An Integrative *In Silico* System for Predicting Dysregulated Genes in the Human Epileptic Focus: Application to SLC Transporters

Nasir Mirza,[1] Olga Vasieva,[2] Richard Appleton,[3] Sasha Burn,[4] Daniel Carr,[1] Daniel Crooks,[5] Daniel du Plessis,[6] Roderick Duncan,[7] Jibril Osman Farah,[8] Vivek Josan,[9] Fabio Miyajima,[1] Rajiv Mohanraj,[10] Arif Shukralla,[10] Graeme J Sills,[1] Anthony G Marson[1,*] and Munir Pirmohamed[1,*]

1 Department of Molecular & Clinical Pharmacology, University of Liverpool, Liverpool L69 3GL, UK.

2 Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK.

3 The Roald Dahl EEG Unit, Paediatric Neurosciences Foundation, Alder Hey Children's NHS Foundation Trust, Liverpool L12 2AP, UK.

4 Department of Neurosurgery, Alder Hey Children's NHS Foundation Trust, Liverpool L12 2AP, UK.

5 Department of Neuropathology, the Walton Centre NHS Foundation Trust, Liverpool L9 7LJ, UK.

6 Department of Cellular Pathology, Salford Royal NHS Foundation Trust, Salford M6 8HD, UK.

7 Department of Neurology, Christchurch Hospital, Private Bag 4710, Christchurch 8140, New Zealand.

8 Department of Neurosurgery, the Walton Centre NHS Foundation Trust, Liverpool L9 7LJ, UK.

9 Department of Neurosurgery, Salford Royal NHS Foundation Trust, Salford M6 8HD, UK

10 Department of Neurology, Salford Royal NHS Foundation Trust, Salford M6 8HD, UK

Correspondence to: Dr Nasir Mirza,

Department of Molecular and Clinical Pharmacology,

University of Liverpool,

Block A: Waterhouse Building,

1-5 Brownlow Street,

Liverpool L69 3GL,

UK.

Tel +44 151 794 5557

Fax +44 151 794 5059

Email nasir.mirza@liv.ac.uk

*Equally contributing senior authors

Running title: Predicting disease genes

Keywords: Computational gene prioritization; Convergent functional genomics; Hippocampal sclerosis; Epilepsy; Solute Carrier Transporters; Microarray; Pharmacoresistance.

Number of text pages: 16

Number of words: 3,525

Number of references: 29

Number of figures: 1

Number of tables: 5

Summary

Objective

Many different gene families are currently being investigated for their potential role in epilepsy and in the response to antiepileptics. A common research challenge is identifying the members of a gene family that are most significantly dysregulated within the human epileptic focus, before taking them forward for resource-intensive functional studies. Published data about transcriptomic changes within the human epileptic focus remains incomplete. A need exists for an accurate *in silico* system for the prediction of dysregulated genes within the epileptic focus. We present such a bioinformatic system. We demonstrate the validity of our approach by applying it to the solute carrier (SLC) gene family. There are >400 known SLCs. SLCs have never been systematically studied in epilepsy.

Methods

Using our *in silico* system, we predicted the SLCs likely to be dysregulated in the epileptic focus. We validated our *in silico* predictions by identifying *ex vivo* the SLCs dysregulated in epileptic foci, and determining the overlap between our *in silico* and *ex vivo* results. For the *ex vivo* analysis, we used a custom oligonucletoide microarray containing exon probes for all known SLCs to analyse 24 hippocampal samples obtained from surgery for pharmacoresistant mesial temporal lobe epilepsy and 24 hippocampal samples from normal post-mortem controls.

Results

There was a highly significant ($p < 9.99 \times 10^{-7}$) overlap between the genes identified by our *in silico* and *ex vivo* strategies. The SLCs identified were either metal ion exchangers or

neurotransmitter transporters, which are likely to play a part in epilepsy by influencing neuronal excitability.

Significance

The identified SLCs are most likely to mediate pharmacoresistance in epilepsy by enhancing the intrinsic severity of epilepsy, but further functional work will be needed to fully evaluate their role. Our successful *in silico* strategy can be adapted in order to prioritize genes relevant to epilepsy from other gene families.

Key Points:

1. We have developed a novel *in silico* strategy for predicting dysregulated genes in the human epileptic focus.
2. A novel and appealing feature of our proposed system is the ability to validate the predicted genes *in silico*.
3. Using this approach, followed by *ex vivo* validation, we have identified the SLC genes dysregulated in the epileptic human hippocampus.

Background

Epilepsy is amongst the most common neurological disorders, with a prevalence of up to 1% of the population.[1] Epilepsy is a complex genetic disease: genes from a number of different functional categories are likely to be involved in its pathogenesis, and many different gene families are currently being actively investigated for their potential role in epilepsy and in the response to antiepileptics[2] A common research challenge is identifying the most dysregulated genes belonging to a particular functional group or gene family, before taking them forward for more arduous and resource-intensive functional studies. However, the low power of microarray studies on brain tissue from epilepsy surgery, because of their small sample sizes, means that published data about transcriptomic changes within the human epileptic focus is still incomplete.[3] Identifying dysregulated genes in human epileptic brain tissue is far from a trivial undertaking: obtaining sufficient amounts of precious brain tissue and the cost of quantifying the expression of potentially hundreds of constituents of a gene family is a major challenge. It will be of use, therefore, to have an easily applicable strategy that uses publically available data and computational tools to identify and then validate *in silico* genes dysregulated in the epileptic focus. The identified genes could then be further verified *ex vivo.* In the present work, we present such a strategy, and demonstrate its validity by applying it to the solute carrier (SLC) superfamily.

Most patients with epilepsy become seizure-free with antiepileptic drug (AED) therapy. However, approximately 30% of epilepsy patients are pharmacoresistant: they continue to experience seizures even if treated with various AEDs at maximal dosages.[4] A number of hypotheses have been put forward to explain pharmacoresistance;[5] one of these is the multidrug transporter (MDT) hypothesis. According to this hypothesis, pharmacoresistance is a consequence of decreased drug concentrations at the epileptic focus secondary to a

localized dysregulation of drug transporters,[5] which could either increase drug efflux from or reduce influx into the epileptic focus. The largest superfamily of MDTs is the SLC superfamily which are mainly influx transporters.[6] Down regulation of SLC transporters that are normally expressed at significant levels in the epileptic focus could thus potentially contribute to the development of pharmacoresistance. There are approximately 400 known SLC proteins in total, but very few of these have been studied in epilepsy. Mesial temporal lobe epilepsy (MTLE), in which seizures originate from the hippocampus, is the most common cause of refractory epilepsy and the most common indication for epilepsy surgery.[7] We focused, therefore, on SLCs most relevant to this phenotype. We devised a comprehensive *in silico* approach that exploited relevant published data and built upon them using computational tools in order to identify the SLCs which are downregulated in the pharmacoresistant epileptic focus. We then validated the output from the *in silico* approach using an analysis of human brain tissue *ex vivo*.

We anticipate that our *in silico* strategy will also be adaptable to other pathologies and gene families.

Methods:

In silico analysis

Our *in silico* strategy was as follows:

1. Rank SLCs based on the strength of the published evidence of their downregulation in pharmacoresistant epileptic foci, if such data were available.
2. For SLCs with no such published data, rank genes based on the computationally-determined likelihood of their down regulation in pharmacoresistant epileptic foci.

The individual steps and processes are illustrated schematically in Fig. 1.

Integrative analysis of microarray studies: We started by integrating the results of large-scale gene expression profiling studies on brain tissue from epilepsy surgery—details of this analysis can be found in our previously published study.[3] Data relating to SLCs from this analysis were carried forward into a convergent functional genomics approach.

Convergent functional genomics (CFG): CFG is an approach for prioritizing candidate genes for complex disorders by integrating and tabulating multiple lines of evidence, such as human and animal-model gene and protein expression data.[8] In order to implement the CFG approach, each SLC transporter was scored on the strength of the evidence of down-regulation of the gene or protein in epileptic foci of humans or of animal models of pharmacoresistant epilepsy. The search strategy and search terms used in order to find this data is detailed in the Supplementary Material. Details of the scoring system are given in Table 1.

Computational gene prioritization (CGP): To perform CGP, we used a freely-available bioinformatics software application called Endeavour[9] whose key feature is that it uses multiple genomic data sources (e.g. sequence, expression, literature and annotation) to estimate how promising a candidate gene is by measuring its similarity with a set of training genes. A detailed explanation of the principles and protocols underlying Endeavour can be found in its original publication.[9] Our training gene set comprised the top scoring SLCs from the CFG step, with strong replicated published evidence in support of a role in epilepsy pharmacoresistance. The validity of the CGP result is critically dependent on the quality of the training set. Therefore, to ensure the quality of the training set, we performed a comprehensive literature review of our training genes to show that they are highly likely to be involved in epilepsy pharmacoresistance, and performed leave-one-out analysis (LOOA); please see details in the Supplementary Material.

Our candidate gene set comprised SLCs that are most abundantly expressed in normal human hippocampus. In order to identify the most abundant SLCs, we performed a 'meta-analysis' of publically available genome-wide gene expression studies. The methods of this meta-analysis are detailed in the supplementary material and summarized here. Briefly: (1) We downloaded genome-wide gene expression data for 73 normal hippocampal tissue samples from four different published Affymetrix GeneChip Human Genome U133 Plus 2.0 Array studies; (2) We filtered out non-expressed genes based on 'fraction present' according to Affymetrix Microarray Suite 5.0 (MAS5) detection call—we only retained genes present in more than 50% of the samples. (3) The retained genes were then ranked by transcript abundance—we used rank product (a non-parametric statistic that detects items that are consistently highly ranked in a number of lists) to create one combined hierarchical list of transcript abundance. (4) From this analysis, the 50 most abundant SLCs were chosen as the candidate gene set—all these genes had a false discovery rate (FDR) of less than 5% for being ranked amongst the 50 most abundant SLC genes.

Computational verification of CGP results: To validate our CGP approach and prioritized lists, the following steps were undertaken:

We set up an *in silico* experiment. For this experiment, we created (i) a 'positive control' training set comprising the most consistently down-regulated genes from our previously mentioned integrative analysis of gene-expression profiling studies on brain tissue from epilepsy surgery[3] and (ii) a 'negative control' training set consisting of randomly chosen genes from our integrative analysis. Both the control training sets were used to perform CGP of the 50 candidate SLC genes. The list of genes comprising the control training sets can be found in the Supplement. We hypothesized that the 'positive control ranking' would be significantly similar to the original ranking, whereas the 'negative control ranking' would be dissimilar to the original ranking. We compared the original CGP ranking to the positive

control and negative control CGP rankings in turn, using objective validated statistical methods: hypergeometric equation to determine the statistical significance of the size of the overlap between lists, and Bioconductor package OrderedList[10] to calculate the statistical significance of rank order similarity (see details in the Supplementary Material).

Creation of the final *in silico* gene list: To create the final *in silico* list of 20 genes, we amalgamated the CFG and CGP results. Giving precedence to genes with direct published evidence of down regulation, CFG genes were placed at the top of the list in rank order, followed by the CGP genes, with a cut-off being applied at a total of 20 genes. As our chosen cut-off of 20 total genes could be deemed arbitrary, we created two further lists of 10 and 30 total genes for testing.

Ex vivo analysis

For *ex vivo* validation, we performed exon array analysis of 24 hippocampal samples from surgery for pharmacoresistant mesial temporal lobe epilepsy and obtained 24 histologically-normal post-mortem hippocampal samples from donors with no known brain disease. The whole-transcript amplification protocol of exon arrays allows more accurate measurement of gene expression than standard microarrays.[11-13] Details of the *ex vivo* methods are provided in the Supplementary Material.

Overlap between computational and microarray results

The statistical significance of the overlap between the computationally-generated gene list and the list of SLCs found to be significantly downregulated in our microarray was calculated using the R function phyper.

Results

In silico analysis

Integrative analysis of microarray studies: The full results of this analysis can be found in our previously published study.[3] Data relating to SLCs from this analysis were carried forward into a convergent functional genomics step.

Convergent functional genomics (CFG): The results of the CFG analysis showed that the final score was 1.5 for *SLC1A2, SLC1A3, SLC17A7;* 1.0 for *SLC24A3, SLC2A3, SLC12A5;* and 0.5 for *SLC30A3, SLC8A2, SLC47A1, SLC4A8, SLC15A2, SLC8A1, SLC17A1, SLC6A20, SLC16A1, SLC1A6, SLC1A1*. References for these data used can be found in the Supplementary Material.

There was data in the published literature relating only to these 17 SLC transporters. We also wished to obtain a prioritized list of SLC transporters that have hence far been overlooked in epilepsy research. To do this, we used the software 'Endeavour' to perform computational gene prioritization.

Computational Gene Prioritization (CGP): To create a training gene set, we utilized the six SLC transporters (*SLC1A2, SLC1A3, SLC2A3, SLC12A5, SLC17A7* and *SLC24A3*) with the highest CFG scores and with replicated evidence in support of a role in epilepsy pharmacoresistance. To assess the quality of our chosen training set, we performed leave-one-out cross-validation. The results of this analysis showed that four defector genes were ranked first, while two were ranked second (see Supplementary Material). This demonstrates that the training set is homogenous. We also performed a comprehensive literature review of our training genes to show that they are highly likely to be involved in epilepsy (see Supplementary Material). For example, studies in animal models reveal that a reduction in *SLC12A5* expression results in an increased susceptibility to the development of seizures: complete deletion of *SLC12A5* is incompatible with life,[14] a 95% reduction in *SLC12A5*

expression results in handling-induced seizure behaviour,[15] and heterozygous animals have a lower threshold for epileptic seizures—electrophysiological measurements in the hippocampus show hyperexcitability and animals demonstrate a twofold increase in pentylenetetrazole-induced seizures.[15] *SLC2A3* heterozygous null mice have a lower seizure threshold and all develop clinical or subclinical electroencephalographically-evident spontaneous seizures after hypoxic-ischaemia while wildtype mice do not.[16; 17] Monozygous mice deficient in *SLC1A2* show lethal spontaneous seizures.[18]

The candidate gene set for CGP comprised the 50 most abundant SLC transporters in the normal human hippocampus, identified by performing a microarray meta-analysis as described in the Methods section.

The results of the CGP are tabulated in the Supplementary Material. As *in silico* verification, this CGP ranking was compared with a 'positive control' CGP ranking, generated using a 'positive control' training set comprising the 10 genes most consistently downregulated in an integrative analysis of microarray studies on brain tissue from epilepsy surgery:[3] there was a statistically significant overlap in the top 20 genes of the two lists (p=0.004) and significant similarity between the order of genes in the two lists (p=0.01). In contrast, when the original CGP ranking was compared with a 'negative control' CGP ranking, generated using a 'negative control' training set of 10 randomly chosen genes from our integrative analysis, there was no significant overlap in the top 20 genes of the two lists (p=0.614) and no significant similarity between the order of genes in the two lists (p=0.10).

Creation of the final *in silico* gene list: To create the final *in silico* list of 20 genes, we amalgamated the CFG and CGP results. Giving precedence to genes with direct published evidence of down regulation, CFG genes were placed at the top of the list in rank order, followed by the CGP genes, with a cut-off being applied at a total of 20 genes. As our chosen

cut-off of 20 total genes could be deemed arbitrary, we created two further lists of 10 and 30 total genes for testing.

Ex vivo analysis

For *ex vixo* validation, we collected 24 hippocampal samples from surgery for pharmacoresistant mesial temporal lobe epilepsy and obtained 24 histologically-normal post-mortem hippocampal samples from donors with no known brain disease. Important patient and donor individual and sample characteristics are summarized in Table 3. Eighteen SLC genes (Table 4) were found to be significantly downregulated (FDR < 0.05, fold change ≥ 1.5). Detailed results of the *ex vivo* analysis are provided in the supplementary material. MIAME-compliant data has been deposited in ArrayExpress (E-MTAB-3123).

Overlap between computational and microarray results: We calculated the size of the overlap between the 18 SLC genes shown to be down-regulated by our microarray and our *in silico* gene lists of 10, 20 and 30 SLC genes respectively. All three overlaps were highly significant, with p-values of 8.41×10^{-7}, 7.86×10^{-6} and 9.99×10^{-7} respectively.

When creating the candidate gene list for CGP, we adopted a robust filtering approach, which has been described in detail above. We determined if our robust approach to candidate gene set selection improved the results of the CGP. If no filters were applied to the candidate gene set, the overlap between the top CGP and microarray results was not statistically significant (p=0.386). But by filtering into the candidate set genes which were present in more than 50% of meta-analysis samples and were amongst the 50 most abundant SLCs, the overlap of the top CGP and microarray results became statistically significant (p=0.009).

Discussion:

In silico strategy

We have developed a novel *in silico* pipeline for the prediction of dysregulated genes in the human epileptic focus. This methodical stepwise pipeline systematically integrates complementary strategies: integrative analysis of large-scale transcriptomic studies, CFG and CGP. Although this approach might be deemed intuitive, it has never been formalized, validated and presented before. This pipeline was highly effective in predicting significantly downregulated SLCs in the pharmacoresistant epileptic human hippocampus. We highlight the following noteworthy elements of our *in silico* strategy that are particularly responsible for its success:

This strategy is based on a foundation of good research practice, as the first steps are to collate and systematically integrate all relevant data already published, including the results of whole-genome microarray studies. When compiling the final *in silico* gene list, we give precedence to genes with direct published evidence of dysregulation; most biologists will agree with this strategy. Systematically collecting data and then scoring it using a pre-defined scoring system (CFG) is also important for creating an unbiased training set for CGP. The methodical stepwise nature of the pipeline will appeal to researchers. Another appealing aspect is that the output from each step is validated before carrying it forward into the next step. In particular, the prioritized gene list from the CGP step is validated using a novel approach, by performing an *in silico* experiment with positive and negative control training sets. The latter validation step is an especially valuable feature as it may alert researchers to erroneous CGP results before resource-intensive wet-lab experiments are undertaken.

An important feature of the pipeline is being selective in creating the candidate gene-set for CGP. This point has not been sufficiently highlighted in previous work on CGP. Our robust

approach to candidate gene set selection improved the results of the CGP—by filtering out SLCs from the candidate set that were less likely to be relevant, more relevant SLCs were ranked more favourably.

We acknowledge that the performance of this system with different phenotypes and gene groups will vary depending on the quality and quantity of published data available. However, when published data is sparse, adopting the robust systematic approaches advocated here becomes even more important.

Ex vivo validation

To ensure that our *ex vivo* validation was robust, we chose to make use of an exon array. Table 5 summarizes the functions of the 18 SLC transporters found to be downregulated in our exon array. As can be seen from Table 5, the vast majority of the identified SLC proteins are either small metal ion exchangers or transporters of neurotransmitters, particularly glutamate. Given the key roles played by ionic transport and glutamatergic transmission in neuronal function, it stands to reason that these should be the most important SLC proteins in the epileptic hippocampus. The potential role of these transporters in epilepsy is also supported by evidence from published literature. For example, *SLC12A5* is a neuron-specific potassium-chloride symporter expressed throughout the central nervous system.[19-21] Studies in animal models reveal that a reduction in SLC12A5 expression results in an increased susceptibility to the development of seizures: complete deletion of SLC12A5 is incompatible with life,[14] a 95% reduction in SLC12A5 expression results in handling-induced seizure behaviour,[15] and heterozygous animals have a lower threshold for epileptic seizures—electrophysiological measurements in the hippocampus show hyperexcitability and animals demonstrate a twofold increase in pentylenetetrazole-induced seizures.[15] An example of a neurotransmitter transporter is the glutamate transporter *SLC25A22*; mutations in this gene

cause various epilepsy syndromes.[22-27] Further examples of the potential role in epilepsy of the identified SLCs are provided in the Supplementary Material.

It should be noted that for SLC transporters of metal ions, there is as yet no evidence of non-endogenous substrate transport—their small endogenous ionic substrates are markedly dissimilar to most xenobiotics, so it might be expected that they are not readily involved in xenobiotic transport.[28] Similarly, there is no evidence as yet that the transporters of glutamate or of other neurotransmitters are able to transport therapeutic drugs. Therefore, while these SLCs potentially mediate pharmacoresistance in epilepsy, it is unlikely that they do this through altered transport of AEDs, but rather by enhancing the intrinsic severity of epilepsy.[29] It is of course possible that some of the genetic changes identified in this analysis are the consequence, rather than the cause, of refractory seizures. The causative role of the genes identified as potentially playing a part in the development of pharmacoresistance will need to be established with future functional studies.

Conclusions

We have developed a novel *in silico* strategy that can aid epilepsy research by prioritizing specific genes for study from within large gene sets. We have identified the most significantly downregulated SLCs in the pharmacoresistant epileptic human hippocampus. The role of these SLCs in the epileptic hippocampus will need to be defined through future functional studies.

Disclosure

None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Acknowledgements

NM was funded by a Medical Research Council Research Training Fellowship. MP is an NIHR Senior Investigator, and acknowledges support from the NHS Chair of Pharmacogenetics and the MRC Centre for Drug Safety Science.

Supplementary material

Supplementary material is available online.

References

1. Sander JW. The natural history of epilepsy in the era of new antiepileptic drugs and surgical treatment. *Epilepsia* 2003;44 Suppl 1:17-20.
2. Dibbens LM, Heron SE, Mulley JC. A polygenic heterogeneity model for common epilepsies with complex genetics. *Genes Brain Behav* 2007;6:593-597.
3. Mirza N, Vasieva O, Marson AG, et al. Exploring the genomic basis of pharmacoresistance in epilepsy: an integrative analysis of large-scale gene expression profiling studies on brain tissue from epilepsy surgery. *Hum Mol Genet* 2011;20:4381-4394.
4. Shorvon SD. The epidemiology and treatment of chronic and refractory epilepsy. *Epilepsia* 1996;37 Suppl 2:S1-S3.
5. Chayasirisobhon S. The mechanisms of medically refractory temporal lobe epilepsy. *Acta Neurol Taiwan* 2009;18:155-160.
6. Huang Y, Sadee W. Membrane transporters and channels in chemoresistance and -sensitivity of tumor cells. *Cancer Lett* 2006;239:168-182.
7. Engel J, Jr. A Greater Role for Surgical Treatment of Epilepsy: Why and When? *Epilepsy Curr* 2003;3:37-40.
8. Le-Niculescu H, McFarland MJ, Mamidipalli S, et al. Convergent Functional Genomics of bipolar disorder: from animal model pharmacogenomics to human genetics and biomarkers. *Neurosci Biobehav Rev* 2007;31:897-903.
9. Tranchevent LC, Barriot R, Yu S, et al. ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucleic Acids Res* 2008;36:W377-384.
10. Lottaz C, Yang X, Scheid S, et al. OrderedList--a bioconductor package for detecting similarity in ordered gene lists. *Bioinformatics* 2006;22:2315-2316.
11. Kapur K, Xing Y, Ouyang Z, et al. Exon arrays provide accurate assessments of gene expression. *Genome Biol* 2007;8:R82.
12. Xing Y, Ouyang Z, Kapur K, et al. Assessing the conservation of mammalian gene expression using high-density exon arrays. *Mol Biol Evol* 2007;24:1283-1285.
13. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical software. *Brief Bioinform* 2011;12:634-644.
14. Hubner CA, Stein V, Hermans-Borgmeyer I, et al. Disruption of KCC2 reveals an essential role of K-Cl cotransport already in early synaptic inhibition. *Neuron* 2001;30:515-524.
15. Woo NS, Lu J, England R, et al. Hyperexcitability and epilepsy associated with disruption of the mouse neuronal-specific K-Cl cotransporter gene. *Hippocampus* 2002;12:258-268.

16. Fung C, Evans E, Shin D, et al. Hypoxic-ischemic brain injury exacerbates neuronal apoptosis and precipitates spontaneous seizures in glucose transporter isoform 3 heterozygous null mice. *J Neurosci Res* 2010;88:3386-3398.
17. Zhao Y, Fung C, Shin D, et al. Neuronal glucose transporter isoform 3 deficient mice demonstrate features of autism spectrum disorders. *Mol Psychiatry* 2010;15:286-299.
18. Tanaka K, Watase K, Manabe T, et al. Epilepsy and exacerbation of brain injury in mice lacking the glutamate transporter GLT-1. *Science* 1997;276:1699-1702.
19. Payne JA, Stevenson TJ, Donaldson LF. Molecular characterization of a putative K-Cl cotransporter in rat brain. A neuronal-specific isoform. *J Biol Chem* 1996;271:16245-16252.
20. Lu J, Karadsheh M, Delpire E. Developmental regulation of the neuronal-specific isoform of K-Cl cotransporter KCC2 in postnatal rat brains. *J Neurobiol* 1999;39:558-568.
21. Williams JR, Sharp JW, Kumari VG, et al. The neuron-specific K-Cl cotransporter, KCC2. Antibody development and initial characterization of the protein. *J Biol Chem* 1999;274:12656-12664.
22. Gursoy S, Ercal D. Diagnostic Approach to Genetic Causes of Early-Onset Epileptic Encephalopathy. *J Child Neurol* 2016;31:523-532.
23. Cohen R, Basel-Vanagaite L, Goldberg-Stern H, et al. Two siblings with early infantile myoclonic encephalopathy due to mutation in the gene encoding mitochondrial glutamate/H+ symporter SLC25A22. *Eur J Paediatr Neurol* 2014;18:801-805.
24. Poduri A, Heinzen EL, Chitsazzadeh V, et al. SLC25A22 is a novel gene for migrating partial seizures in infancy. *Ann Neurol* 2013;74:873-882.
25. McTague A, Appleton R, Avula S, et al. Migrating partial seizures of infancy: expansion of the electroclinical, radiological and pathological disease spectrum. *Brain* 2013;136:1578-1591.
26. Molinari F, Kaminska A, Fiermonte G, et al. Mutations in the mitochondrial glutamate carrier SLC25A22 in neonatal epileptic encephalopathy with suppression bursts. *Clin Genet* 2009;76:188-194.
27. Molinari F, Raas-Rothschild A, Rio M, et al. Impaired mitochondrial glutamate transport in autosomal recessive neonatal myoclonic epilepsy. *Am J Hum Genet* 2005;76:334-339.
28. Dobson PD, Kell DB. Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat Rev Drug Discov* 2008;7:205-220.
29. Schmidt D, Loscher W. New developments in antiepileptic drug resistance: an integrative view. *Epilepsy Curr* 2009;9:47-52.

Legend to Figure

Figure 1 Our *in silico* strategy. We extracted data relating to SLCs from (1) an integrative analysis of microarray studies on brain tissue from epilepsy surgery, and (2) a comprehensive review of published literature on epilepsy pharmacoresistance, and integrated all the data using 'convergent functional genomics (CFG)'—a validated technique for prioritizing genes involved in complex diseases by collating evidence using a pre-defined scoring system. To identify SLCs which have not been studied in epilepsy pharmacoresistance but could potentially be involved, we employed a computational gene prioritization tool called Endeavour, using SLCs with the highest CFG-scores as the training genes. We validated this computationally prioritized gene list by (1) prioritizing the same candidate genes using a robust independent 'positive control' training set comprising the 10 most consistently downregulated genes from the aforementioned integrative analysis and, then, demonstrating significant similarity with the original list, and (2) prioritizing the same candidate genes using a 'negative control' training set comprising 10 randomly chosen genes from the aforementioned integrative analysis and demonstrating no significant rank order similarity with the original list.

Table 1 Convergent Functional Genomics scoring scheme for SLC transporters

Evidence	Score	Reason
Gene/protein expression data		
Downregulated in ≥2 studies on pharmacoresistant MTLE epileptic foci	1	Maximum weight given to confirmed downregulation in MTLE, which is our phenotype of interest
Downregulated in only 1 study on pharmacoresistant MTLE epileptic foci	0.50	Lower score given to non-replicated evidence
Downregulated in ≥1 study on pharmacoresistant non-MTLE epileptic foci	0.50	Lower score given to evidence in other phenotypes
Evidence of downregulation in brain tissue from animal models of pharmacoresistant epilepsy	0.25	Lowest score for animal models

Table 2 *In silico* gene list of 30 genes.

***In silico* list of 30 genes**
SLC1A2
SLC1A3
SLC17A7
SLC24A3
SLC2A3
SLC12A5
SLC30A3
SLC8A2
SLC47A1
SLC4A8
SLC15A2
SLC8A1
SLC17A1
SLC6A20
SLC16A1
SLC1A6
SLC1A1
SLC1A4
SLC6A1
SLC24A2
SLC12A7
SLC25A22
SLC6A8
SLCO1C1
SLC7A5
SLCO3A1
SLC7A11
SLC25A11
SLC4A3
SLC20A1

Table 3 Sample characteristics. RIN=RNA Integrity Number

Sample	Phenotype	Age	Sex	RIN
D1	Case	41	F	7.4
D2	Case	23	F	6.8
D3	Case	51	M	6
D4	Case	49	F	6.9
D5	Case	50	F	7.8
D6	Case	45	F	6.6
D7	Case	12	M	7
D8	Case	29	F	7.8
D9	Case	33	M	6.8
D10	Case	25	F	7.1
D11	Case	34	M	7
D12	Case	33	M	8.8
D13	Case	33	M	8.6
D14	Case	22	F	7.3
D15	Case	48	M	7.6
D16	Case	39	F	7.4
D17	Case	29	F	7.1
D18	Case	44	M	7.9
D19	Case	40	F	6.6
D20	Case	48	M	8.5
D21	Case	23	M	8.4
D22	Case	63	M	7.9
D23	Case	31	M	8.2
D24	Case	27	F	8.2
N1	Control	81	M	6.5
N2	Control	78	F	6.2
N3	Control	84	F	6.6
N4	Control	91	F	6.7

N5	Control	88	M	6.3
N6	Control	38	M	6.1
N7	Control	50	M	6.2
N8	Control	45	M	6.3
N9	Control	39	M	6.1
N10	Control	40	M	6
N11	Control	61	M	6.2
N12	Control	63	F	6.2
N13	Control	66	M	6.2
N14	Control	22	F	6.3
N15	Control	27	M	6.3
N16	Control	45	M	6.9
N17	Control	44	F	6.7
N18	Control	50	M	6.5
N19	Control	43	M	6.6
N20	Control	46	M	6.8
N21	Control	51	M	6.2
N22	Control	48	M	6
N23	Control	43	M	6.6

Table 4 Microarray results. FDR=false discovery rate

Gene Symbol	Microarray results	
	FDR	Fold-change
SLC24A3	1.65×10^{-4}	3.7
SLC47A1	1.49×10^{-3}	2.6
SLC25A23	3.38×10^{-4}	2.3
SLC8A2	1.64×10^{-5}	2.0
SLC17A7	3.89×10^{-3}	1.8
SLC25A41	1.20×10^{-5}	1.6
SLC26A10	5.28×10^{-5}	1.6
SLC4A3	8.02×10^{-5}	1.6
SLC4A7	2.26×10^{-2}	1.6
SLC12A5	1.24×10^{-2}	1.6
SLC7A1	6.81×10^{-5}	1.6
SLC16A2	3.38×10^{-4}	1.6
SLC25A22	1.88×10^{-4}	1.6
SLC29A4	1.22×10^{-3}	1.5
SLC8A1	1.24×10^{-2}	1.5
SLC35E2	6.80×10^{-4}	1.5
SLC4A8	1.83×10^{-4}	1.5
SLC18A2	2.89×10^{-2}	1.5

Table 5 Functions of the 18 SLC proteins found to be downregulated in the exon microarray.

Data from Gene (http://www.ncbi.nlm.nih.gov/gene) and UniPort

(http://www.uniprot.org/help/uniprotkb).

Gene Symbol	Function
SLC24A3	Calcium, potassium:sodium antiporter activity
SLC47A1	Monovalent cation:hydrogen antiporter activity
SLC25A23	Calcium-dependent mitochondrial solute carrier activity
SLC8A2	Sodium:calcium exchange activity
SLC17A7	L-glutamate transmembrane transporter activity
SLC25A41	Mitochondrial ATP-magnesium:phosphate carrier activity
SLC26A10	Anion exchange activity
SLC4A3	Anion exchange activity
SLC4A7	Sodium bicarbonate cotransporter activity
SLC12A5	Potassium:chloride transporter activity
SLC7A1	Transport of the cationic amino acids (arginine, lysine and ornithine)
SLC16A2	Thyroid hormone transporter
SLC25A22	Transport of glutamate across the inner mitochondrial membrane
SLC29A4	Reuptake of monoamines into presynaptic neurons
SLC8A1	Sodium:calcium exchanger activity
SLC35E2	Transport of nucleotide sugars into endoplasmic reticulum and Golgi bodies
SLC4A8	Sodium- and carbonate-dependent chloride:bicarbonate exchange activity
SLC18A2	Vesicular transport of biogenic amine neurotransmitters