RESEARCH ARTICLE

# Assessment of burden and segregation profiles of CNVs in patients with epilepsy

Claudia Moreau[1] iD, Frédérique Tremblay[1], Stefan Wolking[2] iD, Alexandre Girard[1],
Catherine Laprise[1] iD, Fadi F. Hamdan[3,4] iD, Jacques L. Michaud[3,5], Berge A. Minassian[6,7] iD,
Patrick Cossette[8,9] & Simon L. Girard[1,10] iD

[1]Department of Fundamental Sciences, University of Quebec in Chicoutimi, Chicoutimi, Canada
[2]Department of Neurology and Epileptology, University Hospital RWTH Aachen, Aachen, Germany
[3]CHU Sainte-Justine Research Center, Montreal, Canada
[4]Department of Pediatrics, University of Montreal, Montreal, Canada
[5]Department of Neurosciences and Department of Pediatrics, University of Montreal, Montreal, Canada
[6]Department of Pediatrics, Hospital for Sick Children and University of Toronto, Toronto, Canada
[7]Department of Pediatrics, University of Texas Southwestern, Dallas, Texas, USA
[8]CHUM Research Center, Montreal, Canada
[9]Department of Neurosciences, University of Montreal, Montreal, Canada
[10]CERVO Research Center, Laval University, Quebec, Canada

**Correspondence**
Simon L. Girard, Département des sciences fondamentales, Université du Québec à Chicoutimi, Office P4-2130, 555, boulevard de l'Université, Chicoutimi, Québec G7H 2B1, Canada. Tel: +1 418 545-5011 ext 2595; Fax: +1 800 463-9880 ext 2595 (toll free); E-mail: simon2_girard@uqac.ca

## Abstract

**Objective:** Microdeletions are associated with different forms of epilepsy but show incomplete penetrance, which is not well understood. We aimed to assess whether unmasked variants or double CNVs could explain incomplete penetrance. **Methods:** We analyzed copy number variants (CNVs) in 603 patients with four different subgroups of epilepsy and 945 controls. CNVs were called from genotypes and validated on whole-genome (WGS) or whole-exome sequences (WES). CNV burden difference between patients and controls was obtained by fitting a logistic regression. CNV burden was assessed for small and large (>1 Mb) deletions and duplications and for deletions overlapping different gene sets. **Results:** Large deletions were enriched in genetic generalized epilepsies (GGE) compared to controls. We also found enrichment of deletions in epilepsy genes and hotspots for GGE. We did not find truncating or functional variants that could have been unmasked by the deletions. We observed a double CNV hit in two patients. One patient also carried a de novo deletion in the 22q11.2 hotspot. **Interpretation:** We could corroborate previous findings of an enrichment of large microdeletions and deletions in epilepsy genes in GGE. We could also replicate that microdeletions show incomplete penetrance. However, we could not validate the hypothesis of unmasked variants nor the hypothesis of double CNVs to explain the incomplete penetrance. We found a de novo CNV on 22q11.2 that could be of interest. We also observed GGE families carrying a deletion on 15q13.3 hotspot that could be investigated in the Quebec founder population.

# Introduction

Epilepsy has a prevalence of ~3% and a high socio-economic burden.[1] About half of the affected individuals experience the first seizures during childhood. About 30%–40% of epilepsy syndromes are thought to have a genetic background. Yet, monogenic forms of the disease are rare[2–5] and represent less than 2% of all cases. The larger share of genetic epilepsy syndromes is thought to be polygenic, which has been substantiated by large-scale genetic studies in the past years.[6,7]

Copy number variants (CNVs) are implicated in the etiology of epilepsy, especially in developmental epileptic encephalopathies (DEE) and genetic generalized epilepsies (GGE).[8–19] These rare CNVs are either occurring at new sites or at genomic hotspots. Most studies on CNVs in epilepsy focused on microdeletions, although microduplications have also been reported in some cases.[20,21] Moreover, except for non-acquired focal epilepsy (NAFE), large CNVs (generally larger than 1 Mb) are significantly enriched in individuals with epilepsy compared to controls.[11,22–24]

The genetic mechanisms by which these CNVs could cause epilepsy or other developmental disorders remain unclear. In the case of microdeletions, several mechanisms have been proposed to explain their incomplete penetrance, including the unmasking of a recessive allele,[25] a non-coding regulatory variant present in the deletion region[26] or the presence of a second large CNV that could contribute to a more severe phenotype.[27] The advent of whole-genome sequencing (WGS) makes it possible to address these hypotheses more systematically.

Here, we investigated CNVs as well as deletions in different sets of genes. The burden of CNVs was assessed in individuals with epilepsy, their unaffected family members and population controls using whole-genome genotyping data. The patients and controls were mostly derived from the Quebec founder population.[28] This could maximize our odds of identifying events that would be deemed rare or very rare in populations without founder effect.[29,30] Our extensive familial data collection was used to check for segregation and variant dissemination in larger familial clusters. In addition, we validated the identified large microdeletions and analyzed the homologous chromosome for unmasked variants that could explain the reduced penetrance in patients with WGS or whole-exome (WES) sequencing data.

# Subjects and Methods

This study was approved by the CHUM Research Center (CRCHUM) ethics committee and by the University of Quebec in Chicoutimi ethics board. Written informed consent was obtained from all patients (or their legal guardians for patients under 18) and adult controls.

## Phenotyping

The epilepsy cohort was composed of extended families comprising affected and unaffected individuals with GGE or NAFE as well as DEE trios with unaffected parents previously collected in CHUM Research Center and CHU Ste-Justine in Montreal and in the Hospital for Sick Children in Toronto as part of the Canadian Epilepsy Network (CENet) and diagnosed by neurologists. The clinical epilepsy phenotype was classified according to the current classification by the International League against Epilepsy (ILAE).[31] Detailed phenotyping is reported in Moreau et al.[28] Certain cases were found with an epilepsy phenotype different from the other affected family members (families marked as "mixed"). The unaffected GGE and NAFE family members and DEE trio parents were used as familial controls in addition to French-Canadian controls from the Quebec Reference Sample.[32]

## Genotyping

Samples were processed on either the Illumina Omni Express (n.SNVs = 710,000) or the Illumina Omni 2.5 (n.SNVs = 2,500,000 including the Omni Express core). Genotypes of all samples were merged and only positions present on both chips were kept. We further removed SNVs with more than 2% missing sites over all individuals and with HWE p-value <0.001 using PLINK software[33] as well as individuals with more than 2% missing SNVs. Individuals with ambiguous sex were removed from the analysis.

## CNV calling and filtering and batch correction

A file was generated by the Genome Quebec Innovation Center in Montreal for each genotyped sample including Log-R ratio (LRR) and B allele frequency (BAF) for all SNVs. PennCNV software[34] was used for CNV calling. Only filtered SNVs were used to generate a custom population B-allele frequency file before calling CNVs. First CNV calling (−-qclrrsd 0.3 --qcbafdrift 0.01 --qcwf 0.05) was performed to remove low-quality samples, then principal components analysis and batch correction (PC-correction) was applied to LRR as described in Cooper et al.[35] using filtered SNPs outside of telomeric, centromeric, and immunoglobulin regions (Fig. S1). Second, CNV calling was performed on the corrected LRR using --qclrrsd 0.3 --qcbafdrift 0.01 --qcwf 0.05 --numsnp 10 --length 20 k --qcnumcnv 50, telomeric, centromeric, and

immunoglobulin regions were removed and CNVs were merged using default fraction argument of 0.2. Total number of samples, males and females after QC in addition to available WGS and WES are presented in (Table 1). CNVs were also called on 135 complete DEE, GGE, NAFE, or mixed trios to look for de novo CNV hits. 4460 CNVs were called for 1548 samples.

We only considered rare CNVs (<=1%) for further analyses. There were 2698 such CNVs in our dataset (Table S1). The CNV frequency was obtained using PLINK[33] v1.07 –cnv-freqmethod2 0.5 option.

## CNV validation

CNVs were validated using either whole-genome (WGS) or whole-exome (WES) depending on the availability of such sequences and/or segregation in the family. For segregation, CNVs were considered as being the same if they overlapped at least 50%. Duplications and deletions were considered separately. Detailed sequencing methods for WGS and WES are described in Moreau et al.[28] and in Wolking et al.,[36] respectively. CNVs on WGS and WES were called using two software, CNVkit[37] and Control-FREEC.[38] A CNV was considered as validated if called by one of these software and overlap at least 50%. We did not consider a CNV as validated if the length of the WGS or WES call was more than twice the length of the genotyping call to avoid spurious calls.

## CNV annotation

PennCNV was used to determine if CNVs were spanning genes (hg19). A CNV was considered to be in the coding region if it overlapped at least 80% of a gene. We also identified 152 genes that were previously associated with epilepsy[39,40] and 1804 genes intolerant for protein-truncating variants defined as probability of loss-of-function (lof) intolerance (pLI) score > 0.99. We also looked for CNVs overlapping epilepsy hotspots previously identified in epileptic patients[41] (Table S2). A CNV was considered to be in a hotspot if it overlapped at least 50% of a hotspot.

## CNV burden

We measured CNV burden for all epilepsy phenotypes for small and large (>1 Mb) rare deletions and duplications separately to evaluate relative contribution on epilepsy type risk. We also looked at rare deletions overlapping genes, epilepsy-associated genes, genes with pLi >0.99 and known epilepsy hotspots (Table S2). To assess for a CNV burden difference between epilepsy cases and controls, we fitted a logistic binomial regression model with sex as covariate using the geekin function of the MESS package (https://cran.r-project.org/web/packages/MESS/index. html) to account for familial relationships. The familial relationships were obtained using PLINK –genome option after pruning. For all burden analyses, odds ratios, 95% confidence intervals (CIs), and significance were calculated by taking the exponential of the logistic regression coefficients. We removed the unaffected DEE parents from the DEE burden analyses. Bonferroni multiple-testing was calculated for 16 tests for both groups of analyses and threshold for significance was 0.003.

## Variant calling and annotation

SNVs in microdeletions were called WGS or WES. We performed joint calling of gvcf files that were merged into a single vcf file using GATK version 3.7-0 (https://gatk. broadinstitute.org/hc/en-us). The vcf file was recalibrated and filtered following the GATK best practice guidelines. SnpEff and SnpSift[42,43] were used to annotate SNVs. An SNV was considered to have a lof or nonsense-mediated mRNA decay (nmd) effect if this effect was seen in more than 90% of the transcripts. All missense variants were considered. We cross-referenced these SNVs in ClinVar (version of June 9th, 2021)[44] to identify known pathogenic variants. To assess whether non-coding SNVs could have a functional effect, we used ExPecto,[45] a deep learning algorithm that computes the tissue-specific effect of variants on gene expression using WGS and WES (although WES are not expected to include many non-coding variants). The computed expression fold change resulting from Expecto analysis was used to identify deleterious variants. We calculated a variation potential

**Table 1.** Number of individuals in each group.

| Phenotype | Samples | Trio/fam | Females | Males | WGS | WES |
|---|---|---|---|---|---|---|
| GGE | 349 | 247 | 218 | 131 | 107 | 130 |
| NAFE | 165 | 138 | 84 | 81 | 94 | 35 |
| Mixed | 30 | 28 | 8 | 22 | 10 | 3 |
| DEE trio patients | 59 | 59 | 21 | 38 | 59 | 0 |
| Unaffected DEE trio parents | 118 | 59 | 59 | 59 | 116 | 0 |
| Unaffected familial ctrls (GGE and NAFE families) | 283 | 107 | 152 | 131 | 0 | 0 |
| Population ctrls | 544 | NA | 293 | 251 | 0 | 0 |

GGE = genetic generalized epilepsies; NAFE = non-acquired focal epilepsy; Mixed = cases with an epilepsy phenotype different from the other affected family members; DEE = developmental epileptic encephalopathies; ctrls = controls.

directionality score for each gene for three tissues related to epilepsy (amygdala, cortex, and hippocampus). Then a constraint violation score was obtained by computing the product of the variation potential directionality score and the predicted expression change for a given SNV. The higher this score is, the more deleterious the SNV is.
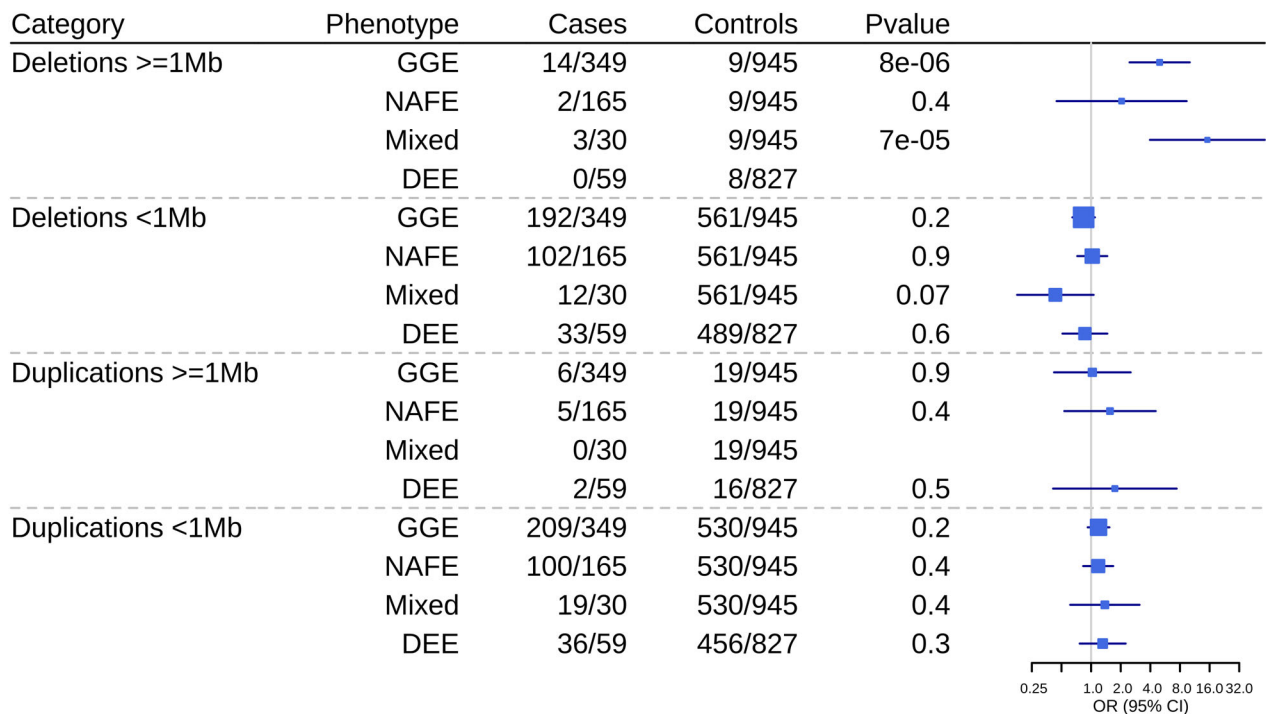
## Results

Burden analysis revealed a greater proportion of deletions >1 Mb in GGE individuals resulting in significant OR (4.97; 95% CI 2.5–10.1) against controls (Fig. 1). Moreover, we also observed an excess of large deletions compared to large duplications in GGE. No such proportions were observed for duplications. Since microdeletions seem more important in epilepsy,[24] burden analysis was performed only on deletions for different gene sets and epilepsy hotspots (Fig. 2). We found an enrichment of deletions in epilepsy genes and hotspots for GGE compared to controls (Tables S3 and S4 for detailed CNV description).

We further analyzed individuals carrying a large deletion (Table 2 and Table S5). All deletions found in patients were located within a gene whereas only 30% of the deletions among the population controls were located in the coding region. Half of the large deletions were

found in an epilepsy gene or hotspot (Table 2). All 12 deletions for which we had WGS or WES in addition to genotypes were validated. The remaining deletions were validated by looking at the segregation in the family. Almost all validated deletions for which we had family information were transmitted either by an affected (four transmissions + one plausible transmission) or an unaffected parent (four transmissions) (Table S5). Only one de novo large deletion could not be validated because the patient did not have WGS data. As reported previously, we document here several cases where known pathological hotspots CNV were either transmitted from an unaffected family member or to a yet unaffected sibling warranting the need to be cautious when using these findings in clinical settings.

Among individuals carrying a deletion validated by WGS or WES, we looked for variants of interest on the other chromosome that could have been unmasked by the deletion. Missense variants were found (Table S6) and were re-validated in IGV.[46] They had to be homozygote, as expected given a deletion on the other chromosome. Most of the missense variants were frequent, with only one variant at less than 1% allele frequency in gnomAD (rs762560584). Moreover, the UNEECON scores,[47] that predict how deleterious a missense variant is, were under 0.15 (not deleterious) for all variants. One non-coding

| Category | Phenotype | Cases | Controls | Pvalue |
|---|---|---|---|---|
| Deletions >=1Mb | GGE | 14/349 | 9/945 | 8e-06 |
| | NAFE | 2/165 | 9/945 | 0.4 |
| | Mixed | 3/30 | 9/945 | 7e-05 |
| | DEE | 0/59 | 8/827 | |
| Deletions <1Mb | GGE | 192/349 | 561/945 | 0.2 |
| | NAFE | 102/165 | 561/945 | 0.9 |
| | Mixed | 12/30 | 561/945 | 0.07 |
| | DEE | 33/59 | 489/827 | 0.6 |
| Duplications >=1Mb | GGE | 6/349 | 19/945 | 0.9 |
| | NAFE | 5/165 | 19/945 | 0.4 |
| | Mixed | 0/30 | 19/945 | |
| | DEE | 2/59 | 16/827 | 0.5 |
| Duplications <1Mb | GGE | 209/349 | 530/945 | 0.2 |
| | NAFE | 100/165 | 530/945 | 0.4 |
| | Mixed | 19/30 | 530/945 | 0.4 |
| | DEE | 36/59 | 456/827 | 0.3 |

OR (95% CI): 0.25  1.0  2.0  4.0  8.0 16.0 32.0

**Figure 1.** Burden of CNVs by length in epilepsy subgroups. GGE = genetic generalized epilepsies; NAFE = non-acquired focal epilepsy; Mixed = cases with an epilepsy phenotype different from the other affected family members; DEE = developmental epileptic encephalopathies.

variant (chr15:31195835CAG > C) in a GGE patient had a negative constraint score for the three tested tissues, which implies that it is not likely to be deleterious and is also quite frequent in gnomAD (0.42). The other non-coding variants did not have a constraint score meaning that they are not likely to have any functional effect.
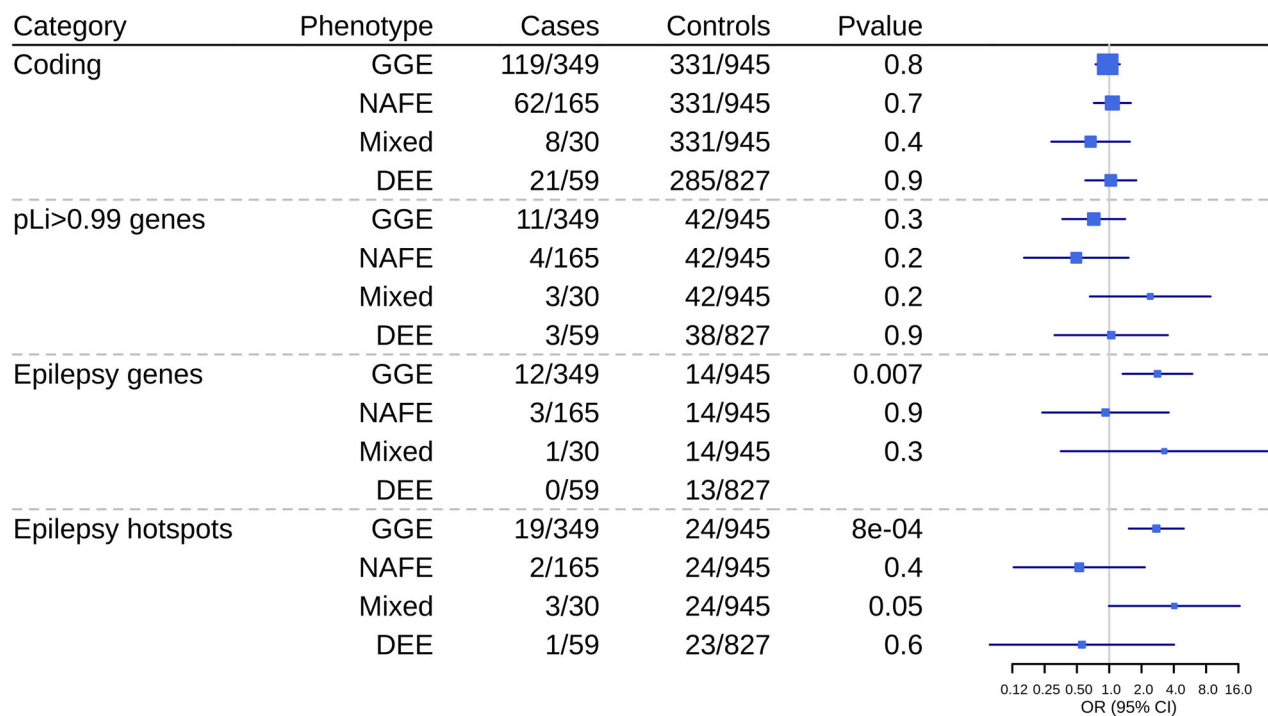
We identified two NAFE patients with a double CNV hit. One had a duplication transmitted by the unaffected mother and a deletion transmitted by the affected father, both on chromosome 7 (chr7:88161734-89838707dup and chr7:108854537-109969407del) and validated by segregation. The second patient had one duplication followed immediately by a 13 Mb deletion on chromosome 18 (chr18:63151948-64412293dup and chr18:64525217-77553173del), both validated by WES, but with no family information.

## Discussion

In the present work, we found an excess of deletions of more than 1 Mb in GGE patients compared to controls, and to a lesser extent, in individuals from mixed families, comparable to previous findings.[24] We also found an excess of large deletions compared to duplications in GGE patients, again comparable to previous findings.[18,22]

Most of these deletions were located in epilepsy genes or hotspots. Moreover, we found an excess of deletions in epilepsy genes and hotspots in GGE patients which is mostly driven by a deletion on the 15q13.3 recurrent site which is also spanning an epilepsy gene, *CHRNA7* and has been reported previously in GGE patients[8,13] (OMIM 612001). This deletion in the 15q13.3 hotspot region was present exclusively in seven GGE patients from six different families and two unaffected family members and was not reported in any population control nor in other epilepsy types. It is the only deletion in an epilepsy hotspot that was restricted to patients and their relatives in this study. This could be a variant linked to the founder effect in the Quebec population[48] and propagated mostly to the affected descendants of a given ancestor. This would need further family and population analyses to validate the transmission scheme of a variant associated with a disease compared to one that is only resulting in the expected transmission in a founder population without any disease association.

The only de novo large deletion was found in a patient from a mixed family (DEE in a NAFE family, Fig. S2). Interestingly, it was found in an epilepsy hotspot, 22q11.2 (OMIM 611867). In addition to DEE, the patient presented a severe intellectual disability and autism-like

| Category | Phenotype | Cases | Controls | Pvalue | |
|---|---|---|---|---|---|
| Coding | GGE | 119/349 | 331/945 | 0.8 | |
| | NAFE | 62/165 | 331/945 | 0.7 | |
| | Mixed | 8/30 | 331/945 | 0.4 | |
| | DEE | 21/59 | 285/827 | 0.9 | |
| pLi>0.99 genes | GGE | 11/349 | 42/945 | 0.3 | |
| | NAFE | 4/165 | 42/945 | 0.2 | |
| | Mixed | 3/30 | 42/945 | 0.2 | |
| | DEE | 3/59 | 38/827 | 0.9 | |
| Epilepsy genes | GGE | 12/349 | 14/945 | 0.007 | |
| | NAFE | 3/165 | 14/945 | 0.9 | |
| | Mixed | 1/30 | 14/945 | 0.3 | |
| | DEE | 0/59 | 13/827 | | |
| Epilepsy hotspots | GGE | 19/349 | 24/945 | 8e-04 | |
| | NAFE | 2/165 | 24/945 | 0.4 | |
| | Mixed | 3/30 | 24/945 | 0.05 | |
| | DEE | 1/59 | 23/827 | 0.6 | |

0.12 0.25 0.50 1.0 2.0 4.0 8.0 16.0
OR (95% CI)

**Figure 2.** Burden of deletions across different gene sets or hotspots in epilepsy subgroups. GGE = genetic generalized epilepsies; NAFE = non-acquired focal epilepsy; Mixed = cases with an epilepsy phenotype different from the other affected family members; DEE = developmental epileptic encephalopathies.

**Table 2.** Number of individuals carrying deletions >1 Mb.

| Phenotype | Dels >1 Mb | Coding | pLi >0.99 | Epilepsy genes | Epilepsy hotspots | Validated by WGS | Validated by WES | Validated by segregation | Validated overall |
|---|---|---|---|---|---|---|---|---|---|
| GGE | 14 | 14 | 2 | 7 | 9 | 5 | 4 | 8 | 14 |
| NAFE | 2 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 2 |
| Mixed | 3 | 3 | 1 | 0 | 1 | 1 | 0 | 1 | 2 |
| DEE trio patients | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Unaffected DEE trio parents | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Unaffected familial ctrls (GGE and NAFE families) | 5 | 5 | 2 | 2 | 4 | 0 | 0 | 5 | 5 |
| Population ctrls | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Dels = deletions; pLi = genes intolerant to truncating variants; GGE = genetic generalized epilepsies; NAFE = non-acquired focal epilepsy; Mixed = cases with an epilepsy phenotype different from the other affected family members; DEE = developmental epileptic encephalopathies; ctrls = controls.

symptoms which are associated with the 22q11.2 deletion. Interestingly, it has been shown that 11% of the 22q11.2 deletion carriers have epilepsy and an additional 59% have seizures or seizure-like symptoms.[49] This could also explain the DEE phenotype within a NAFE family for this patient.

Most of the identified large deletions in epileptic patients were transmitted either by an affected or an unaffected parent, denoting incomplete penetrance[50] with only one deletion that occurred de novo. The validation rate was high in the present study, thanks to the variety of data available for these patients. To test whether the incomplete penetrance of epilepsy-related deletions could be explained by the unmasking of a variant on the other chromosome, we looked at the deletion regions in available WGS and WES for lof and missense variants in addition to variants in ClinVar and variants predicted to affect gene expression using Expecto (see Methods for details). We did not find any evidence of lof or other variants classified as probably pathogenic in ClinVar or affecting gene expression. We found missense variants that are not predicted deleterious according to the annotations in gnomAD and the UNEE-CON scores[47] (Table S6).

Another hypothesis that has been proposed to explain incomplete penetrance is the double CNV hit hypothesis.[50] Two NAFE patients had a double CNV, one patient had both CNVs on chromosome 7 and the other both on chromosome 18. The former patient's duplication on chromosome 7 is the most frequent duplication and has been seen in 12 patients and controls in the present dataset. Both CNVs on chromosome 7 affect coding regions but do not include genes intolerant to truncating variants nor known epilepsy genes or hotspots, so we do not have evidence that these would be associated with the disease. However, the second NAFE patient had both CNVs, a duplication and a deletion, adjacent to chromosome 18. The deletion was the largest found in our dataset,

spanning 13 Mb and two genes intolerant to truncating variants, *ZNF236* and *ZNF407* that were associated with chromosome 18q deletion syndrome (OMIM 601808), neurodevelopmental disorders, and intellectual disability,[51] among others. The finding of two CNVs in this case does not necessarily support the double CNV hypothesis since it cannot be ruled out that the deletion alone caused the phenotype.

In conclusion, we found an excess of large deletions in GGE patients compared to unaffected familial controls from the CENet cohort and population controls from the Quebec Reference Sample and also compared the number of duplications in GGE patients. Most of the deletions are located at genomic hotspots in GGE, especially at the 15q13.3 site which could have been brought and disseminate by an ancestor of the Quebec founder population. We also found one de novo deletion that could explain the patient's phenotype and be of interest for the medical follow-up. We could not find evidence of deleterious or regulatory variants on the homologous chromosome that would explain the incomplete penetrance of the disease among individuals having large deletions. The double CNV hypothesis could not be supported neither although we found two large CNVs in two NAFE patients including one deletion of 13 Mb that could be of interest for the patient and the clinician. We found missense variants within the deletion regions that seem not sufficient to explain the disease. Therefore, we think that there might be other genomic or epigenomic causes in addition to large deletions that would explain the incomplete penetrance of epilepsy-related microdeletions, although we need more sequencing data to validate these findings.

All authors participated in the study design and reviewed the manuscript. CM, FT, and AG performed the analyses. CM wrote the manuscript. SLG supervised the analyses. CL built and manages the SLSJ family cohort which provided the Saguenay population control samples.

## Acknowledgments

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. Allers K, Essue BM, Hackett ML, et al. The economic impact of epilepsy: a systematic review. BMC Neurol. 2015;15(245).

2. Cossette P, Liu L, Brisebois K, et al. Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. Nat Genet. 2002;31(2):184-189.

3. Baulac S, Gourfinkel-An I, Picard F, et al. A second locus for familial generalized epilepsy with febrile seizures plus maps to chromosome 2q21-q33. Am J Hum Genet. 1999;65(4):1078-1085.

4. Beck C, Moulard B, Steinlein O, et al. A nonsense mutation in the α4 subunit of the nicotinic acetylcholine receptor (CHRNA4) cosegregates with 20q-linked benign neonatal familial convulsions (EBNI). Neurobiol Dis. 1994;1(1-2):95-99.

5. Kalachikov S, Evgrafov O, Ross B, et al. Mutations in LGI1 cause autosomal-dominant partial epilepsy with auditory features. Nat Genet. 2002;30(3):335-341.

6. Feng Y-CA, Howrigan DP, Abbott LE, et al. Ultra-rare genetic variation in the epilepsies: a whole-exome sequencing study of 17,606 individuals. Am J Hum Genet. 2019;105(2):267-282.

7. The International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9(1):5269.

8. Dibbens LM, Mullen S, Helbig I, et al. Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. Hum Mol Genet. 2009;18(19):3626-3631.

9. Mefford HC, Muhle H, Ostertag P, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. 2010;6(5):e1000962.

10. Olson H, Shen Y, Avallone J, et al. Copy number variation plays an important role in clinical epilepsy. Ann Neurol. 2014;75(6):943-958.

11. Lal D, Ruppert AK, Trucks H, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet. 2015;11(5):e1005226.

12. Pérez-Palma E, Helbig I, Klein KM, et al. Heterogeneous contribution of microdeletions in the development of common generalised and focal epilepsies. J Med Genet. 2017;54(9):598-606.

13. Helbig I, Mefford HC, Sharp AJ, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet. 2009;41(2):160-162.

14. De Kovel CGF, Trucks H, Helbig I, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain. 2010;133(1):23-32.

15. Mullen SA, Carvill GL, Bellows S, et al. Copy number variants are frequent in genetic generalized epilepsy with intellectual disability. Neurology. 2013;81(17):1507-1514.

16. Addis L, Rosch RE, Valentin A, et al. Analysis of rare copy number variation in absence epilepsies. Neurol Genet. 2016;2(2):e56.

17. Lal D, Pernhorst K, Klein KM, et al. Extending the phenotypic spectrum of RBFOX1 deletions: sporadic focal epilepsy. Epilepsia. 2015;56(9):e129-e133.

18. Heinzen EL, Radtke RA, Urban TJ, et al. Rare deletions at 16p13.11 predispose to a diverse Spectrum of sporadic epilepsy syndromes. Am J Hum Genet. 2010;86(5):707-718.

19. Hamdan FF, Myers CT, Cossette P, et al. High rate of recurrent de novo mutations in developmental and epileptic encephalopathies. Am J Hum Genet. 2017;101(5):664-685.

20. Piccione M, Vecchio D, Cavani S, et al. The first case of myoclonic epilepsy in a child with a de novo 22q11.2 microduplication. Am J Med Genet A. 2011;155A(12):3054-3059.

21. Gourari I, Schubert R, Prasad A. 1q21.1 duplication syndrome and epilepsy. Neurology. Genetics. 2018;4(1):e219.

22. Mefford HC, Yendle SC, Hsu C, et al. Rare copy number variants are an important cause of epileptic encephalopathies. Ann Neurol. 2011;70(6):974-985.

23. Striano P, Coppola A, Paravidino R, et al. Clinical significance of rare copy number variations in epilepsy: a case-control survey using microarray-based comparative genomic hybridization. Arch Neurol. 2012;69(3):322-330.

24. Niestroj LM, Perez-Palma E, Howrigan DP, et al. Epilepsy subtype-specific copy number burden observed in a

genome-wide study of 17458 subjects. Brain. 2020;143 (7):2106-2118.

25. Bearden CE, Jawad AF, Lynch DR, et al. Effects of a functional COMT polymorphism on prefrontal cognitive function in patients with 22q11.2 deletion syndrome. Am J Psychiatry. 2004;161(9):1700-1702.

26. Albers CA, Paul DS, Schulze H, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet. 2012;44(4):435-439.

27. Girirajan S, Rosenfeld JA, Cooper GM, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet. 2010;42(3):203-209.

28. Moreau C, Michaud JL, Hamdan FF, et al. Global prevalence of potentially pathogenic short-tandem repeats in an epilepsy cohort [Internet]. bioRxiv 2020;2020.08.20.259168. Available from: http://biorxiv.org/content/early/2020/08/21/2020.08.20.259168.abstract

29. Lencz T, Yu J, Khan RR, et al. Novel ultra-rare exonic variants identified in a founder population implicate cadherins in schizophrenia. Neuron. 2021;109(9):1465-1478.

30. Morin A, Madore AM, Kwan T, et al. Exploring rare and low-frequency variants in the Saguenay–lac-saint-Jean population identified genes associated with asthma and allergy traits. Eur J Hum Genet. 2019;27(1):90-101.

31. Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):512-521.

32. Roy-Gagnon MH, Moreau C, Bherer C, et al. Genomic and genealogical investigation of the French Canadian founder population structure. Hum Genet. 2011;129 (5):521-531.

33. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-575.

34. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665-1674.

35. Cooper NJ, Shtir CJ, Smyth DJ, et al. Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. Hum Mol Genet. 2015;24(6):1774-1790.

36. Wolking S, Moreau C, McCormack M, et al. Assessing the role of rare genetic variants in drug-resistant, non-lesional focal epilepsy. Ann Clin Transl Neurol. 2021;8(7):1376-1387.

37. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol. 2016;12(4):e1004873.

38. Boeva V, Popova T, Bleakley K, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28 (3):423-425.

39. Berkovic SF, Scheffer IE, Petrou S, et al. A roadmap for precision medicine in the epilepsies. Lancet Neurol. 2015;14(12):1219-1228.

40. Coppola A, Cellini E, Stamberger H, et al. Diagnostic implications of genetic copy number variation in epilepsy plus. Epilepsia. 2019;60(4):689-706.

41. Watson CT, Marques-Bonet T, Sharp AJ, Mefford HC. The genetics of microdeletion and microduplication syndromes: an update. Annu Rev Genomics Hum Genet. 2014;15:215-244.

42. Cingolani P, Patel VM, Coon M, et al. Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front Genet. 2012;3:35.

43. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w 1118; iso-2; iso-3. Fly. 2012;6:80-92.

44. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D1067.

45. Zhou J, Theesfeld CL, Yao K, et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat Genet. 2018;50(8):1171-1179.

46. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24-26.

47. Huang YF. Unified inference of missense variant effects and gene constraints in the human genome. PLoS Genet. 2020;16(7):e1008922.

48. Scriver CR. Human genetics: lessons from Quebec populations. Annu Rev Genomics Hum Genet. 2001;2:69-101.

49. Eaton CB, Thomas RH, Hamandi K, et al. Epilepsy and seizures in young people with 22q11.2 deletion syndrome: prevalence and links with other neurodevelopmental disorders. Epilepsia. 2019;60(5):818-829.

50. Carvill GL, Mefford HC. Microdeletion syndromes. Curr Opin Genet Dev. 2013;23(3):232-239.

51. Zahra Q, Çakmak Ç, Koprulu M, et al. Biallelic ZNF407 mutations in a neurodevelopmental disorder with ID, short stature and variable microcephaly, hypotonia, ocular anomalies and facial dysmorphism. J Hum Genet. 2020;65 (12):1115-1123.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1** LRR's PCA before (left panel) and after (right panel) PC correction. Symbols represent the different batches and colors of the different plates (many plates were sent for genotyping within one batch).

**Figure S2.** Pedigree of the mixed patient (DEE in a NAFE family) carrying the de novo deletion. Unaffected individuals are in black.

**Table S1** Rare CNVs in epileptic patients and controls.

**Table S2** Recurrent deletions' description from Watson *et al.*

**Table S3** Deletions in epilepsy genes.

**Table S4** Deletions in epilepsy hotspots.

**Table S5** Description of deletions >1 Mb.

**Table S6** Unmasked missense variants.