

Uncovering New Therapeutic Targets for Amyotrophic Lateral Sclerosis and Neurological Diseases Using Real-World Data

Mohammadali Alidoost¹ , Jeremy Y. Huang², Georgia Dermentzaki³, Anna S. Blazier³, Giorgio Gaglia², Timothy R. Hammond³, Francesca Frau⁴, Mary Clare McCorry⁵ , Dimitry Ofengeim³ and Jennifer L. Wilson^{1,*} 

Although attractive for relevance to real-world scenarios, real-world data (RWD) is typically used for drug repurposing and not therapeutic target discovery. Repurposing studies have identified few effective options in neurological diseases such as the rare disease, amyotrophic lateral sclerosis (ALS), which has no disease-modifying treatments available. We previously reclassified drugs by their simulated effects on proteins downstream of drug targets and observed class-level effects in the EHR, implicating the downstream protein as the source of the effect. Here, we developed a novel ALS-focused network medicine model using data from patient samples, the public domain, and consortia. With this model, we simulated drug effects on ALS and measured class effects on overall survival in retrospective EHR studies. We observed an increased but non-significant risk of death for patients taking drugs with complement system proteins downstream of their targets and experimentally validated drug effects on complement activation. We repeated this for six protein classes, three of which, including multiple chemokine receptors, were associated with a significantly increased risk for death, suggesting that targeting proteins such as CXCR5, CXCR3, chemokine signaling generally, or neuropeptide Y (NPY) could be advantageous therapeutic targets for these patients. We expanded our analysis to the neuroinflammatory condition, myasthenia gravis, and neurodegenerative disease, Parkinson's, and recovered similar effect sizes. We demonstrated the utility of network medicine for testing novel therapeutic effects using RWD and believe this approach may accelerate target discovery in neurological diseases, addressing the critical need for new therapeutic options.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Currently, the impact of real-world data (RWD) on drug-gable targets with translational impact is limited to drug-repurposing. However, we hypothesized that network models of drug-induced signaling could elevate the impact of RWD for identifying novel drug targets.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ This study addressed the feasibility of that approach by simulating the effects of novel targets and testing these effects in neurological conditions.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ We demonstrated that drugs that alter complement activation have deleterious effects on overall survival, and these results are consistent with the effects of recently FDA-approved drugs, although none of our patients were taking them.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ We believe this paradigm will change translational science in how RWD may be used to identify novel, druggable targets.

We need novel treatment identification methods for rare, difficult-to-treat diseases, especially those with insufficient experimental systems. Big data approaches and advanced algorithms have the power to capture patient data and infer new disease drivers in ways previously not possible. Many approaches suffer from limitations

that were well established a decade ago: animal models are valuable for understanding mechanisms but have limited relevance to human conditions, human epidemiology is highly relevant to human disease but is insufficient for explaining causality, and natural conditions are the most relevant and may explain causality

¹Department of Bioengineering, University of California Los Angeles, Los Angeles, California, USA; ²Precision Medicine & Computational Biology, Sanofi Research US, Cambridge, Massachusetts, USA; ³Rare & Neurologic Diseases, Sanofi Research US, Cambridge, Massachusetts, USA; ⁴Evidence Generation & Decision Sciences, Sanofi Development, Frankfurt, Germany; ⁵Scientific Relations & Initiatives, Sanofi Research US, Cambridge, Massachusetts, USA. *Correspondence: Jennifer L. Wilson (jenniferwilson@g.ucla.edu)

Received October 15, 2024; accepted March 26, 2025. doi:10.1002/cpt.3682

but are rare.¹ We hypothesized that a model-informed analysis of patient data may overcome these hurdles by predicting causality first and later testing causal predictions in relevant human disease contexts.

Retrospective analyses of electronic health records (EHRs) are attractive for difficult-to-treat conditions, including many neurological conditions, where animal and model systems have failed to deliver life-sparing treatments. However, most EHR studies are restricted to the repurposing of known druggable targets. For instance, the discovery of phosphodiesterase inhibitors as protective for Parkinson's² and amyotrophic lateral sclerosis (ALS),³ peroxisome proliferator-activated receptor (PPAR) agonists as protective in Alzheimer's Disease,⁴ and the potential of multiple drugs, diazoxide, gefitinib, paliperidone, and dimethyltryptamine, as high-quality candidates for repurposing in ALS⁵ all leverage 'omics data, network medicine, and EHRs, but they uncovered evidence for known, drugged targets. Indeed, reviews of the field emphasized that real-world data can support clinical trial emulation, sub-phenotyping, or image analysis,⁶ but the authors failed to consider the historical record as sufficient for supporting *novel, untested* drug targets. Despite the status quo, we had compelling evidence to more directly vet novel targets in EHRs. Specifically, we had learned that disease-specific network models improved the prediction of drug effects⁷ and that drug-effect predictions based on shared protein networks were detectable in EHRs.⁸ We considered that applications in neurology—especially in neuroinflammatory and neurodegenerative diseases—could be an exciting and impactful application.

Debilitating neuroinflammatory and neurodegenerative conditions without substantially effective treatments have motivated creative and compelling approaches to advance treatment options. For ALS, a rare, neuroinflammatory condition, there are seven Food and Drug Administration (FDA)-approved treatments, of which, none are disease-modifying⁹ or life-sparing. Multiple studies have identified distinct molecular disease mechanisms including the role of the innate immune system (Table S4), among other genetic drivers.^{9,10} Targeting these disrupted pathways or causal genetic drivers in the clinic has had limited efficacy and coverage due to the time to develop new medicines and initiate new trials. Since ALS is a rare disease, clinical trial cohorts often struggle to meet statistically significant thresholds, leading to increased interest in leveraging new insights from real-world data. Thus, we pursued our approach in neurology, and first considered neuroinflammatory conditions like ALS and myasthenia gravis (MG) because of the potential impact on these patients. Because neurodegenerative diseases generally lack preclinical platforms, we considered also testing our platform across a broad set of conditions and included neurodegenerative conditions like Parkinson's disease (PD).

In this study, we tested our network inference platform on multiple conditions and corroborated our model predictions with EHR investigation and experimental analysis. Intriguingly, our results synergize with other investigations of dysregulated pathways in ALS and MG; specifically, we recovered increased risks for death associated with complement system activation. Even though ALS patients were not prescribed any complement-targeted drugs, our retrospective analysis provided rationale for directly inhibiting this

pathway. We also asked whether these insights were generalizable across additional diseases with larger patient populations, such as MG and PD. In total, we built and tested six network classes with different impacts on overall patient survival. Encouragingly, our pipeline independently identified therapeutic targets currently in clinical development and suggested additional targets with stronger effects. By classifying approved therapies by their effects on downstream proteins, rather than labeled primary targets, we can leverage large-scale cohort studies to identify novel targets for ALS and other neurological diseases.

MATERIALS AND METHODS

Curating disease pathways—overview

To customize our network analysis, we curated novel disease (“phenotype”) pathways from multiple sources. Altogether, we included 195,252 gene-phenotype relationships across 1,096 unique disease pathways and 24,441 unique genes/transcripts. Further, a list of all derived pathway names, their sources, and the total number of genes is included in **Supplemental File 2** (GitHub repository). For a detailed list of pathway sources and processing for each source, and integration into our network modeling platform, please see **Supplemental Methods**.

Running ALS-PathFX on all approved drugs

Like previous versions of PathFX, the ALS version required drug-binding proteins as inputs. We used drug target information from DrugBank version 5.1.6 as inputs for ALS-PathFX analysis. We generated networks and association tables for all drugs and then searched the PathFX-generated association tables for phenotypes of interest. The association tables are standard outputs from PathFX and include all network-associated phenotypes, the multiple-hypothesis-corrected p-values for the associations, and which disease pathway proteins were discovered in the networks. All ALS-PathFX predictions, created by merging all associations tables for drugs analyzed, are included in **Supplemental File 2** (GitHub repository).

Generating network classes

After creating a merged table of all ALS-PathFX predicted drug-phenotype associations, we counted how often a network protein was used to make a drug-phenotype prediction (e.g., CXCL13 was used to predict a network association from gabapentin to the “MS CSF” and “Bulk RNAseq ALS vs Control” pathways, and C3 was used to predict a network association from gabapentin to the “Bulk RNAseq ALS vs. Control” pathway). After identifying network proteins occurring in the most drug-ALS PathFX associations, we developed new drug classes by grouping drugs that shared network proteins (e.g., CXCL13 or C3). We completed this analysis using the pandas module in Python. As we wanted to study patient data, we also filtered out experimental or unapproved drugs and counted the number of approved active ingredients with each network protein.

In a few instances, we observed several related network proteins and made multi-protein classes based on qualitative and quantitative assessment of the ranked proteins, such as in the case of the complement system proteins and chemokine proteins. Grouping drugs based on network proteins yielded groups for the “target” cohorts (“network” class). To generate comparable “comparator” cohorts, we assessed all remaining approved drugs that had network associations with the same ALS phenotypes but lacked a network protein in the association (“non-network” class). For instance, all drugs with an association with the “MS CSF” or “Bulk RNAseq ALS vs. Control” pathways would be considered in the non-CXCL13 comparator class. After generating these classes, we ranked them based on having a relatively high and similar number of approved drugs in “network” and “non-network” classes

because insufficient numbers of drugs would yield insufficient patient cohorts. Within the top of this ranking, we observed several complement inhibitors and chemokine-related proteins and grouped these into single classes. All network proteins and their presence in drug networks are contained in **Supplemental File 3** (Github repository). The code for this analysis is included in our GitHub.

Clinical data—Optum market clarity

We used Optum's Market Clarity data: Optum® Market Clarity Data is an innovative dataset of integrated multi-source medical and pharmacy claims and electronic health record (EHR) data that enables detailed evaluation and assessment of the patient journey. The Market Clarity Data Asset combines robust transactional pharmacy and medical claims data with best-of-breed EHR data. It is a fully HIPAA-compliant, statistician-certified, de-identified precision data set that uncovers hidden intelligence to turn insights into quick action. The Market Clarity Dataset links EHR data with historical, linked administrative claim data, pharmacy claims, physician claims, facility claims (with clinical information), and is inclusive of medications prescribed and administered. Clinically rich and specific data elements sourced from the EHR include lab results, vital signs and measurements, diagnoses, procedures, and information derived from unstructured clinical notes using natural language processing.

We specifically assessed patient records available through February 2023 and accessed the following data tables—the membership info for collecting date of birth, sex, race, ethnicity, and gender, when available; the RX claims table to assess filled prescriptions as a proxy for drug exposures; the diagnosis table for identifying ALS patients and other comorbid conditions; and the date of death table for collecting the month/year of death, when available.

Observational studies in the electronic health record

To measure the effects of our network-derived drug classes, we used a standard target-comparator approach using best practices in propensity score estimation and weighted treatment effects. We have used a similar approach in our own work⁸ and describe the approach in greater detail in **Supplemental Methods**.

Stem cell culture

Human iPSC astrocytes (iCell astrocytes, Fujifilm Cellular Dynamics: Madison, Wisconsin, Cat# C1037) were cultured in Matrigel-coated (Corning (Fischer Scientific): Atlanta, Georgia, 354277) 96-well plates with DMEM/F12 (Life Technologies: Chicago, Illinois, 11330057) media supplemented with heat-inactivated FBS (Gibco (Life Technologies): Chicago, Illinois, A38400-01) and N-2 supplement (Gibco 17502-048).

Stimulation and drug treatment

Three days post-plating, iCell astrocytes were treated with selective compounds (**Table S2**) at 100 nM, 1 μM, and 10 μM concentrations. For astrocyte stimulation (A1 subtype) we used the A1 cocktail: IL-1α (R&D Systems: Minneapolis, Minnesota, Cat# 200-LA, 0.3 ng/mL), TNF α (R&D Systems, Cat# 210-TA, 3 ng/mL) and C1q (MyBiosource: San Diego, California, Cat# MBS147305, 40 ng/mL) factors. Both the compounds, at multiple doses (100 nM, 1 μM, and 10 μM), and the A1 cocktail factors were added simultaneously to the cultures for overnight treatment (24 hours).

ELISA for complement system activation

Media were collected 24 hours post-treatment, and complement levels were assessed with the Milliplex Human Complement Panel 1 (Millipore: St. Louis, Missouri, HCMP1MAG-19 K / Factors: C1q, C3, C3b/iC3b, C4, Factor B, Factor H) and 2 (Millipore, HCMP2MAG-19 K / Factors: C2, C4b, C5, C5a, Factor D, Mannose-binding lectin (MBL), Factor I) according to the kit's instructions. The assay was run in a Bio-Plex 2000 instrument (Biorad).

RESULTS

ALS-PathFX predicted drug network associations to ALS through distinct proteins

Our overall approach used pathway modeling to identify novel drug classes and observational studies in the health record to test the utility of these drug classes (**Figure 1**). To build a custom network medicine platform, we curated several high-quality datasets from the TargetALS data portal and several patient-derived datasets¹¹ (see **Material and Methods**) to create an innovative and unique network medicine platform. We modified our highly flexible PathFX platform (released with¹²) to emphasize these gene lists (referred to as “pathway phenotypes”) into PathFX^{12–14} using the original bias reduction methods published in¹³ and described in methods. We then used our custom network algorithm, ALS-PathFX, to model drug effects and derive new drug classes based on predicted effects on nearby, but untargeted (not bound by drug) proteins (**Figure 1b**).

We generated ALS-PathFX networks for 7,013 approved and experimental drugs using targets from DrugBank.^{15,16} Briefly, ALS-PathFX first identified high-confidence proteins downstream of druggable targets, and after generating a protein network for each drug, ALS-PathFX identified network-associated phenotypes, by assessing disease proteins enriched near druggable targets (as described in¹³ and in **Materials and Methods**). In total, we discovered 13,967 drug-phenotype associations containing 2,984 unique drugs and 90 unique disease pathways. For example, PathFX generated a network of 235 proteins and 238 edges for the antipsychotic, acetophenazine, and connected the network to 138 phenotypes, including 4 neurodegenerative phenotypes (subnetwork associated with four phenotypes shown in **Figure 2**). Acetophenazine's network contained the complement system proteins Complement C3 (C3), Complement C5 (C5), Complement C5a Receptor 1 (C5AR1), and Complement C3a Receptor 1 (C3AR1). Network proteins may be associated with more than one phenotype, and for instance, C3 and C3AR1 were associated with “Multiple sclerosis” and “sequencing data 1.” We assigned acetophenazine to the “complement-associated drug class” or “CS-class” for brevity. We considered any drugs predicted to affect the same phenotypes (e.g., “Multiple sclerosis” or “sequencing data 1”) but without complement proteins in their networks, as sufficient comparator drugs and considered them in the “non-CS-class” of drugs. Like the complement system, we quantified how often a network protein connected a drug-phenotype prediction and looked for protein classes with a reasonable number of approved drugs with and without the protein (see **Materials and Methods**). We provide more detail about ALS-PathFX results and assessment of network-protein drug classes in the **Supplemental Results**. The complement proteins were among the top balanced protein classes, and further, they were an exciting test case given precedent for the utility as drug targets in neuroinflammatory conditions.

Complement-system-associated drugs are associated with changes in survival dynamics

We hypothesized that drugs with complement system proteins in their networks would have distinct clinical effects. This hypothesis is further supported by multiple recent developments of

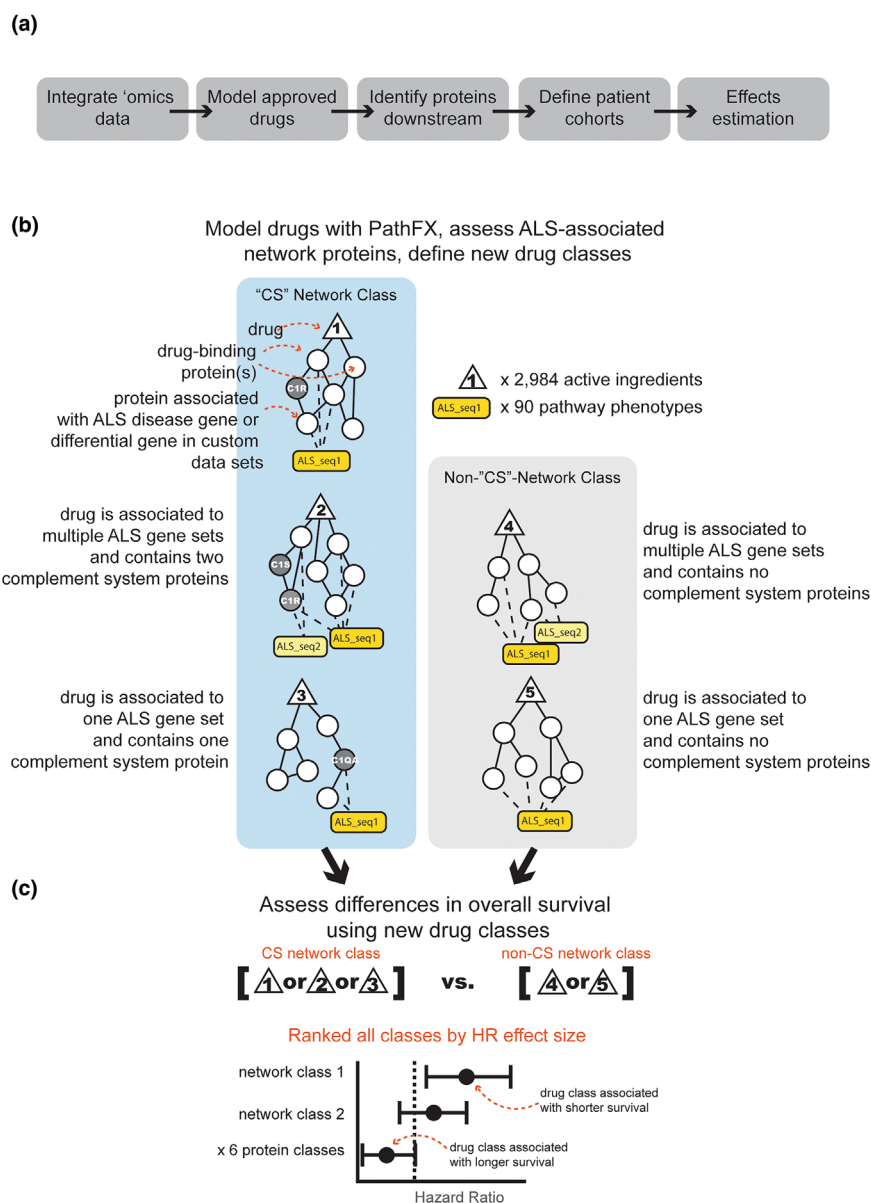


Figure 1 A combined pathways and observational study approach for identifying pathways with protective or harmful effects on overall survival in ALS. (a) Our overall workflow consisted of 'omics data integration using several high-quality data sets, network modeling, identification of downstream proteins, identifying patient cohorts, and effects estimation (b) For network inference, we modeled all approved, active ingredients using ALS-PathFX and then created novel drug classes based on downstream proteins included in drug networks. In this example, we assessed drugs with network associations to multiple ALS sequencing datasets and included them in the “complement system” (“CS”) drug class if they had any complement-system proteins in their networks. We highlighted five drugs that are associated with any one of two disease pathways (i.e., “ALS sequencing data 1”, or “ALS sequencing data 2”). In our schematic, we also highlighted that drugs “1”, “2”, and “3” contain downstream associations to complement-system proteins and drugs “4” and “5” are connected to the same pathway phenotypes, but through different proteins. Drugs, network proteins, and disease phenotypes are represented by triangles, circles, or yellow boxes, respectively. (c) Later, we measured overall survival for patients exposed to at least one of the “complement class” or “non-complement class” drugs using observational studies in the Optum Market Clarity dataset. In this example, we grouped drugs “1”, “2”, and “3” into the “target” class, and drugs “4” and “5” into the “comparator” class. We repeated this analysis for 6 drug classes and ranked classes by their effects on overall survival. Three complement proteins are shown in the diagram, but a total of nine were considered when assigning drugs to the complement system class: “Complement component 1, R subcomponent (C1R)”; “Complement component 1, S subcomponent (C1S)”; “Complement component 1, Q subcomponent, alpha polypeptide (C1QA)”; “Complement component 1, Q subcomponent, beta polypeptide (C1QB)”; “Complement component 1, Q subcomponent, alpha polypeptide (C1QA)”; “Complement component 1, Q subcomponent, gamma polypeptide (C1QC)”; “Complement C3 (C3)”; “Complement C5 (C5)”; “Complement C3a receptor 1 (C3AR1)”; & “Complement C5a receptor 1 (C5AR1)”.

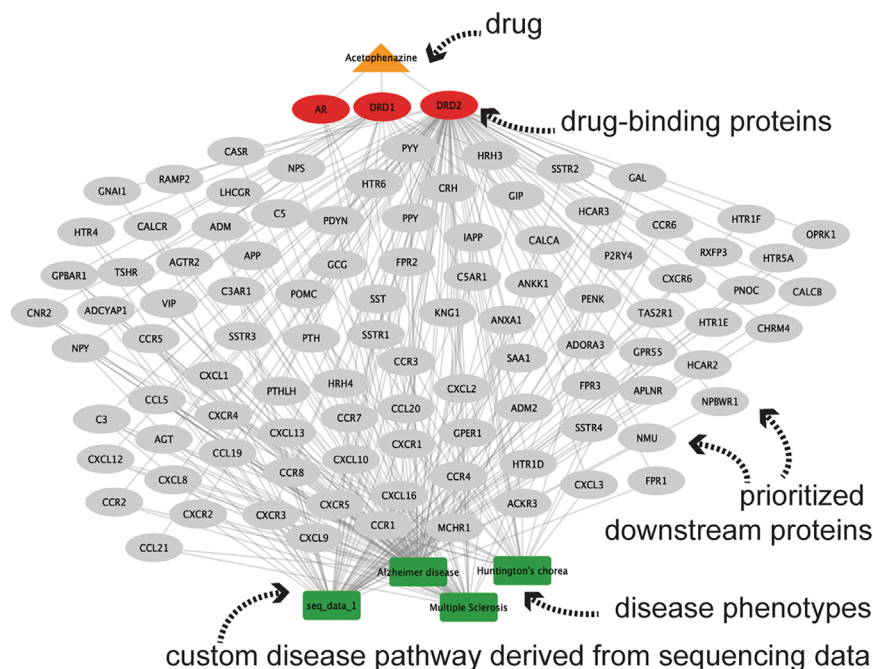


Figure 2 ALS-PathFX connected acetophenazine to multiple disease phenotypes. PathFX connected the drug, acetophenazine (orange triangle) to four relevant disease phenotypes (green boxes): Alzheimer disease, Huntington's chorea, seq_data_1, and Multiple Sclerosis. PathFX used three drug target proteins (red circles): androgen receptor (AR), and the dopamine receptors D1 and D2 (DRD1, DRD2), and 99 adjacent proteins (gray ellipses) when predicting associations to these phenotypes.

complement system inhibitors for neuroinflammation disease, specifically myasthenia gravis (Table S4). Aside from recent discoveries, complement system inhibitors have been approved since 2007 for several other non-neurological diseases, including autoimmune and blood disorders (Table S4). Despite several approved compounds, none were prescribed in our ALS study patient population (*described below and depicted in Figure S6A*, all compounds listed in Supplemental File 4) (Github repository), further emphasizing the need for a way to anticipate their effects on overall survival.

To test whether ALS-PathFX-predicted drugs affected overall survival, we conducted a novel observational study using standard treatment-comparator study design and large-scale propensity scoring to adjust for possible confounding variables. More specifically, we accessed the de-identified Optum Market Clarity data (longer description in Materials and Methods). The novelty in our approach was deriving network-protein-based drug classes instead of using traditional drug classifications such as anatomical therapeutic category (ATC) classification systems (more description in the Supplemental Results).

After controlling for possible confounding (Figure S6), we estimated the differences in overall survival and observed a marginal, decreased, but not significant survival time for patients with an exposure to CS-class drugs (HR = 1.031 95% confidence interval (0.937, 1.134), $P = 0.534$) (Figure 3a, Table 1). The study discovered a detrimental effect early, where CS-class drugs are separated from non-CS-class drugs before ~2,000 days (~5.5 years). At later timepoints, ~10,000 days (~27.4 years), overall survival was indistinguishable. While the result was insufficient to directly change

clinical practice, this result affirmed that network-derived drug classes could predict early differential effects on ALS survival in historical patient cohorts; suggested the possibility of testing other network drug classes. Yet, as mentioned previously, the results do not confirm whether drugs activate or inhibit the complement system.

Network-derived drug classes activate complement system proteins

We next sought additional data sources and experiments to better understand possible drug-induced effects on the complement system. We first sought drug-induced gene expression changes from two public data sources, PharmOmics¹⁷ and LINCS¹⁸ (Supplemental methods), and saw little or ambiguous effects in these data sources (Supplemental results). We next hypothesized that protein-level changes may better inform the observed effect on overall survival. We investigated complement levels using ELISA, specifically measuring C3, Complement 4 (C4), and factor B levels released from healthy stem cells following exposure to drugs from our network and non-network classes (Figure 4). Healthy stem cells were a sufficient system because we started following patients before their ALS diagnosis, and this could indicate whether study drugs induced or suppressed complement system proteins. We prioritized compounds for experimental testing based on those that had the highest number of unique patients and total exposure days per patient (reflecting repeat prescription claims). The CS-class drugs, mirtazapine and methylprednisolone, increased activation of both C3 and C4, and the CS-class drug, gabapentin, appeared to have no effect on either C3 or C4. Of the non-CS-class drugs

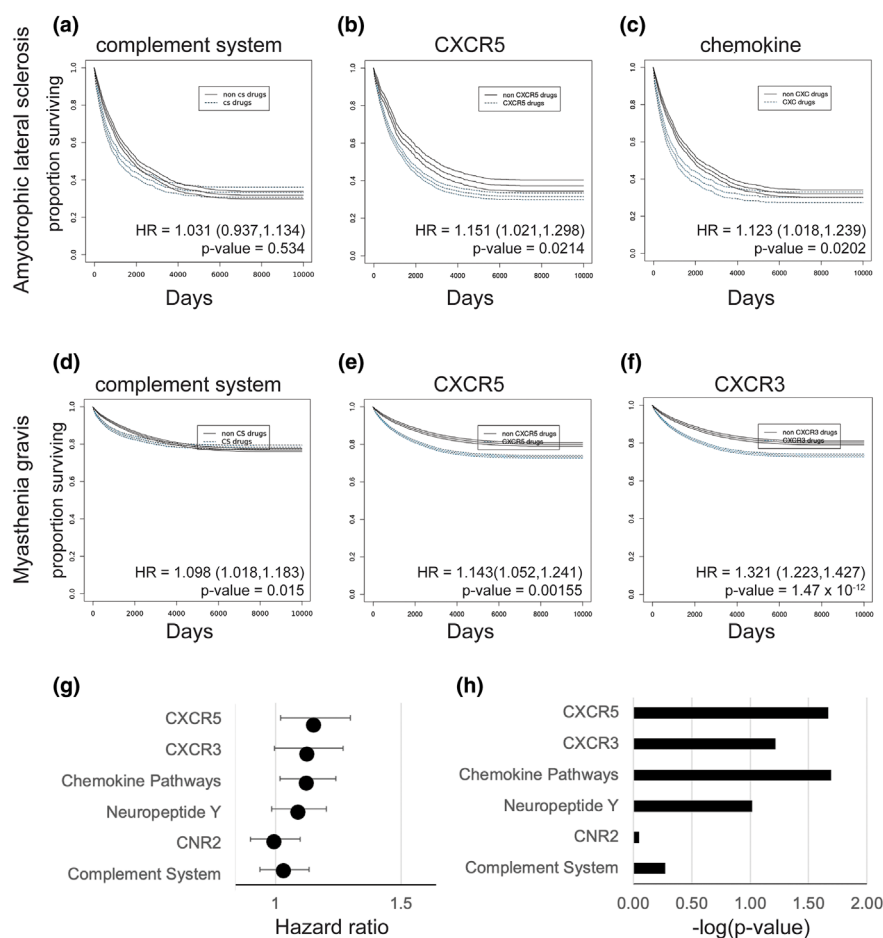


Figure 3 Multiple network drug classes have effects on overall survival dynamics in ALS and myasthenia gravis. We plotted Kaplan-Meier survival curves for the complement system class (a), the chemokine receptor 5 (CXCR5) class, a relatively high-risk class (b), and the chemokine pathway class, also a relatively high-risk class (c) in ALS patients. We also plotted Kaplan-Meier plots for three network classes in myasthenia gravis patients: the complement system (d), chemokine receptor 5 (CXCR5) (e), and the chemokine receptor 3 (CXCR3) (f) classes. In all plots, the proportion of the surviving population was plotted against time in days. Patients exposed to a “network class” or “Non-network class” drug are represented with a dotted blue or solid line, respectively. Hazard ratio values with the 95% confidence interval for six network-drug classes (g) and their $-\log(P\text{-values})$ as estimated from the Wald test (h) as measured with weighted Cox proportional hazards (coxph from the R survival package).

Table 1 Hazard ratios and confidence intervals for 6 drug classes across 3 disease indications

Pathway name	ALS -HR	ALS -CI low	ALS -CI high	MG - HR	MG - CI low	MG - CI high	PD - HR	PD - CI low	PD - CI high
CXCR5	1.151	1.021	1.298	1.143	1.052	1.241			
CXCR3	1.124	0.995	1.269	1.321	1.223	1.427			
Chemokine pathways	1.123	1.018	1.239						
Neuropeptide Y	1.088	0.985	1.203						
CNR2	1.099	1.890	1.097						
Complement system	1.031	0.937	1.134	1.098	1.018	1.183	0.992	0.965	1.020

Each of the colors represents one of the diseases.

ALS, amyotrophic lateral sclerosis (blue); CI, 95% confidence interval; HR, hazard ratio; MG, myasthenia gravis (green); PD, Parkinson's disease (yellow).

tested, cephalexin, amoxicillin, cefuroxime, nitroglycerin, and solatol had no effect on C3 or C4 levels; however, the non-CS-class drugs, famotidine, and erythromycin, increased C3 and C4 levels (Figure 4). Drugs from both classes seemed to induce Factor B, but to a lesser effect. Some of these effects are supported

by additional literature: different methylprednisolone regimens in MG patients led to an increase in C3 and C4 levels.¹⁹ This increase in C3, C4, and Factor B for the above compounds indicates that these compounds induce further complement activation once complement activation is initiated since treatment of human iPSC

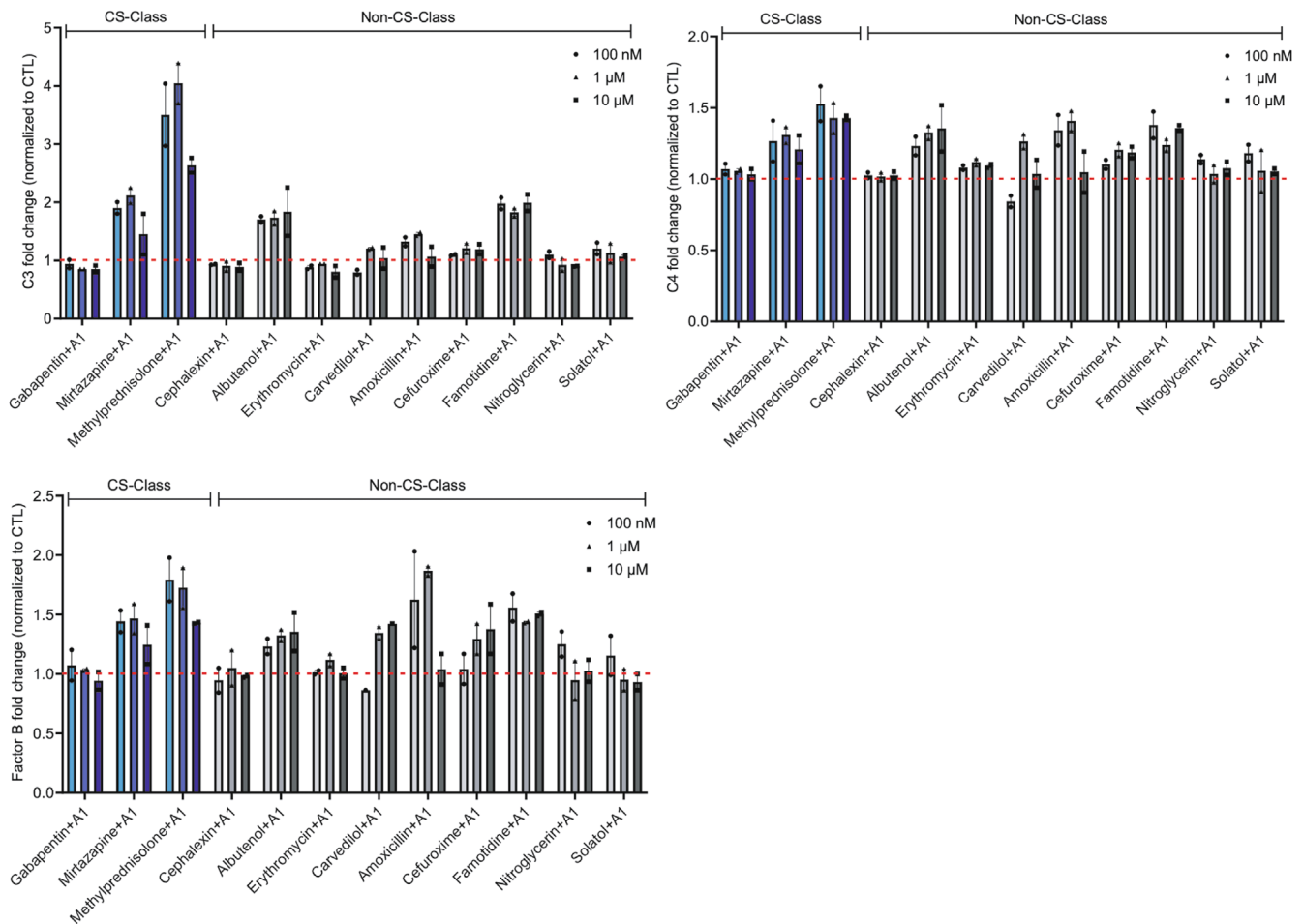


Figure 4 Complement-associated drugs activate multiple components of the complement pathway in ELISA assays. Normalized values (to CTL: A1 cocktail factors) of fold change for the complement factors C3, C4, and Factor B following treatment of human iPSC astrocytes with the A1 cocktail factors and selective classes of compounds (100 nM, 1 μM and 10 μM). Addition of methylprednisolone, mirtazapine, and albuterol showed the strongest effect in increasing the complement levels (especially C3).

astrocytes alone without the A1 cocktail factors did not elicit any complement response (undetectable levels).

Regarding the differential activation across drugs and complement factors, we can only speculate about differences based on the connections identified by our model. In our network model, methylprednisolone, gabapentin, and mirtazapine have distinct network connections to the complement system proteins, C3, C5, complement C3a receptor (C3AR1), and complement C5a receptor (C5AR1). Specifically, methylprednisolone, gabapentin, and mirtazapine are connected through their primary targets annexin A1 (ANXA1), the adenosine A1 (ADORA1) receptor, and the opioid receptor kappa 1 (OPRK1), respectively (all pathway connections in [Supplemental File 2](#) (Github repository)). This suggests that distinct drug targets may have different downstream effects on the complement system.

Several network classes alter overall survival in ALS-patients

We repeated our observational study analysis pipeline for an additional 5 drug classes and ranked their overall effects on survival ([Figure 3b-h](#), [Table 1](#)). We observed many prescription days per

study drug like before; however, the exact brand names changed per study ([Table S6](#), [Figures S9](#), [S11](#), [S13](#), [S15](#) and [S17](#)). Again, we saw a low predictive performance of the logistic regression, suggesting good covariate balance ([Table S7](#)) and similar propensity score plots, suggesting sufficient class balance ([Figures S10](#), [S12](#), [S14](#), [S16](#) and [S18](#)). Across all drug classes, we saw an increased risk for death in the network drug class compared to the non-network classes ([Figure 3g,h](#), [Table 1](#)). The CXCR5 class had the greatest increased risk of death (HR = 1.151, 95% confidence interval (1.021–1.298), $P = 0.0214$) ([Figure 3b](#)), and the chemokine class had the next highest risk (HR = 1.123, 95% confidence interval (1.018–1.239), $P = 0.0202$) ([Figure 3c](#)). The CNR2 class had a marginally decreased but not significant risk (HR = 0.994, 95% confidence interval (0.890–1.097), $P = 0.897$), most like the value observed for the complement system drug class. All other survival curves are provided in [Figures S19–S21](#). All of the retained brand names, the number of unique patients with an RX claim, the prescription days per drug, and patient demographics are included in the extended supplement for each network class comparison (Github repository). We also investigated the extent to which network drug classes shared brand names and observed

shared but distinct brand names in each drug class comparison as measured by Jaccard similarity (Figure S22C,D, Supplemental methods and results).

Network-drug classes affect overall survival in multiple neurodegenerative diseases

An intriguing utility for our analysis could be in the identification of key nodes of pathophysiology across multiple diseases in the nervous system. We first demonstrated that neurodegenerative and neuroinflammatory diseases shared pathways information using network models (Figure S23, Supplemental methods and results). Intriguingly, these diseases shared more genes near druggable targets than when comparing all disease genes. We thus repeated measurements for some network classes in multiple disease indications. We specifically considered the effects of the complement system in myasthenia gravis and Parkinson's disease and the CXCR3, and CXCR5 classes in myasthenia gravis. We again assessed prescription days per drug for all studies (Table S6, Figures S24, S26, S28 and S30), measured low predictive performance of a logistic regression analysis (Table S7), and saw similar propensity score plots between target and comparator cohorts (Figures S25, S27, S29 and S31). Interestingly, we observed a similar trend for MG patients taking the complement system drugs where early survival (before ~2,000 days) showed separation between the CS-class and non-CS-class drugs, however, the curves crossed over at later timepoints (~10,000 days), suggesting a detrimental effect of complement-associated drugs at early timepoints. Unlike our analysis in ALS patients, this analysis yielded a slight, but significantly increased risk for death (HR 1.098, 95% confidence interval (1.018–1.183), $P = 0.015$) (Figure 3d). This finding combined with our experimental results and the recent approvals of complement inhibitors for MG further suggested that activated complement is detrimental for overall survival in neuroinflammatory conditions, if only at early timepoints. The effect of CXCR5 drugs was similar in MG patients relative to ALS patients (HR 1.143, 95% confidence interval (1.052–1.241), $P = 0.00155$) (Figure 3e) and the effect of CXCR3 drugs was greater than in ALS patients (HR 1.321, 95% confidence interval (1.223–1.427), $P = 1.47 \times 10^{-12}$) (Figure 3f). The effect of complement-system drugs in Parkinson's patients showed similar but starker crossover effects like we observed in ALS and MG patients. We observed no overall survival benefit or risk (HR 0.992, 95% confidence interval (0.965–1.020), $P = 0.579$) (Figure S32), suggesting that not all pathophysiology is shared between neuroinflammatory and neurodegenerative conditions. Like before, all extended and supporting data is documented in the Github repository.

DISCUSSION

We presented an approach for using proteins downstream of druggable targets to define new drug classes and tested the effects of these drug classes on overall survival in ALS and two additional neurological diseases. We used observational studies as exploratory analyses, and while they would require further, dedicated clinical investigations to confirm their effects, they are compelling for their ability to recover known effects—specifically the effect of complement activation as detrimental to early survival in

ALS, MG, and PD. Our preliminary experimental data combined with recent approvals of complement inhibitors for MG patients (Table S4) further support the potential of complement inhibitors in ALS. Overall, this suggests a novel and impactful approach for identifying disease drivers in rare and difficult-to-treat diseases and may provide valuable insights into therapeutic target selection.

Our results are further compelling for their utility in difficult-to-treat populations. ALS is insufficiently explained by genetics, with less than 10% of familial and sporadic ALS cases explained by genetic drivers.^{9,10} Without sufficient molecular targets, novel approaches are desperately needed to identify treatment strategies for the disease. Further, the ALS population is relatively small and is typically insufficient for observational studies. Indeed, through a related project, the Veterans Affairs dataset, which contained nearly 20,000 ALS patients, is the largest repository of historical ALS patient data and is nearly double the size of the cohort in this study.²⁰ Yet, we still resolved statistically significant clinical effects by emphasizing proteins downstream of druggable targets (e.g., CXCR3, CXCR5), which allowed the inclusion of more patients.

While we believe our paradigm can greatly enhance model-informed drug development efforts, further work could overcome limitations in our current study. For instance, we had insufficient network connections to all possible druggable targets and were limited to deriving network classes for proteins that were “close” to approved targets. Our network drug classes also shared many approved drugs, indicating that our current interaction network was only sufficiently connected to some pathways. However, the differences in network phenotypes effectively separated drugs to discern relatively stronger effects, such as the increased risks associated with CXCR5 networks. This emphasizes the importance of deriving high-quality, disease pathway phenotypes. We also tested these network classes in a single patient population (i.e., Optum) and further testing using data from additional health systems could elevate the impact of these findings. Indeed, previous meta-analyses across healthcare systems have successfully discovered drug-class effects on cardiovascular disease outcomes.²¹ Further, an intriguing opportunity, or possible challenge, of historical data analysis is how to interpret the outcomes in the context of EHR timescales. We followed patients for nearly two decades when most clinical trials are months to years in duration. Of the four FDA-approved therapies for ALS, the riluzole trials measured outcomes at 12 months²² or 18 months follow-up²³ and the trials for edavarone, phenylbutyrate-taurursodiol (PB-TUDCA), and tofersen measured changes in the revised ALS functional rating scale²⁴ at shorter durations—ranging from 24 to 28 weeks.^{9,25–28} Our measured differences in proportions of surviving patients at short durations (~2 years) are like effect sizes in the riluzole trials, suggesting that future observational studies, which leverage drug network classes, could shorten observation times to better anticipate effect sizes measurable in the duration of a clinical study. Additionally, it could be possible to corroborate early effect sizes with long-term outcomes, especially because we observed that many of our drug classes had effects at early but not late timepoints. Another point for consideration is the selection of endpoints—we emphasized overall survival because patients otherwise have no life-sparing treatments. However, we cannot disambiguate if the predicted drug targets are directly

reversing disease progression or buffering other organ functions despite ongoing pathophysiology. Another limitation, and opportunity, would be to validate these findings, especially the durations of the survival curves, in another patient cohort.

Our discovery of disease drivers complements an evolving landscape of shared molecular underpinnings of clinically distinct diseases. Specifically, we predicted that activated complement would increase the risk of death across multiple disease indications. Our pathways analysis of shared dysregulation is supported by other findings; specifically, Arneson et al.²⁹ noted that when disease genes and proteins weren't shared across neurodegenerative diseases, gene and protein functional information was shared, further emphasizing shared dysregulation across neurodegenerative diseases. Separate investigations of post-mortem brain samples further corroborated shared gene-level changes across patients with four diseases (Alzheimer's Disease, Parkinson's disease, Huntington's disease, and ALS) and discovered that shared genes were involved in functional processes such as inflammation, mitochondrial dysfunction, and oxidative stress.^{29,30} Additionally, multiple canonical inflammatory reactivity inducers, including complement subcomponent C1q, interleukin 1 alpha (IL-1 α), and tumor necrosis factor (TNF), convert astrocytes into A1 reactive astrocytes,³¹ and these proinflammatory, reactive astrocytes secrete multiple complement factors, reduce synaptic and neuronal connectivity, and are linked extensively to neurodegenerative disease.³² Taken together, this further supports the possibility of shared molecular changes in these diseases; however, further experimental validation would be required to confirm this general finding.

Our approach motivates extending the integration of clinical health data into Model Informed Drug Development (MIDD). Already, the FDA has provided guidance about the use of real-world data (RWD) for generating real-world evidence (RWE) from sources other than randomized, controlled trials (RCTs) (Table S9). Indeed, many have generated RWE to support regulatory decision making, including accelerated approvals, learning intrinsic factors, and optimizing dosing (Table S9). Yet these studies did not leverage RWD for novel target support.

Other sources have identified novel targets from RWD; however, they have focused directly on target-driven effects instead of downstream effects. Genetic data repositories have helped identify lead candidates from patient data, such as in the case of Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors, where inactivating and gain-of-function mutants were associated with elevated or reduced cholesterol, respectively.^{33,34} Mendelian randomization is another method of inferring drug-target effects in new populations,³⁵ but not every suitable target has sufficient genetic connections to support later therapeutic development. Additionally, analysis of PheWAS and expression quantitative trait loci (eQTLs) has also coupled gene variation with observable clinical effects, especially side effects.^{36,37} In contrast to our approach, these approaches overlook downstream effects and require sufficient genetic variation in patient populations to advance therapeutic discovery.

For the most part, prior work combining molecular models of drug effects with patient data has been limited to drug repurposing studies, and we believe our approach greatly expands the

possibilities for identifying promising therapeutic targets from real-world patient data. Knowing that electronic health records will be increasingly useful in supporting new target identification, we anticipate that network medicine will extend capabilities beyond target-centric approaches.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

The authors would like to acknowledge the Sanofi real-world data team, and especially Thomas Verrier, the RWD-Platform Support team for their invaluable support navigating data access, without which, the study could not be completed.

FUNDING

This project was supported by a Sanofi iDEA-TECH award.

CONFLICTS OF INTEREST

The authors Jeremy Y. Huang, Georgia Dermentzaki, Anna S. Blazier, Giorgio Gaglia, Timothy R. Hammond, Francesca Frau, Mary Clare Mccorry, and Dmitry Ofengeim are Sanofi employees and may hold shares and/or stock options in the company. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

J.L.W., M.A., J.Y.H., G.D., A.S.B., F.F., and D.O. wrote the manuscript. J.L.W., J.Y.H., G.D., A.S.B., and D.O. designed the research. J.L.W., J.Y.H., M.A., and G.D. performed the research. J.L.W., M.A., and G.D. analyzed the data. A.S., G.G., T.R.H., F.F., and M.C.M. contributed new reagents/analytical tools.

DATA AVAILABILITY STATEMENT

The Market Clarity data used were licensed from Optum and are not publicly available. To increase the rigor and reproducibility of our work, we have released code and data for the network analysis, and anonymized code for accessing clinical data through a GitHub repository: https://github.com/jenwilson521/network_drug_classes_als/. All supplemental Excel files are included in their original format in the GitHub repository to increase readability and reusability of the data.

© 2025 The Author(s). *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. Plenge, R.M., Scolnick, E.M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
2. Adesuyan, M. et al. Phosphodiesterase type 5 inhibitors in men with erectile dysfunction and the risk of Alzheimer disease. *Neurology* **102**, e209131 (2024).
3. Fang, J. et al. Endophenotype-based in silico network medicine discovery combined with insurance record data mining identifies sildenafil as a candidate drug for Alzheimer's disease. *Nat. Aging* **1**, 1175–1188 (2021).
4. Fang, J. et al. Artificial intelligence framework identifies candidate targets for drug repurposing in Alzheimer's disease. *Alzheimers Res. Ther.* **14**, 7 (2022).
5. Cheng, F. et al. Network medicine informed multi-omics integration identifies drug targets and repurposable medicines for amyotrophic lateral sclerosis (2024) <https://doi.org/10.21203/rs.3.rs-4271755/v1>.

6. Cheng, F. *et al.* Artificial intelligence and open science in discovery of disease-modifying medicines for Alzheimer's disease. *Cell Rep. Med* **5**, 101379 (2024).
7. Wilson, J.L., Gravina, A. & Grimes, K. From random to predictive: a context-specific interaction framework improves selection of drug protein–protein interactions for unknown drug pathways. *Integr. Biol.* **14**, 13–24 (2022).
8. Wilson, J.L., Steinberg, E., Racz, R., Altman, R.B., Shah, N. & Grimes, K. A network paradigm predicts drug synergistic effects using downstream protein–protein interactions. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 1527–1538 (2022).
9. Ilieva, H., Vullaganti, M. & Kwan, J. Advances in molecular pathology, diagnosis, and treatment of amyotrophic lateral sclerosis. *BMJ* **383**, e075037 (2023).
10. Ghasemi, M. & Brown, R.H. Genetics of amyotrophic lateral sclerosis. *Cold Spring Harb. Perspect. Med.* **8**, a024125 (2018).
11. Zelic, M. *et al.* Single-cell transcriptomic and functional studies identify glial state changes and a role for inflammatory RIPK1 signaling in ALS pathogenesis. *Immunity* **58**, 961–979 (2024) <https://doi.org/10.1016/2024.04.12.589201>.
12. Wilson, J.L., Wong, M., Stepanov, N. & Petkovic, D.P. PhenClust, a standalone tool for identifying trends within sets of biological phenotypes using semantic similarity and the unified medical language system. *JAMIA Open* **4**, ooab079 (2021).
13. Wilson, J.L. *et al.* PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development. *PLoS Comput. Biol.* **14**, e1006614 (2018).
14. Wilson, J.L. *et al.* PathFXweb: a web application for identifying drug safety and efficacy phenotypes. *Bioinformatics* **11**, 265–4506 (2019).
15. Wishart, D.S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
16. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
17. Chen, Y.-W. *et al.* PharmOmics: a species- and tissue-specific drug signature database and gene-network-based drug repositioning tool. *iScience* **25**, 104052 (2022).
18. Keenan, A.B. *et al.* The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* **6**, 13–24 (2018).
19. Lin, X. & Qi, G. Observation on the efficacy of different methylprednisolone regimens in the treatment of myasthenia gravis. *Pak. J. Méd. Sci* **38**, 910–915 (2022).
20. Reimer, R.J. *et al.* An electronic health record cohort of veterans with amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Front. Degener.* **24**, 695–701 (2023).
21. Suchard, M.A. *et al.* Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* **394**, 1816–1826 (2019).
22. Bensimon, G., Lacomblez, L. & Meininger, V. A controlled trial of Riluzole in amyotrophic lateral sclerosis. *N. Engl. J. Med.* **330**, 585–591 (1994).
23. Lacomblez, L., Bensimon, G., Leigh, P.N., Guillet, P. & Meininger, V. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. *Lancet* **347**, 1425–1431 (1996).
24. Cedarbaum, J.M. *et al.* The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *J. Neurol. Sci.* **169**, 13–21 (1999).
25. Abe, K. *et al.* Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* **16**, 505–512 (2017).
26. Paganoni, S. *et al.* Trial of sodium phenylbutyrate–Taurursodiol for amyotrophic lateral sclerosis. *N. Engl. J. Med.* **383**, 919–930 (2020).
27. Miller, T.M. *et al.* Trial of antisense oligonucleotide Tofersen for SOD1 ALS. *N. Engl. J. Med.* **387**, 1099–1110 (2022).
28. Brooks, B.R. *et al.* Edaravone efficacy in amyotrophic lateral sclerosis with reduced forced vital capacity: post-hoc analysis of study 19 (MCI186-19) [clinical trial NCT01492686]. *PLoS One* **17**, e0258614 (2022).
29. Arneson, D., Zhang, Y., Yang, X. & Narayanan, M. Shared mechanisms among neurodegenerative diseases: from genetic factors to gene networks. *J. Genet.* **97**, 795–806 (2018).
30. Li, M.D., Burns, T.C., Morgan, A.A. & Khatri, P. Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. *Acta Neuropathol. Commun.* **2**, 93 (2014).
31. Liddel, S.A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
32. Lawrence, J.M., Schardin, K., Wigdahl, B. & Nonnemacher, M.R. Roles of neuropathology-associated reactive astrocytes: a systematic review. *Acta Neuropathol. Commun.* **11**, 42 (2023).
33. Cohen, J.C., Boerwinkle, E., Mosley, T.H. & Hobbs, H.H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
34. Horton, J.D., Cohen, J.C. & Hobbs, H.H. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem. Sci.* **32**, 71–77 (2007).
35. Bowden, J. & Holmes, M.V. Meta-analysis and mendelian randomization: a review. *Res. Synth. Methods* **10**, 486–496 (2019).
36. Duffy, Á. *et al.* Tissue-specific genetic features inform prediction of drug side effects in clinical trials. *Sci. Adv.* **6**, eabb6242 (2020).
37. Nguyen, P.A., Born, D.A., Deaton, A.M., Nioi, P. & Ward, L.D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat. Commun.* **10**, 1579 (2019).