Heterogeneous contribution of microdeletions in the development of common generalised and focal epilepsies

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi:10.1136/jmedgenet-2016-104495

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:34491987

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Heterogeneous contribution of microdeletions in the development of common generalised and focal epilepsies

Eduardo Pérez-Palma,1,2 Ingo Helbig,3,4 Karl Martin Klein,5,6 Verneri Anttila,7,8 Heiko Horn,9 Eva Maria Reinthaler,10 Padhraig Gormley,7,11 Andrea Ganna,7 Andrea Byrnes,7 Katharina Pernhorst,12 Mohammad R Toliat,2 Elmo Saarentaus,7,8 Daniel P Howrigan,7,8 Per Hoffman,13 Juan Francisco Miquel,14 Giancarlo V De Ferrari,1 Peter Nürnberg,2 Holger Lerche,15 Fritz Zimprich,10 Bern A Neubauer,16 Albert J Becker,12 Felix Rosenow,5,6 Emilio Perucca,17,18 Federico Zara,19 Yvonne G Weber,14 Dennis Lal2,7,8

ABSTRACT

Background Microdeletions are known to confer risk to epilepsy, particularly at genomic rearrangement ‘hotspot’ loci. However, microdeletion burden not overlapping these regions or within different epilepsy subtypes has not been ascertained.

Objective To decipher the role of microdeletions outside hotspot loci and risk assessment by epilepsy subtype.

Methods We assessed the burden, frequency and genomic content of rare, large microdeletions found in a previously published cohort of 1366 patients with genetic generalised epilepsy (GGE) in addition to two sets of additional unpublished genome-wide microdeletions found in 281 patients with rolandic epilepsy (RE) and 807 patients with adult focal epilepsy (AFE), totalling 2454 cases. Microdeletions were assessed in a combined and subtype-specific approaches against 6746 controls.

Results When hotspots are considered, we detected an enrichment of microdeletions in the combined epilepsy analysis (adjusted p=1.06×10−5, OR 1.89, 95% CI 1.51 to 2.35). Epilepsy subtype-specific analyses showed that hotspot microdeletions in the GGE subgroup contribute most of the overall signal (adjusted p=9.79×10−12, OR 7.45, 95% CI 4.20–13.5). Outside hotspots, microdeletions were enriched in the GGE cohort for neurodevelopmental genes (adjusted p=9.13×10−12, OR 2.85, 95% CI 1.62–4.94). No additional signal was observed for RE and AFE. Still, gene-content analysis identified known (NRXN1, RBF1X1 and PCDH7) and novel (LOC102273362) candidate genes across epilepsy subtypes that were not deleted in controls.

Conclusions Our results show a heterogeneous effect of recurrent and non-recurrent microdeletions as part of the genetic architecture of GGE and a minor contribution in the aetiology of RE and AFE.

INTRODUCTION

Epilepsies comprise a clinically complex group of neurological disorders defined by recurrent spontaneous seizures,1 with an age-adjusted global prevalence estimated in the range of 2.7–17.6 per 1000 individuals.2 The most common types of epilepsies are the heterogeneous group of acquired and non-acquired adult focal epilepsies (AFE), the non-lesional genetic generalised epilepsies (GGE) and the non-lesional childhood focal rolandic epilepsies (RE). The genetic architecture of these common epilepsies is presumed to be complex as it has been described by a wide range of syndrome-specific variant associations, as well as a few shared seizure susceptibility variants (for a review, see Noebels3). Twin studies have shown strong but differential concordance rates among epilepsies, including GGE and AFE with −80% and −40%, respectively.4 In contrast, the genetic contribution to RE may be related to the underlying electroencephalogram (EEG) pattern rather than the seizures themselves.6 Despite these differences, familial enrichment for seizure disorders has been demonstrated, as well as genetic risk factors in a single gene (GRIN2A) in patients with RE and other RE-related syndromes.7

CNVs are genomic segments between 50bp and 3Mb, which may result in loss and/or gain of genomic sequence relative to the reference genome based on the number of copies present.10 CNVs are a significant source of genetic variation between two individuals and can, depending on their location on the chromosomes, cause changes in gene dosage, alternative splicing or lead to gene fusion events.11 Microdeletions, defined as large (ie, >400kb) and rare deletions (minor allele frequency <1%), are more likely to have damaging effects than duplications.10 In addition, from an analysis standpoint, duplications are more prone to false-positive calls as they are more difficult to identify in genotype array data.

Microdeletions are associated with a broad spectrum of neurological diseases such as autism spectrum disorder (ASD),12 schizophrenia (SCZ)14 and intellectual disability (ID).15 Previously, 12 genomic regions prone to exhibit copy number rearrangements or CNV ‘hotspot loci’ have been reported to
increase the risk for neurodevelopmental disorders, including seven CNVs associated with epilepsy, namely, 1q21.1, 15q11.2, 15q13.3, 16p11.2, 16p12, 16p13.11 and 22q11.2 with GGE, and also 16p11.2 with RE and AFE. The relationship to epilepsy of the remaining five loci, 3q29, 10q22q23, 15q24, 17q12 and 17q21.3, remains to be elucidated. Overall, how these large and polygenic microdeletions increase risk for neurodevelopmental disorders is not fully understood. Similarly, the contribution of additional genomic regions outside these hotspots towards common types of epilepsy remains to be elucidated. Recent evidence from patients with GGE showed an enrichment of genes within genomic boundaries of recurrent microdeletions associated with neurodevelopmental disorders including ASD, ID and SCZ collectively named ‘neurodevelopmental genes’. Whether this enrichment is also found in other types of common epilepsies has not yet been evaluated. Previous results support non-recurrent microdeletions in RBFOX1,22 24 25 NRXN124 and GRIN2A24 in candidate gene studies in various epilepsies. However, a genome-wide comparison for shared or subtype-specific deleted genes in GGE, RE and AFE has not yet been conducted.

Due to the low frequency of microdeletions, large sample sizes are required to identify novel susceptibility genes and to decipher syndrome-specific patterns. Considering that previous microdeletion associations are generally reported only within a particular type of epilepsy, our goal was to evaluate the global and specific contribution of microdeletions across GGE, RE and AFE. In order to investigate the microdeletion burden landscape for common types of epilepsy, we examined a total of 9200 individuals including 2454 epilepsy cases and 6746 controls. We combined the microdeletions found in a published GGE cohort20 with unpublished genome-wide microdeletions found in RE21 and AFE22 studies. We investigated the combined and subtype-specific burden of microdeletions, as well as their frequency, distribution and gene content. Finally, we assessed the protein–protein interactions and tissue-specific expression pattern of genes that were only deleted in patients.

**MATERIALS AND METHODS**

**Case–control cohorts**

All patient and control cohorts included in the analysis are of European origin have been genetically matched with their respective controls and have been described previously in detail20–22 (table 1). The epilepsy subtype classifications were based on the terminology proposed by the Commission on Classification and Terminology of the International League against Epilepsy. For an extended description of phenotypes, sample recruitment, genotyping and CNV calling, see online supplementary information files. Briefly, the epilepsy cohort was composed of 1366 GGE cases genotyped with the Genome-Wide Human SNP Array 6.0 platform (Affymetrix, Santa Clara, California, USA)20 plus 281 and 807 RE and AFE cases, respectively, which were genotyped using the Human OmniExpress BeadChip platform (Illumina, San Diego, California, USA). The control set was composed of 5234 samples extracted from the original GGE study and genotyped with the Genome-Wide Human SNP Array 6.0 platform (n=5234)20 plus 1512 controls extracted from the original RE study21 and genotyped using the Human OmniExpress BeadChip platform, totalling 6746 control individuals (table 1).

**Microdeletion calling**

The software PennCNV was used as the CNV calling algorithm following author recommendations for the corresponding Affymetrix and Illumina arrays.27 Microdeletions were defined as autosomal deletions spanning at least 400 kbp in both Affymetrix and Illumina arrays to enrich for likely pathogenic variants.28 Additionally, to ensure highly reliable calls, we only considered microdeletions involving at least 200 probes for the Affymetrix 6.0 platform.20 28 Considering that the amount of probes of the Illumina Omni express arrays (n=740 000 markers) was less than half (40%) of the Affymetrix 6.0 platform (~1 850 000 markers), but at the same time less prone to false positives,28 a threshold of 100 markers was applied to the Illumina calls. Only rare microdeletions, defined by a cohort-specific frequency below 1%, were considered for further analysis. Microdeletion frequency was calculated considering only the control cohort (n=6746).

**Burden statistics**

For the combined epilepsy data set, we evaluated patient and control autosomal microdeletion burden using a binomial logistic regression model implemented in the R statistical software.29 To account for possible bias introduced due to the different genotyping platforms, we adjusted for this factor in the regression by including ‘platform’ as a covariate in the model. Corresponding ORs and 95% CIs were estimated from the log-likelihood function, whereas the association p values for the combined regression coefficients were calculated using a Wald’s test with 1 df. For the epilepsy subtype analysis, each group (GGE, RE and AFE) was tested individually for microdeletion enrichment in comparison with platform-matched controls. GGE samples were compared with the 5234 Affymetrix controls, whereas RE and AFE samples were compared individually to the 1512 Illumina controls (table 1). The p values, corresponding ORs and 95% CIs, were calculated with a two-sided Fisher’s exact test. Because eight gene sets (see below) were interrogated over four case–control configurations (combined, GGE, AFE and RE) with and without hotspot loci inclusion, nominal p values were adjusted with Bonferroni method considering 64 comparisons. Adjusted two-sided p values <0.05 were considered significant.

**Microdeletions subsets interrogated**

We evaluated burden enrichment within eight microdeletion subsets: (1) all microdeletions, (2) microdeletions overlapping hotspot loci, (3) microdeletions outside hotspot loci, (4) microdeletions overlapping ‘constrained’ genes, (5) microdeletions overlapping ‘neurodevelopmental’ genes, (6) microdeletions overlapping ‘ASD-related’ genes, (7) microdeletions overlapping ‘DDG2P’ genes, and (8) microdeletions overlapping ‘loss of function intolerant’ genes.

Known CNV hotspot loci were extracted from a previous review on microdeletion syndromes28 (no of genes=330). For subset 4, we define ‘constrained’ genes as those not overlapped by a CNV of the CNV control map reported by Zarrei et al.28

---

**Table 1** Data sets main features.

<table>
<thead>
<tr>
<th>Epilepsy subtype</th>
<th>Code</th>
<th>Cases</th>
<th>Controls</th>
<th>Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic generalised epilepsy</td>
<td>GGE</td>
<td>1366</td>
<td>5234</td>
<td>Affymetrix GenomeWide 6.0</td>
</tr>
<tr>
<td>Rolandic epilepsy</td>
<td>RE</td>
<td>281</td>
<td>1512</td>
<td>Illumina Omni Express Array</td>
</tr>
<tr>
<td>Adult focal epilepsy</td>
<td>AFE</td>
<td>807</td>
<td>–</td>
<td>Illumina Omni Express Array</td>
</tr>
<tr>
<td>Total</td>
<td>2454</td>
<td>6746</td>
<td>9200</td>
<td></td>
</tr>
</tbody>
</table>

---

which constitutes a curated version of the database of genomic variants on healthy individuals (no of genes=20,208). For subset 5, ‘neurodevelopmental’ genes were extracted from the original GGE study and defined based on literature and database queries (no of genes=1683). For subset 6, ‘ASD-related’ genes were extracted from Uddin et al and defined as those enriched for deleterious exonic de novo mutations in comparison with healthy siblings (no of genes=15,591). For subset 7, ‘DDG2P’ genes were extracted from a curated list of genes reported to be associated with developmental disorders, compiled by clinicians as part of the DDD study to facilitate clinical feedback of likely causal variants (no of genes=294; https://decipher.sanger.ac.uk/ddd#ddgenes). For subset 8, ‘loss of function intolerant’ genes were extracted from and defined by having less than expected loss-of-function (LoF) variants within the 60,706 unrelated individuals from the exome aggregation consortium (no of genes=2,506). The complete gene sets are available as a publicly available resource in https://github.com/dlal-group/gene-sets).

**Tissue-specific expression analysis**

Gene expression analysis was performed using GTEx project resource (http://www.gtexportal.org/home/; version3), which includes gene expression of data 42 tissues from 1,561 human samples. Filtered candidate genes were used as a query to evaluate significant enrichment of tissue-specific expression against a background distribution derived from multiple permutations. Extended description of the implemented methodology is provided in the online supplementary information files.

**Network analysis**

We evaluated brain-expressed genes exclusively deleted in cases (ie, genes overlapped by at least one microdeletion in cases and never in controls). Brain-expressed genes were extracted from the BrainSpan RNA-Seq transcriptome data set (>4.5 log2 of reads per kilo base per million reads). Five gene input sets were generated: (1) genes exclusively deleted in the combined epilepsy analysis (GGE+RE+AFE), (2) genes exclusively deleted in patients with GGE, (3) genes exclusively deleted in patients with RE, (4) genes exclusively deleted in patients with AFE and finally (5) an additional control set consisting of genes exclusively deleted in controls. The Disease Association Protein-Protein Link Evaluator (DAPPLE) software (available at https://genepattern.broadinstitute.org/) was run with each of the five gene list as input, using 1000 iterations for each set. Network enrichment was calculated with the hypergeometric test over MSigDB pathways adjusting for genes being member of the InWeb network.

**RESULTS**

The analysis was performed in two stages. First, the combined autosomal burden analysis for the entire epilepsy cohort (GGE+RE+AFE) was performed. Second, we performed a subtype burden analyses on GGE, RE and AFE independently. In both strategies, we tested burden enrichment among eight microdeletion subtypes: (1) all microdeletions, (2) microdeletions overlapping hotspot loci and (3) microdeletions outside hotspot loci. Hotspot loci included collectively 12 loci: 1q21.1, 3q29, 10q22q23, 15q11.2, 15q13.3, 15q24, 16p11.2, 16p12, 16p13.11, 17q12, 17q21.3 and 22q11.2. Subsequently, considering epilepsy aetiology and comorbidities, microdeletions were filtered to include (4) microdeletions overlapping ‘constrained’ genes, (5) microdeletions overlapping ‘neurodevelopmental’ genes, (6) microdeletions overlapping ‘ASD-related’ genes, (7) microdeletions overlapping ‘developmental disorders’ genes (DDG2P) and finally (8) microdeletions overlapping ‘loss of function intolerant’ genes.

**Microdeletion frequency: combined burden analysis**

To investigate the overall contribution of microdeletions in the aetiology of common types of epilepsy, we combined all the published microdeletions found in the GGE cohort with the genome-wide microdeletions identified in RE and AFE cohorts (table 1). In total, we analysed 134 microdeletions in 2454 patients with epilepsy compared with 219 microdeletions in 6746 controls. Only 24 microdeletions (6.74%) were found in more than one sample, with a maximum frequency of 0.059% (n=4 out of 6746 control samples), which was reached by only two microdeletions. Thus, all included microdeletions were rare (<1%) and mostly singletons (93.04%). Genomic context and sample annotation of all microdeletions found in the combined epilepsy sample is provided in online supplementary table S1.

We compared the frequency of cases with microdeletions against the frequency of controls with microdeletions, interrogating their burden while controlling for batch (ie, platform) effects with a linear regression model (see the Materials and methods section). Overall, 5.42% of cases (n=133) carried at least one microdeletion compared with 3.46% (n=234) in controls (adjusted p=1.60×10−6, OR 1.89, 95%CI 1.51 to 2.35). The number of individuals carrying at least two microdeletions did not differ significantly between cases (no of cases=4 out of 2454) and controls (no of controls=7 out 6746, p=0.498, OR 1.57, 95%CI 0.33 to 6.18), suggesting that enrichment in patients for single microdeletions was not due to an excessive number of microdeletions per patient, which may arise due to batch effects in CNV calling. For the combined results, the most significant enrichment was observed for microdeletions overlapping 1 of the 12 hotspot loci (adjusted p=1.59×10−11, OR 6.99, 95%CI 4.2 to 11.97), which is in agreement with previous observations in the GGE cohort that considered only seven hotspot loci. In this regard, we decided to examine the contribution of microdeletion burden in patients with epilepsy outside these known hotspot loci. Thus, a total of 58 microdeletions were filtered out from the analysis (online supplementary table S1). The contribution of microdeletions inside these regions is substantial, and the overall enrichment did not reach significance after hotspot loci had been removed (adjusted p=1, OR 1.34, 95%CI 1.03 to 1.73).

**Microdeletion distribution: epilepsy subtype burden analysis**

We hypothesised that microdeletions outside of the hotspot loci are also conferring risk for the disease but are more heterogeneously distributed across the genome and epilepsy subtype specific. In this regard, we subsequently compared the microdeletion burden, including hotspot loci, for each epilepsy subtype (GGE, AFE and RE; online supplementary figure S2) and then investigated whether we can identify enrichment for candidate microdeletion subsets not overlapping these regions (figure 1). As expected, the analysis showed that patients with GGE were most significantly enriched for microdeletions overlapping hotspot loci (adjusted p=9.79×10−15, OR 7.45, 95%CI 4.21 to 13.3) and overlapping neurodevelopmental genes as previously shown by Lal et al. Interestingly, microdeletions overlapping neurodevelopmental genes not overlapping hotspot loci remained significant (adjusted p=9.13×10−3, OR 2.85, 95%CI 1.62 to 4.94). In contrast, patients with RE showed nominal significance and a large effect size of microdeletions overlapping
Microdeletions gene content: epilepsy candidate gene identification outside hotspot loci

To extract additional epilepsy genes of interest overlapped by microdeletions outside hotspot loci (already examined by Lal et al20), we generated a short list of potential epilepsy candidate genes that can be followed up in future studies. First, we performed a gene-oriented burden analysis outside hotspot loci comparing the number of cases carrying a microdeletion within a gene against the corresponding number of controls. We detected nominal association for NRXN1 and RBFOX1 genes (table 2, nominal p=0.019, no of cases=3, no of controls=0). Due to the low frequency of microdeletion events, large sample sizes are required to identify significant enrichment for a particular gene. Second, we report genes outside hotspot loci boundaries, affected more than once by microdeletions in cases and not in controls (eg, case-only genes). Finally, we also highlight genes with at least one case affected by a microdeletion belonging to the DDG2P gene list for developmental disorders,31 namely, SKI, KCNA2, GCH1 and DVL1. Nominally associated genes, candidate genes and DDG2P genes are summarised in table 2.

Overall, NRXN1, PCDH7, TSNARE1 and RBFOX1 have been previously reported to be associated with GGE.20 23 Although microdeletions overlapping RBFOX1 were expected to be present also in the AFE cohort based on the original study,22 we also identified additional microdeletions overlapping NRXN1.
and PCDH7 carried by patients with RE and AFE as well. CNVs overlapping LOC102723362 have been reported in patients with neurodevelopmental phenotypes, including ASD and ID (DECIPHER entries 278594, 260750 and 249773, database version 9.3).31 Microdeletions overlapping PACRG were observed in patients with AFE and GGE. Furthermore, two patients with GGE had microdeletions overlapping the ADGRB1 as well as three non-coding genes (LINC00051, MIR1302-7, MIR4472-1 and MIR4539). For LOC102723362, PACRG and LOC101928137, no microdeletions were reported in the curated CNV map of healthy individuals36 in the Database of Genomic Variants (table 2).

### Microdeletions gene content: tissue-specific expression and network analysis of candidate genes

To further evaluate the plausible involvement of the selected candidate genes in neuronal processes and epilepsy, we performed a global gene set enrichment analysis using expression data from the GTEx portal (http://www.gtexportal.org/home/; version3). We compared the expression patterns of the 12 genes deleted more than twice in patients (table 2) versus the 92 genes deleted more than once in controls. The analysis did not show significant differences after correction for multiple testing (online supplementary figure S4). Finally, to evaluate the entire pool of genes exclusively deleted in patients in a network context, we used the DAPPLE36 to assess protein–protein interactions networks with a higher number of interconnections than expected. As DAPPLE uses a non-brain-specific network, we filtered the input genes based on brain expression to improve specificity of the analysis (see the Materials and methods section). Five gene sets were tested: 116 genes exclusively deleted in the combined analysis, 105 genes exclusively deleted in GGE, 116 genes exclusively deleted in RE, and 105 genes exclusively deleted in AFE samples, respectively. As a control, the 110 genes exclusively deleted in controls samples was also tested (see online supplementary table S5). Genes exclusively deleted in the combined epilepsy cohort showed a significant interconnection (figure 2; direct network p=0.00099), whereas genes exclusively deleted in controls did not show any signal (p=0.46). Only the GGE epilepsy subgroup resulted in a significant interconnection (figure 2; direct network p=0.0019).

Among the genes with significant interconnections (figure 2, red nodes), we highlight GRM1, PLCB1 and MAPK3 as members of the KEGG long-term depression and long-term potentiation gene sets (hsa04730 and hsa04720, adjusted p=2.89×10⁻³).

### DISCUSSION

We have previously shown an enriched burden of microdeletions overlapping neurodevelopmental genes in patients with GGE in comparison with controls and that the signal was particularly concentrated in seven hotspot loci.26 In the present work, we extended our analysis to additional epilepsy subtypes. Although we were able to replicate the original GGE signal, we were unable to identify specific associations in the RE and AFE cohorts. In addition, we refined our analysis with novel analysis tools, including the evaluation of 12 hotspot loci, microdeletions overlapping constrained genes, neurodevelopmental disorders genes (DDG2P) and LoF intolerant genes. Moreover, focusing on genes exclusively deleted in cases, we identified additional genes with plausible pathogenic behaviour.

### Burden of microdeletions in patients with epilepsy

In the combined analysis, we found a 1.39-fold excess of patients with microdeletions compared with controls, which translates into 4.85% of patients with epilepsy carrying at least one microdeletion compared with 3.47% of controls. Accordingly, microdeletions have a small but significant contribution towards the overall genetic susceptibility for common epilepsy types, which is particularly strong for GGE. The spectrum of 134 microdeletions identified in patients of the combined epilepsy analysis contained a high proportion of recurrent microdeletions at genomic rearrangement hotspot loci (32.34%). Again, the highest fraction was seen in patients with GGE (90.47%). Although we controlled for array differences (Affymetrix vs Illumina) in the combined analysis, we cannot be certain that platform bias was completely removed in our analysis. Therefore, we cannot conclude that RE or AFE have no additional enrichment of microdeletions, especially considering that individuals with microdeletions only account for a minor proportion of patients. Similarly, in the subtype-specific analysis, we only compared platform-matched cases with controls. Nevertheless, the complete contribution of microdeletions to the presentation

---

**Table 2** Candidate gene identification outside hotspot loci

<table>
<thead>
<tr>
<th>Nominally associated genes</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Patients (n=2454)</th>
<th>Controls (n=6746)</th>
<th>Nominal p value</th>
<th>Constrained gene*</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBFOX1</td>
<td>16</td>
<td>6069131</td>
<td>7763340</td>
<td>3 (AFE, AFE, GGE)</td>
<td>0</td>
<td>0.019</td>
<td>No</td>
</tr>
<tr>
<td>NrxN1</td>
<td>2</td>
<td>50145642</td>
<td>51259674</td>
<td>3 (GGE, GGE, RE)</td>
<td>0</td>
<td>0.019</td>
<td>No</td>
</tr>
<tr>
<td>Candidate genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC102723362</td>
<td>2</td>
<td>22759350</td>
<td>2276159</td>
<td>2 (GGE, RE)</td>
<td>0</td>
<td>0.0711</td>
<td>Yes</td>
</tr>
<tr>
<td>PCDH7</td>
<td>4</td>
<td>30722036</td>
<td>3148423</td>
<td>2 (AFE, GGE)</td>
<td>0</td>
<td>0.0711</td>
<td>No</td>
</tr>
<tr>
<td>PACRG</td>
<td>6</td>
<td>163148164</td>
<td>163736524</td>
<td>2 (GGE, AFE)</td>
<td>0</td>
<td>0.0711</td>
<td>Yes</td>
</tr>
<tr>
<td>ADGRB1, TSNAE1, LINC00051, MIR1302-7, MIR4472-1, MIR4539</td>
<td>8</td>
<td>142867603</td>
<td>14362638</td>
<td>2 (GGE, GGE)</td>
<td>0</td>
<td>0.0711</td>
<td>No, No, Yes, Yes, No, Yes</td>
</tr>
<tr>
<td>LOC101928137</td>
<td>12</td>
<td>73552969</td>
<td>73602097</td>
<td>2 (GGE, GGE)</td>
<td>0</td>
<td>0.0711</td>
<td>Yes</td>
</tr>
<tr>
<td>DDG2P genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKI, DVL1</td>
<td>1</td>
<td>1270658</td>
<td>2241652</td>
<td>1 (AFE)</td>
<td>0</td>
<td>0.267</td>
<td>Yes, No</td>
</tr>
<tr>
<td>KCNA2</td>
<td>1</td>
<td>111136202</td>
<td>111174096</td>
<td>1 (GGE)</td>
<td>0</td>
<td>0.267</td>
<td>Yes</td>
</tr>
<tr>
<td>GCH1</td>
<td>14</td>
<td>55308724</td>
<td>55369542</td>
<td>1 (RE)</td>
<td>0</td>
<td>0.267</td>
<td>No</td>
</tr>
</tbody>
</table>

*Constrained gene according to the CNV map of healthy individuals.36

AFE, adult focal epilepsy; GGE, genetic generalised epilepsy; RE, rolandic epilepsy.

Single gene and candidate gene enrichment analysis. Associated genes: genes identified due to nominal enrichment in patients. Candidate genes: epilepsy candidate genes deleted at least twice in patients and not deleted in the control cohort.
Neurogenetics

of each epilepsy subtype could be underestimated because our analysis was restricted only to high confidence calls and large microdeletions based on high-throughput genotyping. This study reports on a large spectrum of epilepsy types in three broad groupings, namely, GGE, AFE and RE. Each phenotypic group includes a wide range of heterogeneous subtypes, which have been grouped based on EEG pattern, age of onset and other clinical aspects. We cannot rule out that the individual microdeletion contribution does not follow this grouping and, for example, only a subgroup of patients with GGE (like JME) contributed to the association. Furthermore, our cohort size, in particular the RE cohort, is still too small to identify microdeletions with small disease risk. We also acknowledge the high heterogeneity of the AFE cohort (see definition in online supplementary information files) spatient cohort with focal epilepsy, we may miss specific subsyndromes with rare pathogenic events. Larger sample size, prospective studies and deep phenotyping will allow for the evaluation of other rare variants. Still, considering the above caveats, our results enable of us to elucidate the followings topics.

The landscape of microdeletions of overlapping and not overlapping hotspot loci is epilepsy subtype specific

In line with the original GGE study, the strongest combined enrichment was found for microdeletions overlapping with known microdeletion syndrome hotspot loci. The level of association did not increase with the inclusion of additional five hotspot loci to the analysis. These microdeletions are commonly found in patients with other highly comorbid traits such as ASD, ID and SCZ giving rise to a complex network of neurodevelopmental phenotypes. The specific effect of each of the hotspot loci evaluated has been difficult to determine, and the outcome of microdeletions overlapping these regions is not epilepsy specific.

Figure 2  Protein–protein interaction network of brain-expressed genes exclusively deleted in cases. Top panel: direct connection network obtained with for the combined epilepsy analysis. Bottom panel: direct connection network obtained with for the GGE subtype analysis. Each gene (node) in the network is coloured based on the significance of having more-than-expected edges (interactions) following the q value legend provided at the bottom-right corner.
When these regions are considered, subtype analyses show that for the GGE cohort, the strongest signal falls within genes overlapping with neurodevelopmental genes (as previously shown in Lal et al28). For RE in contrast, the frequency of microdeletions was similar to controls in all examined subsets, with the exception of a modest enrichment and a large OR observed for microdeletions overlapping ASD-related genes (online supplementary figure S2). Although the latter observations was not significant, we cannot rule out that future larger studies will identify a small but significant microdeletion burden. The AFE subtype analysis did not show enrichment for any of the investigated microdeletion subsets with hotspots included.

In this regard, we observed that the contribution of these regions is compelling and subtype specific. Notably, by removing hotspot regions from the analysis, the only remaining significant enrichment was for genes associated with neuronal development in the GGE subtype (figure 1).

The enrichment for GGE with hotspot loci has previously been shown to be more significant in patients with GGE and ID.36 The microdeletion burden for common epilepsy patients with normal social and intellectual skills (ie, high functioning) is expected to be modest in contrast to severe neurodevelopmental disorders with or without seizures.37 Accordingly, we cannot rule out that patients with GGE in our cohort with hotspot microdeletions represent those at the lower boundaries of the IQ distribution in the general population. It has previously been shown that CNVs at hotspot loci affect cognition in patients and controls.38 39 Given that the majority of the identified hotspot CNVs have been also associated with ID, well-phenotyped cohorts are needed in future studies to investigate whether seizures are also an independent consequence of these particular microdeletions.

Established disease and candidate genes only deleted in patients outside of microdeletion hotspots

Four patients carried microdeletions overlapping established developmental disorder genes, including SKI, KCNA2, GCH1 and DVL1. As an additional support for a haploinsufficiency mechanism for these genes, we reviewed LoF variants for these genes in >60 000 population controls (http://exac.broadinstitute.org) and found that all four genes are depleted for LoF variants. We identified nominal enrichment for the gene encoding the adhesion molecule neurexin 1 (NRXN1) and the splicing regulator RNA-binding protein fox-1 homolog (RBFOX1), which were previously implicated in epilepsy and neurodevelopmental disorders.22 25 26 40 41 To identify plausible candidate genes for epilepsy with potentially large effects, we selected genes exclusively overlapped by at least two microdeletions in patients with epilepsy. We identified 10 candidate genes at five loci (table 2). These autosomal microdeletions overlapped several genes previously implicated in epilepsy and neurodevelopmental disorders. Specifically, the genes encoding T-SNARE domain containing 1 (T3NARE1) and protocadherin 7 (PCDH7) have been highlighted previously in our GGE microdeletion analysis20 (1366 patients and 5234 controls), which was a subset of our present study. Furthermore, with the analysis of other types of epilepsies, we detected one patient with RE with a partial NRXN1 microdeletion and one patient with AFE with a complete PCDH7 deletion. These observations suggest their role as broader epilepsy risk factors rather than syndrome-specific variants of high effect. Furthermore, our gene-centric (compared with microdeletion-centric) analysis could narrow down four large microdeletions to PCDH7, PACRG, LOC102723362 and LOC101928137 as the only remaining genes not deleted in controls, respectively. We acknowledge the limitations of this analysis because we do not have the power to assign a meaningful p value to these detected genes.

Expression and network analysis

In the expression analysis of candidate genes, we did not observe significant brain tissue enrichment. We hypothesise that the number of included genes was small and possibly too heterogeneous. Although global or individual-gene brain expression patterns would have been informative, the results are not conclusive, and thus we cannot rule out candidates based on gene expression filtering. The network analysis resulted in significant interconnection for the brain-expressed genes exclusively deleted in the combined epilepsy cohort as well as in the GGE subgroup (figure 2). We do not observe significant interactions for the control gene list, which is similar in size compared with the combined and GGE subgroup lists. Our results expand previous observations on GGE20 and show that interconnections become more enriched when considering genes exclusively deleted in cases. The enriched KEGG long-term depression (hsa04730) and long-term potentiation (hsa04720) networks represent plausible neuronal enriched networks in patients with epilepsy.

In summary, we show that the microdeletion enrichment in patients with epilepsy includes genes involved in neurodevelopmental processes. Patients with GGE syndrome exhibit the highest microdeletion frequency, especially at hotspot loci. Apart from these loci, ultrarare heterogeneous deletions contribute significantly to GGE, whereas microdeletion frequency and distribution in AFE are indistinguishable from controls. The RE cohort is the smallest and therefore has the lowest statistical power for association discovery. However, the RE cohort shows nominal enrichment for hotspot loci microdeletions. There was some support for ultrarare microdeletions as plausible epilepsy candidate genes. Our study demonstrates that the contribution of microdeletions in common epilepsies is subtype specific. With increasing cohort sizes, the genetic architecture of the epilepsies and the contribution of microdeletions will become more evident.

Despite these differences, candidate genes can be found commonly deleted in more than one epilepsy type. Thus, the present findings contribute to our understanding of the structural genetic architecture of epilepsies and general and with respect to epilepsy subtypes.

Author affiliations

1Faculty of Biological Sciences and Medicine, Center for Biomedical Research, Universidad Andres Bello, Santiago, Chile
2Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany
3Department of Neuropediatrics, University Medical center Schleswig-Holstein, Kiel, Germany
4Division of Neurology, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
5Department of Neurology, Epilepsy Center Frankfurt Rhine-Main, Center of Neurology and Neurosurgery, University Hospital, Goethe-University Frankfurt, Frankfurt, Germany
6Department of Neurology, Epilepsy Center Hessen, University Hospitals Giessen & Marburg, and University of Marburg, Marburg, Germany
7Stanley Center for Psychiatric Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
8Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
9Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA
10Department of Neurology, Medical University of Vienna, Vienna, Austria
11Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA


