateral Sclerosis. *Genes (Basel), 11*(6), doi:10.3390/genes11060668.

Beard, J. D., Engel, L. S., Richardson, D. B., Gammon, M. D., Baird, C., Umbach, D. M., et al. (2016). Military service, deployments, and exposures in relation to amyotrophic lateral sclerosis etiology. *Environment International, 91*, 104-115, doi:http://dx.doi.org/10.1016/j.envint.2016.02.014.

Berger, M. M., Kopp, N., Vital, C., Redl, B., Aymard, M., & Lina, B. (2000). Detection and cellular localization of enterovirus RNA sequences in spinal cord of patients with ALS. *Neurology, 54*(1), 20-25.

Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., et al. (2008). The NIFSTD and BIRNLex Vocabularies: Building Comprehensive Ontologies for Neuroscience. *Neuroinformatics, 6*(3), 175-194, doi:10.1007/s12021-008-9032-z.

Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., et al. (2017a). Ontop: Answering SPARQL Queries over Relational Databases. *Semantic Web, 8*(3), 471-487.

Calvanese, D., Kalayci, T. E., Montali, M., & Tinella, S. (2017b). Ontology-Based Data Access for Extracting Event Logs from Legacy Data: The onprom Tool and Methodology. *Business Information Systems (Bis 2017), 288*, 220-236, doi:10.1007/978-3-319-59336-4_16.

Campos, C. F., Gromicho, M., Uysal, H., Grosskreutz, J., Kuzma-Kozakiewicz, M., Oliveira Santos, M., et al. (2021). Delayed Diagnosis and Diagnostic Pathway of ALS Patients

in Portugal: Where Can We Improve? *Front Neurol, 12*, 761355, doi:10.3389/fneur.2021.761355.

Campos, C. F., Gromicho, M., Uysal, H., Grosskreutz, J., Kuzma-Kozakiewicz, M., Oliveira Santos, M., et al. (2023). Trends in the diagnostic delay and pathway for amyotrophic lateral sclerosis patients across different countries. *Front Neurol, 13*, 1064619, doi:10.3389/fneur.2022.1064619.

Cardoso, S., Meneton, P., Aimé, X., Meininger, V., Grabli, D., Guezennec, G., et al. (2021). Use of a modular ontology and a semantic annotation tool to describe the care pathway of patients with amyotrophic lateral sclerosis in a coordination network. *PLoS One, 16*(1), e0244604, doi:10.1371/journal.pone.0244604.

Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., et al. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). *J Neurol Sci, 169*(1-2), 13-21.

Cengiz, B., & Kuruoğlu, R. (2020). A new parameter to discriminate amyotrophic lateral sclerosis patients from healthy participants by motor cortical excitability changes. *Muscle Nerve, 61*(3), 354-362, doi:10.1002/mus.26786.

Charcot, J.-M., & Joffroy, A. (1869). *Deux cas d'atrophie musculaire progressive: avec lésions de la substance grise et des faisceaux antérolatéraux de la moelle épinière*. Paris: Masson.

Chen, H., Richard, M., Sandler, D. P., Umbach, D. M., & Kamel, F. (2007). Head Injury and Amyotrophic Lateral Sclerosis. *American Journal of Epidemiology, 166*(7), 810-816, doi:10.1093/aje/kwm153.

Chen, S., Sayana, P., Zhang, X., & Le, W. (2013). Genetics of amyotrophic lateral sclerosis: an update. *Molecular Neurodegeneration, 8*, 28-28, doi:10.1186/1750-1326-8-28.

Chiò, A., Calvo, A., Moglia, C., Mazzini, L., & Mora, G. (2011). Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study. *J Neurol Neurosurg Psychiatry, 82*(7), 740-746, doi:10.1136/jnnp.2010.235952.

Chiò, A., Hammond, E. R., Mora, G., Bonito, V., & Filippini, G. (2015). Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry, 86*(1), 38-44, doi:10.1136/jnnp-2013-306589.

Courtot, M., Gibson, F., Lister, A. L., Malone, J., Schober, D., Brinkman, R. R., et al. (2011). MIREOT: the minimum information to reference an external ontology term. *Applied Ontology, 6*(1), 23-33.

Cragg, J. J., Johnson, N. J., & Weisskopf, M. G. (2017). Military Service and Amyotrophic Lateral Sclerosis in a Population-based Cohort: Extended Follow-up 1979-2011. *Epidemiology (Cambridge, Mass.), 28*(2), e15-e16, doi:10.1097/EDE.0000000000000589.

Czaplinski, A., Yen, A. A., Simpson, E. P., & Appel, S. H. (2006). Slower Disease Progression and Prolonged Survival in Contemporary Patients With Amyotrophic Lateral Sclerosis: Is the Natural History of Amyotrophic Lateral Sclerosis Changing? *Arch Neurol, 63*(8), 1139-1143, doi:10.1001/archneur.63.8.1139.

Das, K., Nag, C., & Ghosh, M. (2012). Familial, Environmental, and Occupational Risk Factors in Development of Amyotrophic Lateral Sclerosis. *North American Journal of Medical Sciences, 4*(8), 350-355, doi:10.4103/1947-2714.99517.

de Carvalho, M., Dengler, R., Eisen, A., England, J. D., Kaji, R., Kimura, J., et al. (2008). Electrodiagnostic criteria for diagnosis of ALS. *Clin Neurophysiol, 119*(3), 497-503, doi:10.1016/j.clinph.2007.09.143.

de Carvalho, M., Gromicho, M., Andersen, P., Grosskreutz, J., Kuzma-Kozakiewicz, M., Petri, S., et al. (2020). Peripheral neuropathy in ALS: phenotype association. *J Neurol Neurosurg Psychiatry*, doi:10.1136/jnnp-2020-325164.

de Carvalho, M., Ryczkowski, A., Andersen, P., Gromicho, M., Grosskreutz, J., Kuźma-Kozakiewicz, M., et al. (2017). International Survey of ALS Experts about Critical Questions for Assessing Patients with ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 18*(7-8), 505-510, doi:10.1080/21678421.2017.1349150.

de Carvalho, M., & Swash, M. (2009). Awaji diagnostic algorithm increases sensitivity of El Escorial criteria for ALS diagnosis. *Amyotroph Lateral Scler, 10*(1), 53-57, doi:10.1080/17482960802521126.

De Marchi, F., Mazzini, L., Mogna, L., Amoruso, A., Pane, M., Aloisio, I., et al. (2018). Potential role of gut microbiota in Amyotrophic Lateral Sclerosis (ALS) pathogenesis (P1.318). *Neurology, 90*(15 Supplement), P1.318.

den Dunnen, J. T., & Antonarakis, S. E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat, 15*(1), 7-12, doi:10.1002/(sici)1098-1004(200001)15:1<7::aid-humu4>3.0.co;2-n.

Diekmann, K., Kuzma-Kozakiewicz, M., Piotrkiewicz, M., Gromicho, M., Grosskreutz, J., Andersen, P. M., et al. (2020). Impact of comorbidities and co-medication on disease onset and progression in a large German ALS patient group. *Journal of Neurology, 267*(7), 2130-2141, doi:10.1007/s00415-020-09799-z.

Djedidi, R., & Aufaure, M. Medical Domain Ontology Construction Approach: A Basis for Medical Decision Support. In *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), 20-22 June 2007 2007* (pp. 509-511). doi:10.1109/CBMS.2007.69.

El-Sappagh, S., Franda, F., Farman, A., & Kwak, K.-S. (2018). SNOMED CT standard ontology based on the ontology for general medical science. *BMC Medical Informatics and Decision Making, 76*(18), 1-19.

Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). *METHONTOLOGY: From Ontological Art Towards Ontological Engineering.* Paper presented at the AAAI-97 Spring Symposium Series, Stanford University, EEUU,

Ferraiuolo, L., Kirby, J., Grierson, A. J., Sendtner, M., & Shaw, P. J. (2011). Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat Rev Neurol, 7*(11), 616-630, doi:10.1038/nrneurol.2011.152.

Freedman, D. M., Curtis, R. E., Daugherty, S. E., Goedert, J. J., Kuncl, R. W., & Tucker, M. A. (2013). The association between cancer and amyotrophic lateral sclerosis. *Cancer Causes Control, 24*(1), 55-60, doi:10.1007/s10552-012-0089-5.

Fukushima, K., Takamatsu, N., Yamamoto, Y., Yamazaki, H., Yoshida, T., Osaki, Y., et al. (2022). Early diagnosis of amyotrophic lateral sclerosis based on fasciculations in muscle ultrasonography: A machine learning approach. *Clinical Neurophysiology, 140*, 136-144, doi:https://doi.org/10.1016/j.clinph.2022.06.005.

Fundel, K., & Zimmer, R. (2006). Gene and protein nomenclature in public databases. *BMC Bioinformatics, 7*, 372, doi:10.1186/1471-2105-7-372.

Ghasemi, M., Fatehi, F., Asadi, B., & Khorvash, F. (2012). Amyotrophic lateral sclerosis with laboratory abnormalities of uncertain significance: A brief review. *Iranian journal of neurology, 11*(1), 25-27.

Gibson, S. B., Abbott, D., Farnham, J. M., Thai, K. K., McLean, H., Figueroa, K. P., et al. (2016). Population-based risks for cancer in patients with ALS. *Neurology, 87*(3), 289-294, doi:10.1212/wnl.0000000000002757.

Gkoutos, G. V., Mungall, C., Dolken, S., Ashburner, M., Lewis, S., Hancock, J., et al. (2009). Entity/quality-based logical definitions for the human skeletal phenome using PATO. *Conf Proc IEEE Eng Med Biol Soc, 2009*, 7069-7072, doi:10.1109/IEMBS.2009.5333362.

Goutman, S. A., Hardiman, O., Al-Chalabi, A., Chió, A., Savelieff, M. G., Kiernan, M. C., et al. (2022). Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol, 21*(5), 465-479, doi:10.1016/s1474-4422(21)00414-2.

Gromicho, M., Figueiral, M., Uysal, H., Grosskreutz, J., Kuzma-Kozakiewicz, M., Pinto, S., et al. (2020). Spreading in ALS: The relative impact of upper and lower motor neuron involvement. *Annals of Clinical and Translational Neurology, 7*(7), 1181-1192, doi:10.1002/acn3.51098.

Gromicho, M., Kuzma-Kozakiewicz, M., Szacka, K., Nieporecki, K., Andersen, P. M., Grosskreutz, J., et al. (2021). Motor neuron disease beginning with frontotemporal dementia: clinical features and progression. *Amyotroph Lateral Scler Frontotemporal Degener, 22*(7-8), 508-516, doi:10.1080/21678421.2021.1910309.

Hamosh, A., Scott, A. F., Amberger, J., Bocchini, C., Valle, D., & McKusick, V. A. (2002). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research, 30*(1), 52-55.

Henriques, A. R., Gromicho, M., Grosskreutz, J., Kuzma-Kozakiewicz, M., Petri, S., Uysal, H., et al. (2023). Association of the practice of contact sports with the development of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener*, 1-8, doi:10.1080/21678421.2023.2189911.

Hertel, N., Kuzma-Kozakiewicz, M., Gromicho, M., Grosskreutz, J., de Carvalho, M., Uysal, H., et al. (2022). Analysis of routine blood parameters in patients with amyotrophic lateral sclerosis and evaluation of a possible correlation with disease progression—a multicenter study. [Original Research]. *Frontiers in Neurology, 13*, doi:10.3389/fneur.2022.940375.

Ivanović, M., & Budimac, Z. (2014). An overview of ontologies and data resources in medical domains. *Expert Systems with Applications, 41*(11), 5158-5166, doi:https://doi.org/10.1016/j.eswa.2014.02.045.

Kim, G., Gautier, O., Tassoni-Tsuchida, E., Ma, X. R., & Gitler, A. D. (2020). ALS Genetics: Gains, Losses, and Implications for Future Therapies. *Neuron, 108*(5), 822-842, doi:https://doi.org/10.1016/j.neuron.2020.08.022.

Korner, S., Kammeyer, J., Zapf, A., Kuzma-Kozakiewicz, M., Piotrkiewicz, M., Kuraszkiewicz, B., et al. (2019). Influence of Environment and Lifestyle on Incidence and Progress of Amyotrophic Lateral Sclerosis in A German ALS Population. *Aging and Disease, 10*(2), 205-216, doi:10.14336/Ad.2018.0327.

Korner, S., Kollewe, K., Ilsemann, J., Muller-Heine, A., Dengler, R., Krampfl, K., et al. (2013). Prevalence and prognostic impact of comorbidities in amyotrophic lateral sclerosis. *Eur J Neurol, 20*(4), 647-654, doi:10.1111/ene.12015.

Kuraszkiewicz, B., Goszczynska, H., Podsiadly-Marczykowska, T., Piotrkiewicz, M., Andersen, P., Gromicho, M., et al. (2020). Potential Preventive Strategies for Amyotrophic Lateral Sclerosis. *Front Neurosci, 14*, 428, doi:10.3389/fnins.2020.00428.

Kuraszkiewicz, B., Podsiadły-Marczykowska, T., Goszczyńska, H., & Piotrkiewicz, M. (2018). Are There Modifiable Environmental Factors Related to Amyotrophic Lateral Sclerosis? [Opinion]. *Frontiers in Neurology, 9*(220), doi:10.3389/fneur.2018.00220.

Malhotra, A., Younesi, E., Gündel, M., Müller, B., Heneka, M. T., & Hofmann-Apitius, M. (2014). ADO: A disease ontology representing the domain knowledge specific to

Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 10*(2), 238-246, doi:10.1016/j.jalz.2013.02.009.

Manjaly, Z. R., Scott, K. M., Abhinav, K., Wijesekera, L., Ganesalingam, J., Goldstein, L. H., et al. (2010). The sex ratio in amyotrophic lateral sclerosis: A population based study. *Amyotroph Lateral Scler, 11*(5), 439-442, doi:10.3109/17482961003610853.

Marin, B., Boumédiene, F., Logroscino, G., Couratier, P., Babron, M. C., Leutenegger, A. L., et al. (2017). Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. *Int J Epidemiol, 46*(1), 57-74, doi:10.1093/ije/dyw061.

Matos, J., Pires, S., Aidos, H., Gromicho, M., Pinto, S., de Carvalho, M., et al. Unravelling Disease Presentation Patterns in ALS Using Biclustering for Discriminative Meta-Features Discovery. In I. Rojas, O. Valenzuela, F. Rojas, L. J. Herrera, & F. Ortuño (Eds.), *Bioinformatics and Biomedical Engineering, Cham, 2020// 2020* (pp. 517-528): Springer International Publishing

Mejzini, R., Flynn, L. L., Pitout, I. L., Fletcher, S., Wilton, S. D., & Akkari, P. A. (2019). ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Front Neurosci, 13*, 1310, doi:10.3389/fnins.2019.01310.

Miller, T. M., Cudkowicz, M. E., Genge, A., Shaw, P. J., Sobue, G., Bucelli, R. C., et al. (2022). Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS. *N Engl J Med, 387*(12), 1099-1110, doi:10.1056/NEJMoa2204705.

Miltenberger-Miltenyi, G., Conceição, V. A., Gromicho, M., Pronto-Laborinho, A. C., Pinto, S., Andersen, P. M., et al. (2019). C9Orf72 expansion is associated with accelerated decline of respiratory function and decreased survival in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 90*(1), 118, doi:10.1136/jnnp-2018-318032.

Morgan, S., & Orrell, R. W. (2016). Pathogenesis of amyotrophic lateral sclerosis. *Br Med Bull, 119*(1), 87-98, doi:10.1093/bmb/ldw026.

Nielsen, F. Å. (2014). Brede Tools and Federating Online Neuroinformatics Databases. *Neuroinformatics, 12*(1), 27-37, doi:10.1007/s12021-013-9183-4.

Okamoto, K., Kihira, T., Kondo, T., Kobashi, G., Washio, M., Sasaki, S., et al. (2009). Lifestyle Factors and Risk of Amyotrophic Lateral Sclerosis: A Case-Control Study in Japan. *Annals of Epidemiology, 19*(6), 359-364, doi:http://dx.doi.org/10.1016/j.annepidem.2009.01.015.

Oluwole, S. O., Yao, Y., Conradi, S., Kristensson, K., & Karlsson, H. (2007). Elevated levels of transcripts encoding a human retroviral envelope protein (syncytin) in muscles from patients with motor neuron disease. *Amyotroph Lateral Scler, 8*(2), 67-72, doi:10.1080/17482960600864207.

Osborne, J. D., Flatow, J., Holko, M., Lin, S. M., Kibbe, W. A., Zhu, L. J., et al. (2009). Annotating the human genome with Disease Ontology. *BMC Genomics, 10 Suppl 1*, S6, doi:10.1186/1471-2164-10-s1-s6.

Park, R. M., Schulte, P. A., Bowman, J. D., Walker, J. T., Bondy, S. C., Yost, M. G., et al. (2005). Potential occupational risks for neurodegenerative diseases. *Am J Ind Med, 48*(1), 63-77, doi:10.1002/ajim.20178.

Pasinelli, P., & Brown, R. H. (2006). Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci, 7*(9), 710-723, doi:10.1038/nrn1971.

Pereira, M., Gromicho, M., Henriques, A., Pronto-Laborinho, A. C., Grosskreutz, J., Kuzma-Kozakiewicz, M., et al. (2021). Cardiovascular comorbidities in amyotrophic lateral sclerosis. *J Neurol Sci, 421*, 117292, doi:10.1016/j.jns.2020.117292.

Prell, T., Gaur, N., Steinbach, R., Witte, O. W., & Grosskreutz, J. (2020). Modelling disease course in amyotrophic lateral Sclerosis: pseudo-longitudinal insights from cross-

sectional health-related quality of life data. *Health and Quality of Life Outcomes, 18*(1), 117, doi:10.1186/s12955-020-01372-6.
Ravits, J. M., & La Spada, A. R. (2009). ALS motor phenotype heterogeneity, focality, and spread: deconstructing motor neuron degeneration. *Neurology, 73*(10), 805-811, doi:10.1212/WNL.0b013e3181b6bbbd.
Robinson, P. N., Köhler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *American journal of human genetics, 83*(5), 610-615, doi:10.1016/j.ajhg.2008.09.017.
Roche, J. C., Rojas-Garcia, R., Scott, K. M., Scotton, W., Ellis, C. E., Burman, R., et al. (2012). A proposed staging system for amyotrophic lateral sclerosis. *Brain, 135*(Pt 3), 847-852, doi:10.1093/brain/awr351.
Rosenfeld, J., & Ellis, A. (2008). Nutrition and Dietary Supplements in Motor Neuron Disease. *Physical Medicine and Rehabilitation Clinics, 19*(3), 573-589, doi:10.1016/j.pmr.2008.03.001.
Rosse, C., & Mejino, J. L., Jr. (2003). A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform, 36*(6), 478-500, doi:10.1016/j.jbi.2003.11.007.
Rowin, J., Xia, Y., Jung, B., & Sun, J. (2017). Gut inflammation and dysbiosis in human motor neuron disease. *Physiol Rep, 5*(18), doi:10.14814/phy2.13443.
Ruttenberg, A., Rees, J. A., Samwald, M., & Marshall, M. S. (2009). Life sciences on the Semantic Web: the Neurocommons and beyond. *Brief Bioinform, 10*(2), 193-204, doi:10.1093/bib/bbp004.
Ryan, M., Heverin, M., McLaughlin, R. L., & Hardiman, O. (2019). Lifetime Risk and Heritability of Amyotrophic Lateral Sclerosis. *JAMA Neurol, 76*(11), 1367-1374, doi:10.1001/jamaneurol.2019.2044.
Sabatelli, M., Conte, A., & Zollino, M. (2013). Clinical and genetic heterogeneity of amyotrophic lateral sclerosis. *Clin Genet, 83*(5), 408-416, doi:10.1111/cge.12117.
Sánchez, D., & Moreno, A. Learning Medical Ontologies from the Web. In D. Riaño (Ed.), *Knowledge Management for Health Care Procedures, Berlin, Heidelberg, 2008// 2008* (pp. 32-45): Springer Berlin Heidelberg
Sanderson, A. B., Novak, J. C., Nash, S. M., Kolb, S. J., & Kissel, J. T. (2015). Laboratory evaluation of suspected motor neuron disease: A survey of physicians. *Muscle Nerve, 52*(1), 83-87, doi:10.1002/mus.24639.
Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., et al. (2012). Disease Ontology: a backbone for disease semantic integration. *Nucleic acids research, 40*(Database issue), D940-D946, doi:10.1093/nar/gkr972.
Segura, T., Medrano, I. H., Collazo, S., Maté, C., Sguera, C., Del Rio-Bermudez, C., et al. (2023). Symptoms timeline and outcomes in amyotrophic lateral sclerosis using artificial intelligence. *Scientific Reports, 13*(1), 702, doi:10.1038/s41598-023-27863-2.
Sennfält, S., Kläppe, U., Thams, S., Samuelsson, K., Press, R., Fang, F., et al. (2023). The path to diagnosis in ALS: delay, referrals, alternate diagnoses, and clinical progression. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 24*(1-2), 45-53, doi:10.1080/21678421.2022.2053722.
Shah, N. H., Cole, T., & Musen, M. A. (2012). Chapter 9: Analyses Using Disease Ontologies. *PLOS Computational Biology, 8*(12), e1002827, doi:10.1371/journal.pcbi.1002827.
Shaw, P. (2005). Molecular and cellular pathways of neurodegeneration in motor neurone disease. *J Neurol Neurosurg Psychiatry, 76*(8), 1046-1057, doi:10.1136/jnnp.2004.048652.

Silachan, K., & Tantatsanawong, P. Domain Ontology Health Informatics Service from Text Medical Data Classification. In *2011 Annual SRII Global Conference, 29 March-2 April 2011 2011* (pp. 357-362). doi:10.1109/SRII.2011.48.

Souza, P. V., Pinto, W. B., Chieia, M. A., & Oliveira, A. S. (2015). Clinical and genetic basis of familial amyotrophic lateral sclerosis. *Arq Neuropsiquiatr, 73*(12), 1026-1037, doi:10.1590/0004-282x20150161.

Spackman, K. (2000). SNOMED RT and SNOMEDCT. Promise of an international clinical terminology. *M.D. computing : computers in medical practice, 17*(6), 29.

Staab, S., & Studer, R. (2010). *Handbook on ontologies*: Springer Science & Business Media.

Steiner, C. M., Albert, D., & Wang, S. (2017). Validating domain ontologies: A methodology exemplified for concept maps. *Cogent Education, 4*(1), doi:10.1080/2331186X.2016.1263006

Su, F. C., Goutman, S. A., Chernyak, S., Mukherjee, B., Callaghan, B. C., Batterman, S., et al. (2016). Association of Environmental Toxins With Amyotrophic Lateral Sclerosis. *JAMA Neurol, 73*(7), 803-811, doi:10.1001/jamaneurol.2016.0594.

Tai, H., Cui, L., Shen, D., Li, D., Cui, B., & Fang, J. (2017). Military service and the risk of amyotrophic lateral sclerosis: A meta-analysis. *J Clin Neurosci, 45*, 337-342, doi:10.1016/j.jocn.2017.08.035.

Tao, C., Pathak, J., Solbrig, H. R., Wei, W. Q., & Chute, C. G. (2013). Terminology representation guidelines for biomedical ontologies in the semantic web notations. *J Biomed Inform, 46*(1), 128-138, doi:10.1016/j.jbi.2012.09.003.

Taylor, J. P., Brown, R. H., Jr., & Cleveland, D. W. (2016). Decoding ALS: from genes to mechanism. *Nature, 539*(7628), 197-206, doi:10.1038/nature20413.

Traynor, B. J., Codd, M. B., Corr, B., Forde, C., Frost, E., & Hardiman, O. (2000). Amyotrophic lateral sclerosis mimic syndromes: a population-based study. *Arch Neurol, 57*(1), 109-113.

Tsarkov, D., & Horrocks, I. (2006). FaCT++ Description Logic Reasoner: System Description. In U. Furbach, & N. Shankar (Eds.), (IJCAR 2006 ed., Vol. 4130): Springer, Berlin, Heidelberg.

Turner, M. R. (2013). Increased premorbid physical activity and amyotrophic lateral sclerosis: born to run rather than run to death, or a seductive myth? *J Neurol Neurosurg Psychiatry, 84*(9), 947, doi:10.1136/jnnp-2013-304935.

Turner, M. R., Bowser, R., Bruijn, L., Dupuis, L., Ludolph, A., McGrath, M., et al. (2013a). Mechanisms, models and biomarkers in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener, 14 Suppl 1*, 19-32, doi:10.3109/21678421.2013.778554.

Turner, M. R., Hardiman, O., Benatar, M., Brooks, B. R., Chiò, A., de Carvalho, M., et al. (2013b). Controversies and priorities in amyotrophic lateral sclerosis. *Lancet Neurol, 12*(3), 310-322, doi:10.1016/s1474-4422(13)70036-x.

Turner, M. R., & Swash, M. (2015). The expanding syndrome of amyotrophic lateral sclerosis: a clinical and molecular odyssey. *J Neurol Neurosurg Psychiatry, 86*(6), 667-673, doi:10.1136/jnnp-2014-308946.

van Blitterswijk, M., van Es, M. A., Hennekam, E. A., Dooijes, D., van Rheenen, W., Medic, J., et al. (2012). Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet, 21*(17), 3776-3784, doi:10.1093/hmg/dds199.

Van Daele, S. H., Moisse, M., van Vugt, J. J. F. A., Zwamborn, R. A. J., van der Spek, R., van Rheenen, W., et al. (2023). Genetic variability in sporadic amyotrophic lateral sclerosis. *Brain, 146*(9), 3760-3769, doi:10.1093/brain/awad120.

Verburgh, K. (2014). Nutrigerontology: why we need a new scientific discipline to develop diets and guidelines to reduce the risk of aging-related diseases. *Aging Cell, 14*(1), 17-24, doi:10.1111/acel.12284.

Verstraete, E., & Foerster, B. R. (2015). Neuroimaging as a New Diagnostic Modality in Amyotrophic Lateral Sclerosis. *Neurotherapeutics, 12*(2), 403-416, doi:10.1007/s13311-015-0347-9.

Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W., & Povey, S. (2002). Guidelines for human gene nomenclature. *Genomics, 79*(4), 464-470, doi:10.1006/geno.2002.6748.

Wang, M.-D., Little, J., Gomes, J., Cashman, N. R., & Krewski, D. (2017). Identification of risk factors associated with onset and progression of amyotrophic lateral sclerosis using systematic review and meta-analysis. *NeuroToxicology, 61*, 101-130, doi:https://doi.org/10.1016/j.neuro.2016.06.015.

Weisskopf, M. G., Morozova, N., O'Reilly, E. J., McCullough, M. L., Calle, E. E., Thun, M. J., et al. (2009). Prospective study of chemical exposures and amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry, 80*(5), 558-561, doi:10.1136/jnnp.2008.156976.

Wiesenfarth, M., Dorst, J., Brenner, D., Elmas, Z., Parlak, Ö., Uzelac, Z., et al. (2024). Effects of tofersen treatment in patients with SOD1-ALS in a "real-world" setting - a 12-month multicenter cohort study from the German early access program. *EClinicalMedicine, 69*, 102495, doi:10.1016/j.eclinm.2024.102495.

Witzel, S., Wagner, M., Zhao, C., Kandler, K., Graf, E., Berutti, R., et al. (2022). Fast versus slow disease progression in amyotrophic lateral sclerosis–clinical and genetic factors at the edges of the survival spectrum. *Neurobiology of Aging, 119*, 117-126, doi:https://doi.org/10.1016/j.neurobiolaging.2022.07.005.

Wright, M. L., Fournier, C., Houser, M. C., Tansey, M., Glass, J., & Hertzberg, V. S. (2018). Potential Role of the Gut Microbiome in ALS: A Systematic Review. *Biol Res Nurs, 20*(5), 513-521, doi:10.1177/1099800418784202.

Younesi, E., Malhotra, A., Gündel, M., Scordis, P., Kodamullil, A. T., Page, M., et al. (2015). PDON: Parkinson's disease ontology for representation and modeling of the Parkinson's disease knowledge domain. [journal article]. *Theoretical Biology and Medical Modelling, 12*(1), 20, doi:10.1186/s12976-015-0017-y.

Zhang, H., Guo, Y., Li, Q., George, T. J., Shenkman, E., Modave, F., et al. (2018). An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC Med Inform Decis Mak, 18*(Suppl 2), 41, doi:10.1186/s12911-018-0636-4.

Zhang, H., Guo, Y., Li, Q., George, T. J., Shenkman, E. A., & Bian, J. (2017). Data Integration through Ontology-Based Data Access to Support Integrative Data Analysis: A Case Study of Cancer Survival. *Proceedings (IEEE Int Conf Bioinformatics Biomed), 2017*, 1300-1303, doi:10.1109/BIBM.2017.8217849.

Zhou, T., Ahmad, T. K., Gozda, K., Truong, J., Kong, J., & Namaka, M. (2017). Implications of white matter damage in amyotrophic lateral sclerosis (Review). *Mol Med Rep, 16*(4), 4379-4392, doi:10.3892/mmr.2017.7186.