High-throughput discovery of *SLC6A1* variants affecting GABA transport in neurological disorders

Marena Trinidad, Steven Froelich, Geoffrey Berguig, William Wallace, Lorenzo Bomba, Jon H. LeBowitz, Karol Estrada, Arthur Wuster

Abstract

SLC6A1 encodes the neuronal GABA transporter GAT-1. Pathogenic variants in *SLC6A1* lead to haploinsufficiency and are associated with epilepsy, autism and schizophrenia. Most variants observed in patients are rare missense mutations whose functional impact and pathogenicity remain unknown. We quantified GABA uptake of 182 *SLC6A1* variants using plasmid expression constructs in *SLC6A1*-deficient HEK293T cells. Cells were incubated with deuterated GABA, then analyzed by mass spectrometry to assess transporter activity. Of the 102 variants with partial or complete loss of transporter function compared to wildtype, only 24 (23.5%) had previously been classified as Pathogenic or Likely pathogenic in the ClinVar database. This suggests an increase in the number of *SLC6A1* variants that are likely pathogenic, implying a 40% increased diagnostic yield for *SLC6A1* haploinsufficiency. *SLC6A1* variants associated with Schizophrenia did not lead to complete loss of function, suggesting either an association between hypomorphic variants and Schizophrenia or an alternative mechanism of pathogenesis.

Introduction

Variants in *SLC6A1* (Solute carrier family 6 member 1), the gene encoding the GABA transporter GAT-1, are associated with pediatric onset epilepsy, developmental disorders, autism and schizophrenia[1,2,3]. Cumulatively, mutations in *SLC6A1* affect 2.4-2.9 in 100,000 births[4], making *SLC6A1* mutations one of the most prevalent monogenic causes of developmental disorders.

SLC6A1 is expressed in all brain regions[5] but not all cell types, with expression being highest in GABAergic neurons, followed by astrocytes[6]. Expression increases during embryonic development, reaching a stable plateau around birth[7]. GAT-1, the protein product of *SLC6A1*, is located in the neuronal membrane, where it transports the inhibitory neurotransmitter GABA from the synaptic cleft back into the cytoplasm.

The association between *SLC6A1* and epilepsy was established by targeted sequencing of 644 patients with epileptic encephalopathies, of which seven had *SLC6A1* mutations and exhibited childhood onset myoclonic-atonic seizures (MAE)[1]. Another survey of *SLC6A1* patients determined epilepsy was the most common phenotype (92 out of 101 patients; 91.1%), followed by developmental delay, cognitive impairment (46 out 56 patients; 82.1%) and autistic traits (20 out of 92 patients; 22.8%)[8]. Another study of 34 *SLC6A1* patients and family members found impaired cognitive development in 33 cases (97.1%) and epilepsy in 31 (91.1%) with a mean age of onset of 3.4 years[9].

In the largest sequencing-based association study of autism to date[3], *SLC6A1* was among the genes with the largest enrichment of *de novo* mutations in autism cases compared to controls. In total, 9 out of 11,986 autism patients (0.08%) had likely causative *SLC6A1* variants. The ratio of *de novo* missense variants to protein truncating variants in cases was higher for *SLC6A1* than for most of the other 101 autism genes identified in the study, leading to the question of whether autism-associated *SLC6A1* missense variants are loss or gain of function.

A sequencing-based association study[2] of schizophrenia trios also found a significant enrichment of *de novo* mutations in *SLC6A1*. Three *SLC6A1* patients were identified in a cohort of 3,444 schizophrenia trios (0.09%), suggesting a similar proportion of *SLC6A1* patients among schizophrenia and autism patients. An additional 13 missense variants were not *de novo* but had suggestive pathogenicity scores. Of the two schizophrenia *SLC6A1* patients with a *de novo* mutation for which data was available, neither had a history of seizures or neurodevelopmental delay. With the age of onset for schizophrenia typically being adulthood, compared to childhood age of onset for epilepsy and autism, this suggests that schizophrenia-associated *SLC6A1* variants may have a different mechanism of pathogenesis than autism or epilepsy associated ones.

Pathogenic mutations in *SLC6A1* have been hypothesized to decrease GABA transport by GAT-1[10], lead to an accumulation of GABA in the intercellular space and disruptions in GABAergic signaling. However, only a small proportion of *SLC6A1* variants have so far been functionally tested[10,11,12,13,14] and it is not known if most pathogenic mutations lead to loss of transporter function. In addition, the pathogenicity of most *SLC6A1* variants seen in patients has not been established. We quantified GABA reuptake using plasmid expression constructs for 182 variants in *SLC6A1*-deficient HEK293T cells. We incubated cells with deuterated GABA and analyzed them by mass spectrometry to assess SLC6A1 transporter activity. These results provide functional evidence for the pathogenicity of more than 100 *SLC6A1* variants. They also for the first time provided functional data for Schizophrenia associated *SLC6A1* variants.

Results

We identified individuals with *SLC6A1* variants from 15 cohorts with epilepsy, developmental disorders or schizophrenia, as well as negative controls with silent variants, variants that had been classified as Benign or Likely benign or variants that were more common in healthy controls (Methods and Supplementary Table 1). After excluding copy number variants, large insertion or deletion variants (indels) and noncoding variants, we selected 182 variants for testing in our cellular model (Supplementary Table 2). Of those, 156 were missense, 8 stop gained, 7 frameshift, 5 inframe insertions or deletions and 6 synonymous. 168 variants were single nucleotide variants and 14 were indels (Supplementary Table 4).

GABA reuptake activity assay. We developed an assay which combines high-throughput automated transfection with highly sensitive and scalable mass spectrometry assays (Figure 1A). Our approach builds on previously developed methods[11,12,13,14] of *SLC6A1* variant characterization using radiolabeled GABA to create a high-throughput GABA trafficking assay. Although our method measures the same activity as previously established reuptake assays, the major advancement is the ability to readily scale to determine GABA reuptake for 182 variants. Briefly, *SLC6A1*-deficient HEK293T cells were transfected with plasmids harboring each

variant and incubated with deuterated GABA before quantifying reuptake by mass spectrometry. The variants tested exhibited a wide range of activity, from complete loss of function to potential gain of function (Figure 1B, Figure 2). All truncating variants, along with inframe insertion and deletion variants, were effectively null at less than 10% wildtype activity (Figure 3). Synonymous variants, with one exception discussed below, showed no evidence of reduced function, while missense variants span the spectrum from complete loss of function to augmented activity above wildtype levels.

The synonymous variant with the lowest function was c.1563A>G at 53.4% (confidence interval 46.7-60.3%) and therefore significantly different from wildtype. This variant was included because it was observed in epilepsy cases included in a cohort of 9,170 epilepsy patients but not in any controls[15]. Its allele frequency in gnomAD is also low at 7.9×10^{-6}. A potential reason for the reduced function we observe is that it changes the codon for Proline from CCA to CCG, which is less frequently observed in the human genome. The frequency of the wildtype CCA codon for proline is 27%, while the frequency of the CCG codon is less than half at 11%[16].

Derivation of threshold for loss of function. Variants that were predicted to be synonymous (n=6) or classified as benign (n=5) in ClinVar were used to establish a cutoff that demarcates high-activity variants from low-activity variants associated with disease in the literature. Briefly, we used the t-interval to calculate the 99% confidence interval of percent-wildtype activity for each benign or synonymous variant and the lowest of these lower-bounds, 42.8% wildtype activity, was selected as the cutoff for deficient or low activity. Therefore, any variant whose activity is less than the 42.8% threshold for benign mutations was categorized as low-activity.

Comparison of activity assay with literature. We identified 27 variants from the literature whose GABA reuptake activities were previously quantified using radiolabeled ^{3}H-GABA and liquid scintillation counting[10,11,12,13,14]. Three of these variants (Gly234Ser, Gly550Arg, Pro361Thr) were assayed using ^{3}H-GABA in multiple publications and exhibited variability between studies, achieving a Pearson correlation of 0.84, with a p-value of 0.36. We observed a Pearson correlation coefficient of 0.68 between our percent-wildtype activities and those from literature. While the correlation of our activities with that of prior studies is modest, former publications did not leverage MSMS, used a variety of cell lines without the CRISPR-Cas9 *SLC6A1*-knockout present in our cellular model and did not normalize for variable expression efficiencies with our BLA-reporter. Because of these differences in methodology, we did not consider discrepancies with previous percent-wildtype levels concerning.

For instance, Val125Met shows partial loss of function at 9.0% of wildtype in our assay. Reduced but not eliminated function as measured by a GABA reuptake assay has also been reported elsewhere for this variant[14]. This variant may destabilize SLC6A1 protein structure, resulting in lower transporter levels at the synapse[14]. Additionally, in our assay, Ser295Leu shows almost complete loss of function at 2.1% of wildtype and classifies as deficient. Ser295Leu was also characterized in a mouse model and found to be deleterious[17]. Compared to wildtype mice, *SLC6A1*$^{\text{Ser295Leu/+}}$ animals have frequent spike wave discharges on chronic wireless electrocochleography (ECoG) and electromyographic (EMG) recordings, in addition to demonstrating behavioral changes. All observed changes are consistent with the Ser295Leu variant being haploinsufficient and the mouse data is therefore consistent with our functional

results. This underscores the ability of high-throughput cellular models like ours to serve as an accurate and cost-effective alternative for mouse models in the functional characterization of Variants of Unknown Significance.

Gain of function variants. The average percent wildtype activity across replicates was greater than 100% for 11 variants, all of which were missense or synonymous (Figure 3). For two of those, c.960C>T (Silent) and Gly561Ser, average function was more than two standard deviations above wildtype and the lower bound of their 95% confidence interval therefore exceeded 100%. c.960C>T is classified as benign and Gly561Ser is a Variant of Unknown Significance according to ClinVar. Codon usage for the silent c.960C>T variant does not provide a clear explanation as the mutation changes the codon for Serine from TCC to TCT, which are observed for 22% and 18% of Serine codons respectively in the human genome[16]. Other limitations of the cell-based model include there could be cell type specific mechanisms like proteases that are not expressed in HEK293T cells and consequently SLC6A1 protein production and GABA uptake could vary by cell type. Finally, some missense variants could be acting through splicing mechanisms which would not be captured in an SLC6A1 CDS construct used in this study.

Reclassification of variants. Of the 182 variants we tested, 143 (78.6%) were not classified as Pathogenic, Likely pathogenic, Benign or Likely benign according to the ClinVar database. 102 variants showed partial or complete loss of transporter function compared to wildtype, of which 24 (23.5%) had previously been classified as Pathogenic or Likely pathogenic. The 78 remaining variants with loss of transporter function had previously not been classified or been classified as Likely benign, as Variants of Uncertain Significance or with Conflicting evidence. Of the 80 variants that did not show loss of function below the threshold of 42.8%, 13 (16.3%) had previously been classified as Benign/Likely benign (Figure 1B, Figure2, Supplementary Table 5A).

Of the 84 variants which had been classified in ClinVar, 25 (30%) were Pathogenic or Likely pathogenic. Our data suggest that the number of variants that should be considered as pathogenic is larger than this. We observed partial or complete loss of function for 48 out of 84 variants (57%), almost doubling the number of variants that are potentially pathogenic (Table 1).

To determine the increase in diagnostic yield to be expected if all variants that show reduced function in our assay were pathogenic, we counted to number of patients with those variants in the cohorts we analyzed. Of the 138 patients carrying a variant classified in ClinVar, 63 (46%) had a variant previously classified as Pathogenic or Likely pathogenic. There were 88 patients (64%) that carried a variant with loss of or reduced function in our assay (Table 1). Assuming GABA transporter function predicts pathogenicity, this implies an increase of the diagnostic yield for SLC6A1 haploinsufficiency by 40%.

Schizophrenia-associated variants are not loss of function. Two of the three variants that have been observed as *de novo* mutations in schizophrenia patients (Ala93Thr, Arg211Cys, Trp495Leu)[2] do not show a reduction in function at the level seen in most pathogenic epilepsy or developmental delay patients. The third shows an incomplete reduction in function to 24.7% of wildtype. This, together with the later age of disease onset for schizophrenia and the differing

phenotypic characteristics of patients, suggests that variants in *SLC6A1* may contribute to Schizophrenia in a different way than to epilepsy.

We also tested 11 variants that were potentially associated with Schizophrenia as they had suggestive pathogenicity scores and were only observed in a Schizophrenia cohort, but not in controls[2]. The association of those variants with Schizophrenia was therefore less clear as that of the three *de novo* variants discussed above. 4 out of 11 had a function below the 42.8% threshold (Gly442Arg at 5.4%, Asp410Asn at 6.5%, Asn137Asp at 23.1% and Arg417Cys at 36.5% of wildtype). The other 7 variants ranged between 57.0% and 82.1% of wildtype.

Correlation with predictors of severity. Our assay results correlate well with other predictors of disease severity. Variants that have previously been observed in the gnomAD database[18] are less likely to be loss of function in our assay (Figure 5). Transporter activity measured by our assay was also used to validate two *in silico* models that predict variant pathogenicity, SIFT[19] and PolyPhen2[20]. Only *SLC6A1* variants whose clinical significances were labelled as Unclassified in ClinVar (n=82) were used for analysis as SIFT and PolyPhen predictions are calculated from ClinVar annotations. The GABA-reuptake activity of each *SLC6A1* variant was compared to SIFT and PolyPhen scores (Supplementary Figure 1). Correlations of reuptake activity with SIFT and PolyPhen were moderate, with Pearson's Coefficients of 0.45 and -0.58, respectively. Classifier performance, predicting whether a VUS was Pathogenic or Non-Pathogenic, is summarized in a Venn Diagram (Supplementary Figure 1) and exhibits that neither algorithm had perfect accuracy. The inadequacy of *in silico* models to predict variant pathogenicity agrees with previous studies[21,22] and highlights the utility of high-throughput activity screens such as ours.

Mapping of variants to the protein structure. Our assay quantifies reuptake activity and additional follow up is required to determine the mechanistic causes underlying changes in SLC6A1 function. For example, a variant may affect the amount of SLC6A1 protein available at the synapse by impacting endoplasmic reticulum retention and degradation[13], by affecting protein stability[14] or by directly impacting the protein's ability to transport GABA. To explore potential mechanisms, we mapped the location of missense variants reducing transporter function to the protein structure. A stylized representation of SLC6A1's topological domains is shown in Figure 6 and an interactive PyMol (version 2.3.5) session with the protein's predicted structure from AlphaFold[24,25] is available on request.

Residues that reduce transporter functions are predominantly located in the transmembrane helices, with fewer on the cytoplasmic face. In the transmembrane helices, loss of function residues are more likely to line the pore region and less likely to face outwards. Overall, 33 out of 56 transmembrane variants (59%) result in reduced function, compared to 34 out of 56 (39%) for cytoplasmic and 27 out of 50 (54%) for extracellular variants (Supplementary Table 6). In addition, residues located in cytoplasmic loop 2 and extracellular loop 5 result in loss of function without exception (Figure 6, Supplementary Table 6).

Discussion

There is an unmet need to characterize the functional impact of *SLC6A1* variants. Additionally, no targeted therapies for patients with *SLC6A1* haploinsufficiency have been approved, although non-targeted therapies like 4-phenylbutyrate show promise by counteracting some of the effects of mutant *SLC6A1* in model systems[26].

Our results allowed us to classify 143 variants that had not previously been classified (Figure 1B). For the 78 variants for which a classification has previously been available, our results agreed well with prior knowledge, with two exceptions: One missense variant, Val342Met, that had been annotated as Pathogenic in ClinVar and has been observed two cohorts of epilepsy patients[9,27], had 44.3% function compared to the wildtype (95% confidence interval 33.1-55.6) and therefore was just above the threshold of 42.8%. It is unclear why the function of this variant is higher than what has been observed for other pathogenic variants. Another missense variant, Pro262Ser, had been annotated as Likely benign in ClinVar and had 42.3% function compared to the wildtype, just below the 42.8% threshold.

Gly234Ser, while unclassified on ClinVar, has previously been shown to reduce *SLC6A1* cell surface expression and GABA uptake as measured by a ^{3}H-GABA uptake assay in HeLa cells (33.5% of wildtype ± 6.4%)[12]. Our results did not replicate this finding and did not show a reduction in GABA uptake (88.6% of wildtype, 95% confidence interval 54.2-122.9). The patient where the Gly234Ser variant has been identified has been diagnosed with Lennox-Gastaut syndrome, with tonic, generalized tonic-clonic, complex partial and myoclonic-atonic seizures. His age of onset was 9.5 years, which is substantially lower than what is seen for other *SLC6A1* patients. Our results suggest that Gly234Ser is either not pathogenic or leads to relatively late onset disease.

The results of our functional assay agreed well with prior knowledge regarding the variant's pathogenicity (Figure 2) as well as with the results of previously published functional assays, where available (Supplementary Figure 3). This suggests that our results are also informative of the pathogenicity of previously unclassified variants. This is despite several caveats, such as our cellular system being based on HEK293 cells rather than the GABAergic neurons where *SLC6A1* is physiologically expressed. Assuming that GABA transporter function predicts pathogenicity, we estimate that incorporating the results of our assay into pathogenicity prediction will increase the diagnostic yield for *SLC6A1* haploinsufficiency by 40%.

We show that *SLC6A1* variants that are likely associated with Schizophrenia do not lead to strongly reduced GABA transport. This suggests either an association between hypomorphic variants and *SLC6A1*-Schizophrenia or an alternative mechanism of pathogenesis different from *SLC6A1* epilepsy or autism. An ongoing clinical trial of the SLC6A1 inhibitor Tiagabine is scheduled to complete in September 2022 (https://clinicaltrials.gov/ct2/show/NCT00179465). Our results, not showing evidence for Schizophrenia-associated variants being gain of function, do not support inhibition of SLC6A1 for Schizophrenia.

We observe that *SLC6A1* missense variants located in the protein's transmembrane domain are more likely to result in reduced GABA transport than those located in the cytoplasmic or extracellular portions of the protein and that within the transmembrane domain, loss of function

variants cluster around the pore region. This suggests that those variants result in reduced GABA transport by disrupting protein function directly, rather than resulting in reduced localization of the channel to the synapse.

We developed a high-throughput GABA trafficking assay to characterize functions of *SLC6A1* variants resulting in improvement of the prediction of pathogenicity for variants of unknown significance. Most variants have very limited available patient data and most are currently unclassified or classified as being Variants of Unknown Significance. Overall, the method we developed enabled the collection of experimental data on all clinically reported variants detected to date. These results enable the improvement of prediction of variant pathogenicity, substantially increase the predicted patient population, and provide a major advancement in the understanding of disease pathogenesis which may lead to improved therapeutic strategies.

Methods

Initial variant selection. We identified individuals with potentially pathogenic *SLC6A1* variants from multiple sources[1,2,3,8, 9,11,12,15,18,27,28, 29,30] (Supplementary Table 1). The individuals and organizations who collected and generated this original data bear no responsibility for further analysis or interpretation in this publication. After taking into account individuals represented in more than one cohort, the total number of carriers was 213. After taking into account variants present in multiple individuals and after excluding large insertion or deletion variants (indels) and most noncoding variants, we arrived at 182 variants we successfully tested. Of those, 90 were variants seen in epilepsy and developmental delay patients, 11 were seen in schizophrenia patients (3 of those with a high probability of being pathogenic), 73 were variants from various sources which are seen in epilepsy cases but for which pathogenicity has not been claimed and the remaining 5 were likely benign and included as negative controls (Supplementary Table 1).

Creation of *SLC6A1* constructs. The *SLC6A1* consensus coding-sequence (CCDS: CCDS2603.1) was synthesized by Genewiz (South Plainfield, NJ) and cloned into the pUC57-Kan plasmid with a CMV promoter driving the production of *SLC6A1*, opposite a reverse-oriented EF1-alpha promoter driving a Beta-lactamase reporter. Genscript (Piscataway, NJ) performed site-specific mutagenesis to generate 184 distinct variants. Mutagenesis was confirmed by Sanger sequencing.

CRISPR/Cas9 knockout of *SLC6A1* in HEK293T cells. We generated a *SLC6A1* deficient cell line, using HEK293T transfected with CRISPR/Cas9 RNP using the following crRNAs targeting exon 5 with PAM in bold:

```
Hs.Cas9.SLC6A1.1.AD:  5’-[AGTGGCCAGCGGATCTGACCTGG]-3’
Hs.Cas9.SLC6A1.1.AQ: 5’-[ATACACAAGGATCCAGGCGATGG]-3’
```

Briefly, we annealed equimolar ratios of ATTO 550 labeled tracrRNA (IDT Cat. 1075928) and the above crRNA. Twelve picomoles of the resulting gRNA were combined with 104 pmol HiFi Cas9 (IDT 1081061) to form RNP’s, according to the manufacturer’s protocol. 2E5 HEK293T cells were transfected with RNP using a 96-well Shuttle nucleofector (Lonza) with Amaxa SF solution (Lonza) and program CM-138. Transfected cells were seeded into a 96-well plate and cultured for 3 days at 37°C, with 5% CO2, in 200 µL of high-glucose DMEM (Invitrogen, Cat.

11995-065) supplemented with 10% (v/v) Fetal Bovine Serum (VWR, Cat. 97068-085) and 1X GlutaMAX-I™ (GIBCO, Cat. 35050-061).

Next, single cells were sorted into 96-well plates containing 200µL media using a FACSMelody cell sorter (BD Biosciences). Clones were monitored by imaging (Cell Metric, Solentim) until confluent, then expanded into six-well plates and genotyped. Genomic DNA was extracted using Quanta Extracta solution (Quanta Biosciences, Cat. 95091). DNA amplicons containing the gRNA target-region were produced using the following primers (*SLC6A1*-binding sequences in bold):

```
FW:  5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTCCCACCAGCTCTGTGTA-3'
REV: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATCCAGTGCCCTCCTGTCC-3'
```

A second round of PCR incorporated NGS barcodes and Illumina sequencing-adapters via Nextera™–Compatible Indexing Primers, which annealed to the adapters (underlined above) added during the first PCR. Amplicons were purified and pooled according to the manufacturer's instructions (Illumina Document # 15031942 v05), then paired-end sequenced (149x149) on a MiSeq 550 with a v2 reagent kit (Illumina, Cat. MS-102-2002). Indels were quantified using CRISPResso (version 2.0.27) to identify a clonal population with homozygous *SLC6A1* deficiency. The clonal cell line selected for experimentation was homozygous for a 37bp deletion in exon 5:

```
5'[GCGCAACATGCATCAGATGACGGACGGGCTGGATAAGCCAGGTCCTGGATCCTTGTGTATTTCTGTATCTGGAAG
GGTGTTGGCTGGACTGGAAAG]-3'
```

Resulting in the following truncated *SLC6A1* protein coding sequence:

```
MATNGSKVADGQISTEVSEAPVANDKPKTLVVKVQKKAADLPDRDTWKGRFDFLMSCVGYAIGLGNVWRFPYLCGKNG
GGAFLIPYFLTLIFAGVPLFLLECSLGQYTSIGGLGVWKLAPMFKGVGLAAAVLSFWLNIYYIVIISWAIYYLYNSFT
TTLPWKQCDNPWNTDRCFSNYSMVNTTNMTSAVVEFWERNMHQMTDGLDKPGPGSLCISVSGRVLAGLERWSTFQPHT
PTSC*
```

The GABA uptake assay was used to confirm loss of SLC6A1 activity in the knockout line.

Transfection of Variants and GABA reuptake assay. HEK293T *SLC6A1*-deficient cells were transfected in sextuplicate, using a 96-well Shuttle nucleofector (Lonza) integrated onto a MICROLAB STAR liquid-handling robot (Hamilton Robotics, Reno, NV). Briefly, 250ng of plasmid per 300,000 cells was transfected using Amaxa solution SF (Lonza) and program CM-138. Cells were seeded onto 96-well plates, then cultured for 3 days in 200µL media, as described above.

Prior to the reuptake assay, cells were centrifuged at 300xg for 5 min and all media was aspirated. Cells were then incubated at 37°C for 15 min with 80µL of Pre-Incubation solution containing 140mM NaCl, 5mM KCl, 2mM CaCl2, 1mM MgSO4, 2nM Glucose, 2.5mM HEPES at pH 7.4 and osmolarity 310. Pre-incubation solution was aspirated, following centrifugation at 300xg for 5min. Cells were incubated for 30 min at room temperature with 100µL of 4800nM d_6-GABA (SIGMA, Cat. 615587) in pre-incubation solution. After centrifuging for 5 min at 300xg,

the uptake solution was aspirated, and cells were washed twice with 80µL preincubation solution. Supernatant was aspirated and plates were flash frozen at -80°C.
Cells were harvested by adding 60 µL MPER (Thermo Fisher Scientific, Cat. 78501) and lysed with rigorous pipetting. 10 µL of lysate was used to determine protein concentrations with a Pierce BCA Protein Assay (Thermo Fisher Scientific, Cat. 23227), according to the manufacturer's instructions.

Beta-lactamase (BLA) activity was determined by measuring hydrolysis of d7-penicillin G (Toronto Research Chemicals, Cat. B288600) into d7-5R,6R-benzylpenicilloic acid (d7-BPA). The BLA Assay Cocktail consisted of 60 µM d7-penicillin G substrate and 1.5 µM 5R,6R-benzylpenicilloic acid (Toronto Research Chemicals, Cat. B288593), in 50 mM Tris-HCl (pH 7.5±0.02). For absolute quantification, separate BPA and d6-GABA standard curves were prepared with the following ranges: 120µM, 60µM, 20µM, 10µM, 2µM, 1µM and 200nM for BPA; and 6µM, 3µM, 1µM, 500nM, 100nM, 50nM, 10nM, 1nM and 500pM for GABA. To avoid hydrolysis, d7-penicillin G and BPA were stored dry, under inert gas at -20°C, and the BLA Assay Cocktail and BPA standards were prepared fresh. BPA standards and lysates were processed in parallel with sample lysates. Briefly, 20 µL of BLA Assay Cocktail was combined with 20 µL lysate or BPA standard, then incubated for 1 hour at 37°C, on an orbital shaker set to 500 rpm. The reaction was quenched with 160 µL of acetonitrile containing 25nM d2-GABA (Sigma, Cat. 617458), then vortexed at 1500 rpm for 5 min and centrifuged at 3000×G for 10 min. 100 µL of supernatant was combined with an equal volume of HPLC-grade water and reserved for the BLA-assay injection. An additional 20µL of supernatant was combined with 10 µL 100 mM Sodium Carbonate in a 96-well plate, then vortexed for 5 min at 1500 rpm. GABA was derivatized with 10 µL 6% (v/v) Benzoyl-Chloride in acetonitrile, vortexing at 1500 rpm for 5 min. The reaction was quenched with 0.2% Formic Acid (v/v) in water, then centrifuged at 4000 rpm for 4 min, to produce the final GABA analyte.

UPLC-MS/MS analysis was performed on an Agilent 6495 Triple-Quadrupole LC/MS, equipped with an Agilent 1290 Infinity II series HPLC. Separation of analytes was achieved using an Acquity UPLC BEH C18 1.7um, 2. x 50 mm column (Waters Crop., Cat. 176001692) at 40°C. Mobile phase A (MPA) contained water with 0.1% formic acid, and mobile phase B (MPB) was acetonitrile with 0.1% formic acid. The BLA and GABA analytes were analyzed in separate, 10 µL injections, utilizing the same LC-MSMS instruments and mobile phases, but different gradients. For the BLA assay, a flowrate of 0.4 mL/min was applied with the following gradient: 0-1 min, 5% mobile phase B; 1-2 min, 5%-95% MPB; 2-2.9 min, 95% MPB; 2.9-3 min 5% MPB. The enzymatic product (d7-5R,6R-benzylpenicilloic acid) and internal standard (5R,6R-benzylpenicilloic acid) were detected through multiple reaction monitoring (MRM) using the following transitions: 360.3 > 160 and 353.3 > 160, respectively. For the GABA assay, the flowrate was set to 0.4 mL/min with the following gradient: 0-4 min, 5-42% mobile phase B; 4-4.4 min, 42%-95% MPB; 4.4-4.9 min 95% MPB; 4.9-5% MPB. We monitored the GABA transported into cells by SLC6A1 (d6-GABA-BZ, following derivatization) and the internal standard (benzoylated d2-GABA) by MRM, using distinct transitions: d6-GABA-BZ, 214.1 > 105; d2-GABA-BZ, 210.1 > 105. In both assays, the dwell time was set to 50 msec. The cone voltage in both assays was 4500 V and the collision energy varied, using 17V in the GABA assay and 11V for the BLA method. Method scheduling was performed in Mass Hunter (Agilent,

version 10.0). Peak calling and standard-curve calculations were executed in QQQ Quantitative Analysis (Agilent, version 10.1).

Data analysis. BLA activity (nmol/hr/mg_protein) was determined using nanomoles d7-BPA, as calculated from the BPA standard curve, then divided by the incubation time (1 hour). This rate was further divided by the milligrams of lysate protein, as a proxy for variable cell counts between wells. SLC6A1 re-uptake activity (nmol/hr/mg_protein) was calculated similarly, by determining the nanomoles of d6-GABA from the d2-GABA standard curve, then dividing by both the incubation time (0.5 hours) and milligrams of protein in the lysate. To account for variable transfection and expression efficiencies, SLC6A1 re-uptake activity (nmol/hr/mg_protein) was then divided the BLA activity, to derive the "GABA_Uptake_Normalized_to_BLA_Activity" reported in Supplementary Table 3. While the BLA-normalized GABA activity is an accurate measure of a variant's transport activity, the functional consequence of a mutation is best described in relation to wildtype-SLC6A1 transport—to convey this, we calculated "Percent Wildtype" activities, where each variant's BLA-normalized SLC6A1-activity was divided by the average BLA-normalized SLC6A1 activity of all wildtype replicates present within an assay plate. Figures were generated in R apart from Figure 1, which was generated using SankeyMATIC (https://sankeymatic.com/) and Figure 6, which was generated using Protter (version 1.0)[31].

Distribution of measurement variation. To account for variability in transfection efficiencies, GABA re-uptake activity was normalized to activity of a Beta-lactamase (BLA) reporter present on the *SLC6A1* expression construct. BLA activity was determined by measuring hydrolysis of d7-penicillin G. While the BLA-normalized GABA re-uptake activity is an accurate measure of a variant's transport activity, there exists experimental variability that is biased towards *SLC6A1* variants with high activity. The experimental variation is heightened due to the high sensitivity of the mass spectrometry assays which are able to detect even slight variations that could be caused by miniscule pipetting errors. Ultimately biological replicates were included to access the variability in transporter activity of SLC6A1 variants which was normalized to activity in wildtype cells.

Protein structure mapping. Two figures were used to visualize the percent-wildtype reuptake of each variant in the context of protein structure. The predicted structure of SLC6A1 (AF-P30531-F1) was first acquired from AlphaFold (AlphaFold Monomer v2.0)[24,25]. For each variant tested in the assay, we overlayed a heatmap of percent-wildtype activities onto the corresponding residues within the AlphaFold image using PyMol (version 2.3.5). The resulting PyMol document is available as a .pse file on request. Secondly, a stylized representation of SCL6A1's topological domains was generated using Protter (version 1.0)[31] and domain annotations from UniProt (ID P30531) (Figure 6). Variants that displayed low reuptake activity (i.e. percent-wildtype activity less than 42.8%) in the assay were shaded in red and variants with high levels of reuptake activity (i.e. percent-wildtype activity greater than the 42.8%) were shaded in blue. The wild-type amino acid identities are preserved in the figure and the amino acid substitutions and reuptake activities for each variant are available in Supplementary Table 3.

Acknowledgements

We thank Amber Freed of SLC6A1Connect, Prof. Dennis Lal and Prof. Jing-Qiong Kang for review of the experimental design. This study incorporates data generated by the DECIPHER community. A full list of centers who contributed to the generation of the data is available from https://deciphergenomics.org/about/stats and via email from contact@deciphergenomics.org. Funding for the DECIPHER project was provided by the Wellcome Trust. We used the Invitae Explorer to explore variants and aggregate testing data from Invitae patients, including genomic and demographic data. We thank Alliance Pharma for support developing and running the BLA and GABA mass spectrometry assays.

Figures and Tables

Figure 1. (A) Schematic of GABA reuptake assay*. SLC6A1 deficient cells were transfected with plasmids producing SLC6A1 variants and assayed for GABA concentrations relative to a BLA reporter normalized to wild-type transfected cells.* ***(B) Summary of variant annotations from ClinVar and activity-assay classifications****. After quantifying reuptake activity in our assay, we demonstrate the decrease in VUS and updated classification of variants.*

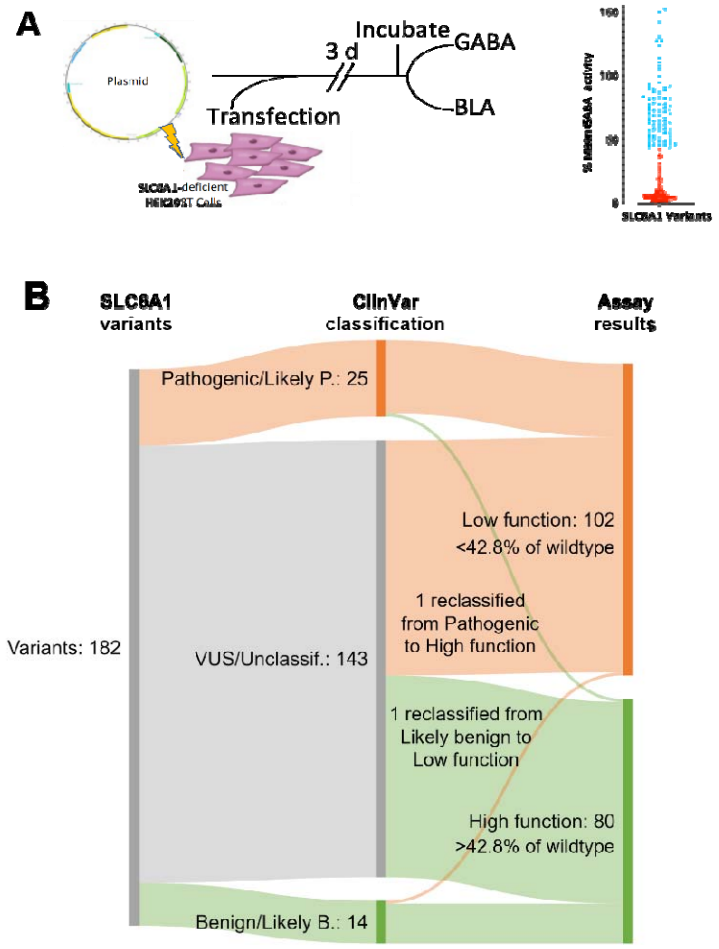

Figure 2. Variant reuptake activities per ClinVar annotation. *Reuptake activities for each pathogenicity group in ClinVar are summarized by boxplot. The cutoff (42.8% wildtype reuptake activity) for demarcating low-activity variants from high-activity variants is plotted as a dashed vertical line.*

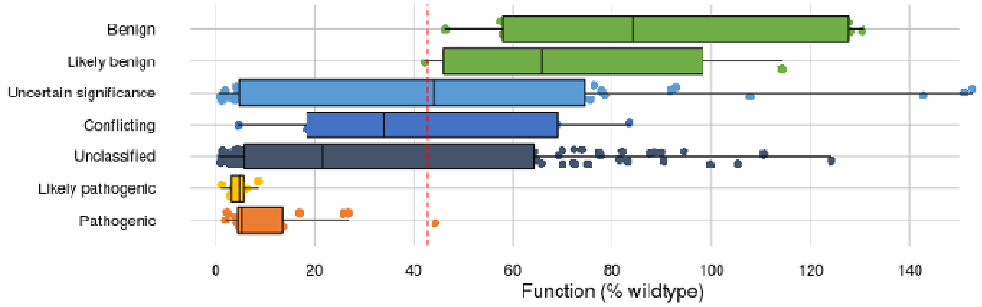

Figure 3. Variant reuptake activities per category for variant impact. *Variant reuptake activities are plotted by variant-impact, as labelled in ClinVar. The dashed, vertical line represents the threshold (42.8% wildtype reuptake activity) differentiating low activity from high-activity variants.*

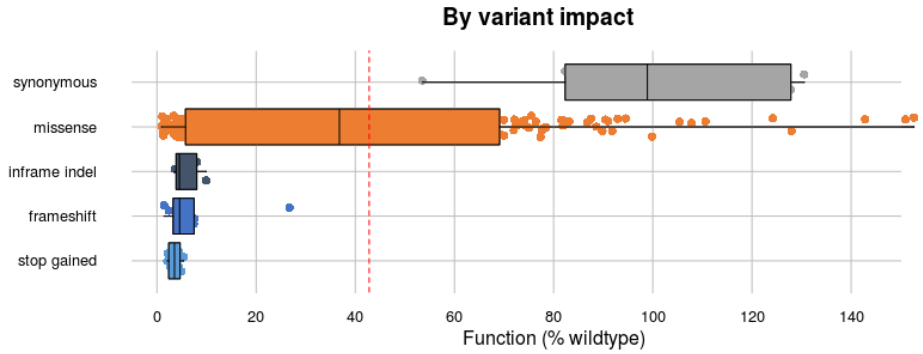

Figure 4. Variant reuptake activities by disease-association from literature. *Variant reuptake activities are plotted based on disease associations from patient literature. The threshold separating low activity from high-activity variants is represented by a vertical, dashed line.*

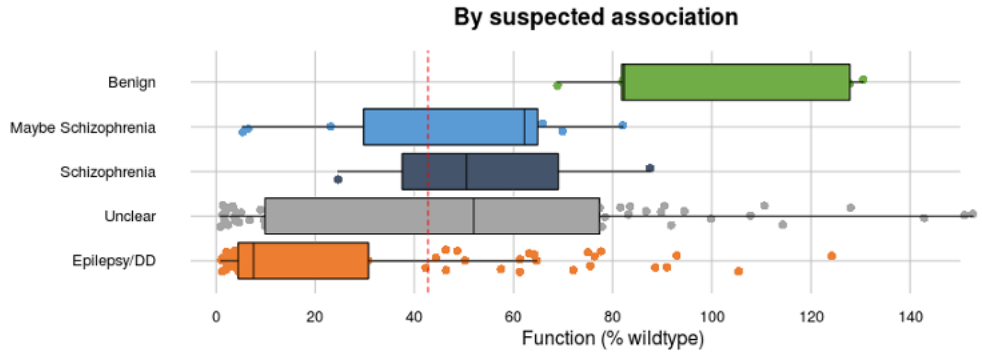

Figure 5. Variant reuptake activities by gnomAD allele count*. For each tier of allele counts in gnomAD, boxplots of variant reuptake activities are provided. The vertical, dashed line represents our cutoff (42.8% wildtype reuptake activity) that defines low-activity from high-activity variants. Definitions for suspected associations are defined in Supplementary Table 1B.*

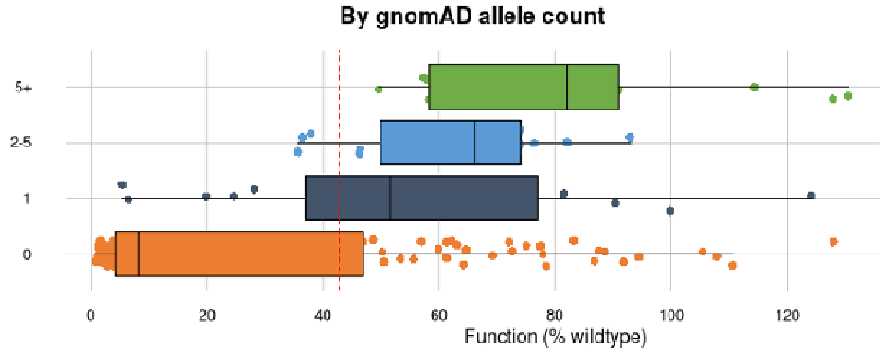

Figure 6. Overview of reuptake assay and activities mapped to SLC6A1 structure*. SLC6A1 topological domain annotations from UniProt demonstrate the relationship between variant localization and impact on percent wildtype reuptake activity in our assay. For each variant, the wild-type residues are displayed and the corresponding amino acid substitutions are available in Supplementary Table 3 with percent wildtype activities.*

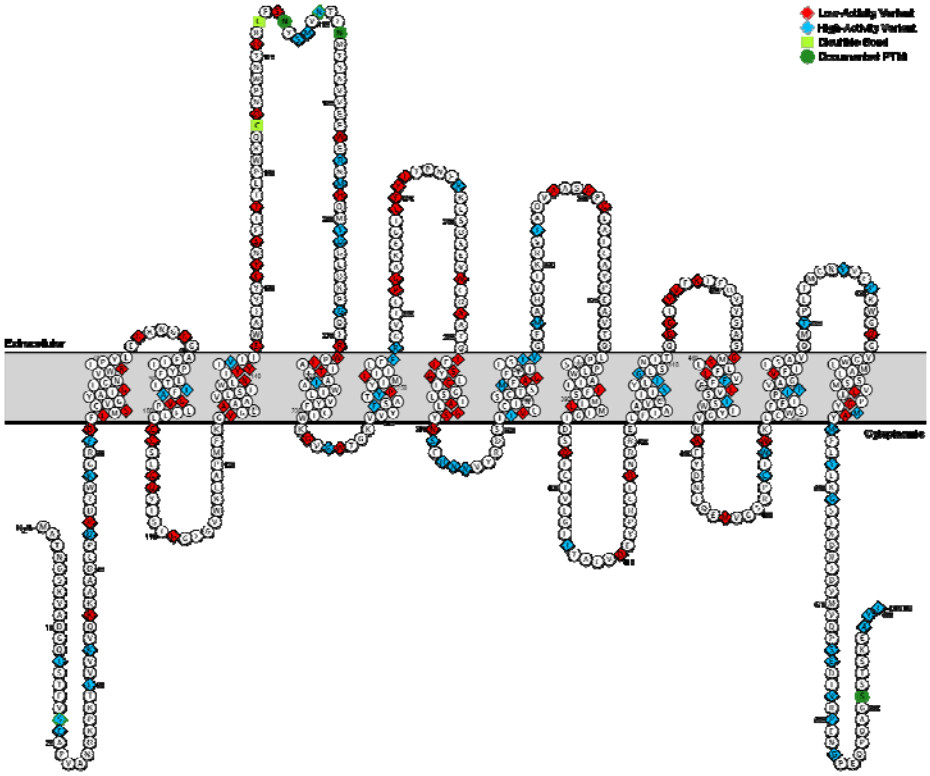

Table 1. Number of variants classified in ClinVar and the number of patients with those variants across all cohorts. Functional assay data suggests an increase in the number of pathogenic variants from 25 to 48 and an increase in the number of patients with SLC6A1 haploinsuficiency from 63 to 88.

	Variants	**Patients**
Variant classified in ClinVar	84	138
ClinVar classification Pathogenic/Likely pathogenic	25 (29.8%)	63 (45.7%)
Low function (<42.8% of wildtype)	48 (57.1%)	88 (63.8%)

Conflicts of interest

All authors are current or past employees of BioMarin Pharmaceutical.

References

1. Carvill, G. L. *et al.* Mutations in the GABA Transporter SLC6A1 Cause Epilepsy with Myoclonic-Atonic Seizures. *Am J Hum Genet* **96**, 808-815, doi:10.1016/j.ajhg.2015.02.016 (2015).
2. Rees, E. *et al.* De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat Neurosci* **23**, 179-184, doi:10.1038/s41593-019-0565-2 (2020).
3. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e523, doi:10.1016/j.cell.2019.12.036 (2020).
4. Lopez-Rivera, J. A. *et al.* A catalogue of new incidence estimates of monogenic neurodevelopmental disorders caused by de novo variants. *Brain* **143**, 1099-1105, doi:10.1093/brain/awaa051 (2020).
5. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science, doi:https://doi.org/10.1126/science.aaz1776 (2020).
6. Cheng, J. Z. & Carvill, G. L. Pathogenic mechanisms underlying SLC6A1 variant-mediated neurodevelopmental disorders. *Brain* **144**, 2237-2239, doi:10.1093/brain/awab259 (2021).
7. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489, doi:10.1038/nature10523 (2011).
8. Goodspeed, K. *et al.* Current knowledge of SLC6A1-related neurodevelopmental disorders. *Brain Commun* **2**, fcaa170, doi:10.1093/braincomms/fcaa170 (2020).
9. Johannesen, K. M. *et al.* Defining the phenotypic spectrum of SLC6A1 mutations. *Epilepsia* **59**, 389-402, doi:10.1111/epi.13986 (2018).
10. Mermer, F. *et al.* Common molecular mechanisms of SLC6A1 variant-mediated neurodevelopmental disorders in astrocytes and neurons. *Brain* **144**, 2499-2512, doi:10.1093/brain/awab207 (2021).
11. Mattison, K. A. *et al.* SLC6A1 variants identified in epilepsy patients reduce gamma-aminobutyric acid transport. *Epilepsia* **59**, e135-e141, doi:10.1111/epi.14531 (2018).
12. Cai, K. *et al.* A missense mutation in SLC6A1 associated with Lennox-Gastaut syndrome impairs GABA transporter 1 protein trafficking and function. *Exp Neurol* **320**, 112973, doi:10.1016/j.expneurol.2019.112973 (2019).
13. Wang, J. *et al.* Endoplasmic reticulum retention and degradation of a mutation in SLC6A1 associated with epilepsy and autism. *Mol Brain* **13**, 76, doi:10.1186/s13041-020-00612-6 (2020).
14. Poliquin, S. *et al.* Genetic mosaicism, intrafamilial phenotypic heterogeneity, and molecular defects of a novel missense SLC6A1 mutation associated with epilepsy and ADHD. *Exp Neurol* **342**, 113723, doi:10.1016/j.expneurol.2021.113723 (2021).
15. Epi25 Collaborative. Electronic address, s. b. u. e. a. & Epi, C. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. Am J Hum Genet 105, 267-282, doi:10.1016/j.ajhg.2019.05.020 (2019).
16. Genscript. GenScript Codon Usage Frequency Table(chart) Tool, <https://www.genscript.com/tools/codon-frequency-table> (February 8, 2022)
17. Lindquist, B. E., Voskobiynyk, Y. & Paz, J. T. Electrobehavioral phenotype and seizure pharmacosensitivity in a novel mouse model of patient-derived SLC6A1 S295L

mutation-associated neurodevelopmental epilepsy. bioRxiv, doi:10.1101/2021.12.17.473036 (2021).
18. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434-443, doi:10.1038/s41586-020-2308-7 (2020).
19. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. Nat Protoc 11, 1-9, doi:10.1038/nprot.2015.123 (2016).
20. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
21. Ernst, C. et al. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. BMC Med Genomics 11, 35, doi:10.1186/s12920-018-0353-y (2018).
22. Pshennikova, V. G. et al. Comparison of Predictive In Silico Tools on Missense Variants in GJB2, GJB6, and GJB3 Genes Associated with Autosomal Recessive Deafness 1A (DFNB1A). ScientificWorldJournal 2019, 5198931, doi:10.1155/2019/5198931 (2019).
23. The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC.
24. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589, doi:10.1038/s41586-021-03819-2 (2021).
25. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50, D439-D444, doi:10.1093/nar/gkab1061 (2022).
26. Nwosu, G. et al. 4-phenylbutyrate restored GABA uptake and reduced seizures in SLC6A1 variants-mediated disorders. bioRxiv, doi:10.1101/2021.12.08.471799 (2021).
27. Truty, R. et al. Possible precision medicine implications from genetic testing using combined detection of sequence and intragenic copy number variants in a large cohort with childhood epilepsy. Epilepsia Open 4, 397-408, doi:10.1002/epi4.12348 (2019).
28. Heyne, H. O. et al. De novo variants in neurodevelopmental disorders with epilepsy. Nat Genet 50, 1048-1053, doi:10.1038/s41588-018-0143-7 (2018).
29. Firth, H. V. et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84, 524-533, doi:10.1016/j.ajhg.2009.03.010 (2009).
30. Symonds, J. D. et al. Incidence and phenotypes of childhood-onset genetic epilepsies: a prospective population-based national cohort. Brain 142, 2303-2318, doi:10.1093/brain/awz195 (2019).
31. Omasits, U., Ahrens, C. H., Muller, S. & Wollscheid, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30, 884-886, doi:10.1093/bioinformatics/btt607 (2014).