ARTICLE   OPEN ACCESS

Polygenic risk scores of several subtypes of epilepsies in a founder population

Claudia Moreau, MSc, Rose-Marie Rébillard, MD, Stefan Wolking, MD, Jacques Michaud, MD,
Frédérique Tremblay, BSc, Alexandre Girard, BSc, Joanie Bouchard, BSc, Berge Minassian, MD,
Catherine Laprise, PhD, Patrick Cossette, MD, PhD, and Simon L. Girard, PhD

Correspondence
Dr. Girard
simon2_girard@uqac.ca

Abstract

Objective
Polygenic risk scores (PRSs) are used to quantify the cumulative effects of a number of genetic variants, which may individually have a very small effect on susceptibility to a disease; we used PRSs to better understand the genetic contribution to common epilepsy and its subtypes.

Methods
We first replicated previous single associations using 373 unrelated patients. We then calculated PRSs in the same French Canadian patients with epilepsy divided into 7 epilepsy subtypes and population-based controls. We fitted a logistic mixed model to calculate the variance explained by the PRS using pseudo-R^2 statistics.

Results
We show that the PRS explains more of the variance in idiopathic generalized epilepsy than in patients with nonacquired focal epilepsy. We also demonstrate that the variance explained is different within each epilepsy subtype.

Conclusions
Globally, we support the notion that PRSs provide a reliable measure to rightfully estimate the contribution of genetic factors to the pathophysiologic mechanism of epilepsies, but further studies are needed on PRSs before they can be used clinically.

From the Centre Intersectoriel en Santé Durable (C.M., F.T., A.G., J.B., C.L., S.L.G.), Université du Québec à Chicoutimi, Saguenay; Axe Neurosciences (R.-M.R., S.W., P.C.), Centre de recherche de l'Université de Montréal, Université de Montréal; Centre de recherche du CHU Ste-Justine (J.M.), Université de Montréal, Canada; and Department of Pediatrics and Neurology and Neurotherapeutics (B.M.), UT Southwestern Medical Center, TX.

Glossary

CAE = childhood absence epilepsy; **FC** = French Canadian; **IGE** = idiopathic generalized epilepsy; **ILAE** = International League Against Epilepsy; **MAF** = minor allele frequency; **NAFE** = nonacquired focal epilepsy; **PC** = principal component; **PRS** = polygenic risk score.

In the last decade, many groups have been working on different genetic techniques and statistics to better understand the complex genetic mechanisms underlying epilepsy.[1] Last year, a large genome-wide association study on epilepsy identified 16 loci associated with the disease, and many of these were already known or suspected.[2] Despite these efforts, there is still a substantial missing heritability component in epilepsy genetics.[3]

It is likely that a wide spectrum of genetic factors is in play, ranging from very rare mutations with large effects to relatively rare variants with medium effect sizes and finally to common variants with smaller risk effects. Polygenic risk scores (PRSs) aim to quantify the cumulative effects of a number of variants, which may individually have a very small effect on susceptibility. They have been used previously in many common traits and diseases such as heart disease[4–6] and in neurologic disorders.[7–10]

In this study, we aim to use PRSs to see whether this method can explain more of the epilepsy genetics than the classical methods did. We take advantage of the recent meta-analysis GWAS metrics[2] to calculate PRSs in 373 unrelated French Canadian (FC) patients with epilepsy divided into 7 subtypes. The population is known for its well-documented recent (400 years) founder effect and its particular genetic background, which makes it an ideal population for genetic studies. French Canadians are also closely related to the European population, which is predominant in the GWAS.[2]

Methods

Standard protocol approvals, registrations, and patient consents

This study was approved by the Centre de Recherche du Centre Hospitalier Universitaire de Montréal ethics committee, and written informed consent was obtained for all patients.

Phenotyping of patients

The epilepsy cohort was composed of families with at least 3 affected individuals with idiopathic generalized epilepsy (IGE) or nonacquired focal epilepsy (NAFE) previously collected and diagnosed by neurologists. The clinical epilepsy phenotype is defined based on the Classification of the Epilepsy Syndromes established by the International League Against Epilepsy (ILAE).[11]

More specifically, the operational definitions of the epilepsy phenotypes studied in the project for NAFE were as follows: (1) patients were aged at least 5 years and had experienced at least 2 unprovoked seizures with focal EEG abnormalities or clear clinical focal semiology in the 6 months before starting treatment AND (2) an MRI scan of the brain that did not demonstrate any potentially epileptogenic lesion (no lesion) OR (3) documented hippocampal sclerosis and lesion other than mesial temporal sclerosis (other lesion).

For IGE, patients were at least 4 years of age at the diagnosis and the IGE subtype (childhood absence epilepsy, juvenile absence epilepsy, juvenile myoclonic epilepsy or IGE not otherwise specified) was determined according to the 1989 ILAE syndrome definitions using clinical and EEG characteristics. In IGE, we also included patients with epilepsy with eyelid myoclonia (Jeavons), which is an idiopathic generalized form of reflex epilepsy characterized by childhood-onset, unique seizure manifestations, striking light sensitivity, and possible occurrence of generalized tonic-clonic seizures alone.

Our epilepsy cohort consisted of 643 patients diagnosed with familial epilepsy. We used 1 patient per family for the analyses for a total of 373 not closely related FC patients, 192 with IGE and 132 with NAFE (and 49 unclassified epilepsies because of a lack of information). We validated that the remaining patients

Table 1 Basic statistics for different epilepsy subtypes

Epilepsy subtype	n	Women, n	Mean age
All epilepsies[a]	373	213	46
IGE[a]	192	114	43
CAE	34	18	35
JAE	18	10	44
JME	81	52	48
GTCS	24	13	41
NAFE[a]	132	70	48
NAFE HS	22	10	55
NAFE no lesion	71	40	46
NAFE other lesion	13	10	51

Abbreviations: CAE = childhood absence epilepsy; GTCS = generalized tonic-clonic seizures alone; HS = documented hippocampal sclerosis; IGE = idiopathic generalized epilepsy; JAE = juvenile absence epilepsy; JME = juvenile myoclonic epilepsy, NAFE = nonacquired focal epilepsy; NAFE HS = nonacquired focal epilepsy documented hippocampal sclerosis; NAFE no lesion = nonacquired focal epilepsy no documented epileptogenic lesion; NAFE other lesion = nonacquired focal epilepsy lesion other than mesial temporal sclerosis.
[a] The difference between the total numbers of patients and of IGE and NAFE broad subtypes and the sum of their respective subtypes comes from the patients who could not be classified into any of the narrow subtypes because of a lack of clinical information (49 unclassified patients, 35 unclassified IGE, and 26 unclassified NAFE).

were not closely related (first cousins or more related) using PLINK. FC ancestry was assessed by self-declared ethnicity and principal component (PC) analysis (figure e-1, links.lww.com/NXG/A253). Table 1 shows the different subtypes of epilepsy that are represented in our cohort. In addition, we selected 954 FC individuals from a reference population data set.[12]

Data availability

The patients' genotype data used in the present study will be available on request.

Genotyping and imputation

For this study, we used whole-genome genotyping data for the patient and the French Canadian control[12] cohorts. All samples were processed on either the Illumina Omni Express (number of single nucleotide polymorphisms = 710,000) or the Illumina Omni 2.5 (number of single nucleotide polymorphisms = 2,500,000 including the Omni Express core) depending on the availability of the chip regardless if they were controls or patients. Genotypes of all samples were merged, and only positions present on both chips were kept. We performed cleaning steps to remove individuals having more than 2% missing genotypes among all SNPs, SNPs with more than 2% missing SNPs over all individuals, and SNPs with Hardy-Weinberg p value <0.001 using PLINK

software.[13] We then removed 121 individuals of non-FC descent using the first 2 principal components in addition to self-identification of patients whenever this information was available. Principal component analysis (figure e-1, links.lww.com/NXG/A253) was performed using Eigensoft[14] on pruned SNPs (pairwise $r^2 < 0.2$ in sliding windows of size 50 shifting every 5 SNPs) at 5% minor allele frequency (MAF). We finally aligned the data set to the GRCh37 genome build for further imputation following the method described here.[15]

The Sanger Imputation Service was used to conduct whole-genome imputation of SNPs.[16] We selected the Human Reference Consortium data set as the reference panel. Postimputation quality control filters were applied to remove SNPs within imputed data with an imputation info score of <0.9 or HWE p value of $<1\ e^{-6}$, and only biallelic SNPs at MAF 1% or higher were kept for further analyses.

Association analysis

We used PLINK software for the logistic association analysis with the first 10 PCs and sex as covariates. Associations were only tested for the 20 SNPs found significant in the ILAE study and only in the epilepsy subtypes in which they were originally

Figure 1 PRS density for broad epilepsy types
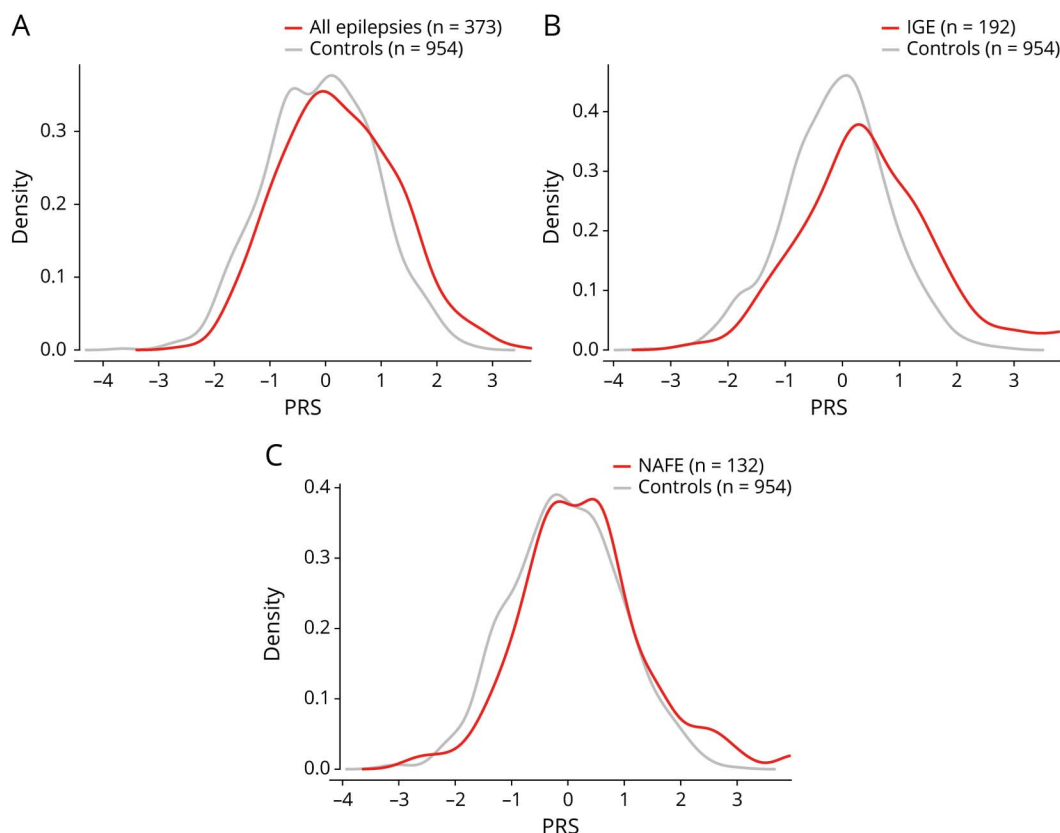
PRS density plots for French Canadian controls and (A) all patients with epilepsy, (B) patients with IGE, and (C) patients with NAFE. IGE = idiopathic generalized epilepsy; NAFE = nonacquired focal epilepsy; PRS = polygenic risk score.

Figure 2 PRS density for IGE subtypes
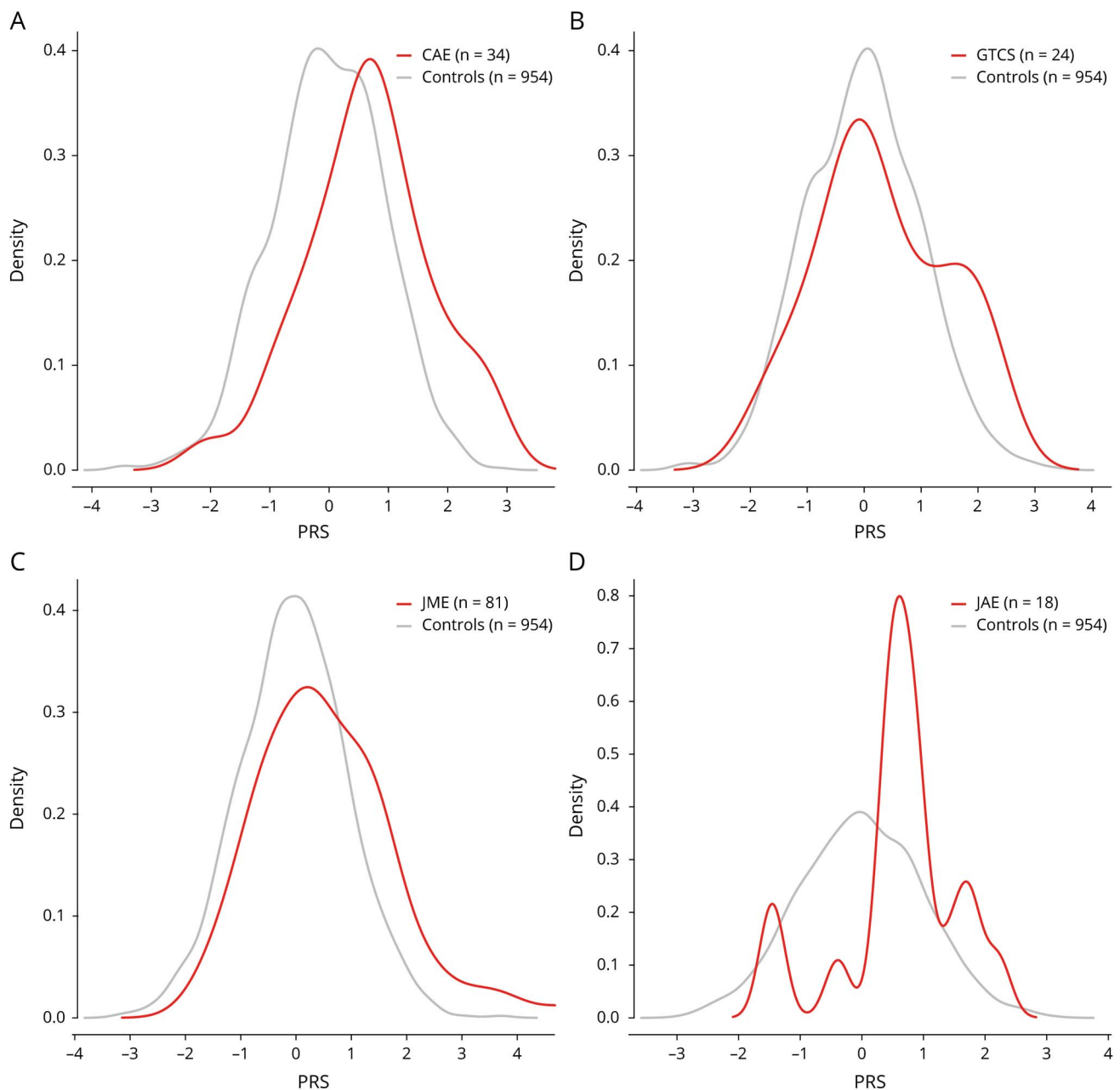
PRS density plots for French Canadian controls and patients with IGE syndrome (A) CAE, (B) GTCS, (C) JME, and (D) JAE. CAE = childhood absence epilepsy; GTCS = generalized tonic-clonic seizures alone; IGE = idiopathic generalized epilepsy; JAE = juvenile absence epilepsy; JME = juvenile myoclonic epilepsy; PRS = polygenic risk score.

reported. We used a p value threshold of 0.0025 to account for multiple testing (n = 20).

PRS calculation

PRSs were calculated with PRSice software[17] using ILAE meta-analysis on epilepsy summary statistics.[2] Because the BETA was not provided for the METAL analyses (all, generalized, and focal epilepsies' analyses), we used the formula from reference 18 to calculate it. We used the first 10 PCs in addition to sex as covariates, recalculating eigenvectors for each patient subset including controls using SNPs at MAF 0.05 pruned (as described above). PRSs were standardized for graphs.

Statistical analyses

We fitted a logistic regression mixed model using R. We then calculated the Nagelkerke pseudo-R^2 (using the PRS at the p value threshold that best predicts the phenotype) with and without the PRS as the full and null model. Note that pseudo-R^2 is reported on the observed scale to avoid overfitting.

Results

We used whole-genome genotyping on 373 unrelated patients having epilepsy and 954 population controls. All individuals

Figure 3 PRS density for NAFE subtypes
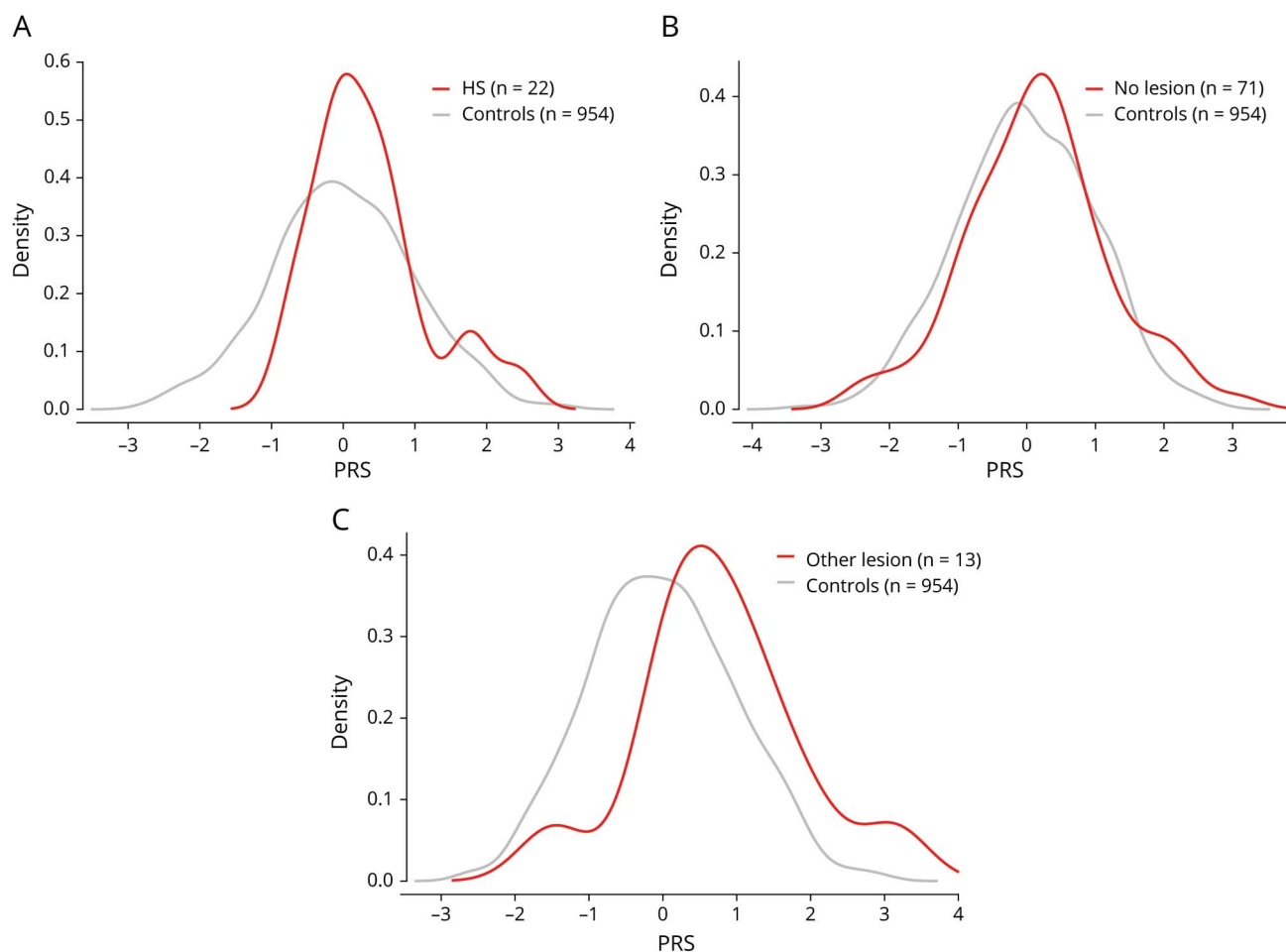
PRS density plots for French Canadian controls and patients with NAFE with (A) HS, (B) no documented lesion, and (C) lesions other than HS. HS = documented hippocampal sclerosis; NAFE = nonacquired focal epilepsy; PRS = polygenic risk score.

were confirmed with French Canadian ancestry. FC control individuals used in this study have already been demonstrated to cluster with Western Europeans.[19] First, we wanted to assess whether the associations found by the ILAE study[2] were valid for our cohort. Table e-1, links.lww.com/NXG/A253, presents the statistics of the association analysis. One locus was found to be significant, and 3 were close to significance (p value threshold = 0.0025). These results show that our founder FC population shares a portion of the epilepsy genetic risks with the Western European populations studied by the ILAE.

Next, to assess whether SNPs taken together could explain a portion the epilepsy phenotype, we used the basic statistics of the ILAE study[2] to construct the PRS. Figure 1 shows the density plots of standardized PRS values of patients compared with controls for the 3 broad epilepsy types; best-fit p values are shown in figure e-2 (links.lww.com/NXG/A253). Our first observation was that the PRS distribution is more shifted to the right in IGE than in NAFE, which is consistent with the heritability estimates reported in the ILAE study.[2] Figures 2

and 3 show the same analysis for the IGE and NAFE subtypes. The best-fits are shown in figures e-3 and e-4 (links.lww.com/NXG/A253).

The next logical question was to investigate whether the PRS could be used to discriminate between a patient with epilepsy and a control. Table 2 presents the logistic mixed model statistics and the variance explained by the PRS calculated using the Nagelkerke pseudo-R^2 (on the observed scale) for patients and controls for all epilepsy subtypes. The variance explained by the PRS varies among epilepsy subtypes, but is generally higher for IGE types than for NAFE types. This is reflected by the higher odds ratio and variance explained by the PRS in IGE broad subtype compared with NAFE and corroborates what was found in a recent study.[20]

Discussion

The strongest association in our FC cohort was observed with the SNP rs1402398. This SNP is located in the noncoding

Table 2 PRS and pseudo-R^2 statistics based on the logistic mixed model

Epilepsy subtype	PRS p value threshold	No. of SNPs	OR (95% CI)	p Value (Wald)	Pseudo-R^2
All epilepsies	0.1606	45,634	1.12 (1.07–1.17)	<0.001	0.031
IGE	0.06265	24,293	1.24 (1.17–1.32)	<0.001	0.087
CAE	0.079	25,616	4.85 (2.31–10.19)	<0.001	0.094
JAE	0.0044	2,705	74.88 (2.82–1,990.77)	0.01	0.042
JME	0.0041	2,941	8.47 (3.67–19.58)	<0.001	0.057
GTCS	0.046	15,624	4.41 (0.69–28.31)	0.118	0.012
NAFE	0.09695	33,200	1.12 (1.05–1.19)	0.001	0.019
NAFE HS	1	95,225	1.82 (0.92–3.58)	0.084	0.016
NAFE no lesion	0.00065	686	2.56 (0.68–9.74)	0.167	0.0047
NAFE other lesion	0.0007	756	43.36 (2.86–658.04)	0.007	0.058

Abbreviations: CAE = childhood absence epilepsy; GTCS = generalized tonic-clonic seizures alone; HS = documented hippocampal sclerosis; IGE = idiopathic generalized epilepsy; JAE = juvenile absence epilepsy; JME = juvenile myoclonic epilepsy; NAFE = nonacquired focal epilepsy; NAFE HS = nonacquired focal epilepsy documented hippocampal sclerosis; NAFE no lesion = nonacquired focal epilepsy no documented epileptogenic lesion; NAFE other lesion = nonacquired focal epilepsy lesion other than mesial temporal sclerosis; PRS = polygenic risk score.

region surrounding genes *FANCL* and *BCL11A*. These genes have been linked with epilepsies through association studies,[21] but no other functional or clinical evidence highlights their roles in the disease.

Although we successfully replicated associations, we believe that the biggest contribution of our study lies in the PRSs that were established for each epilepsy type. This is, to our knowledge, one of the first documented examples of how PRSs can be used for epilepsy genetic studies for different subtypes of epilepsy, although a recent study has shown that for broad epilepsy subtypes.[20] Although this measure cannot yet be translated into clinical use, our analysis shows that the additive value of common variants can be used to better understand the disease.

One definite pitfall of our study is the small size of our cohort. The initial GWAS was performed on more than 15,000 patients with epilepsy. Our study only included 373 patients with epilepsy and thus cannot have the same outreach as the initial one. This is why we did not report genome-wide association statistics and focused only on the replication of associated SNPs. We believe that the small size of our cohort also affects the PRS calculations, but to a smaller degree.

For these reasons, we have to take the variance explained by the PRS with caution. However, for the broad phenotypes, we explain 4 times more of the variance for patients with IGE than what we explain for patients with NAFE, as shown elsewhere.[20] This also supports the fact that epilepsy should be divided into subtypes when studying the genetic mechanism underlying the disease, as some epilepsy types were reasonably well explained by the PRS (i.e., childhood absence epilepsy).

This study was conducted on a documented founder population. The FC population is well known for its high prevalence of specific disease-causing mutations.[22,23] For epilepsy, although we cannot exclude that some of the associations found were driven by rare haplotypes, we show here that the genetic etiology of the disease is consistent with that of the general European population. In future work, we will try to assess whether the strong PRS found in some epilepsy subtypes could be explained by rarer haplotypes, as we would expect in a founder population.

Globally, we support the notion that PRSs provide a reliable measure to rightfully estimate the contribution of genetic factors to the pathophysiologic mechanism of epilepsies.

Disclosure
C. Moreau, R.-M. Rébillard, S. Wolking, J. Michaud, F. Tremblay, A. Girard, J. Bouchard, B. Minassian, C. Laprise, P. Cossette, and S.L. Girard report no disclosure. This study

Appendix Authors

Name	Location	Contribution
Claudia Moreau, MSc	Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay	Designed and conceptualized the study, analyzed the data, and drafted the manuscript for intellectual content
Rose-Marie Rébillard, MD	Axe Neurosciences, Centre de recherche de l'Université de Montréal, Université de Montréal, Canada	Determined patients' epilepsy subtype
Stefan Wolking, MD	Axe Neurosciences, Centre de recherche de l'Université de Montréal, Université de Montréal, Canada	Determined patients' epilepsy subtype
Jacques Michaud, MD	Centre de recherche du CHU Ste-Justine, Université de Montréal, Canada	Collected and genotyped patients with epilepsy
Frédérique Tremblay	Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay	Interpreted the data and revised the manuscript for intellectual content
Alexandre Girard	Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay	Interpreted the data and revised the manuscript for intellectual content
Joanie Bouchard	Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay	Interpreted the data and revised the manuscript for intellectual content
Berge Minassian, MD	Department of Pediatrics and Neurology and Neurotherapeutics, UT Southwestern Medical Center, TX	Collected and genotyped patients with epilepsy
Catherine Laprise, PhD	Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay	Contributed in the control group
Patrick Cossette, MD, PhD	Axe Neurosciences, Centre de recherche de l'Université de Montréal, Université de Montréal, Canada	Collected and genotyped patients with epilepsy

Appendix (continued)

Name	Location	Contribution
Simon L. Girard, PhD	Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Saguenay	Designed and conceptualized the study, supervised the analysis of the data, and finalized the manuscript

References

1. Koeleman BPC. What do genetic studies tell us about the heritable basis of common epilepsy? Polygenic or complex epilepsy? Neurosci Lett 2018;667:10–16.
2. 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;9:5269.
3. Thomas RH, Berkovic SF. The hidden genetics of epilepsy-a clinically important new paradigm. Nat Rev Neurol 2014;10:283–292.
4. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219–1224.
5. Wünnemann F, Lo KS, Langford-Avelar A, et al. Validation of genome-wide polygenic risk scores for coronary artery disease in French Canadians. Circ Genomic Precis Med 2019;12:243–248.
6. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. J Am Coll Cardiol 2018;72:1883–1893.
7. Fullerton JM, Koller DL, Edenberg HJ, et al. Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young at-risk Individuals. Am J Med Genet B Neuropsychiatr Genet 2015;168:617–629.
8. Dima D, de Jong S, Breen G, Frangou S. The polygenic risk for bipolar disorder influences brain regional function relating to visual and default state processing of emotional information. Neuroimage Clin 2016;12:838–844.
9. Ahn K, An SS, Shugart YY, Rapoport JL. Common polygenic variation and risk for childhood-onset schizophrenia. Mol Psychiatry 2016;21:94–96.
10. Boies S, Mérette C, Paccalet T, Maziade M, Bureau A. Polygenic risk scores distinguish patients from non-affected adult relatives and from normal controls in schizophrenia and bipolar disorder multi-affected kindreds. Am J Med Genet Part B Neuropsychiatr Genet 2018;177:329–336.
11. Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. Epilepsia 2010;51:676–685.
12. Quebec Reference Sample. Website [online]. Available at: www.quebecgenpop.ca/home.html. Accessed August 6, 2019.
13. 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;81:559–575.
14. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:304–309.
15. McCarthy Group Tools. Website [online]. Available at: www.well.ox.ac.uk/~wrayner/tools/. Accessed June 14, 2019.
16. Sanger Imputation Service. Website [online]. Available at: imputation.sanger.ac.uk/). Accessed March 18, 2019.
17. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. Bioinformatics 2015;31:1466–1468.
18. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48:481–487.
19. Roy-Gagnon MH, Moreau C, Bherer C, et al. Genomic and genealogical investigation of the French Canadian founder population structure. Hum Genet 2011;129:521–531.
20. Leu C, Stevelink R, Smith AW, et al. Polygenic burden in focal and generalized epilepsies. Brain 2019;142:3473–3481.
21. The International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol 2014;13:893–903.
22. Scriver CR. Human Genetics : lessons from Quebec populations. Annu Rev Genomics Hum Genet 2001;2:69–101.
23. Laberge AM, Michaud J, Richter A, et al. Population history and its impact on medical genetics in Quebec. Clin Genet 2005;58:287–301.