scientific reports

OPEN

SynaptopathyDB integrates synaptic proteomes, genetic and phenotypic data to advance research on nervous system disorders

Oksana Sorokina[1,8], Digin Dominic[2,8], Àlex Bayés[3,4], J. Douglas Armstrong[1,5] & Seth G. N. Grant[2,6,7 ✉]

Synaptic dysfunction resulting from pathogenic variants in genes encoding synaptic proteins is a major contributor to brain and behavioural disorders, collectively termed synaptopathies. To facilitate research into the genetic basis and clinical manifestations of synaptopathy we have created SynaptopathyDB, an online resource that integrates data from 64 mammalian synapse proteomic studies and multiple genetic and phenotypic resources (www.synaptopathyDB.org). We identified a consensus set of 3,437 mammalian synapse proteins from presynaptic and postsynaptic compartments, which have wide application in genetic and omic studies. Mutations in 954 genes encoding 28% of the consensus synapse proteome were associated with 1,266 OMIM diseases of the central and peripheral nervous system. We present findings that underscore the pervasive role of synaptic gene variants in the phenotypes of neurological, psychiatric, developmental, and systemic disorders highlighting the significant burden they impose on individuals and healthcare systems. SynaptopathyDB is a versatile platform and discovery tool for understanding the role of synapse proteins and genetic variants in human disease phenotypes.

Synapse proteome datasets have had an immense impact on our understanding of human brain disease. The realisation that mutations in genes coding for synaptic proteins cause human brain disorders arose from proteomic studies of the postsynaptic proteome of excitatory synapses in 2000, which linked synaptic proteins to three genetic disorders causing severe neurological symptoms[1]. By 2011, mutations in 199 genes encoding postsynaptic proteins had been linked to 133 nervous system disorders[2]. In addition to revealing causal mechanisms for many rare monogenic disorders, the synapse proteome played a critical role in unravelling the biological mechanisms of polygenic disorders, including schizophrenia[3–8], which is now recognised as a synaptic disorder. Recent large-scale genetic studies of depression have leveraged synapse proteome data to reveal that, like schizophrenia, the disorder arises from genetic variants targeting many diverse postsynaptic proteins[7,9–12]. In addition to revealing the mechanisms of common psychiatric disorders, synapse proteome datasets have been instrumental in understanding the biological basis of the genetic variation underlying Intelligence Quotient (IQ), which also shows an enrichment in genes encoding synaptic proteins[13–15].

Since 2011, there have been significant advances in the depth, scale and throughput of human genome sequencing, leading to the discovery of many more disease associated mutations. Concurrently, numerous proteomic studies have mapped the protein composition of mammalian synapses, expanding the catalogue of known synaptic proteins[16]. However, there are no centralised and readily searchable resources that integrate data on synaptic proteomes, genetics, diseases, and phenotypes. The development of such a resource would greatly

[1]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK. [2]Genes to Cognition Programme, Institute of Neuroscience and Cardiovascular Research, University of Edinburgh, Edinburgh EH16 4SB, UK. [3]Molecular Physiology of the Synapse Laboratory, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain. [4]Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain. [5]Computational Biomedicine Institute (IAS-5 / INM-9), Forschungszentrum Jülich, Jülich, Germany. [6]Euan MacDonald Centre, University of Edinburgh, Edinburgh EH16 4SB, UK. [7]Simons Initiative for the Developing Brain (SIDB), Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh EH8 9XD, UK. [8]Oksana Sorokina and Digin Dominic contributed equally to this work. ✉email: seth.grant@ed.ac.uk

facilitate both simple and complex analyses. At present, it is not straightforward to retrieve detailed associations between proteins involved in synaptopathies and their associated diseases or phenotypes. Generating consensus sets of synaptic proteins across multiple studies would greatly assist genetic association and gene enrichment analyses by providing a common reference. A comprehensive and accessible resource that would support systematic investigations into the relationships between synaptic proteins and phenotypes is lacking.

Here, we present SynaptopathyDB (www.synaptopathyDB.org), a comprehensive platform designed to facilitate exploration of the synaptic basis of brain and peripheral nervous system disorders via three intuitive search interfaces: Gene Search, Disease Search, and Publication Search. Users can efficiently retrieve relevant information using autocomplete suggestions, flexible query terms and refine results with filters such as subsynaptic protein localization. Expandable results panels provide detailed annotations and direct links to external databases including Online Mendelian Inheritance in Man (OMIM)[17,18], International Classification of Diseases (ICD-11)[19], Human Phenotype Ontology (HPO)[20], Gene Ontology (GO)[21], Synaptic Gene Ontology (SynGO)[22], UniProt[23] and PubMed. In addition to search functionalities, SynaptopathyDB offers interactive visualizations for tracking disease associations, discovery rates, and publication trends over time through customizable plots. For advanced users the platform provides an application programming interface (API), enabling programmatic access to raw data and complex queries. Together, these features position SynaptopathyDB as both a reference database and an integrated research tool for connecting synaptic proteins, genetic variation, and human disease, providing a new tool that will help development of treatments for synaptopathies.

Results

Identifying a consensus set of synapse proteins

We previously reported a database comprising 58 proteomic studies characterising the protein composition of mammalian synaptic fractions[16]. This has now been expanded to include 64 studies, collectively reporting over 8,000 unique proteins[24]. These investigations have primarily focused on human and rodent models, whose synaptic proteomes are known to be highly conserved across species[2,25–28]. However, many of the proteins identified in these studies have been reported only once and might represent experimental artefacts or contaminants. To improve data confidence, we defined a highly-reproducible consensus set of 3,437 synaptic proteins, each supported by evidence from at least five independent studies (see methods). This curated set is hereafter referred to as SynProteome^{CONS}.

The synapse comprises two principal substructures: the presynaptic and postsynaptic terminals, which are interconnected by proteins within the synaptic cleft. The presynaptic bouton contains synaptic vesicles loaded with neurotransmitters, while the postsynaptic terminal houses multiprotein complexes that anchor neurotransmitter receptors, cell adhesion molecules, scaffolding proteins, and signalling molecules to the cytoskeleton. The proteomes of synaptosomes — isolated whole synapses — as well as presynaptic and postsynaptic compartments and their constituent complexes and substructures have been characterized in the 64 proteomic studies. To produce consensus sets of proteins for the presynaptic and postsynaptic terminals within SynProteome^{CONS} we compared the proteins characterised by the different methodologies and found eight groups (Supplementary Fig. 1A), which we reduced to three (Supplementary Fig. 1B): SynProteome^{POST} containing 1,354 proteins only found in postsynaptic preparations; SynProteome^{PRE+} comprising 456 presynaptic proteins plus proteins found in both presynaptic and postsynaptic fractions; and SynProteome^{SYN} comprising 1,627 proteins isolated from synaptosomes but not specifically allocated to presynaptic or postsynaptic compartments.

Disease-relevant genetic variants target all synapse compartments

To identify disease-relevant (coding and non-coding) variants targeting synaptic proteins, we cross-referenced each of the 3,437 SynProteome^{CONS} genes with the OMIM database[17,18], revealing 1,236 (35.9%) genes with variants associated with one or more diseases (Supplementary Table S1). A high percentage of genes in each compartment were targeted by mutations with OMIM phenotypes: 45% (206/456) of SynProteome^{PRE+}, 38% (516/1354) of SynProteome^{POST} and 32% (514/1627) of SynProteome^{SYN} genes (Table 1).

We next asked how many of these disease-relevant genes were linked to diseases that involve the central (CNS) and peripheral (PNS) nervous system or to non-nervous system disorders (Table 1). 78% (954 genes) of SynProteome^{CONS} disease-relevant genes have either a CNS or PNS phenotype, with similar percentages in each of the synaptic compartments. Of the different synaptic compartments, SynProteome^{PRE+} showed the highest

Proteome set	Unique genes	Unique genes with OMIM ID	CNS	PNS	PNS&CNS	Total NS	nonNS
SynProteome^{CONS}	3437	1236	709 (57%)	25 (2%)	220 (17.8%)	954 (78%)	282 (22.8%)
SynProteome^{POST}	1354	516	296 (57.3%)	15 (2.9%)	93 (18%)	404 (78.2%)	112 (21.7%)
SynProteome^{PRE+}	456	206	129 (62%)	2 (1%)	43 (20.1)	174 (84.4%)	32 (15%)
SynProteome^{SYN}	1627	514	284 (55.2)	8 (1.5%)	84 (16.3)	376 (73.1)	138 (26.8%)

Table 1. Synapse proteome subsets and their associated disease genes. For each proteome set the number of genes encoding that set (Unique genes) and the number with OMIM IDs (Unique genes with OMIM ID) are shown. Additional columns show the number and percentage of disease genes within each proteome set associated with phenotypes with CNS, PNS, combined CNS and PNS. Total NS column, combining the PNS, CNS and PNS&CNS, corresponds to the number of OMIM diseases that have any of the nervous system phenotypes. Genes associated with non-nervous system phenotypes (nonNS) are shown.

levels of enrichment (total nervous system, 84.4%, 174 genes, $p = 1.4 \times 10^{-7}$; CNS-associated, 62%, 129 genes, $p = 1.7 \times 10^{-5}$, hypergeometric test).

Types of brain diseases and their phenotypes

To further understand the types of brain disorders caused by synaptic mutations we integrated SynProteome[CONS] with ICD-11[19]. ICD-11 organizes disorders into numbered chapters, several of which are directly relevant to the nervous system, including Chap. 6 (Mental, behavioural or neurodevelopmental disorders) and Chap. 8 (Diseases of the nervous system). As shown in Fig. 1A, ranking the 20 ICD-11 chapters by the number of associated synaptic gene variants revealed that the majority of these mutations are concentrated in four chapters. The chapter with the highest number of associated genes is Chap. 8, comprising 343 of the 1,236 genes, with Chap. 6 also prominently represented (227 genes). A substantial number (227 genes) of disease-associated synaptic genes were also mapped to Chap. 20 (Developmental anomalies), which includes disorders affecting brain structure and nervous system formation, and (316 genes) to Chap. 5 (Endocrine, nutritional or metabolic diseases), which includes metabolic dysfunctions — such as impaired energy metabolism — that often compromise neuronal function and may manifest with diverse nervous system phenotypes. Plotting the contribution of CNS and PNS diseases to each of these chapters shows the major contribution of synaptic pathogenic variants to the CNS (Fig. 1A).

We also asked if any of these disease categories are more closely associated with any specific synapse compartments (Table 2). We found that SynProteome[POST] shows a significant enrichment in genes associated with mental/neurodevelopmental disorders (Chap. 6) ($p = 7.9 \times 10^{-6}$) and that SynProteome[PRE+] shows significant enrichment for nervous system diseases (Chap. 8, $p = 2.4 \times 10^{-8}$) and modest significance in developmental disorders (Chap. 20) ($p = 0.049$). Next, we examined the subcategories within diseases of the nervous system (Chap. 8), which describes many classes of disorders that traditionally fall under the large family of neurological diseases (Fig. 1B). Epilepsies and movement disorders contain the highest number of synapse gene variants and both are dominated by CNS-acting gene mutations.

Diseases are classified from constellations of phenotypes, which clinically are referred to as symptoms and signs. To identify the human phenotypes caused by pathogenic variants in genes encoding synaptic proteins, we mapped SynProteome[CONS] genes involved with disease to the Human Phenotype Ontology (HPO)[20], identifying a total of 3,986 phenotypes (Supplementary Table S2). Next, using gene set enrichment analysis we identified
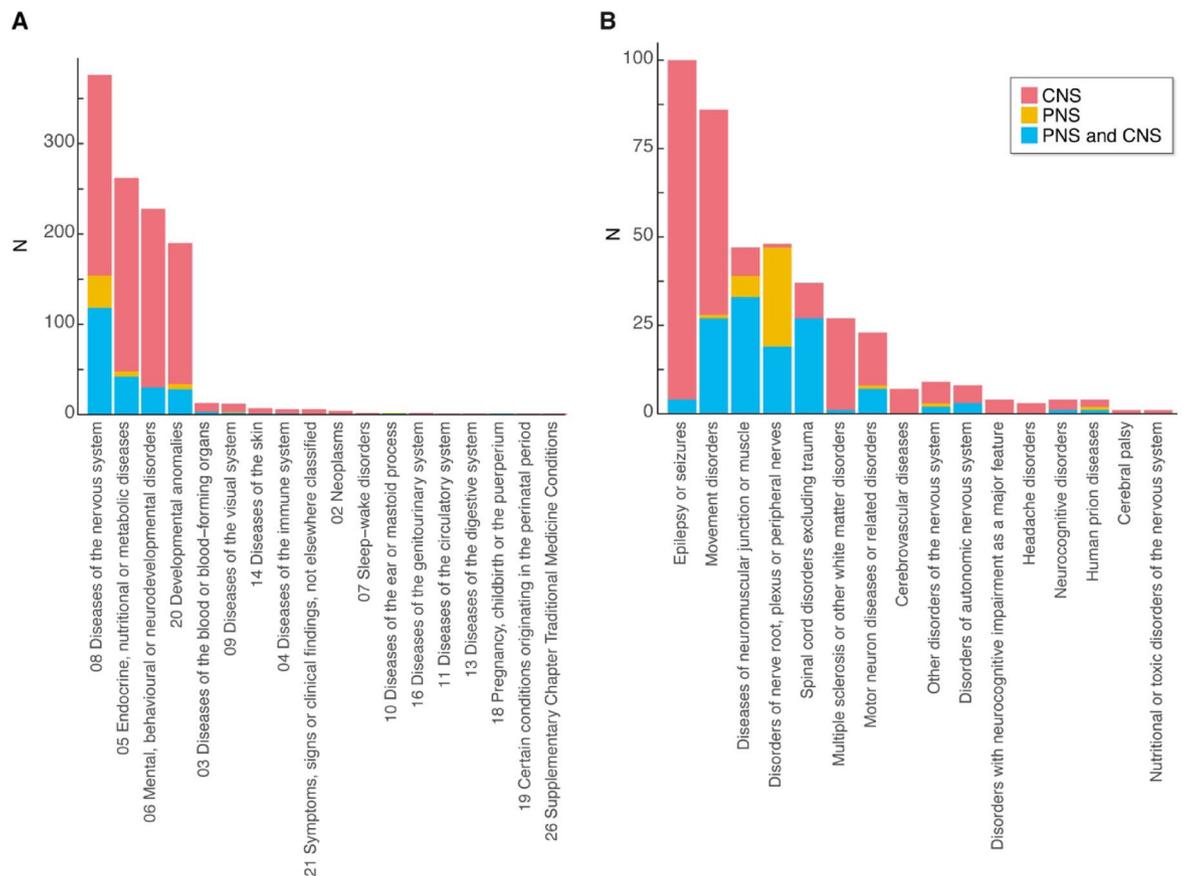

Fig. 1. Diseases associated with synaptic gene mutations. (**A**) The number of OMIM-associated genes associated with each ICD-11 chapter subdivided into those associated with CNS, PNS, or both, and their phenotypes. (**B**) The number of OMIM-associated genes associated with disorders within ICD-11 Chap. 8 (Diseases of the nervous system).

Proteome set	Unique genes	Unique genes with OMIM ID	Chapter 8 disease of nervous system	Chapter 6 mental, behavioural or neurodegenerative	Chapter 5 endocrine, nutritional metabolic	Chapter 20 developmental anomalies
SynProteome^{CONS}	3437	1236	343 (28%)	227 (18.3%)	316 (25.5%)	227 (18.4%)
SynProteome^{POST}	1354	516	153 (27.5%)	121 (23.4%)	96 (18.6%)	100 (19.3%)
SynProteome^{PRE+}	456	206	81 (33.9%)	44 (21.3%)	41 (19.9%)	39 (18.9%)
SynProteome^{SYN}	1627	514	109 (21.1%)	62 (11.9%)	179 (33.4%)	88 (17.1%)

Table 2. Synapse proteome subsets and their associated disorders. For each proteome set the number of genes encoding that set (Unique genes) and the number with OMIM IDs (Unique genes with OMIM ID) are shown. Additional columns show the number and percentage of disease genes in ICD-11 Chaps. 5, 6, 8 and 20.
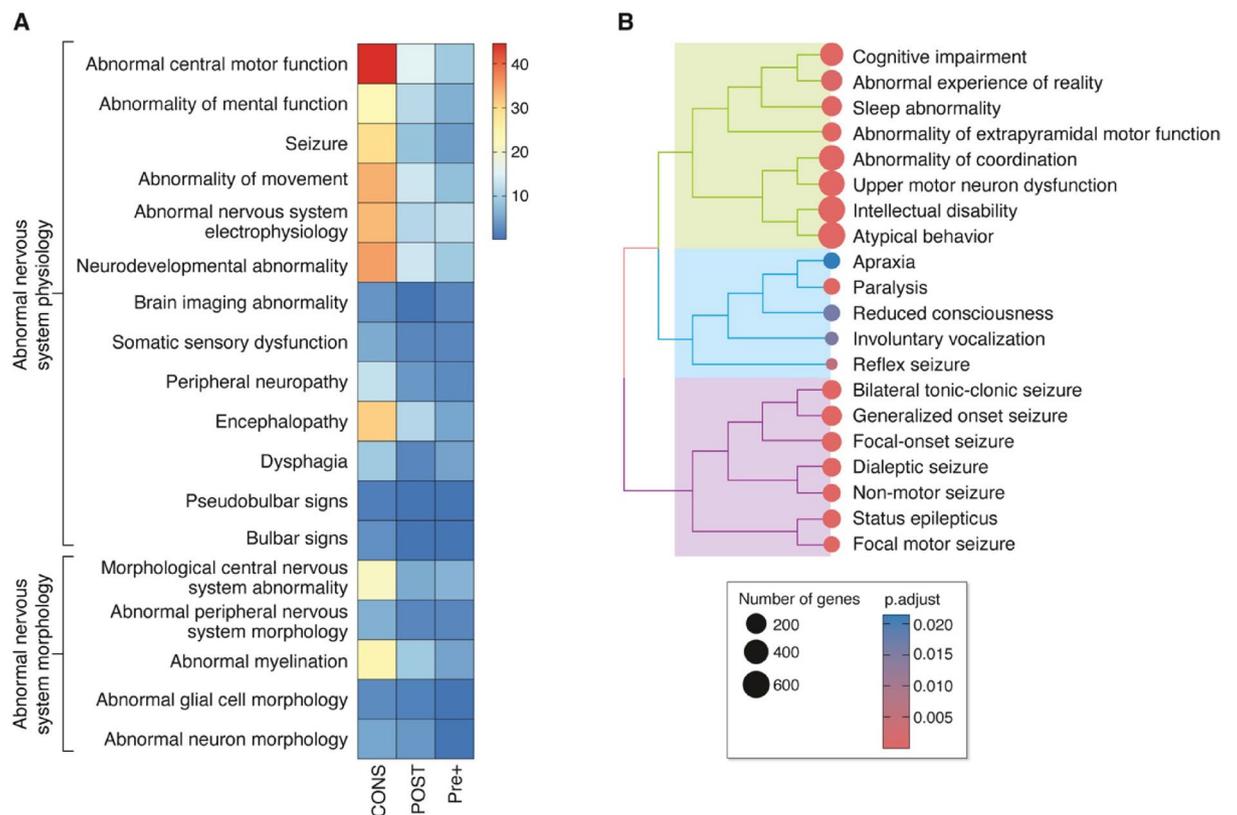

Fig. 2. Human phenotypes of synaptic mutations. (**A**) Highly enriched HPO phenotypes within Abnormalities of the Nervous System phenotypes. Key, log(p.adj) enrichment value. CONS, SynProteome^{CONS}; POST, SynProteome^{POST}; Pre+, SynProteome^{PRE+}. (**B**) Clustering of highly enriched cognitive, motor and epileptic phenotypes within the Abnormal Nervous System Physiology category.

the phenotypes most relevant for the SynProteome^{CONS} and found 739 significantly enriched terms. The major enriched phenotype was Abnormality of the Nervous System and within this two other terms Abnormal Nervous System Physiology and Abnormal Nervous System Morphology, were also highly enriched. Figure 2A shows 18 terms from these two categories which are highly significantly enriched ($P < 10^{-5} - 10^{-45}$) in SynProteome^{CONS}, SynProteome^{PRE+} and SynProteome^{POST} genes. Cognitive, motor and epileptic phenotypes are highly enriched within the Abnormal Nervous System Physiology category (Fig. 2B).

Declining rate of discovery of diseases linked to synapse mutations
How many more diseases arising from genetic variants in genes coding for synaptic proteins remain to be discovered? To evaluate this, we built a display in SynaptopathyDB that produces a chronology of the number of diseases linked to the 1,236 genes (Fig. 3A). This shows that the total (nervous and non-nervous system) number of diseases discovered per year reached a peak in 2013 and since 2017 has been steadily declining. This trend is reflected in the cumulative disease discovery plot shown in Fig. 3B. These data show that 1,768 (1,266 nervous and 502 non-nervous system) diseases were identified by 2024, suggesting that the spectrum of diseases and phenotypes associated with synaptic gene variants for this set of genes is reaching completion. However, at
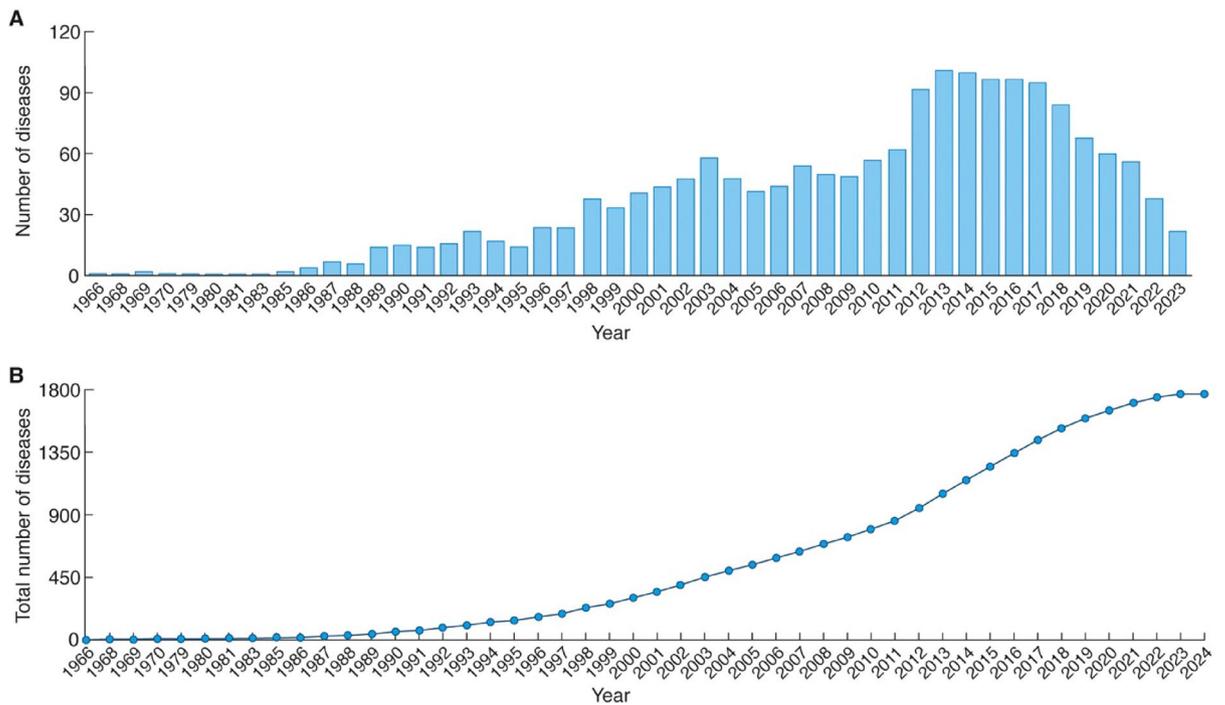
Fig. 3. Declining rate of discovery of diseases linked to synapse mutations. (**A**) The number of diseases discovered each year that were later reported to be linked to synaptic gene mutations. (**B**) The cumulative number of diseases discovered per year.

present, as 64% of SynProteome^{CONS} is not associated with disease variants, it thus seems plausible that many genes encoding this subset will be candidates for future discovery.

Using the SynaptopathyDB website

The website features a comprehensive search system with three main entry points: 'Gene Search', which allows users to identify which nervous system disorders are caused by a given gene or set of genes; 'Disease Search', which allows users to perform the opposite search, identifying which synaptic genes are associated with specific disorders; and 'Publication Search', which retrieves the scientific publications with gene-disease connections. The platform supports quick search with autocomplete[29–31], multiple identifier queries, and highlighted results. Filtering options enable users to refine searches based on protein localization within the synapse or specific gene mentions in publications. Expandable panels, link-outs to external databases, and rich metadata, including subcellular locations and mutation details (for epilepsy and ASD) enhance utility. Data visualization tools provide interactive disease trend analyses, track discovery rates over time, and display annual mentions of disease-associated genes through various chart types. Additionally, the platform offers API endpoints for exporting structured datasets, including synapse proteome lists, in CSV format for further analysis.

Discussion

From 64 proteomic studies, we defined a consensus set of 3,437 proteins, filtered from more than 8,000 candidates using a replication threshold of five or more studies, chosen based on the distribution of gene identifications[16]. We consider this threshold sufficient to capture the complexity of the synapse proteome. Since the discovery rate of new synapse proteins has plateaued, it is unlikely that the overall complexity of the synapse proteome will expand substantially. We also generated consensus proteomes for the presynaptic and postsynaptic compartments, the principal structural domains of the synapse. Continued refinement of proteomic methodologies, particularly the purification and characterisation of complexes and supercomplexes[32], will be essential for understanding how mutations disrupt nanoscale synaptic molecular organisation.

Our meta-analysis of the high-confidence set (3437 proteins) of the mammalian synaptic proteome reveals pathogenic variants in 954 of the corresponding genes linked to 1,266 nervous system disorders causing 739 different phenotypes. These findings highlight the central role of synaptic protein structure and function in sustaining normal neural processes and behaviour.

Linking synaptic protein mutations to disease and phenotype sheds light on the pervasive functions of synapses in the nervous system. Of the 20 broad ICD-11 disease chapters, four related to nervous system disorders accounted for the largest burden. These disorders encompass virtually all neural functions, including motor control, sensation, perception, learning, sleep, consciousness, and metabolic regulation, and they manifest throughout the lifespan, from infancy to old age. However, each gene or protein typically regulates specific constellations of human phenotypes[2]. Consistent with this, large-scale phenotyping of mice with synaptic mutations reveals that individual genes affect subsets of behavioural components, whereas combinations of

genes converge on specific behavioural traits[33]. One mechanism that could explain this specificity is the spatial and temporal expression patterns of synaptic proteins. Regional proteomic and transcriptomic studies of the human brain demonstrate distinct synaptic protein compositions with enriched phenotype associations[34,35]. Synaptome mapping studies in the mouse brain at single-synapse resolution have revealed extraordinary diversity in synaptic composition, protein turnover, and nanoscale architecture[36,37]. This diversity, termed the synaptome, and its 3D organisation—the synaptome architecture—provide a framework for understanding how synaptic gene mutations target specific synapse types in defined brain regions. Labelling studies show that proteins implicated in related cognitive disorders exhibit both overlapping and unique expression across synapse types, with distributions varying across brain regions. Moreover, synaptome mapping across the lifespan shows dynamic changes in synapse distributions, potentially explaining why certain mutations produce distinct phenotypes at different ages[37,38]. These observations emphasise that spatial and temporal expression of synaptic proteins is a critical determinant of phenotype, and they underscore the importance of synaptome mapping in human brain tissue, a now feasible approach[39].

In conclusion, the consensus proteome lists have broad utility in genetic, transcriptomic, and proteomic research, and potentially in drug discovery. We plan annual updates to SynaptopathyDB, along with the continued expansion of its functionality, to support classification and mechanistic understanding of the synaptic basis of brain disease. By integrating large-scale datasets that link synapse proteomes and their spatial organisation to genetic variation, we aim to advance understanding of the molecular foundations of the human behavioural repertoire and its disorders.

Methods
Synaptic proteome database
Synaptic studies for presynaptic, postsynaptic, synaptic vesicle and whole synaptosome compartment were identified by manual curation of PubMed, starting from year 2000, which resulted in 64 studies[24]. Protein/gene lists from the published synaptic proteome studies were combined into the total list of synaptic components. Identifiers from each study were mapped to stable IDs including: Entrez Human and Mouse, Uniprot and MGI IDs. Each protein is associated with literature ID and curated list of metadata such as species (mouse, human, rat), respective synaptic compartment, brain, experimental methods, Gene Ontology (GO) and disease associations (see below). The synaptic proteome datasets are available in SQLite format[24]. Users can convert gene lists from other species using widely available orthology mapping tools in NCBI or Bioconductor packages such as Orthogene.

Building the consensus synaptic protein list
Using the synapse proteome database[24] we have selected the human genes that were identified at least five times in any of the following synaptic compartments: postsynaptic, presynaptic, whole synaptosome, synaptic vesicle. The choice of this threshold was informed by the distribution of gene identifications in Sorokina et al., 2021[16] where there was a clear inflection around the value 5. We assume that, given threshold guarantees, the specific protein is highly likely to be associated with this compartment, and excluded the proteins with a few identifications across multiple compartments as potential contaminants.

Disease annotation
We used gene-disease annotation data collected from the OMIM[17], GeneRIF[40] and Ensembl variation[41] and Alliance Genome Resource[42] databases. Using the topOnto package, the annotation data were standardised using MetaMap[43] and NCBO Annotator[44,45] to recognize terms found in the Human Disease Ontology (HDO)[46,47]. Recognised disease ontology terms were then associated with gene identifiers and stored locally.

For each of the consensus genes we extracted gene and phenotype entries from OMIM using the mim2gene file linking via their Entrez Gene identifiers. Diseases associated with SynProteome^CONS proteins were further classified using information in the OMIM entry (Supplementary Table S1), into CNS, PNS, both, or other. All diseases were then grouped into chapters of the International Classification of Disease (ICD-11) developed by the World Health Organisation (WHO). Those OMIM diseases that could not readily be found in ICD-11 were assigned to one of the ICD-11 chapters based on their symptoms, if possible. Diseases in Chap. 8 (Diseases of the nervous system) were divided into more specific blocks within that chapter.

Human phenotype ontology gene set enrichment analysis
Gene lists representing human illnesses and conditions were downloaded from the Human Phenotype Ontology (HPO). The HPO database curates OMIM diseases and genes into a hierarchical format with more higher-level terms representing more general clusters of phenotypes. Genes linked to each condition are also documented in the HPO. This dataset was downloaded on January 20, 2025. Enrichment of phenotype genes in the SynProteome^CONS relative to the genome was assessed on phenotypes that had more than 20 genes in the genome. Significances of overrepresentation of disease conditions in the SynProteome^CONS were assessed by hypergeometric statistics corrected by the Benjamini-Hochberg false discovery rate, implemented as *fora* function in *gsea* R package. For comparison, a non-synaptic protein list was built by extraction of our latest total synaptic list (10,500 genes, not shown) from whole human genome (NCBI database) downloaded in January 2025, which resulted in 9,667 genes. Overrepresentation of each phenotype relative to the genome was computed separately for synaptic (SynProteome^CONS) and for the non-synaptic gene lists.

Web-based resource

SynaptopathyDB is implemented as a comprehensive web application with a user-friendly interface designed to facilitate exploration of synaptic protein-disease relationships. The platform provides the following primary search interfaces:

Unified search

The homepage features a unified search functionality that allows users to simultaneously search across genes, diseases, and publications. The autocomplete feature provides real-time suggestions, supporting multiple identifier types including:

- Gene names and identifiers (Human/Mouse/Rat Entrez, Protein, MGI, Ensembl ID, Alias)[29–31].
- Disease names and ontology identifiers.
- Publication details (PubMed ID, title, year).

Gene search

The gene search interface enables comprehensive exploration of individual genes, providing:

- Detailed gene information including multiple species identifiers[29–31].
- Associated publications.
- Disease associations/mutation details.
- Localization within synaptic compartments.
- External cross-references to databases including AllianceGenome, Disease Ontology, HGNC, NCBIGene, HPO, PubMed, QuickGo and OMIM.
- Filtering options by consensus status.
- Downloadable data in CSV format.

Disease search

The disease search interface allows users to:

- Explore diseases and their associated synaptic genes.
- View detailed mutation information.
- Filter genes by consensus status.
- External cross-references to authoritative databases including AllianceGenome, Disease Ontology, NCBIGene, HPO and PubMed.
- Access publication links.
- Generate downloadable reports.

Publication search

The publication search feature provides:

- Comprehensive details of synapse proteomics studies.
- Filtering of associated genes and mutations.
- Localization and method information.
- Interactive pagination and csv export capabilities.

Catalogues

The database provides catalogues of all genes, diseases, and publications:

- Synapse Proteome Catalogue.
- Diseases Catalogue.
- Publications Catalogue.

Data visualization and analysis tools

SynaptopathyDB offers several chart visualizations:

- Disease classification trends.
- Protein localization history.
- Discovery rate analyses.
- Compartment-specific protein distributions.

Technical infrastructure

- Frontend: React.js with Tailwind CSS.
- Backend: Flask (Python).

- Database: SQLite.
- Interactive components powered by Recharts and other visualization libraries.

API and data access
The platform provides RESTful API endpoints for:

- Programmatic data retrieval.
- CSV export of comprehensive datasets.
- Flexible querying across genes, diseases, and publications.

The web application is designed to be intuitive, responsive, and accessible across different devices, supporting researchers in exploring the complex relationships between synaptic proteins, genetic variations, and human diseases.

Data availability

For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Synapse proteome data are available from Edinburgh Data-Share at [https://doi.org/10.7488/ds/3771].

References

1. Husi, H., Ward, M. A., Choudhary, J. S., Blackstock, W. P. & Grant, S. G. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nat. Neurosci.* **3**, 661–669. https://doi.org/10.1038/76615 (2000).
2. Bayes, A. et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21. https://doi.org/10.1038/nn.2719 (2011).
3. Fernandez, E. et al. Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol. Syst. Biol.* **5**, 269. https://doi.org/10.1038/msb.2009.27 (2009).
4. Fromer, M. et al. De Novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184. https://doi.org/10.1038/nature12929 (2014).
5. Kirov, G. et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry*. **17**, 142–153. https://doi.org/10.1038/mp.2011.154 (2012).
6. Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190. https://doi.org/10.1038/nature12975 (2014).
7. Song, L., Chen, W., Hou, J., Guo, M. & Yang, J. Spatially resolved mapping of cells associated with human complex traits. *Nature* **641**, 932–941 https://doi.org/10.1038/s41586-025-08757-x (2025).
8. Skene, N. G., Roy, M. & Grant, S. G. A genomic lifespan program that reorganises the young adult brain is targeted in schizophrenia. *Elife* **6**, e17915 https://doi.org/10.7554/eLife.17915 (2017).
9. Als, T. D. et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844. https://doi.org/10.1038/s41591-023-02352-1 (2023).
10. Goes, F. S. et al. Large-scale transcriptomic analyses of major depressive disorder reveal convergent dysregulation of synaptic pathways in excitatory neurons. *Nat. Commun.* **16**, 3981. https://doi.org/10.1038/s41467-025-59115-4 (2025).
11. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352. https://doi.org/10.1038/s41593-018-0326-7 (2019).
12. Adams, M. J. et al. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell* **188**, 640–652 e649. https://doi.org/10.1016/j.cell.2024.12.002 (2025).
13. Coleman, J. R. et al. Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Mol. Psychiatry*. **24**, 182–197 (2019).
14. Deary, I. J., Cox, S. R. & Hill, W. D. Genetic variation, brain, and intelligence differences. *Mol. Psychiatry*. **27**, 335–353 (2022).
15. Hill, W. D. et al. Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl. Psychiatry*. **4**, e341–e341 (2014).
16. Sorokina, O. et al. A unified resource and configurable model of the synapse proteome and its role in disease. *Sci. Rep.* **11**, 9967. https://doi.org/10.1038/s41598-021-88945-7 (2021).
17. Online Mendelian Inheritance in Man, OMIM. (2024).
18. Amberger, J., Bocchini, C. A., Scott, A. F. & Hamosh, A. McKusick's online mendelian inheritance in man (OMIM). *Nucleic Acids Res.* **37**, D793–796. https://doi.org/10.1093/nar/gkn665 (2009).
19. Harrison, J. E., Weber, S., Jakob, R. & Chute, C. G. ICD-11: an international classification of diseases for the twenty-first century. *BMC Med. Inf. Decis. Mak.* **21**, 1–10 (2021).
20. Köhler, S. et al. The human phenotype ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).
21. Aleksander, S. A. et al. The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
22. Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234 (2019).
23. UniProt. The universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
24. Synapse proteome database. (https://datashare.ed.ac.uk/handle/10283/4757). (2025).
25. Bayes, A. et al. Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLoS One*. **7**, e46683. https://doi.org/10.1371/journal.pone.0046683 (2012).
26. Bayes, A. et al. Evolution of complexity in the zebrafish synapse proteome. *Nat. Commun.* **8**, 14613. https://doi.org/10.1038/ncomms14613 (2017).
27. Emes, R. D. et al. Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.* **11**, 799–806. https://doi.org/10.1038/nn.2135 (2008).
28. Ramos-Vicente, D., Grant, S. G. & Bayes, A. Metazoan evolution and diversity of glutamate receptors and their auxiliary subunits. *Neuropharmacology* **195**, 108640. https://doi.org/10.1016/j.neuropharm.2021.108640 (2021).
29. Lelong, S. et al. BioThings SDK: a toolkit for building high-performance data APIs in biomedical research. *Bioinformatics* **38**, 2077–2079. https://doi.org/10.1093/bioinformatics/btac017 (2022).
30. Wu, C., Macleod, I. & Su, A. I. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.* **41**, D561–565. https://doi.org/10.1093/nar/gks1114 (2013).

31. Xin, J. et al. High-performance web services for querying gene and variant annotation. *Genome Biol.* **17**, 91. https://doi.org/10.1186/s13059-016-0953-9 (2016).
32. Frank, R. A. et al. NMDA receptors are selectively partitioned into complexes and supercomplexes during synapse maturation. *Nat. Commun.* **7**, 11264 (2016).
33. Komiyama, N. H. et al. Synaptic combinatorial molecular mechanisms generate repertoires of innate and learned behavior. *bioRxiv* 500389 https://doi.org/10.1101/500389 (2018).
34. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
35. Roy, M. et al. Proteomic analysis of postsynaptic proteins in regions of the human neocortex. *Nat. Neurosci.* **21**, 130–138 (2018).
36. Zhu, F. et al. Architecture of the mouse brain synaptome. *Neuron* **99**, 781–799e710 (2018).
37. Bulovaite, E. et al. A brain atlas of synapse protein lifetime across the mouse lifespan. *Neuron* **110**, 4057–4073e4058 (2022).
38. Cizeron, M. et al. A brainwide atlas of synapses across the mouse life span. *Science* **369**, 270–275 (2020).
39. Curran, O. E. et al. A single-synapse resolution survey of PSD95-positive synapses in twenty human brain regions. *Eur. J. Neurosci.* **54**, 6864–6881 (2021).
40. Mitchell, J. et al. (ed, A.) Gene indexing: characterization and analysis of NLM's GeneRIFs. *AMIA Annu. Symp. Proc.* **2003** 460–464 (2003).
41. Chen, Y. et al. Ensembl variation resources. *BMC Genomics* **11**, 293. https://doi.org/10.1186/1471-2164-11-293 (2010).
42. Kishore, R. et al. Automated generation of gene summaries at the alliance of genome resources. *Database* https://doi.org/10.1093/database/baaa037 (2020).
43. Aronson, A. R. & Lang, F. M. An overview of metamap: historical perspective and recent advances. *J. Am. Med. Inf. Assoc.* **17**, 229–236. https://doi.org/10.1136/jamia.2009.002733 (2010).
44. Musen, M. A. et al. The National center for biomedical ontology. *J. Am. Med. Inf. Assoc.* **19**, 190–195. https://doi.org/10.1136/amiajnl-2011-000523 (2012).
45. Whetzel, P. L. et al. BioPortal: enhanced functionality via new web services from the National center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Res.* **39**, W541–545. https://doi.org/10.1093/nar/gkr469 (2011).
46. Schriml, L. M. & Mitraka, E. The disease ontology: fostering interoperability between biological and clinical human disease-related data. *Mamm. Genome.* **26**, 584–589. https://doi.org/10.1007/s00335-015-9576-9 (2015).
47. Kibbe, W. A. et al. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.* **43**, D1071–1078. https://doi.org/10.1093/nar/gku1011 (2015).

Author contributions
O.S. built the database and analyzed data. D.D. built the website. A.B. contributed to disease analysis. J.D.A. and S.G.N.G. supervised the project. S.G.N.G. wrote the manuscript and all authors edited the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.G.N.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.