DOI:
10.1038/s41525-023-00370-z

Document Version
Peer reviewed version

SLCO5A1 and synaptic assembly genes contribute to impulsivity in juvenile myoclonic epilepsy

Delnaz Roshandel[1], Eric J. Sanders[1, 2], Amy Shakeshaft[3, 4], Naim Panjwani[1], Fan Lin [1], Amber Collingwood[3], Anna Hall[3], Katherine Keenan[1], Celine Deneubourg[3], Filippo Mirabella[3], Simon Topp[3], Jana Zarubova[5], Rhys H. Thomas[6, 7], Inga Talvik[8], Marte Syvertsen[9], Pasquale Striano[10, 11], Anna B. Smith[3], Kaja K. Selmer[12, 13], Guido Rubboli[14, 15], Alessandro Orsini[16], Ching Ching Ng[17], Rikke S. Møller[14, 18], Kheng Seang Lim[19], Khalid Hamandi[20, 21], David A. Greenberg[22], Joanna Gesche[23], Elena Gardella[14, 18], Choong Yi Fong[24], Christoph P. Beier[23], Danielle M. Andrade[25], Heinz Jungbluth[26, 27] , Mark P. Richardson[3,28], Annalisa Pastore[3], Manolis Fanto[3], Deb K. Pal[3, 4, 28]*, Lisa J. Strug[1, 2, 29, 30]* and the BIOJUME Consortium[†]

[1] Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, Canada

[2] Division of Biostatistics, Dalla Lana School of Public Health, The University of Toronto, Toronto, Canada

[3] Department of Basic & Clinical Neurosciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[4] MRC Centre for Neurodevelopmental Disorders, King's College London, London, United Kingdom

[5] Department of Neurology, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

[6] Newcastle upon Tyne NHS Foundation Trust, Newcastle, United Kingdom

[7] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, United Kingdom

[8] Tallin Children's Hospital, Tallin, Estonia

[9] Department of Neurology, Drammen Hospital, Vestre Viken Health Trust, Oslo, Norway

[10] IRCCS Istituto 'G. Gaslini', Genova, Italy

[11] University of Genova, Genova, Italy

[12] Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway

[13] National Centre for Epilepsy, Oslo University Hospital, Oslo, Norway

[14] Danish Epilepsy Centre, Dianalund, Denmark

[15] University of Copenhagen, København, Denmark

[16] Pediatric Neurology, Azienda Ospedaliero-Universitaria Pisana, Pisa University Hospital, Pisa, Italy

[17] Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

[18] Department of Regional Health Research, University of Southern Denmark, Odense, Denmark

[19] Division of Neurology, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[20] The Welsh Epilepsy Unit, Department of Neurology Cardiff & Vale University Health Board, Cardiff, United Kingdom

[21] Department of Psychological Medicine and Clinical Neuroscience, Cardiff University, Cardiff, United Kingdom

[22] Nationwide Children’s Hospital, Ohio, USA

[23] Odense University Hospital, Odense, Denmark

[24] Division of Paediatric Neurology, Department of Pediatrics, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[25] Adult Epilepsy Genetics Program, Krembil Research Institute, University of Toronto, Toronto, Canada

[26] Randall Centre for Cell and Molecular Biophysics, Muscle Signalling Section, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

[27] Department of Paediatric Neurology, Neuromuscular Service, Evelina's Children Hospital, Guy's & St. Thomas' Hospital NHS Foundation Trust, London, United Kingdom

[28] King's College Hospital, London, United Kingdom

[29] Departments of Statistical Sciences and Computer Science, The University of Toronto, Toronto, Canada

[30] The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada.

[†]A list of authors and their affiliations appears at the end of the paper.

***Correspondence to:**

Professor Deb K Pal
Maurice Wohl Clinical Neurosciences Institute
Institute of Psychiatry, Psychology & Neuroscience,
King's College London
5 Cutcombe Road, London SE5 9RX
deb.pal@kcl.ac.uk
@palneurolab
+44 (0)207 848 5162

Professor Lisa J Strug
Departments of Statistical Sciences and Computer Science
The University of Toronto
Senior Scientist
The Hospital for Sick Children
555 University Avenue**,** Toronto M5G 0A4
Lisa.Strug@utoronto.ca
(416) 813-7654 Ext: 301762

Abstract

Elevated impulsivity is a key component of attention-deficit hyperactivity disorder (ADHD), bipolar disorder and juvenile myoclonic epilepsy (JME). We performed a genome-wide association, colocalization, polygenic risk score, and pathway analysis of impulsivity in juvenile myoclonic epilepsy (n = 381). Results were followed up with functional characterization using a drosophila model. We identified genome-wide associated SNPs at 8q13·3 ($P = 7{\cdot}5 \times 10^{-9}$) and 10p11·21 ($P = 3{\cdot}6 \times 10^{-8}$). The 8q13·3 locus colocalizes with *SLCO5A1* expression quantitative trait loci in cerebral cortex ($P = 9{\cdot}5 \times 10^{-3}$). *SLCO5A1* codes for an organic anion transporter and upregulates synapse assembly/organization genes. Pathway analysis demonstrates 12·7-fold enrichment for presynaptic membrane assembly genes (P = 0.0005) and 14·3-fold enrichment for presynaptic organization genes (P = 0.0005) including *NLGN1* and *PTPRD*. RNAi knockdown of *Oatp30B*, the *Drosophila* polypeptide with the highest homology to *SLCO5A1*, causes over-reactive startling behaviour ($P = 8{\cdot}7 \times 10^{-3}$) and increased seizure-like events ($P = 6{\cdot}8 \times 10^{-7}$). Polygenic risk score for ADHD genetically correlates with impulsivity scores in JME ($P = 1{\cdot}60 \times 10^{-3}$). *SLCO5A1* loss-of-function represents an impulsivity and seizure mechanism. Synaptic assembly genes may inform the aetiology of impulsivity in health and disease.

Keywords

GWAS; eQTL; generalised epilepsy; Barratt Impulsiveness Scale (BIS)

INTRODUCTION

Impulsivity is a heritable behavioural trait leading to actions that are "poorly conceived, prematurely expressed, unduly risky, or inappropriate to the situation and that often result in undesirable consequences".[1] Estimates of heritability for impulsivity from a study of twins were between 33% and 56% at ages 11-13 years, and between 19% and 44% at ages 14-16.[2] Raised impulsivity is a key endophenotype of attention-deficit hyperactivity disorder (ADHD),[3] bipolar disorder[4] and juvenile myoclonic epilepsy.[5-7] ADHD is characterized by inattention, hyperactivity and impulsivity. Individuals with ADHD show more signs of impulsivity (attentional, non-planning and motor) compared to controls.[8] A previous genome-wide association study (GWAS) of impulsive personality traits (UPPS-P Sensation Seeking, Drug Experimentation and UPPS-P Negative Urgency) in 22,861 healthy individuals of European ancestry demonstrated two significant associated loci at 3p12·1 and 22q13·1.[9] Variants at the 3p12·1 locus correlated with predicted *Cell Adhesion Molecule–2* (*CADM2*) expression, in the putamen,[10] and the 22q13·1 locus near *CACNA1I* has been previously implicated in schizophrenia.[11] CADM2 mediates synaptic signalling and is highly expressed in the human cerebral cortex and cerebellum.[12] Given impulsivity is elevated in neuropsychiatric disorders, there may be shared genetic mechanisms across disorders and/or with impulsivity in the general population, however to our knowledge there has been no GWAS of impulsivity in any neuropsychiatric disorder.

Impulsivity is elevated in different epilepsies, but the evidence across multiple dimensions of impulsivity is strongest in juvenile myoclonic epilepsy (JME).[5-7] JME is a common adolescent-onset syndrome characterized by awakening myoclonic, generalized tonic-clonic and absence seizures, often triggered by sleep deprivation. Trait impulsivity in JME is associated with the frequency of both myoclonic and absence seizures,[6] but it is not clear if this indicates a causal relationship or a common mechanism regulating both impulsivity and

seizures, though convergent lines of evidence suggest the involvement of overlapping prefrontal-striatal networks in both JME and impulsivity.[13-20] Finding a shared aetiology would offer new therapeutic approaches for drug-resistant epilepsy.

The overall syndrome of JME has complex inheritance with few replicated susceptibility loci,[21,22] and other loci with less support.[22-24] A major challenge for epilepsies of complex inheritance is to explain the wide variation in phenotypic expression and treatment response between individuals. Forty-percent experience antiseizure medication (ASM) resistance or intolerance.[25] In addition, no current ASM modifies the lifelong disease course of JME and the pharmacological options are sparse, especially for women.[25] Hence novel treatments based on genetic disease mechanisms, such as those emerging for monogenic channelopathy and mTOR pathway epilepsies, are urgently needed.[26,27] Our methodological approach is to carry out genome-wide analysis of endophenotypes in JME such as impulsivity and clinically relevant outcomes such as ASM resistance, a strategy with predicted advantages for reducing heterogeneity, increasing statistical power[28,29] and improving direct clinical translation for precision medicine.

RESULTS

Genome-wide association analysis with BIS-Brief

We investigated the influence of 8,950,360 variants on impulsivity in European ancestry patients with JME (n = 324) and a mega-analysis including all ancestries (n = 372), who self-rated their trait impulsivity using the Barratt Impulsivity Scale, eight-item BIS-Brief version.[30] We conducted a GWAS of BIS-Brief score (Supplementary Figure 2) in the European subset, adjusted for sex, genotyping batch, age at consent, population stratification, and seizure frequency (Supplementary Table 2). We discovered two genome-wide significant

loci, one on chromosome 8 (rs73293634 (G/T)) and one on chromosome 10 (rs75042057 (T/G) (Figure 1, Table 1, Supplementary Figure 3 & 4). Given the distribution of BIS-Brief was slightly right skewed, for sensitivity analysis we tested the SNP associations using an inverse rank normal transformed BIS-Brief phenotype as well. Qualitatively similar results were obtained with rs73293634 and rs75042057 demonstrating association with $p = 3{\cdot}1 \times 10^{-8}$ and $p = 1{\cdot}4 \times 10^{-7}$, respectively (Supplementary Table 3). The distribution of BIS-Brief by rs73293634 and rs75042057 genotypes are provided in the Supplementary Figure 5. In a mega-analysis comprised of all ancestral groups (Supplementary Figure 6), these loci were further supported including by a nearby chromosome 8 SNP (rs146866040, $r^2 = 0.89$) with stronger evidence of association in the combined ancestry mega-analysis as measured by the p-value ($P = 1{\cdot}57 \times 10^{-9}$; Table 1), providing cross-ancestral support for the locus.

rs73293634 falls in an intergenic region near *SLCO5A1*. The phenotypic variation explained (PVE) for rs73293634 was 10·1% in the European analysis. Although a second JME cohort with impulsivity measured is not available for replication, Watanabe et al[31] reported a rs73293634 association with risk taking in the UK Biobank, where they asked the question "Would you describe yourself as someone who takes risks?" (OR (95% CI) = 1.032 (1.001-1.065), p = 0.04, minor allele frequency (MAF) = 0.03, N = 371,049). Association results posted on the same data by the Neale Lab[32] with ~23K fewer participants, provides a similar qualitative conclusion (β (SE) = 0.005 (0.003), p = 0.09, MAF = 0.03, N = 348,549). Two individuals with large structural deletions that include *SLCO5A1* are reported in the Decipher Genomics database with seizures and neurodevelopmental disorder (www.deciphergenomics.org/gene/SLCO5A1/patient-overlap/cnvs).

The significant genome-wide association on chromosome 10 (rs75042057) falls in intron 22 of *PARD3* (NM_001184785·2). The PVE by the SNP is 9·3%, although there are no variants in linkage disequilibrium with this SNP so further interrogation and confirmation of this

locus is required. We note, however, that significant linkage (multipoint max LOD 4·23, alpha 0·34) was previously reported to this locus in French-Canadian families with idiopathic generalized epilepsy (IGE),[33] of which JME is a common subtype. As well, rs75042057 was also associated with risk taking in the UK Biobank (OR (95% CI) = 1.067 (1.029-1.106), p = 4.79E-4, MAF = 0.02, N = 371,049)[31].

Colocalization analysis with gene expression

Since the GWAS-associated variants are not exonic, we next assessed whether the variants impact gene expression, and for which gene in which tissue of origin, by assessing colocalization of the genome-wide significant peaks with expression quantitative trait loci (eQTL) in brain tissues. We used eQTLs from the Genotype-Tissue Expression project (GTEx) v8,[12] PsychENCODE,[34] and human fetal brains[35] and combined them with the GWAS summary statistics from the mega-analysis, for colocalization analysis adjusting for multiple hypothesis testing.[36] Colocalization analysis with eQTLs from GTEx brain and tibial nerve tissues for genes at the locus (chr8:69,650,000-70,000,000, hg38) shows significant colocalization with *SLCO5A1* in the cerebral cortex, and no colocalization with other genes in the region (Figure 2A and Supplementary Figure 7; Simple Sum 2 colocalization P = 9·5 × 10^{-3}). The minor allele for the lead SNP rs73293634 (T) decreases expression in GTEx cerebral cortex (Figure 2C). We found no significant colocalization with eQTLs from PsychENCODE[34] and fetal brains,[35] although nearby variants in the locus in adult brains in PsychENCODE have, in general, a clear influence on *SLCO5A1* expression (Figure 2B). According to BrainSpan,[37,38] *SLCO5A1* is highly expressed prenatally, with expression dropping after birth but remains detectable throughout adulthood (Figure 2D). We did not observe significant colocalization at the chromosome 10 locus with eQTLs from adult brains in GTEx,[12] PsychENCODE[34] or fetal brains.[35]

Functional characterization of *SLCO5A1*

SLCO5A1 is a membrane-bound organic anion transporter with no known substrate[39] (Figure 3). We performed a full protein BLAST (BLASTp) search of the SLCO5A1 polypeptide sequence (NP_112220·2) on *Drosophila melanogaster* to identify the closest matching sequence alignment. While several members of the Oatp family were found to have significant homology, *Oatp30B* was the family member with the highest homology and a 37·66% identity and E-value of 2×10^{-150} (NP_995667·1). SLCO5A1 was the closest human analog of Oatp30B also in a reverse BLASTp. Indeed, BLASTp of Oatp30B polypeptide sequence (Q9VLB3) across all species for conserved domains reveals this gene has conserved major facilitator superfamily (MFS), OATP, and Kazal domains (Figure 3 and Supplementary Figure 8). We therefore used an effective RNAi transgenic line (Supplementary Figure 9A) to assess the effect of pan-neuronal adult knockdown of *Oatp30B/SLCO5A1*. Flies with reduced *Oatp30B* levels displayed a small but significant shortening of their lifespan (Supplementary Figure 9B) and a striking over-reaction to vibration stimuli applied through the automated Drosophila Arousal Tracking (DART) system,[40] which elicit an otherwise modest activity response in two separate control fly genotypes (Figure 4A). Additional analysis of locomotor behaviour clarifies that *Oatp30B* knockdown did not alter the speed of flies or the duration of each activity bout or the interval in between bouts of action (Supplementary Figure 9C-E), indicating a specific defect in excessive response to stimuli. Furthermore, *Oatp30B* knockdown led to a dramatic increase in the frequency of seizure-like events (Figure 4B) when exposed to hyperthermia, a trigger for seizures in *Drosophila*.[41] Recovery to full motility after seizure-like events was also significantly slower in flies with *Oatp30B* knockdown (Figure 4C). These data establish a common causal link between *Oatp30B/SLCO5A1* downregulation, startling behaviour, and susceptibility to seizure-like events.

Gene enrichment analyses

We next sought to assess whether there was additional signal in the GWAS where sub-GWAS significant variants could inform additional contributing genes or pathways and whether there were shared genetic contributions with other psychiatric or epilepsy phenotypes. We selected all variants displaying $P \leq 5 \times 10^{-4}$ and annotated these variants to the transcription start site of the nearest gene resulting in 810 unique genes. Gene enrichment analysis using one-sided hypergeometric tests [42] identified a 12·7-fold enrichment of associated genes from the presynaptic membrane organisation gene set (five out of nine genes; gene ontology (GO):0097090) and a 14·3-fold enrichment of associated genes from the presynaptic membrane assembly gene set (five out of eight genes; GO:0097105). These genes were *PTEN, NLGN1, PTPRD, IL1RAPL1*, and *NLGN4X* (Table 2). The combined PVE for the lead variants annotated to these five genes was 15·6% (25·8% with the addition of rs73293634 from the *SLCO5A1* locus and rs75042057 from the *PARD3* locus).

The permutation tests of presynaptic membrane organisation (GO:0097090) over-enrichment and of presynaptic membrane assembly (GO:0097104) over-enrichment both produced permutation-based p-values of 0.0005.

Investigation of these 810 genes revealed further [43-45] that there was a significant overlap with genes reported in the GWAS catalog that contribute to phenotypes relevant to the predominance of JME seizures on awakening, impulsivity and metabolism: chronotype (66 out of 522 genes overlap, $P = 2{\cdot}92 \times 10^{-12}$), obesity-related traits (77 out of 662 overlap, $P = 2{\cdot}69 \times 10^{-12}$), general risk tolerance (30 out of 238 overlap, $P = 2{\cdot}30 \times 10^{-5}$), and adventurousness (21/134, $P = 3{\cdot}70 \times 10^{-5}$).

Polygenic risk score analysis

Given impulsivity is a major component of ADHD, risk taking, bipolar disorder and epilepsy, we tested and found that a higher ADHD polygenic risk score (PRS) was significantly associated with a higher BIS-Brief score ($p = 1{\cdot}60 \times 10^{-3}$) (Supplementary Figure 10). It should be noted that the lead *SLCO5A1* SNP, rs73293634, was not present in the ADHD GWAS from which the PRS was calculated, but rs146866040 which is in high LD did not show evidence of association itself with ADHD (OR (SE) = 0·9481 (0·0562), p = 0·34)[46]. The rs75042057 SNP on Chr10 was also not present in the ADHD dataset nor was there a proxy with $R^2 > 0{\cdot}6$ available. The risk taking PRS was also nominally associated with a higher BIS-Brief score (p = 0·018). PRSs for bipolar disorder, generalized and focal epilepsy did not reach statistical significance for association with BIS-Brief score at the 5% or Bonferroni corrected level of 1% (P = 0·08, 0·33 and 0·96, respectively) (Supplementary Table 4). Altogether this suggests that the impulsive trait seen in JME is an endophenotype that shares genetic architecture with impulsivity in the general population as well as with individuals diagnosed with ADHD.

DISCUSSION

This is a GWAS of trait impulsivity in a neuropsychiatric disorder and we present convergent evidence for the role of SLCO5A1 in impulsivity and seizure susceptibility through triangulation[47] with GWAS, independent replication, colocalization with gene expression and functional evaluation in *Drosophila.*[48] While several Oatp family members display significant homology to SLCO5A1, the identified Oatp30B was the closest polypeptide in a BLASTp search and SLCO5A1 was the human polypeptide with the highest homology in a reverse BLASTp search. Therefore, whereas our analysis does not rule out some contribution by other closely related Oatp genes, for instance *Oatp26F*, it has identified a major role of *Oatp30B* in regulating startling and seizure-like behaviour in *Drosophila*. In contrast to

human *SLCO5A1*, *Oatp30B* is expressed in the nervous system at constant low to moderate levels throughout fly stages, from development to adulthood. This enables investigation of gene function *in vivo*, in adult flies, although it limits generalization as an *SLCO5A1*-linked disease model.

One GWAS of impulsive traits in the general population identified genome-wide significant association with variants in the *CADM2* gene. *CADM2* encodes a cell adhesion protein from the SynCam Immunoglobulin superfamily of recognition molecules, important for synaptic organisation and specificity; association of variants at the *CACNA1I* locus has been observed in previous studies with schizophrenia.[9] Our GWAS did not show significant association with these previously reported general population associated variants at the *CADM2* and *CACNA1I* loci[9] (P = 0·152, *beta* = -0.52 for rs139528938; and P = 0·32, *beta* = -0·35 for rs4522708; the latter a SNP with r^2=0·87 with the reported SNP, rs199694726, in our BIS-Brief dataset). Genome-wide summary statistics were not available to make additional comparisons. Genome-wide summary statistics were available for the risk-taking phenotype in the UK Biobank[31], in which we observed replication of our lead genome-wide significant *SLCO5A1* variant, rs73293634.

Previous expression studies show that *SLCO5A1* upregulates gene sets implicated in cell adhesion, synapse assembly and organization, principally belonging to the cadherin superfamily[39]; and the enrichment for presynaptic membrane assembly and organisation pathways in our dataset includes genes encoding trans-synaptically interacting proteins that are implicated in a wide range of neuropsychiatric disorders.[49,50] Genetic correlation between ADHD and the BIS-Brief score suggests converging genetic influences across ADHD and epilepsy. Taken together, these results support an important role for specific cell recognition molecules in the organisation of synaptic connections as a mechanism for variation in impulsivity across health and disease.[51]

While prefrontal-striatal inhibitory control networks are implicated in impulse control, specifically between mPFC and nucleus accumbens,[18,20] a role for these limbic networks has only been hinted at in epilepsy. Striato-nigral circuits, preferentially involving the ventral striatum, have long ago been implicated in the *regulation* of generalised seizures in rodent models of generalised epilepsy.[19] Recently, an *initiating* role for cortico-striatal networks in absence seizures with generalized spike-and-wave discharges has been shown in the mouse model of the genetic epilepsy caused by haploinsufficiency of *STXBP1*,[52] specifically by reduced cortical excitatory transmission onto striatal fast spiking interneurons. The startling and the seizure-like phenotype of the *SLCO5A1/Oatp30B* knockdown in *Drosophila* suggests the genetic co-causality of startling and seizures. While it is not possible to define startling as the *Drosophila* equivalent of impulsivity, the two traits share some commonality in the lack of moderation in behaviour. This offers some additional support to the idea that excitatory-inhibitory imbalance in the prefrontal-striatal network may predispose simultaneously to epilepsy and impulsivity substrates and invites new approaches to neuromodulation of generalised seizures.

METHODS

Human Participants

We collected cross-sectional clinical and genetic data from the Biology of Juvenile Myoclonic Epilepsy (BIOJUME) consortium study, which focuses on gathering cases with JME (n = 864).[25] Inclusion criteria have been discussed previously.[6] BIOJUME is a study across 50 sites in 10 countries (Appendix). Furthermore, all participants' medical history was reviewed by a phenotyping committee to validate the diagnosis of JME. Written informed consent was obtained from all participants prior to inclusion in the study and ethical approval from the UK Health Research Authority, South Central Oxford C Research Ethics Committee

(16/SC/0266) and all other collaborating sites was obtained. The SickKids Research Ethics Board of The Hospital for Sick Children (1000033784) also gave ethical approval for this work.

Barratt Impulsivity Scale-Brief (BIS-Brief)

We collected self-rating of trait impulsivity through the BIS-brief. [6,30] The BIS-Brief is a short version of the BIS, one of the most commonly used measures of impulsiveness. The current version of BIS (BIS-11) includes 30-items measuring 3 theoretical subtraits: attentional, motor, and non-planning impulsiveness. BIS-Brief is a unidimensional scale including 8 of the original BIS-11 items generating a total score ranging from 8 to 32. BIS-Brief demonstrated similar indices of construct validity observed for the BIS-11 total score. Using BIS-Brief in large epidemiological studies of psychiatric disorders reduces the burden on respondents without loss of information. [29]

Genotyping quality control

DNA was extracted from blood by each consortium site and sent to The Centre for Applied Genomics at The Hospital for Sick Children in Toronto for genotyping. We genotyped participants' DNA in four batches (n = 702) using the Illumina Omni 2·5 array. SNPs were called using the self-clustering method in Genome Studio. We performed quality control (QC) for each genotyped batch using PLINK v1·90b6·18[53] and custom in-house scripts. Briefly, we removed individuals and variants with call rates below 90%; samples with sex mismatches and/or high heterozygosity; males with heterozygous calls for X chromosome markers (non-pseudoautosomal region); and females with non-missing calls for markers on the Y chromosome. We retained heterozygous calls for mitochondrial markers in both sexes (i.e., due to heteroplasmy). We obtained an unrelated sample by using KING v.2·2·4 software's[54] --unrelated option (that is, those with estimated kinship coefficient less than

0·088). We corrected and updated the ped file with all found relationships, and identified markers with Mendelian errors using PEDSTATS 0·6·12.[55] We flagged 399 markers but did not remove those out of Hardy-Weinberg Equilibrium ($P < 10^{-4}$). We conducted principal component analysis adjusted using the kinship matrix output by KING using PC-AiR in the GENESIS v2·16·0 package.[56]

We performed quality control on each genotyping batch separately, followed by removal of ambiguous A/T, G/C SNPs, chr0 SNPs, indels, monomorphic variants, and duplicate variants; and performed strand alignment using Will Rayner's alignment files (www.well.ox.ac.uk/~wrayner/strand/), then merged all batches. We re-analysed and removed cryptic relationships across batches. The final merged set contained 1,489,917 variants, 695 individuals (241 males, 454 females) including 23 related pairs (for association analyses however, an unrelated set was selected).

Genotype imputation

We used the McCarthy Tools v4·3·0 to prepare the genotype data for imputation (www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.3.0.zip) using TOPMED as the reference panel (r2@1·0·0) on the TOPMED imputation server.[57-59] We converted coordinates from hg37 to hg38 coordinates using strand files (www.well.ox.ac.uk/~wrayner/strand/InfiniumOmni2-5-8v1-4_A1-b38-strand.zip). We merged the pseudoautosomal region (PAR) using PLINK's --merge-x option and checked variants using the HRC checking tool. We removed a total of 282,660 variants due to no matches in the reference (but still analyzed for association with BIS-Brief afterwards), and 1,739,329 variants remained for imputation on the server. We used Eagle v2·4 for phasing, and minimac v4 v1·0·2 for imputation. We kept variants with imputation quality score $r^2 >$

0·4 and MAF > 1% for analysis. A total of 8,950,360 variants remained for association analysis.

Genome-wide association analysis

We included for analysis 381 individuals who passed phenotype QC with complete BIS-Brief rating. From these, four failed genotyping QC, and one individual was removed due to cryptic relatedness (n = 376). The mega-GWAS analysis consisted of a total of 372 unrelated individuals adjusted for sex, genotyping batch, and population stratification (Supplementary Figure 1). The mega-GWAS was used for colocalization analysis of the genome-wide association peak on chromosome 8. We identified 329 patients as European ancestry (defined as within 6 standard deviations from the 1000 Genomes[60] European cluster in a principal component analysis). Among these, five patients had missing information on seizure frequency, so we used 324 individuals for the genome-wide association analysis. The current sample size is sufficient to detect genetic variants that explain 12% of the variance in the BIS-Brief score with 80% power after adjusting for multiple hypothesis testing at the genome-wide significance level. We adjusted for sex, genotyping batch, age at consent, population stratification, and the frequency of myoclonus or absence seizures. The relationship of the frequency of myoclonus or absence seizures, and its relationship with anti-seizure medication and sex with trait impulsivity in JME, has been described previously and was thus adjusted for in current regression analyses.[6,25] All analyses were conducted in the European subset unless noted otherwise. Chromosome X (non-pseudoautosomal region) was analysed with males coded as zero for the reference allele and two for the alternate allele, under the assumption of X-inactivation.[61] Genome-wide significant loci were further investigated for replication of association with risk taking phenotypes in the general population using publicly available summary statistics [31,32].

Gene enrichment analysis

Variants with $P \leq 5 \times 10^{-4}$ were annotated to the gene with the nearest transcription start site using the Ensembl Variant Effect Predictor (v94). [62] This gene set was used as input in a GO enrichment analysis,[63,64] to test for enrichment in annotated pathways. One-sided hypergeometric tests were completed to identify over-representation of pathways. [42] To reduce the risk of false positive results, a permutation procedure [65] was employed by randomly shuffling GWAS p-values 2000 times, each time re-applying the $P \leq 5 \times 10^{-4}$ threshold and calculating the hypergeometric test statistics. For one pathway, the final permutation-based p-value was calculated as the percentage of the 2000 permutations that produced a p-value less than or equal to the p-value calculated from the non-permuted data. A pseudo count was added during this calculation to prevent calculating p-values equal to 0.

Phenome-wide association study (PheWAS) analysis

We queried the top associated genome-wide variant and the top associated variant for each of the nine presynaptic assembly enriched genes across PheWAS databases: GWAS Atlas (https://atlas.ctglab.nl/), Global Biobank Engine,[66] PheWeb,[67] and Gene Atlas.[68]

We used PheWeb portals:

- UK Biobank: https://pheweb.org/MGI-freeze2/
- Oxford Brain Imaging Genetics (BIG) Project: http://big.stats.ox.ac.uk/
- fastGWA: https://yanglab.westlake.edu.cn/resources/ukb_fastgwa/imp/
- https://pheweb.org/UKB-SAIGE/

PRS analysis

Clumping and thresholding were used to calculate ADHD, risk taking, bipolar disorder, generalized epilepsy, and focal epilepsy PRS in individuals of European ancestry using

PLINK v1·9.[53] Five PRS were calculated. A Bonferroni-corrected critical value for significance would therefore be $p < 0{\cdot}05/5=0.01$. The source of summary statistics used, variant filtering, clumping and thresholding details are summarized in Supplementary Table 1. PRS values were generated by weighting selected SNPs after clumping and thresholding by the additive scale effect ($\log_{10}$(OR)/Beta), and then summing over the variants. The PRS values were then centred to the mean. Association of PRSs with BIS-Brief was tested using linear regression with age, sex, and frequency of absence/myoclonic seizure as covariates in the model.

Colocalization analysis

We used the Simple Sum 2[36] and COLOC2[69] colocalization methods as implemented in LocusFocus[70] (v1·4·9) to test for colocalization of the genome-wide peaks with eQTL analyses in brain tissues in GTEx v8,[12] PsychENCODE,[34] and fetal brain.[35] For the genome-wide associated locus on chromosome 8, we performed colocalization analysis using both the mega-GWAS and Europeans-only GWAS. The required significance threshold, after multiple testing of all colocalization datasets analyzed was 0·01.

Domain architecture of SLCO5A1

A BLAST search against the entire Protein Data Base (PDB) identified only one hit with a convincingly high E-value (1e-55) that pointed to the Chain L of the Kazal-like domain containing mice protein (7EEB). The search had a 26% identity and a coverage of 74%. After this hit, the other four identified sequences had E-values > 0·002, clearly distinguishing between significant and non-significant hits. 7EEB is a large complex containing several subunits, among which is *SLCO6C1,* which is the region scoring for *SLCO5A1*.

Phenotypic variance explained

To assess the PVE by a SNP or a group of SNPs, we calculated the partial r^2 as the proportion of the residual sum of squares (RSS) reduced when adding the SNP (or group of SNPs) to the base regression model with all covariates.

siRNA probe design and knockdown of *Oatp30B* in *Drosophila melanogaster*

Drosophila

Flies were maintained and crossed at 18°C. All ageing was done in a controlled environment of 29°C and 60% humidity.

Stocks

ubiGal80^{ts} // *UAS-Oatp30B^{IR}* (GD12775) obtained from the VDRC // *w^{1118}*, *nSybGal4, TubGal4* and *UAS-GFP^{IR}* obtained from the BDSC.

Lifespan

Lifespan analysis was performed as previously reported.[41] All crosses were maintained at 18°C during the developmental stages of the progeny. Newly eclosed adult flies were collected within 5 days at 18°C. Females and males were pooled together and equally distributed within vials.

Motor behaviour assay

Single fly tracking was carried out as previously described.[41] In each of 3 experiments, up to 12 flies per genotype, aged 15 days (adult stage) at 29°C to allow RNAi expression and knock-down, were placed into individual round 6-wells arenas. The protocol used consisted of 6 stimuli events equally split during a period of 2 h and 15 min, the first one starting after 30 min of recording, and the last one 30 min before the end of the protocol. Each stimuli event was composed of 5 vibrations of 200 ms spaced by 500 ms. The x/y position of each

single fly was tracked and analysed using the DART software in order to evaluate the relative speed and activity before, during and after the stimuli event. The speed analysis was used for the “Stimuli Response Trace” and the general activity used to deduce “Active Speed”, “Mean Bout Length” and “Inter-Bout Interval”, using a custom-made modification of the DART software.[40]

Heat-induced seizure assay

Flies aged 15 days at 29°C to allow RNAi expression and knock-down were isolated into new plastic vials without food for 10-20 min before immersion in a 42°C water bath for 120 seconds. Each tube was video recorded during and post immersion and seizures were defined as a period of brief leg twitches, convulsions, and failure to maintain standing posture. Flies were, thereafter, allowed to recover at room temperature and the time to recover from seizure was calculated only for flies that had undergone seizures. All experiments were randomised and double-blinded.

RNA extraction and qPCR

RNA was extracted as previously reported[71] from 15 adult flies of both sexes, aged 15 days at 29°C to allow RNAi expression and knock-down, using TriZol (Thermo-Fischer). cDNA was generated using SuperScript III Reverse Transcriptase (Thermo-Fischer). Quantitative PCR was performed in combination with qPCRBIO SyGreen Blue mix (PCR Biosystems) on Quantstudio 7 from real-time PCR system (Thermo-Fischer). *eIF4a* was used as housekeeping control. The following oligos were used: *Oatp30B* Fw (GAATCCGACCAACCGCCTGA), *Oatp30B* Rv (ATGGATTCCTGCCGCCTGTG), *eIF4a* Fw (CGTGAAGCAGGAGAACTGG), *eIF4a* Rv (CATCTCCTGGGTCAGTTG).

Data Availability

eQTL data are available for download from GTEx (https://gtexportal.org/home), PsychENCODE (http://resource.psychencode.org/), and fetal brains (https://doi.org/10.6084/m9.figshare.6881825). GWAS summary statistics for this study are available for download from our website (https://lab.research.sickkids.ca/strug/softwareandresources/).

Acknowledgements

This work was supported by the Canadian Institutes of Health Research (CIHR) Operating Grant, FRN: 142405 (DKP, LJS) and CIHR Foundation Grant, FRN: 167282 (LJS); UK Medical Research Council, Centre for Neurodevelopmental Disorders MR/N026063/1 (DKP, MPR); UK Medical Research Council, Programme Grant MR/K013998/1, (MPR); PhD stipend from UK Medical Research Council and the Sackler Institute for Translational Neurodevelopment (AS); NIHR Specialist Biomedical Research Centre for Mental Health of South London and Maudsley National Health Service Foundation Trust (DKP, MPR); UK Engineering and Physical Sciences Research Council, Centre for Predictive Modelling in Healthcare (EP/N014391/1 (MPR)); DINOGMI Department of Excellence of MIUR 2018–2022 (legge 232 del 2016 (PS)); Wales BRAIN Unit and Research Delivery Staff funded by Welsh Government through Health and Care Research Wales (KH); Biomarin srl, ENECTA srl, GW Pharmaceuticals, Kolfarma srl. and Eisai (PS); South-Eastern Regional Health Authority, Norway (Project Number 2016129 (KKS)); The Research Council of Norway (Project Number 299266 (MS)); Epilepsy Research UK (RHT, MPR); Health & Care Research Wales (MPR), Wales Gene Park (MPR), Abertawe Bro Morgannwg University NHS R&D (MPR); UCB (GR); Nationwide Children's Hospital (DAG); Odense University Hospital (JG); University of Southern Denmark (17/18517 (CPB)); Grants NC/V001051/1

from the NC3Rs (MF), European Union's Horizon 2020 Research and Innovation Programme (765912 - DRIVE - H2020-MSCA-ITN-2017 (HJ)) and Action Medical Research (GN2446 (HJ, MF)). LJS is a Canada Research Chair and this research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

Author Contributions

LJS and DKP contributed to conception and study design. DR, NP, AS, AC, FL, AH, KK, DKP and LJS contributed to data management and project administration. DMA, CPB, CYF, EG, JG, DAG, CD, FM, KH, KSL, RSM, CCN, AO, KKS, GR, PS, MS, IT, RHT, JZ, MPR, DKP and LJS contributed to acquisition of study data. DR, EJS, NP, AS, CD, FM, ST, HJ, MPR, AP, MF, DKP, and LJS contributed to analysis of data. DR, EJS, NP, AS, MF, LJS, and DKP contributed to drafting the manuscript. Members of the BIOJUME consortium are listed in the appendix.

Competing interests

DA, KKS, RHT, and JZ report honoraria from UCB Pharma (manufacturer of levetiracetam) and RHT reports honoraria from Sanofi (manufacturer of sodium valproate). KH reports honoraria from UCB Pharma, Eisai and GW Pharma. MS reports honoraria from UCB Pharma and Eisai. GR reports honoraria from UCB Pharma (manufacturer of levetiracetam), from EISAI (manufacturer of perampanel), from Angelini Pharma (manufacturer of cenobamate). RHT reports honorarium from Arvelle/Angelini, Bial, Eisai, GW Pharma/Jazz, Zogenix. All other authors report no conflicts of interest.

References

1 Daruna, J. H. & Barnes, P. A. in *The impulsive client: theory, research and treatment* (eds W G McCown, J L Johnson, & M B Shure) 23-37 (American Psychological Association, 1993).

2 Niv, S., Tuvblad, C., Raine, A., Wang, P. & Baker, L. A. Heritability and longitudinal stability of impulsivity in adolescence. *Behavior genetics* **42**, 378-392, doi:10.1007/s10519-011-9518-6 (2012).

3 Dalley, J. W. & Robbins, T. W. Fractionating impulsivity: neuropsychiatric implications. *Nat Rev Neurosci* **18**, 158-171, doi:10.1038/nrn.2017.8 (2017).

4 Ramirez-Martin, A., Ramos-Martin, J., Mayoral-Cleries, F., Moreno-Kustner, B. & Guzman-Parra, J. Impulsivity, decision-making and risk-taking behaviour in bipolar disorder: a systematic review and meta-analysis. *Psychol Med* **50**, 2141-2153, doi:10.1017/S0033291720003086 (2020).

5 Smith, A., Syvertsen, M. & Pal, D. K. Meta-analysis of response inhibition in juvenile myoclonic epilepsy. *Epilepsy Behav* **106**, 107038, doi:10.1016/j.yebeh.2020.107038 (2020).

6 Shakeshaft, A. *et al.* Trait impulsivity in Juvenile Myoclonic Epilepsy. *Ann Clin Transl Neurol*, doi:10.1002/acn3.51255 (2020).

7 Wandschneider, B. *et al.* Risk-taking behavior in juvenile myoclonic epilepsy. *Epilepsia* **54**, 2158-2165, doi:10.1111/epi.12413 (2013).

8 Malloy-Diniz, L., Fuentes, D., Leite, W. B., Correa, H. & Bechara, A. Impulsive behavior in adults with attention deficit/ hyperactivity disorder: characterization of attentional, motor and cognitive impulsiveness. *Journal of the International Neuropsychological Society : JINS* **13**, 693-698, doi:10.1017/s1355617707070889 (2007).

9 Sanchez-Roige, S. *et al.* Genome-Wide Association Studies of Impulsive Personality Traits (BIS-11 and UPPS-P) and Drug Experimentation in up to 22,861 Adult Research Participants Identify Loci in the CACNA1I and CADM2 genes. *J Neurosci* **39**, 2562-2572, doi:10.1523/JNEUROSCI.2662-18.2019 (2019).

10 Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825, doi:10.1038/s41467-018-03621-1 (2018).

11 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

12 GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).

13 O'Muircheartaigh, J. *et al.* Focal structural changes and cognitive dysfunction in juvenile myoclonic epilepsy. *Neurology* **76**, 34-40, doi:76/1/34 [pii] 10.1212/WNL.0b013e318203e93d (2011).

14 Keller, S. S. *et al.* Microstructural and volumetric abnormalities of the putamen in juvenile myoclonic epilepsy. *Epilepsia* **52**, 1715-1724, doi:10.1111/j.1528-1167.2011.03117.x (2011).

15 Landvogt, C., Buchholz, H.-G., Bernedo, V., Schreckenberger, M. & Werhahn, K. J. Alteration of dopamine D2/D3 receptor binding in patients with juvenile myoclonic epilepsy: Alteration of Dopamine D2/D3 Receptor Binding in JME. *Epilepsia* **51**, 1699-1706, doi:10.1111/j.1528-1167.2010.02569.x (2010).

16 Ciumas, C. *et al.* Reduced dopamine transporter binding in patients with juvenile myoclonic epilepsy. *Neurology* **71**, 788-794, doi:10.1212/01.wnl.0000316120.70504.d5 (2008).
17 Dalley, Jeffrey W., Everitt, Barry J. & Robbins, Trevor W. Impulsivity, Compulsivity, and Top-Down Cognitive Control. *Neuron* **69**, 680-694, doi:https://doi.org/10.1016/j.neuron.2011.01.020 (2011).
18 Dalley, J. W. & Roiser, J. P. Dopamine, serotonin and impulsivity. *Neuroscience* **215**, 42-58, doi:10.1016/j.neuroscience.2012.03.065 (2012).
19 Deransart, C., Vercueil, L., Marescaux, C. & Depaulis, A. The role of basal ganglia in the control of generalized absence seizures. *Epilepsy Res* **32**, 213-223, doi:10.1016/s0920-1211(98)00053-9 (1998).
20 Cho, S. S. *et al.* Morphometric correlation of impulsivity in medial prefrontal cortex. *Brain Topogr* **26**, 479-487, doi:10.1007/s10548-012-0270-x (2013).
21 Santos, B. P. D. *et al.* Genetic susceptibility in Juvenile Myoclonic Epilepsy: Systematic review of genetic association studies. *PLoS One* **12**, e0179629, doi:10.1371/journal.pone.0179629 (2017).
22 International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat Commun* **9**, 5269, doi:10.1038/s41467-018-07524-z (2018).
23 Bai, D. *et al.* DNA variants in coding region of EFHC1: SNPs do not associate with juvenile myoclonic epilepsy. *Epilepsia* (2008).
24 Bailey, J. N. *et al.* Variant Intestinal-Cell Kinase in Juvenile Myoclonic Epilepsy. *N Engl J Med* **378**, 1018-1028, doi:10.1056/NEJMoa1700175 (2018).
25 Shakeshaft, A. *et al.* Sex-specific disease modifiers in juvenile myoclonic epilepsy. *Sci Rep* **12**, 2785, doi:10.1038/s41598-022-06324-2 (2022).
26 Li, M. *et al.* Antisense oligonucleotide therapy reduces seizures and extends life span in an SCN2A gain-of-function epilepsy model. *J Clin Invest* **131**, doi:10.1172/JCI152079 (2021).
27 Karalis, V. & Bateup, H. S. Current Approaches and Future Directions for the Treatment of mTORopathies. *Dev Neurosci* **43**, 143-158, doi:10.1159/000515672 (2021).
28 Hall, M.-H. & Smoller, J. W. A New Role for Endophenotypes in the GWAS Era: Functional Characterization of Risk Variants. *Harv Rev Psychiatry* **18**, 67-74, doi:10.3109/10673220903523532 (2010).
29 Manchia, M. *et al.* The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases. *PLoS One* **8**, e76295, doi:10.1371/journal.pone.0076295 (2013).
30 Steinberg, L., Sharp, C., Stanford, M. S. & Tharp, A. T. New tricks for an old measure: the development of the Barratt Impulsiveness Scale-Brief (BIS-Brief). *Psychol Assess* **25**, 216-226, doi:10.1037/a0030550 (2013).
31 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).
32 *Neale's Lab UK Biobank GWAS Results Round 2 (Imputed v3 - File Manifest Release 20180731)*, <http://www.nealelab.is/uk-biobank> (
33 Kinirons, P. *et al.* A novel locus for idiopathic generalized epilepsy in French-Canadian families maps to 10p11. *Am J Med Genet A* **146A**, 578-584, doi:10.1002/ajmg.a.32139 (2008).
34 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).

35 O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol* **19**, 194, doi:10.1186/s13059-018-1567-1 (2018).
36 Wang, F., Panjwani, N., Wang, C., Sun, L. & Strug, L. J. A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus. *Am J Hum Genet* **109**, 253-269, doi:10.1016/j.ajhg.2021.12.012 (2022).
37 *BrainSpan Atlas of the Developing Human Brain [Internet]*, <http://brainspan.org> (
38 Sunkin, S. M. *et al.* Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* **41**, D996-D1008, doi:10.1093/nar/gks1042 (2013).
39 Sebastian, K. *et al.* Characterization of SLCO5A1/OATP5A1, a solute carrier transport protein with non-classical function. *PLoS One* **8**, e83257, doi:10.1371/journal.pone.0083257 (2013).
40 Faville, R., Kottler, B., Goodhill, G. J., Shaw, P. J. & van Swinderen, B. How deeply does your mutant sleep? Probing arousal to better understand sleep defects in Drosophila. *Sci Rep* **5**, 8454, doi:10.1038/srep08454 (2015).
41 Mazaud, D. *et al.* Transcriptional Regulation of the Glutamate/GABA/Glutamine Cycle in Adult Glia Controls Motor Activity and Seizures in Drosophila. *J Neurosci* **39**, 5269-5283, doi:10.1523/JNEUROSCI.1833-18.2019 (2019).
42 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258, doi:10.1093/bioinformatics/btl567 (2007).
43 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
44 Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770, doi:10.1038/nature07107 (2008).
45 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).
46 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).
47 Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int J Epidemiol* **45**, 1866-1886, doi:10.1093/ije/dyw314 (2016).
48 Parker, L., Howlett, I. C., Rusan, Z. M. & Tanouye, M. A. Seizure and epilepsy: studies of seizure disorders in Drosophila. *Int Rev Neurobiol* **99**, 1-21, doi:10.1016/B978-0-12-387003-2.00001-X (2011).
49 Uhl, G. R. & Martinez, M. J. PTPRD: neurobiology, genetics, and initial pharmacology of a pleiotropic contributor to brain phenotypes. *Ann N Y Acad Sci* **1451**, 112-129, doi:10.1111/nyas.14002 (2019).
50 Hu, Z., Xiao, X., Zhang, Z. & Li, M. Genetic insights and neurobiological implications from NRXN1 in neuropsychiatric disorders. *Mol Psychiatry* **24**, 1400-1414, doi:10.1038/s41380-019-0438-9 (2019).
51 Sanes, J. R. & Zipursky, S. L. Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits. *Cell* **181**, 536-556, doi:10.1016/j.cell.2020.04.008 (2020).
52 Miyamoto, H. *et al.* Impaired cortico-striatal excitatory transmission triggers epilepsy. *Nat Commun* **10**, 1917, doi:10.1038/s41467-019-09954-9 (2019).
53 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

54 Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873, doi:10.1093/bioinformatics/btq559 (2010).
55 Wigginton, J. E. & Abecasis, G. R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21**, 3445-3447, doi:10.1093/bioinformatics/bti529 (2005).
56 Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol* **39**, 276-293, doi:10.1002/gepi.21896 (2015).
57 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
58 Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-784, doi:10.1093/bioinformatics/btu704 (2015).
59 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299, doi:10.1038/s41586-021-03205-y (2021).
60 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
61 Chen, B., Craiu, R. V., Strug, L. J. & Sun, L. The X factor: A robust and powerful approach to X-chromosome-inclusive whole-genome association studies. *Genet Epidemiol* **45**, 694-709, doi:10.1002/gepi.22422 (2021).
62 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
63 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
64 Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334, doi:10.1093/nar/gkaa1113 (2021).
65 Backes, C. *et al.* Systematic permutation testing in GWAS pathway analyses: identification of genetic networks in dilated cardiomyopathy and ulcerative colitis. *BMC genomics* **15**, 622, doi:10.1186/1471-2164-15-622 (2014).
66 McInnes, G. *et al.* Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics* **35**, 2495-2497, doi:10.1093/bioinformatics/bty999 (2019).
67 Gagliano Taliun, S. A. *et al.* Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat Genet* **52**, 550-552, doi:10.1038/s41588-020-0622-5 (2020).
68 Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat Genet* **50**, 1593-1599, doi:10.1038/s41588-018-0248-z (2018).
69 Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* **102**, 1169-1184, doi:10.1016/j.ajhg.2018.04.011 (2018).
70 Panjwani, N. *et al.* LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. *PLoS Comput Biol* **16**, e1008336, doi:10.1371/journal.pcbi.1008336 (2020).
71 Napoletano, F. *et al.* Polyglutamine Atrophin provokes neurodegeneration in Drosophila by repressing fat. *EMBO J* **30**, 945-958, doi:10.1038/emboj.2011.1 (2011).
72 Pao, S. S., Paulsen, I. T. & Saier, M. H., Jr. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1-34, doi:10.1128/MMBR.62.1.1-34.1998 (1998).
73 Walmsley, A. R., Barrett, M. P., Bringaud, F. & Gould, G. W. Sugar transporters from bacteria, parasites and mammals: structure-activity relationships. *Trends Biochem Sci* **23**, 476-481, doi:10.1016/s0968-0004(98)01326-7 (1998).

74 Madej, M. G., Dang, S., Yan, N. & Kaback, H. R. Evolutionary mix-and-match with MFS transporters. *Proc Natl Acad Sci U S A* **110**, 5870-5874, doi:10.1073/pnas.1303538110 (2013).

FIGURE LEGENDS

Figure 1: Manhattan plot showing GWAS with BIS-Brief score

Linear regression was used to test association of each SNP with BIS-Brief. Sex, genotyping batch, age at consent, first 3 PCs, and the frequency of myoclonus or absence seizures were included as covariates in the model. We found two significant genome-wide associations on chromosome 8 (rs73293634 (G/T)) and 10 (rs75042057 (T/G)) in the analysis of 324 European individuals with JME. Variants below $-\log_{10}P < 1$ were omitted in the plot.

Figure 2: LocusFocus[70] plot for the GWAS with BIS-Brief in JME (circles) and eQTLs in GTEx[12] brain and tibial nerve tissues for the *SLCO5A1* gene (lines)

The Simple Sum 2[36] and COLOC2[69] colocalization methods implemented in LocusFocus (v1·4·9)[70] were used to test for colocalization of the BIS-Brief genome-wide peaks with eQTL analyses brain tissues from GTEx v8,[12] PsychENCODE,[34] and fetal brain.[35]. (**A**) Colocalization figure from LocusFocus for the *SLCO5A1* gene. Lines depict the minimum *P*-value trace in a sliding window for *SLCO5A1* eQTLs from GTEx, one line per tissue. Circles depict the GWAS with BIS-Brief, with the lead SNP in purple and pairwise LD with the lead SNP marked as shown in the legend, calculated using the 1000 Genomes Project[60] European subset. Significant colocalization is observed for *SLCO5A1* eQTLs in GTEx v8 for the cerebral cortex after increasing sample size in a mega-GWAS (n=367, $-\log_{10}$ *Simple Sum 2*[36] $P = 9{\cdot}5 \times 10^{-3}$). Colocalization analysis with only the Europeans is provided in Supplementary Figure 7. Colocalization was also tested for all other nearby genes shown in the figure, but no other genes' eQTLs colocalized with BIS-Brief GWAS (not shown). (**B**) Colocalization analysis with PsychENCODE eQTLs in the dorsolateral prefrontal cortex (DLPFC) (n = 1,866),[34] and eQTLs derived from second trimester fetal brains (n = 120),[35] with GTEx's brain cortex eQTL as in A provided for reference. Colocalization analysis results suggest no colocalization with either PsychENCODE (*Simple Sum 2* P = 0·985) or fetal brain eQTLs (does not pass first stage test in Simple Sum 2 for having significant eQTLs in the region). (**C**) Violin plot for the eQTL effect of rs73293634 SNP on *SLCO5A1* expression in the cerebral cortex from GTEx v8. (**D**) Expression change of *SLCO5A1* from brains in various developmental stages from BrainSpan.[37,38] pcw, post conception weeks; preadolescence, 2-12 years old (inclusive); adolescence, 13-19 years old; adult, ≥ 20 years old (oldest samples are 40 years old). The center lines represent the 50th percentile (median) and the bounds of the boxes are the 75th and 25th percentiles (interquartile range) with the whiskers being the largest value within 1.5 times the interquartile range above the 75th percentile and smallest values within 1.5 times the interquartile range below the 25th percentile.

Figure 3: Domain architecture of human SLCO5A1

(**A**) Schematic representation of the protein with the indication of recognised domains. A SMART analysis to identify structural domains confirmed the presence of two modules, Major Facilitator Superfamily (MFS) and a Kazal domain, interspaced with potentially unstructured sequences. The MFS transporters are membrane proteins capable of transporting small solutes in response to chemiosmotic ion gradients.[72,73] They are represented in many

organisms from *Archaea* to *Homo sapiens*. MFS proteins target a wide range of substrates, including ions, carbohydrates, lipids, amino acids and peptides, nucleosides and other small molecules and transport them in both directions across the membrane.[74] The Kazal domain is an evolutionary conserved module usually acting as a serine-protease inhibitor. (**B**) Predicted model of the monomeric form of SLCO5A1 from amino acids 115-766, built using the SwissModel homology server (https://swissmodel.expasy.org) and utilising the template structure pdb:7eeb. Red: alpha helices; Yellow: Beta strands; Green: Loops.

Figure 4: Startling reaction to trains of vibrations, increased seizure prevalence and increased post-seizure recovery time in flies with *Oatp30B* knock down.

(A) Startling reaction to trains of vibrations. The *UAS-Oatp30B^{IR}* (GD12775) transgenic or the control *UAS-GFP^{IR}* were driven with *nSyb-Gal4* and *Ubi-Gal80ts*. The *w^{1118}* strain is a control for the genetic background in absence of transgenes. Mean +/- SEM ** $P < 0{\cdot}01$, One Way ANOVA, Tukey's post-hoc test. Units are the vibration events experienced 6 times for each fly, n = 174-210. (**B) Increased seizure prevalence.** The *UAS-Oatp30B^{IR}* (GD12775) transgenic or the control *UAS-GFP^{IR}* were driven with *nSyb-Gal4* and *Ubi-Gal80ts*. Percent +/- SE **** $P < 0{\cdot}0001$, Log-rank (Mantel-Cox) test, χ^2 24·68 for 1 df, n = 34-36. (**C**) **Increased post-seizure recovery time.** The *UAS-Oatp30B^{IR}* (GD12775) transgenic or the control *UAS-GFP^{IR}* were driven with *nSyb-Gal4* and *Ubi-Gal80ts*. Mean +/- SEM * $P < 0{\cdot}05$, Mann Whitney non- parametric test, two tails, n = 10-26. Only flies that displayed a seizure within 120 s as in Fig. 4B have been included in the analysis.

BIOJUME Consortium

Delnaz Roshandel[1], Eric J. Sanders[1, 2], Naim Panjwani[1], Fan Lin [1], Lisa J. Strug[1, 2, 29, 30], Danielle M. Andrade[25], Jana Zarubova[5], Zuzana Šobíšková[5], Cechovaz Pracoviste[5], Michaela Kajsova[5], Guido Rubboli[14, 15], Rikke S. Møller[14], Elena Gardella[14], Christoph P. Beier[23], Joanna Gesche[23], Maria Miranda[23], Inga Talvik[8], Pasquale Striano[10,] Alessandro Orsini[10], Choong Yi Fong[24], Ching Ching Ng[17], Kheng Seang Lim[19], Kaja K. Selmer[12, 13], Marte Syvertsen[9], Pronab Bala[31], Amy Kitching[31], Kate Irwin[32], Lorna Walding[32], Lynsey Adams[32], Uma Jegathasan[33], Rachel Swingler[33], Rachel Wane[33], Julia Aram[34], Nikil Sudarsan[34], Dee Mullan[34], Rebecca Ramsay[34], Vivien Richmond[34], Mark Sargent[34], Paul Frattaroli[34], Matthew Taylor[35], Marie Home[35], Sal Uka[35], Susan Kilroy[35], Tonicha Nortcliffe[35], Halima Salim[35], Kelly Holroyd[35], Khalid Hamandi[20, 21], Alison McQueen[20], Dympna Mcaleer[20], Dina Jayachandran[36], Dawn Egginton[36], Bridget MacDonald[37], Michael Chang[37], David Deekollu[38], Alok Gaurav[38], Caroline Hamilton[38], Jaya Natarajan[38], Inyan Takon[39], Janet Cotta[39], Nick Moran[40], Jeremy Bland[40], Rosemary Belderbos[41], Heather Collier[41], Joanne Henry[41], Matthew Milner[41], Sam White[41], Michalis Koutroumanidis[42], William Stern[42], Mark P. Richardson[3,28], Jennifer Quirk[28], Javier Peña Ceballos[28], Anastasia Papathanasiou[28], Ioannis Stavropoulos[28], Dora Lozsadi[43], Andrew Swain[43], Charlotte Quamina[43], Jennifer Crooks[43], Tahir Majeed[44], Sonia Raj[44], Shakeelah Patel[44], Michael Young[44], Melissa Maguire[45], Munni Ray[45], Caroline Peacey[45], Linetty Makawa[45], Asyah Chhibda[45], Eve Sacre[45], Shanaz Begum[45], Martin O' Malley[45], Lap Yeung[46], Claire Holliday[46], Louise Woodhead[46], Karen Rhodes[46], Rhys H. Thomas[6, 7], Shan Ellawela[6], Joanne Glenton[6], Verity Calder[6], John Davis[6], Paul McAlinden[6], Sarah Francis[6], Lisa Robson[6], Karen Lanyon[47], Graham Mackay[47], Elma Stephen[47], Coleen Thow[47], Margaret Connon[47], Martin Kirkpatrick[48], Susan MacFarlane[48], Anne Macleod[48], Debbie Rice[48], Siva Kumar[49], Carolyn Campbell[49], Vicky Collins[49], William Whitehouse[50], Christina Giavasi[50], Boyanka Petrova[50], Thomas Brown[50], Catie Picton[50], Michael O'Donoghue[50], Charlotte West[50], Helen Navarra[50] Seán J. Slaght[51], Catherine Edwards[51], Andrew Gribbin[51], Liz Nelson[51], Stephen Warriner[51], Heather Angus-Leppan[52], Loveth Ehiorobo[52], Bintou Camara[52], Tinashe Samakomva[52], Rajiv Mohanraj[53], Vicky Parker[53], Rajesh Pandey[54], Lisa Charles[54], Catherine Cotter[54], Archana Desurkar[55], Alison Hyde[55], Rachel Harrison[55], Markus Reuber[56], Rosie Clegg[56], Jo Sidebottom[56], Mayeth Recto[56], Patrick Easton[56], Charlotte Waite[56], Alice Howell[56], Jacqueline Smith[56], Rosie Clegg[56], Shyam Mariguddi[57], Zena Haslam[57], Elizabeth Galizia[58], Hannah Cock, Mark Mencias[58], Samantha Truscott[58], Deirdre Daly[58], Hilda Mhandu[58], Nooria Said[58], Mark Rees[59], Seo-Kyung Chung[59], Owen Pickrell[59], Beata Fonferko-Shadrach[59], Mark Baker[59], Fraser Scott[60], Naveed Ghaus[60], Gail Castle[60], Jacqui Bartholomew[60], Ann Needle[60], Julie Ball[60], Andrea Clough[60], Shashikiran Sastry[61], Charlotte Busby[61], Amit Agrawal[62], Debbie Dickerson[62], Almu Duran[62], Muhammad Khan[63], Laura Thrasyvoulou[63], Eve Irvine[63], Sarah Tittensor[63], Jacqueline Daglish[63], Sumant Kumar[64], Claire Backhouse[64], Claire Mewies[64], Julia Aram[65], Nikil Sudarsan[65], Dee Mullan[65], Rebecca Ramsay[65], Vivien Richmond[65], Denise Skinner[65], Mark Sargent[65], Rahul Bharat[66], Sarah-Jane Sharman[66], Arun Saraswatula[67], Helen Cockerill[67], David A. Greenberg[22]

[31] Airedale NHS Foundation Trust, Keighley, UK

[32] Ashford and St. Peter's Hospitals NHS Foundation Trust, Chertsey, UK

[33] Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

[34] Brighton and Sussex University Hospitals NHS Trust, Brighton, UK

[35] Calderdale and Huddersfield Foundation Trust, Huddersfield, UK

[36] County Durham and Darlington NHS Foundation Trust, Darlington, UK

[37] Croydon Health Services NHS Trust, Croydon, UK

[38] Cwm Taf Morgannwg University Health Board, Mountain Ash, UK

[39] East and North Hertfordshire NHS Trust, Stevenage, UK

[40] East Kent Hospitals University NHS Foundation Trust, Canterbury, UK

[41] East Lancashire Hospitals NHS Trust, Burnley, UK

[42] Guy's and St Thomas' NHS Foundation Trust, London, UK

[43] Kingston Hospital NHS Foundation Trust, Kingston upon Thames, UK

[44] Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK

[45] Leeds Teaching Hospitals NHS Trust, Leeds, UK

[46] Manchester University NHS Foundation Trust, Manchester, UK

[47] NHS Grampian, Aberdeen, UK

[48] NHS Tayside, Dundee, UK

[49] North Tees and Hartlepool NHS Foundation Trust, Stockton-on-Tees, UK

[50] Nottingham University Hospitals NHS Trust, Nottingham, UK

[51] Portsmouth Hospitals NHS Trust, Portsmouth, UK

[52] Royal Free London NHS Foundation Trust, London, UK

[53] Salford Royal NHS Foundation Trust, Salford, UK

[54] Sandwell & West Birmingham Hospitals NHS Trust, Birmingham, UK

[55] Sheffield Children's NHS Foundation Trust, Sheffield, UK

[56] Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

[57] Southport and Ormskirk Hospital NHS Trust, Southport, UK

[58] St George's University Hospitals NHS Foundation Trust, London, UK

[59] Swansea University Medical School and Swansea Bay University Health board, Swansea, UK

[60] The Mid Yorkshire Hospitals NHS Trust, Wakefield, UK

[61] The Royal Wolverhampton NHS Trust, Wolverhampton, UK

[62] The Walton Centre NHS Foundation Trust, Liverpool, UK

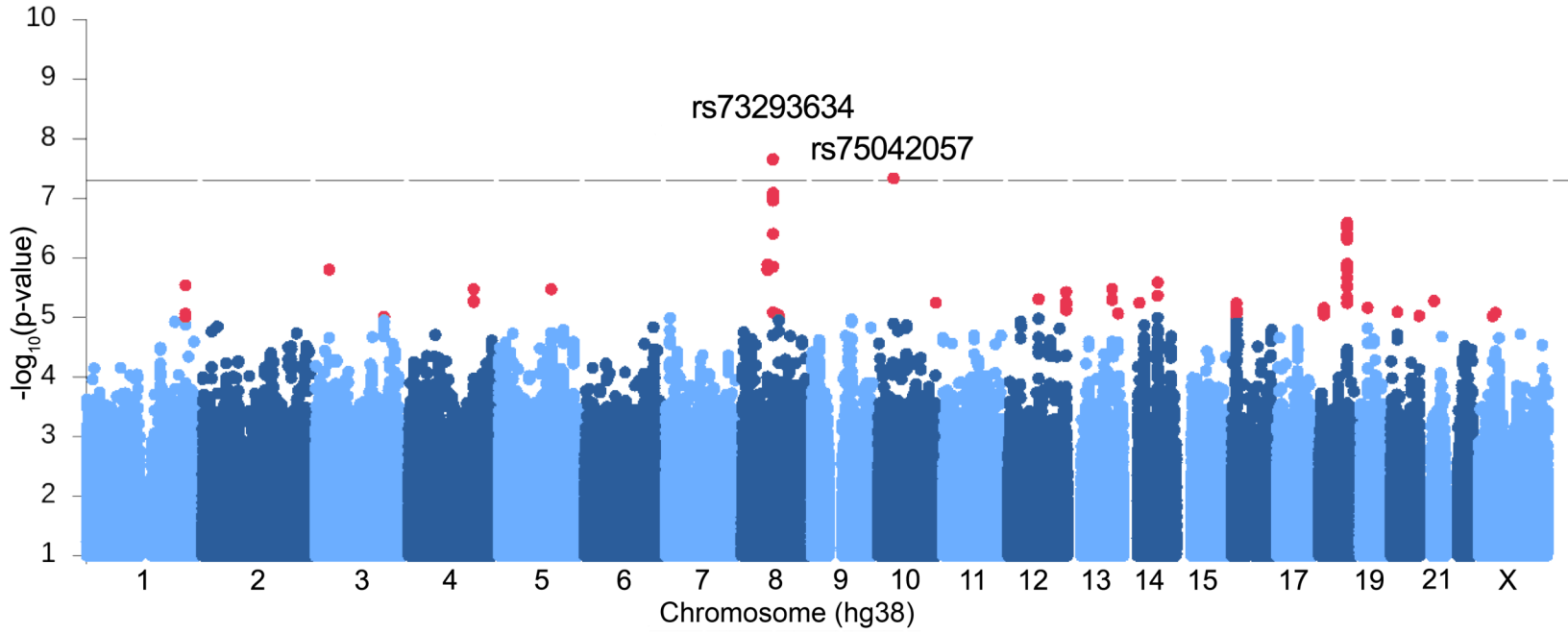
[63] University Hospitals Birmingham NHS Foundation Trust, Birmingham, London

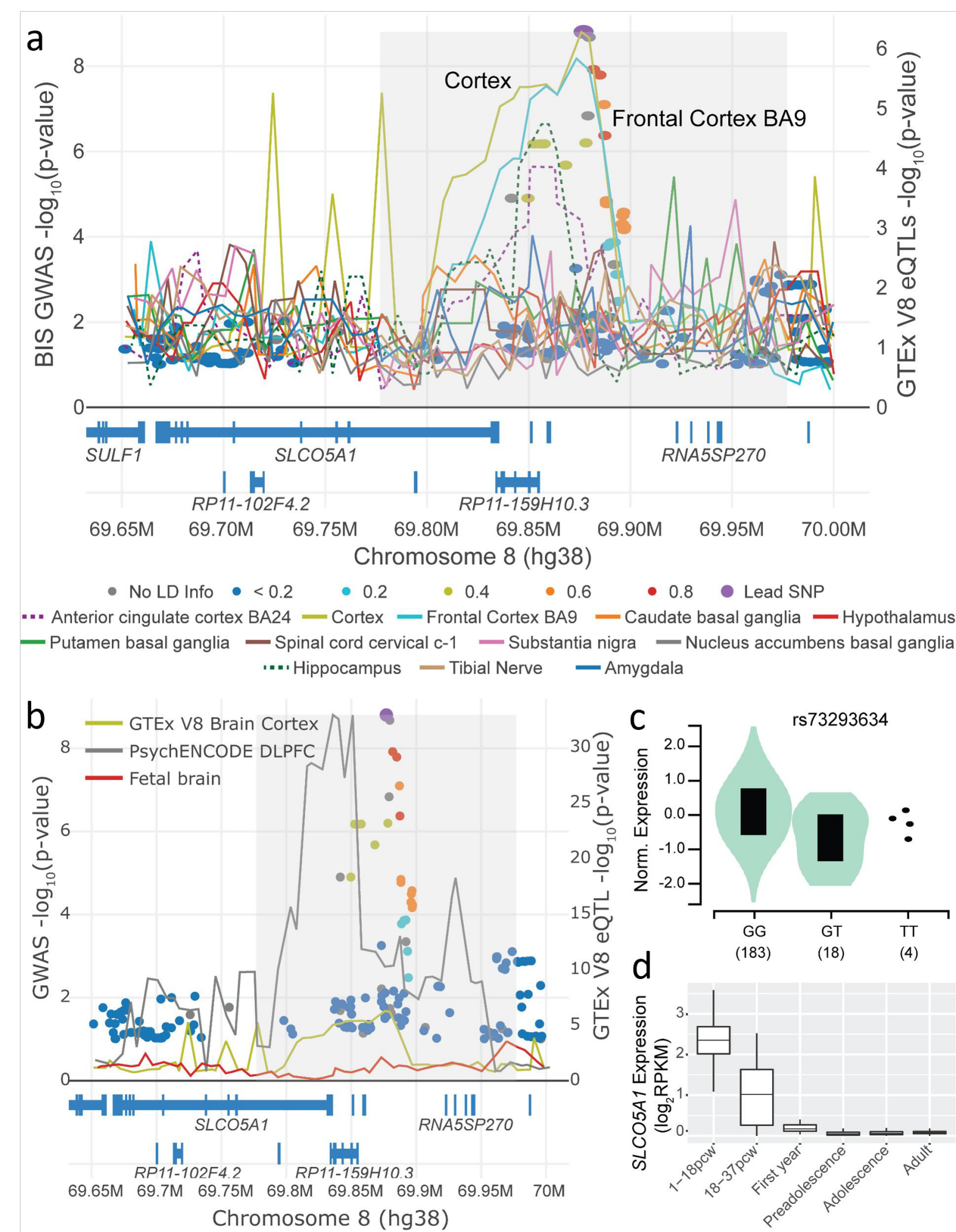
[64] University Hospitals of Derby and Burton NHS Foundation Trust, Derby, UK

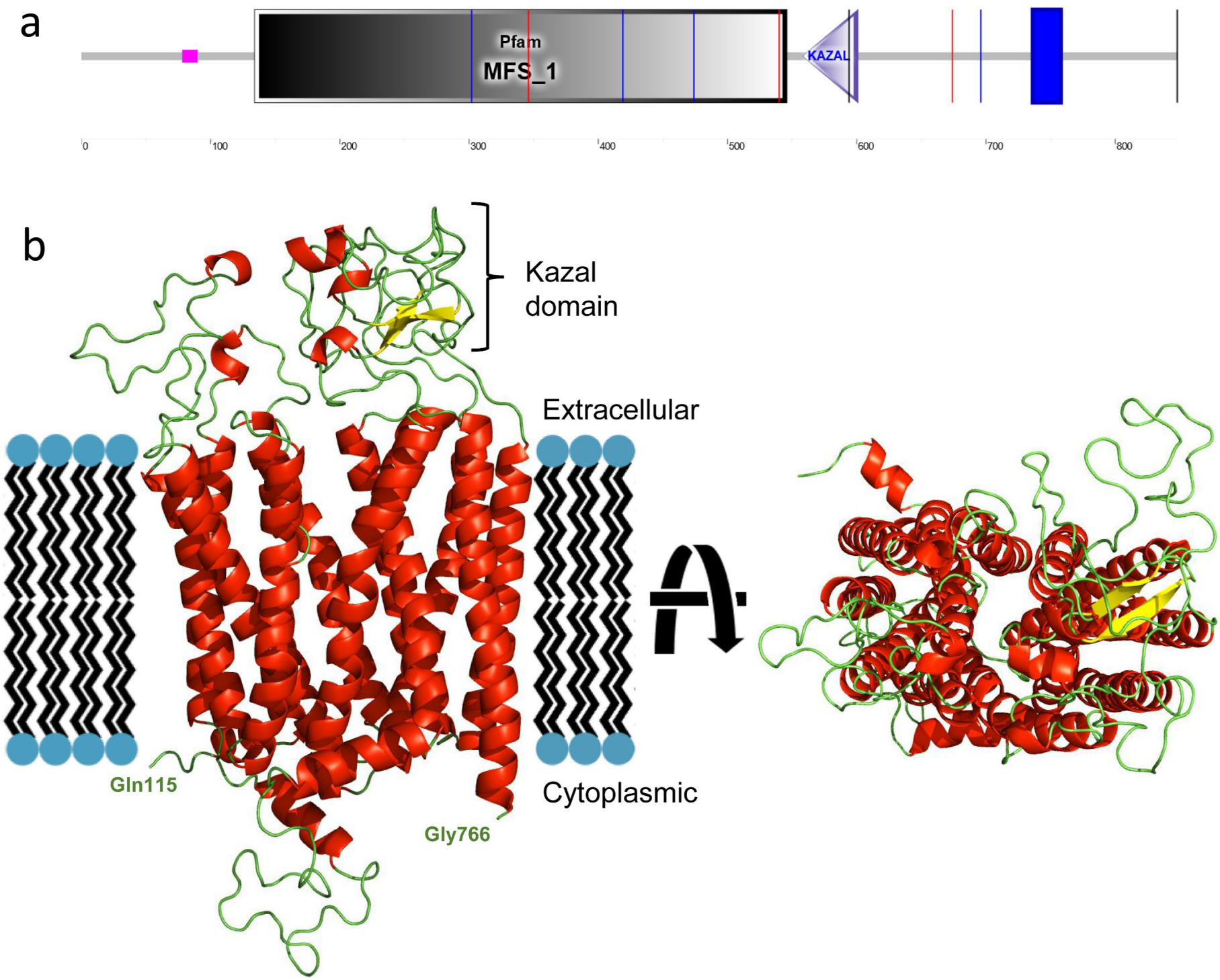
[65] University Hospitals Sussex Hospitals NHS Foundation Trust, Brighton, UK

[66] University Hospitals Plymouth NHS Trust, Plymouth, UK

[67] West Suffolk NHS Foundation Trust, Bury Saint Edmunds, UK


rs73293634
rs75042057
-log10(p-value)
10
9
8
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
X
Chromosome (hg38)


a
Cortex
Frontal Cortex BA9
BIS GWAS -log10(p-value)
GTEx V8 eQTLs -log10(p-value)
SULF1
SLCO5A1
RNA5SP270
RP11-102F4.2
RP11-159H10.3
69.65M
69.70M
69.75M
69.80M
69.85M
69.90M
69.95M
70.00M
Chromosome 8 (hg38)
No LD Info
< 0.2
0.2
0.4
0.6
0.8
Lead SNP
Anterior cingulate cortex BA24
Cortex
Frontal Cortex BA9
Caudate basal ganglia
Hypothalamus
Putamen basal ganglia
Spinal cord cervical c-1
Substantia nigra
Nucleus accumbens basal ganglia
Hippocampus
Tibial Nerve
Amygdala
b
GTEx V8 Brain Cortex
PsychENCODE DLPFC
Fetal brain
GWAS -log10(p-value)
GTEx V8 eQTL -log10(p-value)
SLCO5A1
RNA5SP270
RP11-102F4.2
RP11-159H10.3
69.65M
69.7M
69.75M
69.8M
69.85M
69.9M
69.95M
70M
Chromosome 8 (hg38)
c
rs73293634
Norm. Expression
GG
(183)
GT
(18)
TT
(4)
d
SLCO5A1 Expression
(log2 RPKM)
1–18pcw.
18–37pcw.
First year
Preadolescence
Adolescence
Adult


a
Pfam
MFS_1
KAZAL
0
100
200
300
400
500
600
700
800
b
Kazal
domain
Extracellular
Cytoplasmic
Gln115
Gly766

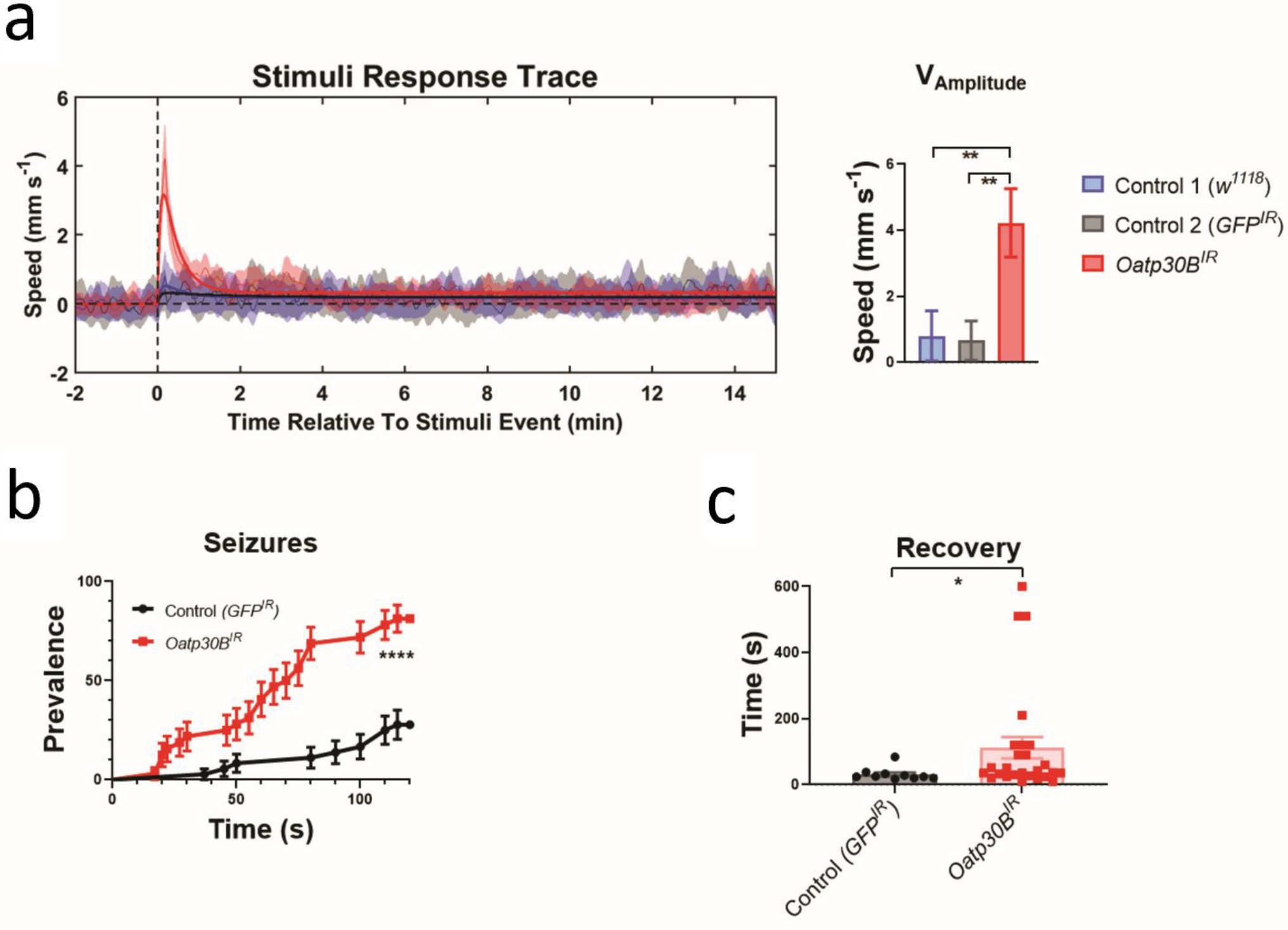

a
Stimuli Response Trace
Speed (mm s-1)
Time Relative To Stimuli Event (min)
VAmplitude
Speed (mm s-1)
**
**
Control 1 (w1118)
Control 2 (GFPIR)
Oatp30BIR
b
Seizures
Prevalence
Control (GFPIR)
Oatp30BIR
****
Time (s)
c
Recovery
Time (s)
*
Control (GFPIR)
Oatp30BIR

Table 1: Summary of genome-wide associated variants for the GWAS of BIS scores in JME (n=324)

		European GWAS (n=324)				Mega-GWAS (n=372)			
Variant ID (hg38)	**Imputation r^2**	**MAF**	**Beta**	**SE**	**P-value**	**MAF**	**Beta**	**SE**	**P-value**
chr8:69,884,968* rs73293634 (G/T)	0·961	0.036 (T)	5·42	0·91	$7{\cdot}5 \times 10^{-9}$	0·041	4·55	0·79	$1{\cdot}61 \times 10^{-8}$
chr8:69,876,965 rs146866040 (A/G)	0·979	0.032 (G)	5·38	0·94	$2{\cdot}5 \times 10^{-8}$	0·031	5·60	0·90	$1{\cdot}57 \times 10^{-9}$
chr10:34,202,650 rs75042057 (T/G)	0·878	0.019 (G)	7·51	1·33	$3{\cdot}6 \times 10^{-8}$	0·022	6·60	1·19	$4{\cdot}99 \times 10^{-8}$

Linear regression was used to test association of each SNP with BIS-Brief. Sex, genotyping batch, age at consent, first 3 PCs, and the frequency of myoclonus or absence seizures were included as covariates in the model in the European analysis. Sex, genotyping batch, and population stratification were included as covariates in the mega-GWAS.
All observed sample allele frequencies are comparable to those seen in the European 1000 Genomes (phase 3) [60].
*The lead SNP for the mega-GWAS was rs146866040. The LD between them is r^2=0·89 or D'=1·0.

Table 2: List of top variants annotated to the five presynaptic assembly genes enriched in the European GWAS of BIS in JME (n=324)

Gene	Location	Size	rsid	Beta	P-value	PVE
PTPRD	chr9:8,314,246-10,613,002	2,298,757	rs1781264	1·827	1·19E-04	0·042
NLGN1	chr3:173,398,448-174,286,644	888,197	rs73177088	6·191	9·95E-04	0·044
NLGN4X	chrX:5,890,042-6,228,867	338,826	rs146813567	-2·898	3·06E-04	0·039
IL1RAPL1	chrX:28,587,446-29,956,718	1,369,273	rs5943492	1·039	8·73E-04	0·043
PTEN	chr10:87,862,563-87,971,930	109,368	rs112050451	5·158	1·27E-03	0·041

Variants with $P \leq 5 \times 10^{-4}$ were annotated to the gene with the nearest transcription start site using the Ensembl Variant Effect Predictor (v94). [62] This gene set was used as input in a GO enrichment analysis,[63,64] to test for enrichment in annotated pathways. One-sided hypergeometric tests were completed to identify over-representation of pathways. [42] To reduce the risk of false positive results, a permutation procedure [65] was employed by randomly shuffling GWAS p-values 2000 times, each time re-applying the $P \leq 5 \times 10^{-4}$ threshold and calculating the hypergeometric test statistics. For each pathway, the final permutation-based p-value was calculated as the percentage of the 2000 permutations that produced a p-value less than or equal to the p-value calculated from the non-permuted data. A pseudo count was added during this calculation to prevent calculating p-values equal to 0.

PTPRD, Protein Tyrosine Phosphatase Receptor Type D; *NLGN1*, Neuroligin 1; *NLGN4X*, Neuroligin 4 X-Linked; *IL1RAPL1*, Interleukin 1 Receptor Accessory Protein Like 1; *PTEN*, Phosphatase and Tensin Homolog.