# *SLC6A1* variant pathogenicity, molecular function, and phenotype: a genetic and clinical analysis

Arthur Stefanski,[1] Eduardo Pérez-Palma,[2] Tobias Brünger,[3] Ludovica Montanucci,[1] Cornelius Gati,[4] Chiara Klöckner,[5] Katrine M. Johannesen,[6,7] Kimberly Goodspeed,[8,9] Marie Macnee,[3] Alexander T. Deng,[10] Ángel Aledo-Serrano,[11] Artem Borovikov,[12] Maina Kava,[13,14] Arjan M. Bouman,[15] M. J. Hajianpour,[16] Deb K. Pal,[17,18] Marc Engelen,[19] Eveline E. O. Hagebeuk,[20] Marwan Shinawi,[21] Alexis R. Heidlebaugh,[22] Kathryn Oetjens,[22] Trevor L. Hoffman,[23] Pasquale Striano,[24,25] Amanda S. Freed,[26] Line Futtrup,[27] Thomas Balslev,[27,28] Anna Abulí,[29] Leslie Danvoye,[30] Damien Lederer,[31] Tugce Balci,[32,33] Maryam Nabavi Nouri,[34] Elizabeth Butler,[35] Sarah Drewes,[36] Kalene van Engelen,[37] Katherine B. Howell,[38,39,40] Jean Khoury,[1] Patrick May,[41] Marena Trinidad,[42] Steven Froelich,[42] Johannes R. Lemke,[5,43] Jacob Tiller,[44] Amber N. Freed,[44] Jing-Qiong Kang,[45,46,47,48,49] Arthur Wuster,[42] Rikke S. Møller[6,50] and Dennis Lal[1,51]

## Abstract

Genetic variants in the *SLC6A1* gene can cause a broad phenotypic disease spectrum by altering the protein function. Thus, systematically curated clinically relevant genotype-phenotype associations are needed to understand the disease mechanism and improve therapeutic decision-making.

We aggregated genetic and clinical data from 172 individuals with likely pathogenic/ pathogenic (lp/p) *SLC6A1* variants and functional data for 184 variants (14.1% lp/p). Clinical and functional data were available for a subset of 126 individuals. We explored the potential associations of variant positions on the GAT1 3D structure with variant pathogenicity, altered molecular function, and phenotype severity using bioinformatic approaches.

The GAT1 transmembrane domains 1, 6, and extracellular loop 4 (EL4) were enriched for patient over population variants. Across functionally tested missense variants ($n$ = 156), the spatial proximity from the ligand was associated with loss-of-function in the GAT1 transporter activity. For variants with complete loss of *in vitro* GABA uptake, we found a 4.6-fold

enrichment in patients having severe disease vs. non-severe disease ($P$ = 2.9e$^{-3}$, 95% CI: 1.5 - 15.3).

In summary, we delineated associations between the 3D structure and variant pathogenicity, variant function, and phenotype in *SLC6A1*-related disorders. This knowledge supports biology-informed variant interpretation and research on GAT1 function. All our data can be interactively explored in the SLC6A1 Portal (https://slc6a1-portal.broadinstitute.org/).

**Author affiliations:**

1 Genomic Medicine Institute & Epilepsy Center, Cleveland Clinic, Cleveland, OH 44195, USA

2 Universidad del Desarrollo, Centro de Genética y Genómica, Facultad de Medicina Clínica Alemana, Santiago 7610658, Chile

3 Cologne Center for Genomics (CCG), Medical Faculty of the University of Cologne, University Hospital of Cologne, Cologne 50931, Germany

4 Department of Biological Sciences, Bridge Institute, USC Michelson Center for Convergent Bioscience, University of Southern California, Los Angeles, CA 90089, USA

5 Institute of Human Genetics, University of Leipzig Medical Center, Leipzig 04103, Germany

6 Department of Epilepsy Genetics and Personalized Medicine, The Danish Epilepsy Centre, Dianalund 4293, Denmark

7 Department of Genetics, University Hospital of Copenhagen, Rigshispitalet, Copenhagen 2100, Denmark

8 Children's Health, Medical Center, Dallas, TX 75235, USA

9 University of Texas Southwestern Medical Center, Department of Pediatrics, Dallas, TX 75390, USA

10 Clinical Genetics, Guys and St Thomas NHS Trust, Great Maze Pond, London SE19RT, UK

11 Epilepsy Program, Neurology Department, Hospital Ruber Internacional, Madrid 28034, Spain

12 Research Centre for Medical Genetics, Moscow 115478, Russia

13 Department of Neurology and Metabolic Medicine, Perth Children's Hospital, Perth 6009, Australia

14 School of Paediatrics and Child Health, UWA Medical School, University of Western Australia, Perth 6009, Australia

15 Department of Clinical Genetics, Erasmus MC, University Medical Center, Rotterdam 3015GD, The Netherlands

16 Department of Pediatrics, Division of Medical Genetics and Genomics, Albany Medical College, Albany Med Health System, Albany, NY 12208, USA

17 Department of Basic & Clinical Neurosciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London SE58AF, UK

18 King's College Hospital, London SE59RS, UK

19 Department of Pediatric Neurology, Amsterdam Public Health, Amsterdam University Medical Center, Amsterdam 1081HV, The Netherlands

20 Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede and Zwolle 2103SW, The Netherlands

21 Division of Genetics and Genomic Medicine, Department of Pediatrics, St. Louis Children's Hospital, Washington University School of Medicine, St. Louis, MO 63110, USA

22 Autism & Developmental Medicine Institute, Geisinger, Danville, PA 17837, USA

23 Southern California Kaiser Permanente Medical Group, Department of Regional Genetics, Anaheim, CA 92806, USA

24 Pediatric Neurology and Muscular Diseases Unit, IRCCS Istituto Giannina Gaslini, Genoa 16147, Italy

25 Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genoa 16132, Italy

26 Department of Clinical Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA 91101, USA

27 Department of Paediatrics, Regional Hospital of Central Jutland, Viborg 8800, Denmark

28 Centre for Educational Development, Aarhus University, Aarhus 8200, Denmark

29 Department of Clinical and Molecular Genetics and Medicine Genetics Group, VHIR, University Hospital Vall d'Hebron, Barcelona 08035, Spain

30 Department of Neurology, Université catholique de Louvain, Cliniques universitaires Saint-Luc, Brussels 1200, Belgium

31 Centre for Human Genetics, Institute for Pathology and Genetics, Gosselies 6041, BE

32 Department of Pediatrics, Division of Medical Genetics, Western University, London, ON N6A3K7, Canada

33 Medical Genetics Program of Southwestern Ontario, London Health Sciences Centre and Children's Health Research Institute, London, ON N6A5A5, Canada

34Department of Paediatrics, Division of Pediatric Neurology, London Health Sciences Centre, London, ON N6A5W9, Canada

35 GeneDx, Gaithersburg, MD 20877, USA

36 Department of Medical Genetics, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA 15224, USA

37 Medical Genetics Program of Southwestern Ontario, London Health Sciences Centre, London, ON N6A5W9, Canada

38 Department of Neurology, Royal Children's Hospital, Melbourne, VIC 3052, Australia

39 Department of Pediatrics, University of Melbourne, Melbourne, VIC 3052, Australia

40 Murdoch Children's Research Institute, Melbourne, VIC 3052, Australia

41 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4362, Luxembourg

42Translational Genomics, BioMarin Pharmaceutical Inc., Novato, CA 94949, USA

43 Center for Rare Diseases, University of Leipzig Medical Center, Leipzig 04103, Germany

44 SLC6A1 Connect, Frisco, TX 75034, USA

45 Department of Neurology, Vanderbilt University Medical Center, Nashville, TN 37240, USA

46 Neuroscience Graduate Program, Vanderbilt University, Nashville, TN 37235, USA

47 Vanderbilt Brain Institute, Nashville, TN 37235, USA

48 Department of Pharmacology, Vanderbilt University, Nashville TN 37232, USA

49 Vanderbilt Kennedy Center of Human Development, Nashville, TN 37203, USA

50 Department of Regional Health Research, University of Southern Denmark, Odense 5000, Denmark

51 Stanley Center of Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

Correspondence to: Dennis Lal, PhD

Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic

Cleveland, OH 44195, USA

E-mail: lald@ccf.org

**Running title**: Variant pathogenicity in *SLC6A1*

**Keywords:** autism; epilepsy; neurodevelopmental disorder; genetics; SLC6A1

# Introduction

*SLC6A1* encodes for the GABA transporter protein type 1 (GAT1), a membrane protein responsible for GABA neurotransmitter reuptake from the synaptic cleft in inhibitory synapses.[1] *SLC6A1*-related developmental and epileptic encephalopathy (DEE) is an autosomal dominant genetic disorder. Clinical manifestation of *SLC6A1* DEE is characterized by childhood onset seizures and mild to severe intellectual disability. Seizure types include absence, myoclonic and atonic. Language impairment and behavioral problems have also been observed.[2–4] Other frequently observed *SLC6A1*-related phenotypes include autism spectrum disorder (ASD) and motor dysfunction, encompassing stereotypies and ataxia. A fraction of patients have shown

intellectual disability or ASD without epilepsy (3%).[2] Recent GAT1 analyses support complete or partial loss-of-function (LoF) as the primary disease-associated molecular pathology, which disrupts the reuptake of GABA.[5–9]

Despite recent aggregation efforts,[2,5,10,11] there is a need for systematically curated clinically relevant genotype-phenotype associations to understand the disease mechanism and possibly guide genetic counseling, patient management, and, ultimately, treatment. It has been shown that 32 out of the 88 (36.4%) described *SLC6A1* patient variants are located in the helical-transmembrane segments and inter-helical hinges. In contrast, general population variants cluster in the cytoplasmic domain.[2] An analysis using the GAT1 3D structure may increase the granularity of these preliminary observations and identify clinically relevant variant-to-phenotype or variant-to-function associations. 3D structure analysis has previously been successful in elucidating genotype-phenotype associations in various genes.[12–17] An investigation into gene variant effects across sodium channelopathies showed clustering of pathogenic missense variants in functional domains.[18–22] However, due to limited available patient data for most *SLC6A1* variants, meaningful associations have been difficult to establish. Currently, *SLC6A1* variant interpretation is still challenging as, to date, there is no single resource with aggregated and curated data for *SLC6A1*-related disorders. Previous studies have suggested that transmembrane segments are important for protein function.[2,4,5,10,11,23] However, clear guidance on which segments or subdomains are particularly affected is lacking. A recent study on the molecular mechanism of *SLC6A1* variants, investigating 182 variants, showed that LoF variants are found predominantly around the proteins' vertical axis.[10,11] A relationship between transporter activity and literature-based disease association has been recently proposed.[11] However, statistical confirmation of phenotype and variant location needs yet to be established.

The complexity and heterogeneity of *SLC6A1*-related disorders pose difficulties in disease recognition, diagnosis, prognosis, and care. The spatial analysis of genetic variants on 3D protein structures has the potential to identify genotype-phenotype correlations, as has been shown in other related neurodevelopmental disorders.[24–30] As 'phenotype' for the analysis, clinical data of variant carriers or molecular readouts generated for the variant can be used to study the effect of different variants.[30] However, this type of work requires large datasets from various sources. In our study, we build upon previous data aggregation efforts and bioinformatic

methods and present the currently most extensive effort to investigate genotype to phenotype associations for *SLC6A1*-related disorders.

Here, we aggregated the currently largest collection of individuals with *SLC6A1*-related disorders and implemented a 3D-based framework[28] to evaluate genetic, clinical, and functional features. Our study compiled a comprehensive dataset of pathogenic and likely pathogenic *SLC6A1* variants from the literature, ClinVar,[31] and our clinical research network. We also incorporated population variants from gnomAD as controls for comparative analysis (gnomAD, public release 2.1.1). Subsequently, we performed linear sequence and 3D protein structure-based genotype-phenotype analysis using in vitro assay and clinical phenotype data to identify structure, to function to phenotype relationships for *SLC6A1*-related disorders. Finally, we deployed all data and analysis tools into the SLC6A1 Portal, a joint effort of clinical and basic science investigators in collaboration with advocacy groups, to enhance further analysis, awareness, and variant interpretation of *SLC6A1*-related disorders.

# Materials and methods

## Genotype and phenotype data from patients with *SLC6A1*-related disorders

We aggregated published genetic and corresponding phenotype data from *SLC6A1*-related disorder studies.[2,5,6,11] Investigators provided unpublished genetic and phenotype data ($n$ = 51) from the Danish Epilepsy Centre, Filadelfia, Denmark (Dr. Katrine M Johannesen and Dr. Rikke S Møller). We also included genetic and syndrome-level data from the Epi25 Collaborative for Large-Scale Whole Exome Sequencing in the Epilepsy Collaborative database.[32] Epi25 data are limited to genotype and International League Against Epilepsy (ILAE) syndrome categorization. The data for all patient variants ($n$ = 172) that was evaluated, curated, and harmonized in collaboration with clinical experts, including comprehensive annotations, can be viewed in Supplementary Table 1.

The functional data were aggregated from two recent studies.[5,11] One study quantified GABA uptake for 182 variants from 15 cohorts, including individuals with epilepsy,

developmental disorders, and healthy controls. The dataset contains pathogenic and likely pathogenic variants, variants of uncertain significance, variants that had been classified as benign or likely benign, and variants that were unclassified or had conflicting annotations.[11] Additionally, we included functional readouts of two variants (p.Pro361Thr & p.Leu73Phe) from a recent publication.[5] Patients or their legal guardians provided signed informed consent according to the Declaration of Helsinki and local IRB requirements.

## Genotype data from public repositories

We retrieved general population *SLC6A1* missense variants ($n$ = 158) from the genome Aggregation Database (gnomAD, public release 2.1.1) in Variant Call Format.[33] Missense variant annotation was performed with Variant Effect Predictor (VEP)[34], including information from public repositories.[33] Pathogenic variation in *SLC6A1*-related disorders is mostly de novo and rarely expected to be found in general population repositories such as gnomAD.[2] Thus, we used general population variants from the gnomAD database as controls. Although most variants are expected to be fully penetrant, we also calculated a gnomAD frequency cut-off for ultra-rare *SLC6A1* disorder variants with incomplete penetrance using the cardiodb allele frequency app.[35] We accessed pathogenic and likely pathogenic ClinVar[31] missense variants from the FTP site (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/) (ClinVar, July 2021). We obtained ClinVar variants classified as of uncertain significance (VUS) from the FTP site (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/) (ClinVar, December 2022). All genetic variants were mapped onto the canonical isoform, P30531, as defined by the UniProt database (The UniProt Consortium, 2021).

## Domain-specific analysis: mapping variants onto the 3D protein structure

We obtained the human wild-type GABA transporter type 1 3D structure from the Protein Data Bank (PDB ID: 7SK2).[23] The variants were mapped onto the structure using PyMOL.[36] For each residue, we calculated a normalized functional score. First, we annotated the functional scores on

the GAT1 protein structure using the bio3d R-package.[37] Second, we normalized the functional activity by calculating the average functional score reported across all residues located within a 5Å radius. We define the distance from the ligand as the distance in Angstrom (Å) between the variant wild-type residue and the ligand. Since no GABA is bound to the GAT1 7SK2 protein structure, we calculated the minimum distance in Angstrom (Å) between the variant wild-type residue and Tiagabine. This GAT1 inhibitor is bound to the GAT1 structure at the GABA binding site (https://www.rcsb.org/structure/7SK2). We considered all atoms of each protein residue and the Tiagabine for the minimum distance calculation.

## Functional data curation

We next aggregated data for 184 electrophysiologically tested variants,[5,11] for which the average transporter activity has been experimentally measured. Both studies have employed a radiolabeled assay to measure the GABA reuptake activity in HEK293T cells. However, Mermer[5] used scintillation counting to quantify the amount of radiolabeled GABA taken up by the cells, whereas Trinidad[11] used mass spectrometry to create a high-throughput GABA trafficking assay. Despite the methodological variations between the two studies, the deviations from previous percent-wild-type (WT) levels were minimal. Mermer[5] conducted their analysis without utilizing two mass spectrometry detectors (MSMS), employed different cell lines that lacked the CRISPER-Cas9 *SLC6A1*-knockout present in Trinidad's[11] research, and did not account for variable expression efficiencies using the Beta-lactamase (BLA)-reporter, as done by Trinidad.[11] Each variants' transporter activity is reported as a percentage of the wild-type activity.[5,11] All functionally tested pathogenic variants but one (p.Val342Met) showed an LoF effect with an average WT activity below 42.8%, relative to WT activity. The threshold for LoF (42.8%) has been derived from the observed behavior of ClinVar variants predicted to be synonymous or classified as benign.[11] To date, pathogenic or likely pathogenic gain-of-function has not been reported.[11] We stratified all variants into three activity groups based on their average relative-to-WT GABA uptake activity: (1) 0-10% - nearly complete LoF, (2) 10-42.8% - low activity, and (3) >42.8% - WT.

## Phenotypic data curation

In collaboration with clinical experts, we summarized and harmonized the cognitive and syndrome level data into six cognitive and twelve epilepsy syndrome categories, respectively (Cognitive level: Severe DD/ID, Moderate DD/ID, Mild DD/ID, Learning disability, Unclassified DD, Normal; Epilepsy syndrome: Childhood Absence Epilepsy (CAE), Developmental and epileptic encephalopathy (DEE), Epilepsy with Myoclonic-Atonic Seizures (EMAS), Intractable primary Generalized Epilepsy, Generalized Epilepsy, Genetic Generalized Epilepsy (GGE), Intractable Absence Epilepsy, Lennox-Gastaut syndrome (LGS), Non-Acquired Focal Epilepsy (NAFE), Temporal Lobe Epilepsy (TLE), unclassified epilepsy, and No seizures. The classifications were regrouped for consistency by experienced epileptologists (K.M.J. and K.G.). We opted for a binary categorization for disease severity and activity into (1) severe disease and (2) non-severe disease. A clinical diagnosis of one of the following syndromes indicates a severe disease: DEE, EMAS, LGS, and intractable absence epilepsy. All the patients diagnosed with one of the four syndromes were considered severe because they impose serious life challenges due to their high seizure burden, often with significant developmental delay, and are typically resistant to many seizure medications. There was no normal cognition reported in individuals with EMAS after seizure onset. Individuals with no seizures, a diagnosis of CAE, unclassified epilepsy, generalized epilepsy, GGE, and TLE or NAFE were classified as having a non-severe disease. Individuals for which no syndrome level data were available or a binary categorization impossible were classified as 'other'. The complete regrouping and reclassification can be found in Supplementary Table 1.

## Portal Design

The SLC6A1 Portal utilizes the Shiny R framework from RStudio (https://shiny.rstudio.com/) to build the interactive web Portal for compatibility, expendability, and portability. The Portal is publicly available and hosted at the Broad Institute and was deployed with Google Cloud service using a self-contained Docker image (https://slc6a1-portal.broadinstitute.org/). The Portal code is available on GitHub (https://github.com/LalResearchGroup/SLC6A1_Portal). We produced a short educational video with whiteboard animation and utilized the software *VideoScribe* to

increase the accessibility to knowledge about *SLC6A1*-related disorders (VideoScribe 3.9.5, Sparkol 2012: https://www.videoscribe.co/en/download/).

# Results

## Data aggregation and description

We present genotype and phenotype data from the largest cohort of individuals with *SLC6A1*-disorders to date, including 172 individuals (DS-172) with 94 unique variants, three copy number variants (CNVs), and 19 recurring variants found in 75 patients (For details on the cohort, see Supplementary Tables 1, 2, and 3). Our clinical dataset of 172 *SLC6A1* variants contains 51 variants that have not been previously published (Supplementary Table 1 & 2). The most frequent variant in our cohort is p.Val342Met, which was identified in 11 patients. In addition, we report on *in vitro* GAT1 functional readouts for 184 unique variants (DS-184) from two sources.[5,11] For 70 variants from 126 individuals (DS-126), clinical and *in vitro* transporter function data was available. For 57 variants from 79 individuals (DS-79), clinical and variant information, including syndrome classification, was available together with functionally tested variants to investigate the relationship between disease severity and function (Supplementary Figure 1 & Supplementary Table 1). Additionally, we mapped 162 out of 195 ClinVar missense variants of uncertain significance (VUS) onto the GAT1 3D protein structure and observed that those variants are dispersed throughout the protein's structure (Supplementary Figure 2).

We obtained general population missense variants from gnomAD ($n$ = 158) as controls for our analyses. The majority of gnomAD variants included do not overlap with patient variants, and only 11 variants (4%) overlap between our patient dataset ($n$ = 172) and gnomAD. To further explore these variants, we calculated the maximum allele frequency in gnomAD (Maximum credible population AF = $6.05e^{-05}$) and added the results to Supplementary Table 1.[35] The maximum credible population AF was determined using an estimated disease prevalence of 1/619.6 (161.38/100k),[38] an incidence of 2.65 per 100,000 live births,[39] and a disease duration of 60.9 years[40] (1.65 x 60.9 = 100.485). Because the lifespan of individuals with *SLC6A1*-related

disorders is unavailable, we selected a population-based estimation of life years lost. In our cohort, the largest proportion of cases is attributable to the missense variant p.Val342Met, found in 11 of 172 *SLC6A1* disorder patients. The allelic heterogeneity was therefore estimated as 0.06 (11/172), while the genetic heterogeneity, representing the number of genes associated with the disorder, was set to 1 as our study inclusion criterion is a variant in *SLC6A1*. For the above calculation, we estimated an 80% variant penetrance given that few patients have been reported with an inherited variant. Using these parameters to calculate the maximum credible population frequency, we identified a gnomAD variant cut-off of $6.05e^{-05}$. None of the eleven patient variants that were also present in gnomAD exceeded the threshold set by the maximum credible allele filter.

## Data sharing through the SLC6A1 Portal

All the aggregated datasets are integrated to enable scientists to use our rich data source for their research studies and educate providers and families on *SLC6A1*-related disorders. The datasets can be explored in the SLC6A1 Portal (https://slc6a1-portal.broadinstitute.org/) (Figure 1), an interactive and user-friendly web application that combines genetic and clinical data of individuals with *SLC6A1*-related disorders with experimental functional and annotation data on variants. Users can navigate through four sections: (1) Basic Information, (2) Educational Resources, (3) Variant Analysis, and (4) Research. By utilizing these data within the Portal infrastructure, we enable the exploration of genotype to structure, function, and phenotype associations (Supplementary Figure 4). Here, we present the SLC6A1 Portal, which provides access to the largest cohort of patients with *SLC6A1*-related disorders, including their clinical phenotypes and the largest dataset for in vitro GAT1 functional readouts. We include a versatile variant entry interface and a visual comparison tool that shows variant location and molecular activity within the GAT1 3D protein structure. Additionally, the portal includes a domain-wide comparison of patient vs. population variants and a functional interface for data analysis, including tools to display each variants' distance from Tiagabine vs. GABA uptake rate and to identify hot zones on the GAT1 3D structure based on user-selected variant filters. All aggregated data is shared according to the FAIR principles to make it findable, accessible, interoperable, and reusable.[41]

## Structure to clinical phenotype relationship

To elucidate critical regions for transporter function, we retrieved, for each GABA transporter type 1 (GAT1) residue, its functional domain as classified on the recently published human structure (PDB ID: 7SK2).[23] We investigated the GABA uptake activity of *SLC6A1* variants by domain and found varying activity levels.[5,11] Our final dataset contains functional readouts for 156 missense variants (DS-156). It shows that the transmembrane helices 1/6 (TM1/6), scaffold, and extracellular loop 4 (EL4) regions harbor 83.3% of variants with low activity (<42.8%), and 16.7% of variants with WT activity (>42.8%) ($n$ = 40 variants vs. $n$ = 8 variants with *in vitro* transporter activity below 42.8%).[11] The scaffold in GAT1 comprises helices H3 and H8 and Linkers H3-H4 and H8-H9 primarily located in the transmembrane region and serves scaffolding functions.[23,42] The domain with the lowest average transporter activity of tested variants was scaffold (18.6% average GAT1 activity); the second and third lowest were TM1/6 and EL4, with 23% and 31.1% average activity, respectively. The fourth lowest is Extracellular loop 3 (EL3), with 38.2% average activity. The Linker, TMD-other, and N-terminal domains have an average activity above the threshold of WT activity at 42.8% with 44%, 45.9%, and 49.4%, respectively, but harbor variants with a wide range of activity levels ranging from nearly complete LoF to WT. In contrast, Extracellular loop 2 (EL2) and C-terminal regions showed no change in their activity, with 56.9% and 90.5% of average activity, respectively (Figure 2 A). We found no enrichment truncating/frameshift variants across different GAT1 domains. Next, we compared the enrichment of patient vs. population variants per region in the *SLC6A1* gene to identify those regions that are predominantly affected by patient variants. For variants that are in the TM1/6 and EL4 regions ($n$ = 51), we found an 8.7 and 8.5-fold enrichment of patient vs. population variants (TM1/6: 95% CI: 3.6 - 23.9, $P = 7.5e^{-9}$ and EL4: 95% CI: 1.9 - 78.7, $P = 1.1e^{-3}$; Figure 2 B). Variants in the TM1/6 and EL4 were in agreement with their functional data presented in Figure 2 A. Variants in both the N- and C-terminal regions were depleted for patient variants with an odds ratio of 0.22 for N-terminal (95% CI: 0.06 - 0.61, $P = 1.3e^{-3}$) and 0.14 for C-terminal (95% CI: 0.03 - 0.5, $P = 2.8e^{-4}$). However, only variants in the C-terminal region were concordant with the functional data, as the N-terminal region harbored variants with a wide range of average transporter activity. We observed no enrichment for patient variants in all other

regions. The domain-wide analysis of patient vs. population variants identified the TM1/6 and EL4 regions as the most essential and the N-/C-terminal as least essential for GABA transport.

## Structure to molecular function relationship

To explore the association between the pathogenicity of a variant and its transporter activity, we localized the missense variants, stratified them into severe LoF (0-10% activity), moderate LoF (10-42.8% activity), and WT (>42.8% activity), onto the GAT1 3D structure. Annotated functional scores were normalized over proximal residues across the entire GAT1 3D structure (see Methods). A visual inspection of the GAT1 3D structure suggests regions that harbor predominantly severe LoF mutations (Figures 3 A & B). A similar analysis on 162 out of 195 ClinVar missense VUS for which a 3D normalized score (For details on the normalized score, see methods) was available did not reveal the same pattern. The variants were dispersed, and only 29/162 (17.9%) and 64/162 (39.5%) fall into the complete LoF and moderate LoF group, respectively (Supplementary Figure 2 & 3). We used a spatial distance scoring framework to explore spatial position-to-function relationships that measure each variant's distance from the ligand.[28] We calculated the ligands' distance from each variant in all three activity groups (Figure 3 C) and performed a two-tailed Wilcoxon rank sum test. The <10% WT activity group showed the lowest mean ligand distance, and the >42.8% WT activity group was the highest ($P$ = $2.2e^{-10}$). Overall, we observed that the higher the average transporter activity, the greater its distance from the ligand in a sub-analysis with removed missense variants whose wild-type residue is glycine or proline residues which are known to be "helix breakers".[43,44] We found the lowest mean ligand distance for variants in the <10% WT activity group and the highest for the >42.8% WT group ($P$ = $1.1e^{-8}$) (Supplementary Figure 5). When considering variants whose wild-type residue was either glycine or proline, an insufficient number of variants remaining impeded a meaningful result (Supplementary Figure 6). We also investigated the distance from the GABA transporter axis (defined as the distance in Angstrom (Å) of each variant from the GAT1 axis), from TM1 and from TM6. Similarly, we observed an increasing average transporter activity with increased distance from the transporter axis, TM1 and TM6. Variants in the <10% WT activity group showed the lowest mean transporter axis distance, and the >42.8% WT

activity group showed the highest for the transporter axis, TM1, and TM6. The strongest signal observed was the distance from TM6 ($P = 8.6e^{-9}$) (Supplementary Methods and Figures 7-9).

## Molecular function to clinical phenotype relationship

The clinical spectrum of *SLC6A1*-related disorders is broad. To determine if *in vitro* functional readouts are associated with disease severity, we compared the number of patients with severe vs. non-severe disease for variants grouped by their average functional activity in relation to WT (0-10% - complete LoF, 10-42.8% - moderate LoF, and >42.8% - WT). Severe disease is defined by a reported high seizure burden, often with significant developmental delay and refractory seizures resulting in pronounced life challenges and inability to support themselves as prevalent in epilepsy syndromes such as DEE or LGS (see Methods). For variants with complete loss of *in vitro* GAT1 reuptake, we found a 4.6-fold enrichment of patients with severe vs. non-severe disease ($P = 2.9e^{-3}$, 95% CI: 1.5 - 15.3) compared to variants with moderate LoF and WT function. Other comparisons were not significant.

# Discussion

We aggregated *SLC6A1* data from various sources to perform joined genetic, protein structure, molecular, and clinical data analysis to study genotype-phenotype relationships in *SLC6A1*-related disorders. We show that a reduced distance from the ligand was associated with a greater reduction in transporter activity and that lower GAT1 transporter function is associated with more severe phenotypic *SLC6A1*-related disorder presentations marked by significant life challenges (i.e., intractable epilepsy, moderate to severe ID, and ASD). To make all aggregated data available for educational purposes and research projects, we developed the SLC6A1 Portal (https://slc6a1-portal.broadinstitute.org/). In addition to data access, we provide tools that allow the evaluation of genetic, clinical, and functional features of *SLC6A1*-related disorders on the GAT1 3D structure.

Patient and population variant positions along its protein sequence or 3D protein structure can inform whether a particular variant within a specific region is more likely to cause disease.[45–

[48] Previous studies suggested that pathogenic missense variants primarily cluster near the GABA binding pocket, located around the sixth and seventh transmembrane domains of the GAT1 protein.[4,10,11] However, quantification and validation using a statistical approach were lacking. We demonstrate for the first time that the variants within the transmembrane region of GAT1 are predominantly LoF. Additionally, we observe that missense variants classified as VUS are dispersed throughout the protein's structure and that 42.6% of those variants fall into the WT group (>42.8% WT activity) and 17.9% fall into the complete LoF group (<10%), indicating that potentially a subset of VUS might be pathogenic. Future patient variant reports, and functional analysis will likely resolve the clinical significance of this subset of VUS. As previously suggested, we could not confirm any cluster of low-activity variants within the seventh transmembrane domain.[4,10] Further, we refine the association using two orthogonal approaches. First, we investigated domain-wide changes in the transporter activity of tested patient variants. Second, we compared the enrichment of patient vs. population variants across all domains and demonstrated that the TM1/6 and EL4 domains harbor variants exclusively with a high decrease in *in vitro* transporter assay activity (<42.8%) compared to WT. Both TM1/6 and EL4 were enriched for patient variants. In line with our observation, previous studies on paralogous *SLC6A1-14* genes also pinpoint the TM1/6 region as crucial based on substantial sequence conservation in central regions of the protein structure and the sodium ion binding sites.[49–54]

We show for the first time a spatial association of transporter activity and the distance between the variant position and the ligand Tiagabine. We could statistically quantify that the distance from the ligand was significantly different for the nearly complete LoF activity group of variants (<10%) compared to the 10-42.8% and >42.8% average activity groups to WT groups. Overall, the greater the distance of a variant from the ligand, the closer the transporter activity was to WT activity and vice versa. Our results align with previous observations from our team and other research groups that indicated that missense variants with a higher decrease in activity compared to WT were predominantly located in the protein's transmembrane domain and suggested upon visual inspection that LoF variants might be enriched near the transporter axis.[11] Concordant with the suggestive preliminary data by Trinidad,[11] our 3D variant mapping also shows that variants with very low to low activity, compared to WT, tend to be closer to the proteins' vertical axis. In contrast, variants with WT-like activity tend to face outwards, meaning that those variants tend to be consistently on the exterior of the protein structure and might have

less interference with the protein's transporter function. This pattern agrees with GAT1 function as a gateway for GABA, and anything that could obstruct that gateway is potentially impairing its function as a transporter.[11,23,55] We observed individual variants clustering around the vertical protein axis, causing a decreased channel function through defective transporter function. However, it's important to note that impaired trafficking to the cell surface in cell culture has been demonstrated also to cause channel dysfunction.[5,6,11] Variants that cause channel dysfunction through impaired trafficking are, however, not captured in our analysis. Also, an assay in HEK cells is not a model for investigating cellular trafficking. Further trafficking-specific functional evaluation of these variants is required to determine a trafficking defect.

Additionally, it has been shown for related proteins from the *SLC6A1*-14 transport family that the transmembrane region around transmembrane helices 1 and 6 is a crucial element for a functioning transporter protein.[49–51,56–59] Although previous studies could show a relation between function and pathogenicity[5,11] in selected *SLC6A1* variants, no study has yet established an association with disease severity. This study could show now that *in vitro* functional readouts of *SLC6A1* variants were associated with disease severity. We observe an enrichment of LoF variants near the ligand. As this area is essential for GABA transport, disruptions within the transmembrane domain can cause devastating disease, potentially due to complete LoF – instead of partial LoF. Other research groups also identified crucial regions within the transmembrane domain near the GAT1 axis.[2,11,23,54] We found that variants with low activity compared to WT and disease severity were associated with a 4.6-fold enrichment of patients with severe disease vs. non-severe disease. Here, severe disease is defined as having major challenges in life, such as any report of refractory seizures, developmental delay, or intellectual disability. Children with non-severe phenotypes have fewer challenges, such as no refractory or milder seizures (see Methods). We could not delineate an association of the ligand distance with the age of seizure onset as this feature was only reported for a small subset of individuals ($n$ = 27). This was expected considering previous information from studies on *SLC6A1* that were all limited by small cohort sizes and sparse clinical information.[5,10,11] Researchers had previously encountered this same challenge with other genes but could overcome these limitations over time with more data.[60] Examples from the literature show a clear path for exploiting the relationship between variant location and patient phenotype.[60–64]

Our study has several limitations. First, although our *SLC6A1* patient cohort is the largest to date, it is still small, which may prevent findings from being extrapolated. Further, the clinical data has not been ascertained through standardized procedures, instead was *post-hoc* curated by clinical experts who ascertained data from many sites. For rare diseases, like *SLC6A1*-related disorders, that only recently have been identified with specific phenotypes, such as specific EEG features, the task of correctly coding phenotypic information in routine care represents a challenge.[65–67] Cohort size and data standardization both affect statistical analysis power. Notably, the variant position-based analysis performed in this study does not account for trafficking defects that have been shown to contribute to *SLC6A1* deficiency.[5]

Nevertheless, we validated previous suggestive evidence and identified novel genotype-phenotype associations. Data from the current prospective natural history of disease studies[68] and larger retrospective data aggregations will likely identify additional genotype-phenotype associations and potentially enable risk prediction models.[20,69,70] Another limitation of this study is the lack of complete genome and environmental data. We noticed that several patients with the same recurrent variant had heterogeneous (non-severe vs. severe) expressions of *SLC6A1*-related disorders. For example, the most frequent recurring variant p.Val342Met is classified as CAE for one individual, EMAS for three, and unclassified epilepsy in four individuals with cognition ranging from normal to severe DD/ID (Supplementary Table 1). Due to the clinical *SLC6A1* heterogeneity, the categorization into non-severe vs. severe is limited. This was confirmed in personal discussions with the treating physicians to rule out a coding bias. Future studies should investigate genetic modifiers such as rare variation, the polygenic risk for epilepsy, autism, or low intelligence quotient[71–74] since several recent studies showed that genomic background could modify the expression of the disease.[75–78] Similarly, environmental factors, including drug history, need to be incorporated into statistical models.[75]

In summary, our results show the relationship between each variant's distance from the ligand and the level of average transporter activity in *SLC6A1*-related disorders. Future functional characterization of variants is needed to investigate the hypothesis presented in this study and determine whether the association between genetic location and disease severity found in this study can also be found in other clinical phenotypes, such as age at seizure onset. More data needs to be aggregated to develop a reliable pathogenicity predictor, as this would be a major step forward in improving the clinical management of patients with *SLC6A1*-related

disorders. Our SLC6A1 Portal will contribute to this endeavor. Future studies could potentially elucidate the relationship between variant location and treatment response, paving the way for a personalized medicine approach.

## Data availability

All relevant data and methods are reported in the article and the Supplementary material.

## Funding

Dennis Lal's work was supported by funds from the Dravet Syndrome Foundation (grant number, 272016), the BMBF (Treat-ION grant, 01GM1907), NIH NINDS (Channelopathy-Associated Epilepsy Research Center, 5-U54-NS108874). Eduardo Pérez-Palma is supported by Chilean National Agency for Investigation and Development, ANID Fondecyt grant 1221464 and the FamilieSCN2A foundation 2020 Action Potential Grant. PM received support by the German BMBF (Treat-Ion2, 01GM2210B) and the FNR in Luxembourg (Research Unit FOR-2715, FNR grant NTER/DFG/21/16394868 MechEPI2).

## Competing interests

D.L. consults for Encoded Therapeutics Inc., San Francisco, USA. A.W. and S.F. are employees of BioMarin Pharmaceutical and hold common stock. M.T. is a former BioMarin employee. E.B. is an employee of GeneDx, LLC. All other authors declare no competing interests.

## Supplementary material

Supplementary material is available at *Brain* online.

# References


1. Scimemi A. Structure, function, and plasticity of GABA transporters. *Front Cell Neurosci*. 2014;8:161. doi:10.3389/fncel.2014.00161
2. Goodspeed K, Pérez-Palma E, Iqbal S, et al. Current knowledge of SLC6A1-related neurodevelopmental disorders. *Brain Commun*. doi:10.1093/braincomms/fcaa170
3. Johannesen KM, Gardella E, Linnankivi T, et al. Defining the phenotypic spectrum of SLC6A1 mutations. *Epilepsia*. 2018;59(2):389-402. doi:10.1111/epi.13986
4. Carvill GL, McMahon JM, Schneider A, et al. Mutations in the GABA Transporter SLC6A1 Cause Epilepsy with Myoclonic-Atonic Seizures. *Am J Hum Genet*. 2015;96(5):808-815. doi:10.1016/j.ajhg.2015.02.016
5. Mermer F, Poliquin S, Rigsby K, et al. Common molecular mechanisms of SLC6A1 variant-mediated neurodevelopmental disorders in astrocytes and neurons. *Brain J Neurol*. Published online May 24, 2021. doi:10.1093/brain/awab207
6. Mattison KA, Butler KM, Inglis GAS, et al. SLC6A1 variants identified in epilepsy patients reduce γ-aminobutyric acid transport. *Epilepsia*. 2018;59(9):e135-e141. doi:10.1111/epi.14531
7. Cai K, Wang J, Eissman J, et al. A missense mutation in SLC6A1 associated with Lennox-Gastaut syndrome impairs GABA transporter 1 protein trafficking and function. *Exp Neurol*. 2019;320:112973. doi:10.1016/j.expneurol.2019.112973
8. Wang J, Poliquin S, Mermer F, et al. Endoplasmic reticulum retention and degradation of a mutation in SLC6A1 associated with epilepsy and autism. *Mol Brain*. 2020;13(1):76. doi:10.1186/s13041-020-00612-6
9. Poliquin S, Hughes I, Shen W, et al. Genetic mosaicism, intrafamilial phenotypic heterogeneity, and molecular defects of a novel missense SLC6A1 mutation associated with epilepsy and ADHD. *Exp Neurol*. 2021;342:113723. doi:10.1016/j.expneurol.2021.113723

10. Kahen A, Kavus H, Geltzeiler A, et al. Neurodevelopmental phenotypes associated with pathogenic variants in SLC6A1. *J Med Genet*. Published online May 18, 2021:jmedgenet-2021-107694. doi:10.1136/jmedgenet-2021-107694

11. Trinidad MI, Froelich S, Berguig G, et al. High-throughput discovery of SLC6A1 variants affecting GABA transport in neurological disorders. Published online March 12, 2022:2022.03.09.22271804. doi:10.1101/2022.03.09.22271804

12. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends Neurosci*. 2020;43(8):608-621. doi:10.1016/j.tins.2020.05.004

13. Lahiry P, Torkamani A, Schork NJ, Hegele RA. Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat Rev Genet*. 2010;11(1):60-74. doi:10.1038/nrg2707

14. Kaufman L, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. *J Neurodev Disord*. 2010;2(4):182-209. doi:10.1007/s11689-010-9055-2

15. Moretto E, Murru L, Martano G, Sassone J, Passafaro M. Glutamatergic synapses in neurodevelopmental disorders. *Prog Neuropsychopharmacol Biol Psychiatry*. 2018;84(Pt B):328-342. doi:10.1016/j.pnpbp.2017.09.014

16. Weiss K, Lazar HP, Kurolap A, et al. The CHD4-related syndrome: a comprehensive investigation of the clinical spectrum, genotype-phenotype correlations, and molecular basis. *Genet Med Off J Am Coll Med Genet*. 2020;22(2):389-397. doi:10.1038/s41436-019-0612-0

17. Lisé MF, El-Husseini A. The neuroligin and neurexin families: from structure to function at the synapse. *Cell Mol Life Sci CMLS*. 2006;63(16):1833-1849. doi:10.1007/s00018-006-6061-3

18. Platzer K, Yuan H, Schütz H, et al. GRIN2B encephalopathy: novel findings on phenotype, variant clustering, functional consequences and treatment aspects. *J Med Genet*. 2017;54(7):460-470. doi:10.1136/jmedgenet-2016-104509

19. Kapplinger JD, Tester DJ, Alders M, et al. An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart Rhythm*. 2010;7(1):33-46. doi:10.1016/j.hrthm.2009.09.069

20. Heyne HO, Baez-Nieto D, Iqbal S, et al. Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci Transl Med*. 2020;12(556). doi:10.1126/scitranslmed.aay6848

21. Strehlow V, Heyne HO, Vlaskamp DRM, et al. GRIN2A-related disorders: genotype and functional consequence predict phenotype. *Brain J Neurol*. 2019;142(1):80-92. doi:10.1093/brain/awy304

22. Pan X, Li Z, Jin X, et al. Comparative structural analysis of human Nav1.1 and Nav1.5 reveals mutational hotspots for sodium channelopathies. *Proc Natl Acad Sci U S A*. 2021;118(11):e2100066118. doi:10.1073/pnas.2100066118

23. Motiwala Z, Aduri NG, Shaye H, et al. Structural basis of GABA reuptake inhibition. *Nature*. 2022;606(7915):820-826. doi:10.1038/s41586-022-04814-x

24. Olson HE, Demarest ST, Pestana-Knight EM, et al. Cyclin-dependent kinase-like 5 (CDKL5) deficiency disorder: clinical review. *Pediatr Neurol*. 2019;97:18-25. doi:10.1016/j.pediatrneurol.2019.02.015

25. Balak C, Benard M, Schaefer E, et al. Rare De Novo Missense Variants in RNA Helicase DDX6 Cause Intellectual Disability and Dysmorphic Features and Lead to P-Body Defects and RNA Dysregulation. *Am J Hum Genet*. 2019;105(3):509-525. doi:10.1016/j.ajhg.2019.07.010

26. Salles PA, Mata IF, Brünger T, Lal D, Fernandez HH. ATP1A3-Related Disorders: An Ever-Expanding Clinical Spectrum. *Front Neurol*. 2021;12:637890. doi:10.3389/fneur.2021.637890

27. Schwarz N, Seiffert S, Pendziwiat M, et al. Spectrum of Phenotypic, Genetic, and Functional Characteristics in Patients With Epilepsy With KCNC2 Pathogenic Variants. *Neurology*. 2022;98(20):e2046-e2059. doi:10.1212/WNL.0000000000200660

28. Brünger T, Pérez-Palma E, Montanucci L, et al. Conserved patterns across ion channels correlate with variant pathogenicity and clinical phenotypes. *Brain J Neurol*. Published online August 29, 2022:awac305. doi:10.1093/brain/awac305

29. Lelieveld SH, Wiel L, Venselaar H, et al. Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *Am J Hum Genet*. 2017;101(3):478-484. doi:10.1016/j.ajhg.2017.08.004

30. Johannesen KM, Iqbal S, Guazzi M, et al. Structural mapping of GABRB3 variants reveals genotype-phenotype correlations. *Genet Med Off J Am Coll Med Genet*. 2022;24(3):681-693. doi:10.1016/j.gim.2021.11.004

31. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(Database issue):D1062-D1067. doi:10.1093/nar/gkx1153

32. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. *Am J Hum Genet*. 2019;105(2):267-282. doi:10.1016/j.ajhg.2019.05.020

33. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

34. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4

35. Whiffin N, Minikel E, Walsh R, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med Off J Am Coll Med Genet*. 2017;19(10):1151-1158. doi:10.1038/gim.2017.26

36. Schrödinger, L., & DeLano, W. PyMOL. Published online 2020. http://www.pymol.org/pymol

37. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinforma Oxf Engl*. 2006;22(21):2695-2696. doi:10.1093/bioinformatics/btl461

38. Tenny S, Hoffman MR. Prevalence. In: *StatPearls*. StatPearls Publishing; 2023. Accessed May 17, 2023. http://www.ncbi.nlm.nih.gov/books/NBK430867/

39. López-Rivera JA, Pérez-Palma E, Symonds J, et al. A catalogue of new incidence estimates of monogenic neurodevelopmental disorders caused by de novo variants. *Brain*. 2020;143(4):1099-1105. doi:10.1093/brain/awaa051

40. Dreier JW, Laursen TM, Tomson T, Plana-Ripoll O, Christensen J. Cause-specific mortality and life years lost in people with epilepsy: a Danish cohort study. *Brain*. 2023;146(1):124-134. doi:10.1093/brain/awac042

41. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. doi:10.1038/sdata.2016.18

42. Imoukhuede PI, Moss FJ, Michael DJ, Chow RH, Lester HA. Ezrin Mediates Tethering of the γ-Aminobutyric Acid Transporter GAT1 to Actin Filaments Via a C-Terminal PDZ-Interacting Domain. *Biophys J*. 2009;96(7):2949-2960. doi:10.1016/j.bpj.2008.11.070

43. Levitt M. Conformational preferences of amino acids in globular proteins. *Biochemistry*. 1978;17(20):4277-4285. doi:10.1021/bi00613a026

44. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol*. 1978;47:45-148. doi:10.1002/9780470122921.ch2

45. Iqbal S, Brünger T, Pérez-Palma E, et al. Genomic analysis of AlphaFold2-predicted structures identifies maps of 3D essential sites in 243 neurodevelopmental disorder-associated proteins. *Biophys J*. 2022;121(3):165a-166a. doi:10.1016/j.bpj.2021.11.1909

46. Iqbal S, Pérez-Palma E, Jespersen JB, et al. Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants. *Proc Natl Acad Sci U S A*. 2020;117(45):28201-28211. doi:10.1073/pnas.2002660117

47. Laskowski RA, Stephenson JD, Sillitoe I, Orengo CA, Thornton JM. VarSite: Disease variants and protein structure. *Protein Sci Publ Protein Soc*. 2020;29(1):111-119. doi:10.1002/pro.3746

48. Glusman G, Rose PW, Prlić A, et al. Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: a proposed framework. *Genome Med*. 2017;9(1):113. doi:10.1186/s13073-017-0509-y

49. Ng J, Zhen J, Meyer E, et al. Dopamine transporter deficiency syndrome: phenotypic spectrum from infancy to adulthood. *Brain J Neurol*. 2014;137(Pt 4):1107-1119. doi:10.1093/brain/awu022

50. Field JR, Henry LK, Blakely RD. Transmembrane domain 6 of the human serotonin transporter contributes to an aqueously accessible binding pocket for serotonin and the psychostimulant 3,4-methylene dioxymethamphetamine. *J Biol Chem*. 2010;285(15):11270-11280. doi:10.1074/jbc.M109.093658

51. Bai X, Moraes TF, Reithmeier RAF. Structural biology of solute carrier (SLC) membrane transport proteins. *Mol Membr Biol*. 2017;34(1-2):1-32. doi:10.1080/09687688.2018.1448123

52. Singh SK, Piscitelli CL, Yamashita A, Gouaux E. A competitive inhibitor traps LeuT in an open-to-out conformation. *Science*. 2008;322(5908):1655-1661. doi:10.1126/science.1166777

53. Yamashita A, Singh SK, Kawate T, Jin Y, Gouaux E. Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. *Nature*. 2005;437(7056):215-223. doi:10.1038/nature03978

54. Navratna V, Gouaux E. Insights into the mechanism and pharmacology of neurotransmitter sodium symporters. *Curr Opin Struct Biol*. 2019;54:161-170. doi:10.1016/j.sbi.2019.03.011

55. Coleman JA, Yang D, Zhao Z, et al. Serotonin transporter-ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature*. 2019;569(7754):141-145. doi:10.1038/s41586-019-1135-1

56. Pramod AB, Foster J, Carvelli L, Henry LK. SLC6 transporters: structure, function, regulation, disease association and therapeutics. *Mol Aspects Med*. 2013;34(2-3):197-219. doi:10.1016/j.mam.2012.07.002

57. Hahn MK, Robertson D, Blakely RD. A mutation in the human norepinephrine transporter gene (SLC6A2) associated with orthostatic intolerance disrupts surface expression of mutant and wild-type transporters. *J Neurosci Off J Soc Neurosci*. 2003;23(11):4470-4478.

58. Reith MEA, Kortagere S, Wiers CE, et al. The dopamine transporter gene SLC6A3: multidisease risks. *Mol Psychiatry*. 2022;27(2):1031-1046. doi:10.1038/s41380-021-01341-5

59. Bowton E, Saunders C, Reddy IA, et al. SLC6A3 coding variant Ala559Val found in two autism probands alters dopamine transporter function and trafficking. *Transl Psychiatry*. 2014;4:e464. doi:10.1038/tp.2014.90

60. Baruteau AE, Kyndt F, Behr ER, et al. SCN5A mutations in 442 neonates and children: genotype-phenotype correlation and identification of higher-risk subgroups. *Eur Heart J*. 2018;39(31):2879-2887. doi:10.1093/eurheartj/ehy412

61. Zuberi SM, Brunklaus A, Birch R, Reavey E, Duncan J, Forbes GH. Genotype-phenotype associations in SCN1A-related epilepsies. *Neurology*. 2011;76(7):594-600. doi:10.1212/WNL.0b013e31820c309b

62. Mannucci I, Dang NDP, Huber H, et al. Genotype-phenotype correlations and novel molecular insights into the DHX30-associated neurodevelopmental disorders. *Genome Med*. 2021;13(1):90. doi:10.1186/s13073-021-00900-3

63. Todd JJ, Sagar V, Lawal TA, et al. Correlation of phenotype with genotype and protein structure in RYR1-related disorders. *J Neurol*. 2018;265(11):2506-2524. doi:10.1007/s00415-018-9033-2

64. Ottenhoff MJ, Rietman AB, Mous SE, et al. Examination of the genetic factors underlying the cognitive variability associated with neurofibromatosis type 1. *Genet Med Off J Am Coll Med Genet*. 2020;22(5):889-897. doi:10.1038/s41436-020-0752-2

65. Boycott KM, Rath A, Chong JX, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet*. 2017;100(5):695-705. doi:10.1016/j.ajhg.2017.04.003

66. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet*. 2015;47(7):717-726. doi:10.1038/ng.3304

67. Splinter K, Adams DR, Bacino CA, et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *N Engl J Med*. 2018;379(22):2131-2139. doi:10.1056/NEJMoa1714458

68. Bain JM, Snyder LG, Helbig KL, Cooper DD, Chung WK, Goodspeed K. Consistency of parent-report SLC6A1 data in Simons Searchlight with Provider-Based Publications. *J Neurodev Disord*. 2022;14(1):40. doi:10.1186/s11689-022-09449-7

69. Brunklaus A, Pérez-Palma E, Ghanty I, et al. Development and Validation of a Prediction Model for Early Diagnosis of SCN1A-Related Epilepsies. *Neurology*. 2022;98(11):e1163-e1174. doi:10.1212/WNL.0000000000200028

70. Boßelmann CM, Hedrich UBS, Müller P, et al. Predicting the functional effects of voltage-gated potassium channel missense variants with multi-task learning. *EBioMedicine*. 2022;81:104115. doi:10.1016/j.ebiom.2022.104115

71. Leu C, Stevelink R, Smith AW, et al. Polygenic burden in focal and generalized epilepsies. *Brain J Neurol*. 2019;142(11):3473-3481. doi:10.1093/brain/awz292

72. Campbell C, Leu C, Feng YCA, et al. The role of common genetic variation in presumed monogenic epilepsies. *EBioMedicine*. 2022;81:104098. doi:10.1016/j.ebiom.2022.104098

73. Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM. Predicting Polygenic Risk of Psychiatric Disorders. *Biol Psychiatry*. 2019;86(2):97-109. doi:10.1016/j.biopsych.2018.12.015

74. Niemi MEK, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*. 2018;562(7726):268-271. doi:10.1038/s41586-018-0566-4

75. Tukker AM, Royal CD, Bowman AB, McAllister KA. The Impact of Environmental Factors on Monogenic Mendelian Diseases. *Toxicol Sci Off J Soc Toxicol*. 2021;181(1):3-12. doi:10.1093/toxsci/kfab022

76. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun*. 2020;11(1):3635. doi:10.1038/s41467-020-17374-3

77. Davies RW, Fiksinski AM, Breetvelt EJ, et al. Using common genetic variation to examine phenotypic expression and risk prediction in 22q11.2 deletion syndrome. *Nat Med*. 2020;26(12):1912-1918. doi:10.1038/s41591-020-1103-1

78. Cleynen I, Engchuan W, Hestand MS, et al. Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. *Mol Psychiatry*. 2021;26(8):4496-4510. doi:10.1038/s41380-020-0654-3

# Figure legends

**Figure 1 SLC6A1 Portal: User Interface and Functionality (https://slc6a1-portal.broadinstitute.org/).** Four main menu items hold different functionalities of the SLC6A1 Portal. Basic Information: Key milestones in *SLC6A1* research and summary statistics of the clinical information. Educational Resources: Resources, links to family advocacy groups, and our in-house produced an animated whiteboard explainer video. Variant Analysis: Clinical significance according to ClinVar and comparative information on the selected variant with other similar variants. Research: Visualizations of variant annotations, clinical phenotypes, and functional data based on multiple filter options. Efforts to add more data to our online resource are motivated by an increased ability to understand the logic of structure to function to phenotype relations. Furthermore, easy access, the ability to explore the data, and educational resources are additional features of our web Portal. Given that very few clinicians and caregivers can collect data and perform bioinformatics analyses, the Portal enables anyone with access to the internet to explore the data, understand, and develop hypotheses.

**Figure 2 Domain-wide analysis of patient and population variants. (A)** The TM1/6, scaffold, and EL4 regions harbored mainly variants with low-activity, whereas the N-/C-terminal domain contained mostly variants with wild-type activity. The dotted line represents the cutoff that separates WT activity variants (>42.8% of WT activity) from low activity variants (<42.8% of WT activity). **(B)** A domain-wide comparison of patient vs. population variants shows enrichment of patient variants in TM1/6 and EL4. The N- and C-terminal regions are depleted for patient variants. The dotted line represents a balance between patient and population variants. Abbreviations: TM1/6: Transmembrane helix 1/6; TMD-other: Transmembrane domain other; EL2/3/4: Extracellular loops 2/3/4; OR: Odds ratio; WT: wild-type.

**Figure 3 The spatial relation of *SLC6A1* variants is associated with function and position within the GAT1 protein structure 3D structure (PDB ID: 7SK2).**[23] **(A)** GAT1 3D structure

color-coded in red regions with nearly complete LoF variants (<10% of normalized WT activity) and in yellow regions with WT activity variants (>42.8% of normalized WT activity). **(B)** Side view of the GAT1 3D structure. *SLC6A1* variants were categorized in the same three activity groups [0-10% (red), 10-42.8% (orange), >42.8% (yellow)] and mapped onto the GAT1 3D structure. The variants with the lowest and medium average functional activity tend to be closer to the ligand, whereas variants with WT activity tend to face outwards. **(C)** Box plot showing the quantification of each variant's distance (Å) from the ligand by the three activity groups. Abbreviations: **: Significant after Bonferroni multiple test correction; *: Nominally significant; n.s.: not significant; LoF: loss-of-function; WT: wild-type.

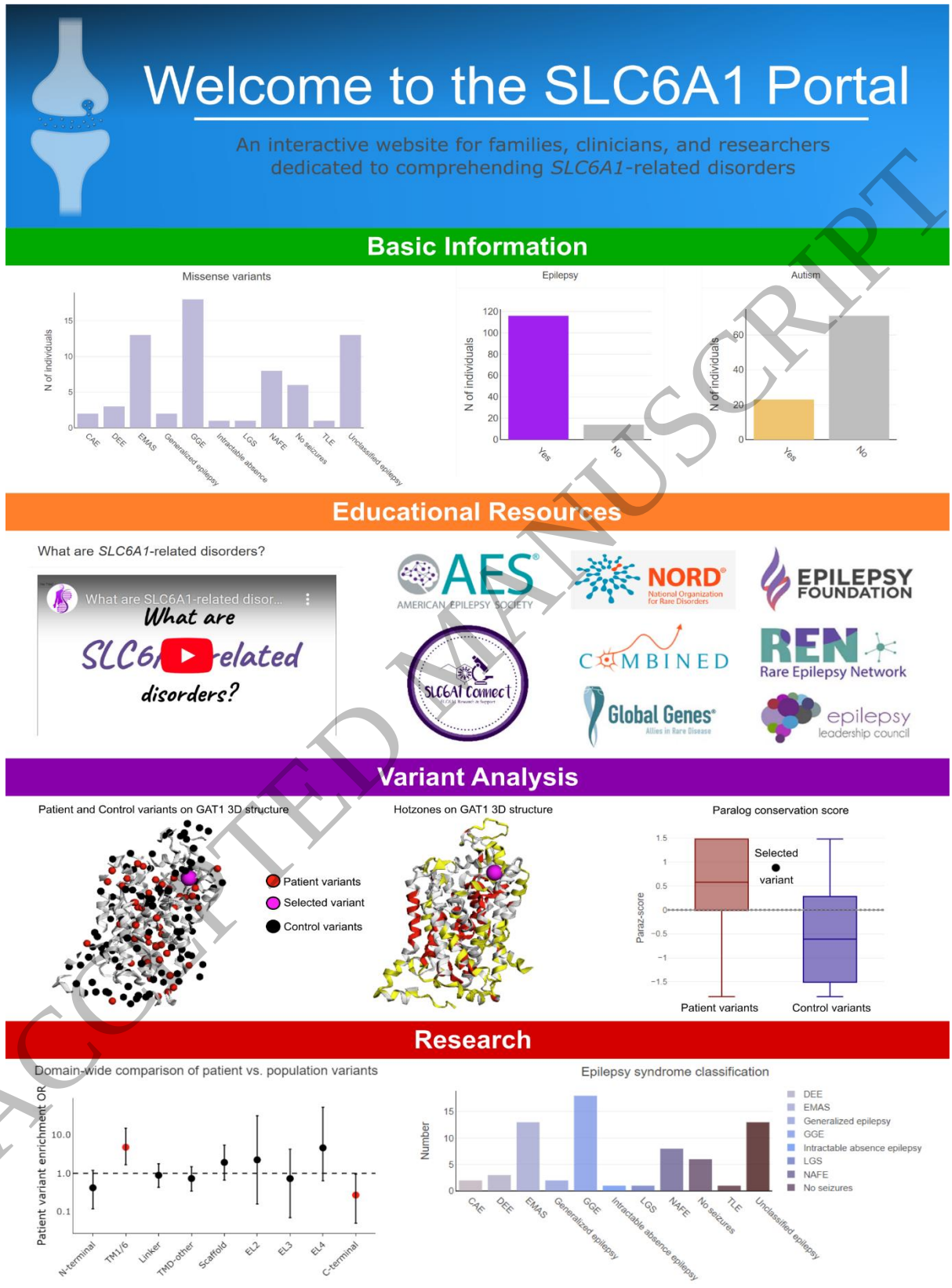


*Figure 1*
*159x225 mm ( x DPI)*

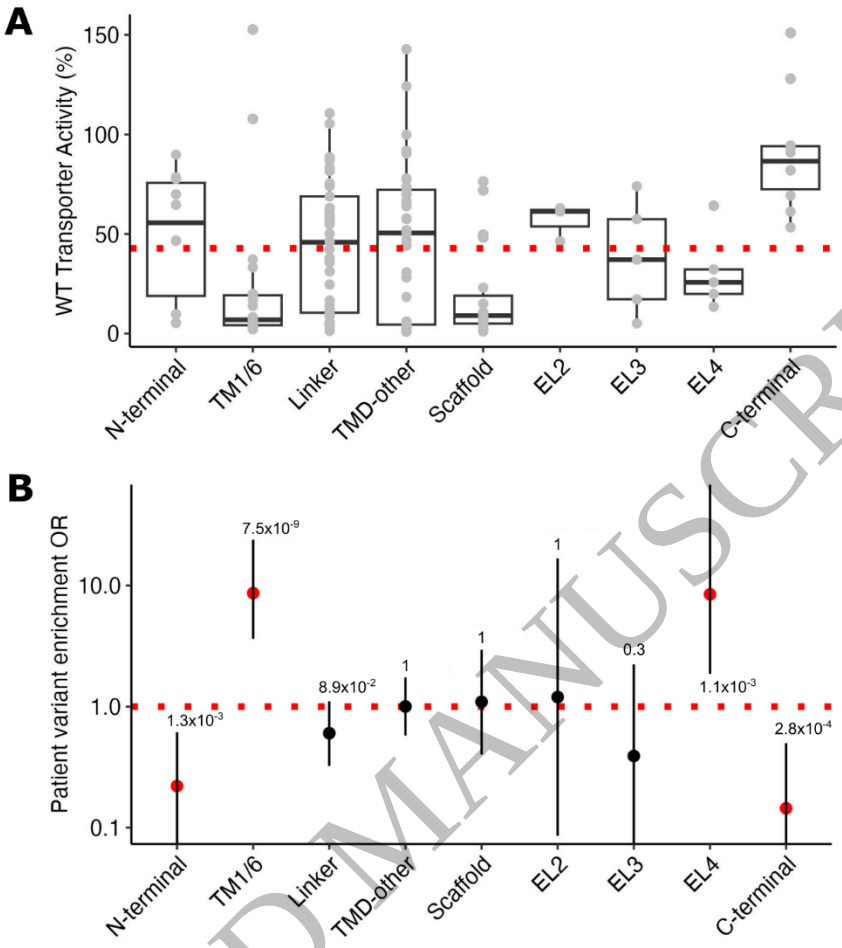


*Figure 2*
*112x129 mm ( x DPI)*

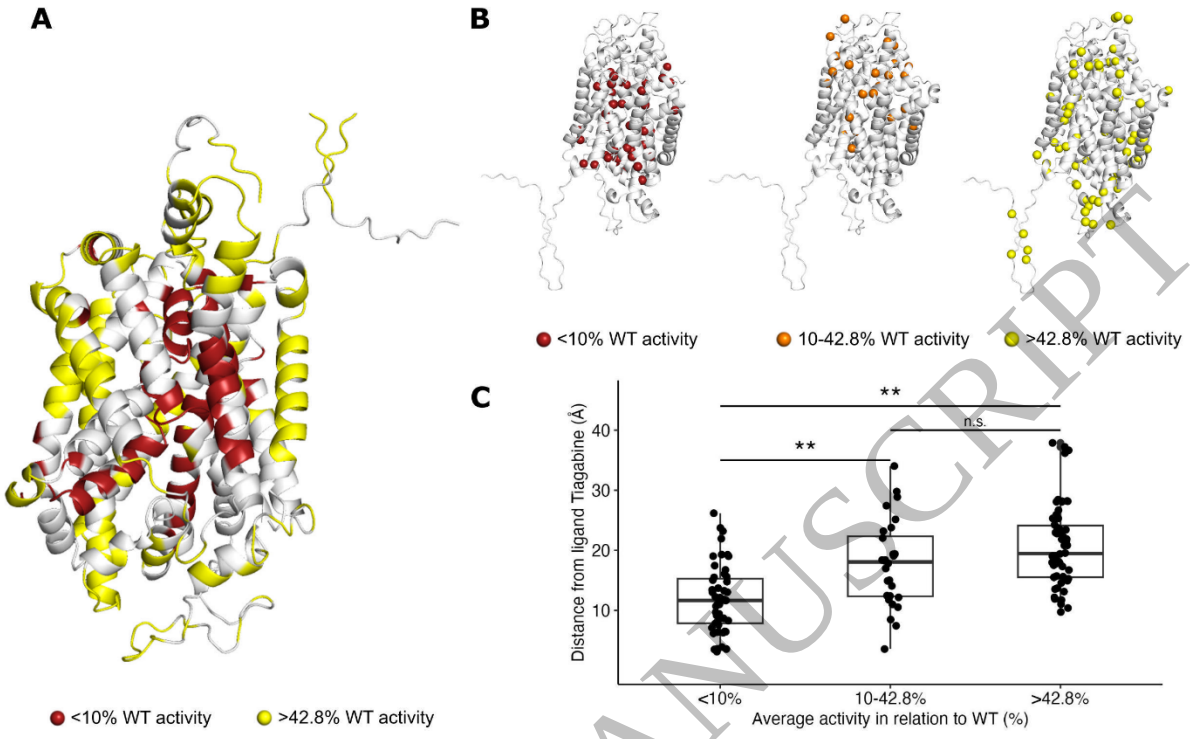


*Figure 3*
*159x97 mm ( x  DPI)*