INTRODUCTION

Focal epilepsy (FE) has a prevalence of 2.99 in 1000 people.\(^1\) The presentation of FE is clinically highly diverse. Cognitive impairments, such as learning and memory difficulties, and mood disorders are frequently comorbid with FE. FE can develop at any point in life, but associated etiologies vary by age, with congenital anomalies often associated with earlier seizure onset, and stroke/neurodegenerative disorders associated with older age at seizure onset.\(^2\) Several studies have reported rare forms of FE believed to be caused by a single
genetic variant of large effect.\textsuperscript{3} Genome-wide association studies (GWASs) for common forms of epilepsy have identified common genetic risk variants with individually small effects for FE.\textsuperscript{4–6} Recently our team showed that polygenic risk scores (PRS) for FE that combine the small effect sizes of thousands of common genetic variants could stratify individuals with FE from population controls.\textsuperscript{7} Despite such examples, the number of genes and loci identified by GWASs and the observed single-nucleotide polymorphism (SNP)–based heritability are low compared to generalized epilepsy or other neurological disorders.\textsuperscript{6,8} The phenotypic category of FE in large-scale genetic studies is defined entirely on a confirmed focal origin of seizures.\textsuperscript{6,9,10} Various clinical phenotypes, for example, structural-metabolic lesions that can be caused by acquired disorders that precede epilepsy (eg, traumatic brain injuries, central nervous system infections, vascular malformations, or stroke),\textsuperscript{11} are possibly driving factors for the high genetic heterogeneity and the low power to discover novel associations.\textsuperscript{7} However, no studies have directly investigated whether the heterogeneity among clinical subtypes of FE correlates with genetic heterogeneity on a common risk variant level. Here, we use two different approaches (categorical classification by routinely assessed clinical features; hierarchical cluster analysis) to classify subgroups of FE, as recommended by Berg et al.\textsuperscript{11} We then use PRS for FE (FE-PRS) to explore whether the defined FE subgroups have a heterogeneous burden of common risk variants for FE.

## 2 | METHODS

### 2.1 | Study cohort

Our study design is presented in Figure 1. Data from individuals of European ancestry who had a diagnosis of FE were obtained from an institutional review board (IRB)–approved Epilepsy Biorepository and Data Registry maintained by the Cleveland Clinic Epilepsy Center (Table S2). FE was diagnosed according to clinical criteria (clinical interview, neurological examination, electroencephalography [EEG], and imaging data). All participants provided written informed consent for participation in the registry and provided blood and/or saliva samples for genetic screening. Genetic ancestry-matched population controls were combined from several in-house projects and the Partners HealthCare Biobank.\textsuperscript{12} The control cohorts,

![Figure 1](https://example.com/figure1.png)

**Figure 1** Study design. (A), Data included 414 European-ancestry patients with focal epilepsy (FE) seen at Cleveland Clinic who had complete phenotype data and single-nucleotide polymorphism (SNP) genotype data after quality control. Genotype data of 20,435 ancestry-matched population controls after quality control were obtained from in-house projects and a publicly available database.\textsuperscript{12} (B), Individuals with FE were stratified into subgroups via two different methods: (I) grouping by clinical features: known/unknown etiology, early/late-onset seizures, and presence/absence of a psychiatric comorbidity, and (II) grouping by hierarchical cluster analysis of the three clinical features. (C) Polygenic risk scoring (PRS) for FE of all individuals with FE and the population controls. (D) Evaluation of the FE-PRS in each of the phenotypic subgroups in comparison to population controls.
genotyping, quality control, and imputation are described in the Supporting Information.

### 2.2 Phenotypic variables

We classified individuals with FE based on the hypothesis that individuals with different etiologies (ie, cause for seizures), seizure onsets, and comorbid psychiatric disorders have varying degrees of genetic components for FE. Individuals with FE were defined as having a known etiology when: (a) there was clear evidence for seizures secondary to an acquired disorder: venous malformation, cavernous malformation, brain aneurysm, intrauterine stroke, prenatal stroke, stroke later in life, transient ischemic attack, white matter small vessel disease; or (b) when medical history was notable for one of the following conditions that preceded seizure onset: head trauma, brain tumor, encephalitis, or meningitis. Individuals were considered to have a psychiatric comorbidity if a history of depression and/or anxiety was documented in their medical chart (preceding or following seizure onset). A total of 414 of 535 genotyped and quality-controlled individuals with FE from our previous non-subgroup-specific study (FE-Cleveland-EUR cohort) had complete data among our phenotypic variables and were included in the analysis. The study data were collected and reviewed by epileptologists using the REDCap electronic data capture tools (https://projectredcap.org) hosted at Cleveland Clinic Epilepsy Biorepository.

### 2.3 Classification based on clinical features and cluster analysis

First, we performed a categorical classification of each patient into binary groups based on three clinical features: known/unknown etiology, early/late-onset seizures, and presence/absence of psychiatric comorbidities (Figure 1). This resulted in six overlapping subgroups of individuals with: (a) known etiology, (b) unknown etiology, (c) age at seizure onset <21 years (early onset), (d) age at seizure onset ≥21 years (late-onset), (e) psychiatric comorbidity, and (f) no psychiatric comorbidity. Next we performed a hierarchical cluster analysis using all three clinical features to partition the individuals with FE into clusters with similar clinical characteristics. The analysis was carried out on all 414 individuals with FE. The variables known/unknown etiology and presence/absence of psychiatric comorbidity were considered as binary features, whereas age at seizure onset was included as a continuous variable. To account for pairwise dissimilarity between individuals with FE, Gower distance was calculated using the daisy function implemented in the cluster package of R (https://cran.r-project.org/web/packages/cluster/cluster.pdf). Hierarchical clustering was performed using Ward’s minimum variance method with the hclust function in base R. The optimal number of clusters was selected with the average silhouette width method, using the silhouette function in the cluster package in R.

### 2.4 PRS calculation

To generate a FE-PRS for each sample, we followed the protocol from Leu et al. Briefly, single-nucleotide polymorphism (SNP) weights were derived from the summary statistic of the largest GWAS meta-analysis for FE and pruned based on \( P \leq .5 \). All remaining SNPs from the GWAS summary statistic were extracted or imputed from the genotype data and subsequently pruned to a subset of uncorrelated SNPs. Finally, FE-PRS for each individual was generated using the allelic scoring function, as implemented in PLINK v1.9 (https://www.cog-genomics.org/plink). None of the individuals with FE included in this study was part of the GWAS meta-analysis for FE, which was used to generate the PRS.

### 2.5 Statistical analyses

We used logistic regression, adjusted for sex and for the first four principal components of ancestry, to correct for population structure, and to determine whether subgroups were associated with polygenic risk for FE compared to controls. The logistic regression models of the 10 tested groups are detailed in the Supporting Information (Table S1). The threshold for statistical significance after Bonferroni correction was set to \( \alpha = 5 \times 10^{-3} \) (10 groups tested: 6 categorical subgroups and 4 cluster-derived subgroups). Following common practices for PRS, Nagelkerke’s pseudo-\( R^2 \) was calculated to measure the proportion of phenotypic variance explained by the FE-PRS, by comparing the full model of the logistic regression (PRS plus all covariates: sex and the first four principal components of ancestry) to the null model (covariates only). In a separate analysis, we investigated if patients with very high FE-PRS are enriched for epilepsy with an unknown etiology, early onset seizures, or psychiatric comorbidities compared to the opposing category. We tested the top 5% of the FE-PRS distribution against the remainder, using a logistic regression model adjusted for sex and for the first four principal components of ancestry. For this analysis, the threshold for statistical significance after Bonferroni correction was set to \( \alpha = .017 \) (three groups tested). All statistical analyses were carried out in R (version 3.5.1).
3 | RESULTS

3.1 | Heterogeneous FE-PRS burden among phenotypic subgroups of individuals with FE

To explore if clinical phenotypes can identify individuals with FE that have different burdens of previously identified FE risk variants, we explored the PRS for FE in 10 clinically defined FE subgroups. We observed that four of the six categorical subgroups had significantly greater FE polygenic burden than 20,435 population controls: individuals with FE and an unknown etiology ($P = 5.2 \times 10^{-5}$, 0.73% phenotypic variance explained by the PRS), early onset seizures ($P = 4.7 \times 10^{-5}$, 0.64% phenotypic variance explained), with a psychiatric comorbidity (depression and/or anxiety) ($P = 1.3 \times 10^{-4}$, 0.66% phenotypic variance explained), and without a psychiatric comorbidity ($P = 3 \times 10^{-3}$, 0.4% phenotypic variance explained) (Figure 2A and Table S2). In contrast, individuals with a known etiology and those with late-onset seizures were not different from population controls after Bonferroni correction for multiple testing ($P > 5 \times 10^{-3}$). The average age at onset was 10.0 years.
(95% confidence interval [CI] 9.3-10.8) in the group with an early age at onset (n = 254) and 38.4 years (95% CI 36.4-40.5) in the group with a late age at onset (n = 160). We also performed an analysis where we scored all patients for FE-PRS and explored if individuals with high-risk variant load were enriched for specific clinical variables. We observe in the top 5% of the FE-PRS distribution a significant 5.33-fold enrichment of individuals with a psychiatric comorbidity (P = 5 × 10^−3, Figure 2C).

### 3.2 Improved power of the FE-PRS in subgroup defined by hierarchical clustering

The above-tested clinical variables are not discrete, and each individual with FE was binned into several groups. In the second approach, we explored if a combination of multiple clinical data points improves the identification of homogeneous FE subgroups. We used hierarchical clustering to partition individuals with FE into nonoverlapping clusters with similar clinical characteristics. The optimal cluster solution identified four clusters (maximum average silhouette width = 0.81). One of the four identified clusters (cluster 4, n = 105) showed significantly higher FE-PRS compared to controls (P = 4.3 × 10^−4, Figure 2A) and had the highest explained phenotypic variance among all classification groups (0.97%, Table S2). The cluster was enriched with individuals with early onset seizures, whereas all individuals had an unknown etiology and psychiatric comorbidities (Figure 2B). In contrast, the remaining three clusters did not show significantly higher FE-PRS compared to controls. One of the clusters (cluster 1, n = 105) was enriched with individuals with late-onset seizures, whereas all individuals had a known etiology and psychiatric comorbidities. The other two clusters (cluster 2, n = 105; cluster 3, n = 99) were enriched for individuals with early onset seizures, whereas all individuals had no psychiatric comorbidities and an unknown (cluster 2) or known etiology (cluster 3), respectively.

### 4 DISCUSSION

We could show recently that individuals with FE have a higher burden of previously identified common genetic variants associated with FE. Here we found that the polygenic burden is enriched in specific clinically defined subgroups of FE. Patients with an unknown etiology, early onset seizures, and psychiatric comorbidities (depression and/or anxiety) have a higher common genetic risk for FE than patients with opposed clinical features. The highest phenotypic variance explained by FE-PRS was observed in a subgroup defined by hierarchical clustering (cluster 4, Table S2). In this group, the phenotypic variance explained by FE-PRS was 2.9-fold higher than the variance explained in the remainder of the cohort (0.97% vs 0.34% weighted average), and 1.9-fold higher than the variance explained in all 535 patients of the FE-Cleveland-EUR cohort (0.51%), before consideration of phenotypic variables. The majority of all individuals in this subgroup had early onset seizures, whereas all had an unknown etiology and psychiatric comorbidities. Our results are in line with that of a family study in which relatives of affected individuals were found to have a higher risk for both focal and generalized epilepsies if they had an early age at onset and idiopathic epilepsy. Our results are also in line with evidence from other common disorders that show association of age of onset and PRS. The observation that all individuals of the cluster with the highest FE-PRS have psychiatric comorbidities, together with the 5.33-fold enrichment of individuals with psychiatric comorbidities in the top 5% highest FE-PRS, may be explained by shared genetically perturbed networks and pathways between common neurological and psychiatric disorders (including epilepsy, depression, and anxiety). In agreement with this hypothesis, we recently found that high FE-PRS is associated with psychiatric traits.

Our results should be interpreted in light of several limitations. First, the FE-PRS were derived from GWAS based on individuals with European ancestry. Our analyses are therefore restricted to individuals of European ethnicity, and generalizability to individuals of non-European ancestry remains to be determined. Second, because the age was unknown for a large proportion of the control population, we did not include age as a covariate in our analyses. Future studies that include age as a covariate could potentially generate different results because age can influence the effect sizes. Third, in the FE-GWAS that was used to generate the FE-PRS in this study, no information about comorbid psychiatric disorders or the age at onset was given. Thus we cannot rule out any potential effects of GWAS composition on our results. However, considering that the lesional, nonlesional, and unspecified FE cases represent each one-third of the FE-GWAS sample (Table S3), no specific FE subgroup is likely to be a source of bias. PRS derived from a GWAS in structural-metabolic epilepsy may lead to different results in our subgroup comparison. For example, stroke, which would be enriched in such a GWAS, has its own common genetic component, with suggestive evidence for genetic factors associated with post-stroke seizures. Future subtype-specific epilepsy GWASs should explore this hypothesis.

Future research should also investigate the interplay of common genetic variation with other factors such as epigenetic modifications. Differential patterns of epigenetic modifications across patients could further improve explanations of phenotypic differences observed in the different FE subgroups.

In summary, phenotypic features can be used categorically or in clustering analyses, to potentially identify genetically
more homogeneous groups of patients with FE. Identifying phenotype subgroups could increase the power of genetic studies. This approach has been shown successful in epilepsy before. More loci have been identified in generalized epilepsies compared to generalized and focal epilepsies combined. Similar observations have been made for many disorders and traits, for example neuroticism. Although our results are novel and indicate genetic heterogeneity in common variant space across clinically defined FE subgroups, clinical variant testing to decipher the clinical heterogeneity of FE is not possible. Larger GWAS studies in well-characterized cohorts are needed to develop better powered FE-PRS for clinical FE subtyping prior to the onset of comorbidities.

REFERENCES
1. Fiest KM, Saurø KM, Wiebe S, Patten SB, Kwon C-S, Dykeman J, et al. Prevalence and incidence of epilepsy. Neurology. 2017;88:296–303.
2. Skidmore C. Adult focal epilepsies. Continuum: lifelong learning. Neurology. 2016;22:94–115.
3. Perucca P. Genetics of focal epilepsies: what do we know and where are we heading? Epilepsy Curr. 2018;18:356–62.
4. Kasperavičiūtė D, Catarino CB, Matarin M, Leu C, Novy J, Tostevin A, et al. Epilepsy, hippocampal sclerosis and febrile seizures linked by common genetic variation around SCN1A. Brain. 2013;136:3140–50.
5. Guo Y, Baum LW, Sham PC, Wong V, Ng PW, Lui CHT, et al. Two-stage genome-wide association study identifies variants in CAMSAP1L1 as susceptibility loci for epilepsy in Chinese. Hum Mol Genet. 2012;21:1184–9.
6. ILAE Consortium on Complex Epilepsies. Genome-wide meta-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9:5269.
7. Leu C, Stevelink R, Smith AW, Goleva SB, Kanai M, Ferguson L, et al. Polygenic burden in focal and generalized epilepsies. Brain. 2019;142:3473–81.
8. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat Genet. 2019;51:404–13.
9. Kasperavičiūtė D, Catarino CB, Heinzen EL, Depondt C, Cavalleri GL, Caboclo LO, et al. Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study. Brain. 2010;133:2136–47.
10. ILAE Consortium on Complex Epilepsies, Avbersek A, Balding D, Baum L, Becker F, Berkovic SF, et al. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. 2014;13:893–903.
11. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, 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;51:676–85.
12. Karlsson EW, Boutin NT, Hoffnagle AG, Allen NL. Building the partners HealthCare biobank at partners personalized medicine: informed consent, return of research results, recruitment lessons and operational considerations. J Pers Med. 2016;6:2.
13. Choi SW, Mak TSH, O’Reilly PF. A guide to performing polygenic risk score analyses. bioRxiv. 2018. https://doi.org/10.1101/416545
14. Ottman R, Annegers JF, Hauser WA, Kurland LT. Seizure risk in offspring of parents with generalized versus partial epilepsy. Epilepsia. 1989;30:157–61.
15. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat Med. 2020;26:549–57.
16. Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360:eaap8757.
17. Leu C, Richardson TG, Kaufmann T, van der Meer D, Andreassen OA, Westlye LT, et al. Pleiotropy of polygenic factors associated with focal and generalized epilepsy in the general population. PLoS One. 2020;15:e0232292.
18. Dichgans M, Pulit SL, Rosand J. Stroke genetics: discovery, biology, and clinical applications. Lancet Neurol. 2019;18:587–99.
19. Sarecká-Hujar B, Kopyta I. Poststroke epilepsy: current perspectives on diagnosis and treatment. Neuropsychiatr Dis Treat. 2019;15:95–103.
20. Kobow K, Ziemann M, Kaipanickichal H, Khurana I, Mühlebner A, Feucht M, et al. Genomic DNA methylation distinguishes subtypes of human focal cortical dysplasia. Epilepsia. 2019;60:1091–103.
21. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. Item-level analyses reveal genetic heterogeneity in neuroticism. Nat Commun. 2018;9:905.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.