Supplementary Material for NeDRex-Web

1 Methods

1.1 Implementation

The main technologies used to create NeDRex-Web are: i) vue.js (v2.5.11) as a primary component-based web-framework with vuetifyjs (v2.6.10) for component styling, fontawesome icons (v5.15.3) and vis-network (v9.1.9) of vis.js for interactive network visualization. ii) Spring Boot (v3.1.39 as a Java-based backend framework for the handling of GET and POST requests, WebSocket connections for request status updates and database access management. iii) A MySQL (v8.0.29) database to store the large heterogeneous network and user results. All three components are deployed using Docker, allowing to quickly change the host server if necessary, to distribute the applications on separate servers, or to deploy a locally running version. NeDRex-Web was tested on all common combinations of Chrome, Firefox, Safari, and Edge running on Linux, MacOS, or Windows installations (Table S1).

1.2 Interactive Network Visualization

A key part of any web-based network explorer is the interactive visualization. In NeDRex-Web, vis.js is used as a base framework to display up to 100,000 entities at once. A force-directed layout is initially prepared for each network on the backend server to save resources on the user's machine. Users have the option to select other network layout options, that are then computed on-demand. We use "portrait", "topographic", and "geodesic" layouts from the cartoGRAPHs package (1) allowing the efficient layouting of larger networks. As "portrait" and "topograpic" layouts are provided for 3D layouting we provide all relevant binary combinations of the dimensions. All layouts are saved and reloaded in case the user decides to return to a network that has been visualized before. The network can be explored interactively by identifying first-degree neighbours and detailed information for individual nodes or by using a rectangular selection tool to select multiple nodes for follow-up analysis.

1.3 Input

Generally, users have three options to provide input in NeDRex-Web. These options are mostly standardized between the different exploration paths and assist in the selection of nodes or edges of different types from the knowledge graph. An input list can be defined using a combination of the available methods:

- **Manual ID input**: ID lists can either be pasted to the corresponding text area, where the delimiter will be automatically recognized, or uploaded as a file. The file upload supports single column lists or tsv tables as long as the IDs are located in the first column. The ID space for each node is the primary identifier type of source database (Table 1 rows 2, 3). The actual IDs can be prefixed by a lower case version of the ID space name. (Table 1 row 4).
- **Keyword search**: Most text-based node attributes are pre-processed and indexed to efficiently execute a substring search on all nodes of one type. This allows to provide suggestions that match the query in any attribute value responsively. Suggestions indicate the attribute and value that match the keyword as well as the main display name of the entry. There are different sorting styles for the list of suggestions and an option to add all suggestions at once.

Further, the keyword search is able to retrieve all entries connected to the search result and this way can add all genes associated with a specific disease or all pathways that include a specific protein, to name two examples.

- **Module identification based on gene expression data**: A special case of input is the expression file that the BiCoN algorithm needs to execute a module identification (MI) task. The matrix can be in tsv or csv format with IDs in the first column. Supported ID spaces are Symbols, EntrezIDs and Ensembl gene IDs.

1.4 Output

For the convenience of users, NeDRex-Web provides results in various ways and file formats:

- **On-page result summaries**: In NeDRex-Web result networks are visualized in the interactive graph browser and node or edge tables. The list of functions and degree of details provided at the primary result page differs between the exploration paths but any result can be loaded in the advanced view with the full list of functions available. Any entry of a table or visualized network can be inspected in detail, giving an overview over all relevant information stored in the NeDRexDB (2). IDs and other values are linked to corresponding external primary sources to enable access to the original entry.

- **Network export**: The resulting network can be downloaded in graphml format, an XML-based format supported by multiple network analysis libraries and Cytoscape (3). Also, extracted networks can be exported in form of node or edge tables in tsv format. Apart from the primary ID and the display name, users have the full control over attributes that should be included in the exported table.

- **List downloads**: In many cases users are able to save lists of entries. These can be manually created input lists, the identified disease modules, a ranked list of drugs, or a connector set in a connectivity search. In all cases, the lists can be uploaded as input to the tool in future queries.

- **Algorithms original outputs**: For all module identification and drug prioritization (DP) algorithm executions, the original result files are stored and provided on request. These may be provided as zip archive for approaches with multiple output files.

1.5 Drug Repurposing

NeDRex-Web implements a two-stage drug repurposing workflow consisting of disease module identification and drug prioritization steps and offers an easy-to-use interactive pipeline.

In the MI step, a set of user-defined seed genes or proteins is extended with the entities of the same type identified through the according interactome. This allows to identify additional drug targets and achieve better coverage of the genetic mechanism underlying the disease. NeDRex-Web provides users with the choice of six MI methods: BiCoN (4), DIAMOnD (5), DOMINO (6), KeyPathwayMiner (7), ROBUST (8), and MuST (9). An explanation of the algorithms ideas and criteria to select an approach fitting the task can be found below (Section 1.9).

In the DP step, drugs that target nodes in the constructed modules are identified serving as potential drug candidates for the disease of interest. The candidate drugs are ranked using either TrustRank (10) or Closeness Centrality (11).

The drug repurposing result overview provides a table for seeds and module genes including algorithm-specific scores, assigned by the module identification method. A similarly structured drug table contains all identified candidates. All these entries are jointly presented in an interactive network visualisation. Drugs that are already in use for the queried disease are highlighted and enriched by information from ClinicalTrials.gov (12). To evaluate the statistical significance of the whole module set or drug ranking, empirical p-values can be computed (see Section 1.9.3).

1.6 Guided Connectivity Search

Guided Connectivity Search allows identifying direct or indirect connections between two sets of source and target entities. Both sets are defined in the same way as seed sets for drug repurposing queries, without the restriction of them being of type gene or protein only. The user specifies the desired type of connection, selecting from a list of available paths. Additionally, it is possible to filter edges and nodes by attributes and define an exclusion or inclusion list of connector nodes. The result overview provides the lists of source, target and on request the connector nodes with an interactive network representation of the extracted network.

1.7 Free Exploration

The Free Exploration provides a way for mostly unrestricted exploratory analysis of the large hetergeneous NeDRex-Web knowledge graph. A base network can be extracted by selecting node and edge types and setting filter options. Based on this initial selection, the sub-network is defined and provided in a visual network browser as well as in a tabular representation. Both views support the selection of nodes and edges for the creation of a standalone network. Further, all MI and DP algorithms are available to be applied on all or a selection of genes or proteins. Entries can be added by adding unused connection types and induced edge creation is possible. These options allow for the iterative analyses to answer each users specific questions without running into an exploratory dead-end after each query.

1.8 Result caching and History

One information that NeDRex-Web needs to save persistently in the user's browser is a cookie with a unique user ID. This ID is linked to all queries, algorithm executions and results of this specific user to provide an individual research history. Each result is also uniquely identifiable through a specific URL, allowing to share it and continue exploration at any time. The unique ID is further used to save hierarchical relationships between networks. Such a relationship is established when one "base" network is used to derive another one with the initial network being the "parent" and the derived one being "child" in the resulting hierarchy. These research hierarchies are the result of iterative explorations in NeDRex-Web which are presented in the user specific research history. The goal is to help users to trace their exploration steps and find old results quickly. This is further supported by allowing custom titles and descriptions of networks to save additional information about their context.

1.9 Algorithms

1.9.1 Disease module identification algorithms

NeDRex-Web provides the user with the choice of five seed-based MI approaches. DOMINO (6) and KeyPathwayMiner (KPM) (7) identify maximal connected subnetworks highly enriched with seed genes. KPM focuses on the whole seed input at once, where DOMINO iteratively refines the enriched regions resulting in potentially multiple disconnected modules. Together with ROBUST (8), which aims to identify only the most reproducible gene additions, these three approaches might be used for input genes, which describe a disease where the mechanism is not fully understood or multiple distinct disease mechanisms exist.

In contrast, MuST (9), the method the ROBUST algorithm was built on, and also DIAMOnD (5) are better used to enrich the genetic background of an already well-defined disease by including first degree interaction partners as potentially druggable targets.

For unsupervised extraction of key genes that describe a disease mechanism, BiCoN (4) can be used. It biclusters expression data and uses the network to limit results to connected sets of genes.

BiCoN

BiCoN (Biclustering Constrained by Networks) (4) enriches general expression data-based biclustering of genes and patients/conditions by gene interaction information. It 'restricts biclusters to functionally related genes connected in molecular interaction networks' while 'maximizing the expression difference between two

subgroups of patients'. Both clusters of genes are seen as relevant for the analyzed condition and thus combined in the resulting module.
Parameters:

- **ID type**: Defines the expected ID type used in the gene expression matrix. Can be one of ['Symbol', 'Entrez', 'Ensembl'].
- **Expression matrix**: Defines the file to be uploaded containing the expression matrix.
- **Min & max**: Defines the minimal and maximal solution network size in genes that are used to bicluster patients or conditions.

DIAMOnD

DIAMOnD (DIseAse MOdule Detection) (5) is an algorithm that iteratively expands a disease module by identifying proteins that have significant connectivity patterns to known disease proteins (seeds). Rather than simply counting connections or focusing on dense communities, DIAMOnD evaluates the statistical significance of the connections between proteins and the current module. For each candidate protein, the algorithm calculates the probability that its observed or higher number of connections to seed proteins (k_s) could occur by chance, given its total number of connections (k) in the network. The connectivity significance is determined by a $P - value$ defined as:

$$P - value(k, k_s) = \sum_{k_i=k_s}^{k} p(k, k_i) \quad (5)$$

where $p(k, k_i)$ is given by the hypergeometric distribution, representing the probability that a protein with k total connections has exactly k_i connections to seed proteins in a random configuration. At each iteration, the protein with the lowest $P - value$ is added to the module, becoming a new seed for subsequent iterations. An extension of the algorithm introduces differential weighting, where connections to original seed proteins receive a weight $w = \alpha$ while connections to proteins added during the iterative process receive a weight $w = 1$. This modification allows for fine-tuning the relative importance of original disease proteins versus newly added module members.
Parameters:

- **N**: Defines the maximum number of iterations and thus the number of nodes to be added to the module.
- **Alpha**: Defines the seed node influence in the significance calculation. Connections to seed nodes are weighted according to α (default: 1) whereas non-seeds are weighted 1. An $\alpha > 1$ will prefer nodes with seed node connections over nodes with non-seed connections to the module.
- **P-cutoff**: Sets a threshold value for the module node p-value (default: 10^{-3}).

DOMINO

Discovery of Modules In Networks using Omic (6) is a quite recently developed active MI approach with the ability to resolve the issue of the over-representation of certain gene ontology terms in result modules. Further, the authors observed results to be too unspecific regarding the biological context. To combat this, DOMINO does not try to connect all active nodes by force but allows for completely independent clusters as long as active nodes are over-represented. NeDRex-Web uses DOMINO as MI method for user-defined sets of seed nodes.

Key Pathway Miner

Key Pathway Miner (KPM) (7) aims to identify functional modules through extracting connected subnetworks. This method relies on an indicator matrix based on which functionally active genes are extracted to identify key pathways within the gene-gene interactome, so network modules in which these genes are enriched. It will uncover all possible, maximal pathways that contain at most K non-active genes. Marking seed genes as active ones and adjusting the K parameter allows to provide a flexible method for MI that, depending on the configuration, is able to uncover various, unconnected disease mechanisms. In

NeDRex-Web the indicator matrix contains one column only which is defined by the set of input seeds.
Parameters:

- **K**: Defines how many non-seeds are allowed to be added in each identified pathway. Influences how freely KPM adds module nodes.

MuST

Minimal Steiner trees (9) are generated, meaning a minimal scoring tree is identified that is still only one connected component. This way, minimal cost paths within the network are extracted. The reasoning behind using multiple Steiner trees (MuST) is that PPIs are highly connected leading to non-unique solutions for the Steiner tree problem. Thus, selecting only one tree does not identify a disease mechanism but merely proposes one part of the overall module. Therefore, multiple trees are used to derive a prioritized list of target nodes the seeds are connected to, where the number of 'visits' of a non-seed node indicates the priority or significance of it.
Parameters:

- **Hub-penalty**: Defines a penalty for hubs and preferring lower network degree nodes (default: $0 \rightarrow$ no penalty).
- **Multiple**: Enables the repeated execution and consensus creation mode. This helps to eliminate the random factor if multiple minimal scoring Steiner trees exist.
- **Trees**: Only shows if 'Multiple' enabled. Defines the number of equally parametrized trees to be constructed and ultimately combined into one.
- **Iterations**: The number of iterations of the Steiner tree construction (default: 10).

ROBUST

Robust disease module mining via enumeration of diverse prize-collecting Steiner trees (ROBUST) (8) combats the issue of robustness that MI methods face regarding input permutation, through the incorporation of non-deterministic steps. Like MuST, ROBUST generates various Steiner trees with the difference of using prize-collecting Steiner trees and seeds having higher priority value than non-seeds. This way the algorithm favours the addition of non-seed nodes only if they are far from being only random occurrences. This method is provided as an approach that is able to lead to the identification of significant and possibly distinct mechanisms.
Parameters:

- **Initial Fraction**: The parameter $\alpha \in (0, 1]$ defines how expensive the Steiner trees are allowed to become. A larger α will allow more diverse results and $\alpha = 0$ would for example only return the best Steiner trees.
- **Reduction Factor**: The parameter $\beta \in [0, 1)$ dictates how many non-seed nodes are allowed to be integrated into Steiner trees. Each non-seed value will decrease with the amount of Steiner trees it was captured with. $\beta = 0$ sets a non-seeds value to 0 once it was used the first time. β close to 1 can lead to limited exploration, because a non-seeds value will only slowly decrease over the integration in many Steiner trees.
- **Threshold**: The threshold $\tau \in (0, 1]$ determines the robustness/explorativeness of the result. The explorativeness increases (and robustness decreases) for lower τ, and *vice versa* for higher τ.
- **Trees**: Number of individual Steiner trees to be computed for the combination step.

1.9.2 Drug ranking algorithms

Closeness centrality

Closeness centrality (CC) (11) is one of several centrality measurements. These are used in network analysis to derive scores that translate to the importance of the nodes in the specific network. Closeness, in particular, is derived by calculating the sum over all shortest paths between the specific node n and any other o to the power of the negative one:

$$CC(n) = \frac{1}{\sum_o d(n,o)}$$

A low CC is then used as an indicator for possible candidates that target the module.

TrustRank

TrustRank(10) is derived from the well known PageRank algorithm (13). Its original purpose is to filter spam-sites that are not highly supported by other pages. This is achieved by assigning weights based on in- and out-going edges to favor highly and punish weakly supported nodes. In theory, these weights are iteratively propagated through the network until the assigned scores or ranks stabilize. In a biological application, the ranked websites are drugs, and the most trusted ones are drugs that have high chances of targeting a large number of seeds.
Parameters:

- **Damping factor**: The damping factor for the network propagation step known from other PageRank implementations.

1.9.3 Validation algorithms

To measure significance of MI and DP results, NeDRex-Web provides permutation-based empirical p-value computation. Drugs known to be associated with the overall context of the search, i.e., mostly the queried disease that is under study, are needed for the validation, as those are used as reference to determine the empirical p-values. If the seed input was derived through the search for diseases by adding all associated genes using the NeDRex-Web suggestion function, this disease is saved and drugs are extracted using the drug-disorder indication edges. Through literature search or expert knowledge, this drug list may be manually extended or curated to provide a set representative for the context. For both algorithm groups, two p-values are derived. For MI results, these are the "empirical p-value" and "empirical (precision-based) p-value"; for DP results, the "empirical DCG-based p-value" and "empirical p-value without considering ranks". They are executed using the NeDRexAPI, and more information about the different validation methods is provided in the NeDRex publication (2).

Module validation

For the calculation of an empirical P-value, n random modules with the same number (c) and combined size (s) of connected components are generated. For this, first c random nodes are selected as starting points for each connected component, assuring that those are not first-degree neighbors to any other starting point or to their neighbors (i.e., not first- or second-degree neighbors). Then, for each connected component, random first-degree neighbors are picked iteratively and added to the connected component set, except if they would create connections to other connected components (either directly or through shared neighbors). This expansion continues until either the total module size reaches s or no more nodes can be added while maintaining the separation of components (with a limit of 20 unsuccessful iterations as a failsafe). This process is repeated n times to create n random modules. The lists of drugs targeting the result modules and each random module are determined and used to generate the following p-values:

- **Empirical p-value**: This value is determined by counting the occurrences of random modules that directly link to a higher number of reference drugs in comparison to the result module. This number of occurrences is then divided by the number of permutations (with a pseudo count of 1) to derive a p-value.

- **Empirical (precision-based) p-value**: This value is derived in the same way as the normal empirical p-value, but the definition of "better performing module" changes. Precision-based means that the module with the better precision with respect to connected drugs is determined as "better". This is calculated by dividing the number of reference drugs targeting the module by the number of all drugs targeting the module.

Drug Ranking Validation

For empirical p-value calculation for drug prioritization results, n randomly from the whole database selected and randomly ranked drug lists are derived matching the length of the result list. The random ones and the observed (result) drug lists are then scored using the reference list. It is evaluated in how many instances a random list performs better than the observed one. The performance measures are:

- **Empirical DCG-based p-value**: Here, the Discounted Cumulative Gain (14) is used to evaluate how good the ranking of the drug list is. With this function, drugs from the unranked reference list that occur at a smaller rank, thus high up in the drug list that is scored improve the score more than worse ranked occurrences. The cumulative score of the observed result and the random ones is then compared to derive the number of random results that perform better than the algorithmic result. Finally, the number of occurrences performing better than the observed result (with a pseudo count of 1) is divided by n to derive an empirical p-value.
- **Empirical p-value without considering ranks**: This measure only compares the number of reference drug occurrences in the observed ranking and the random rankings. The number of random drug lists with more reference drugs is counted to derive the p-value as before.

1.10 Runtime and Scaling

To evaluate runtime and scaling behavior differences between algorithms, we followed a standardized methodology using the API. To ensure realistic test cases, diseases from the database were categorized into bins (10, 25, 50, 100, and 250) based on their number of associated genes, allowing for a 15% deviation. Five diseases were randomly selected from each bin and their gene lists were saved as seed sets. These seed lists served as input for both module identification algorithms (DIAMOnD, MuST, KPM, DOMINO, ROBUST) and drug ranking algorithms (TrustRank, Closeness Centrality). Each algorithm combination was executed sequentially, with execution times measured and logged. The results are visualized in line graphs with error areas representing standard deviation across examples, shown in Figures S1 and S2, respectively. While KPM, DOMINO, and DIAMOnD demonstrate constant runtime independent of seed set size, ROBUST, Closeness Centrality, and TrustRank show slight input size dependency. In contrast, MuST exhibits apparent polynomial scaling with respect to input seed count. Code and data for reproducibility are available at: `https://github.com/repotrial/nedrex-web-runtime-scaling`.

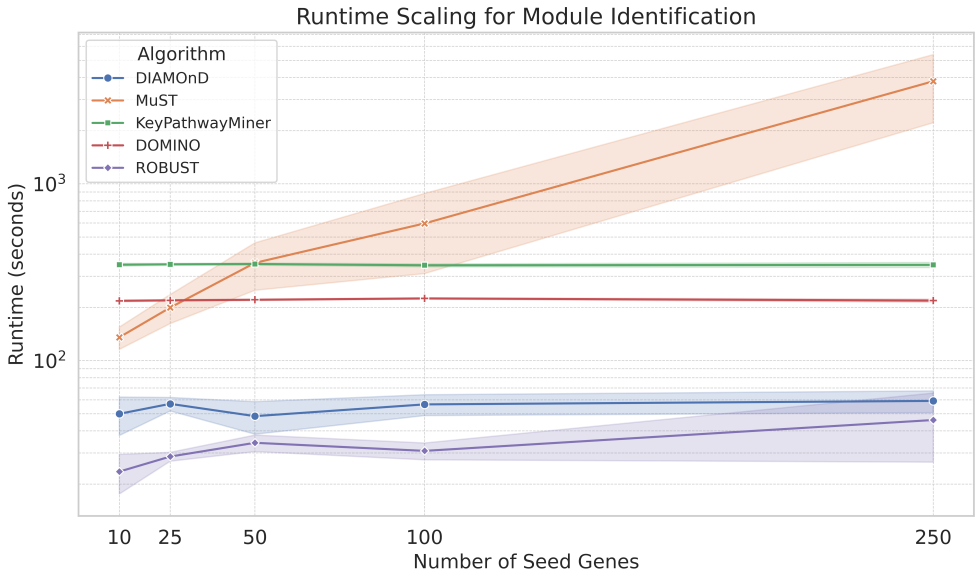

Figure S1: Runtime evaluation of module identification algorithms. KeyPathwayMiner and DOMINO demonstrate constant runtime regardless of seed count. While DIAMOnD hints to general seed count independence, some runtime variation is observed. ROBUST executes rapidly overall but exhibits linear scaling with input gene count. MuST, initially among the slower algorithms, displays polynomial runtime scaling with input size, becoming notably the slowest algorithm for seed sets exceeding 50 genes.

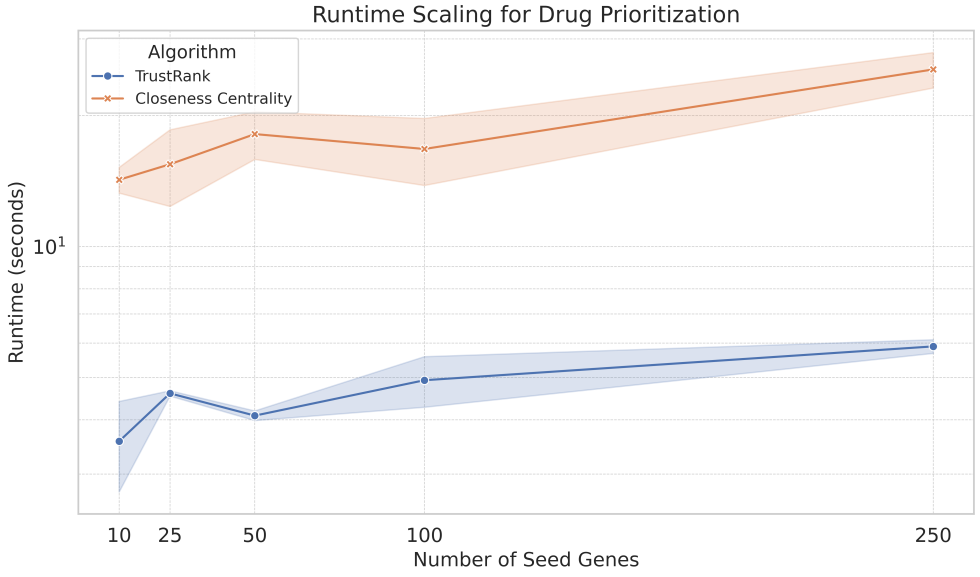

Figure S2: Runtime evaluation of drug prioritization algorithms. TrustRank and Closeness Centrality exhibit similar runtime scaling patterns with respect to input seed set size. TrustRank consistently performs approximately an order of magnitude faster than Closeness Centrality.

OS	Version	Chrome	Firefox	Safari	Edge
Linux	Mint 22 Cinnamon	127.0.6533.119	129.0	n/a	n/a
MacOS	Sonoma 14.5	127.0.6533.119	(123.0.1)	17.5	n/a
Windows	10/11	127.0.6533.120	192.0.1	n/a	127.0.2651.98

Table S1: NeDRex-Web tested browser compatibility table. Versions in brackets mean that the page is generally usable but minor unintended behaviour might be experience.

1.11 Additional Material

Additional information on NeDRex-Web is provided in form of a documentation[1] and tutorial videos[2], demostrating the different exploration paths on simple example cases. NeDRex-Web is compatible with four most popular browsers and tested under three operating systems (Table S1).

2 Data & Attributes

In the following an overview about the five node and eleven edge types in the NeDRex-Web knowledge graph will be given. This includes the source databases and attributes of entries. The source database versions used to produce presented results are listed in Table S4

2.1 Nodes

An overview about the integrated node types and sources is given in Table S2.

Table S2: NeDRex-Web comprises 439,712 nodes of five different types.

Category	**Node Types**				
	Disorder	***Drug***	***Gene***	***Pathway***	***Protein***
Entries	24,254	15,235	193,456	2,679	204,088
Sources	MONDO	DrugBank	NCBI gene	Reactome	UniProt
ID space	MONDO	DrugBank	Entrez	Reactome	UniProt
ID example	mondo. 0009061	drugbank. DB09038	entrez.6524	reactome. R-HSA-77111	uniprot. F2WII1
Attributes	'ID' 'display name' 'ICD-10' 'synonyms' 'description'	'ID' 'display name' 'category' 'drug group' 'inchi' 'cas number' 'mol. formula' 'IUPAC name' 'SMILES' 'sequences' 'synonyms' 'indication' 'description'	'ID' 'symbol' 'alt. symbols' 'chromosome' 'display name' 'gene type' 'genomic location' 'synonyms' 'description'	'ID' 'display name'	'ID' 'display name' 'gene name' 'Review Status' 'synonyms' 'sequence' 'comments'

[1] https://docs.google.com/document/d/1BGp0wovJk_ERonojc6s9XQ7W2sV_Bgb-1jtjI7yj2bg/
[2] https://www.youtube.com/watch?v=NDx2YVN7TpQ&list=PLY5U8UsdXZRncOAO5f9QjCMZSbO65WKfV

2.2 Edges

NeDRex-Web imports eight edge types from NeDRexDB. Three edge types are *inferred* from applying the protein to gene mapping provided through the protein - [encoded by] - gene edge on the according edge types. An overview over all edge types, the corresponding number of entries and their sources are combined in Table S3.

- **Disorder** - [*is comorbid with*] - **Disorder**: This edge provides information about co-occurrence of diseases in the same individual based of data provided by the Estonian Biobank, resulting in 4,414,177 comorbidity connections.
- **Disorder** - [*is subtype of*] - **Disorder**: This edge provides disease hierarchy information all integrated from MONDO (15). This directed acyclic ontology graph contains 36,529 connections.
- **Drug** - [*has contraindication*] - **Disorder**: DrugCentral (16) provides information about strong adverse side-effects of drugs applied in specific diseases. 13,809 contra-indications are imported into NeDRexDB.
- **Drug** - [*has indication*] - **Disorder**: The indication for drugs having a positive effect on a disease is encoded in 17,672 indication edges. Those are combined from DrugCentral (16) and the Comparative Toxicogenomics Database (CTD) (17).
- **Drug** - [*has target*] - **Gene**: This connection describes the genetic target of drugs. Normally this information is directed towards proteins, because drugs mostly bind to proteins to have some effect on patients system. Using the protein to drug mapping given by the protein - [encoded by] - gene edge, this edge is created on bases of the original drug - [has target] - protein data. This preprocessing step produces 29,866 unique edges and summarizes the actions of the protein targets.
- **Drug** - [*has target*] - **Protein**: The information about what is a drugs genetic target is essential in several steps of NeDRex-Web. The information is gathered from DrugCentral (16) and DrugBank (18) and combines to 30,088 unique entries. Information about the action of the drug on the protein is included.
- **Gene** - [*associated with*] - **Disorder**: The underlying genetic mechanism of a disease can be partially described by its known associations to genes. By integrating DisGeNET (19) and Online Mendelian Inheritance in Man (OMIM) (20) 33,304 association edges are established. A score provides the strength of the association.
- **Gene** - [*interacts with*] - **Gene**: The gene based interactome is a translated version of the normal protein-based interactome by using the protein to gene mapping edge protein - [encoded by] - gene. This process creates 1,382,888 unique edges and collects the entries for the attribute list 'evidence type' [predicted, orthologue, experimental] and conditions under which the interaction is expected including 'tissue', 'brain tissue', development stage', 'joint tissue' and 'subcellular location'.
- **Protein** - [*associated with*] - **Disorder**: This edge is created based on the gene - [associated with] - disorder edges using the protein - [encoded by] - gene edge to map between genes and proteins. Association scores are transferred to the 63,877 derived edges.
- **Protein** - [*encoded by*] - **Gene**: Information about the origin gene of a protein is extracted from UniProt (21) generating 30,389 mapping entries.
- **Protein** - [*in*] - **Pathway**: The pathways that proteins plays a role in are derived from Reactome (22). 126,564 edges are this way created in the NeDRexDB.
- **Protein** - [*interacts with*] - **Protein**: The interactome is essential for almost all analyses done on NeDRex-Web. 2,769,759 protein - protein interaction (PPI) edges are a combination of the integration of BioGRID (23), Inactive Ingredients Database (IID) (24) and IntAct (25). The source databases are stored as attributes alongside the annotated methods to determine the interaction (e.g biochemical) and the list of evidence types (predicted, experimental, orthologoues). A fourth resource the Human Protein Atlas (HPA) (26) is integrated to allow the construction of condition specific interactomes.

The condition classes are 'tissues', 'brain tissues', 'joint tissues', development stages' and 'subcellular locations'.

Table S3: A summary table containing all 12 edge types in the NeDRex-Web knowledge graph.

Edge Type	**Entries**	**Sources**
Disorder - [is comorbid with] - Disorder	4,414,177	Estonian Biobank
Disorder - [is subtype of] - Disorder	36,529	MONDO
Drug - [has contraindication] - Disorder	13,809	DrugCentral
Drug - [has indication] - Disorder	17,672	DrugCentral, CTD
Drug - [has target] - Gene	29,866	*inferred*
Drug - [has target] - Protein	30,088	DrugCentral, DrugBank
Gene - [associated with] - Disorder	33,304	DisGeNET, OMIM
Gene - [interacts with] - Gene	1,545,159	*inferred*
Protein - [associated with] - Disorder	63,877	*inferred*
Protein - [encoded by] - Gene	30,389	UniProt
Protein - [in] - Pathway	126,564	Reactome
Protein - [interacts with] - Protein	2,769,759	BioGRID, IID, IntAct, (HPA)
Total	9,111,193	-

3 Use cases

3.1 Drug Repurposing for Cystic Fibrosis

For the first show case, cystic fibrosis (OMIM 219700, MONDO 0009061) was investigated for potential drug repurposing candidates. The seed list was generated through NeDRex-Web using all cystic fibrosis associated genes. A visualization of the result network can be found in Figure S3.
For module identification, DIAMOnD was employed with the following parameters:

- P-value threshold: 10^{-35}
- Rest: default

TrustRank was selected for drug ranking with the following configuration:

- Approved drugs only: disabled
- Rest: default

Source	Version
BioGRID (23)	v4.4.236
CTD (17)	release July 2024
DisGeNET (19)	v7.0
DrugCentral (16)	2023-11-10
DrugBank (18)	v5.1
HPA (26)	v23
IID (24)	2021-09
IntAct (25)	2023-05
MONDO (15)	release 2024-08-06
NCBI gene (27)	state 2024-08-15
OMIM (20)	release 2024-08-14
Reactome (22)	v89
UniProt (21)	release 2024_07

Table S4: All source database integrated in NeDRexDB and used to construct the NeDRex-Web knowledge graph.

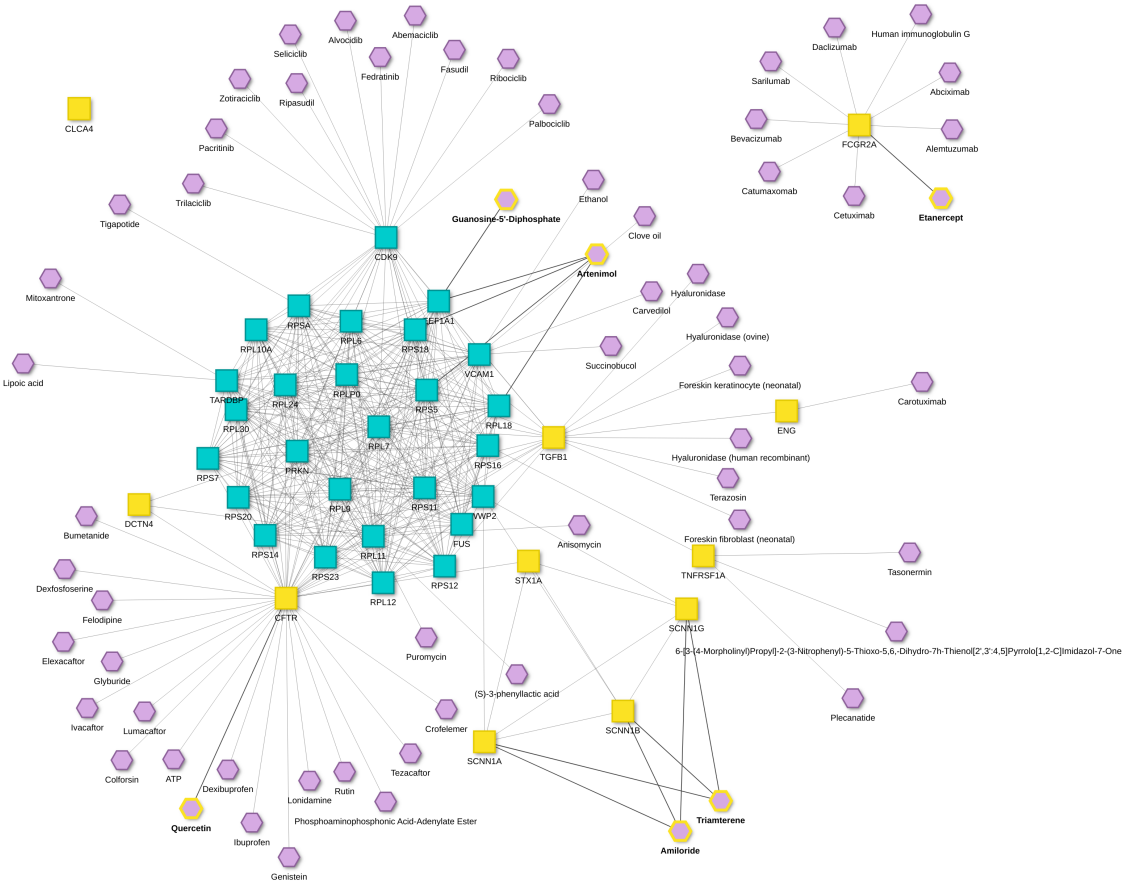

Figure S3: Full-sized version of Figure 3A - The resulting subnetwork from the disease module and drug repurposing pipeline performed for Cystic Fibrosis (CF). Yellow nodes depict the seeds associated with the CF disease in the NeDRexDB. Turquoise nodes are genes incorporated by the DIAMoND module (P-value threshold 10^{-35}), and purple nodes represent top drugs targeting all module genes after TrustRank drug prioritization. Drugs highlighted with a yellow border are drugs identified to have shown potential or might be candidates for use in CF treatment, namely amiloride, artenimol, etanercept, guanosine-5-diphosphate, quercetin, and triamterene.

3.2 Exploratory analysis of the positive outcomes of SGLT2 inhibitors in Heart Failure

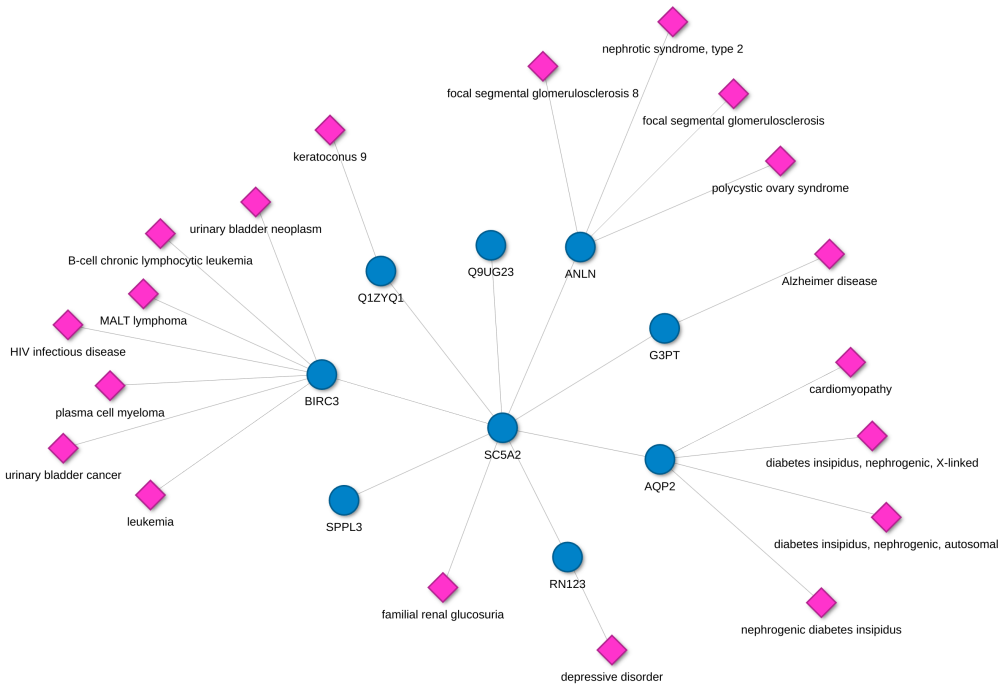

Figure S4: Full-sized version of Figure 3B - First neighbour PPI network of the *SGLT2* co-transporter (*SC5A2*, UniProt P31639) and associated disease phenotypes. On the right, Aquaporin 2 (*AQP2*) appeared directly connected to *SC5A2* and associated to cardiomyopathy. These results, supported also by previous research, suggest *AQP2* as a potential mechanism that could explain the positive outcomes of *SGLT2* inhibitors in heart failure.

3.3 Drug prioritization for the heterogeneous brain lesions in patients with progressive multiple sclerosis

This exploratory study was conducted on a now updated version of the integrated interactome, thus the exact networks might differ in the latest version of NeDRex-Web.

Multiple sclerosis (MS) is a chronic inflammatory and neurodegenerative disease of the central nervous system (CNS) characterized by the formation of focal lesions (28). In the last decades, a rapid increase in treatment options has been developed (29). Despite the success of disease modifying therapies (DMTs) in the early disease phase, the compartmentalized chronic inflammation and neurodegeneration in progressive MS (PMS) are not prevented (30). In PMS, new lesions are rare, but there are still infiltrating immune cells inducing inflammatory demyelinating lesions (31), some of which can spontaneous repair (RL), become chronic active (CA) or subsequently become inactive (IL). We have established the PMS transcriptome of these different lesions (n=100) in the brain of ten patients with PMS versus five non-neurological diseased controls (32, 33). With NeDRex-Web, we used TrustRank (Figures S5.A, S6) to explore general MS treatment options based on expressed genes (FDR<0.05) shared among all the different lesion types. The set of these 264 expressed genes mapped to brain tissue specific PPIs is used as the starting point of the analysis. A gene cluster from the generated interactome contained two approved MS treatments, natalizumab and alemtuzumab, ranked 36 and 53 respecively (Figure S5A). Both lead to immune suppression e.g. natalizumab target the integrin (*ITGA4*) on lymphocytes to reduce infiltration into the brain tissue (Figure S5A). The MS susceptibility gene *HLA-DRB5* (34) and the key chemokine (*CXCL12*) in MS glia immunity (35) were also present, where plerixafor (rank 30) an inhibitor of the *CXCL12* receptor (*CXCR4*) and tinzaparin (rank

28) also targeting the *ITGA4* were suggested (Figure S5A).

Major differences exist between MS patients, especially those that stay in the early relapsing-remitting phase contra those that convert to the progressive phase. Adapting treatment to the progressive phase is critical for patient outcomes. In PMS the number of CA lesions are increased and associated with poor clinical prognosis (36). Therefore with NeDRex-Web, we wanted to identify candidates for PMS specific targets (n=294) unique for CA (FDR<0.001) mapped to brain tissue-specific PPIs (Figure S5B, S7). One of two CA-specific connected gene-drug components was dominated by Eph receptors (*EPHA6, EPHA5, EPHA7, EPHB1*), and the drug suggestions (e.g. bosutinib, sorafenib, erlotinib, pazopanib, ranked 7, 18, 19, 21) were also recommended in an astrocyte-microglia gene-drug module of the common MS lesion seeds (Figure S6B) indicating a synergic effect of same drugs on multiple MS pathological events (Figure S5B). Eph/ephrin bidirectional signaling regulate axon guidance through contact repulsion (37) and inducing axonal growth cone collapse in injured tissue (38). Besides their direct role in inhibition of regeneration by repulsive mechanisms, the Eph/ephrins can also regulate astrocyte function by supporting scar tissue formation (39). Thereby blocking of the Eph/ephrin signaling in MS lesions may play a central role in determining repairing outcomes and can be the direction for regeneration-based therapies of CA in PMS. To sum up, with NeDRex-Web we re-discovered already applied drugs in the MS clinic, but moreover the tool also revealed multiple new candidates for drug repurposing. The idea of using multiple drugs acting in a synergistic manner on the same and/or multiple genes expressed in all types of white matter lesions or even specifically for patients with CA lesions may increase treatment efficacy in PMS.

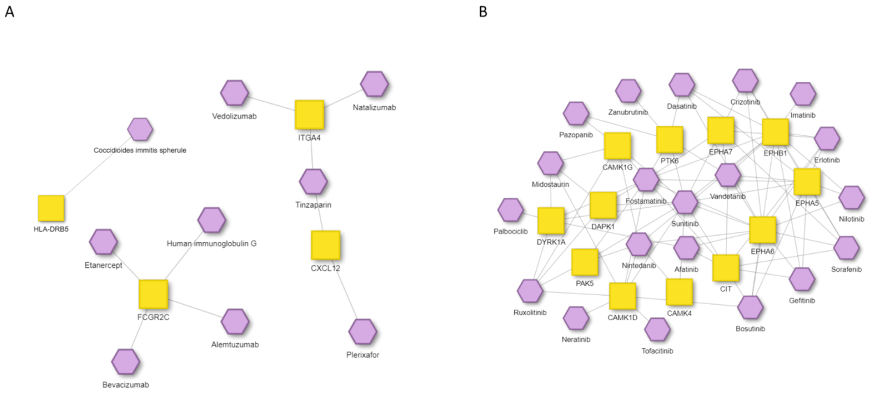

Figure S5: Selected gene clusters of **(A)** seeds(n=264, FDR<0.05) shared among all white matter lesions types in patients with progressive multiple sclerosis. Additionally, four disease modules were produced of the common seeds (Figure S6). **(B)** Seeds (n=294, FDR<0.001) uniquely dysregulated in chronic active lesions produced two connected components (B and Fig S7) in the interactome of patients with progressive multiple sclerosis.

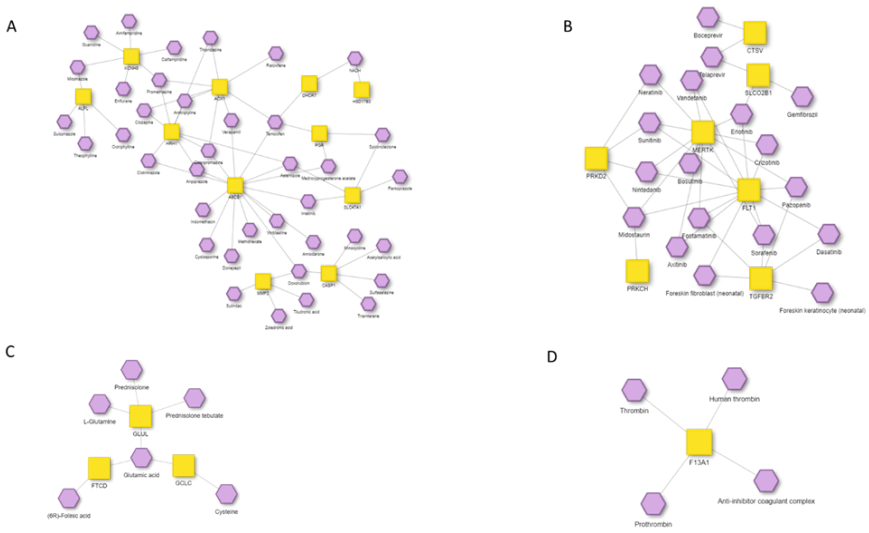

Figure S6: This depicts the connected components in the drug-target network for seeds shared among all white matter lesions types (n=264, FDR<0.05 compared to non-neurological diseased controls). The drugs were identified and ranked with TrustRank. **(A)** The biggest component contained multiple seeds reflecting the pathological consequences of demyelination (*DHCR7*, *HSD17B3*, *MMP2*) and neuronal/axonal ion imbalance (*KCNH8*, *ABCB1*, *SLC47A1*). **(B)** A second component contained known activated receptors in MS representing the dynamic cross-talk between astrocytes and microglia: (i) the phagocytic tyrosine kinase receptor (*MERTK*) on microglia important for myelin clearance activated by transforming growth factor β (*TGF-β*) (40), (ii) the *TGF-β* receptor II (*TGRBR2*), an important glia immune activator (41, 42) and (iii) the FMS-related tyrosine kinase 1 (*FLT-1/VEGFR-1*) important for proliferation and scar formation in astrocyte controlled by microglia (43). Several tyrosine kinase inhibitors are suggested (e.g. bosutinib, sorafenib, erlotinib, pazopanib). However, whether this MS activiated microglia-astrocyte signaling should be activated or inhibited is unclear. **(C)** In another connected group, seeds are coding for enzymes involved in glutamatergic pathways (*FTCD*, *GCLC*, *GLUL*). Glutamate toxicity has long been suggested to contribute to MS disease process, and several compounds targeting the glutamatergic system in animal models of MS have been studied, however clinical trials have not been successful (44), and the drug suggestions are also the substrates of the dysregulated enzymes. **(D)** Lastly, four drug suggestions (rank 89, 90, 91, 97) targeting the coagulation factor XIII, A1 polypeptide (*F13A1*), is a well known CAM gene enriched in microglia in the MS brain (45).

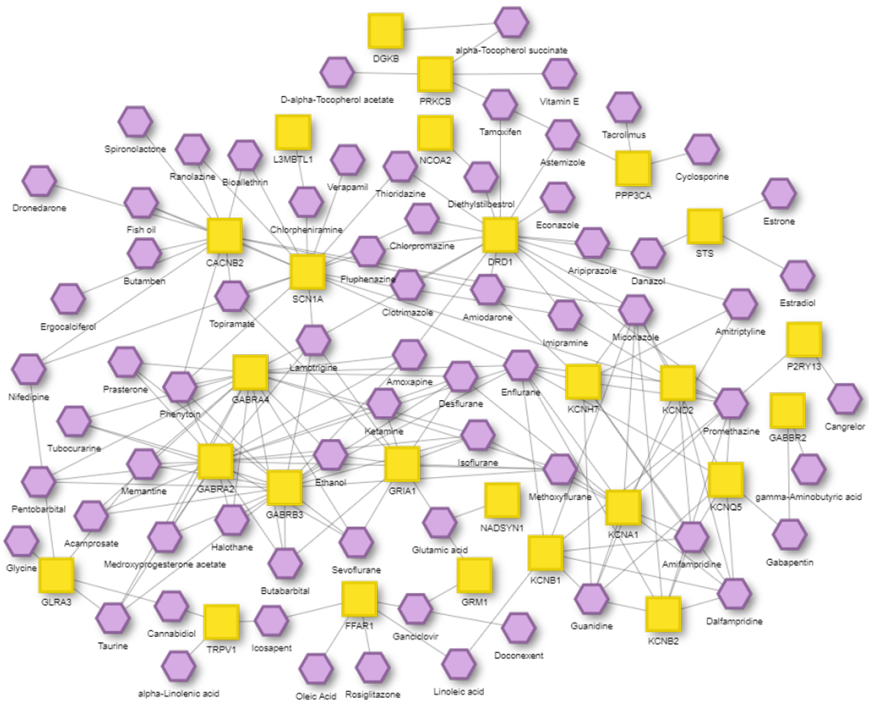

Figure S7: The biggest connected component in the gene-drug network of seeds uniquely expressed in the chronic active lesion type (CA) (n=294, FDR<0.001 compared to non-neurological diseased controls) mapped to the brain tissue-specific experimentally validated PPI network. The drugs were ranked with TrustRank (find parameters below). The subnetwork is dominated by ionotropic glutamate receptors (*GRIA1*), metabotropic glutamate receptors (*GRM1*), modulators of glutamate (*STS*) with multiple suggestive drugs e.g. treatment of major depressive disorders (amoxapine, amitriptyline). Studies have also shown the role of glutamate excitotoxicity in the progression of MS (46, 47). Glutamate is the major excitatory neurotransmitter of the CNS, and it has a central role in a complex cellular interactome. However, an excess release of glutamate leads to the activation of ionotropic and metabotropic receptors, resulting in accumulation of toxic cytoplasmic Ca^{2+} and cell death. The presence of Na^{+} channel (*SCN1A*) in the CA seeds may also reflect the imbalance of ion homeostasis. On contrary, this CA gene-drug component also contained seeds that may prevent excitotoxic neuronal damage as both *GABAA* receptors (*GABRA2, GABRA3, GABRB3*) are though to inhibit depolarization-evoked Ca^{2+} influx, and increased expression of voltage-gated K^{+} channels (*KCNA1, KCNB1, KCNB2, KCND2, KCNQ5, KCNH7*) are believed to limit neuronal excitability (48). Additionally, the microglia (*P2RY13*), fatty acid signaling (*FFAR1*) and candidates targets in Alzheimer Disease (*L3MBTL1* involved in proteasomal digestion) and the dopamine receptor (*DRD1*) are also present in the seeds. Therefore, understanding how the distinct channels and receptors contribute to the MS pathophysiology and excitotoxicity in the CA is still a challenge before therapeutic strategies. However using NeDRex-Web, the importance of excitotoxicity in the MS disease progression was emphasized and identification of functional gene-drug clusters for further investigation.

Parameters and Material

In both drug ranking cases, common shared genes (Table S5) and CA specific genes (Table S6), TrustRank was employed using the following configuration:

- TopX drugs: 100 (default)
- Only experimentally validated interactions: enabled (default)
- Add gene interactions: disabled
- Tissue-specific interactions: Brain
- Direct drugs only: enabled (default)
- Filter Element Drugs: enabled (default)
- damping factor: 0.85 (default)

2	4240	7048	1781	9332	9920	23143	22837	11098	55603	340485	728723
230	9124	7706	1119	8477	7168	84327	55503	23041	54502	221178	116159
834	5793	4864	9882	9987	8418	10457	51059	57664	58475	253827	196383
837	3269	3127	1601	6387	3676	25859	80704	10841	84899	127435	337873
316	5241	3992	3426	1522	9829	55244	57211	26146	54622	374887	143903
730	8832	8797	2752	8740	2035	51176	10018	23533	80833	115361	646755
463	4094	5763	3603	6693	3481	51365	25816	10971	51339	285668	153328
873	5937	8857	3487	3021	2294	10165	92610	10110	116362	283755	100505495
862	1832	4211	1193	6819	9452	84947	51735	80000	221395	197259	100463486
390	2683	1717	1806	8870	1356	91351	10602	63901	126669	153830	100130562
947	1107	1295	2313	4481	11309	25937	54510	51705	259236	116441	101928470
871	5243	3293	5420	4360	23231	55340	53827	28965	255488	131096	101928176
249	4053	9096	2729	2321	51363	84689	55303	84952	171586	338596	100463498
840	6319	3500	2162	2023	89857	27296	54855	58480	439949	201626	100288805
951	4162	9935	9411	3082	25793	23321	57091	10742	286467	284293	100288332
684	9371	3937	8365	3638	10491	64167	23266	54718	728430	400550	100271871
113	1281	5583	5239	7052	25865	10894	10461	26112	219855	162073	100420358
4313	9103	7253	1634	1803	23345	79745	64231	84285	219833	283316	100131320
2078	4773	5523	1278	2296	23208	84735	51703	23224	122953	387751	101928651
7705	7681	8829	2114	1515	51296	89846	26585	23180	171024	285735	100128890
8542	2635	5787	6697	3075	91607	90102	26157	80177	138639	170692	101234262
4121	4647	5151	6778	7439	55803	10788	93663	56895	155038	283349	104266957

Table S5: Table of 264 common MS genes (EntrezIDs)

67	2119	5577	9353	11197	27086	55342	78996	116931	219736	440482
412	2138	5579	9400	11234	27115	55384	79085	120114	222008	641467
670	2259	5688	9446	22854	27315	55531	79145	121260	253152	642236
783	2555	5693	9462	22866	27346	55729	79632	121456	255167	654429
814	2557	5753	9480	23057	27445	55755	79667	125228	261729	654790
1000	2562	5800	9516	23085	29100	55809	79698	127254	283392	677772
1004	2762	5936	9547	23092	29948	56479	79815	127700	283480	677775
1005	2774	6096	9568	23105	29951	56666	79968	134728	284215	729993
1007	2864	6323	9569	23132	50485	56704	80059	134829	284252	780813
1008	2890	6547	9644	23145	50939	57118	80063	138649	285220	100127950
1040	2911	6882	9840	23178	51334	57144	80195	139599	285484	100129696
1123	2967	7111	9851	23228	51361	57168	80742	140767	285905	100130621
1318	3131	7162	9961	23236	51399	57172	81491	145624	285958	100216544
1497	3208	7360	10009	23281	53335	57451	81669	145853	286046	100420155
1593	3736	7415	10129	23329	53829	57484	81849	146713	286826	100422558
1602	3745	7442	10311	23518	53905	57574	83746	148479	317772	100506244
1607	3751	7458	10412	23543	54212	57620	84301	149297	319085	100507043
1612	4017	7761	10499	23551	54477	57633	84451	150084	343990	100507381
1730	4166	7932	10552	23767	54566	57718	84679	151126	346157	100873986
1804	4208	8001	10721	24141	54682	57821	84867	151258	347730	100874113
1812	4338	8395	10777	25759	54715	57835	85377	151354	348487	101410534
1859	4628	8554	10793	25937	54952	60437	90134	151516	349152	101927193
1946	4850	8708	10824	26013	55061	63982	90417	152006	353116	101927972
2044	5121	9079	10891	26050	55154	64061	92715	157574	359822	105370424
2045	5210	9228	10936	26097	55191	64388	93986	157807	387978	
2047	5318	9312	10953	26586	55223	65258	113263	163223	401190	
2115	5530	9315	11113	27020	55257	65982	114880	171523	439921	

Table S6: Table of 294 CA lesion type specific genes (EntrezIDs)

References

1. Hütter, C. V., Sin, C., Müller, F., and Menche, J. (2, 2022) Network cartographs for interpretable visualizations. *Nature Computational Science 2022 2:2,* **2**, 84–89.

2. Sadegh, S., Skelton, J., Anastasi, E., Bernett, J., Blumenthal, D. B., Galindez, G., Salgado-Albarrán, M., Lazareva, O., Flanagan, K., Cockell, S., Nogales, C., Casas, A. I., Schmidt, H. H., Baumbach, J., Wipat, A., and Kacprowski, T. (11, 2021) Network medicine for disease module identification and drug repurposing with the NeDRex platform. *Nat. Commun.,* **12**, 1–12 [PubMed:34824199] [PubMed Central:PMC8617287] [doi:10.1038/s41467-021-27138-2].

3. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (11, 2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.,* **13**, 2498 [PubMed:14597658] [PubMed Central:PMC403769] [doi:10.1101/gr.1239303].

4. Lazareva, O., Canzar, S., Yuan, K., Baumbach, J., Blumenthal, D. B., Tieri, P., Kacprowski, T., and List, M. (8, 2021) BiCoN: network-constrained biclustering of patients and omics data. *Bioinformatics,* **37**, 2398–2404 [PubMed:33367514] [doi:10.1093/bioinformatics/btaa1076].

5. Ghiassian, S. D., Menche, J., and Barabási, A. L. (4, 2015) A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome. *PLoS Comput. Biol.,* **11**, e1004120 [PubMed:25853560] [PubMed Central:PMC4390154] [doi:10.1371/journal.pcbi.1004120].

6. Levi, H., Elkon, R., and Shamir, R. (1, 2021) DOMINO: a network-based active module identification algorithm with reduced rate of false calls. *Mol. Sys. Biol.,* **17**, e9593 [PubMed:33471440] [PubMed Central:PMC7816759] [doi:10.15252/msb.20209593].

7. List, M., Alcaraz, N., Dissing-Hansen, M., Ditzel, H. J., Mollenhauer, J., and Baumbach, J. (7, 2016) KeyPathwayMinerWeb: online multi-omics network enrichment. *Nucleic Acids Res.,* **44**, W98–W104 [PubMed:27150809] [PubMed Central:PMC4987922] [doi:10.1093/nar/gkw373].

8. Bernett, J., Krupke, D., Sadegh, S., Baumbach, J., ndor Fekete, S. P., Kacprowski, T., List, M., and Blumenthal, D. B. (3, 2022) Robust disease module mining via enumeration of diverse prize-collecting Steiner trees. *Bioinformatics,* **38**, 1600–1606 [PubMed:34984440] [doi:10.1093/bioinformatics/btab876].

9. Sadegh, S., Matschinske, J., Blumenthal, D. B., Galindez, G., Kacprowski, T., List, M., Nasirigerdeh, R., Oubounyt, M., Pichlmair, A., Rose, T. D., Salgado-Albarrán, M., Späth, J., Stukalov, A., Wenke, N. K., Yuan, K., Pauling, J. K., and Baumbach, J. (7, 2020) Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *Nat. Commun.,* **11**, 1–9 [PubMed:32665542] [PubMed Central:PMC7360763] [doi:10.1038/s41467-020-17189-2].

10. Gyöngyi, Z., Garcia-Molina, H., and Pedersen, J. O. (2004) Combating Web Spam with TrustRank. *Proc. 2004 VLDB Conf.,* pp. 576–587.

11. Bavelas, A. (6, 1950) Communication Patterns in Task-Oriented Groups. *J. Acoust. Soc,* **22**, 725.

12. Zarin, D. A., Tse, T., Williams, R. J., and Carr, S. (11, 2016) Trial Reporting in ClinicalTrials.gov - The Final Rule. *N Engl J Med,* **375**, 1998–2004 [PubMed:27635471] [PubMed Central:PMC5225905] [doi:10.1056/NEJMsr1611785].

13. Page, L., Brin, S., Motwani, R., and Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. (1998).

14. Järvelin, K. and Kekäläinen, J. (October, 2002) Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.,* **20**(4), 422–446.

15. Mungall, C. J., McMurry, J. A., Kohler, S., Balhoff, J. P., Borromeo, C., Brush, M., Carbon, S., Conlin, T., Dunn, N., Engelstad, M., Foster, E., Gourdine, J. P., Jacobsen, J. O., Keith, D., Laraway, B., Lewis, S. E., Xuan, J. N., Shefchek, K., Vasilevsky, N., Yuan, Z., Washington, N., Hochheiser, H., Groza, T., Smedley, D., Robinson, P. N., and Haendel, M. A. (1, 2017) The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.,* **45**, D712.

16. Avram, S., Bologa, C. G., Holmes, J., Bocci, G., Wilson, T. B., Nguyen, D. T., Curpan, R., Halip, L., Bora, A., Yang, J. J., Knockel, J., Sirimulla, S., Ursu, O., and Oprea, T. I. (1, 2021) DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Res.,* **49**, D1160–D1169 [PubMed:33151287] [PubMed Central:PMC7779058] [doi:10.1093/nar/gkaa997].

17. Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegers, J., Wiegers, T. C., and Mattingly, C. J. (1, 2021) Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Res.,* **49**, D1138–D1143 [PubMed:33068428] [PubMed Central:PMC7779006] [doi:10.1093/nar/gkaa891].

18. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., and Wilson, M. (1, 2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.,* **46**, D1074–D1082 [PubMed:29126136] [PubMed Central:PMC5753335] [doi:10.1093/nar/gkx1037].

19. Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., and Furlong, L. I. (1, 2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.,* **48**, D845–D855 [PubMed:31680165] [PubMed Central:PMC7145631] [doi:10.1093/nar/gkz1021].

20. Amberger, J. S., Bocchini, C. A., Scott, A. F., and Hamosh, A. (1, 2019) OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Res.,* **47**, D1038–D1043 [PubMed:30445645] [PubMed Central:PMC6323937] [doi:10.1093/nar/gky1151].

21. Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Silva, A. D., Giorgi, M. D., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., Castro, E. D., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., and Zhang, J. (1, 2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.,* **45**, D158–D169.

22. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. (1, 2020) The reactome pathway knowledgebase. *Nucleic Acids Res.,* **48**, D498–D503.

23. Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (1, 2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.,* **34**, D535–D539 [PubMed:16381927] [PubMed Central:PMC1347471] [doi:10.1093/nar/gkj109].

24. Kotlyar, M., Pastrello, C., Malik, Z., and Jurisica, I. (1, 2019) IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.,* **47**, D581–D589 [PubMed:30407591] [PubMed Central:PMC6323934] [doi:10.1093/nar/gky1037].

25. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., Del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R. C., Meldal, B., Melidoni, A. N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., Roey, K. V., Cesareni, G., and Hermjakob, H. (1, 2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.,* **42**, D358–D363 [PubMed:24234451] [PubMed Central:PMC3965093] [doi:10.1093/nar/gkt1115].

26. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Åsa Sivertsson, Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., Feilitzen, K. V., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., Heijne, G. V., Nielsen, J., and Pontén, F. (1, 2015) Tissue-based map of the human proteome. *Science,* **347**.

27. Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (1, 2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.,* **35**.

28. Compston, A. and Coles, A. (2008) Multiple sclerosis. *Lancet,* **372**, 1502–1517.

29. Hauser, S. L. and Cree, B. A. (12, 2020) Treatment of Multiple Sclerosis: A Review. *Am. J. Med.,* **133**, 1380–1390.e2.

30. 't Hart, B. A., Luchicchi, A., Schenk, G. J., Killestein, J., and Geurts, J. J. (6, 2021) Multiple sclerosis and drug discovery: A work of translation. *EBioMedicine,* **68** [PubMed:34044219] [PubMed Central:PMC8245896] [doi:10.1016/j.ebiom.2021.103392].

31. Absinta, M., Lassmann, H., and Trapp, B. D. (6, 2020) Mechanisms underlying progression in multiple sclerosis. *Curr. Opin. Neurol.,* **33**, 277–285.

32. Frisch, T., Elkjaer, M. L., Reynolds, R., Michel, T. M., Kacprowski, T., Burton, M., Kruse, T. A., Thomassen, M., Baumbach, J., and Illes, Z. (8, 2020) Multiple Sclerosis Atlas: A Molecular Map of Brain Lesion Stages in Progressive Multiple Sclerosis. *Netw Syst Med.,* **3**, 122 [PubMed:32954379] [PubMed Central:PMC7500075] [doi:10.1089/nsm.2020.0006].

33. Elkjaer, M. L., Frisch, T., Reynolds, R., Kacprowski, T., Burton, M., Kruse, T. A., Thomassen, M., Baumbach, J., and Illes, Z. (12, 2019) Molecular signature of different lesion types in the brain white matter of patients with progressive multiple sclerosis. *Acta Neuropathol. Commun.,* **7**, 1–17 [PubMed:31829262] [PubMed Central:PMC6907342] [doi:10.1186/s40478-019-0855-7].

34. Lang, H. L., Jacobsen, H., Ikemizu, S., Andersson, C., Harlos, K., Madsen, L., Hjorth, P., Sondergaard, L., Svejgaard, A., Wucherpfennig, K., Stuart, D. I., Bell, J. I., Jones, E. Y., and Fugger, L. (2002) A functional and structural basis for TCR cross-reactivity in multiple sclerosis. *Nat. Immunol.,* **3**, 940–943 [PubMed:12244309] [doi:10.1038/ni835].

35. Krumbholz, M., Theil, D., Cepok, S., Hemmer, B., Kivisäkk, P., Ransohoff, R. M., Hofbauer, M., Farina, C., Derfuss, T., Hartle, C., Newcombe, J., Hohlfeld, R., and Meinl, E. (1, 2006) Chemokines in multiple sclerosis: CXCL12 and CXCL13 up-regulation is differentially linked to CNS immune cell recruitment. *Brain,* **129**, 200–211 [PubMed:16280350] [doi:10.1093/brain/awh680].

36. Lassmann, H. (3, 2018) Multiple Sclerosis Pathology. *Cold Spring Harb. Perspect. Med.,* **8**.

37. Gallarda, B. W., Bonanomi, D., Müller, D., Brown, A., Alaynick, W. A., Andrews, S. E., Lemke, G., Pfaff, S. L., and Marquardt, T. (4, 2008) Segregation of axial motor and sensory pathways via heterotypic trans-axonal signaling. *Science,* **320**, 233–236 [PubMed:18403711] [PubMed Central:PMC3158657] [doi:10.1126/science.1153758].

38. Fujita, Y., Takashima, R., Endo, S., Takai, T., and Yamashita, T. (9, 2011) The p75 receptor mediates axon growth inhibition through an association with PIR-B. *Cell Death Dis.,* **2** [PubMed:21881600] [PubMed Central:PMC3186903] [doi:10.1038/cddis.2011.85].

39. Goldshmit, Y., McLenachan, S., and Turnley, A. (9, 2006) Roles of Eph receptors and ephrins in the normal and damaged adult CNS. *Brain Res. Rev.,* **52**, 327–345 [PubMed:16774788] [doi:10.1016/j.brainresrev.2006.04.006].

40. Shen, K., Reichelt, M., Kyauk, R. V., Ngu, H., Shen, Y. A. A., Foreman, O., Modrusan, Z., Friedman, B. A., Sheng, M., and Yuen, T. J. (3, 2021) Multiple sclerosis risk gene Mertk is required for microglial activation and subsequent remyelination. *Cell Rep.,* **34**, 108835 [PubMed:33691116] [doi:10.1016/j.celrep.2021.108835].

41. Zhang, S. Z., Wang, Q. Q., Yang, Q. Q., Gu, H. Y., Yin, Y. Q., Li, Y. D., Hou, J. C., Chen, R., Sun, Q. Q., Sun, Y. F., Hu, G., and Zhou, J. W. (11, 2019) NG2 glia regulate brain innate immunity via TGF-β2/TGFBR2 axis. *BMC Med.,* **17**, 1–22.

42. Zöller, T., Schneider, A., Kleimeyer, C., Masuda, T., Potru, P. S., Pfeifer, D., Blank, T., Prinz, M., and Spittau, B. (10, 2018) Silencing of TGFβ signalling in microglia results in impaired homeostasis. *Nat. Commun.,* **9**, 1–13.

43. Rothhammer, V., Borucki, D. M., Tjon, E. C., Takenaka, M. C., Chao, C. C., Ardura-Fabregat, A., Lima, K. A. D., Gutiérrez-Vázquez, C., Hewson, P., Staszewski, O., Blain, M., Healy, L., Neziraj, T., Borio, M., Wheeler, M., Dragin, L. L., Laplaud, D. A., Antel, J., Alvarez, J. I., Prinz, M., and Quintana, F. J. (5, 2018) Microglial control of astrocytes in response to microbial metabolites. *Nature,* **557**, 724–728 [PubMed:29769726] [PubMed Central:PMC6422159] [doi:10.1038/s41586-018-0119-x].

44. Macrez, R., Stys, P. K., Vivien, D., Lipton, S. A., and Docagne, F. (9, 2016) Mechanisms of glutamate toxicity in multiple sclerosis: biomarker and therapeutic opportunities. *Lancet Neurol.,* **15**, 1089–1102 [PubMed:27571160] [doi:10.1016/S1474-4422(16)30165-X].

45. Miedema, A., Gerrits, E., Brouwer, N., Jiang, Q., Kracht, L., Meijer, M., Nutma, E., Peferoen-Baert, R., Pijnacker, A. T., Wesseling, E. M., Wijering, M. H., Gabius, H. J., Amor, S., Eggen, B. J., and Kooistra, S. M. (12, 2022) Brain macrophages acquire distinct transcriptomes in multiple sclerosis lesions and normal appearing white matter. *Acta Neuropathol. Commun.,* **10**, 1–18 [PubMed:35090578] [PubMed Central:PMC8796391] [doi:10.1186/s40478-021-01306-3].

46. Srinivasan, R., Sailasuta, N., Hurd, R., Nelson, S., and Pelletier, D. (5, 2005) Evidence of elevated glutamate in multiple sclerosis using magnetic resonance spectroscopy at 3 T. *Brain,* **128**, 1016–1025 [PubMed:15758036] [doi:10.1093/brain/awh467].

47. Azevedo, C. J., Kornak, J., Chu, P., Sampat, M., Okuda, D. T., Cree, B. A., Nelson, S. J., Hauser, S. L., and Pelletier, D. (2014) In vivo evidence of glutamate toxicity in multiple sclerosis. *Ann. Neurol.,* **76**, 269–278 [PubMed:25043416] [PubMed Central:PMC4142752] [doi:10.1002/ana.24202].

48. Boscia, F., Elkjaer, M. L., Illes, Z., and Kukley, M. (6, 2021) Altered Expression of Ion Channels in White Matter Lesions of Progressive Multiple Sclerosis: What Do We Know About Their Function?. *Front. Cell. Neurosci,* **15**, 214 [PubMed:34276310] [PubMed Central:PMC8282214] [doi:10.3389/fncel.2021.685703].