Rare gene deletions in genetic generalized and Rolandic epilepsies

Jabbari, Kamel

2018-08-27


http://hdl.handle.net/10138/247098
https://doi.org/10.1371/journal.pone.0202022

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
RESEARCH ARTICLE

Rare gene deletions in genetic generalized and Rolandic epilepsies

Kamel Jabbari1,2,4, Dheeraj R. Bobbili3,4, Dennis Lal1,4,5,6, Eva M. Reinthaler7, Julian Schubert6, Stefan Wolking6, Vishal Sinha9, Susanne Motameny1, Holger Thiele1, Amit Kawalia1, Janine Altmüller1,10, Mohammad Reza Toliat1, Robert Kraaij11, Jeroen van Rooij11, André G. Uitterlinden11, M. Arfan Ikram12, EuroEPINOMICS CoGIE Consortium8, Federico Zara13, Anna-Elena Lehesjoki14,15, Roland Krause3, Fritz Zimprich7, Thomas Sander1, Bernd A. Neubauer16, Patrick May9, Holger Lerche6, Peter Nürnberg1,17,18,*

1 Cologne Center for Genomics, University of Cologne, Cologne, Germany, 2 Cologne Biocenter, Institute for Genetics, University of Cologne, Cologne, Germany, 3 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg, 4 Psychiatric and Neurodevelopmental Genes Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, 5 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, 6 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, 7 Department of Neurology, Medical University of Vienna, Vienna, Austria, 8 Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany, 9 Institute for Molecular Medicine FIMM, University of Helsinki, Helsinki, Finland, 10 Institute of Human Genetics, University of Cologne, Cologne, Germany, 11 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands, 12 Departments of Epidemiology, Neurology, and Radiology, Erasmus Medical Center, Rotterdam, The Netherlands, 13 Laboratory of Neurogenetics and Neuroscience, Institute G. Gaslini, Genova, Italy, 14 Folkhälsoan Institute of Genetics, Folkhälsoan Research Center, Helsinki, Finland, 15 Neuroscience Center and Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland, 16 Department of Neuropediatrics, Medical Faculty University Giessen, Giessen, Germany, 17 Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany, 18 Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany.

These authors contributed equally to this work. *Membership of the EuroEPINOMICS CoGIE Consortium is provided in the Acknowledgments. nuernberg@uni-koeln.de

Received: March 16, 2018

Accepted: July 26, 2018

Published: August 27, 2018

Copyright: © 2018 Jabbari et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: H.L., P.N., T.S., F.Z. received grants by the EuroEPINOMICS programme (DFG grant numbers: HL: LE1030/11-1, PN: NU50/8-1, TS: SA434/5-1, FWF grant number: FZ: 1643-B09) within the EUROCORES framework of the European Science Foundation (ESF). D.R.B. and P.M. were supported by the JPND Courage-PD research grant. P.M. also from the NCER FNR grant. H.L., P.N., T.S., and R.K. received further support from the NPdenovo database.

Abstract

Genetic Generalized Epilepsy (GGE) and benign epilepsy with centro-temporal spikes or Rolandic Epilepsy (RE) are common forms of genetic epilepsies. Rare copy number variants have been recognized as important risk factors in brain disorders. We performed a systematic survey of rare deletions affecting protein-coding genes derived from exome data of patients with common forms of genetic epilepsies. We analysed exomes from 390 European patients (196 GGE and 194 RE) and 572 population controls to identify low-frequency genic deletions. We found that 75 (32 GGE and 43 RE) patients out of 390, i.e. ~19%, carried rare genic deletions. In particular, large deletions (>400 kb) represent a higher burden in both GGE and RE syndromes as compared to controls. The detected low-frequency deletions (1) share genes with brain-expressed exons that are under negative selection, (2) overlap with known autism and epilepsy-associated candidate genes, (3) are enriched for CNV intolerant genes recorded by the Exome Aggregation Consortium (ExAC) and (4) coincide with likely disruptive de novo mutations from the NPdenovo database. Employing several knowledge...
Introduction

Epilepsies are among the most widespread neurological disorders with a lifetime incidence of ~3% [1]. They represent a heterogeneous group of different disease entities that, with regard to aetiology, can be roughly divided in epilepsies with an exogeneous/symptomatic cause and those with a genetic cause. Genetic generalized epilepsies (GGE; formerly idiopathic generalized epilepsies) are the most common genetic epilepsies accounting for 30% of all epilepsies. They comprise syndromes such as juvenile myoclonic epilepsy, childhood absence epilepsy and juvenile absence epilepsy. In general, they tend to take a benign course and show a good response to pharmacotherapy. Among focal genetic epilepsies, benign epilepsy with centro-temporal spikes or Rolandic epilepsy (RE) is the most common form. RE has its onset in childhood or early adolescence and usually tapers off around the age of 15.

High-throughput genomic studies raised the number of epilepsy-associated candidate genes to hundreds; nowadays, frequently mutated ones are included in diagnostic gene panels (for recent reviews see [2,3]). Large consortia initiatives such as Epi4k [4] enrolled 1,500 families, in which two or more affected members displayed epilepsy, as well as 750 individuals, including 264 trios, with epileptic encephalopathies and infantile spasms, Lennox-Gastaut syndrome, polymicrogyria or periventricular heterotopias. In addition to the detection of known and unknown risk factors, the consortium found a significant overlap between the gene network of their epilepsy candidate genes and the gene networks for autism spectrum disorder (ASD) and intellectual disability. Intriguingly, epilepsy is the medical condition most highly associated with genetic autism syndromes [5].

Genomic disorders associated with copy number variations (CNVs) appear to be highly penetrant, occur on different haplotype backgrounds in multiple unrelated individuals and seem to be under strong negative selection [6–8]. A number of chromosomal locations suspected to contribute to epilepsy have been identified [9–11][12,13].

A genome-wide screen for CNVs using array comparative genomic hybridization (aCGH) in patients with neurological abnormalities and epilepsy led to the identification of recurrent microdeletions on 6q22 and 1q22.31 [14]. A deletion on 15q13.3 belongs to the most frequent recurrent microdeletions in epilepsy patients; it is associated with intellectual disability, autism, schizophrenia, and epilepsy [15,16]. The recurrence of some CNVs seems to be triggered by the genome structure, namely by the chromosomal distribution of interspersed repetitive sequences (like Alu transposons) or recently duplicated genome segments (large blocks of sequences >10 kbp with >95% sequence identity that constitute five to six percent of the genome) that give rise to nonallelic homologous recombination [6,17].

CNV screening in large samples showed that 34% of heterozygous deletions affect genes associated with recessive diseases [18]. CNVs are thought to account for a major proportion of human genetic variation and have an important role in genetic susceptibility to common disease, in particular neuropsychiatric disorders [19]. Genome-wide surveys have demonstrated that rare CNVs altering genes in neuro-developmental pathways are implicated in epilepsy, autism spectrum disorder and schizophrenia [3,20].

Considering all types of CNVs across two analysed cohorts, the total burden was not significantly different between subjects with epilepsy and subjects without neurological disease [21];
however, when considering only genomic deletions affecting at least one gene, the burden was significantly higher in patients. Likewise, using Affymetrix SNP 6.0 array data, it has recently been shown that there is an increased burden of rare large deletions in GGE [13]. The drawback of the latter approach is that smaller CNVs cannot be detected. Systematic searches of CNVs in epilepsy cohorts using whole-exome sequencing (WES) data, which provides the advantage to identify smaller deletions along with the larger ones, are still missing.

In the present study, we provide the CNV results of the largest WES epilepsy cohort reported so far. We aimed at (1) identifying the genome-wide burden of large deletions (>400kb), (2) studying the enrichment for deletions of brain-expressed exons, in particular those under negative selection, (3) detecting deletions that overlap with previously defined autism and epilepsy candidate genes, and (4) browsing knowledge databases to help understand the disease aetiology.

Materials and methods

Patient cohorts

All patients or their representatives, if participants were under age 18, and included relatives, gave their informed consent to this study. All procedures were in accordance with the Helsinki declaration and approved by the local ethics committees/internal review boards of the participating centers. The leading institution was the Ethics Commission of the University and the University Clinic of Tübingen.

GGE cohort: This cohort included 196 subjects with genetic generalized epilepsy. All subjects were of European descent (Italian 81, German 54, Finnish 22, Dutch 11, British 9, Danish 8, Turkish 6, Swedish 3, French 1, Greek 1). The cohort included 117 female subjects (60%). The GGE-diagnoses were childhood absence epilepsy (CAE; n = 94), juvenile absence epilepsy (JAE; 21), juvenile myoclonic epilepsy (JME; 47), genetic generalized epilepsy with generalized tonic-clonic seizures (EGTCS, 27), early-onset absence epilepsy (EOAE, 4), epilepsy with myoclonic absences (EMA, 1), and unclassified GGE (2). Age of epilepsy onset ranged from 1 year to 38 years with a median of 8 years. The majority of subjects derived from multiplex families with at least 2 affected family members (n = 183), thereof 90 families with 3 or more affected members.

RE cohort: This cohort included 204 unrelated Rolandic patients of European ancestry which were recruited from centers in Austria (n = 107), Germany (n = 84), and Canada (n = 13).

Control cohort: We used 445 females and 283 males (728 in total) from the Rotterdam Study as population control subjects [22]. The same cohort was recently used for the screening of 18 GABA<sub>A</sub>-receptor genes in RE and related syndromes [23].

Workflow for CNV detection

Our primary analysis workflow included three major steps as shown in Fig 1. These are 1) data pre-processing, 2) SNV/INDEL analysis and 3) copy number variant analysis.

Data pre-processing: Sequencing adapters were removed from the FASTQ files with cutadapt [24] and sickle [25]. GATK best practices were followed for the next steps of data pre-processing and variant calling [26]. Alignment to the GRCh37 human reference genome was performed using BWA-MEM [27] with default parameters. Conversion of SAM to BAM files was done with SAMtools [28]. Sorting of BAM files, marking of duplicate reads due to PCR amplification and addition of read group information were done using Picard (https://github.com/broadinstitute/picard) tools with default parameters. Base quality score recalibration and local realignment for INDELs was performed using GATK version 3.2.
Fig 1. Flowchart of the analysis steps. Parameters used in each step are described in detail in the methods section.

https://doi.org/10.1371/journal.pone.0202022.g001
Coverage: Mean depth of coverage and target coverage of exons were calculated from the BAM files using the depth of coverage tool from GATK. The same files were also used as input for calling of CNVs.

Variant calling: The GATK haplotype caller (version 3.2) was chosen to perform multiple sample variant calling and genotyping with default parameters. To include splice site variants in the flanking regions of the exons, exonic intervals were extended by 100 bp each upstream and downstream. Multiple sample calling is advantageous in deciding whether a variant can be identified confidently as it provides the genotype for every sample. It allows filtering variants based on the rate of missing genotypes across all samples and also according to the individual genotype.

Sample QC: Samples were excluded from the analysis based on the following criteria: 1) Samples with a mean depth $< 30x$ or $< 70\%$ of exon targets covered at $< 20x$ were excluded from further analysis; 2) samples with $> 3$ standard deviations from mean in number of alternate alleles, number of heterozygotes, transition/transversion ratio, number of singletons and call rate as calculated with the PLINK/SEQ i-stats tool (https://atgu.mgh.harvard.edu/plinkseq); 3) call rate $< 97\%$; 4) ethnically unmatched samples as identified by multi-dimensional scaling analysis with PLINK version 1.9 [29]; 5) PI_HAT score $> 0.25$ as calculated by PLINK version 1.9 to exclude related individuals.

Variant QC: Initial filtering of variants was performed based on quality metrics over all the samples with the following parameters for VQSR: Tranches chosen, VQSRTrancheSNV99. 90to100.00. QC over all samples (INFO column) was done as follows: a) for SNVs, variants were filtered for QD $< 2.0$, FS $> 60.0$, MQ $< 40.0$, MQRankSum $< -12.5$, ReadPosRankSum $< -8.0$, DP $< 10.0$, GQ MEAN $< 20.0$, VQSLOD $< 0$, more than 5\% missingness, ABHet $> 0.75$ or $< 0.25$ and deviation from Hardy-Weinberg equilibrium (Phred scale p-value of $> 20$); b) for INDELs, the same was done as for SNVs except for the following parameters for variant filtration: QD $< 2.0$, FS $> 200.0$, ReadPosRankSum $< -20.0$, DP $< 10.0$, GQ MEAN $< 20.0$, missingness $< 5\%$, Hardy-Weinberg Phred scale value of $> 20$, VQSLOD $> 0$.

To further exclude low quality variants, we also applied filtering based on quality metrics for each genotype using read depth and quality of individual genotypes. Genotypes with a read depth of $< 10$ and GQ of $< 20$ were converted to missing by using BCFtools [28]. Multi-allelic variants were decomposed using variant-tests [30] and left-normalized using BCFtools [28].

Variant annotation: Variants were annotated with ANNOVAR [31] version 2015, Mar22 using RefSeq and Ensembl versions 20150322 and the dbNSFP [32] version 2.6 annotations including nine scores for missense mutations (SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaSVM, MetaLR), the CADD score, and three conservation-based scores from GERP++, PhyloP and SiPhy. Splicing variants were defined to include 2 bp before and after the exon boundary position. To obtain rare variants, we filtered the variants for a minor allele frequency (MAF) of $< 0.005$ in public databases such as 1000 genomes [33], dbSNP [34], ExAC (release 0.3) and the exome variant server (EVS). We defined deleterious variants as those variants that fulfil any of the following three criteria: 1) all the variants except the synonymous variants predicted to be deleterious by at least 5 out of 8 missense prediction scores, CADD score $> 4.5$, or 2 out of 3 conservation scores (GERP$> 3$, PhyloP$> 0.95$, SiPhy$> 10$) show high conservation; 2) variants annotated as "splicing", "stop gain" or "stop loss"; 3) any insertion or deletion.

CNV detection: In the remaining high quality samples, CNVs were detected by using XHMM as described in [35]. In the current study, we focused only on deletions, as the false positive rate for duplications is too high to allow for meaningful interpretation. CNV calls
were annotated using bedtools version 2.5 [36]. NCBI RefSeq (hg19, 20150322) was used to identify the genes that lie within the deletion boundaries.

CNV filtering: The detected deletions were filtered based on the following criteria: 1) Z score < -3, given by XHMM; 2) Q_SOME score ≥ 60, given by XHMM.

**Burden analysis of large and rare deletions**

Excess deletion rate of the large deletions (length >400 kb) in subjects with epilepsy compared to the controls was measured as described in [13] using PLINK version 1.9 [29]. We set the overlap fraction to 0.7 (70%) and the internal allele frequency cut-off <0.5% and evaluated the significance empirically by 10,000 case-control label permutations.

**Case-only CNVs**

The CNVs that are unique for cases (not present in any of the in-house controls) and occur at a low frequency, i.e., present in ≤2 independent cases, while having a frequency of ≤1% in the CNVmap, the DGV gold standard dataset [37] and 1000 genomes SV [38] were selected and subjected to further analysis as described below.

**Validation of CNVs**

We proceeded by visual inspection of depth variation across exons of the filtered deletions; we also performed qPCR validations of three small deletions, two of which, NCAPD2 and CAPN1, stood the filtering procedure (see Table 1). For RE patients, genomic DNA samples were analysed using the Illumina OmniExpress Beadchip (Illumina, San Diego, CA, USA) [13]. Twenty-three of 60 CNVs present in the RE patients were validated by available array data (S1 Table). Generally, small CNVs cannot be reliably identified with SNP arrays [39]. Indeed, of the 37 CNVs that were not identified in the beadchip data, 23 have a size of <10 kb, whereas only 2 of the 23 validated CNVs have a size of less than 10 kb according to the array data.

**Compound heterozygous mutations and protein-protein interactions**

We checked for concurrence of a deletion in one allele and a deleterious variant in the second allele. We included the first order interacting partners from the protein-protein interaction network (PPIN) in this analysis [40] and assessed if any gene or its first order interacting partner carries a deletion in one allele and a deleterious variant in the other. We excluded all genes that had no HGNC (HUGO Gene Nomenclature Committee) entry resulting in a network of 13,364 genes and 140,902 interactions. This network was then further filtered for interactions likely to occur in brain tissues using a curated data set of brain-expressed genes [41]. The final brain-specific PPIN consisted of 10,469 genes and 114,533 interactions.

**Gene-set enrichment analysis**

Genes that were expressed in brain [42] and located within deletion boundaries were used as input for an enrichment analysis using the Ingenuity Pathway Analysyer (IPA®) [43]. We performed the enrichment analysis with all deleted genes from the RE and GGE samples together as well as for each phenotype separately.

**Over-representation analysis**

To assess whether the deleted set of genes were enriched in known epilepsy-associated genes, we retrieved genes that were associated with the disease term “epilepsy” from the DisGeNET database [44]. Then we compared the overlap between the brain-expressed genes that are
### Table 1. Epilepsy associated microdeletions.

<table>
<thead>
<tr>
<th>Type</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Z-score</th>
<th>Length</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE</td>
<td>1</td>
<td>32671406</td>
<td>32673183</td>
<td>-3.78</td>
<td>1777</td>
<td>IQCC</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>43296070</td>
<td>43317481</td>
<td>-6.52</td>
<td>2141</td>
<td>ERMAP, ZNF691</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>53320120</td>
<td>53329849</td>
<td>-3.94</td>
<td>9729</td>
<td>ZYG11A</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>55586222</td>
<td>55591447</td>
<td>-4.46</td>
<td>5225</td>
<td>USP24</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>115137047</td>
<td>11516853</td>
<td>-3.13</td>
<td>31483</td>
<td>DENTD2C</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>150252003</td>
<td>15025922</td>
<td>-3.3</td>
<td>7249</td>
<td>Clorf54, CIART</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>153658548</td>
<td>15366204</td>
<td>-3.44</td>
<td>3499</td>
<td>NPR1</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>160061571</td>
<td>16006497</td>
<td>-4.46</td>
<td>3426</td>
<td>IGSF8</td>
</tr>
<tr>
<td>RE</td>
<td>1</td>
<td>249144392</td>
<td>24921259</td>
<td>-3.25</td>
<td>68199</td>
<td>PGBD2, ZNF692</td>
</tr>
<tr>
<td>RE</td>
<td>2</td>
<td>44502637</td>
<td>44539912</td>
<td>-4.48</td>
<td>37275</td>
<td>SLC3A1</td>
</tr>
<tr>
<td>RE</td>
<td>3</td>
<td>44037776</td>
<td>4562816</td>
<td>-3.79</td>
<td>159040</td>
<td>ITPR1, ITPR1-AS1, SUMF1</td>
</tr>
<tr>
<td>RE</td>
<td>4</td>
<td>169362457</td>
<td>16939393</td>
<td>-3.01</td>
<td>31473</td>
<td>DDX60L</td>
</tr>
<tr>
<td>RE</td>
<td>5</td>
<td>71519462</td>
<td>71533975</td>
<td>-5.36</td>
<td>14513</td>
<td>MRPS27</td>
</tr>
<tr>
<td>RE</td>
<td>5</td>
<td>75858199</td>
<td>75914495</td>
<td>-3.32</td>
<td>56296</td>
<td>COL10A1, NTSDC1</td>
</tr>
<tr>
<td>RE</td>
<td>6</td>
<td>74197233</td>
<td>74212576</td>
<td>-3.34</td>
<td>15343</td>
<td>GT2IRD2, NCF1</td>
</tr>
<tr>
<td>RE</td>
<td>6</td>
<td>100146395</td>
<td>10015339</td>
<td>-4.98</td>
<td>6998</td>
<td>AGFG2</td>
</tr>
<tr>
<td>RE</td>
<td>7</td>
<td>82571539</td>
<td>82752251</td>
<td>-3.13</td>
<td>180712</td>
<td>CHMP4C, IMPA1, SLC10A5, SNX16, ZFAND1</td>
</tr>
<tr>
<td>RE</td>
<td>8</td>
<td>146028239</td>
<td>146033207</td>
<td>-4.62</td>
<td>4968</td>
<td>ZNF517</td>
</tr>
<tr>
<td>RE</td>
<td>9</td>
<td>35800178</td>
<td>35801935</td>
<td>-4.58</td>
<td>1757</td>
<td>NPF2</td>
</tr>
<tr>
<td>RE</td>
<td>9</td>
<td>140234513</td>
<td>140250835</td>
<td>-3.24</td>
<td>7322</td>
<td>EXD3</td>
</tr>
<tr>
<td>RE</td>
<td>10</td>
<td>5920045</td>
<td>5926074</td>
<td>-5.17</td>
<td>6029</td>
<td>ANKRD16</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>49383834</td>
<td>49420140</td>
<td>-4.44</td>
<td>36306</td>
<td>FRMPD2</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>2549103</td>
<td>2606577</td>
<td>-3.38</td>
<td>57474</td>
<td>KCNQ1</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>4930992</td>
<td>4929495</td>
<td>-4.46</td>
<td>26403</td>
<td>OR5A1, OR5I1</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>7727796</td>
<td>7781510</td>
<td>-3.74</td>
<td>90714</td>
<td>OR5P2, OVCH2</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>17533429</td>
<td>17546119</td>
<td>-3.17</td>
<td>12690</td>
<td>USH1C</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>47600393</td>
<td>47608426</td>
<td>-5.11</td>
<td>8033</td>
<td>FAM180B, KBTBD4, NDUFS5</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>59189760</td>
<td>59211596</td>
<td>-3.31</td>
<td>21836</td>
<td>OR5A1, OR5A2</td>
</tr>
<tr>
<td>RE</td>
<td>11</td>
<td>64977256</td>
<td>64981526</td>
<td>-6.72</td>
<td>4270</td>
<td>CAPN1, SLC22A20</td>
</tr>
<tr>
<td>RE</td>
<td>12</td>
<td>6625993</td>
<td>6627159</td>
<td>-5.01</td>
<td>1166</td>
<td>NCAPD2</td>
</tr>
<tr>
<td>RE</td>
<td>12</td>
<td>130922883</td>
<td>130927235</td>
<td>-3.85</td>
<td>4352</td>
<td>RIMBP2</td>
</tr>
<tr>
<td>RE</td>
<td>14</td>
<td>54863694</td>
<td>55907289</td>
<td>-3.32</td>
<td>1043595</td>
<td>ATG14, CDKN3, CGRRF1, CNHI1, DLGAP5, FBXO34, GCH1, GMFB, LGALS3, MAPK1IP1L, MIR4308, SAMD4A, SOCS4, TBPL2, WDHD1</td>
</tr>
<tr>
<td>RE</td>
<td>14</td>
<td>77302503</td>
<td>77327178</td>
<td>-3.27</td>
<td>24675</td>
<td>LRRC74A</td>
</tr>
<tr>
<td>RE</td>
<td>14</td>
<td>92900208</td>
<td>92920437</td>
<td>-4.01</td>
<td>20229</td>
<td>SLC24A4</td>
</tr>
<tr>
<td>RE</td>
<td>15</td>
<td>23811123</td>
<td>28525396</td>
<td>-4.21</td>
<td>4714273</td>
<td>ATP10A, GABRA5, GABRB3, GABRG3, GABRG3-AS1, HERC2, IPW, LINC00929, LOC100128714, MAGEL2, MIR4715, MKRN3, NDN, NPA1, OCA2, PWAR1, PWARS, PWARSN, PWRN1, PWRN2, PWRN3, PWRN4, SNORD107, SNORD108, SNORD109A, SNORD109B, SNORD115-1, SNORD115-10, SNORD115-11, SNORD115-12, SNORD115-13</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Type</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Z-score</th>
<th>Length</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE</td>
<td>15</td>
<td>29346087</td>
<td>32460550</td>
<td>-4.57</td>
<td>3114463</td>
<td>APBA2, ARHGPAP11B, FAM7A, NA7, DKFZP434L87, FAM189A1, FAN1, GOLGA8H, GOLGA8J, GOLGA8R, GOLGA8T, HRC2P10, KLF13, LOC100288687, LOC283710, MR211, MTMR10, NDNL2, OTUD7A, TJP1, TRPM1, ULK4P1, ULK4P2, ULK4P3</td>
</tr>
<tr>
<td>RE</td>
<td>16</td>
<td>9856958</td>
<td>10032248</td>
<td>-5.26</td>
<td>175290</td>
<td>GRIN2A</td>
</tr>
<tr>
<td>RE</td>
<td>16</td>
<td>70564998</td>
<td>70573138</td>
<td>-4.22</td>
<td>12640</td>
<td>SF3B3, SNORD111, SNORD111B</td>
</tr>
<tr>
<td>RE</td>
<td>17</td>
<td>7010272</td>
<td>7017572</td>
<td>-3.43</td>
<td>7300</td>
<td>ADPRM, MAGOH2P, MYH1, MYH2, MYH3, MYHAS, SCO1, TMEM220</td>
</tr>
<tr>
<td>RE</td>
<td>17</td>
<td>38346658</td>
<td>38350074</td>
<td>-4.63</td>
<td>3416</td>
<td>MIR6867, RAPGEFL1</td>
</tr>
<tr>
<td>RE</td>
<td>17</td>
<td>73623470</td>
<td>73661285</td>
<td>-4.39</td>
<td>37815</td>
<td>SF3B3, SNORD111, SNORD111B</td>
</tr>
<tr>
<td>RE</td>
<td>17</td>
<td>76967650</td>
<td>76970921</td>
<td>-4.45</td>
<td>3271</td>
<td>LGALS3BP, SMIM5, SMIM6</td>
</tr>
<tr>
<td>RE</td>
<td>18</td>
<td>30873076</td>
<td>30928981</td>
<td>-3.86</td>
<td>55905</td>
<td>CCDC178</td>
</tr>
<tr>
<td>RE</td>
<td>19</td>
<td>9014087</td>
<td>9054377</td>
<td>-4.25</td>
<td>40290</td>
<td>MIR211, MIR212, NDUFB7, TECR</td>
</tr>
<tr>
<td>RE</td>
<td>19</td>
<td>14673265</td>
<td>14677799</td>
<td>-4.49</td>
<td>4514</td>
<td>MUC16</td>
</tr>
<tr>
<td>RE</td>
<td>19</td>
<td>14854191</td>
<td>14884892</td>
<td>-3.24</td>
<td>30701</td>
<td>ADGRE2</td>
</tr>
<tr>
<td>RE</td>
<td>19</td>
<td>35862216</td>
<td>35941102</td>
<td>-3.7</td>
<td>78886</td>
<td>FFAR2, LINC0153</td>
</tr>
<tr>
<td>RE</td>
<td>19</td>
<td>45447959</td>
<td>45463635</td>
<td>-5.89</td>
<td>17406</td>
<td>SF3B3, SNORD111, SNORD111B</td>
</tr>
<tr>
<td>RE</td>
<td>20</td>
<td>51175236</td>
<td>51192575</td>
<td>-3.33</td>
<td>17339</td>
<td>SHANK1</td>
</tr>
<tr>
<td>RE</td>
<td>20</td>
<td>52271871</td>
<td>52327971</td>
<td>-3.68</td>
<td>56100</td>
<td>FPR2, FPR3</td>
</tr>
<tr>
<td>GGE</td>
<td>1</td>
<td>76779478</td>
<td>77094515</td>
<td>-4.34</td>
<td>315037</td>
<td>ST6GALNAC3</td>
</tr>
<tr>
<td>GGE</td>
<td>1</td>
<td>16951023</td>
<td>16951164</td>
<td>-4.9</td>
<td>1407</td>
<td>F5</td>
</tr>
<tr>
<td>GGE</td>
<td>2</td>
<td>16685248</td>
<td>16687227</td>
<td>-3.02</td>
<td>19792</td>
<td>LOC1002724058, SCN1A</td>
</tr>
<tr>
<td>GGE</td>
<td>3</td>
<td>10331397</td>
<td>10359151</td>
<td>-4.04</td>
<td>4518</td>
<td>GHR1, GHRLOS</td>
</tr>
<tr>
<td>GGE</td>
<td>5</td>
<td>21751815</td>
<td>21854929</td>
<td>-6.45</td>
<td>10314</td>
<td>CDH12</td>
</tr>
<tr>
<td>GGE</td>
<td>6</td>
<td>43320067</td>
<td>43323250</td>
<td>-3.68</td>
<td>3183</td>
<td>ZNF318</td>
</tr>
<tr>
<td>GGE</td>
<td>7</td>
<td>13971097</td>
<td>14028735</td>
<td>-3.88</td>
<td>57638</td>
<td>ETV1</td>
</tr>
<tr>
<td>GGE</td>
<td>7</td>
<td>12165116</td>
<td>12165265</td>
<td>-3.43</td>
<td>1524</td>
<td>PTPRZ1</td>
</tr>
<tr>
<td>GGE</td>
<td>7</td>
<td>14671346</td>
<td>14682961</td>
<td>-4.41</td>
<td>358269</td>
<td>CNTNAP2, LOC101928700</td>
</tr>
<tr>
<td>GGE</td>
<td>7</td>
<td>50501839</td>
<td>50558285</td>
<td>-3.04</td>
<td>56446</td>
<td>AOC1, TMEM176</td>
</tr>
<tr>
<td>GGE</td>
<td>8</td>
<td>2944572</td>
<td>3045513</td>
<td>-6.06</td>
<td>100941</td>
<td>CSMD1</td>
</tr>
<tr>
<td>GGE</td>
<td>8</td>
<td>14439174</td>
<td>14440026</td>
<td>-4.33</td>
<td>8712</td>
<td>TOP1MT</td>
</tr>
<tr>
<td>GGE</td>
<td>9</td>
<td>21350268</td>
<td>21409671</td>
<td>-3.7</td>
<td>59403</td>
<td>IFNA13, IFNA2, IFNA6, IFNA8</td>
</tr>
<tr>
<td>GGE</td>
<td>9</td>
<td>97080895</td>
<td>97090973</td>
<td>-5.74</td>
<td>10078</td>
<td>NUTM2F</td>
</tr>
<tr>
<td>GGE</td>
<td>9</td>
<td>113189903</td>
<td>113550109</td>
<td>-7.57</td>
<td>360206</td>
<td>MUSK, SVEP1</td>
</tr>
<tr>
<td>GGE</td>
<td>10</td>
<td>52432177</td>
<td>52506529</td>
<td>-3.39</td>
<td>74352</td>
<td>PLXDC2</td>
</tr>
<tr>
<td>GGE</td>
<td>10</td>
<td>55568487</td>
<td>55582740</td>
<td>-4.38</td>
<td>14253</td>
<td>PCDH15</td>
</tr>
<tr>
<td>GGE</td>
<td>10</td>
<td>90524124</td>
<td>90534348</td>
<td>-5.32</td>
<td>10224</td>
<td>LIPI</td>
</tr>
<tr>
<td>GGE</td>
<td>10</td>
<td>116919814</td>
<td>117026498</td>
<td>-3.09</td>
<td>106684</td>
<td>ATRNL1</td>
</tr>
<tr>
<td>GGE</td>
<td>11</td>
<td>26568916</td>
<td>26587286</td>
<td>-3.41</td>
<td>18370</td>
<td>ANO3, MUC15</td>
</tr>
<tr>
<td>GGE</td>
<td>11</td>
<td>40136003</td>
<td>40137868</td>
<td>-3.76</td>
<td>1865</td>
<td>LRR2C4</td>
</tr>
<tr>
<td>GGE</td>
<td>11</td>
<td>60531165</td>
<td>60621186</td>
<td>-4.85</td>
<td>90021</td>
<td>CCDC36, MS4A10, MS4A15, PTDGR2</td>
</tr>
<tr>
<td>GGE</td>
<td>11</td>
<td>72465895</td>
<td>72794788</td>
<td>-3.31</td>
<td>328893</td>
<td>AG16L2, FCHSD2, MR4459, MR4692, STARD10</td>
</tr>
<tr>
<td>GGE</td>
<td>11</td>
<td>124844950</td>
<td>124850818</td>
<td>-5.07</td>
<td>13068</td>
<td>CCDC15</td>
</tr>
<tr>
<td>GGE</td>
<td>12</td>
<td>40749853</td>
<td>41463875</td>
<td>-4.27</td>
<td>714022</td>
<td>CNTN1, LRK2, MUC19</td>
</tr>
<tr>
<td>GGE</td>
<td>12</td>
<td>53073535</td>
<td>53086673</td>
<td>-3.42</td>
<td>13138</td>
<td>KRT1, KRT77</td>
</tr>
</tbody>
</table>

(Continued)
deleted in RE (n = 85), GGE (n = 49) and RE+GGE (n = 134) against the brain-expressed epilepsy-related genes in DisGeNet (n = 674). We used the total number of brain-expressed genes (n = 14,177) as the background. The R GeneOverlap package (https://bioconductor.org/packages/release/bioc/html/GeneOverlap.html) was used to compute the p-value.

**CNV tolerance score analysis**

The CNV tolerance score was used as defined in [45]. The CNV tolerance and deletion scores for the genes that are deleted in our study were obtained from the ExAC database [46] and their enrichment in GGE and RE cases was assessed by the Wilcoxon rank sum test.

**Overlap with different databases**

The overlap between the different data sets was obtained by gene symbol matches between the detected gene deletions and the gene lists from different databases; more details are given in the discussion section. A workflow depicting the steps above is shown in Fig 1.

**Results**

After quality control, exomes of 390 epilepsy cases (196 GGE, 194 RE) and 572 controls were used for downstream analyses (Fig 1). The final RE and GGE datasets comprised 26,476 and 30,207 variants, respectively.

**Epilepsy-associated microdeletions**

75 out of 390 epilepsy patients (~19%) carried a total of 104 case-only deletions spanning 260 genes (see Table 1), which covered a wide size range between 915 bp and 3.11 Mbp. 43 out of 194 RE patients carried deletions compared to 32 out of 196 patients with GGE, thus, we did not observe any significant difference in the total number of deletions between the two disease

### Table 1. (Continued)

<table>
<thead>
<tr>
<th>Type</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Z-score</th>
<th>Length</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGE</td>
<td>12</td>
<td>56825208</td>
<td>56827994</td>
<td>-5.15</td>
<td>2786</td>
<td>TIMELESS, ANKRD20A19P, CIQTNF9, CIQTNF9B, CIQTNF9B-AS1, LINC00327, MIPEP, MIR2276, SACS, SACS-AS1, SGCC, SPATA13, SPATA13-A S1, TNFRSF19</td>
</tr>
<tr>
<td>GGE</td>
<td>12</td>
<td>91445028</td>
<td>91450028</td>
<td>-3.85</td>
<td>4956</td>
<td>KERA, FANCA, ZNF276</td>
</tr>
<tr>
<td>GGE</td>
<td>13</td>
<td>23777841</td>
<td>24895857</td>
<td>-3.6</td>
<td>1118016</td>
<td>ANKRD20A19P, CIQTNF9, CIQTNF9B, CIQTNF9B-AS1, LINC00327, MIPEP, MIR2276, SACS, SACS-AS1, SGCC, SPATA13, SPATA13-A S1, TNFRSF19</td>
</tr>
<tr>
<td>GGE</td>
<td>16</td>
<td>20471400</td>
<td>20498025</td>
<td>-7.68</td>
<td>26625</td>
<td>ACSM2A</td>
</tr>
<tr>
<td>GGE</td>
<td>16</td>
<td>56659681</td>
<td>56693111</td>
<td>-4.59</td>
<td>33430</td>
<td>MT1A, MT1B, MT1DP, MT1E, MT1F, MT1P, MT1M</td>
</tr>
<tr>
<td>GGE</td>
<td>16</td>
<td>61747707</td>
<td>61859108</td>
<td>-3.57</td>
<td>111401</td>
<td>CDH8</td>
</tr>
<tr>
<td>GGE</td>
<td>16</td>
<td>89804176</td>
<td>89849549</td>
<td>-4.26</td>
<td>45373</td>
<td>FANCA, ZNF276</td>
</tr>
<tr>
<td>GGE</td>
<td>16</td>
<td>36453091</td>
<td>36485777</td>
<td>-3.98</td>
<td>32686</td>
<td>GPR179, MRPL45</td>
</tr>
<tr>
<td>GGE</td>
<td>17</td>
<td>62850638</td>
<td>62856934</td>
<td>-4.04</td>
<td>6296</td>
<td>LRRC37A3</td>
</tr>
<tr>
<td>GGE</td>
<td>18</td>
<td>43496355</td>
<td>43604681</td>
<td>-3.89</td>
<td>108326</td>
<td>EP5G, PSTPIP2</td>
</tr>
<tr>
<td>GGE</td>
<td>19</td>
<td>1056280</td>
<td>1061916</td>
<td>-3.36</td>
<td>5636</td>
<td>ABCA7</td>
</tr>
<tr>
<td>GGE</td>
<td>19</td>
<td>9270761</td>
<td>9272102</td>
<td>-4.04</td>
<td>1341</td>
<td>ZNF317</td>
</tr>
<tr>
<td>GGE</td>
<td>19</td>
<td>37309563</td>
<td>37619956</td>
<td>-3.24</td>
<td>310393</td>
<td>ZNF345, ZNF420, ZNF568, ZNF790, ZNF790-AS1, ZNF829</td>
</tr>
<tr>
<td>GGE</td>
<td>19</td>
<td>58386121</td>
<td>58420835</td>
<td>-3.08</td>
<td>34714</td>
<td>ZNF417, ZNF814</td>
</tr>
<tr>
<td>GGE</td>
<td>20</td>
<td>2463808</td>
<td>2465032</td>
<td>-3.21</td>
<td>1224</td>
<td>ZNF343</td>
</tr>
<tr>
<td>GGE</td>
<td>20</td>
<td>22562576</td>
<td>23016658</td>
<td>-3.9</td>
<td>454082</td>
<td>FOXA2, LINC01384, SSTR4</td>
</tr>
<tr>
<td>GGE</td>
<td>20</td>
<td>58440579</td>
<td>58444005</td>
<td>-3.3</td>
<td>3426</td>
<td>SYCP2</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0202022.t001
entities (p-value = 0.68). In the combined dataset, 35 out of 73 were large multigene deletions. Among them were several recurrent deletions (see Table 1), including those located on 15q13.3 and 16p11.2 that were previously reported to be associated with epilepsy and other brain disorders.

**Comparative analysis of Rolandoic and GGE candidate genes**

Because our cohort is composed of GGE and RE patients, we sought to compare the functional differences between the two subtypes of epilepsies by studying the pathways and functions that are enriched in the respective deleted genes (see Table 2). Initially we performed GO term enrichment without applying any additional filter to the deletion calls and noticed that synaptic and receptor functions are more prominent in RE cases (data not shown). If the deletion calls were filtered for brain-specific gene expression, we observed that, separately and together, GGE and RE-deleted genes are enriched for the functional terms “nervous system development and function”, “behavior” and “tissue morphology”; this functional convergence might have been expected when selecting for brain-expressed genes.

When analysing GGE and RE datasets separately, the top PPIN enriched in GGE is associated with “carbohydrate metabolism”, “small molecule biochemistry” and “cell signaling”, whereas the top network enriched in RE is associated with “neurological disease”, “organismal injury and abnormalities” and “psychological disorders” (see Table 3). The top enriched network including GGE and RE-deleted genes (Fig 2) is described below.

**Deletion burden analysis**

We performed 10,000 case-control label permutations to test whether there is an increased burden of large and rare deletions in cases as compared to the controls (Table 4). We noticed that (1) the deletion rate per individual with at least one deletion in cases compared to the controls showed statistical significance in both GGE and RE (p-value = 1e-04, p-value = 0.011) and (2), considering cumulative length of all large and small deletions, no significant difference between cases and controls was observed in both GGE and RE (p-value = 0.16, p-value = 0.41), indicating that there is no difference in the length of CNVs in cases and controls.

Table 2. Physiological system development and function.

<table>
<thead>
<tr>
<th>Name</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GGE+RE</td>
</tr>
<tr>
<td>Nervous System Development and Function</td>
<td>2.74E-02–3.36E-06</td>
</tr>
<tr>
<td>Tissue Morphology</td>
<td>2.62E-02–4.20E-06</td>
</tr>
<tr>
<td>Behavior Auditory and Vestibular System Development and Function</td>
<td>2.37E-02–3.63E-05</td>
</tr>
<tr>
<td>Organ Morphology</td>
<td>2.43E-02–5.29E-04</td>
</tr>
<tr>
<td></td>
<td>RE</td>
</tr>
<tr>
<td>Nervous System Development and Function</td>
<td>4.90E-02–3.89E-05</td>
</tr>
<tr>
<td>Tissue Morphology</td>
<td>4.90E-02–1.34E-04</td>
</tr>
<tr>
<td>Behavior</td>
<td>4.90E-02–2.56E-04</td>
</tr>
<tr>
<td>Auditory and Vestibular System Development and Function</td>
<td>4.53E-02–2.59E-04</td>
</tr>
<tr>
<td>Organ Morphology and Vestibular System Development and Function</td>
<td>4.90E-02–2.59E-04</td>
</tr>
<tr>
<td></td>
<td>GGE</td>
</tr>
<tr>
<td>Nervous System Development and Function</td>
<td>4.91E-02–2.28E-04</td>
</tr>
<tr>
<td>Tissue Morphology</td>
<td>4.07E-02–2.28E-04</td>
</tr>
<tr>
<td>Behavior</td>
<td>4.47E-02–4.62E-04</td>
</tr>
<tr>
<td>Hematological System Development and Function</td>
<td>3.81E-02–6.79E-04</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0202022.t002
Enrichment for known epilepsy and autism-associated genes

To check the overlap between the deletions detected in our study and genes known to be associated with epilepsy, we searched for overlap with the genes listed (n = 499) in the Epilepsy-Genes database [47]. This led to the following set of 8 genes: CHRFAM7A, CHRNA7, SCN1A, CNTNAP2, GABRB3, GRIN2A, IGSF8, ITPR1. The GRIN2A deletion is from the same patient published earlier [48] and which we used as one of the positive controls in our primary CNV detection pipeline [49]. One should notice that genes such as CHRNA7 and GABRB3 are located within larger deletions containing other genes; so they might be questionable as bona fide epilepsy-associated genes.

Using the core autism candidate genes (n = 455 genes) present in brainspan [50], we identified 13 deleted genes: APBA2, ATP10A, CDH22, CDH8, GABRA5, GABRG3, NDN, NDNL2, CNTNAP2, GABRB3, GRIN2A, SCN1A and SHANK1 (Table 5). This set is particularly enriched in GO terms “neuron parts” and “transporter complexes”. Note that GABRB3 and GABRG3 belong to multigenic large deletions (Table 1).

Deletions of brain-critical exons

Disorders such as autism, schizophrenia, mental retardation and epilepsy impact fecundity and put negative selection pressure on risk alleles. In a recent report [7] exome and transcriptome data from large human population samples were combined to define a class of brain-expressed exons that are under purifying selection. These exons that are highly expressed in brain tissues and characterized by a low mutation burden in population controls were called “brain-critical exons” (n = 3,955); the associated genes were accordingly called “brain-critical genes” (BCG, n = 1,863) [3].

Twenty-two deleted genes are in common with the BCG set (see Table 5). The SHANK1 deletion is found in a single RE case. It spans 17,339 bp (8 exons out of 9). There is only one report on the possible implication of the deletion of this gene in childhood epilepsy [51]. A deletion of ITPR1 is observed in another RE case; this deletion affects also SUMF1, but this gene was filtered...
out by the BCG overlap selection. The deletion of CNTN1 in a GGE patient encompasses in addition MUC19 and LRRK2, the latter is a known Parkinson candidate gene [52].

**Exome Aggregation Consortium deletions**

The ExAC data comprise 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies. Deletions annotated in ExAC (release 0.3.1 of 23/08/16) were identified, similar to the present study, by read depth analysis using XHMM [45]. We sought to compare those CNV calls with the ones detected in the present work. Out of the 260 deleted genes detected in our study, 164 genes (67%) showed deletions in ExAC too (see S2 Table). Several genes highlighted in the previous paragraphs were ranked high using the CNV tolerance score defined by [45]. However, we did not identify a significant difference, neither in CNV tolerance scores (p-value = 0.53) nor in CNV deletion scores (p-value = 0.22), between GGE and RE-deleted genes. This may indicate that GGE and RE deletions are equally likely to fall into the same category of ExAC deletion calls.

**Compound heterozygous and first order protein-protein interaction mutations**

Compound heterozygous mutations play a role in many disease aetiologies such as autism and Parkinson’s disease [53–55]. We searched for possibly deleterious non-synonymous changes in the parental undeleted gene copy, but we did not detect any hemizygous variant that had a critical intolerance score (see Methods). Subsequently, we hypothesised that simultaneous
mutations in proteins which interact directly (first-order protein interactors) may increase the associated deleterious effect. Within a curated brain-specific PPIN (see Methods, [40]), we inspected first order interacting proteins with potentially deleterious mutations or exon losses (see Table 6) and found a few interesting hits, including SPTAN1 that interacts directly with SHANK1; SPTAN1 encodes alpha-II spectrin and is known to be associated with epilepsy [56,57]. A remarkable and unique case of multiple hits was observed in a patient who accumulated four hits: the originally detected ITPR1 deletion and three potentially deleterious non-synonymous SNVs in RYR2, HOMER2 and STARD13. RYR2 (ryanodine receptor 2) and ITPR1 (inositol-1,4,5-trisphosphate receptor 1) have been independently reported to be

<table>
<thead>
<tr>
<th>Gene with deleterious SNV/INDEL</th>
<th>Gene within deletion boundaries</th>
<th>Type</th>
<th>CHR</th>
<th>position</th>
<th>ref</th>
<th>alt</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LACTB</td>
<td>MRPS27</td>
<td>RE</td>
<td>15</td>
<td>63421767</td>
<td>T</td>
<td>C</td>
<td>exonic</td>
</tr>
<tr>
<td>SPEN</td>
<td>SF3B3, SNORD111, SNORD111B</td>
<td>RE</td>
<td>1</td>
<td>16254645</td>
<td>G</td>
<td>A</td>
<td>exonic</td>
</tr>
<tr>
<td>NRG1</td>
<td>SF3B3, SNORD111, SNORD111B</td>
<td>RE</td>
<td>8</td>
<td>3206278</td>
<td>A</td>
<td>G</td>
<td>exonic</td>
</tr>
<tr>
<td>SPTAN1</td>
<td>SHANK1</td>
<td>RE</td>
<td>9</td>
<td>131367308</td>
<td>T</td>
<td>G</td>
<td>splicing</td>
</tr>
<tr>
<td>STARD13</td>
<td>ITPR1, ITPR1-AS1, SUMF1</td>
<td>RE</td>
<td>13</td>
<td>33700223</td>
<td>C</td>
<td>T</td>
<td>exonic</td>
</tr>
<tr>
<td>RYR2</td>
<td>ITPR1, ITPR1-AS1, SUMF1</td>
<td>RE</td>
<td>1</td>
<td>237730032</td>
<td>A</td>
<td>G</td>
<td>exonic</td>
</tr>
<tr>
<td>HOMER2</td>
<td>ITPR1, ITPR1-AS1, SUMF1</td>
<td>RE</td>
<td>15</td>
<td>83561556</td>
<td>G</td>
<td>C</td>
<td>exonic</td>
</tr>
<tr>
<td>EPS15L1</td>
<td>AGFG2</td>
<td>RE</td>
<td>19</td>
<td>16528403</td>
<td>C</td>
<td>T</td>
<td>exonic</td>
</tr>
<tr>
<td>DXD41</td>
<td>U2SURP</td>
<td>GGE</td>
<td>5</td>
<td>176939650</td>
<td>G</td>
<td>C</td>
<td>splicing</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0202022.t006
implicated in brain disorders. *RYR2* de novo mutations have been identified in patients with intellectual disability [58] and activation of *ITPR1* and *RYR2* can lead to the release of Ca$^{2+}$ from intracellular stores affecting propagating Ca$^{2+}$ waves [59]. *HOMER2*, a brain-expressed gene, has been reported to be involved in signalling defects in neuropsychiatric disorders [60]. The *STARD13* locus has been reported to be associated with aneurysm and sporadic brain arteriovenous malformations [61,62].

**Over-representation of gene-disease associations**

DisGeNET is a discovery platform integrating information on gene-disease associations from public data sources and literature [44]. The current version (DisGeNET v4.0) contains 429,036 associations between 17,381 genes and 15,093 diseases ranked according to supporting evidence. Over-representation analysis of genes that are deleted in both GGE and RE together (134 genes) showed significant over-representation (empirical p-value = 0.012) of epilepsy-associated genes (*APBA2, CHRNA7, CNTNAP2, F5, GABRA5, GABRB3, GRIN2A, KCNQ1, MT1E, PTPRZ1, SCN1A, SGCG, SSTR4*). We observed a similar result for GGE (49 genes; empirical p-value = 0.009; overlapping genes: *CNTNAP2, F5, MT1E, PTPRZ1, SCN1A, SGCG, and SSTR4*), but we did not see an over-representation in RE (85 genes; empirical p-value = 0.217; overlapping epilepsy genes are *APBA2, CHRNA7, GABRA5, GABRB3, GRIN2A,* and *KCNQ1*). This may reflect the heterogeneous risk factors in adulthood epilepsies compared to RE.

**Protein-protein interaction network analysis**

We searched for network modules carrying a higher deletion burden with Ingenuity Pathway Analyser (IPA®). Considering GGE and RE together and using brain-expressed genes as an input for IPA we identified a total of 12 networks. The identified network scores ranged from two to 49 and the number of focus molecules in each network ranged from one to 24. Of all the 12 identified networks, the network shown in Fig 2 is the top-ranked network with a score of 49 and 24 focus molecules. It is associated to the terms “Nervous system development and function”, “Neurological disease” and “Behavior” (see Table 3). The network reveals an interesting module where the genes *CAPN1, GRIN2A, ITPR1, SCN1A* and *CHRNA7* are central. Interestingly, *CAPN1* is well ranked (no deletion or duplication) in the ExAC CNV records (S2 Table) and it is not covered by BCG, epilepsy and autism data sets used in this study.

**Enrichment for likely disruptive de novo mutations**

Many studies on neuropsychiatric disorders such as autism spectrum disorder, epileptic encephalopathy, intellectual disability and schizophrenia have utilized massive trio-based whole-exome sequencing (WES) and whole-genome sequencing (WGS). Epilepsy candidate genes with de novo mutations (DNMs) were searched in the NeuroPsychiatric De Novo Database, NPdenovo [63]. DNMs were found in *GABRB3, SHANK1, ITPR1, GRIN2A, SCN1A, PCDHB4* and *IQGAP2*.

**Discussion**

We analysed a WES dataset of 390 epilepsy patients (196 GEE, 194RE) for microdeletions. The deletion rate per individual with at least one deletion in cases compared to 572 controls showed statistical significance in both GGE and RE. Enrichment for known epilepsy and autism genes led to gene sets with synaptic and receptor functions which were mainly represented in Rolandic cases. The top PPIN enriched in GGE was associated with "carbohydrate
metabolism”, “small molecule biochemistry” and “cell signaling”, whereas the top networks associated with RE are “neurological disease”, “organismal injury and abnormalities” and “psychological disorders”, this is reminiscent of our previous attempt to classify metabolic and developmental epilepsies [3].

Among single-gene deletions, CDH22, CDH12 and CDH8 are of particular interest; CDH12 is a cadherin expressed specifically in the brain and its temporal pattern of expression seems to be consistent with a role during a critical period of neuronal development [64]. Moreover, a group of cadherins, CDH7, CDH12, CDH18 and PCDH12, are reported to be associated with bipolar disease and schizophrenia [65]. The smallest deletion (1,166 bp) that we could detect in this study concerns NCAPD2; this gene is annotated in the autismkb database [66]. It is an important component of the chromatin-condensing complex, which is highly conserved across metazoan. This gene was previously found to be associated with Parkinson’s disease [39] and its paralog NCAPD3 is associated with developmental delay [67].

Deletions of brain-critical exons pointed to the ITPR1 deletion, which has been reported to be associated with spinocerebellar ataxia type 16 [68,69]. CNTN1 is another deletion of interest, the gene is highly expressed in fetal brain, it encodes a neural membrane protein which functions as a cell adhesion molecule and may be involved in forming axonal connections/growth and in neuronal migration in the developing nervous system [70,71]. Moreover, its paralogs CNTN2 and CNTN4 are associated with epilepsy [72] and autism [73], respectively. Interestingly, in the ExAC data, the brain-expressed genes ITPR1 and CNTN1 show the third and fourth highest intolerance score ranks, respectively (S2 Table).

Protein-Protein interaction network analysis revealed the CAPN1 deletion as an interesting candidate gene; this is a double gene loss (4,270 bp) spanning CAPN1 (exon 17 to 22 out of 22 exons) and SLC22A1 (exon 1 out of 10 exons). SLC22A1, a transporter of organic ions across cell membranes, is lowly expressed in the brain, whereas CAPN1 is highly expressed in the brain. Calpain1 (CAPN1) belongs to the calcium-dependent proteases, which play critical roles in both physiological and pathological conditions in the central nervous system. They are also recognized for their synaptic and extra-synaptic neurotoxicity and neuro-protection [74]. Several ion channels, including GRIN2A [75] are calpain substrates. Further, a missense mutation in CAPN1 is associated with spino-cerebellar ataxia in the Parson Russell terrier dog breed [76] and has recently been reported in humans with cerebellar ataxia and limb spasticity [77].

Additional candidate genes can be identified on the periphery of the IPA network (see Fig 2): 1) CNTN1 (commented on above), 2) SACS, for which a large deletion (> 1Mb) was found, and 3) the single gene deletion of KCNQ1 (~ 57 kb). For SACS, a SNV is reported to be associated with spastic ataxia [78] and epilepsy [79]. KCNQ1 and its paralog KCNQ3 are subunits forming an expressed neuronal voltage-gated potassium channel. Further, hypomorphic mutations in either KCNQ2, an established epilepsy-associated gene [80], or KCNQ3 are reported to be highly penetrant [81]. KCNQ1 is co-expressed in heart and brain; it is found in forebrain neuronal networks and brainstem nuclei, regions in which a defect in the ability of neurons to repolarize after an action potential can produce seizures and dysregulate autonomic control of the mouse heart [82], yet one should be cautious as no validation is available for human.

Enrichment for likely disruptive de novo mutations in several genes suggests that deletions of these genes could cause a similar phenotype as in the NPdenovo and consequently will be penetrant in the heterozygotic state. This is indeed the case for ITPR1, for which recessive and dominant de novo mutations causing Gillespie syndrome [83], a rare variant form of aniridia characterized by non-progressive cerebellar ataxia, intellectual disability and iris hypoplasia, have been described. Two of the genes, which we have identified as ITPR1 interactors, RYR2 and SPTAN1, are also DNM genes in DPdenovo.
In summary, by filtering and comparison to genes that are (1) evolutionary constrained in the brain, (2) implicated in autism and epilepsy, (3) spanned by ExAC deletions, or (4) affected by neuropsychiatric associated de novo mutations, we observed a significant enrichment of deletions in genes potentially involved in neuropsychiatric diseases, namely GRIN2A, GABRB3, SHANK1, ITPR1, CNTN1, SCN1A, PCDHb4, IQGAP2, SACS, KCNQ1 and CAPN1. Interaction network analysis identified a hub connecting many of the epilepsy candidate genes identified in this and previous studies. The extended search for likely deleterious mutations in the first order protein-protein interactions and NPdenovo database pointed to the potential importance of ITPR1 deletion alone or in combination with RYR2 and SPTAN1 deleterious mutations.

We are aware that the set of epilepsy exomes that we screened for CNVs in the present study, although the largest analyzed so far, is still small given the genetic complexity of the disease and its population frequency. However, this study appears to provide a contrasting view to the genetic bases of childhood and juvenile epilepsies, as the top protein–protein interactions showing that GGE deleted proteins are preferentially associated with metabolic pathways, whereas in RE cases the association is biased towards neurological processes. Scrutinizing of additional patients’ exomes/genomes and transcriptomes should provide an efficient way to understand the disease aetiology and the biological processes underlying it. The results presented here may contribute to the understanding of epilepsy genetics and provide a resource for future validations to improve diagnostics.

Supporting information

S1 Table. Deletions present in array data.

S2 Table. Deletions in common with ExAC CNVs. Data is sorted from low to high deletion score (del.score) and duplication (dup) frequencies. “+” indicates expression in the brain. Deletion score increases with increasing intolerance.

Acknowledgments

The authors are grateful to the study participants, the staff of the Rotterdam Study and the participating general practitioners and pharmacists. We thank the members of the Genomics Lab and the ERGO support team for their help in sampling the data and in creating the database. Parts of the analysis were conducted on the HPC facilities of the University of Luxembourg (http://hpc.uni.lu). The authors would also like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about. Members of the EuroEPI-NOMICS CoGIE Consortium are Aarno Palotie, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK; Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, USA; Anna-Elina Lehesjoki, Folkhalsan Institute of Genetics, Helsinki, Finland; Neuroscience Center, University of Helsinki, Helsinki, Finland; Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland; Ann-Kathrin Ruppert, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Arvid Suls, Neurogenetics group, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; Auli Siren, Outpatient Clinic for...
Persons with Intellectual Disability, Tampere University Hospital, Tampere, Finland; Birgit Neophytou, St. Anna Children’s Hospital, Department of Neuropediatrics, Vienna, Austria; Bobby Koeleman, Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands; Dennis Lal, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Edda Haberlandt, Department of Pediatrics, Medical University of Innsbruck, Innsbruck, Austria; Eva Maria Reinthaler, Department of Neurology, Medical University of Vienna, Vienna, Austria; Federico Zara, Laboratory of Neurogenetics, Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Gaslini Institute, Genova, Italy; Felicitas Becker, Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; Fritz Zimprich, Department of Neurology, Medical University of Vienna, Vienna, Austria; Gabriel M Ronen, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada; Hande Caglayan, Department of Molecular Biology and Genetics, Bogazici University, Istanbul, Turkey; Helle Hjalgrim, Danish Epilepsy Centre, Dianalund, Denmark; Institute for Regional Health Services, University of Southern Denmark, Odense, Denmark; Hiltrud Muhle, University Medical Center Schleswig-Holstein, Christian-Albrechts University, Kiel, Germany; Hannelore Steinböck, Private Practice of Pediatrics, 1150 Vienna, Austria; Herbert Schulz, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Holger Thiele, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Ingo Helbig, University Medical Center Schleswig-Holstein, Christian-Albrechts University, Kiel, Germany; Janine Altmüller, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Julia Geldner, Department of Pediatrics, Hospital SMZ Süd Kaiser—Franz—Josef Spital, Vienna, Austria; Julian Schubert, Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; Kamel Jabbari, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Kate Everett, Cardiovascular and Cell Sciences Research Institute, St George’s University of London, London, UK; Martha Feucht, Department of Pediatrics, Medical University of Vienna, Vienna, Austria; Martina Balestri, Division of Neurology, Bambino Gesu’ Children’s Hospital and Research Institute, Rome, Italy; Michael Nothnagel, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Pasquale Striano, Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, G. Gaslini Institute, Genoa, Italy; Patrick May, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; Peter Nürnberg, Cologne Center for Genomics, University of Cologne, Cologne, Germany, Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany; Rikke S Møller, Danish Epilepsy Centre, Dianalund, Denmark, Institute for Regional Health Services, University of Southern Denmark, Odense, Denmark; Rima Nabbout, Centre de Reference Epilepsies Rares, Inserm U1129, Neuropediatrics Department, Necker-Enfants Malades Hospital, APHP, Paris Descartes University, CEA, Orsay, France; Roland Krause, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; Rudi Balling, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; Stephanie Baulac, Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épineure, ICM, F-75013, Paris, France; Thomas Sander, Cologne Center for Genomics, University of Cologne, Cologne, Germany; Ursula Gruber-Sedlmayr, Department of Pediatrics, Medical University of Graz, Graz, Austria; Wolfram Kunz, Department of Epileptology
and Life & Brain Center, University of Bonn Medical Center, Bonn, Germany; Yvonne G. Weber, Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany. Coordinator of the CoGIE Consortium is Holger Lerche from Tübingen (contact email address: holger.lerche@uni-tuebingen.de).

**Author Contributions**

**Conceptualization:** Kamel Jabbari, Dheeraj R. Bobbili, Thomas Sander, Patrick May, Holger Lerche, Peter Nürnberg.

**Data curation:** Kamel Jabbari, Dheeraj R. Bobbili, Dennis Lal, Eva M. Reinthaler, Julian Schubert, Stefan Wolking, Vishal Sinha, Jeroen van Rooij, Roland Krause.

**Formal analysis:** Dheeraj R. Bobbili, Vishal Sinha, Susanne Motameny, Holger Thiele, Amit Kawalia, Robert Kraaij, Jeroen van Rooij, Roland Krause, Patrick May.

**Funding acquisition:** André G. Uitterlinden, M. Arfan Ikram, Federico Zara, Anna-Elina Lehesjoki, Fritz Zimprich, Bernd A. Neubauer, Holger Lerche, Peter Nürnberg.

**Investigation:** Kamel Jabbari, Dheeraj R. Bobbili, Eva M. Reinthaler, Julian Schubert, Stefan Wolking, Susanne Motameny, Holger Thiele, Amit Kawalia, Janine Altmüller, Mohammad Reza Toliat, Robert Kraaij, Patrick May.

**Methodology:** Kamel Jabbari, Dheeraj R. Bobbili, Dennis Lal, Vishal Sinha, Susanne Motameny, Holger Thiele, Amit Kawalia, Janine Altmüller, Robert Kraaij, Jeroen van Rooij, M. Arfan Ikram, Roland Krause, Thomas Sander, Patrick May.

**Project administration:** André G. Uitterlinden, M. Arfan Ikram, Anna-Elina Lehesjoki, Fritz Zimprich, Bernd A. Neubauer, Holger Lerche, Peter Nürnberg.

**Resources:** Janine Altmüller, André G. Uitterlinden, M. Arfan Ikram, Federico Zara, Anna-Elina Lehesjoki, Fritz Zimprich, Bernd A. Neubauer, Holger Lerche, Peter Nürnberg.

**Software:** Roland Krause, Patrick May.

**Supervision:** André G. Uitterlinden, Roland Krause, Patrick May, Holger Lerche, Peter Nürnberg.

**Validation:** Dennis Lal, Mohammad Reza Toliat, Thomas Sander.

**Visualization:** Dheeraj R. Bobbili.

**Writing – original draft:** Kamel Jabbari.

**Writing – review & editing:** Dheeraj R. Bobbili, Dennis Lal, Eva M. Reinthaler, Julian Schubert, Stefan Wolking, Vishal Sinha, Susanne Motameny, Holger Thiele, Amit Kawalia, Janine Altmüller, Mohammad Reza Toliat, Robert Kraaij, Jeroen van Rooij, André G. Uitterlinden, M. Arfan Ikram, Federico Zara, Anna-Elina Lehesjoki, Roland Krause, Fritz Zimprich, Thomas Sander, Bernd A. Neubauer, Patrick May, Holger Lerche, Peter Nürnberg.

**References**


