OPEN ACCESS

ORIGINAL ARTICLE

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 Lal[2,7,8]

For numbered affiliations see end of article.

Correspondence to
Dr Dennis Lal, Cologne Center for Genomics (CCG), University of Cologne, Stanley Center for Psychiatric Genetics, Broad Institute of MIT and Harvard, 185 Cambridge Street, Massachusetts General Hospital, Boston, MA 02114, USA; dlal@broadinstitute.org

CrossMark

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 hotspots 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published cohort of 1366 patients with genetic generalised epilepsy (GGE) in addition to two sets of additional unpublished genome-wide microdeletions found in 281 patients with rolandic epilepsy (RE) and 807 patients with adult focal epilepsy (AFE), totalling 2454 cases.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 $p=1.06\times10^{-6}$, 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 $p=9.79\times10^{-12}$, OR 7.45, 95% CI 4.20–13.5). Outside hotspots, microdeletions were enriched in the GGE cohort for neurodevelopmental genes (adjusted $p=9.13\times10^{-3}$, 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*, *RBFOX1* and *PCDH7*) and novel (*LOC102723362*) 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 (AFEs), the non-lesional genetic generalised epilepsies (GGEs) and the non-lesional childhood focal rolandic epilepsies (REs). 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 Noebels[3]). Twin studies have shown strong but differential concordance rates among epilepsies, including GGE and AFE with ~80% and ~40%, respectively.[4 5] 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.[7–9]

CNVs are genomic segments between 50 bp and 3 Mb, 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, >400 kb) and rare deletions (minor allele frequencyF <1%), are more likely to have damaging effects than duplications.[10 12] 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),[13] 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,[16] including seven CNVs associated with epilepsy, namely, 1q21.1, 15q11.2, 15q13.3, 16p11.2, 16p12, 16p13.11 and 22q11.2 with GGE,[17–20] and also 16p11.2 with RE and AFE.[21 22] 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[23] collectively named 'neurodevelopmental genes'.[20] 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] NRXN1[26] and GRIN2A[7 8] 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 cohort[20] with unpublished genome-wide microdeletions found in RE[21] and AFE[22] 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 detail[20–22] (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.[1] 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 study[21] and genotyped using the Human OmniExpress Bead-Chip 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 kb 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 microdeletions 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 syndromes[16] (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,[10]

Table 1 Data sets main features.

Epilepsy subtype	Code	Cases	Controls	Array
Genetic generalised epilepsy	GGE	1366	5234	Affymetrix GenomeWide 6.0
Rolandic epilepsy	RE	281	1512	Illumina Omni Express Array
Adult focal epilepsy	AFE	807	–	Illumina Omni Express Array
Total		2454	6746	9200

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[20] and defined based on literature and database queries[20 23] (no of genes=1559). For subset 6, 'ASD-related' genes were extracted from Uddin *et al*[30 30] and defined as those enriched for deleterious exonic de novo mutations in comparison with healthy siblings (no of genes=1683). 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[31] 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[32 32] 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=2506). 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 data of 42 tissues from 1561 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).[33] 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[34] 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 microdeletions subsets: (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.[16] 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.[32]

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[20] with the genome-wide microdeletions identified in RE[21] and AFE[22] 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 $p=1.60 \times 10^{-6}$, 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 $p=1.59 \times 10^{-11}$, 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.[20] 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 $p=9.79 \times 10^{-12}$, OR 7.45, 95% CI 4.21 to 13.5) and overlapping neurodevelopmental genes as previously shown by Lal *et al*.[20] Interestingly, microdeletions overlapping neurodevelopmental genes not overlapping hotspot loci remained significant (adjusted $p=9.13 \times 10^{-3}$, 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
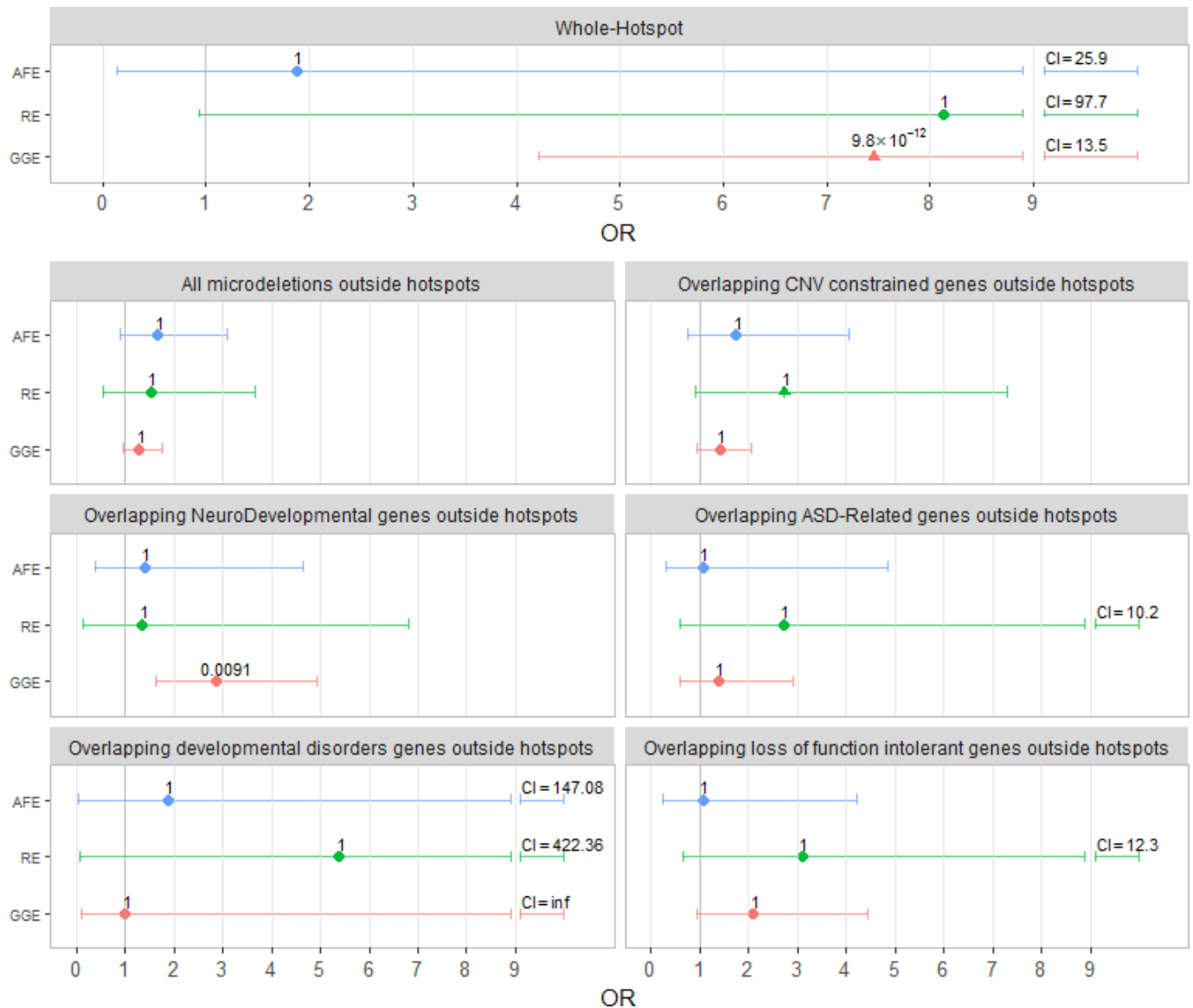
Figure 1 Microdeletion burden analysis in common epilepsy types. For adult focal epilepsies (AFEs, cyan), rolandic epilepsies (RE, green) and genetic generalised epilepsies (GGEs, red) data sets, the burden analysis without genomic rearrangement hotspots loci consideration[16] is shown using the following microdeletion sets: all microdeletions; overlapping at least one: CNV constrained gene, neurodevelopmental gene,[23] autism spectrum disorder (ASD)-related gene,[30] developmental disorders genes[31] and loss-of-function intolerant genes.[32] The effect size observed (OR), CI (horizontal lines) and multiple testing corrected p value obtained is shown for each data set. Triangles denote if the signal is nominally significant. CI values above 9 are shown in numbers.

hotspot loci, but this did not survive multiple testing correction (figure 1; nominal p=0.029, adjusted p=1, OR 8.13, 95% CI 0.92 to 97.79). Patients with AFE did not show significant differences with control samples within hotspot loci (p=1, OR 2.85, 95% CI 0.13 to 25.9). Furthermore, RE and AFE samples were not enriched for any of the microdeletion subsets interrogated outside hotspot loci. Sample count of all the microdeletion subsets evaluated is provided in online supplementary table S3.

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 al[20]), 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 25] 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*

Table 2 Candidate gene identification outside hotspot loci

Nominally associated genes	Chr	Start	End	Patients (n=2454)	Controls (n=6746)	Nominal p value	Constrained gene*
RBFOX1	16	6 069 131	7 763 340	3 (AFE, AFE, GGE)	0	0.019	No
NRXN1	2	50 145 642	51 259 674	3 (GGE, GGE, RE)	0	0.019	No
Candidate genes							
LOC102723362	2	22 759 350	22 761 159	2 (GGE, RE)	0	0.0711	Yes
PCDH7	4	30 722 036	31 148 423	2 (AFE, GGE)	0	0.0711	No
PACRG	6	163 148 164	163 736 524	2 (GGE, AFE)	0	0.0711	Yes
ADGRB1, TSNARE1, LINC00051, MIR1302-7, MIR4472-1, MIR4539	8	142 867 603	143 626 368	2 (GGE, GGE)	0	0.0711	No, No, Yes, Yes, Yes, No
LOC101928137	12	73 552 969	73 602 097	2 (GGE, GGE)	0	0.0711	Yes
DDG2P genes							
SKI, DVL1	1	1 270 658	2 241 652	1 (AFE)	0	0.267	Yes, No
KCNA2	1	111 136 202	111 174 096	1 (GGE)	0	0.267	Yes
GCH1	14	55 308 724	55 369 542	1 (RE)	0	0.267	No

*Constrained gene according to the CNV map of healthy individuals.[10]
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.

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 278 594, 260 750 and 249 773, 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 individuals[10] 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 once 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 DAPPLE[35] 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 epilepsy cohort (GGE+RE+AFE) plus 77, 14 and 53 genes exclusively deleted in GGE, RE and 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 cases 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.[20] 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
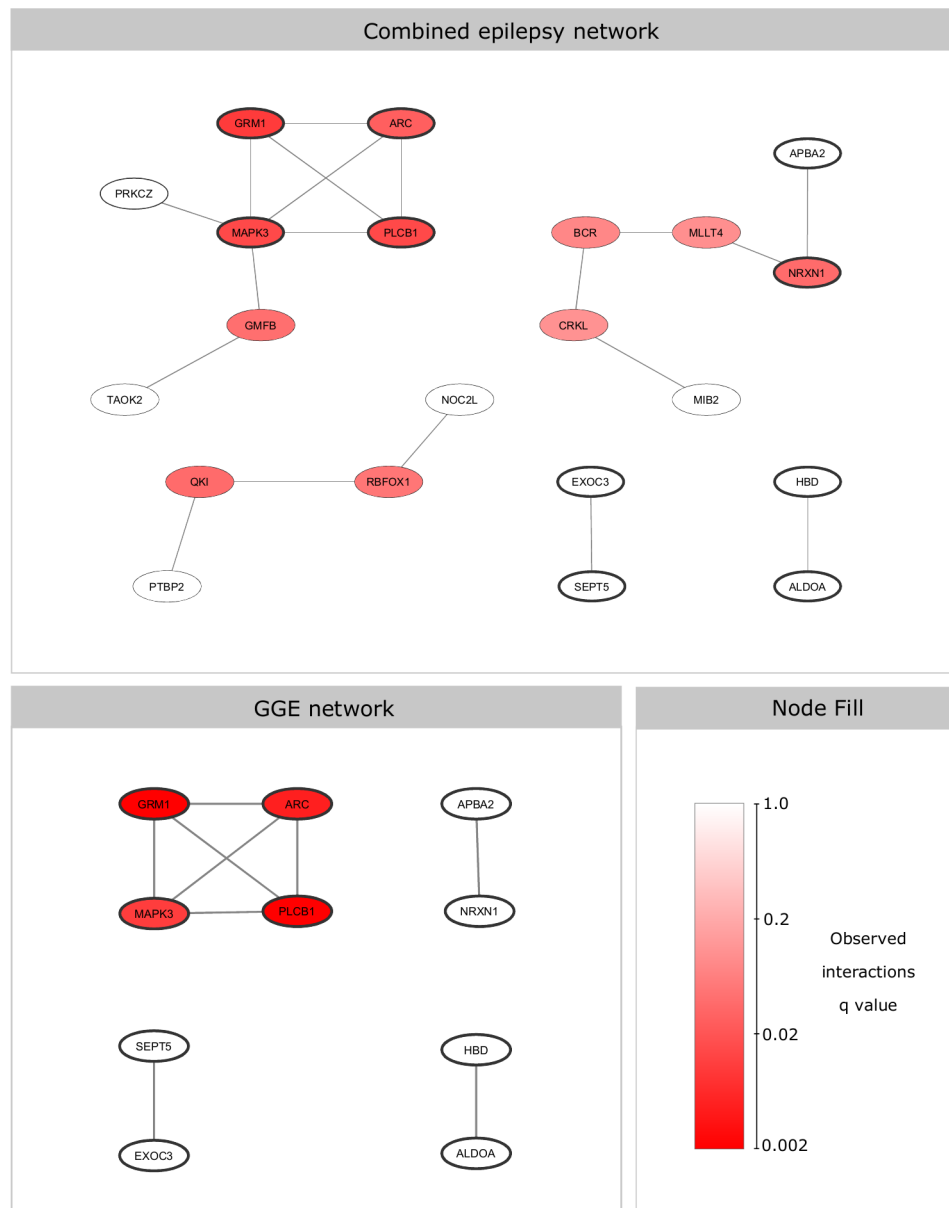
Figure 2   Protein–protein interaction network of brain-expressed genes exclusively deleted in cases. Top panel: direct connection network obtained with for the combined epilepsy analysis. Bottom panel: direct connection network obtained with for the GGE subtype analysis. Each gene (node) in the network is coloured based on the significance of having more-than-expected edges (interactions) following the q value legend provided at the bottom-right corner.

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,[20] 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.[16 18] 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.

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 al[20]). 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 haploinsufficieny 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 (*TSNARE1*) and protocadherin 7 (*PCDH7*) have been highlighted previously in our GGE microdeletion analysis[20] (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 GGE[20] 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
[1]Faculty of Biological Sciences and Medicine, Center for Biomedical Research, Universidad Andres Bello, Santiago, Chile
[2]Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany
[3]Department of Neuropediatrics, University Medical Center Schleswig-Holstein, Kiel, Germany
[4]Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[5]Department of Neurology, Epilepsy Center Frankfurt Rhine-Main, Center of Neurology and Neurosurgery, University Hospital, Goethe-University Frankfurt, Frankfurt, Germany
[6]Department of Neurology, Epilepsy Center Hessen, University Hospitals Giessen & Marburg, and University of Marburg, Marburg, Germany
[7]Stanley Center for Psychiatric Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[8]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
[9]Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA
[10]Department of Neurology, Medical University of Vienna, Vienna, Austria
[11]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[12]Department of Neuropathology, University of Bonn, Bonn, Germany
[13]Division of Medical Genetics Department of Biomedicine, University of Basel, Basel, Switzerland
[14]Departamento de Gastroenterología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile
[15]Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research,University of Tübingen, Tübingen, Germany
[16]Department of Neuropediatrics, University Medical Center Giessen and Marburg, Giessen, Germany
[17]C. Mondino National Neurological Institute, Pavia, Italy
[18]Department of Internal Medicine and Therapeutics, University of Pavia, Pavia, Italy
[19]Laboratory of Neurogenetics, Neuromuscular Disease Unit, Genova, Italy

REFERENCES

1 Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, Engel J, French J, Glauser TA, Mathern GW, Moshé SL, Nordli D, Plouin P, Scheffer IE. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on classification and terminology, 2005-2009. *Epilepsia* 2010;51:676–85.

2 Banerjee PN, Filippi D, Allen Hauser W. The descriptive epidemiology of epilepsy-a review. *Epilepsy Res* 2009;85:31–45.

3 Noebels JL. Single-Gene determinants of epilepsy comorbidity. *Cold Spring Harb Perspect Med* 2015;5:a022756.

4 Berkovic SF, Howell RA, Hay DA, Hopper JL. Epilepsies in twins: genetics of the major epilepsy syndromes. *Ann Neurol* 1998;43:435–45.

5 Vadlamudi L, Milne RL, Lawrence K, Heron SE, Eckhaus J, Keay D, Connellan M, Torn-Broers Y, Howell RA, Mulley JC, Scheffer IE, Dibbens LM, Hopper JL, Berkovic SF. Genetics of Epilepsy: the testimony of twins in the molecular era. *Neurology* 2014;83:1042–8.

6 Vadlamudi L, Kjeldsen MJ, Corey LA, Solaas MH, Friis ML, Pellock JM, Nakken KO, Milne RL, Scheffer IE, Harvey AS, Hopper JL, Berkovic SF. Analyzing the etiology of benign rolandic epilepsy: a multicenter twin collaboration. *Epilepsia* 2006;47:550–5.

7 Lemke JR, Lal D, Reinthaler EM, Steiner I, Nothnagel M, Alber M, Geider K, Laube B, Schwake M, Finsterwalder K, Franke A, Schilhabel M, Jähn JA, Muhle H, Boor R, Van Paesschen W, Caraballo R, Fejerman N, Weckhuysen S, De Jonghe P, Larsen J, Møller RS, Hjalgrim H, Addis L, Tang S, Hughes E, Pal DK, Veri K, Vaher U, Talvik T, Dimova P,

Guerrero López R, Serratosa JM, Linnankivi T, Lehesjoki AE, Ruf S, Wolff M, Buerki S, Wohlrab G, Kroell J, Datta AN, Fiedler B, Kurlemann G, Kluger G, Hahn A, Haberlandt DE, Kutzer C, Sperner J, Becker F, Weber YG, Feucht M, Steinböck H, Neophythou B, Ronen GM, Gruber-Sedlmayr U, Geldner J, Harvey RJ, Hoffmann P, Herms S, Altmüller J, Toliat MR, Thiele H, Nürnberg P, Wilhelm C, Stephani U, Helbig I, Lerche H, Zimprich F, Neubauer BA, Biskup S, von Spiczak S. Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nat Genet* 2013;45:1067–72.

8 Carvill GL, Regan BM, Yendle SC, O'Roak BJ, Lozovaya N, Bruneau N, Burnashev N, Khan A, Cook J, Geraghty E, Sadleir LG, Turner SJ, Tsai MH, Webster R, Ouvrier R, Damiano JA, Berkovic SF, Shendure J, Hildebrand MS, Szepetowski P, Scheffer IE, Mefford HC. GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat Genet* 2013;45:1073–6.

9 Lesca G, Rudolf G, Bruneau N, Lozovaya N, Labalme A, Boutry-Kryza N, Salmi M, Tsintsadze T, Addis L, Motte J, Wright S, Tsintsadze V, Michel A, Doummar D, Lascelles K, Strug L, Waters P, de Bellescize J, Vrielynck P, de Saint Martin A, Ville D, Ryvlin P, Arzimanoglou A, Hirsch E, Vincent A, Pal D, Burnashev N, Sanlaville D, Szepetowski P. GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat Genet* 2013;45:1061–6.

10 Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet* 2015;16:172–83.

11 Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME; Wellcome Trust Case Control Consortium. Origins and functional impact of copy number variation in the human genome. *Nature* 2010;464:704–12.

12 Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, Coe BP, Baker C, Nordenfelt S, Bamshad M, Jorde LB, Posukh OL, Sahakyan H, Watkins WS, Yepiskoposyan L, Abdullah MS, Bravi CM, Capelli C, Hervig T, Wee JT, Tyler-Smith C, van Driem G, Romero IG, Jha AR, Karachanak-Yankova S, Toncheva D, Comas D, Henn B, Kivisild T, Ruiz-Linares A, Sajantila A, Metspalu E, Parik J, Villems R, Starikovskaya EB, Ayodo G, Beall CM, Di Rienzo A, Hammer MF, Khusainova R, Khusnutdinova E, Klitz W, Winkler C, Labuda D, Metspalu M, Tishkoff SA, Dryomov S, Sukernik R, Patterson N, Reich D, Eichler EE, diversity G. Global diversity, population stratification, and selection of human copy-number variation. *Science* 2015;349:aab3761.

13 Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SF, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009;459:528–33.

14 Tam GW, Redon R, Carter NP, Grant SG. The role of DNA copy number variation in schizophrenia. *Biol Psychiatry* 2009;66:1005–12.

15 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI, Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;466:368–72.

16 Carvill GL, Mefford HC. Microdeletion syndromes. *Curr Opin Genet Dev* 2013;23:232–9.

17 Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C, Baker C, von Spiczak S, Kron KL, Steinich I, Kleefuss-Lie AA, Leu C, Gaus V, Schmitz

B, Klein KM, Reif PS, Rosenow F, Weber Y, Lerche H, Zimprich F, Urak L, Fuchs K, Feucht M, Genton P, Thomas P, Visscher F, de Haan GJ, Møller RS, Hjalgrim H, Luciano D, Wittig M, Nothnagel M, Elger CE, Nürnberg P, Romano C, Malafosse A, Koeleman BP, Lindhout D, Stephani U, Schreiber S, Eichler EE, Sander T. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet* 2009;41:160–2.

18 de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, Kluck C, Muhle H, von Spiczak S, Ostertag P, Obermeier T, Kleefuss-Lie AA, Hallmann K, Steffens M, Gaus V, Klein KM, Hamer HM, Rosenow F, Brilstra EH, Trenité DK, Swinkels ME, Weber YG, Unterberger I, Zimprich F, Urak L, Feucht M, Fuchs K, Møller RS, Hjalgrim H, De Jonghe P, Suls A, Rückert IM, Wichmann HE, Franke A, Schreiber S, Nürnberg P, Elger CE, Lerche H, Stephani U, Koeleman BP, Lindhout D, Eichler EE, Sander T. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* 2010;133(Pt 1):23–32.

19 Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, Franke A, Malafosse A, Genton P, Thomas P, Gurnett CA, Schreiber S, Bassuk AG, Guipponi M, Stephani U, Helbig I, Eichler EE. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet* 2010;6:e1000962.

20 Lal D, Ruppert AK, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenité D, Sonsma AC, Koeleman BP, Lindhout D, Weber YG, Lerche H, Kapser C, Schankin CJ, Kunz WS, Surges R, Elger CE, Gaus V, Schmitz B, Helbig I, Muhle H, Stephani U, Klein KM, Rosenow F, Neubauer BA, Reinthaler EM, Zimprich F, Feucht M, Møller RS, Hjalgrim H, De Jonghe P, Suls A, Lieb W, Franke A, Strauch K, Gieger C, Schurmann C, Schminke U, Nürnberg P, Sander T; EPICURE Consortium. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. *PLoS Genet* 2015;11:e1005226.

21 Reinthaler EM, Lal D, Lebon S, Hildebrand MS, Dahl HH, Regan BM, Feucht M, Steinbock H, Neophytou B, Ronen GM, Roche L, Gruber-Sedlmayr U, Geldner J, Haberlandt E, Hoffmann P, Herms S, Gieger C, Waldenberger M, Franke A, Wittig M, Schoch S, Becker AJ, Hahn A, Mannik K, Toliat MR, Winterer G. p11.2 European C, Lerche H, Nurnberg P, mefford H, Scheffer IE, Berkovic SF, Beckmann JS, Consortium E, Euro EC, Sander T, Jacquemont S, Reymond A, Zimprich F, Neubauer BA. 16p11.2 600 kb duplications confer risk for typical and atypical rolandic epilepsy. *Human molecular genetics* 2014;23:6069–80.

22 Lal D, Pernhorst K, Klein KM, Reif P, Tozzi R, Toliat MR, Winterer G, Neubauer B, Nürnberg P, Rosenow F, Becker F, Lerche H, Kunz WS, Kurki MI, Hoffmann P, Becker AJ, Perucca E, Zara F, Sander T, Weber YG. Extending the phenotypic spectrum of *RBFOX1* deletions: sporadic focal epilepsy. *Epilepsia* 2015;56:e129–e133.

23 Krumm N, O'Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, Jacquemont S, Munson J, Bernier R, Eichler EE. Transmission disequilibrium of small CNVs in simplex autism. *Am J Hum Genet* 2013;93:595–606.

24 Lal D, Trucks H, Møller RS, Hjalgrim H, Koeleman BP, de Kovel CG, Visscher F, Weber YG, Lerche H, Becker F, Schankin CJ, Neubauer BA, Surges R, Kunz WS, Zimprich F, Franke A, Illig T, Ried JS, Leu C, Nürnberg P, Sander T, Consortium EM; EMINet ConsortiumEPICURE Consortium. Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. *Epilepsia* 2013;54:265–71.

25 Lal D, Reinthaler EM, Altmüller J, Toliat MR, Thiele H, Nürnberg P, Lerche H, Hahn A, Møller RS, Muhle H, Sander T, Zimprich F, Neubauer BA. RBFOX1 and RBFOX3 mutations in rolandic epilepsy. *PLoS One* 2013;8:e73323.

26 Møller RS, Weber YG, Klitten LL, Trucks H, Muhle H, Kunz WS, Mefford HC, Franke A, Kautza M, Wolf P, Dennig D, Schreiber S, Rückert IM, Wichmann HE, Ernst JP, Schurmann C, Grabe HJ, Tommerup N, Stephani U, Lerche H, Hjalgrim H, Helbig I, Sander T, Consortium E; EPICURE Consortium. Exon-disrupting deletions of NRXN1 in idiopathic generalized epilepsy. *Epilepsia* 2013;54:256–64.

27 Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated Hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17:1665–74.

28 Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, Macdonald JR, Mills R, Prasad A, Noonan K, Gribble S, Prigmore E, Donahoe PK, Smith RS, Park JH, Hurles ME, Carter NP, Lee C, Scherer SW, Feuk L. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 2011;29:512–20.

29 Team RR. *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*, 2010. (01/19).

30 Uddin M, Tammimies K, Pellecchia G, Alipanahi B, Hu P, Wang Z, Pinto D, Lau L, Nalpathamkalam T, Marshall CR, Blencowe BJ, Frey BJ, Merico D, Yuen RK, Scherer SW. Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat Genet* 2014;46:742–7.

31 Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl Resources. *Am J Hum Genet* 2009;84:524–33.

32 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285-91.

33 Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z, Vorstman JA, Thompson A, Regan R, Pilorge M, Pellecchia G, Pagnamenta AT, Oliveira B, Marshall CR, Magalhaes TR, Lowe JK, Howe JL, Griswold AJ, Gilbert J, Duketis E, Dombroski BA, De Jonge MV, Cuccaro M, Crawford EL, Correia CT, Conroy J, Conceição IC, Chiocchetti AG, Casey JP, Cai G, Cabrol C, Bolshakova N, Bacchelli E, Anney R, Gallinger S, Cotterchio M, Casey G, Zwaigenbaum L, Wittemeyer K, Wing K, Wallace S, van Engeland H, Tryfon A, Thomson S, Soorya L, Rogé B, Roberts W, Poustka F, Mouga S, Minshew N, McInnes LA, McGrew SG, Lord C, Leboyer M, Le Couteur AS, Kolevzon A, Jiménez González P, Jacob S, Holt R, Guter S, Green J, Green A, Gillberg C, Fernandez BA, Duque F, Delorme R, Dawson G, Chaste P, Café C, Brennan S, Bourgeron T, Bolton PF, Bölte S, Bernier R, Baird G, Bailey AJ, Anagnostou E, Almeida J, Wijsman EM, Vieland VJ, Vicente AM, Schellenberg GD, Pericak-Vance M, Paterson AD, Parr JR, Oliveira G, Nurnberger JI, Monaco AP, Maestrini E, Klauck SM, Hakonarson H, Haines JL, Geschwind DH, Freitag CM, Folstein SE, Ennis S, Coon H, Battaglia A, Szatmari P, Sutcliffe JS, Hallmayer J, Gill M, Cook EH, Buxbaum JD, Devlin B, Gallagher L, Betancur C, Scherer SW. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014;94:677–94.

34 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.

35 Trost B, Arsenault R, Griebel P, Napper S, Kusalik A. DAPPLE: a pipeline for the homology-based prediction of phosphorylation sites. *Bioinformatics* 2013;29:1693–5.

36 Mullen SA, Carvill GL, Bellows S, Bayly MA, Trucks H, Lal D, Sander T, Berkovic SF, Dibbens LM, Scheffer IE, Mefford HC. Copy number variants are frequent in genetic generalized epilepsy with intellectual disability. *Neurology* 2013;81:1507–14.

37 Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. A copy number variation morbidity map of developmental delay. *Nat Genet* 2011;43:838–46.

38 D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L, Hanson E, Maillard AM, Faucett WA, Mace A, Pain A, Bernier R, Chawner SJ, David A, Andrieux J, Aylward E, Baujat G, Caldeira I, Conus P, Ferrari C, Forzano F, Gerard M, Goin-Kochel RP, Grant E, Hunter JV, Isidor B, Jacquette A, Jonch AE, Keren B, Lacombe D, Le Caignec C, Martin CL, Mannik K, Metspalu A, Mignot C, Mukherjee P, Owen MJ, Passeggeri M, Rooryck-Thambo C, Rosenfeld JA, Spence SJ, Steinman KJ, Tjernagel J, Van Haelst M, Shen Y, Draganski B, Sherr EH, Ledbetter DH, van den Bree MB, Beckmann JS, Spiro JE, Reymond A, Jacquemont S, Chung WK. Cardiff University Experiences of Children With Copy Number Variants S,. p11.2 European C, Simons variation in individuals Project C. defining the effect of the 16p11.2 Duplication on cognition, behavior, and medical comorbidities. *JAMA psychiatry* 2016;73:20–30.

39 Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, Bjornsdottir G, Walters GB, Jonsdottir GA, Doyle OM, Tost H, Grimm O, Kristjansdottir S, Snorrason H, Davidsdottir SR, Gudmundsson LJ, Jonsson GF, Stefansdottir B, Helgadottir I, Haraldsson M, Jonsdottir B, Thygesen JH, Schwarz AJ, Didriksen M, Stensbøl TB, Brammer M, Kapur S, Halldorsson JG, Hreidarsson S, Saemundsen E, Sigurdsson E, Stefansson K. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 2014;505:361–6.

40 Lal D, Pernhorst K, Klein KM, Reif P, Tozzi R, Toliat MR, Winterer G, Neubauer B, Nürnberg P, Rosenow F, Becker F, Lerche H, Kunz WS, Kurki MI, Hoffmann P, Becker AJ, Perucca E, Zara F, Sander T, Weber YG. Extending the phenotypic spectrum of RBFOX1 deletions: sporadic focal epilepsy. *Epilepsia* 2015;56:e129–e133.

41 Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, Lapointe M, Spiegelman D, Noreau A, Lafrenière RG, Fathalli F, Joober R, Krebs MO, DeLisi LE, Mottron L, Fombonne E, Michaud JL, Drapeau P, Carbonetto S, Craig AM, Rouleau GA. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum Genet* 2011;130:563–73.