Prioritization of epileptogenic-associated genes through comparative transcriptome analyses

Sreyoshi Chatterjee
German Aerospace Center
Santiago Alvarez-Cardenas
University of Luxembourg
Veronica Codoni
University of Luxembourg
Zied Landoulsi
Luxembourg Institute of Health
Alexander Skupin
University of Luxembourg
Holger Lerche
University of Tübingen
Patrick May
University of Luxembourg
Reinhard Schneider
University of Luxembourg
Roland Krause
roland.krause@uni.lu

University of Luxembourg

Research Article

Keywords: Focal epilepsy, generalised epilepsy, exploratory analysis, omics data integration, gene level analysis, epileptogenesis

Posted Date: August 5th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7175225/v1

Additional Declarations:

No competing interests reported.

Table S1 is not available with this version.

Prioritization of epileptogenic-associated genes through comparative transcriptome analyses

Sreyoshi Chatterjee[1,2], Santiago Alvarez-Cardenas[1], Veronica Codoni[1], Zied Landoulsi[1,3], Alexander Skupin[1], Holger Lerche[4], Patrick May[1], Reinhard Schneider[1], and Roland Krause[1]

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg
[2] Institut of Air Transport, German Aerospace Center, Germany
[3] Transversal Translational Medicine, Luxembourg Institute of Health, Luxembourg
[4] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

Abstract

Background:

Gene expression in the context of epilepsy syndromes has been studied in a variety of animal epilepsy models and in human patient samples. Joint analyses of this diverse range of data and the integration of transcriptomics with genetic data may identify genes involved across syndromes. These shared genes and processes present the opportunity to classify a core set of molecular pathways relevant to epileptogenesis.

Methods:

Here we performed a comparative analyses of organismal models covering major induction mechanisms, sequencing technologies and animal systems to identify genes common to the two principal syndrome categories of generalized and focal epilepsies, which represent distinct pathophysiological mechanisms of seizure initiation and development. We integrated the differentially expressed genes with genome-wide association studies to identify representative models, deregulated genes across epilepsy models and associations with other diseases.

Results:

Models for generalised epilepsies show fewer genes are differentially expressed than those for focal epilepsies. We found no overlap across all models but a group of 25 genes were detected across all focal models. Of these *LRRC8B* encodes a gliotransmitter previously not implicated in epileptogenesis. Integration with human genetic data highlights two models to be particularly representative for genetic

generalised epilepsy. We also demonstrate the previously recognized relation with cancers across the majority of the focal models in contrast to the generalised models.

Conclusions:

We suggest *LRRC8B* as a new gene involved in the epileptogenesis of focal seizures. Our analyses highlights pathways of particular relevance in the molecular biology of epilepsy and underscores the need for further development of generalised models.

Keywords:

Focal epilepsy, generalised epilepsy, exploratory analysis, omics data integration, gene level analysis, epileptogenesis.

Background

Epilepsy is a group of chronic neurological conditions characterized by spontaneous seizures that affect millions of people [1]. The primary classification of seizures distinguishes focal seizures that have a local region of origin, from generalised seizures that emerge from the cortex as a whole. The diagnosis of a particular epileptic syndrome relies on the specific seizure semiology, epileptic discharges in the electroencephalogram (EEG) and imaging using MRI[2], with focal and generalised onset being used as the primary distinction of syndromes.

Many generalised and some focal epilepsy types are known to be inherited and many genes harboring causal variants have been identified [3]. The number of genes associated with epilepsy ranges between 84 and 977, depending on the stringency of the association and definition of epilepsy[4].

Genome-wide association studies (GWAS) have also identified epilepsy-specific genes [5] and their responses to several antiepileptic drugs[6]. The mere relation of a gene and a trait through genetics is insufficient to establish the implication of the gene in the disease development. Therefore, current GWAS evaluate their results using transcriptomics data as well as other functional information[7].

The study of rare variants through sequencing allows for the identification of the origin of a mutation [8–11]. For some genes, e.g. the sodium channel 1A (*SCN1A*), its role in Dravet syndrome (an epileptic encephalopathy) has been well established [12] through CRISPR/Cas9 mutagenesis; for most genes however, we do not have a good understanding of why a particular genetic variant results in seizures. Larger studies on rare variants were able to implicate only particular groups of genes such as those in the $GABA_A$ receptors pathways [13, 14] or those identified through transcriptomics studies [15].

The relationship between gene and environment over time leading to the establishment of the condition of chronic seizures in the brain is collectively called epileptogenesis. Such process poses as an intervention target to block seizure onset. For most genes or conditions however, we do not have a deep understanding of the these processes [16, 17] but transcriptomics rise as a key tool for epileptogenesis' research.

Research investigating epileptogenesis in primary human brain tissue is naturally limited and dependent on availability of materials after surgery on people with treatment-resistant focal epilepsy [18] or postmortem tissue[19]. Thus, genome-wide RNA expression on animal models are a current standard to study epileptogenesis and many models have been established over time [20–23].

Transcriptomics studies for various models have been conducted individually but a comprehensive study combining the different animal models and technologies is yet to be carried out. Integration of data sets from different epilepsy backgrounds highlights in particular the genes common to all models across the phenotypic spectrum. Using information from a combination of animal models and genetic data may lead to the identification of genes that are deregulated in more than one epilepsy forms.

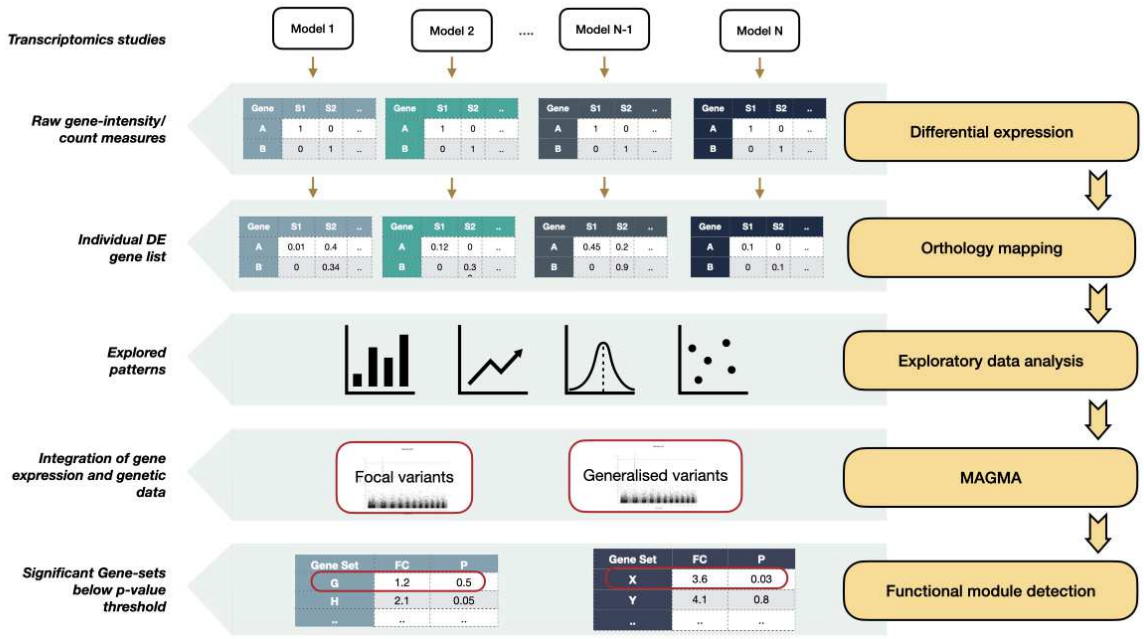


Figure 1: Study design. Datasets were collected from various animals models and transcriptomic technologies. Normalization of raw data were performed according to each models instructions and differential expression was calculated by a uniform method. GWAS data was integrated with orthology mapped DEGs to identify the gene-sets. Fuzzy clustering of the significant gene-sets was performed to identify modules from the protein-protein interaction networks.

In the present study, we compare epileptogenesis data sets from animal models and transcriptomic studies to assess their commonalities and identify genes from the intersection. We complement the transcriptomics results with GWAS data to characterize epilepsy-associated gene sets and further examined whether different parts of the brain are more susceptible to epileptogenic processes [24, 25] (see Figure 1). For simplicity, we refer hereafter to the transcriptomics data sets as *models* and to protein and RNA products of genes as *genes*.

We compare the enrichment of Gene Ontology and Disease Ontology terms in the models to characterize them. Multidimensional scaling allows us to interpret the similarity of the models with respect to key influences such as syndrome type, induction method and animal model.

Methods

Transcriptomic and genetic models

We surveyed the literature for transcriptomic studies of epileptogenesis from the Gene Expression Omnibus (GEO) with the keywords “epileptogenesis”, “transcriptomics”, “RNA-seq” and “sc-RNA-seq” since “year 2017”. Some data were requested and received from external authors [12, 18, 26, 27].

For the genetic data, summary statistics were obtained from the latest GWAS release [28] and the two major phenotype categories “genetic generalised epilepsy” and “focal epilepsy” [28] were considered for subsequent analysis.

Multidimensional scaling

We projected the high dimensional space spanned by the genes in the expression matrix into low dimensional space for visualizing the similarities of the models with multidimensional scaling (MDS). We calculated the similarity by euclidean distance of the genes by their fold-changes between seizure vs. control samples for each model, applying the *cmdscale* function in the *stats* package in R version 4.5.1 [29].

The models were grouped by seizure type, mechanism, model species and sequencing technologies.

Correlation analyses

Since the fold changes were not assumed to follow a normal distribution as they are from different technologies and models, pairwise correlations of gene expression changes were calculated using the non-parametric Spearman’s rank correlation function *corr* in R using the log2-fold changes in the genes.

Differential expression analyses and orthology mapping

We received RNA-seq and scRNA-seq count data, as well as CEL files for microarray analyses. DEGs were identified using packages Limma[30] for microarrays and DESeq2[31] for RNA-Sequencing and single-cell RNA-Sequencing, respectively. Conos [32], a wrapper of DESeq2, was used to process the TLE-EX and TLE-IN samples as described in the original publication[18].

We computed differentially expressed genes (DEG) separately for each model and compared the models on the level of DEGs as this allowed the inclusion of models for which no raw data could be obtained.

For sc-RNA-seq models, differential expression analysis contrasted cases and controls in bulk rather than by cell type to make sc-RNA-seq and RNA-seq results directly comparable, as cell type markers description are not available for bulk RNA-seq. The p-values were adjusted using multiple testing correction by Benjamini and Hochberg and a threshold of 0.05 was considered as the false discovery rate (FDR) cutoff universally. Genes were considered as differentially expressed based on their adjusted p-value cutoff.

PILO, KAI and SSSE models were processed for days 1, 3 and 10 post-induction separately (Table 1) and the combined unique number of genes has been reported for each dataset. For the PILO model, we followed the primary authors and removed samples that were identified as outliers by principal component analysis (PCA) [33].

Similarly, the samples from the SCN1A model [12] samples for days 4 and 7 post-fertilization were processed separately and the total DEG count for both days were analyzed jointly.

For the TLE-EX and TLE-IN models, DEGs were calculated on cases vs. controls of each dataset separately. Apart from this, TLE-EX and TLE-IN were treated as separate models throughout this study as they have different neuronal subtypes. Additionally, for TBI-CX and TBI-TH models, the DEG intersection is represented as a single dataset called TBI. Otherwise, these two subtypes are analyzed as separate models in this work.

Genes from animal species were mapped to the human gene set with orthology information from the ENSEMBL database version(1.8) [34].

Enrichment analyses

Disease ontology

Disease Ontology (DO) terms associated most strongly with DEGs were calculated with the DOSE package [35].

Gene ontology

The GO terms were calculated on the basis of their DEGs for the BP (Biological Processes) component using the enrichGO function from the clusterProfiler [36] package in R.

GWAS Integration

MAGMA [25] version 1.10 was used to calculate significant gene-sets from the gene expression and GWAS summary statistics by testing the joint association of genes present against the association of genes absent in the gene set [25]. Common single nucleotide polymorphisms (SNPs) identified by the GWAS categories FE and GE respectively were mapped to the reference genome (GRCh37, annotation release 105) and annotated with the information of SNP locations to their corresponding gene locations. Next, a SNP-wise mean test was performed along with BED file information from the european subset of 1000 Genomes project (phase 3) used as a reference panel to estimate linkage disequilibrium (LD) between SNPs [38]. The test utilized SNP locations given by chromosome start and stop positions, number of SNPs on gene, measure of gene association ZSTAT and gene p-values.

Tissue-level RNA expression

Tissue and transcript-level expression for the Leucine-Rich Repeat Containing 8 (*LRRC8*) gene family members (*LRRC8A*, *LRRC8B*, *LRRC8C*, *LRRC8D*, *LRRC8E*) were obtained from the Human Protein Atlas (HPA), version 24.0 Ensemble v109 (v24.proteinatlas.org). Specifically, we accessed the RNA consensus data, which represents the maximum transcript expression levels for each gene across 50 different human healthy tissues[39] as million transcripts per total annotated transcripts (nTPM).

To determine the cell-type specific expression of *LRRC8B*, we interrogated the Single Cell Type section of HPA, which provides gene expression values for 81 common human cell types as measured by scRNA-seq on healthy tissue of 31 different tissue types [39].

Single cell-type expression clustering was investigated via the expression cluster functionality of HPA. This analysis identifies genes with similar expression patterns across bulk tissues and scRNA-seq data [40]. We specifically focused on genes coexpressed with *LRRC8B* to identify potential functional relationships. The result of the cluster analysis is presented as a UMAP based on gene expression, where each cluster has been summarized as colored areas containing most of the cluster genes (see Figure S3).

Results

Data models inclusion

In our study we included transcriptomics data from 15 models. We aimed for heterogeneity in the models' induction (induced or genetic), transcriptomics system (microarray, bulk RNA-sequencing (RNA-seq), single cell or single nucleus RNA-sequencing, and biological sources (mice, rats, zebrafish models as well as primary human tissue). The final models were chosen such that a variety of epilepsy models was covered (see Table 1). We excluded data without an analysis manuscript accompanying the data.

Within the models we selected the experimental conditions that captured the seizure onset best. The SCN8A pre-seizure low threshold spike (LTS) models was excluded from analysis as no seizures were observed in the animals [27]. The dataset GABRG2_KO was dropped from differential expression and GWAS integration as no DEGs were found after FDR correction. We processed the models and mapped all genes to the human orthologs as described above (see Table 2).

Multidimensional scaling analyses

To identify the factors that influence the similarity of the expression sets we reduced the dimensionality of the data with MDS.

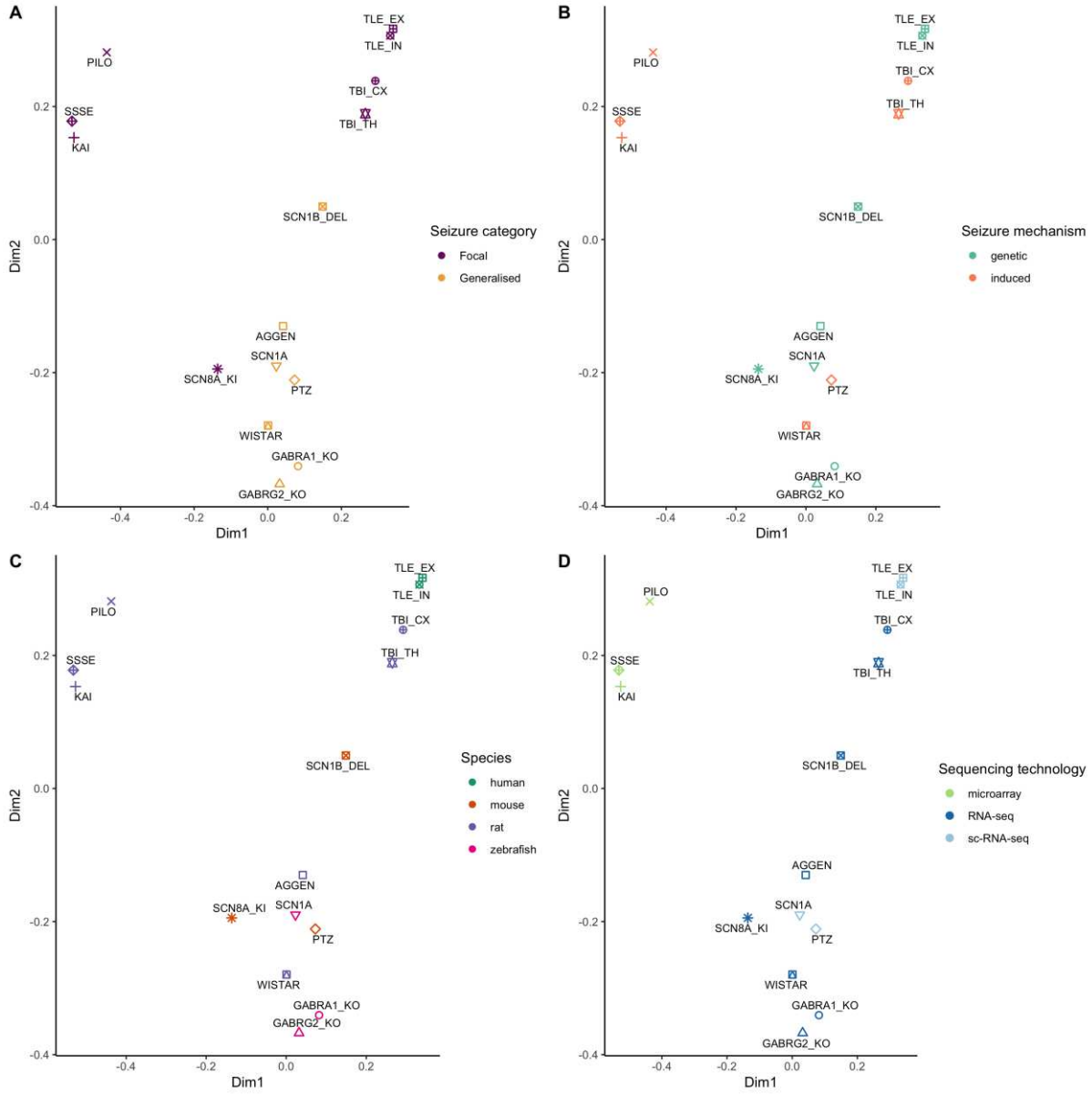


Figure 2: Multi dimensional scaling based on fold changes. Models are labeled according to Table 1 and colored accoring to seizure categories (A), seizure induction mechanisms (genetic or induced (B), sequencing technologies (C) and according to species (D).

The primary distinctive feature of seizures is the the seizure type, focal and generalised sets with the exception of SCN8A_KI (focal) and SCN1B_DEL (generalised), both mouse models. (Figure 2 A and B). This difference can be attributed to the method of seizure induction, as SCN8A_KI and SCN1B_DEL are both genetic models while PTZ, another mouse, generalised, induced is retained closer to the other generalised, induced models. The models cluster by species with the notable exception of rats, as the difference in sequencing technologies and induction and seizure type are strong drivers of differentiation. The sequencing technologies place sc-RNA-seq and bulk RNA-seq in one cluster and the microarray models in another (Figure 2 C and D).

Pairwise correlation of models

We computed the pairwise correlation of all models to find further similarities. We found that models from the same laboratories with similar seizure generation patterns, such as TBI_TH and TBI_CX, showed positive correlations (ρ >= 0.6), as well as PILO, KAI and SSSE which also share in common microarray sequencing and species. TLE_EX and TLE_IN exhibited a high correlation (ρ = 0.76) sharing both species and sequencing technology (Figure 3). No other significant correlations were observed among the remaining models with respect to species, sequencing technologies, or laboratories, including between GABRA1_KO and GABRG2_KO.

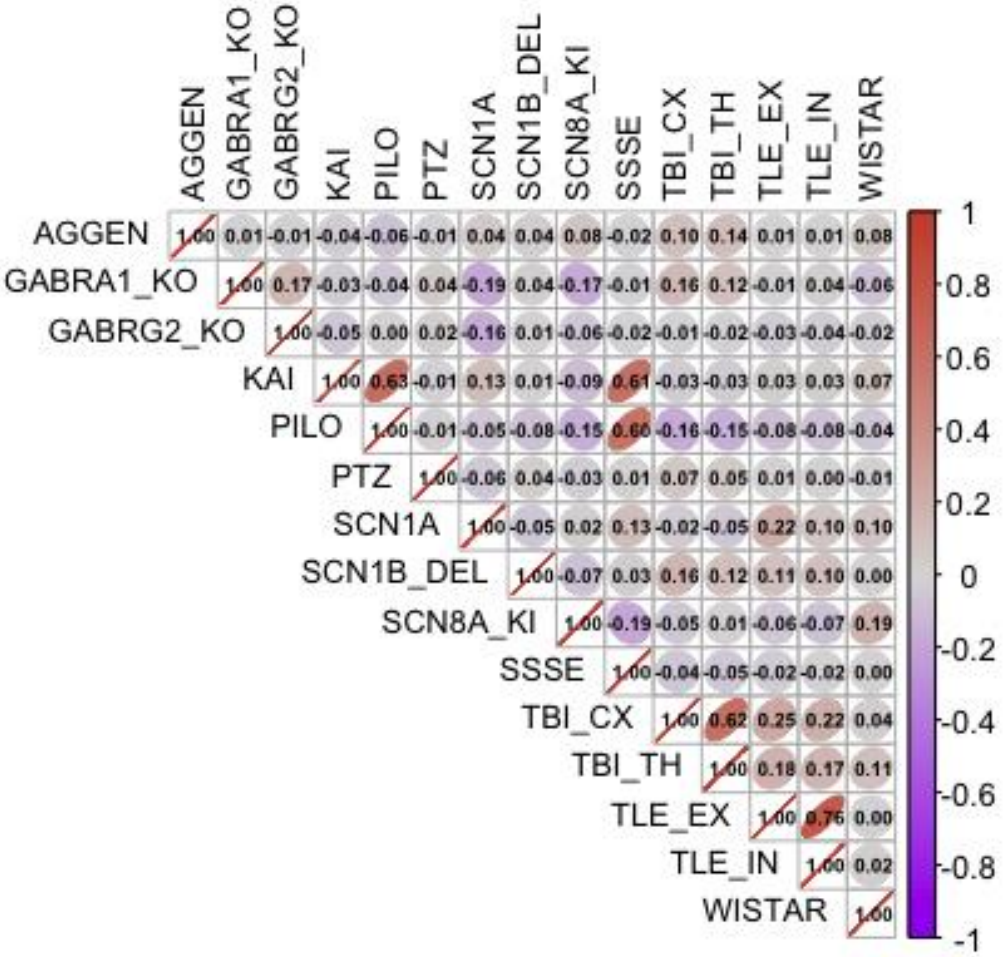


Figure 3: Pairwise correlations of the models. models labels according to Table 1.

Differential expression analysis

The total number of DEGs identified by model ranges from 16 to 8085 (Table 3), excluding GABRG2_KO as it did not include any DEGs.

As no DEG was found across all models, we investigated intersections between the models' major characteristics such as seizure type and seizure induction. We found no gene in common to all generalised seizure models either (Figure 4 A); interestingly, only

two genes were found in common between seizure induction type (induced or genetic) for generalised models (S100A16 and RPL29). When comparing all focal models, we found just TGFBR1, PLXDC2 and KCNA2 intersecting all data sets (Figure 4 C). After merging the two most disjoint sc-RNA-seq data sets into TBI (from TBI_CX and TBI_TH) and TLE_HUMAN (from TLE_EX and TLE_IN) we found 25 genes that are intersecting (Figure 4 B).

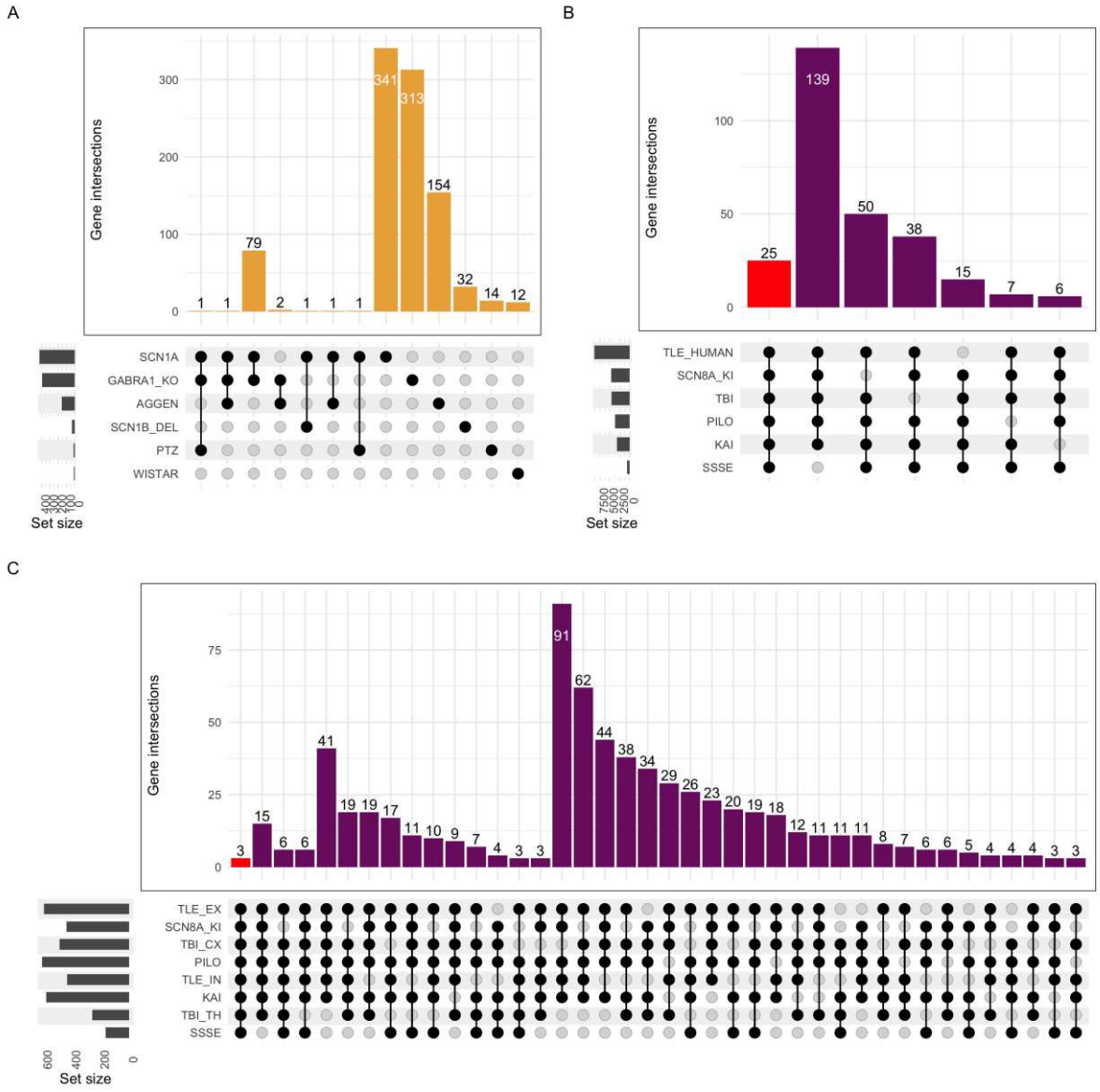


Figure 4: DEGs intersection for focal and generalised epilepsy models. Models are labeled according to Table 1. A: DEG intersections considering all generalised models.B: DEG intersections of the focal models with merged subtypes from the same seizure mechanism. Red bar shows the common intersection among all the models. C: DEG intersections for all the focal models.

All genes except for *LRRC8B* were previously implicated in epilepsy and other neurological diseases (Table 7) [41–43]. Interestingly, the only two genes in common for

both induced and genetic origin in generalised models, *S100A16* and *RPL29*, are related to tumor cell proliferation and invasion [44] and regulation of translation efficiency [45], respectively.

The average number of DEGs in the focal models is substantially higher than in the generalised models. Following the Wilcoxon test (p <3.11e-04) suggesting that there are significant differences between the groups (Figure 5).

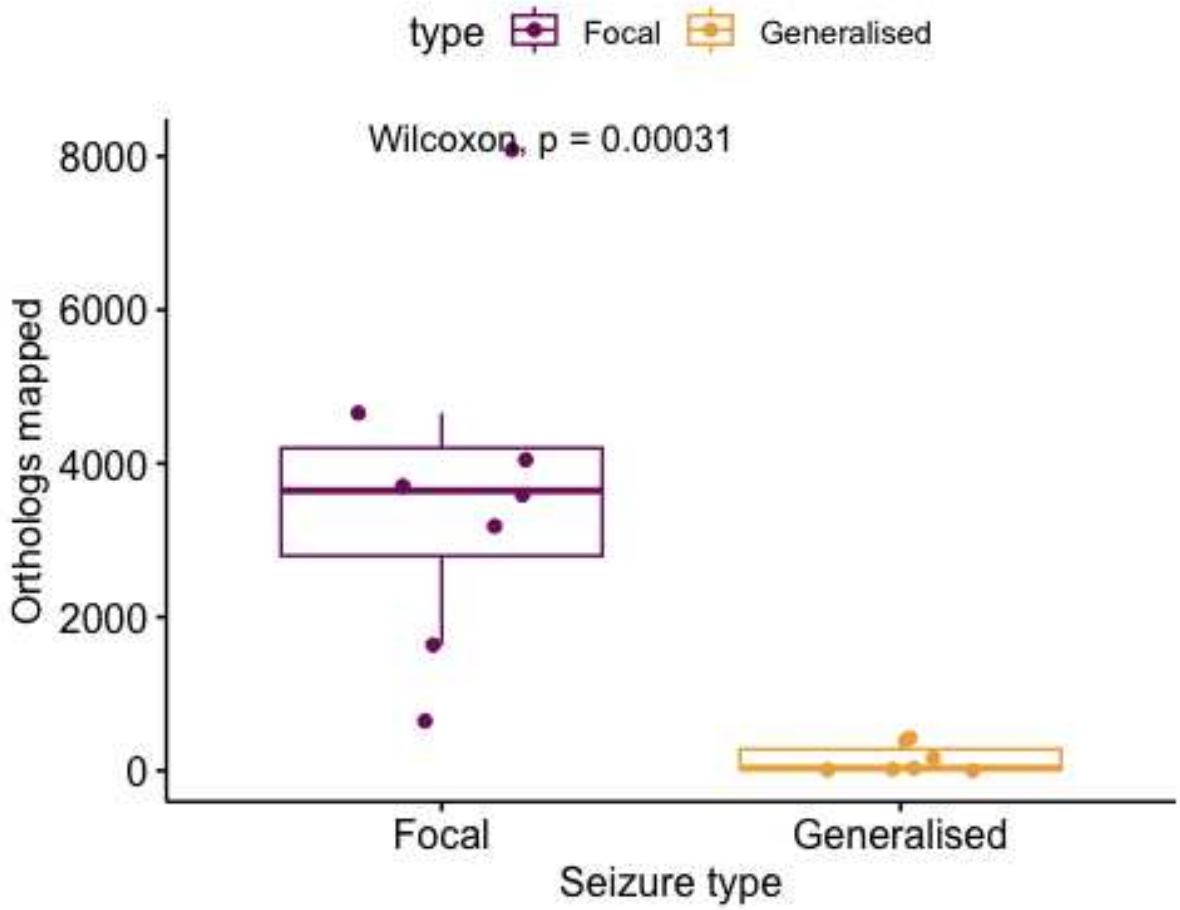


Figure 5: Wilcoxon test for comparison of focal and generalised models.

Enrichment analyses

Disease ontology

To identify associated diseases beyond epilepsy for the models we performed a disease enrichment analysis (see Figure 6).

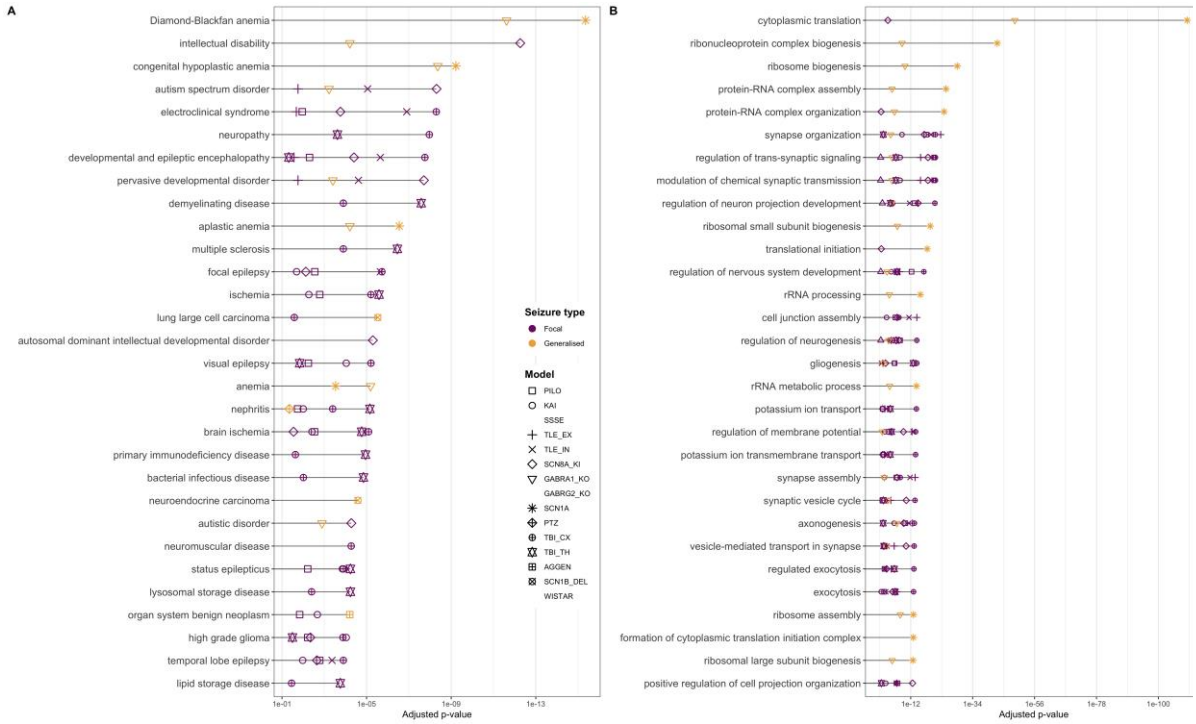


Figure 6: Disease ontology (DO, Panel A) and Gene ontology (GO, Panel B) Ontology terms enrichiched in DEGs by epilepsy model. The top 30 terms with the lowest p-values are depicted.

Epilepsy syndrome and *Developmental disorders* were found to be the most significant terms and are enriched in most focal models while being completely absent in generalised epilepsy.

The generalised models except for PTZ were associated with anemic disorders like *anemia*, *aplastic anemia* and *Diamond-blackfan anemia* (Figure 6) unlike focal models. While terms like *intellectual disability* and *developmental disorder of mental health* are more closely related to the definition of generalised epilepsy [46], they were detected only in the GABRA1_KO model.

Gene ontology analyses

GO analysis using the BP component showed that the generalised models SCN1A and GABRA1_KO were enriched in almost all the top terms including *translation* and *ribosome biogenesis* processes (almost the sole representatives), as well as *Neurogenesis* and *Synaptic Processes* (Figure 6). The *potassium ion transport*, *exocytosis* and *cell junction assembly* were exclusively targeted by most of the focal models only. The SCN1A model was involved in the *formation of cytoplasmic translation initiation complex* term solely. Key features of epilepsies relating to *neurogenesis*, *dendrite development*, *nervous system development*, *axonogenesis* etc. were enriched in many focal models and only in the GABRA1_KO model for the generalised sets.

GWAS integration

We compared the DEGs with genetic data from the most current encompassing epilepsy GWAS [28] study using MAGMA [25].

GABRA1_KO (MAGMA p-value = 0.003) and AGGEN (MAGMA p-value = 0.004) (Table 4), as well as TBI_TH (MAGMA p-value = 0.03) (Table 5) datasets showed significant differential expression during the analysis using MAGMA for the generalised and focal models, respectively.

Expression patterns of LRRC8 family members across human tissues

The consensus transcript expression data from the Human Protein Atlas (v24) shows distinct expression patterns for the five members of the LRRC8 gene family across human tissues (see Figure S2).

Broad expression of LRRC8A and LRRC8D

LRRC8A, encoding the essential subunit for Volume-Regulated Anion Channels (VRACs), exhibited the highest and broadest expression among the family members. High nTPM values were observed in numerous tissues and, notably, across all analyzed brain regions (Figure S2, yellow bars). Brain tissues such as the cerebellum, cerebral cortex, midbrain, hypothalamus, hippocampal formation, amigdala and basal ganglia showed particularly high *LRRC8A* expression. *LRRC8D* also displayed relatively broad expression, albeit generally at lower levels than *LRRC8A*. Moderate *LRRC8D* expression was also detected in some tissues and prominently in spinal cord, cortex and midbrain.

LRRC8B is broadly expressed in the brain

In contrast to *LRRC8A* and *LRRC8D* gene expression, *LRRC8B* had overall lower expression levels. Notably, *LRRC8B* exhibited tissue-enhanced expression in the brain, with mRNA levels at least four-fold higher than the average across all other tissues (Figure S2, yellow bars). Within the brain, *LRRC8B* showed relatively low regional specificity. Outside the central nervous system, notable expression was also seen in the testis and adrenal gland. The relatively specific expression pattern of *LRRC8B* within the brain, particularly in regions like the cerebral cortex (21.7 nTPM), spinal cord (11.8 nTPM) and hippocampus (10.4 nTPM) which are often implicated in seizure generation, highlights its potential specific role in neuronal function and pathophysiology, distinct from the more ubiquitous *LRRC8A*.

LRRC8C and *LRRC8E* generally exhibited low to very low expression across most tissues analyzed (Figure S2). *LRRC8C* showed some detectable expression in smooth muscle, adipose tissue, and low levels in brain regions and lymphoid tissues. *LRRC8E* expression was minimal across almost all tissues, with only trace levels detected in cerebral cortex. *LRRC8B* mRNA was higly detected in both excitatory (171.6 nTPM) and inhibitory neurons (188.9 nTPM), as well as in oligodendrocytes (131 nTPM).

LRRC8B co-expresses with genes involved in synaptic function

Single-cell type clustering analysis revealed that *LRRC8B* gene expression clusters with genes known to be involved in neuronal synaptic function (Figure S3). Among the genes with similar expression patterns were *GABBR2* (Gamma-aminobutyric acid type B receptor subunit 2) and *ADAM22* (A Disintegrin and Metalloproteinase 22), both of which have been previously implicated in epilepsy [19, 47].

Discussion

Our goal was to provide context for the transcriptomics models. A key questions was to identify DEG common to many models as potential targets of intervention. The number of differentially expressed genes varied significantly with the sc-RNA-seq methods yielding many more genes than other methods.

This might be a limitation of our naive treatment of single cell technologies as bulk sequencing but a large number of genes is also seen in the primary analysis of the data [18]. Some models yielded few DEGs (like WISTAR) and GABRG2_KO did not yield any, again in line with the primary analysis [9]. The disparity is supported by the varying numbers of animals or tissues employed in the individual study (see Table 1).

AGGEN, GABRA1_KO and TBI_TH were identified as representative gene-sets for genomics studies on the basis of their significant p-value < 0.05 for focal and generalised models respectively.

No common DEG is found across all models, which would not be too surprising giving that many genes are known to influence epileptogenesis but no single genes had been previously highlighted as universal.

The 25 genes we found across the focal models are a substantial contribution to a core set of genes involved in focal epilepsies. Outstanding is the gliotranmitter gene *LRRC8B* which had not been previously associated with epilepsy[48].

Genes like *TGFBR1*, *PLXDC2* and *KCNA2* related to temporal lobe epilepsy and encephalopathy [8, 49, 50] arise in all the focal models.

LRRC8B (Leucine Rich Repeat Containing 8 VRAC Subunit B) is a protein coding gene from the LRRC8 family required to maintain constant cell volume in response to extra-cellular and intra-cellular osmotic changes [51]. Primarily involved in the brain-synaptic functions, this gene is expressed in neurons and oligodendrocytes [52]. *LRRC8A*, an important paralog of this gene, is the obligatory pore-forming subunit of the volume-regulated anion channel (VRAC) and its function requires heteromeric assembly with at least one other LRRC8 family member (*LRRC8B-E*); The specific subunit composition dictates channel characteristics, including ion selectivity and conductance [53, 54].

LRCC8A is heavily involved in the central nervous system (CNS) and with rat epilepsy models [55, 56]

LRRC8B is among the 25 DEG found across all the focal models. Evidence from the Human Protein Atlas and GTEx demonstrate broad brain expression, and single-cell RNA-seq data further highlight its presence in excitatory and inhibitory neurons, as well as oligodendrocytes. The role of *LRRC8B* containing VRAC, described as gliotransmitter, in cellular processes is multifaceted. Beyond its primary function in cell volume regulation by conducting chloride and organic osmolytes, it also conducts iodide better than chloride [51]. Furthermore, *LRRC8B* has been shown to participate in intracellular Ca2+ homeostasis by acting as a leak channel in the endoplasmic reticulum [57]. Critically, LRRC8 channels, including those containing LRRC8B, are capable of transporting neurotransmitters, including glutamate and aspartate. While LRRC8B or LRRC8C disruption did not decrease transport rates of all substrates, their inclusion into LRRC8 heteromers influenced the substrate preference of VRAC [53]. This suggests a complex, composition-dependent role for VRACs, and potentially LRRC8B-containing channels, in modulating extracellular glutamate levels. Given that glutamate excitotoxicity is a key factor in seizure generation and neuronal damage in epilepsy, the potential contribution of LRRC8B to glutamate release, either directly via VRAC or indirectly through its influence on ER calcium, warrants careful consideration. Furthermore, *LRRC8B*'s expression profile clusters with genes involved in neuronal synaptic function, including *GABBR2* and *ADAM22*, both of which have established links to epilepsy [19, 47]. LRR proteins, are known to play vital roles in synapse formation, differentiation, and synaptic plasticity [58]. This association confirm a potential indirect role for *LRRC8B* in synaptic processes relevant to epileptogenesis. Blood-brain barrier (BBB) dysfunction is also implicated in epileptogenesis, and could be a factor to consider [59]. Although the direct involvement of *LRRC8B* in epilepsy remains to be fully elucidated, its multifaceted roles in cell volume regulation, glutamate transport, calcium homeostasis, and synaptic function, combined with its expression pattern in relevant cell types, suggest that it represents a novel point of interest in understanding the pathophysiology of focal epilepsy.

There is no common gene among the generalised models and can be attributed to the difference in seizure generation of the models. The 81 intersecting genes between GABRA1_KO and SCN1A models points to similarities between both models such as knock-out and zebrafish species which may influence the number of genes. Albeit finding 2 common genes when comparing the epilepsy induction type, only 3 (PTZ, GABRA1_KO and SCN1A) out of the 6 generalised models provided these genes.

Similarity and correlations among models from different backgrounds

Through MDS we demonstrated that models with similar seizure categories i.e focal and generalised had higher similarity based on DEGs.

Exceptionally SCN8A_KI has 4656 intersecting DEGs and was closer to the generalised models. Its differential expression changes must be investigated further to consider the similarity with generalised models.

For most of the models the Spearman's correlations are as expected i.e higher for models coming from the same laboratory or study and lower if they are from contrasting experiments.

Similarly, mild negative correlation (ρ = -0.19) is observed amongst GABRA1_KO and SCN1A, both being genetic generalised zebrafish larvae models.

However, considering the correlation strength of the DEGs of the models only, there is perfect negative correlation (ρ = -1) between SCN1A and AGGEN, although both are generalised models (See Figure S1).

Relation of focal epilepsy and glioma

The GO term *epilepsy syndrome* and *brain disease* have been identified by most of the focal epilepsy models which agrees with the definition of focal seizures [60]. The term *focal epilepsy* itself was identified by TLE_IN and TBI_CX (adjusted p-value < 0.0005).

Among the top DO terms, focal epilepsy models are mostly related to cancer terms like *neuroblastoma*,*glioma*, *connective tissue cancer*. (see Table S1).

The relationship of focal epilepsy to brain tumors has been researched extensively in the last decade [61–63] and brain tumors are the second most common cause of focal intractable epilepsy in epilepsy surgery series, with the greatest proportion related to dysembryoplastic neuroepithelial tumors and gangliogliomas [64].

Moreover, gene *LRRC8B* as a common gene among all the focal models indicates that this interconnection of seizure phenotypes with brain tumors should be followed more closely in future studies.

Concordance between gene expression and genetic studies

For GE and FE, we evaluated the enrichment of every gene-set with the corresponding epilepsy type. We found that GE risk variants are significantly associated with GABRA1_KO (p-value = 0.0038) and AGGEN(p-value = 0.0042) and crucially, no focal epilepsy gene sets were significantly differentiated. Hence, many genes representing these sets are both differentially expressed and genetically associated with epilepsy phenotypes.

A further clustering of these sets may highlight pathways and hubs for epileptogenesis.

Comparing heterogenous epilepsy models at the transcriptomics level requires to use pragmatic solutions. For the DEG analysis, only the p-value cutoff has been considered for filtering the DEGs and not the log fold changes. A log fold change filtering cutoff

would mean that most models would not have contributed any DEG. However, orthology mapping is likely to obscure some details but any solution other than using the standard dataset from ENSEMBL is unlikely to be managable or result in a substantial increase in numbers. For sc-RNA-seq studies, sample information might have been lost as the DEGs were calculated similar to bulk sequencing to maintain dataset DEG uniformity rather than cell-type marker analysis.

Conclusions

Generalised and focal epilepsies are characterized distinctly by the ILAE based on seizure types but have not yet been compared directly in terms of their transcriptomics and genetic constituency.

Different brain tissues can be accounted for these seizure types and a deeper understanding of the genes involved in each of these epilepsy types may help to target these genes for future validation and recovery for patients. We compared these two epilepsy types from evidences on their gene expression and genetic data thereby providing insights to the key genes involved.

Integration of gene expression and GWAS studies highlights the genes which have both genetic attribute to them and additionally get differentially expressed w.r.t controls.

Consequently, enrichment analysis of these gene-sets reveals the chief pathways and genes associated with a particular kind of epilepsy.

As the data selected for the present study come from different animal models, it provides an overview for the focal and generalised seizures. To further extend this, in-depth analysis of multiple studies from the same sequencing technology should be carried out.

Data and materials

Code availability: The workflow using targets pipeline manager in R is openly available under [https://git-r3lab.uni.lu/mechepi/geneintegration_magma]. The public datasets can be downloaded from their respective repositories given in the code. The data sources can also be found here https://sreyoshi.shinyapps.io/shiny/.

Abbreviations

ILAE: International League Against Epilepsy; GO: Gene ontology; DO: Disease ontology; FDR: False discovery rate; FC: Fold change; TLE: Temporal lobe epilepsy; TBI: Traumatic brain injury; KO: Knock out; KI: Knock in; DEG: Differentially expressed gene; GWAS: genome wide association studies; w.r.t: with respect to. nTPM: transcripts per million

Acknowledgements

We thank Dr. Eric Samarut for GABRA1_KO and GABRG2_KO datasets, Dr. Ryan Sprissler for sharing the SCN8A datasets.

Funding

This work was supported by the Research Unit FOR-2715, funded by the German Research Foundation (DFG) and the Fond Nationale de la Recherche (FNR) in Luxembourg (grants Le1030/16-1 to HL, INTER/DFG/17/11583046 and INTER/DFG/21/16394868 to AS/RK/PM).

Author information

SC performed initial bioinformatics and statistical analyses. SAC improved the computational pipeline. VC performed the biological interpretation. RK designed and led the study. ZL, PM and RK supported the interpretation. AS provided experimental data. SC, SAC, VC and RK wrote the manuscript. All authors have read and approved the manuscript.

References

1. Feigin V, Collaborators G2016E, et al. Global, regional, and national burden of epilepsy, 1990-2016: A systematic analysis for the global burden of disease study 2016. The Lancet Neurology. 2019;18:357–75.

2. Scheffer IE, Berkovic SF, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;1–10. doi:10.1111/epi.13709.

3. Mullen SA, Berkovic SF, Commission IG, Berkovic SF, Lowenstein DH, Kato M, et al. Genetic generalized epilepsies. Epilepsia. 2018;59:1148–53.

4. Wang J, Lin Z-J, Liu L, Xu H-Q, Shi Y-W, Yi Y-H, et al. Epilepsy-associated genes. Seizure. 2017;44:11–20.

5. Consortium TILAE et al. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nature communications. 2018;9.

6. Wolking S, Schulz H, Nies AT, McCormack M, Schaeffeler E, Auce P, et al. Pharmacoresponse in genetic generalized epilepsy: A genome-wide association study. Pharmacogenomics. 2020;21:325–35.

7. International League Against Epilepsy Consortium on Complex Epilepsies. GWAS meta-analysis of over 29,000 people with epilepsy identifies 26 risk loci and subtype-specific genetic architecture. Nature genetics. 2023;55:1471–82.

8. Hundallah K, Alenizi A, AlHashem A, Tabarki B. Severe early-onset epileptic encephalopathy due to mutations in the KCNA2 gene: Expansion of the genotypic and phenotypic spectrum. european journal of paediatric neurology. 2016;20:657–60.

9. Liao M, Kundap U, Rosch RE, Burrows DR, Meyer MP, Bencheikh BOA, et al. Targeted knockout of GABA-a receptor gamma 2 subunit provokes transient light-induced reflex seizures in zebrafish larvae. Disease models & mechanisms. 2019;12.

10. Helbig I, Lopez-Hernandez T, Shor O, Galer P, Ganesan S, Pendziwiat M, et al. A recurrent missense variant in AP2M1 impairs clathrin-mediated endocytosis and causes developmental and epileptic encephalopathy. The American Journal of Human Genetics. 2019;104:1060–72.

11. Wallace RH, Marini C, Petrou S, Harkin LA, Bowser DN, Panchal RG, et al. Mutant GABAA receptor γ2-subunit in childhood absence epilepsy and febrile seizures. Nature Genetics. 2001;28:49–52.

12. Tiraboschi E, Martina S, Ent W van der, Grzyb K, Gawel K, Cordero-Maldonado ML, et al. New insights into the early mechanisms of epileptogenesis in a zebrafish model of dravet syndrome. Epilepsia. 2020;61:549–60.

13. May P, Girard S, Harrer M, Bobbili DR, Schubert J, Wolking S, et al. Rare coding variants in genes encoding GABAA receptors in genetic generalised epilepsies: An exome-based case-control study. The Lancet Neurology. 2018;17:699–708.

14. Feng Y-CA, Howrigan DP, Abbott LE, Tashman K, Cerrato F, Singh T, et al. Ultra-rare genetic variation in the epilepsies: A whole-exome sequencing study of 17,606 individuals. The American Journal of Human Genetics. 2019;105:267–82.

15. Koko M, Krause R, Sander T, Bobbili DR, Nothnagel M, May P, et al. Distinct gene-set burden patterns underlie common generalized and focal epilepsies. EBioMedicine. 2021;72:103588.

16. Pitkänen A, Lukasiuk K, Dudek FE, Staley KJ. Epileptogenesis. Cold Spring Harbor perspectives in medicine. 2015;5:a022822.

17. Vezzani A, Friedman A, Dingledine RJ. The role of inflammation in epileptogenesis. Neuropharmacology. 2013;69:16–24.

18. Pfisterer U, Petukhov V, Demharter S, Meichsner J, Thompson JJ, Batiuk MY, et al. Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis. Nature communications. 2020;11:1–19.

19. Van Der Knoop MM, Maroofian R, Fukata Y, Van Ierland Y, Karimiani EG, Lehesjoki AE, et al. Biallelic *ADAM22* pathogenic variants cause progressive encephalopathy and infantile-onset refractory epilepsy. Brain. 2022;145:2301–12. doi:10.1093/brain/awac116.

20. Grone BP, Baraban SC. Animal models in epilepsy research: Legacies and new directions. Nature neuroscience. 2015;18:339–43.

21. Schauwecker PE. The relevance of individual genetic background and its role in animal models of epilepsy. Epilepsy research. 2011;97:1–11.

22. Hansen KF, Sakamoto K, Pelz C, Impey S, Obrietan K. Profiling status epilepticus-induced changes in hippocampal RNA expression using high-throughput RNA sequencing. Scientific reports. 2014;4:1–12.

23. Qureshi IA, Mehler MF. Epigenetic mechanisms underlying human epileptic disorders and the process of epileptogenesis. Neurobiology of disease. 2010;39:53–60.

24. De Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nature Reviews Genetics. 2016;17:353.

25. Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.

26. Samarut É, Swaminathan A, Riché R, Liao M, Hassan-Abdi R, Renault S, et al. γ-aminobutyric acid receptor alpha 1 subunit loss of function causes genetic generalized epilepsy by impairing inhibitory network neurodevelopment. Epilepsia. 2018;59:2061–74.

27. Sprissler RS, Wagnon JL, Bunton-Stasyshyn RK, Meisler MH, Hammer MF. Altered gene expression profile in a mouse model of SCN8A encephalopathy. Experimental neurology. 2017;288:134–41.

28. Berkovic SF, Cavalleri GL, Koeleman BP, et al. Genome-wide meta-analysis of over 29,000 people with epilepsy reveals 26 loci and subtype-specific genetic architecture. medRxiv. 2022.

29. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. https://www.R-project.org/.

30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43:e47–7.

31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15:1–21.

32. Barkas N, Petukhov V, Nikolaeva D, Lozinsky Y, Demharter S, Khodosevich K, et al. Joint analysis of heterogeneous single-cell RNA-seq dataset collections. Nature methods. 2019;16:695–8.

33. Dingledine R, Coulter DA, Fritsch B, Gorter JA, Lelutiu N, McNamara J, et al. Transcriptional profile of hippocampal dentate granule cells in four rat epilepsy models. Scientific data. 2017;4:170061.

34. Andrew D, Premanand A, Wasiu A, James A, Jamie A, Jorge A, et al. Ensembl 2020. Nucleic Acids Res. 2020;48:D682–8.

35. Yu G, Wang L-G, Yan G-R, He Q-Y. DOSE: An r/bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015;31:608–9.

36. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: An r package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16:284–7.

37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102:15545–50.

38. Consortium 1000GP et al. A global reference for human genetic variation. Nature. 2015;526:68.

39. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419. doi:10.1126/science.1260419.

40. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Science Advances. 2021;7:eabh2169.

41. Asadollahi R, Oneda B, Joset P, Azzarello-Burri S, Bartholdi D, Steindl K, et al. The clinical significance of small copy number variants in neurodevelopmental disorders. Journal of Medical Genetics. 2014;51:677–88.

42. Møller R, Hansen C, Jackson G, Ullmann R, Ropers H-H, Tommerup N, et al. Interstitial deletion of chromosome 4p associated with mild mental retardation, epilepsy and polymicrogyria of the left temporal lobe. Clinical genetics. 2007;72:593–8.

43. Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes, Brain and Behavior. 2014;13:686–701.

44. Yang D, Zhao F, Zhou Y, Zhang Y, Shen J, Yu B, et al. S100A16 is a potential target for reshaping the tumor microenvironment in the hypoxic context of liver cancer.

International Immunopharmacology. 2024;134:112076.
doi:https://doi.org/10.1016/j.intimp.2024.112076.

45. Hamidi T, Singh AK, Veland N, Vemulapalli V, Chen J, Hardikar S, et al. Identification of Rpl29 as a major substrate of the lysine methyltransferase Set7/9. Journal of Biological Chemistry. 2018;293:12770–80.
doi:https://doi.org/10.1074/jbc.RA118.002890.

46. Henkin Y, Sadeh M, Kivity S, Shabtai E, Kishon-Rabin L, Gadoth N. Cognitive function in idiopathic generalized epilepsy of childhood. Developmental medicine and child neurology. 2005;47:126–32.

47. Feng Y, Wei Z-H, Liu C, Li G-Y, Qiao X-Z, Gan Y-J, et al. Genetic variations in GABA metabolism and epilepsy. Seizure: European Journal of Epilepsy. 2022;101:22–9.
doi:10.1016/j.seizure.2022.07.007.

48. Bagla S, Cukovic D, Asano E, Sood S, Luat A, Chugani HT, et al. A distinct microRNA expression profile is associated with α [11C]-methyl-l-tryptophan (AMT) PET uptake in epileptogenic cortical tubers resected from patients with tuberous sclerosis complex. Neurobiology of disease. 2018;109:76–87.

49. Han C-L, Zhao X-M, Liu Y-P, Wang K-L, Chen N, Hu W, et al. Gene expression profiling of two epilepsy models reveals the ECM/integrin signaling pathway is involved in epiletogenesis. Neuroscience. 2019;396:187–99.

50. Balak N, Balkuv E, Karadag A, Basaran R, Biceroglu H, Erkan B, et al. Mammillothalamic and mammillotegmental tracts as new targets for dementia and epilepsy treatment. World Neurosurgery. 2018;110:133–44.

51. Voss FK, Ullrich F, Münch J, Lazarow K, Lutter D, Mah N, et al. Identification of LRRC8 heteromers as an essential component of the volume-regulated anion channel VRAC. Science (New York, NY). 2014;344:634–8.

52. García-Rodríguez C, Bravo-Tobar ID, Duarte Y, Barrio LC, Sáez JC. Contribution of non-selective membrane channels and receptors in epilepsy. Pharmacology & Therapeutics. 2021;107980.

53. Lutter D, Ullrich F, Lueck JC, Kempa S, Jentsch TJ. Selective transport of neurotransmitters and modulators by distinct volume-regulated LRRC8 anion channels. Journal of Cell Science. 2017;130:1122–33.

54. Syeda R, Qiu Z, Dubin AE, Murthy SE, Florendo MN, Mason DE, et al. LRRC8 Proteins Form Volume-Regulated Anion Channels that Sense Ionic Strength. Cell. 2016;164:499–511.

55. Wang Z, Li Y, Zeng Z, Guo S, Chen W, Luo Y. Leucine-rich repeat containing 8A contributes to the expansion of the potential role of leucine-rich repeat-containing

protein 8A in central nervous system: Current situation and prospect. Neuroscience. 2022.

56. Zahiri D, Burow P, Großmann C, Müller CE, Klapperstück M, Markwardt F. Sphingosine-1-phosphate induces migration of microglial cells via activation of volume-sensitive anion channels, ATP secretion and activation of purinergic receptors. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2021;1868:118915.

57. Ghosh A, Khandelwal N, Kumar A, Bera AK. Leucine-rich repeat-containing 8B protein is associated with the endoplasmic reticulum Ca2+ leak in HEK293 cells. Journal of Cell Science. 2017;130:3818–28. doi:10.1242/jcs.203646.

58. de Wit J, Ghosh A. Control of neural circuit formation by leucine-rich repeat proteins. Trends in Neurosciences. 2014;37:539–50. doi:https://doi.org/10.1016/j.tins.2014.07.004.

59. Marchi N, Granata T, Ghosh C, Janigro D. Blood–brain barrier dysfunction and epilepsy: Pathophysiologic role and therapeutic approaches. Epilepsia. 2012;53:1877–86. doi:https://doi.org/10.1111/j.1528-1167.2012.03637.x.

60. Engel Jr J. ILAE classification of epilepsy syndromes. Epilepsy research. 2006;70:5–10.

61. Politsky JM. Brain tumor-related epilepsy: A current review of the etiologic basis and diagnostic and treatment approaches. Current neurology and neuroscience reports. 2017;17:1–13.

62. Slegers RJ, Blumcke I. Low-grade developmental and epilepsy associated brain tumors: A critical update 2020. Acta neuropathologica communications. 2020;8:1–11.

63. Zhang X, Zheng L, Duan J, Li Z, Tang Y. Clinical characteristics of brain tumor-related epilepsy and factors influencing the identification of epilepsy-associated tumors. Acta Epileptologica. 2020;2:1–6.

64. Çetin ÖE, İşler C, Uzan M, Özkara Ç. Epilepsy-related brain tumors. Seizure. 2017;44:93–7.

65. Lipponen A, Paananen J, Puhakka N, Pitkänen A. Analysis of post-traumatic brain injury gene expression signature reveals tubulins, Nfe2l2, nfkb, Cd44 and S100a4 as treatment targets. Scientific reports. 2016;6:1–9.

66. Chuvakova LN, Funikov SY, Rezvykh AP, Davletshin AI, Evgen'ev MB, Litvinova SA, et al. Transcriptome of the krushinsky-molodkina audiogenic rat strain and identification of possible audiogenic epilepsy-associated genes. Frontiers in molecular neuroscience. 2021;14.

67. Lacar B, Linker SB, Jaeger BN, Krishnaswami SR, Barron JJ, Kelder MJ, et al. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nature communications. 2016;7:1–13.

68. Haigh JL, Adhikari A, Copping NA, Stradleigh T, Wade AA, Catta-Preta R, et al. Deletion of a non-canonical regulatory sequence causes loss of Scn1a expression and epileptic phenotypes in mice. Genome medicine. 2021;13:1–22.

69. Damasceno S, Souza Fonseca PA de, Rosse IC, Moraes MFD, De Oliveira JAC, Garcia-Cairasco N, et al. Putative causal variant on Vlgr1 for the epileptic phenotype in the model wistar audiogenic rat. Frontiers in Neurology. 2021;12.

70. Vasudevaraja V, Rodriguez JH, Pelorosso C, Zhu K, Buccoliero AM, Onozato M, et al. Somatic focal copy number gains of noncoding regions of receptor tyrosine kinase genes in treatment-resistant epilepsy. Journal of Neuropathology & Experimental Neurology. 2021;80:160–8.

71. Wlaź P, Socała K, Nieoczym D, Żarnowski T, Żarnowska I, Czuczwar SJ, et al. Acute anticonvulsant effects of capric acid in seizure tests in mice. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2015;57:110–6.

72. Jamali S, Bartolomei F, Robaglia-Schlupp A, Massacrier A, Peragut J-C, Régis J, et al. Large-scale expression study of human mesial temporal lobe epilepsy: Evidence for dysregulation of the neurotransmission and complement systems in the entorhinal cortex. Brain. 2006;129:625–41.

73. Mata A, Gil V, Pérez-Clausell J, Dasilva M, González-Calixto MC, Soriano E, et al. New functions of semaphorin 3E and its receptor PlexinD1 during developing and adult hippocampal formation. Scientific reports. 2018;8:1–16.

74. Singh S, Gupta A, Zech M, Sigafoos AN, Clark KJ, Dincer Y, et al. De novo variants of NR4A2 are associated with neurodevelopmental disorder and epilepsy. Genetics in Medicine. 2020;22:1413–7.

75. Hermey G, Plath N, Hübner CA, Kuhl D, Schaller HC, Hermans-Borgmeyer I. The three sorCS genes are differentially expressed and regulated by synaptic activity. Journal of neurochemistry. 2004;88:1470–6.

76. Bot AM, Dębski KJ, Lukasiuk K. Alterations in miRNA levels in the dentate gyrus in epileptic rats. PloS one. 2013;8:e76051.

77. Rawat C, Kutum R, Kukal S, Srivastava A, Dahiya UR, Kushwaha S, et al. Downregulation of peripheral PTGS2/COX-2 in response to valproate treatment in patients with epilepsy. Scientific reports. 2020;10:1–14.

78. Labonne JD, Driessen TM, Harris ME, Kong I-K, Brakta S, Theisen J, et al. Comparative genomic mapping implicates LRRK2 for intellectual disability and autism

at 12q12, and HDHD1, as well as PNPLA4, for x-linked intellectual disability at Xp22. 31. Journal of clinical medicine. 2020;9:274.

79. Taha AY, Huot PS, Reza-López S, Prayitno NR, Kang JX, Burnham WM, et al. Seizure resistance in fat-1 transgenic mice endogenously synthesizing high levels of omega-3 polyunsaturated fatty acids. Journal of neurochemistry. 2008;105:380–8.

80. Pensold D, Reichard J, Van Loo KM, Ciganok N, Hahn A, Bayer C, et al. DNA methylation-mediated modulation of endocytosis as potential mechanism for synaptic function regulation in murine inhibitory cortical interneurons. Cerebral Cortex. 2020;30:3921–37.

81. Blazer LL, Storaska AJ, Jutkiewicz EM, Turner EM, Calcagno M, Wade SM, et al. Selectivity and anti-parkinson's potential of thiadiazolidinone RGS4 inhibitors. ACS chemical neuroscience. 2015;6:911–9.

82. Yap SM, Smyth S. Ryanodine receptor 2 (RYR2) mutation: A potentially novel neurocardiac calcium channelopathy manifesting as primary generalised epilepsy. Seizure. 2019;67:11–4.

83. Sim SC, Ingelman-Sundberg M. Pharmacogenomic biomarkers: New tools in current and future drug therapy. Trends in pharmacological sciences. 2011;32:72–81.

84. Friedrich M, Grahnert A, Klein C, Tschöp K, Engeland K, Hauschildt S. Genomic organization and expression of the human mono-ADP-ribosyltransferase ART3 gene. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression. 2006;1759:270–80.

85. Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, et al. Exome sequencing and the management of neurometabolic disorders. New England Journal of Medicine. 2016;374:2246–55.

86. Burl RB, Clough S, Sendler E, Estill M, Krawetz SA. Sperm RNA elements as markers of health. Systems biology in reproductive medicine. 2018;64:25–38.

87. Hodgson JG, Yeh R-F, Ray A, Wang NJ, Smirnov I, Yu M, et al. Comparative analyses of gene copy number and mRNA expression in glioblastoma multiforme tumors and xenografts. Neuro-oncology. 2009;11:477–87.

88. Fuller TD, Westfall TA, Das T, Dawson DV, Slusarski DC. High-throughput behavioral assay to investigate seizure sensitivity in zebrafish implicates ZFHX3 in epilepsy. Journal of neurogenetics. 2018;32:92–105.

1 Tables

Table 1: Compendium of gene expression models processed for this study. Model GABRG2_KO was excluded from gene-level analysis as it showed no DEGs (See Methods).

Model		Species	Cases: Controls	Short description	Seizures
KAI	[33]	Rattus norvegicus	12:36	Kainaic acid model from days 1,3 and 10 after SE	Focal
PILO	[33]	Rattus norvegicus	12:36	Pilocarpine induced model from days 1,3 and 10 after SE	Focal
SCN8A_KI	[27]	Mus musculus	03:06	Knock-in model after 14 weeks when seizures were observed	Focal
SSSE	[33]	Rattus norvegicus	05:15	Self sustained status epilepticus from days 1,3 and 10 after SE	Focal
TBI_CX	[65]	Rattus norvegicus	05:05	TBI model from the cortex after 3 months of injury	Focal
TBI_TH	[65]	Rattus norvegicus	05:05	TBI model from the ipsilateral thalamus after 3 months of injury	Focal
TLE_EX	[18]	Homo sapiens	10:09	TLE surgery from the hippocampus	Focal
TLE_IN	[18]	Homo sapiens	10:09	TLE surgery from the hippocampus	Focal

AGGEN	[66]	Rattus norvegicus	04:07	Audiogenic epilepsy (AE) strains from the superior and inferior colliculi	Generalised
GABRA1_KO	[26]	Danio rerio	02:02	Knock-out of larvae from 5days post fertilization	Generalised
GABRG2_KO	[12]	Danio rerio	03:03	Knock-out of larvae after 5 days post fertilization	Generalised
PTZ	[67]	Mus musculus	46:36	PTZ induced model from the dentate gyrus	Generalised
SCN1A	[12]	Danio rerio	16:15	Knock-in model of larvae after 5 and 8 days post fertilization	Generalised
SCN1B_DEL	[68]	Mus musculus	02:06	Deletion of 1b non-coding interval of Scn1a	Generalised
WISTAR	[69]	Rattus norvegicus	02:02	Seizures following acoustic stimulation. Only DE data were available for analysis.	Generalised

1

Table 2: Gene counts before and after orthology mapping

Model	Original	Mapped	Reduction %
PILO	15577	12946	16.89
KAI	15577	12946	16.89
SSSE	15577	12946	16.89

TLE_EX	17625	17625	0.00
TLE_IN	14869	14869	0.00
SCN8A_KI	17877	15601	12.73
GABRA1_KO	9798	7977	18.59
GABRG2_KO	25290	13987	44.69
SCN1A	1873	881	52.96
PTZ	14536	11434	21.34
TBI_CX	15571	13927	10.56
TBI_TH	14462	13101	9.41
AGGEN	16281	14993	7.91
SCN1B_DEL	15426	14507	5.96
WISTAR	22398	14573	34.94

1

Table 3: Differentially expressed gene counts before and after orthology mapping

Model	Original	Mapped	Reduction %
PILO	4277	3703	13.42
KAI	3681	3184	13.50
SSSE	747	648	13.25
TLE_EX	8085	8085	0.00
TLE_IN	3588	3588	0.00
SCN8A_KI	4902	4656	5.02
GABRA1_KO	470	396	15.74
SCN1A	687	425	38.14
PTZ	35	16	54.29
TBI_CX	4303	4048	5.93
TBI_TH	1728	1636	5.32
AGGEN	177	158	10.73

SCN1B_DEL	32	33	-3.12
WISTAR	16	12	25.00

1

Table 4: MAGMA enrichment P-values for generalised models reveals GABRA1-KO and AGGEN are significant.

Model	Magma-P
AGGEN	0.004226
GABRA1_KO	0.003788
PTZ	0.926720
SCN1A	0.435920
SCN1B_DEL	0.882070
WISTAR	0.311720

2

Table 5: MAGMA enrichment p-values for focal models showing SCN8A_KI and TBI_TH as significant.

Model	Magma-P
KAI	0.320290
PILO	0.307990
SCN8A_KI	0.067946
SSSE	0.684810
TBI_CX	0.559890
TBI_TH	0.034528
TLE_EX	0.805360
TLE_IN	0.581980

3

Table 6: Significant genes after multiple testing correction (P < 0.05) from the generalised epilepsy models.

SET	GENE	NSNPS	ZSTAT	P	p_adjust
AGGEN	LSMEM2	3	4.953	1.114e-07	5.335e-05
AGGEN	LCN2	4	2.617	1.757e-03	4.949e-02
AGGEN	NFE2	10	3.125	2.884e-04	1.256e-02
AGGEN	PER1	9	4.003	7.751e-06	6.188e-04
AGGEN	VAMP2	2	3.687	3.471e-05	1.722e-03
AGGEN	SLC4A1	3	3.766	2.617e-05	1.567e-03
GABRA1_KO	RPL5	7	3.045	3.836e-04	1.413e-02
GABRA1_KO	CSPG5	11	3.672	3.594e-05	1.722e-03
GABRA1_KO	CACNA2D2	184	4.021	6.559e-06	6.188e-04
GABRA1_KO	PCDH7	795	4.203	7.801e-07	1.246e-04
GABRA1_KO	RPL34	12	3.754	2.214e-05	1.515e-03
GABRA1_KO	HDAC9	1,240	2.314	1.721e-03	4.949e-02
GABRA1_KO	ZMIZ2	18	4.014	6.212e-06	6.188e-04
GABRA1_KO	MAGI2	3,469	2.395	8.906e-04	3.047e-02
GABRA1_KO	KDM2B	18	2.610	1.541e-03	4.922e-02
GABRA1_KO	NOVA1	126	3.054	3.178e-04	1.269e-02
GABRA1_KO	CTC1	40	4.629	3.155e-07	7.556e-05

1

Table 7: Common differentially expressed genes among focal epilepsy models.

Gene symbol	Description	Disease context	References
EGFR	Epidermal growth factor receptor	Treatment resistant epilepsies	[70]
CA10	Capric acid 10	Anticonvulusant properties adds to clinincal efficacy of refactory epilepsy	[71]
IGFBP5	Insulin like growth factor binding protein 5	Mesial temporal lobe epilepsy	[72]
SEMA3E	Semaphorin 3E	Involved in hippocampal formation	[73]
ACVR1C	Activin A receptor type 1C	Deletion observed in epilepsy patients	[74]
SORCS1	Sortilin related VPS10 domain containing receptor 1	Temporal lobe epilepsy	[75]
TGFBR1	TGF-beta receptor type-1	Temporal lobe epilepsy	[49]
CBFB	Core-binding factor subunit beta	Transcription regulation in dentate gyrus	[76]
DNM3	Dynamin-3	Epileptic encephalopathy	[41]
PTGS2	Prostaglandin-Endoperoxide synthase 2	Downregulated in epilepsy patients following treatment	[77]
SLC2A13	Solute carrier family 2 member 13	Involved in intellectual disability	[78]

FAT1	FAT atypical cadherin 1	Involved in induced epilepsy seizure models	[79]
WIPF1	WAS/WASL interacting protein family member 1	Negative correlation of methylation and expression level in TLE patients	[80]
RGS8	Regulator of G protein signalling 8	Dysregulation in Parkinson's disease patients	[81]
SLIT2	Slit guidance ligand 2	Temporal lobe epilepsy	[42]
RYR2	Ryanodine receptor 2	Mutation leads to genetic generalized epilepsy	[82]
CYP4F2	Cytochrome P450 family 4 subfamily F member 2	Biomarker for central nervous system disorders	[83]
PLXDC2	Plexin domain-containing protein 2	Mesial temporal lobe epilepsy	[50]
ART3	ADP-Ribosyltransferase 3	Expressed in heart and skeletal muscles	[84]
MYLK	Myosin light chain kinase	De-novo mutation leads to intellectual development disorder	[85]
CABYR	Calcium binding tyrosine phosphorylation regulated	Reproductive health biomarker	[86]
KCNA2	Potassium voltage-gated channel subfamily A member 2	Early onset epileptic encephalopathy	[8]
SLC35E3	Solute carrier family 35 member E3	Glioblastoma multiforme tumors(GBMs)	[87]

LRRC8B	Volume-regulated anion channel subunit	gliotransmission	[48]
ZFHX3	Zinc finger homeobox 3	Neurodevelopment and epileptic encephalopathy	[88]

Additional files

Supplementary Figures

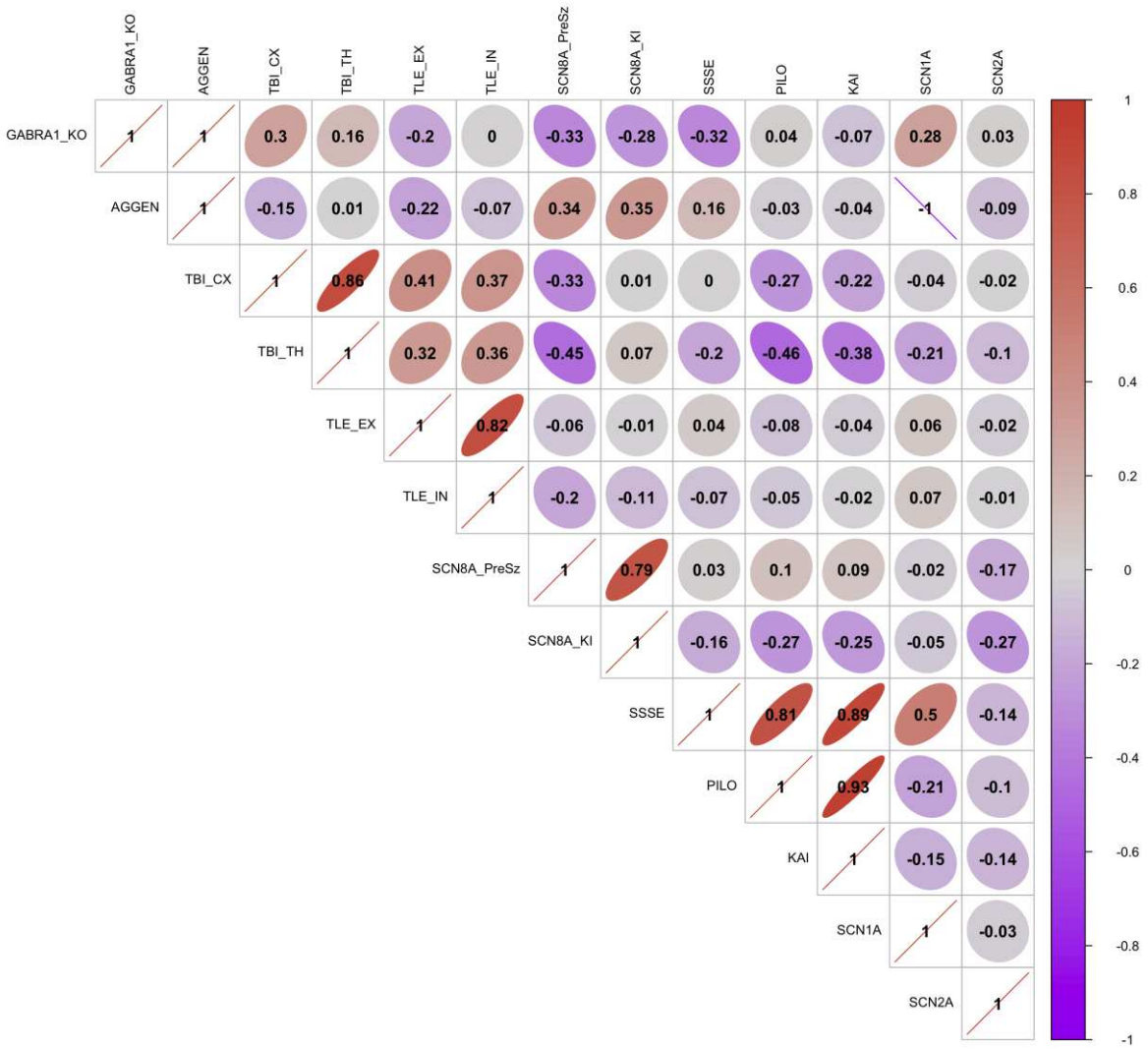


Figure S1: Correlation plot on DEGs

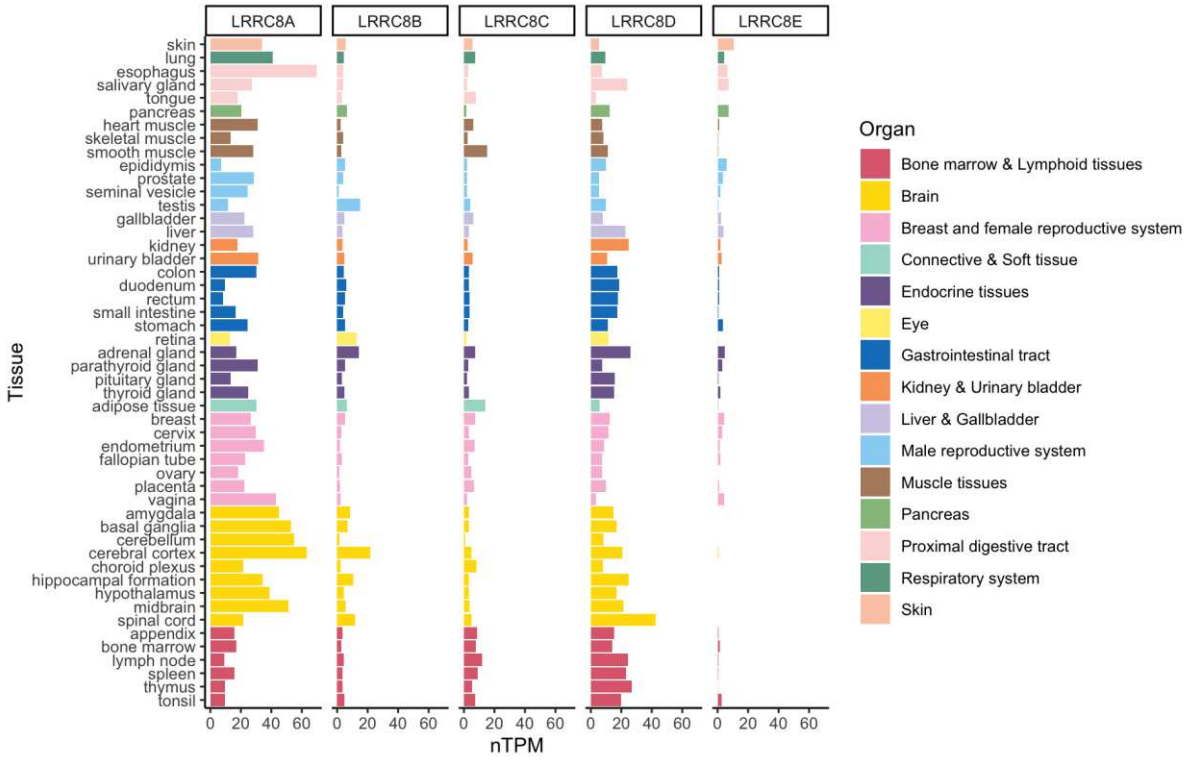
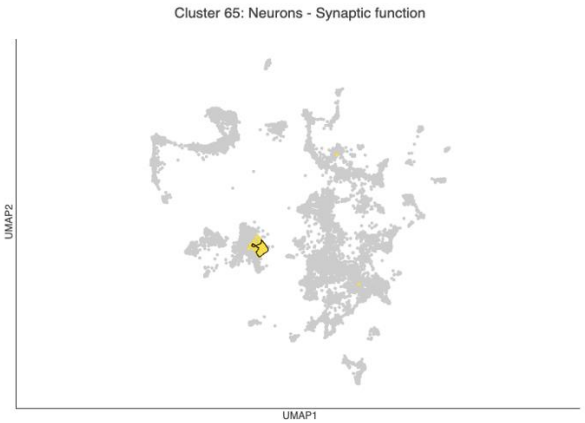


Figure S2: Tissue-specific mRNA expression profiles of LRRC8 gene family members (LRRC8A-E). Data represents consensus normalized expression (nTPM) derived from HPA, GTEx datasets, obtained from the Human Protein Atlas (v24). Bars represent nTPM values for each gene in the indicated tissue. Tissues are listed on the y-axis, and individual LRRC8 genes are shown in separate columns. Colors indicate the organ system category as defined by the HPA legend.

1



15 nearest neighbours based on single cell RNA expression

Neighbour	Description	Correlation	Cluster
MARCHF4	Membrane associated ring-CH-type finger 4	0.7655	65
RIMBP2	RIMS binding protein 2	0.7546	65
ENOX1	Ecto-NOX disulfide-thiol exchanger 1	0.7427	8
PRMT8	Protein arginine methyltransferase 8	0.7417	65
GABRA4	Gamma-aminobutyric acid type A receptor subunit alpha4	0.7374	65
SYN2	Synapsin II	0.7205	65
GABBR2	Gamma-aminobutyric acid type B receptor subunit 2	0.7154	65
GRIA1	Glutamate ionotropic receptor AMPA type subunit 1	0.7139	65
KIRREL3	Kirre like nephrin family adhesion molecule 3	0.7131	65
ADAM22	ADAM metallopeptidase domain 22	0.7112	65
TMEM151B	Transmembrane protein 151B	0.7107	65
CNTN5	Contactin 5	0.7104	65
KCNC2	Potassium voltage-gated channel subfamily C member 2	0.7099	65
GRIN2B	Glutamate ionotropic receptor NMDA type subunit 2B	0.7098	65
PRKCE	Protein kinase C epsilon	0.7088	65

Figure S3: Single-cell type clustering reveals LRRC8B significantly enriched for Synaptic functions. The RNA data was used to cluster genes according to their expression across single cell types. The clustering of 19203 genes expressed in single

cell types resulted in 80 expression clusters, which have been manually annotated to describe common features in terms of function and specificity. The result of the cluster analysis is presented as a UMAP based on gene expression, where each cluster has been summarized as colored areas containing most of the cluster genes. The picture shows the Synaptic function cluster where LRRC8B gene is significantly enriched (left) jointly with the table listing the 15 most similar genes in terms of expression profile (right).

1