Characterizing the clinical and genetic epidemiology of functional seizures

By

Slavina Borisova Goleva

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Molecular Physiology and Biophysics

August 31st, 2021

Nashville, Tennessee

Approved:

Lea K Davis, PhD, Advisor

Nancy J Cox, PhD, Committee Chair

Kevin F Haas, MD, PhD

Bingshan Li, PhD

David C Samuels, PhD

James S Sutcliffe, PhD

To my incredible family:

To my mom, the wisest and most supportive human I know

To my dad, always a source of strength, positivity, and encouragement

To Aton, the king of empathy, always kind, comforting, and able to make me laugh

and

To my grandparents, who have given up so much for us and whose love and support has been unwavering through all these years

and

To my sweetie pies:

To Doug, who has been infinitely supportive during the completion of this PhD, and one of the only people who can consistently make me laugh while I'm crying

and

To Daisy, a constant source of joy in my life

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support of the Vanderbilt Molecular Endocrinology Training Program, the National Science Foundation Graduate Research Fellowship Program, the Vanderbilt Dean's Award, or the Vanderbilt University Graduate Fellowship. I am particularly grateful to Dr. Lea K. Davis, who has been unwaveringly supportive of my career development, academic goals, and life goals, regardless of what they may be. She has been a phenomenal role model for what a good scientist, and person, should be, and I wouldn't be where I am without her.

I am also grateful to all of those with whom I have had the pleasure of working on my thesis project and on other related projects in the lab. Thank you to all my lab members for their constant help, support, and friendship. Thank you also to my Dissertation Committee, who have provided me extensive personal and professional guidance, and who I have always felt were in my corner and there to elevate my research, to make me a better scientist and human, and to support me.

Thank you to my family, loved ones, and friends, who have been my most ardent cheerleaders throughout the pursuit of my Ph.D. You have helped me get through all the rough parts and come out on top for it. To my brother, parents, grandparents, to my partner in crime and in life, Doug, and to my teeny, tiny baby girl (slash dog), Daisy, thank you for all your unwavering, unconditional love and support. It is my favorite part of life on both my best and my worst days.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION*

Clinical and phenotypic characteristics of functional seizures

Functional seizures (FS) are paroxysmal episodes that often include altered awareness or convulsions with presentation like epileptic seizures. FS differ from epileptic seizures (ES) in that no evidence of aberrant electrical signaling patterns is evident on video electroencephalogram (v-EEG) and underlying psychological factors are thought to be involved in the etiology. There is currently no known underlying organic cause for FS.

Approximately eighty percent of FS patients are initially diagnosed with ES, which further delays the proper diagnosis. Many FS patients are prescribed antiepileptic drugs (AED), which appear to be neutral at best and detrimental at worst for treating FS.[1,2] The number of AEDs prescribed is associated with increased delay to diagnosis, particularly for patients in whom the first AED did not work, but who were continued on different AED trials despite the lack of response.[1] The gold standard to distinguish FS from ES is to use video-encephalogram recordings (v-EEGs) to capture both brain activity (by EEG) and convulsions (by video recording) throughout the onset and duration of a "typical seizure" for that patient.[3] A lack of aberrant brain signaling during the seizure episode is required for the diagnosis of FS.[4] Further clouding the proper diagnosis of FS and epilepsy is the reported literature suggesting that around a fifth of FS patients also have concurrent epilepsy.[1,5,6]

*Adapted with permission from Goleva SB et al., JAMA Netw Open, 2020

Research comparing FS and ES has shown that phenotypically, FS patients are more debilitated by their seizure disorder on average, and that the frequency of their seizures is significantly greater.[7] For example, Rawlings et al. described patients who had upwards of three hundred seizures over a four day span, making it difficult for patients to hold down steady jobs and significantly impacting their quality of life.

While the cause of FS remains unknown, it is believed that FS are primarily related to psychiatric distress and trauma is thought to be a primary risk factor. Previous studies, including a clinical epidemiology study from our group, confirmed the significant enrichment of psychiatric diagnoses in FS patients, with posttraumatic stress disorder and sexual assault trauma among two of the primary associations in FS patients.[8–11] Females constitute approximately 75% of FS diagnosed patients, which our own epidemiological studies replicated.[8,10–12]

Historical perspective on the diagnosis of FS

FS is often a diagnosis of clinical exclusion, and has historically been arduous to diagnose, with a current average of over 8 years to diagnosis.[1] In the DSM-V, FS is classified under conversion disorder, subtype "with attacks or seizures," and is defined as neurological symptoms that are incompatible with neuropathology. The lack of evidentiary diagnostic criteria for FS further impedes urgently needed clinical research. Moreover, the absence of consistent nomenclature or specific International Classification of Diseases (ICD) classification additionally hampers clinical recognition and treatment.

Numerous records and clinical publications dating from 1730 to 1901 show that FS were historically thought to be products of hysteria, particularly in females, or conscious manifestations of psychological disorders, rather than uncontrollable experiences.[13] In Mandeville's 1730 book, he writes the first known English description of functional seizures, then described as "hysterical seizures" or simply "hysterics,":

> "As to Fits, some are seized with violent Coughs; others with Hiccups; and abundance of Women are taken with Convulsive laughing. There are Fits that have short Remissions, in which you would think the Woman was going to recover, and yet last many Hours. Some are so slight that the Patients only lose the Use of their Legs and Tongue, but remain sensible; others again are so violent that those who are seized with them foam at the Mouth, rave and beat their Heads against the Ground; but whether they resemble an Apoplex, or are only fainting, or seem to be Epileptic, they all come under the Denomination of Hysteric … ."[14]

In an example of the sex discrimination that has historically been used against women to invalidate their experience of functional seizures, he goes on to describe his rationale that the "imbecility of the contexture of spirits in women" contributes to their development of hysterics:

> "We'll examine how much the Imbecility of the Contexture of Spirits in Women contributes to the cause of [Hysterics]. First, that it renders them all obnoxious to what

> is the immediate Cause of the Disorders in the Functions of the Brain and Nerves, or both ... But besides this Confusion of the Spirits to make the Distemper habitual, and render Women Hysterics of the first Class, there is required, and always observed, another Antecedent Cause, that can bring about the Confusion I speak of, without the Assistance of any external Violence. That is the Deficiency of the finer Spirits, which the stomachic Germent suffers by, whereof I have said so much: To the producing this Effect, the Imbecility of them is likewise so far accessary, that where there is anything to exhaust the Spirits, the Weakness of their contexture occasions it to be sooner accomplished; and the less Force serves to dissipate and destroy them: One Hour's intense Thinking wastes the Spirits more in a Woman, than six in a man. [I have] thus demonstrated how far the Weakness of the spirits disposes Women to the hysteric passion."[14]

As video-encephalograms were not yet available, in 1855, RB Todd described that the movements of "hysterical fits" were "combined and regular, and directed to an end and by a purpose." He claimed that a main distinguishing feature was that the patient "was a highly excitable, hysterical person, who has been subjected to moral, and perhaps physical influences also, well-calculated to keep up that state." In short, RB Todd's diagnosis favored the presence of psychological factors in the etiology.[15,16]

In 1884, Thomas Clifford wrote a scathing take on patients who were predisposed to "hysteria":

> "Take a hysterical person, man or woman, in its common and, so far, proper sense; take it to mean a person of feeble purpose, of limited reason, of foolish impulse, of wanton humors, of irregular or depraved appetites, of indefinite and inconsistent complaints, seeing things as they are not, often fat and lazy, always selfish, or to take it in less degree, one capricious, listless, willful, attractive perhaps, yet having always the chief notes of hysteria – selfishness and feebleness of purpose; and if such persons… have or have had anesthesia, unreal epilepsy, unreal syncope, unreal palsy, unreal cramps, then set down such a person as hysterical, but forget not, nevertheless, to cure mind and body.[17]

In 1923, Pierre Briquet posited that hysteria is predominantly a disease of women, as to fulfill her noble mission in life, woman has been endowed with great sensitiveness and is easily moved emotionally, which predisposes her to hysteria.[18] As the instance of "hysterical seizures" rose, Briquet hypothesized that "It is very probably that in several of these epidemics particular circumstances determined the form of the symptoms of the hysterical persons seized first. Then, through the involuntary influence upon the mind and the tendency to imitation, the hysterical persons who came afterwards had symptoms like those who had them first, and these in turn influenced the others." Thus, spreading the theory that functional seizures could be attributed to a powerful imagination.

Given the history of physician attitudes towards FS, it is not surprising that terminology for FS was overtly pejorative including hystero-epilepsy (1800s) and pseudoseizures (1964-

2000s), which stigmatized patients by implying underlying malingering or factitious disorder.[19] Today FS patient advocates and clinicians are lobbying for new nomenclature.

Nosology and naming of FS

To reduce stigmatization of patients and to fit with nosology of other movement disorders, there is a call to standardize the nomenclature for functional seizures.[19–21] Today, functional seizures are known by one of 7 terms: non-epileptic seizures, non-epileptic attack disorder, psychogenic non-epileptic seizures, conversion disorder with seizures, dissociative seizures, psychogenic non-epileptic attacks, or functional seizures.[19–21]. The inability to agree on a name to describe this disorder creates a difficulty for patients in understanding and accepting their illness without stigma. It additionally impedes doctors in explaining the condition to FS patients and impedes researchers in identifying FS patients for studies and in sharing their research with community audiences.

Throughout this research, I choose to use the term 'functional seizures' due to recent studies suggesting that patients prefer this term to describe their disorder and find it least stigmatizing.[19–21] Specifically, the Functional Neurological Disorder society found in a survey that patients ranked 'functional seizure' as their most favored term, and psychogenic non-epileptic seizures, which is currently the most commonly used name for this disorder among the medical community, as only their fourth favorite term.[19]. Another survey found that only 6% of FS patients found the term 'functional seizures' offensive, compared to 48% of patients who found the term 'hysterical seizure' offensive and 26% who found the term 'psychogenic non-epileptic seizure' offensive.[20]

Treatment modalities

Once patients are diagnosed with FS, treatment modalities available are extremely limited. Once diagnosed, FS patients may be treated with various psychiatric therapeutic modalities including cognitive behavioral therapy. Cognitive behavioral therapy is the only treatment modality linked to significantly improved outcomes, although randomized clinical trials show that it does not significantly affect the frequency of seizures in patients after one year of treatment.[22–24] To date, no pharmaceuticals have passed clinical trials for the improvement of seizures in FS patients.[22,24]

In the FS historical perspective published in Epilesia in 2004, W. Curt LaFrance Jr. and Orrin Devinsky point out that, "Although our treatments are refined in the sense that we psychopharmacologically target neurotransmitter abnormalities in [nonepileptic seizure] patients or treat them with individualized psychotherapies, after >200 years, we have not added one controlled trial to empirically test the outcome of our treatments for NES."[13]

Review of current epidemiological research on FS

FS remains understudied despite the fact that 20-30% of patients referred to epilepsy monitoring units are eventually diagnosed with FS.[25] The few published studies include small sample sizes, and limited epidemiological data.[26] Given this lack of research, the Commission on Neuropsychiatric Aspects of Epilepsy of the International League Against Epilepsy (ILAE) placed FS in the top three neuropsychiatric problems with anxiety/depression and psychotic disorder in 2011[12,27]. Based on this ranking, the "ILAE Psychogenic Non-Epileptic Seizures Task Force" was founded to summarize the current understanding, diagnosis, management, and treatment

for FS.[4,12,28] Given the challenging diagnosis of FS, lack of standardized nomenclature, limited treatment options, lack of awareness, and publications highlighting this issue, there is a compelling need for more research on FS.[12,13,29] There is a particular need to find ways to shorten the time to diagnosis or lead to advanced exploration of treatments for FS patients.

The prevalence of FS was previously indirectly estimated at 2-33 per 100,000 (0.002% - 0.033%),[30] and approximately 17 - 22% of FS patients have concurrent epilepsy.[5,6] FS patients reportedly have a higher rate of psychiatric disorders (e.g. PTSD, anxiety, depression, etc.) than both the general population and epilepsy patients (RR = 1.30, 95% CI = 1.14 - 1.48 across all psychiatric disorders studied).[31] Trauma is also known to be a significant risk factor and a recent study reported that females with FS were eight times more likely to report sexual abuse.[10]

Using electronic health records to study FS

Electronic health records (EHRs) have been used as a strategy to increase the identification of patients with certain disorders or phenotypes for research.[32,33] As EHRs contain structured and unstructured data on all patients who use any given hospital system, these records provide a large amount of data for clinical epidemiology research studies. EHRs contain electronic longitudinal data including demographics, billing codes, procedural codes, and clinical notes collected during a given patients' health care, all of which have associated dates facilitating research on the trajectory of various symptoms and comorbidities with respect to an index condition.

Challenges of using EHRs for research include ensuring and prioritizing patient privacy and data security and creating ways to use unstructured data, such as clinical notes, associated

with EHRs. To this end, VUMC has devised a de-identified version of the VUMC-EHR for research use named the Synthetic Derivative (SD). The SD is a clinical data warehouse for research that allows scientists to use structured data and search through unstructured clinical notes using keywords and regular expressions.[34]

FS studies have historically suffered from small sample sizes and a lack of epidemiological data. Thus, we determined that using EHR data would be a powerful way to quickly collect cases for clinical and genetic epidemiology studies of FS. However, due to stigmatization of the disease, traditional EHR phenotyping methods were difficult in this sample as there are no International Classification of Disease, 9th edition (ICD9) codes that are specific to FS patients. Therefore, the VUMC-EHR SD presented a particularly useful opportunity for us to develop an algorithm to reliably identify FS cases. We also recognized the potential to calculate the first-ever direct prevalence of FS. We were able to calculate period prevalence, or the proportion of people with FS at any time during a specific time interval. In our case, this was October 1989 to October 2018, the dates of records in the EHR at the time of data extraction.

The algorithm we developed to identify FS patients using VUMC-EHR records includes general seizure and conversion disorder ICD codes, FS keywords in patient charts, and Current Procedural Terminology (CPT) codes for v-EEGs.[8] Chart review by a team including a clinical neurologist demonstrated a positive predictive value of 98%. Our clinical epidemiology studies in the VUMC-EHR using these patients as cases confirmed previous reports of sex-bias and psychiatric comorbidities in FS patients.[8] Our direct calculation of FS patients in the Vanderbilt University Medical Center Electronic Health Record (VUMC-EHR) system estimated the period prevalence of FS to be much higher than previously thought, at 0.014%.[8]

Potential role of genetics in FS

The first genetic study on FS presented whole-genome sequencing to determine that six percent of their sample harbored pathogenic or likely pathogenic variants, including in the genes NSD1 and GABRA5.[35] This genetic burden was similar to individuals with generalized epilepsy (GE) (2%) and focal epilepsy (FE) (3%). Previous genome wide association studies (GWAS) have also shown that disorders related to FS, including GE, FE, and PTSD all have a significant, though small, heritable component, with SNP-based heritability most recently reported at 32%, 9%, and 5%, respectively.[36,37] Although the SNP-based heritability estimates for PTSD and FE are low, this establishes that these disorders have some genetic contribution. Thus, we hypothesized that FS is heritable, meaning that a significant portion of the variance in the phenotype of functional seizures can be explained by genetic variance.

To explore this hypothesis, we used a genome-wide association study (GWAS) approach to determine whether any genetic variants are associated with functional seizure cases versus controls. To reduce the cost associated with whole genome sequencing, genotyping is used to genotype only common genetic variants (typically MAF > 0.01). For polygenic traits in which a high number of variants with low effect size contribute to the underlying genetic architecture of the trait; GWAS are an efficient way to scan common genetic variants across the entire genome.[38,39] Many psychiatric and neurological diseases are polygenic and heritable, thus we hypothesized that this technique may be successful in determining the genetic architecture of FS as well.[36,37,40–43] Using GWAS analyses, we aimed to (1) identify the variants and genes associated with FS, (2) calculate a polygenic risk score associated with FS cases vs controls, and (3) determine the genetic heritability of FS. Incorporating the validated VUMC-EHR algorithm,

we conducted a multi-site FS GWAS meta-analysis, which is also the first FS GWAS reported to date.

Studies show that other psychiatric and neurological disorders are polygenic, which have an inherent reduction of discoverability. For example, a book by Ted Abel and Thomas Nickl-Jockschat recently demonstrated that GWAS for Crohn's disease, for which GWAS was shown to be a successful strategy, discovered around 50 loci associated with Crohn's disease with a sample size of around 5,000.[44] In comparison, bipolar disorder and schizophrenia discovered less than 5 associations each with the same sample size. This indicates that BIP and SCZ, which are both psychiatric conditions, have a more polygenic genetic architecture, in which more variants are involved in the disease and each one has a smaller effect size. For SCZ GWAS, there was a critical inflection point around 15k sample size, above which the relationship between sample size and number of novel discoveries became much more linear. Since FS is more similar phenotypically to SCZ and BIP than Crohn's disease, we expected that we would need a significant sample size before being powered to discover variants associated with FS using GWAS. As the number of FS samples available using our own genetic data in BioVU was limited to around 300, we determined that conducting a GWAS meta-analysis using genetic data from six biobanks around the world would be a better strategy to increase our power for discovery.[45] Results include cases and controls identified from our BioVU samples, the Cleveland Clinic, Mount Sinai (BioMe), the Million Veterans Program (MVP), Massachusetts General BioBank (MGBB), and the Danish population registry (iPSYCH).

Overall, this thesis characterizes the clinical and genetic epidemiology of FS in clinical populations. We utilized a database of de-identified electronic health records (EHR) from

Vanderbilt University Medical Center (VUMC) to (1) develop a clinically validated phenotyping algorithm to identify FS cases, (2) estimate the prevalence of FS in a hospital population, (3) systematically identify FS comorbidities, and (4) investigate the relationship between sexual assault-trauma and FS. We then guided phenotyping efforts to incorporate data genomic data on FS cases and controls from six other biobanks and registries to (5) determine the SNP-based heritability of FS, and (6) identify individual single nucleotide polymorphisms (SNPs) and genes associated with FS.

CHAPTER II

CLINICAL EPIDEMIOLOGY OF FUNCTIONAL SEIZURES*

Introduction

In this chapter, we characterize the clinical epidemiology of FS in a hospital population. We utilized a database of de-identified electronic health records (EHR) from Vanderbilt University Medical Center (VUMC) to identify FS cases and controls. This epidemiological study used an automated phenotyping algorithm to identify cases and controls by incorporating International Classification of Diseases (ICD) codes, Current Procedural Terminology (CPT) codes, and regular expressions. Most FS cases identified were ascertained through primary care or the epilepsy monitoring unit. Here, we address the following research questions: (1) what is the prevalence of FS in our hospital sample, (2) do we observe known comorbidities of FS in this sample, (3) what novel comorbidities are identified in FS cases, and (4) what is the relationship between sexual assault trauma and FS in this clinical population?

Methods

Sample and Data Description

The VUMC Electronic Health Record (VUMC-EHR) system contains inpatient and outpatient data from the medical center spanning labs, billing codes, medications, chart notes, and procedural codes, among other medical information, from 1994 onward.[34] The synthetic

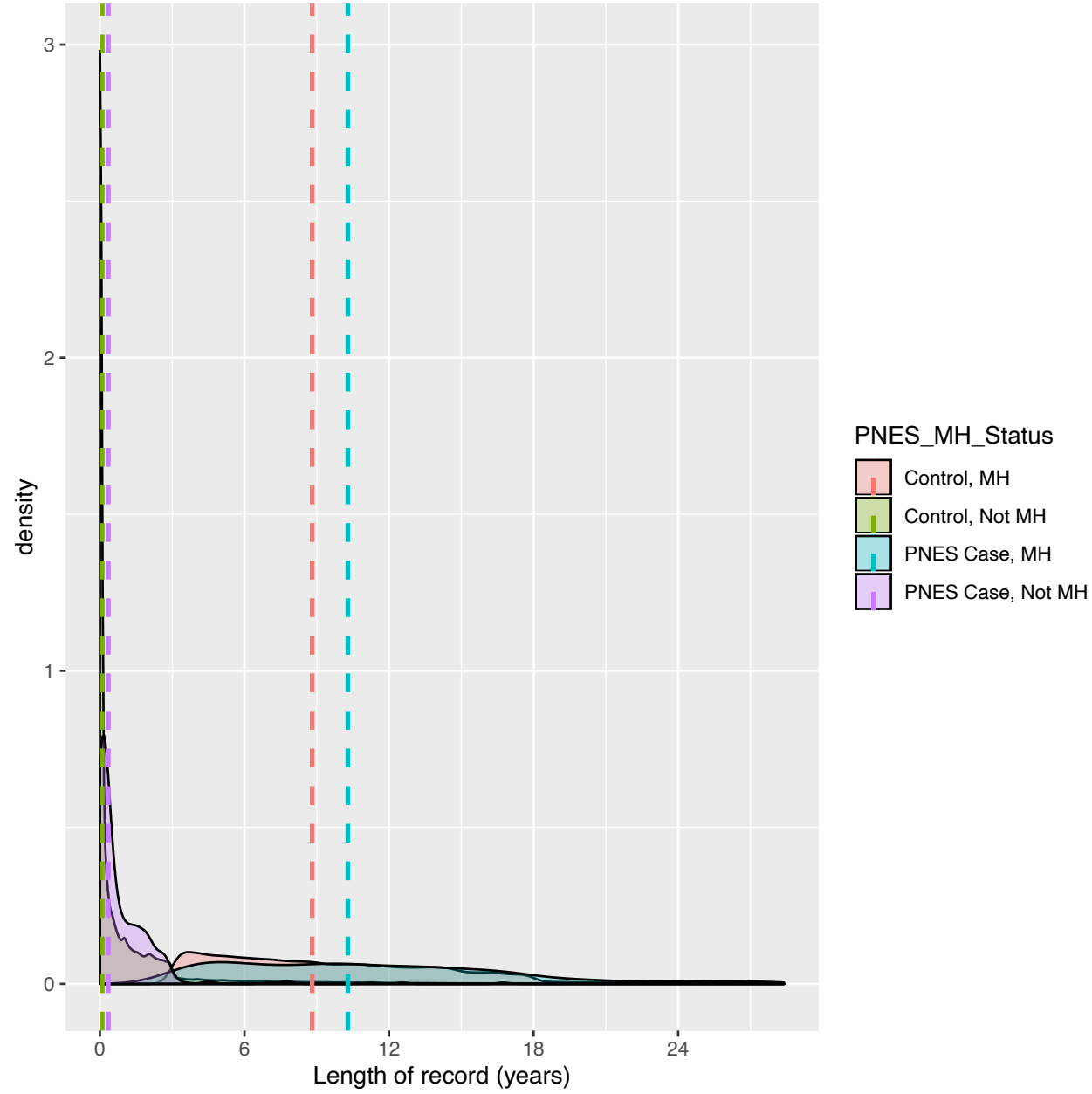
*Adapted with permission from Goleva SB et al., JAMA Netw Open, 2020

derivative (SD) is a mirror image of the VUMC-EHR which is used for research and has been de-identified to protect patient privacy. During this process, the medical record number from each patient's record is removed and replaced with a research unique identifier which cannot be linked to the original medical record number, as no link is maintained between the two. Then, any identifying features in the medical record are scrubbed or replaced with non-identifying replacements. For example, an algorithm is used to scrub any instances of names in the record using census-derived name dictionaries and a function which derives the name information from the header files of the original medical records. Additionally, dates are changed by a randomly generated amount of time within a similar timeframe. The aggregate error rate has been reported to be 1.7%. As de-identification algorithms can never perform perfectly, a further safeguard is implemented in that all researchers who use the SD are mandated by the terms of a data use agreement.

Our study population included all 2,346,808 unique VUMC patients from 1994 to 2019.[34] Demographics, ICD9 and ICD10 codes, CPT codes, and clinical notes were extracted from the synthetic derivative (SD), a de-identified copy of the medical record, and mined for subsequent analyses. For the phenome-wide association study, we restricted the study population to adults (>=18 years at the end of the medical record) with at least five ICD9 or ICD10 codes on different days over at least three years to exclude individuals who did not receive regular health care at VUMC and reduce bias due to nonrandom missingness between cases and controls (medical home; Figure 1A). A total of 1,910 out of 3,341 FS patients did not meet criteria for the medical home definition and were not included in the PheWAS, resulting in a total of 1,431 patients included in the analysis. Related ICD codes were organized into hierarchical code families (i.e.,

"phecodes") that represent broader categories of disease (https://phewascatalog.org/phecodes).[46] This research study was reviewed and approved by VUMC IRB #181185 and #190085.

A.



B.

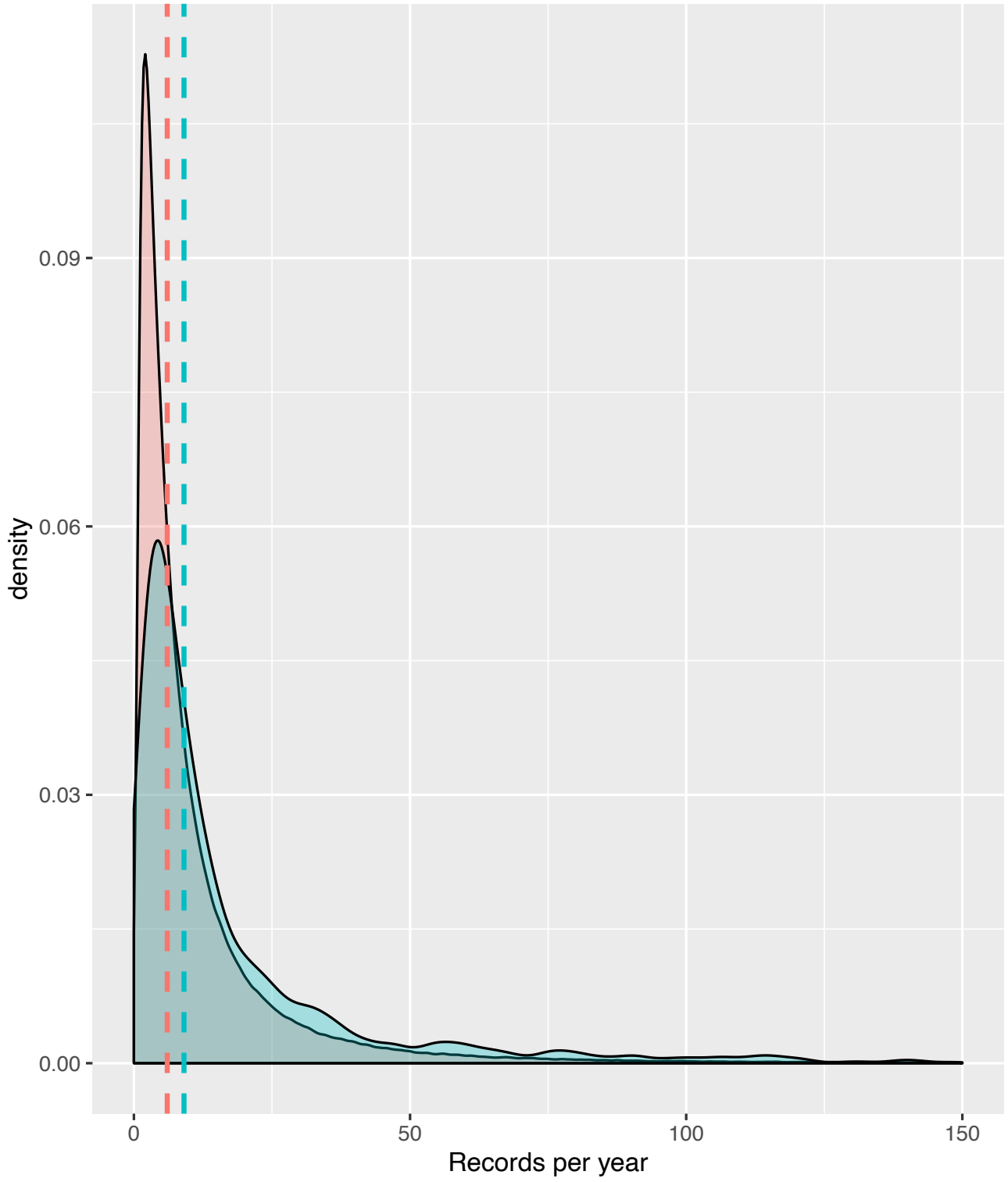


Figure 1. FS patients have longer, more dense medical records as compared to controls. A. Density plot of length of record is plotted in years for FS cases (blue, purple) versus controls (red, green) in both medical home (red, blue)

and outside of it (green, purple). Additionally, the median value for each group is plotted as a vertical dashed line. B. Density plot of records per year in FS cases (blue) and controls (red) within the medical home. Median values are plotted as vertical dashed line for both groups.

FS case/control algorithm

Two separate FS algorithms were developed under supervision of a clinical neurologist, Kevin F. Haas, MD, PhD (KFH), to reduce bias during algorithm development and chart review. The primary algorithm was used to identify FS patients without concurrent epilepsy (Figure 2, Table 1). The secondary algorithm identified FS patients without exclusions for epilepsy (Figure 3, Table 2).[47] All analyses were done in both algorithms independently, but as all results remained similar, only results from the primary algorithm are presented. 100% of patients in the primary algorithm are also included in the secondary algorithm. Future directions may include conducting these analyses on patients who have both epilepsy and FS, which would not have been included in the primary algorithm and would be specific to the secondary algorithm. The primary algorithm was validated by unblinded manual chart review of 50 algorithm-identified cases.

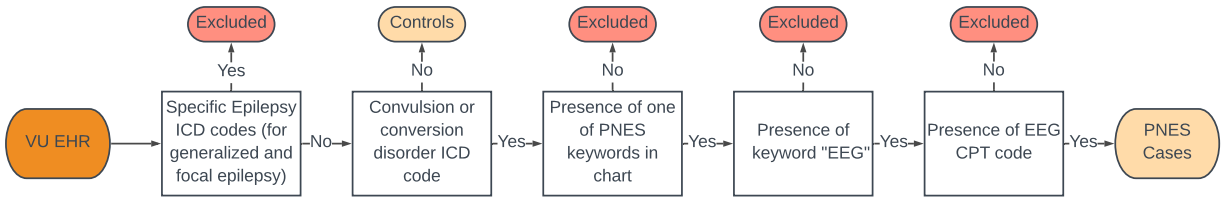


Figure 2. Algorithm for detecting FS cases and controls within the VUMC EHR. All VUMC EHR patients were initially included, then anyone with a generalized or focal epilepsy ICD code was excluded. Patients without convulsion or

conversion disorder ICD codes were considered controls, while anyone with both a FS keyword in their chart and the presence of the keyword EEG was included as a FS case.

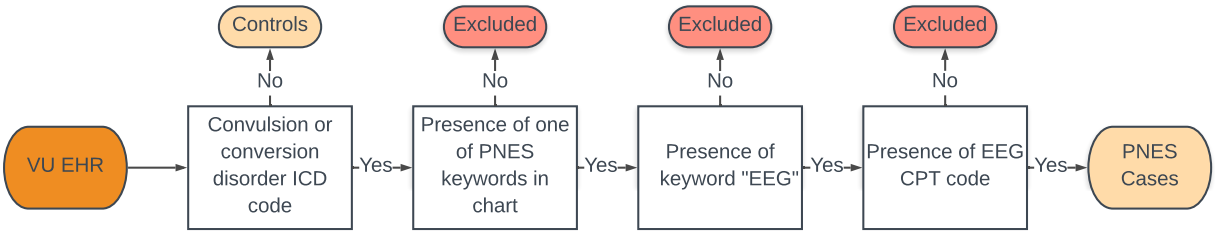


Figure 3. Flow chart of algorithm used to identify FS patients irrespective of their epilepsy case/control status.

FS inclusion keywords (>= 1)	**Inclusion ICD codes (>= 1)**	**Inclusion CPT codes (>= 1)**	**Additional inclusion criteria**	**Exclusion Epilepsy ICD codes in group (>= 1)**
"Psychogenic non-epileptic" *(or nonepileptic or non epileptic)*	300.11 – conversion disorder	Group 95812-95830: Routine Electroencephalography (EEG) Procedures	Genotyping data available	345.1
"Pseudoseizure"	780.39 – other convulsions	Group 95950-95967: Special EEG Testing Procedures	Presence of keyword "EEG" in chart	345.10
"Psychogenic seizure"	R56.9 – unspecified convulsions	..	..	345.4
"Non-epileptic seizure" *(or nonepileptic or non epileptic)*	F44.5 – conversion disorder with seizures or convulsions	..	..	345.40
" PNES "	..	..	..	345.5
..	..	..	..	345.50
..	..	..	..	G40.20
..	..	..	..	G40.30
..	..	..	..	G40.00
..	..	..	..	G40.10

Table 1. Criteria for algorithm to identify FS patients, excluding those who have concurrent epilepsy. Inclusion ICD codes, inclusion keywords and other criteria, and exclusion ICD codes are all listed.

FS inclusion keywords (>= 1)	Inclusion ICD codes (>= 1)	Inclusion CPT codes (>= 1)	Additional inclusion criteria
"Psychogenic non-epileptic" *(or nonepileptic or non epileptic)*	300.11 – conversion disorder	Group 95812-95830: Routine Electroencephalography (EEG) Procedures	Genotyping data available
"Pseudoseizure"	780.39 – other convulsions	Group 95950-95967: Special EEG Testing Procedures	Presence of keyword "EEG" in chart
"Psychogenic seizure"	R56.9 – unspecified convulsions	..	..
"Non-epileptic seizure" *(or nonepileptic or non epileptic)*	F44.5 – conversion disorder with seizures or convulsions	..	..
" PNES "	..	..	..

Table 2. Criteria for algorithm to identify FS patients regardless of epilepsy status. Inclusion ICD codes, inclusion keywords and other inclusion criteria are all listed.

Both algorithms were created by filtering criteria joined with Boolean operators. First, we required one or more convulsion or conversion disorder (ICD9 codes 300.11 OR 780.39 OR ICD10 codes R56.9 OR F44.5)[47] AND the presence of regular expressions indicating FS-related keywords in charts ("pseudoseizure," "psychogenic seizure," "nonepileptic seizure", or " pnes "). Although FS can present as syncopal events, we have not included this phrase in our algorithm because in patient charts such as the ones we mined, this would be a nonspecific symptom and would add noise to our algorithm. Next, we required the presence of both the keyword "EEG" AND one or more EEG CPT code (Group 95812-95830: Routine Electroencephalography (EEG) Procedures and Group 95950-95967: Special EEG Testing Procedures) to ensure that patients had an EEG performed at VUMC. Eighty-two percent of potential FS cases with an "EEG" keyword also had an EEG procedure CPT code present in their chart.

Extraction of sexual assault trauma from the EHR

We identified sexual assault trauma patients in the VUMC-EHR using regular expressions identification in clinical notes (Supplementary Methods). While it is likely to be under-reported at the time of the assault, a history of sexual assault can be extracted from the medical record. Based on an initial review of 25 charts, which were randomly selected, we identified eight inclusion phrases (i.e., "his/her sexual assault", "history of sexual abuse", "history of rape", etc.) and four exclusion phrases (i.e., "denies history of sexual abuse", "no hx of sexual assault", etc.) to identify cases (Table 3). Although ICD codes do exist for sexual assault trauma, they were infrequently used. The primary mode of identification of patients who experienced sexual assault trauma in the EHR was through the identification of regular expressions (e.g., "history of sexual abuse", "was raped") within clinical notes.

Include	**Exclude**
history (OR hx OR h/o) of sexual abuse	no history (OR hx OR h/o) of sexual abuse
history (OR hx OR h/o) of sexual assault	denies history (OR hx OR h/o) of sexual abuse
sexual (OR sexually) abuse(d) by	no history (OR hx OR h/o) of sexual assault
reports (OR reported) a rape	denies history (OR hx OR h/o) of sexual assault
her (OR his) rape	
was raped	
sexually abused him (OR her)	
secondary to rape (OR sexual abuse OR sexual assault)	

Table 3. Criteria to identify sexual assault trauma patients. Patients were exclusively identified using natural language processing of patients' clinical notes using various inclusion phrases and exclusion phrases. The inclusion phrases and exclusion phrases used are both reported here.

Sexual assault trauma mediation of FS:

We fitted a multivariable logistic regression model to adult FS cases and controls to determine whether sexual assault trauma, female sex, or the interaction between them was correlated with FS case status. Additional covariates included median age of record, median BMI, race, and record density. Next, we performed mediation analysis using the 'mediation' R package to test whether females were more likely to develop FS because females are more likely to experience sexual assault trauma.[52]

Because sex is randomly determined at conception, the assumption of random assignment to the "treatment" condition holds, thus the patient sex (male/female) was coded as the "treatment" variable. Sexual assault trauma was then coded as the mediator and FS as the outcome in a model-based mediation analysis.[52] We used 1,000 bootstrap comparisons to determine empirical confidence intervals and determine statistical significance of the mediation model. Additionally, we ran the mediation analysis with four different randomly generated seeds to ensure that the results and empirical p-values of the analysis remained stable across all permutations.

Chart Review

Fifty charts were randomly selected from the FS algorithm identified cases and were reviewed by one rater (SBG), who was trained by a clinical neurologist (KFH).

The charts were reviewed for positive clinical diagnosis of FS based on a publication by LaFrance et al describing the minimum requirements for the diagnosis of FS.[4] Positive predictive values (PPVs) were calculated for two groups based on this paper: (a) documented, clinically

established, or probable FS; and (b) possible FS. Based on this paper, the first criteria for a positive diagnosis of documented, clinically established, or probable FS was defined as having chart history characteristics consistent with FS. The second criteria was having a seizure witnessed by a clinician who either 1) reviewed the video EEG recording and found the semiology typical of FS, or 2) is experienced in diagnosis of seizure disorders and found the seizure to show semiology typical of FS, regardless of whether or not it was on v-EEG. The third criteria was that there was no epileptiform activity in routine or sleep-deprived interictal EEG, in routine or ambulatory ictal EEG during a typical event, or immediately before, during or after ictus captured on ictal video EEG. A positive diagnosis for possible FS was defined as having a history consistent with FS, event matched FS semiology by witness report or self-report/description, and that there was no epileptiform activity in routine or sleep-deprived interictal EEG.[4] KFH was consulted as needed to adjudicate.

Prior to review of these 50 charts, SG reviewed 100 charts, and KH also reviewed 37 of these charts. The goal of this first round of chart review was to train the reviewer (SG) to identify key phrases and language used by VUMC neurologists to interpret video EEG and to confirm diagnoses. SG, KH, and LKD sat in conference for approximately 4 hours to discuss the levels of evidence observed and the certainty of diagnosis given video EEG results.

For the sexual assault trauma analysis, AL manually reviewed 52 charts to identify the presence of confirmatory statements describing a history or current experience of sexual assault trauma.

Standard error (SE) and confidence intervals (CI) for the positive predictive value of each algorithm was calculated using the following formulas, where TP = number true positive, FP = number false positive:

$$\mathrm{SE} = \sqrt{\frac{\mathrm{PPV} * (1 - \mathrm{PPV})}{\mathrm{TP} + \mathrm{FP}}}$$

$$\mathrm{CI} = \mathrm{PPV} \pm 1.96 * \mathrm{SE}$$

Positive Predictive Value (PPV=total true cases/total algorithm identified cases*100) was calculated to determine the accuracy of the phenotyping approach for both FS and sexual assault trauma. The FS charts were reviewed for documented evidence of a clinical diagnosis of FS and classified as either (a) documented, clinically established, or probable FS; or (b) possible FS.[4] Additionally, mean and standard error of the reported time from seizure onset to FS diagnosis were calculated using data extracted during chart review.

Development of Medical Home Population

The total study population used for algorithm development and prevalence calculations included the entire hospital population (N = 2,346,808). However, the phenome-wide association study was restricted to a subset of this total population called the "medical home" population. The medical home is a heuristic definition that restricts the total sample to a subpopulation with more complete medical record data to facilitate the comparison of comorbidity patterns between cases and controls. The definition applied required the presence of at least 5 codes over a period of at least 3 years. The reason for this restriction is to ameliorate the impact of missing data in controls who may have a much shorter medical record

which can result in a biased upward estimate in the regression coefficients in a case vs. control comparison in the PheWAS.

Phenome Wide Association Studies (PheWAS)

Phenome Wide Association Studies (PheWAS) were performed to systematically assess the co-occurrence of FS and all other phenotypes across the adult medical phenome after covariate adjustment. ICD-9 codes were mapped to 1,814 phecodes (outcomes) as described and validated using a hierarchical grouping system by the Phecode Map v1.2 project.20 For each of the 1,653 phecodes , any individual with two or more phecode-mapped ICD-9 codes was assigned case status for the corresponding phecode, while those with no phecode-mapped ICD-9 codes were assigned as controls, and those with only one phecode-mapped ICD-9 code were excluded from the analysis of that phecode. ICD-10 codes were not included as, at the time, there was not yet a validated structure for ICD-10 to phecode mapping, and since most of the data uses the ICD-9 billing data. We required a minimum number of 100 cases for each phecode. We fitted 1,653 logistic regressions to test the relationship between algorithm-defined binary FS case/control status (predictor variable) and each phecode (outcome) after adjusting for potential mediators and confounders including sex, age (defined for each individual as median age across their medical record), density of records (in records per year), and EHR-reported race.[46,51] PheWAS results were considered statistically significant if the p-value of the association passed a Bonferroni corrected threshold for multiple testing ($p<0.05/1,653=3.02E-05$).

PheWAS of FS compared to controls with epilepsy

We compared algorithm-identified FS cases (n = 1,431) to controls with epilepsy (n = 4,715). Controls with epilepsy were defined as anyone who had either an ICD10 code from group G40 or an ICD9 code from group 345 for either generalized or focal epilepsy, as previously reported.[47,48] Then, we conducted a PheWAS where the independent variable was FS cases status vs. controls with epilepsy using all of the same covariates. Finally, we repeated the analysis a final time and compared epilepsy cases (4,175) to controls (n = 496,890). Here, controls were defined as all other patients, excluding FS cases, FS patients meeting exclusion criteria for FS, and any patients with either one or two GE or FE ICD codes.

FS comorbidity sex-differences analysis

We performed a phenome-wide FS by sex interaction analysis to determine whether comorbidities were more common among males or females with FS after accounting for any baseline sex-difference in the prevalence of the comorbidity. A minimum number of 100 cases of any given phenotype were required for inclusion in the analysis. Phenotypes that exceeded a Bonferroni corrected interaction p-value ($p < 3.02E-05$) demonstrated a significant diagnosis by sex interaction. Although all phenotypes were included in the analysis, any significant phenotypes which were sex-specific were manually excluded.

FS date of diagnosis analyses

For each patient identified as a FS case, we established the date of the first clinical suspicion of FS, defined as the first mention of a FS keyword in clinical notes. We used the date

of EEG administration (with a FS keyword appearing in the patient's chart within 30 days) to define the date of the diagnosis of FS. We then determined the average time from the first mention to the diagnosis for algorithm defined FS cases. Furthermore, during the manual chart review we documented patient-reported onset of seizures.

Cerebrovascular disease identification

Based on initial results from the PheWAS, we conducted further follow-up analysis of the class of cerebrovascular disease conditions. Date of diagnosis of cerebrovascular disease was identified using the date of the first ICD code of a list of ICD9 codes that were found to be significantly associated with FS and epilepsy (all of which overlapped). The ICD9 code list was as follows: 430, 433, 433.8, 430.2, 433.2, 433.21, 433.3, 433.31, 430.3, 433.6, 430.1, 433.1, 433.5, 433.11, 433.12.

Temporal analysis between FS and cerebrovascular disease:

We performed additional analyses to further explore the temporal relationship between FS (and epilepsy) diagnosis and coded cerebrovascular disease (CVD) after our PheWAS revealed an association between FS and CVD (Supplementary Methods). We determined whether CVD or FS was more likely to occur in the EHR first, then whether there was any difference in the common comorbidities between those patients who presented with FS first or CVD first. For patients with both FS and CVD, we calculated the median (and interquartile) age at the first ICD code for CVD and the number of years from FS diagnosis to first CVD code. We then binned patients into three groups (a) FS diagnosis >90 days before CVD, (b) FS diagnosis

within 90 of CVD diagnosis, and (c) CVD diagnosis >90 days before FS diagnosis. We compared these three groups of FS+CVD patients on seventeen different clinical and demographic features (Table 7). We performed parallel analyses to examine the temporal relationship between CVD and epilepsy and to test for differences in the same seventeen clinical features that were examined in FS+CVD cases.

Calculation of BMI from the EHR

Previous publications indicate that body mass index (BMI) may be higher in patients with FS.[49] Therefore we included BMI as a covariate in several analyses. BMI data was obtained from the EHR. All individuals had multiple BMI measurements from various visits to the VUMC. BMI in the EHR is prone to recording errors.[50] To remove erroneous BMI records, we first z-score scaled the longitudinal BMI measurements collected within each individual, then removed any measurement with a Z-score below -3 or above 3. Next, we calculated median BMI value for everyone.[50]

Using the cleaned median BMI values for everyone with available data, we calculated the mean and SE of BMI for FS cases without epilepsy as defined by our algorithm (n = 1,605), for FS cases with epilepsy as defined by our second algorithm (n = 1,058), and for controls (n = 488,398) as defined by our algorithm. We also calculated the mean and SE BMI value for epileptic patients (n = 6,186) for comparison and replication of prior studies.[49] We used the R package ggplot2 for plotting, and a one-way ANOVA was performed in R followed by a Tukey post-hoc analysis to determine statistical significance of the difference in BMI between each group.

It is also worth noting that all of these analyses were performed separately with and without median BMI included as a covariate. However, results for analyses excluding the median BMI covariate are not shown as they were very similar to the analyses including median BMI as a covariate.

Results

Calculation of Positive Predictive Value (PPV) from automated EHR-based FS phenotyping algorithm

Thirty-five of 50 (70%; 95% CI = 69.18% - 70.82%) FS designated charts reviewed met criteria to be considered "Documented, Clinically Established, and Probable FS cases", while 14 (28%; 95% CI = 27.21%-28.79%) of cases were "Possible FS cases", and only 1 (2%) case was considered "No FS/Not enough information".

EEG keyword presence in charts and CPT code presence are strongly correlated

Across the entire VUMC-EHR, 66,936 patients had an EEG CPT code, 109,523 patients had "EEG" written in their charts, and 62,249 patients had both an EEG CPT code and keyword EEG in their charts. Approximately 93% of patients with an EEG CPT code also had a clinical note with the EEG description or interpretation and approximately 57% of patients with an EEG keyword also had an EEG code in their chart.

Automated EHR-based sexual assault trauma phenotyping algorithm yields a PPV of 90%

AML manually reviewed 52 charts identified as cases by the sexual assault trauma phenotyping algorithm and calculated the positive predictive value of the algorithm. Charts were identified as true positives if they clearly contained a description of the patient's reported history of sexual assault. In the 5/52 charts identified as false positives, inclusion phrases were present either as negation of sexual assault history (e.g., "she denies a h/o sexual abuse") or in reference to someone other than the patient (e.g., "her sister was sexually abused"). Using this approach, AML determined the PPV of the sexual assault trauma phenotyping algorithm was 90.38% (CI = 82.37% - 98.39%).

Chart review to determine the average time from seizure to diagnosis

Thirty-two charts of the fifty randomly selected for manual chart review included written descriptions of patient-reported date of seizure onset. As SBG reviewed the charts, she made note of the patient-reported date of seizure onset. Based on this data, the mean time from seizure onset to diagnosis of FS was 6.6 (± 1.4) years.

Descriptive medical record statistic in medical home vs. non-medical home cases and controls

Within the entire VUMC-EHR, FS cases (purple) have a longer length of record than controls (green; Figure 1A). This difference remains when restricted to medical home (MH) status, where FS cases (blue) have an even longer length of record on average than controls (red). Additionally, we see that within the medical home, FS patients (blue) have higher density of record (mean=9.07, SE=0.95) than controls (mean=6.03, SE=0.03; red; Figure 1B). In other words, FS cases get diagnosed with more billing codes per year than their control counterparts.

Combined with their longer length of records, this indicates that throughout their medical records, FS cases get coded with more ICD codes than controls.

Calculation of FS prevalence in a clinical population

The FS algorithm identified 3,341 patients >18 years old with FS in the VUMC-EHR, out of a total patient population of 2,346,808 (0.14%). 752,024 VUMC patients met our definition for "medical home". Of the total FS case sample, 1,431 met the definition for medical home (Table 4). The period prevalence of FS patients in the adult medical home population was 0.27% (1,431 out of 523,593) and 74% of the algorithm-identified FS medical home patients were female (Table 4).

	FS Controls	FS Cases	OR (95% CI)	Beta (SE)	p-value
N VUMC-EHR		3,341			
N Medical Home	502,200	1,431			
Female sex, N (%)	299,472 (60)	1,062 (74)	1.75 (1.55-1.99)		1.61E-18*
Race, N (%)					
Caucasian	417,523 (83)	1,264 (88)	1.97 (1.11-3.47)		0.02
African American	56,549 (11)	166 (12)	1.02 (0.51-2.05)		0.96
Asian	7,581 (2)	13 (1)	0.35 (0.09-1.27)		0.11
Unknown/Other	18,097 (4)	30 (2)	0.33 (0.08-1.26)		0.10
Native American	1,992 (0.004)	10 (1)	0.13 (0.01-1.10)		0.06
Ethnicity, N (%)					
Hispanic/Latino(a)	9,644 (2)	23 (2)	0.38 (0.22-0.65)		5.20E-04*
Unknown	19,176 (4)	37 (3)			
Records per year, Median (Q1 – Q3)	6.03 (2.88 – 13.15)	9.07 (4.05 – 20.89)	..	10.26 (0.61)	1.72e-63*
Length of record (years), Median (Q1 – Q3)	8.81 (5.56 – 12.99)	10.29 (6.26 – 14.68)	..	1.13 (.14)	1.68e-16*
Number of ICD codes, Median (Q1 – Q3)	49 (23 – 113)	80 (35 – 215.5)	..	124.25 (5.21)	1.84e-125*
Age, Median (Q1 – Q3)	59.58 (45.02 – 72.48)	49.31 (39.40 – 59.87)	..	-9.77 (0.46)	3.4e-104*

Median Age of Record, Median (Q1 – Q3)	51 (36 – 65)	40 (30 – 51)	..	-1.05 (0.46)	6.6e-124*
Median BMI (Q1 -Q3)	27.8 (24.2 – 32.5)	28.7 (24.0-34.9)	..	1.04 (0.19)	1.27E-7*

Table 4. Algorithm-defined FS cases and controls demographics. Number of adult cases identified by FS algorithm within the VUMC EHR and how many of those meet a medical home definition, or those who have at least 5 ICD codes on different days over the span of at least 3 years. We have further shown the proportion of males and females within each of these categories. Around 74% of FS cases identified by our algorithm are female (compared to 60% females in controls), which mirrors previously reported FS demographics. We have also shown the race, ethnicity, density of records, and age demographics in both algorithm-defined FS cases and controls. * indicates that the statistical analysis exceeded Bonferroni multiple-testing correction.

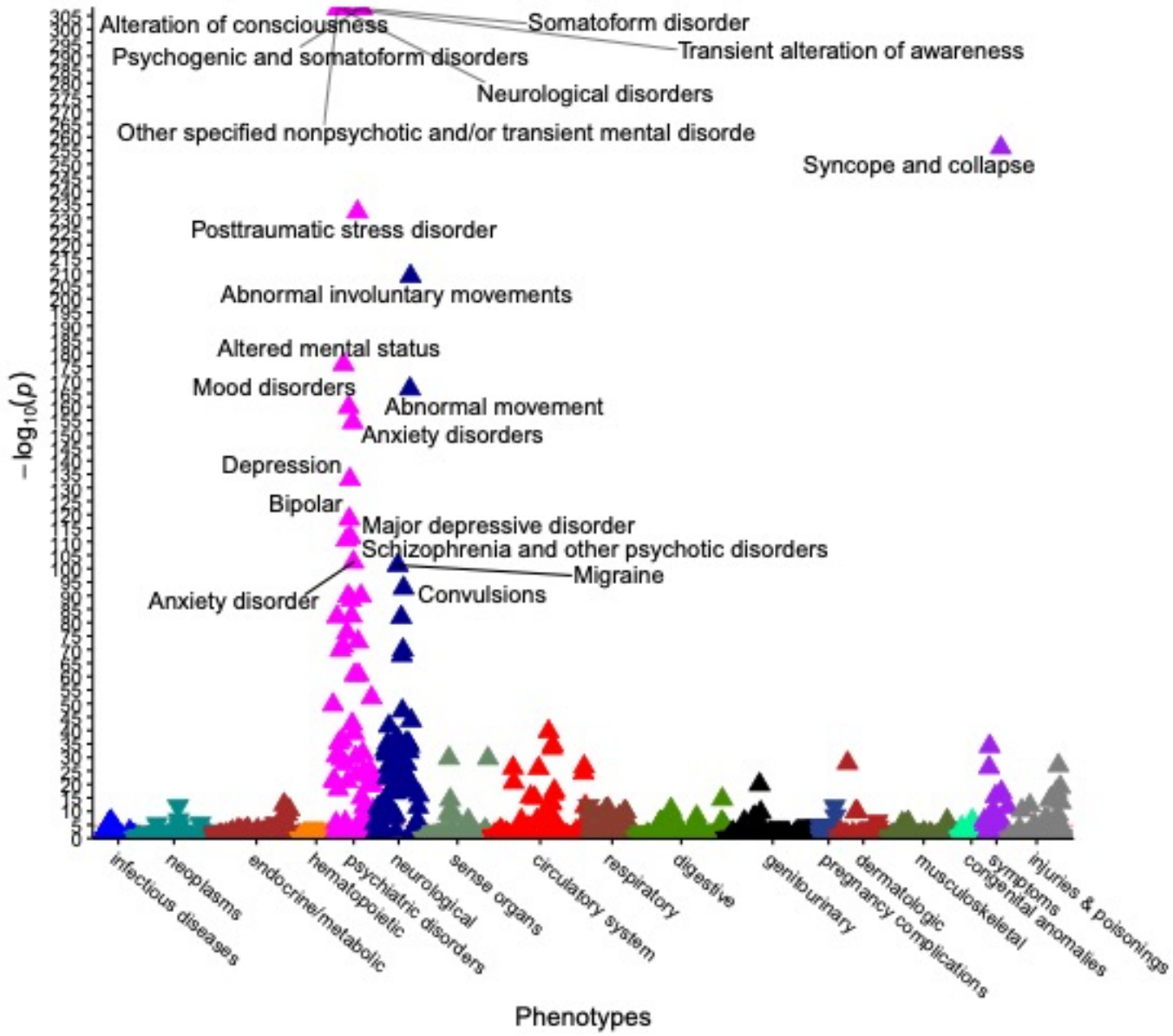
Demographic differences in FS cases and controls

Linear or logistic regression was applied to determine the effects of FS status on each outcome presented in Table 4 while covarying for all other variables presented. Logistic regression was used for binomial variables while linear regression was used for all quantitative variables. Results indicated that FS cases are significantly more likely to be female (OR = 1.75, 95% CI = 1.55-1.99, p = 1.6e-18). FS cases were more likely to be Caucasian (OR = 1.97, 95% CI = 1.11-3.47, p = 0.02) and less likely to be Hispanic (OR = 0.38, 95% CI = 0.22-0.65, p = 5.2e-04). Additionally, FS cases had a longer medical record (median = 10.29, Q1-Q3 = 6.26 – 14.68) than controls (median = 8.81, Q1-Q3 = 5.56 – 12.99, p=3.44e-04), more ICD codes accrued (FS cases median = 80, Controls = 49, p = 3.37e-27), and a greater density of record than controls (FS median = 9.07, Control median = 6.03, p = 0.002). FS cases were significantly younger than controls (median FS age = 49.31, median control age = 59.58, p=3.4e-104) and had a significantly higher BMI (FS median = 28.7, control median = 27.8, p=1.27e-07).

FS comorbidity with psychiatric and neurological disorder codes

Phenotypes most significantly associated with FS when compared to algorithm-identified controls included psychogenic and somatoform disorders (OR 1.37, 95% CI 1.35 – 1.38, p<3.02e-05; somatoform disorder (OR 1.43, 95% CI 1.41 – 1.45, p<3.02e-05); PTSD (OR 1.22, 95% CI 1.21 – 1.24, p<3.02e-05); mood disorders (OR 1.14, 95% CI 1.13 – 1.15, p<3.02e-05), and anxiety disorders (OR 1.14, 95% CI 1.13 – 1.15, p<3.02e-05; Figure 4A). We also identified strong associations with depression (OR 1.14, 95% CI 1.13 – 1.15, p<3.02e-05), and schizophrenia (OR 1.19, 95% CI 1.17– 1.20, p<3.02e-05). Overall, 56/72 (77%) phecodes mapping to the psychiatric disorders category were significantly associated with FS after Bonferroni correction for multiple testing (p = 3.02E-05). We found that 55/82 (67%) of the 'neurological disorder' phecodes were also significantly associated with FS.

A


Alteration of consciousness
Somatoform disorder
Transient alteration of awareness
Psychogenic and somatoform disorders
Neurological disorders
Other specified nonpsychotic and/or transient mental disorde
Syncope and collapse
Posttraumatic stress disorder
Abnormal involuntary movements
Altered mental status
Mood disorders
Abnormal movement
Anxiety disorders
Depression
Bipolar
Major depressive disorder
Schizophrenia and other psychotic disorders
Migraine
Anxiety disorder
Convulsions
−log10(p)
infectious diseases
neoplasms
endocrine/metabolic
hematopoietic
psychiatric disorders
neurological
sense organs
circulatory system
respiratory
digestive
genitourinary
pregnancy complications
dermatologic
musculoskeletal
congenital anomalies
symptoms
injuries & poisonings
Phenotypes

B

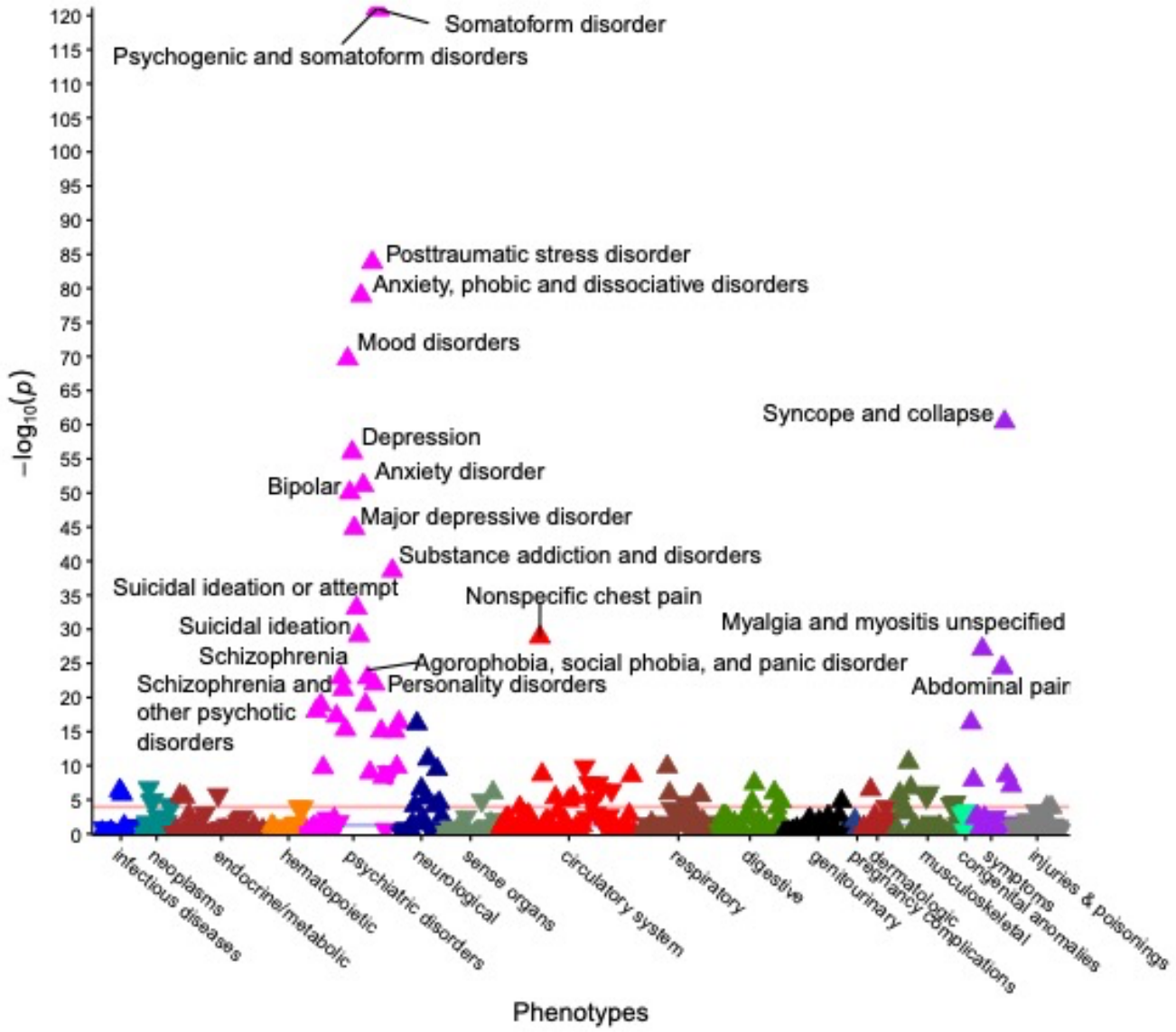


Figure 4. PheWAS of A) VUMC EHR FS algorithm defined cases versus algorithm defined controls and B) VUMC EHR FS algorithm defined cases versus epilepsy controls. Results are plotted by category of phenotypes, with each category shown in a different color. Only the top 20 associations are labeled to increase visibility of the graph. Additionally, upturned triangles represent positive associations with FS case status, while downturned triangles represent negative associations with FS case status. The horizontal red line indicates the Bonferroni corrected p-value.

Table 5 includes the period prevalence of sexual-assault trauma and ten common psychiatric comorbidities among algorithm-defined FS cases. These were ten predominantly adult-afflicting psychiatric comorbidities for which ICD code definitions had previously been validated.[53] The prevalence of most of these psychiatric comorbidities in the VUMC-EHR confirmed previously reported literature.[54–61] However, prevalence of these psychiatric comorbidities was much higher in FS patients ranging from 2.59% (phobia) to 30.19% (MDD).

Phenotype	N Cases Total (N comorbid with FS)	VUMC-EHR MH Prevalence (%)	Prevalence within FS patients (%)	Ratio of prevalence within FS:VUMC-EHR MH
Schizophrenia	10,322 (176)	1.97	12.30	6.24
Major depression	36,069 (432)	6.89	30.19	4.38
Bipolar disorder	14,789 (254)	2.82	17.75	6.28
Insomnia	32,493 (195)	6.21	13.63	2.20
Posttraumatic stress disorder	8,088 (339)	1.54	23.69	15.34
Obsessive-compulsive disorder	2,086 (39)	0.40	2.73	6.84
Alcohol abuse	13,630 (105)	2.60	7.34	2.82
Alcohol use - long term	5,164 (46)	0.99	3.21	3.26
Phobia	2,550 (37)	0.49	2.59	5.31
Anxiety	38,898 (391)	7.43	27.32	3.68
Sexual Assault Trauma	4,357 (188)	0.83	13.14	15.79
Functional seizures	1,431	0.27	..	..

Table 5. Psychiatric phenotypes and sexual assault trauma comorbidities in FS cases. The number of adult cases for each psychiatric comorbidity within the VUMC-EHR medical home is reported as well as the number of cases with FS comorbidity. We have also calculated the prevalence of each disorder in the VUMC-EHR medical home adult population and the FS adult population. Finally, we have calculated the ratio of the prevalence of each disorder within FS patients vs. in the entire VUMC-EHR. The total number of adult patients in the SD medical home was 523,584.

Psychiatric, neurological, and cerebrovascular disorders are associated with epilepsy

We also performed a PheWAS analysis between epilepsy cases and epilepsy controls, which showed 47 out of 71 psychiatric disorders codes were significantly associated with epilepsy, as well as 60 out of 82 neurological phecodes (Figure 5). We also observed that 35 circulatory system phenotypes were significantly associated with epilepsy, several of which were among the strongest associations with epilepsy.

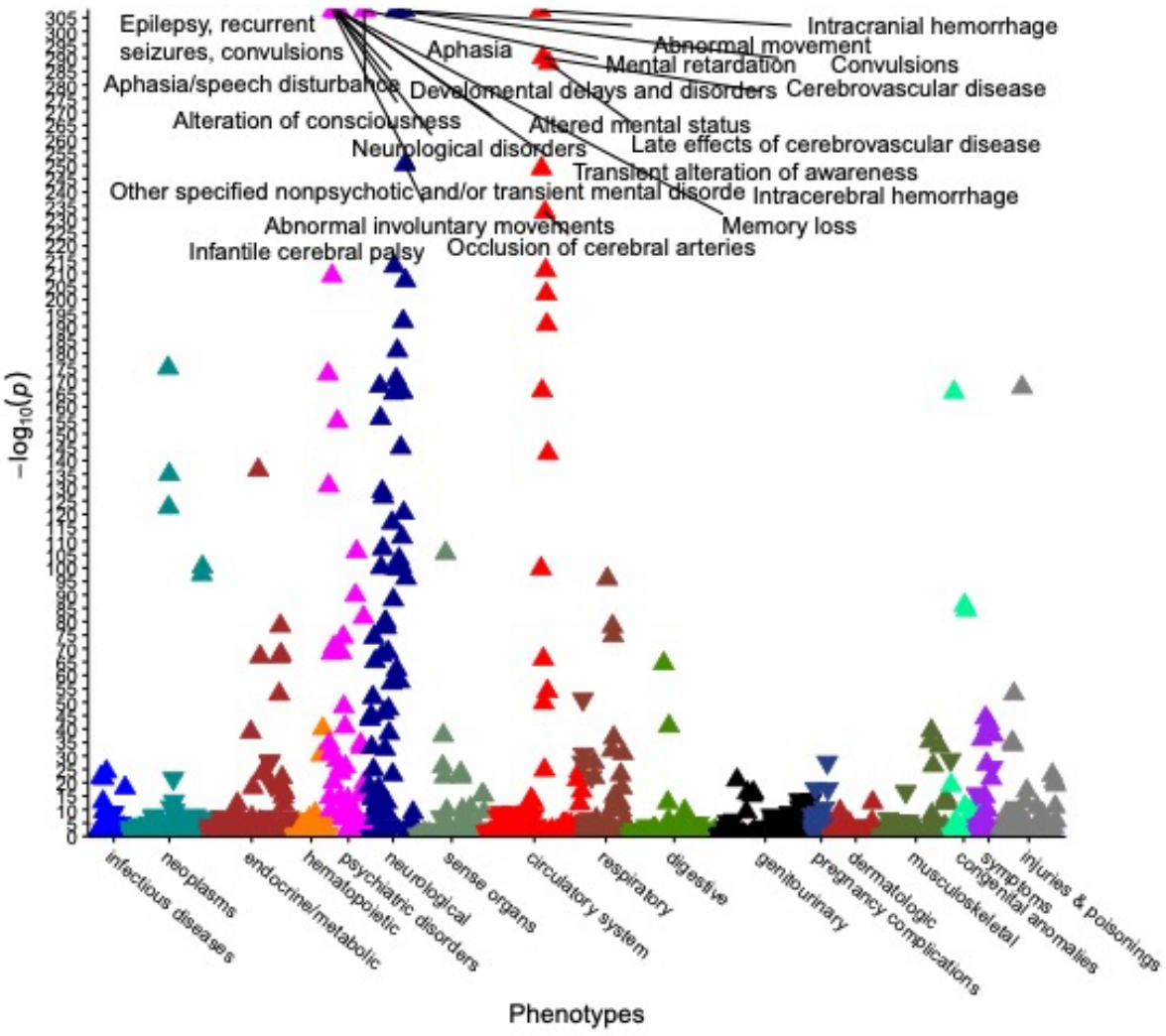


Figure 5. PheWAS of VUMC EHR ICD-code defined epilepsy cases versus controls. Results are plotted by category of phenotypes, with each category shown in a different color. Only the top 20 associations are labeled to increase visibility of the graph. Additionally, upturned triangles represent positive associations with FS case status, while

downturned triangles represent negative associations with FS case status. The horizontal red line indicates the Bonferroni corrected p-value.

Additionally, PheWAS comparing comorbidity patterns between FS cases and epilepsy cases showed that 72% of the psychiatric, 42% of the neurological and 20% of the circulatory system phenotypes that were associated with FS, remained associated with FS even when the comparison group was limited to epilepsy cases (Figure 4B).

FS co-occurrence with cerebrovascular disease

We identified thirteen novel and significant associations between cerebrovascular disease codes and FS. These included cerebrovascular disease (OR 1.08, 95% CI 1.06 – 1.09, p<3.02e-05) and transient cerebral ischemia (OR 1.09, 95% CI 1.08 – 1.11, p<3.02e-05; Table 6). PheWAS analyses comparing FS cases to epilepsy cases (Figure 4B) and epilepsy cases to controls (Figure 5) showed that cerebrovascular disease is associated with epileptic and non-epileptic seizures. CVD (phecode 433) was more strongly associated with epilepsy than FS (FS OR = 1.08, CI =1.06-1.09, p=2.6E-40; Epilepsy OR = 1.20, CI = 1.18 – 1.21, p<3.02e-05).

Phecode	Phecode Description	OR	p	N Cases	N Controls	95% CI
433	Cerebrovascular disease	1.08	2.57E-40	19161	374563	(1.06 - 1.09)
433.31	Transient cerebral ischemia	1.09	2.80E-35	7475	374563	(1.08 - 1.11)
433.3	Cerebral ischemia	1.09	3.24E-34	7886	374563	(1.08 - 1.11)
433.8	Late effects of cerebrovascular disease	1.09	2.56E-18	3510	374563	(1.07 - 1.11)
433.6	Acute, but ill-defined cerebrovascular disease	1.10	2.60E-17	2350	374563	(1.08 - 1.13)
433.2	Occlusion of cerebral arteries	1.07	6.77E-14	6680	374563	(1.05 - 1.09)
433.21	Cerebral artery occlusion, with cerebral infarction	1.07	1.72E-13	6476	374563	(1.05 - 1.09)
430	Intracranial hemorrhage	1.08	1.37E-11	3088	374563	(1.05 - 1.10)
433.5	Cerebral aneurysm	1.08	2.22E-08	1618	374563	(1.05 - 1.11)
430.1	Subarachnoid hemorrhage	1.08	5.68E-08	1187	374563	(1.05 - 1.11)
430.2	Intracerebral hemorrhage	1.07	6.20E-06	1525	374563	(1.04 - 1.10)
430.3	Subdural hemorrhage	1.09	1.80E-05	782	374563	(1.05 - 1.14)
433.1	Occlusion and stenosis of precerebral arteries	1.05	2.86E-05	6741	374563	(1.03 - 1.08)

Table 6. FS-associated cerebrovascular disease phenotypes. Cerebrovascular disease phenotypes in the circulatory system category that were significantly associated with FS case/control status, as well as their associated OR and the number of cases for each phenotype.

Twenty-nine percent of patients were diagnosed with FS prior to the onset of CVD, 23% were diagnosed with CVD and FS within 90 days of each other, and almost half (48%) were diagnosed with FS after CVD (Figure 6, Figure 7, Table 7). However, these results are difficult to disentangle given the complication of the long diagnostic odyssey of FS. Essentially, the date of diagnosis for FS has shown to be on average 8 years after the typical onset of seizures in this patient population. This implies a certain level of uncertainty in the exact order of the CVD vs FS events even in the face of this analysis. A parallel analysis of epilepsy and CVD revealed similar patterns (Table 8).

	CVD and FS codes (n = 92)	FS before CVD (n = 27)	CVD before FS (n = 44)	FS and CVD simultaneous (n = 21)
Age at first cerebrovascular code, Median (Q1 – Q3)	49.78 (39.03 – 55.51)	45.27 (37.29 – 54.79)	50.09 (41.83 – 55.68)	51.27 (38.87 – 58.58)
Age at FS suspicion, Median (Q1 – Q3)	49.03 (38.10 – 57.30)	38.80 (32.95 – 49.10)	53.79 (43.46 – 60.66)	51.30 (38.68 – 58.34)
Age at FS diagnosis, Median (Q1 – Q3)	49.71 (38.41 – 57.82)	38.80 (32.95 – 49.14)	53.79 (45.08 - 61.35)	51.30 (38.68 – 58.58)
Years from cerebrovascular code to FS diagnosis, Median (Q1—Q3)	0.19 (-0.85 - 3.48)	-3.98 (-6.86 to -1.54)	3.89 (1.40 - 5.77)	0.01 (0.00 - 0.07)
PTSD comorbidity, N (%)	20 (22)	6 (22)	10 (23)	4 (19)
Sexual assault trauma comorbidity, N (%)	14 (15)	6 (22)	5 (11)	3 (14)
Migraine comorbidity, N (%)	27 (29)	7 (26)	14 (32)	6 (29)
Schizophrenia comorbidity, N (%)	3 (3)	1 (4)	1 (2)	1 (5)
MDD comorbidity, N (%)	20 (22)	5 (19)	11 (25)	4 (19)
Bipolar disorder comorbidity, N (%)	10 (11)	3 (11)	6 (14)	1 (5)
Insomnia comorbidity, N (%)	9 (10)	2 (7)	4 (9)	3 (14)
PTSD comorbidity, N (%)	0 (0)	0 (0)	0 (0)	0 (0)
Anxiety comorbidity, N (%)	29 (32)	6 (22)	17 (39)	6 (29)
Phobia comorbidity, N (%)	14 (15)	5 (19)	6 (14)	3 (14)
Alcoholism comorbidity, N (%)	5 (5)	2 (7)	2 (5)	1 (5)
OCD comorbidity, N (%)	5 (5)	2 (7)	2 (5)	1 (5)
Female sex, N (%)	64 (70)	17 (63)	33 (75)	14 (67)
Race, N (%)				
Caucasian	81 (88)	23 (85)	38 (86)	20 (95)
African American	11 (12)	4 (15)	6 (14)	1 (5)
Ethnicity, N (%)				
Hispanic/Latino(a)	2 (2)	1 (4)	0 (0)	1 (5)
Median BMI, Median (Q1 – Q3)	28.91 (24.72 – 35.61)	28.27 (24.13 – 32.03)	28.31 (24.03 – 36.12)	32.86 (26.57 – 37.34)
Records per year, Median (Q1 – Q3)	19.26 (8.61 – 61.02)	23.54 (6.79 – 48.33)	26.46 (9.65 – 68.30)	11.52 (7.85 – 34.07)
Length of record, Median (Q1 – Q3)	11.09 (7.38 - 15.75)	12.00 (8.80 - 15.38)	11.11 (7.80 - 15.99)	10.62 (7.13 - 11.85)
Number of ICD codes, Median (Q1 – Q3)	241.50 (68.25 – 648)	212 (67 – 607)	409.5 (102 – 819.5)	100 (56 – 399)
Age, Median (Q1 – Q3)	59.14 (48.94 – 65.91)	52.35 (42.72 – 60.94)	62.85 (53.47 – 68.04)	60.99 (53.35 – 65.84)
Median Age of Record, Median (Q1 – Q3)	51.50 (40 – 59)	46 (37.25 – 55)	54.50 (43.75 – 60.25)	51 (42 – 58)

Cerebrovascular disease subcategory ICD codes (% of total CVD ICD codes)				
Cerebral hemorrhage	12	17	9	10
Cerebral artery occlusion	19	23	18	16
Cerebral ischemia	29	23	33	30
Cerebrovascular disease	35	30	37	41
Cerebral aneurysm	4	5	4	1
Cerebral atherosclerosis	1	2	0	1

Table 7. Demographics for 92 patients who were diagnosed with both cerebrovascular disease and FS. Demographics are also shown for three subgroups of these patients: those who developed FS before CVD (n = 27); those who developed FS and CVD within three months of each other (n = 21); and those who developed CVD before FS (n = 44). The proportion of males and females, race, ethnicity, density of records, median age across the medical record, age at first CVD code, age at FS suspicion, age at FS diagnosis, and percentages of CVD subcategories represented in each group. ICD codes for cerebral hemorrhage included 430, 430.2, 430.3, and 430.1. ICD codes for cerebral artery occlusion included 433.2, 433.21, 433.1, and 433.11. ICD codes for cerebral ischemia included 433.3 and 433.31. ICD codes for cerebrovascular diseases included 433, 433.8, and 433.6. The ICD code for cerebral aneurysm was 433.5, and the ICD code for cerebral atherosclerosis was 433.12.

	CVD and Epi codes (n = 735)	**Epi before CVD (n = 208)**	**CVD before Epi (n = 374)**	**Epi and CVD simultaneous (n = 153)**
Age at first cerebrovascular code, Median (Q1 – Q3)	55.27 (42.72 – 65.18)	52.31 (40.03 – 61.81)	54.31 (43.62 – 65.07)	60.96 (48.26 – 70.49)
Age at first epilepsy code, Median (Q1 – Q3)	55.74 (42.07 - 66.16)	46.07 (33.68 – 57.73)	58.70 (45.91 – 68.55)	61.02 (48.30 – 70.49)
Days from epilepsy code to cerebrovascular, Median (Q1—Q3)	-106 (-1006.5 to 289)	1195 (607.5 – 2739.8)	-990 (-2089.8 to -350.2)	-1 (-28 to 0)
Sexual assault trauma, N (%)	10 (1)	3 (1)	4 (1)	3 (2)
PTSD, N (%)	18 (2)	9 (4)	7 (2)	2 (1)
Female sex, N (%)	357 (49)	104 (50)	183 (49)	70 (46)
Race, N (%)				
Caucasian	599 (81)	180 (87)	293 (78)	126 (82)

African American	127 (17)	25 (12)	77 (21)	25 (16)
Asian	6 (1)	2 (1)	3 (1)	1 (1)
Unknown/Other	1 (0)	0 (0)	1 (0)	0 (0)
Ethnicity, N (%)				
Hispanic/Latino(a)	14 (2)	4 (2)	7 (2)	3 (2)
Median BMI, Median (Q1 – Q3)	27.45 (23.69 – 31.84)	27.14 (23.50 – 31.51)	27.45 (24.03 – 32)	27.75 (23.14 – 31.16)
Records per year, Median (Q1 – Q3)	29.15 (14.40 – 56.46)	24.07 (11.25 – 44.43)	35.03 (17.22 – 64.94)	26.28 (13.67 – 53.67)
Length of record, Median (Q1 – Q3)	12.56 (8.05 – 16.79)	14.06 (9.50 – 17.45)	13.16 (8.48 – 16.84)	9.21 (6.12 – 14.12)
Number of ICD codes, Median (Q1 – Q3)	339 (158 – 676)	288 (142 – 556.8)	413.5 (181.2 – 791.2)	250 (148 – 535)
Age, Median (Q1 – Q3)	66.4 (54.4 – 75.72)	61.32 (48.52 – 71.63)	67.60 (56.01 – 78.25)	71.14 (59.35 – 77.81)
Median Age of Record, Median (Q1 – Q3)	58 (46 – 68)	52.50 (40.00 – 62.00)	59 (48 – 70)	63 (50 – 70)
Cardiovascular subcategory ICD codes (% of total CVD ICD codes)				
Cerebral hemorrhage	20	19	18	27
Cerebral artery occlusion	25	25	26	22
Cerebral ischemia	20	20	21	18
Cerebrovascular disease	33	33	32	32
Cerebral aneurysm	2	3	2	2
Cerebral atherosclerosis	1	1	1	1

Table 8. Demographics for 735 patients who were diagnosed with both cerebrovascular disease and epilepsy. Demographics are also shown for three subgroups of these patients: those who developed Epilepsy before CVD (n = 208); those who developed Epilepsy and CVD within three months of each other (n = 153); and those who developed CVD before Epilepsy (n = 374). The proportion of males and females, race, ethnicity, density of records, median ages across the medical record, age at first CVD code, age at first generalized or focal epilepsy code, and percentages of CVD subcategories are shown for each group. ICD codes for cerebral hemorrhage included 430, 430.2, 430.3, and 430.1. ICD codes for cerebral artery occlusion included 433.2, 433.21, 433.1, and 433.11. ICD codes for cerebral ischemia included 433.3 and 433.31. ICD codes for cerebrovascular diseases included 433,

433.8, and 433.6. The ICD code for cerebral aneurysm was 433.5, and the ICD code for cerebral atherosclerosis was 433.12.

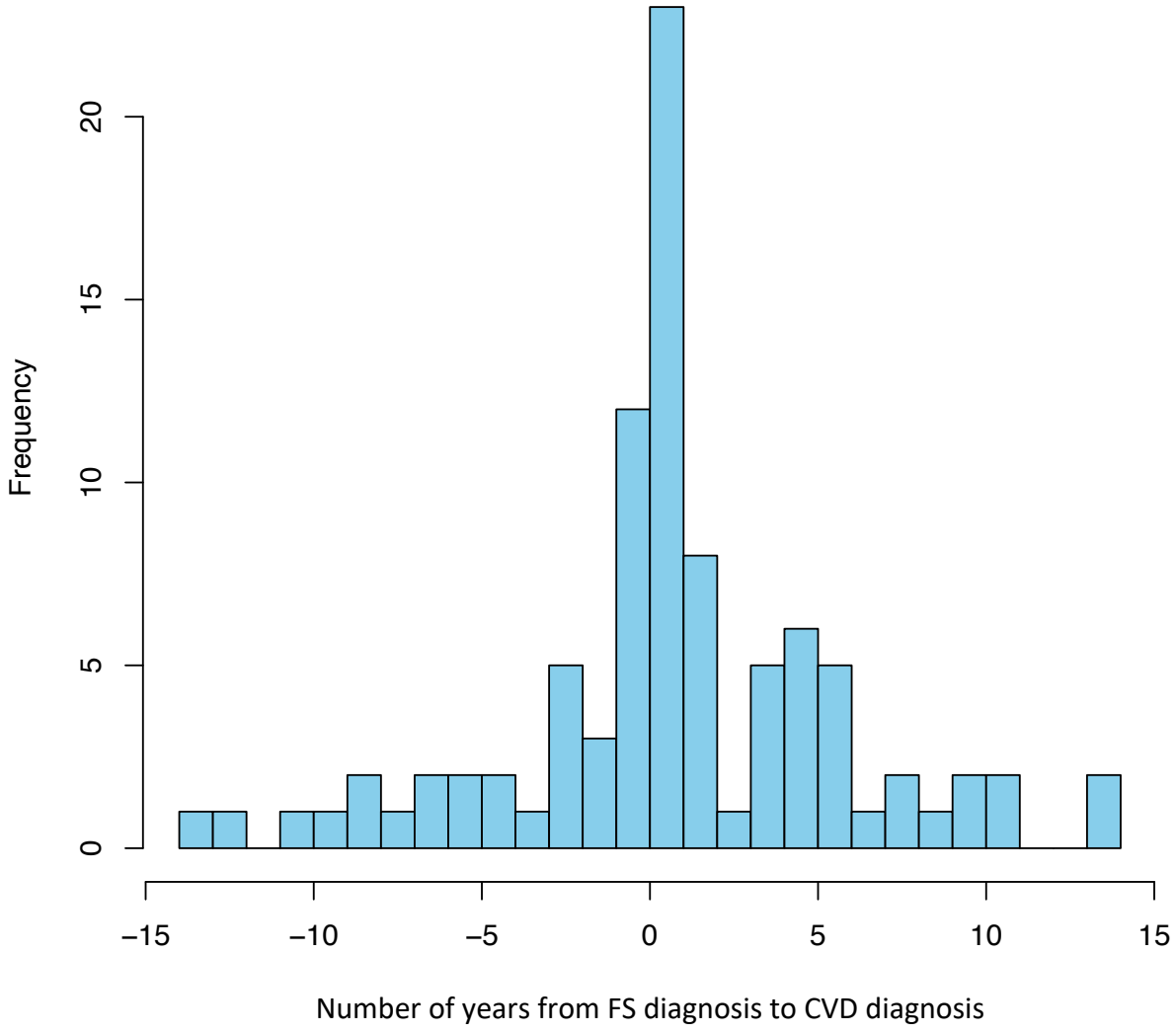


Figure 6. Histogram of the number of years from date of PNES diagnosis to cerebrovascular diagnosis in 92 patients who had both CVD and PNES. Negative values indicate patients who were diagnosed with CVD before PNES, while positive values indicate patients who were diagnosed with PNES before CVD.

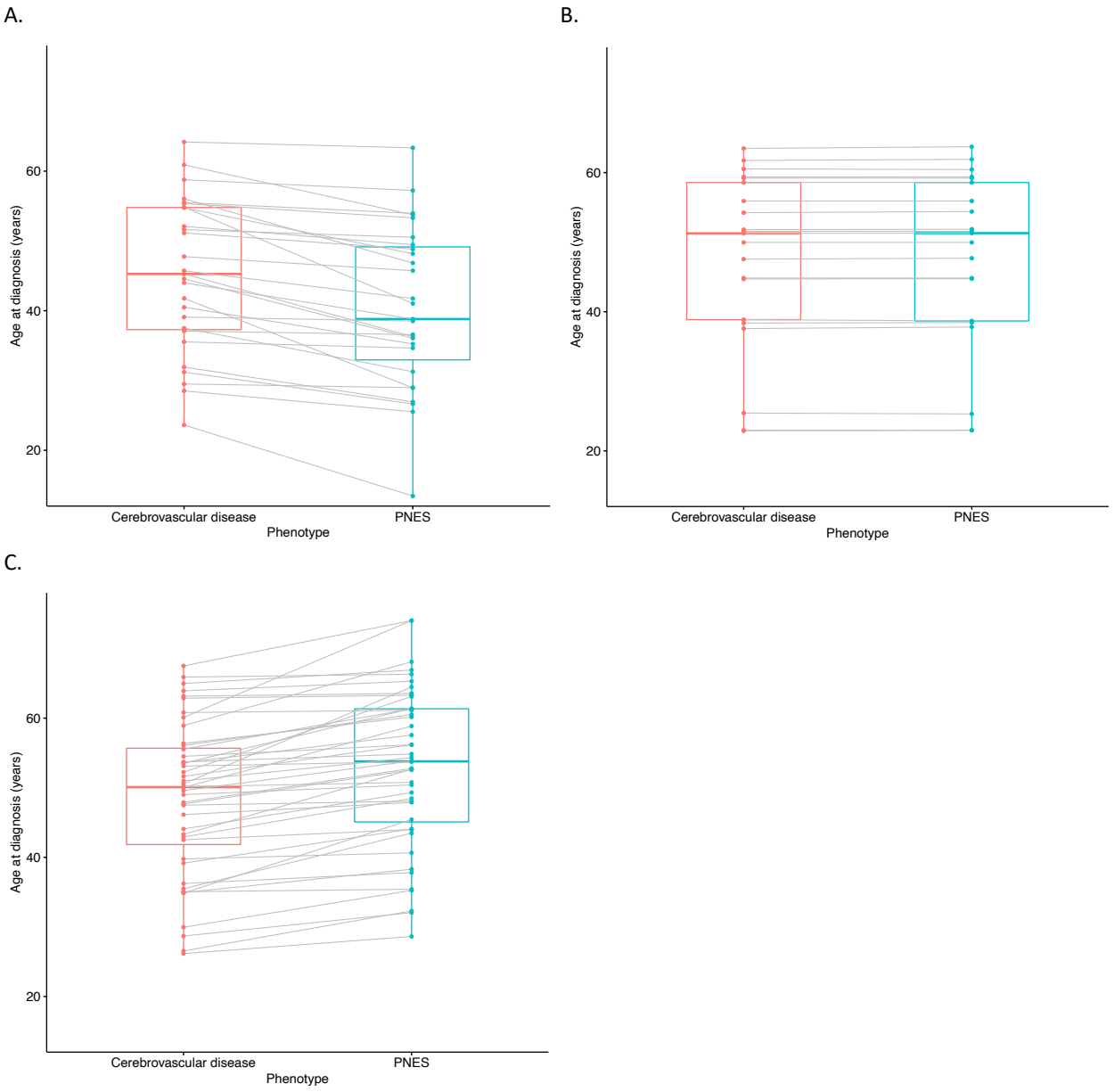


Figure 7. Temporal analysis of the development of FS and cerebrovascular disease in 92 patients. Patients are binned into three groups: A. Patients who developed FS before CVD (n = 27); B. Patients who developed FS and CVD within 3 months of each other (n = 21); and C. Patients who developed CVD before FS (n = 44). The age at the first CVD ICD code each patient was diagnosed with is plotted as red dots, with median and first and third quartiles plotted as a box plot. The corresponding age at FS diagnosis is plotted as blue dots, with a grey line connecting the age of diagnosis of CVD and FS for each patient.

FS association with sexual assault trauma

We identified a total of 10,036 individuals (0.36%) in the EHR who met inclusion criteria for sexual assault trauma including 1,853 males and 8,183 females (Table 3). Among these patients, 4,357 also met criteria for the medical home population and were over the age of 18. Approximately four percent (188/4,357) of patients reporting sexual assault met criteria for FS. Conversely, the prevalence of sexual assault trauma among patients with FS was 13.14%, 15.79 times more prevalent than in the general medical home population. The prevalence of sexual assault trauma among patients with epilepsy was 132 out of 4,715 (2.80%).

Using a multi-variable logistic regression, we found that sexual assault trauma is significantly associated with FS (OR = 10.26, CI = 10.09 – 10.44, p<3.02e-05) (Table 9) after adjusting for age, sex, median BMI, race, and medical record density. While our analysis showed that female sex was also significantly associated with FS (Male OR = 0.64, CI = 0.51 – 0.77, p<3.02e-05), at this time there was no significant interaction between sex and sexual assault trauma on FS diagnosis (OR = 1.30, CI = 0.80 – 1.81, p = 0.31).

Variable	P-value	OR	CI
Sexual Assault Trauma	5.36E-146	10.26	10.09-10.44
Sex (Male)	4.04E-11	0.64	0.51-0.77
Sexual Assault Trauma:Sex	0.31	1.30	0.80-1.81

Table 9. FS case status and sex are significantly associated with sexual assault in the VUMC-EHR. A multivariable logistic regression FS ~ Sexual assault trauma + sex + sexual assault trauma*sex was performed to determine whether FS case/control status, sex, or the interaction of FS case status were correlated with sexual assault case status. P-values, odds ratio (OR) and the 95% confidence interval (CI) are presented for each variable. The

multivariable logistic regression used included these variables and additional covariates for median age of record, median BMI, race, and the density of records.

Mediation of sexual assault trauma and the increased rate of females with FS

Sexual assault trauma mediated 22% ($p<3.02e-05$) of the variance in FS diagnosis associated with female sex.

Comorbidity patterns for FS do not differ by sex

There were no significant differences between males and females in the prevalence of each FS comorbidity. Based on the PheWAS conducted in FS patients with a sex interaction term added to the logistic regression, no phenotypes exceeded the significance threshold for multiple testing.

BMI is significantly higher in FS patients than controls

One-way ANOVA and a post-hoc Tukey test showed that FS cases without concurrent epilepsy had significantly higher median BMI measurement than algorithm-defined or epilepsy patients (Figure 8). Additionally, a greater proportion of algorithm-defined FS cases without concurrent epilepsy were considered obese (BMI>30) (39%) compared to algorithm-defined controls (33%), epileptic patients (27%), and FS patients with concurrent epilepsy (37%)(Table 10).

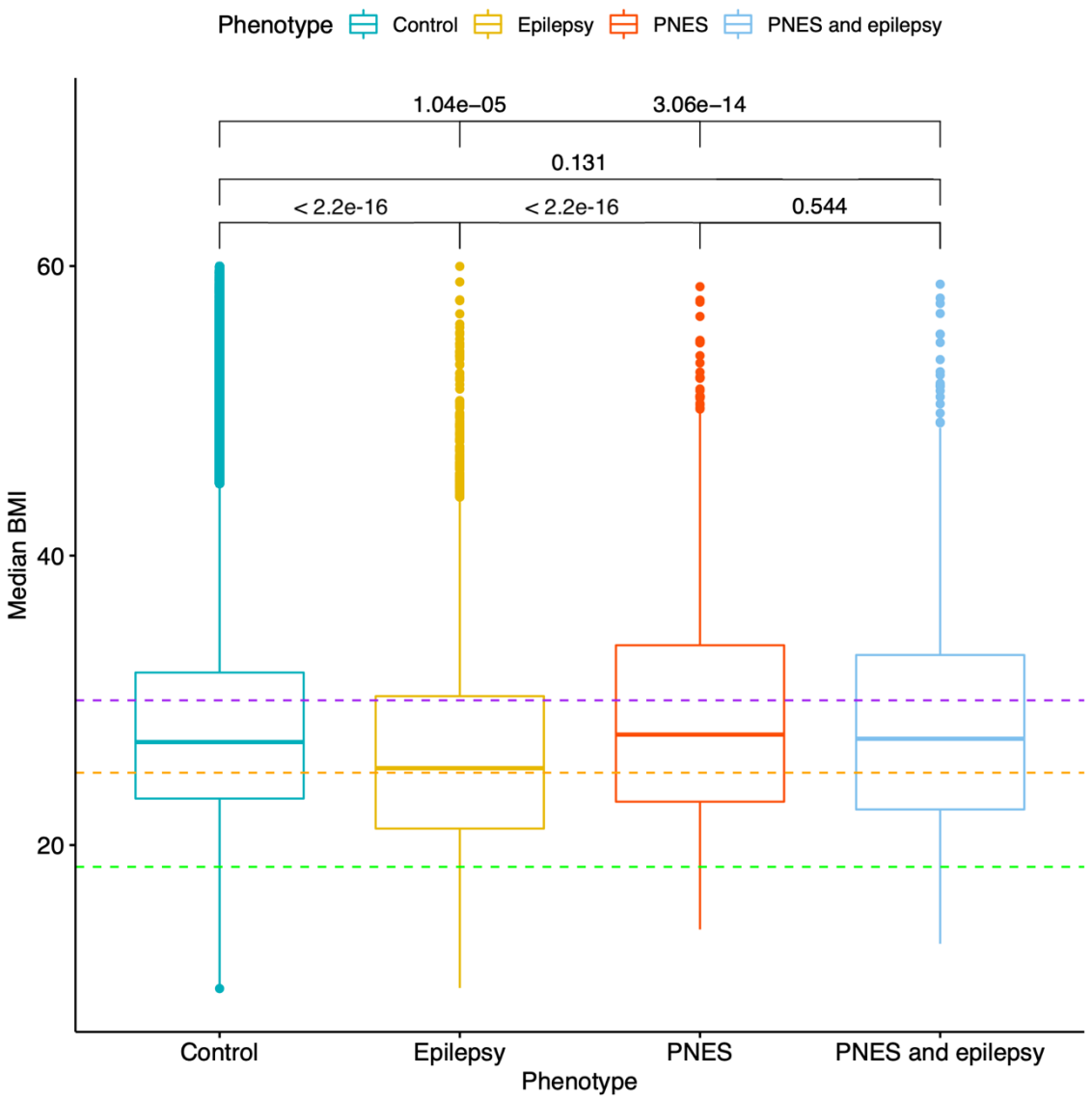


Figure 8. FS patients have greater BMIs than the control group and epilepsy patients. Average and SE of median BMI are shown for controls as identified by our algorithm, epilepsy patients as identified by ICD codes, FS cases as identified by our algorithm, and FS cases with epilepsy as identified by our second algorithm. All participants analyzed had a median age of records over 18. The sample size and percentage of patients from each group that were considered obese (BMI over 30) is also shown.

Phenotype	Sample size	Mean body mass index (SE)	Obese (% BMI over 30)
Control	488,398	28.17 (0.01)	33.46
Epilepsy	6,186	26.29 (0.09)	26.59
FS only	1,605	29.01 (0.20)	39.31
FS and epilepsy	1,058	28.64 (0.24)	37.43

Table 10. Body Mass Index information for FS and epilepsy cases as compared to controls. Average and SE of median BMI are shown for controls as identified by our algorithm, epilepsy patients as identified by ICD codes, FS cases as identified by our algorithm, and FS cases with epilepsy as identified by our second algorithm. All participants analyzed had a median age of records over 18. The sample size and percentage of patients from each group that were considered obese (BMI over 30) is also shown.

Discussion

This study aimed to examine clinical and epidemiological characteristics of FS. First, we identified FS patients in the VUMC-EHR by developing an automated phenotyping algorithm with a PPV of 98%. The complete algorithm (Figure 2, Table 1) for identification of FS cases is provided in supplementary materials and deposited in PheKB for future research use by others. Chart review of these patients also revealed that the average time from first seizure to FS diagnosis in these patients was 6.6 years (± 1.4). Based on the number of patients identified by our algorithm in proportion to the total number of patients in our hospital system, we calculated the period prevalence of FS to be 0.14% in our clinical population. When restricting to the medical home population, the period prevalence was 0.27%. In analyzing comorbidities among FS patients, we confirmed previous reports that FS co-occurs with psychiatric and

neurological disorders, and that FS patients are nearly sixteen times more likely than the average hospital patient to have a documented history of sexual assault trauma. Moreover, we found that sexual assault trauma explains 22% of the increased rate of FS in females. Finally, we also discovered that FS co-occurs with cerebrovascular disease at a rate higher than expected by chance.

FS was previously estimated to occur in approximately 2-33 people per 100,000 (0.002 – 0.033%).[30] The directly calculated period prevalence of FS in the VUMC clinical population was, as expected, much higher at 142 per 100,000 (0.14%). VUMC is home to an epilepsy monitoring unit (EMU), which increases the number of FS patients relative to the general population. Thus, the prevalence in the VUMC EHR may not be generalizable. However, it is critical to understand the prevalence of FS in a medical center setting for multiple reasons. First, efforts aimed at improving clinical care for FS patients are bolstered by awareness of the frequency of FS in a clinical setting. Second, the prevalence in a clinical setting also provides motivation for development of an ICD diagnostic classification specific to FS. The visibility issues that patients with FS face was further substantiated by our chart review which indicated that the average time from first seizure to FS diagnosis was 6.6 years, closely matching older reports.[1]

Several demographic comparisons between FS cases and controls closely matched previously reported literature. We confirmed the association between FS and female sex status (OR = 1.75, 95% CI = 1.55-1.99, p = 1.6e-18).[5,62] We also confirmed that FS patients had a significantly higher BMI (FS median = 28.7, control median = 27.8, p=1.27e-07) than controls.[49] Other demographic analyses revealed currently unreported associations to our knowledge. Our study demonstrated that FS cases were more likely to be Caucasian (OR = 1.97, 95% CI = 1.11-

3.47, p = 0.02), and less likely to be Hispanic (OR = 0.38, 95% CI = 0.22-0.65, p = 5.2e-04). Additionally, FS cases had a longer medical record (median = 10.29, Q1-Q3 = 6.26 – 14.68) than controls (median = 8.81, Q1-Q3 = 5.56 – 12.99, p=3.44e-04), more ICD codes accrued (FS cases median = 80, Controls = 49, p = 3.37e-27), and a greater density of record than controls (FS median = 9.07, Control median = 6.03, p = 0.002). FS cases were significantly younger than controls (median FS age = 49.31, median control age = 59.58, p=3.4e-104). This indicates that while FS patients in our medical system are around a decade younger than the average patient at VUMC, they still accrued more ICD codes and had a longer record than controls by around a decade. This indicates generally worsened health outcomes for FS patients at an earlier age than controls.

Our PheWAS results suggest that patients with FS are at risk for additional chronic health conditions including cerebrovascular disease. While associations between CVD and epilepsy are widely reported, no robust associations between FS and CVD have been reported to our knowledge.[63–67] Diagnoses of cerebral ischemia, occlusion of cerebral arteries, and intracranial hemorrhage were all significantly associated with FS in our data (Table 6). These results are consistent with a previous report detailing FS and co-morbid chronic medical conditions.[68] However, we observed no clear illness trajectory from FS to CVD, and in fact found that CVD often preceded the onset of FS (Figure 6, Figure 7, Table 7). Moreover, we found no difference in the rate of FS risk factors or comorbidities between those patients who experienced FS first compared to those who experienced CVD first. The finding that CVD may precede FS could be explained by brain trauma or psychological distress related to CVD, paralleling prior data demonstrating that stroke is a risk factor for later-onset epilepsy. These

findings have important implications for the management of patients who develop post-stroke seizures. Specifically, diagnostic video-EEG evaluation for post-stroke seizures is critical to confirm the diagnosis of epilepsy and/or FS.

Consistent with previous reports, we found FS are associated with multiple psychiatric and neurological disorders compared to the general hospital population and to epilepsy patients (Figure 4B; Figure 5). These multiple illnesses may be causally linked to FS etiology or pathophysiology. Alternatively, multiple diagnoses may accumulate for a FS patient while undergoing clinical treatment and care since FS has a broad differential diagnosis. Therefore, further research is needed to clarify if any causal relationship exists between FS and the numerous comorbidities identified in this study. Given that FS patients on average are diagnosed with 4.75 different psychiatric diagnoses (compared to 2.6 in epilepsy patients and 0.85 in hospital controls), we suggest a clinical guideline: that patients experiencing seizures with a high burden of psychiatric illness be considered for a FS diagnosis and referred for diagnostic video-EEG monitoring. This is especially important as early FS diagnosis and treatment are associated with better outcome.[69–71] Overall, we believe that this novel EHR-based study provides important rationale and motivation for ongoing EHR-based research to improve the complex and challenging clinical care of patients with FS.

FS has a much higher prevalence in females. Consistent with previous reports, approximately 74% of FS cases in our cohort were female.[10,29,62,72,73] Previous studies also indicate that females with FS were eight times more likely to report sexual assault trauma than males with FS.[10] We found that sexual assault trauma and PTSD were approximately 16 times more frequent in FS patients compared to the general hospital population (Table 5). Among FS

patients, 15.82% of females and 5.42% of males reported a history of sexual assault trauma. Given sexual assault trauma exposure, there was no significant difference between males and females in the rate of FS diagnosis indicating that males and females who experience sexual assault trauma are at equivalent risk of developing FS (Table 9). While females with FS reported more sexual assault trauma than males with FS, this reflected the increased overall rate of sexual assault among females (1.08%) compared to males (0.25%). Mediation analysis results indicated that the overall increased rate of sexual assault trauma among females explains nearly a quarter of the increased rate of FS among females. Taken together, these findings provide evidence for the hypothesis that FS, while influenced by multiple complex factors exhibiting interindividual differences, may be considered a physical manifestation of the neurological damage caused by trauma.

CHAPTER III

GENETIC EPIDEMIOLOGY OF FUNCTIONAL SEIZURES

Introduction

The population prevalence of functional seizures was previously estimated to be 0.002%-0.033%[30] and in Chapter II, we calculated the period prevalence of FS in the Vanderbilt University Medical Center Electronic Health Record (VUMC-EHR) system to be 0.014%.[8] Approximately 20-30% of patients admitted to epilepsy monitoring units (EMUs) are eventually diagnosed with FS,[25] highlighting the benefit of including academic medical center EMUs as recruitment sites for research on FS. Motivated by this rationale, in Chapter II, we developed an algorithm to identify FS patients which yielded a positive predictive value of 98%.[8]

Furthermore, most existing single site studies of FS studies are limited in sample size.[26] However, meta-analysis of EHRs with associated biobanks is a powerful way to quickly collect large numbers of cases and controls for clinical and genetic epidemiology studies.[42,45]

The first genetic study on FS, conducted by Drs. Costin Leu and Dennis Lal, who also collaborated on this project, used whole-genome sequencing to determine that six percent of their sample contained pathogenic or likely pathogenic rare variants in genes such as NSD1 and GABRA5.[35] This genetic burden was similar to that observed in individuals with generalized epilepsy (2%) and focal epilepsy (3%; p=0.3).[35] Previous studies also show that phenotypically related disorders, including generalized epilepsy (GE), focal epilepsy (FE), and PTSD are heritable, with SNP-based heritability most recently reported at 32%, 9%, and 5%,

respectively.[36,37] Thus, we hypothesized that FS are also modestly heritable with a polygenic architecture.

Using the validated VUMC-EHR algorithm, we conducted a multi-site genome-wide association study (GWAS) and meta-analysis of functional seizures. Cases and controls were identified from the EHRs of the Vanderbilt University Medical Center (BioVU), the Cleveland Clinic, Mount Sinai (BioMe), the Million Veteran Program (MVP), Massachusetts General BioBank (MGBB), and the Danish population registry (iPSYCH), resulting in a total of 9,289 FS cases and 417,818 controls of European ancestry.

Methods

Sample Collection

We included cases and controls from six individual biobanks in this meta-analysis of functional seizures. Further information regarding the algorithmic phenotyping (Table 11) and analysis method (Table 12) at each site is described below.

	ICD10 F44.5	**ICD9 300.11 or 780.39**	**ICD10 R56.9**	**EEG CPT Code**	**FS Keyword**	**Medical record review**	**Epilepsy ICD code Exclusion**
BioVU	X	X	X	X	X		X
CC						X	
iPSYCH	X						X
Mt. Sinai	X	X	X	X			X
MVP	X	X					X
MGBB	X	X	X	X			X

Table 11. FS phenotyping strategies across each biobank. Each row represents a biobank included in the FS GWAS meta-analysis, and each column represents a parameter used to select FS cases. An 'X' is used to demark which

phenotyping parameters each biobank used. Epilepsy ICD codes (ICD9 codes 345.1, 345.10, 345.4, 345.40, 345.5, 345.50 and ICD10 codes G40.20, G40.30, G40.00, and G40.10) were excluded.

BioBank	FS Cases	FS Controls	Total Sample
BioVU	309	59,856	60,165
CC	109	17,811	17,920
Mt. Sinai	133	9,744	9,877
MVP	7,111	446,799	453,910
MGBB	1,134	24,562	25,696
iPSYCH	232	2,660	2,892
Meta-analysis	**9,028**	**561,432**	**570,460**

Table 12. Number of FS cases and controls across each biobank. Each row represents a biobank included in the FS GWAS meta-analysis. The number of cases and controls in each individual biobank, the total number of individuals from each biobank, as well as the resulting total number of cases and controls used in the meta-analysis, is shown.

Post-GWAS, pre-meta-analysis Quality Control

After obtaining summary statistics from each individual site, we followed a field standard quality control protocol, EasyQC.[74] SNPs were filtered for MAF <0.05, missing data, non-A, C, T, or G alleles, data entry and analysis errors (e.g. negative betas and standard errors), and imputation scores r2<0.3 or info score <0.8. GWAS effect alleles (EA) were aligned with the 1000 Genomes reference population, and effect allele frequency was calculated (Figure 9).[75]

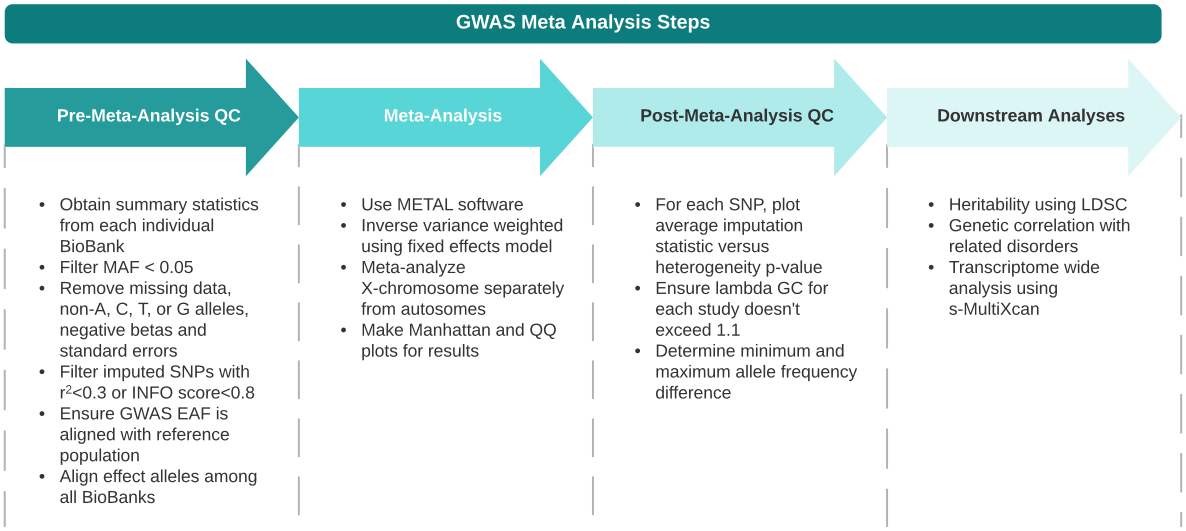


Figure 9. GWAS Meta-Analysis pipeline. Steps for post-GWAS, pre-meta-analysis QC, meta-analysis, post-meta-analysis QC, and post GWAS integrative functional analysis is described.

BioVU Genome-wide Association Study

Phenotyping approach

A subset of individuals in the Vanderbilt University Medical Center EHR also contribute blood left over after routine clinical testing for the purposes of biobanking (BioVU) and genetic analysis. A thorough description of BioVU has been previously published [34]. As of July 2021, BioVU contains 260,002 samples and 94,474 of these samples are genotyped on the Illumina Mega-ex Array, which covers nearly 2 million markers. An automated phenotyping algorithm was used to select FS cases and controls for GWAS analysis as described previously (Figures 2,3; Tables 1, 2, 11).[8] This resulted in the identification of 309 functional seizure cases and 59,856 controls with genotyping data available (Table 12).

Quality Control of Genetic Data

Genotype quality control included SNP precleaning using a SNP call rate of <0.95 using PLINK v1.9 (https://www.cog-genomics.org/plink/1.9/)[76,77]. In other words, individual SNPs were filtered temporarily for the purpose of individual-level filtering, and were then added back to the data. Individual SNPs with MAF <1% were temporarily filtered during the following individual-filter steps: Individuals were filtered who had a call rate <0.98. Next, Fhet values were calculated within EHR-reported races and individuals with |Fhet| >0.2 were filtered out. The individuals to be filtered from those three steps were removed and then a new MAF filter of 0.5% was applied.

SNP ids were then mapped to 1000Genomes reference panel and principal components were calculated using PLINK v1.9 and Eigenstrat (https://github.com/DReichLab/EIG)[78,79]. Next, ancestry cluster boundaries were defined to identify individuals of homogenous ancestry. Closely related individuals within each ancestry group were again filtered out by calculating IBD and filtering PI_HAT > 0.2 using PLINK v1.9. Hardy-Weinberg disequilibrium was calculated within each ancestry group using non-LD pruned data and SNPs with a Hardy-Weinberg p-value < 10^-10 were removed. SNPs with >0.1 MAF difference between study data and the 1KG population reference were filtered.[75]

Next, the genotype data was strand-aligned to the Haplotype Reference Consortium (HRC; http://www.haplotype-reference-consortium.org/)[80] reference panel using Will Rayner's pre-imputation script, HRC-1000G-check-bim-NoReadKey.pl (https://www.well.ox.ac.uk/~wrayner/tools/).[81] Data was then phased with ShapeIt.[82,83] The aligned samples were then submitted for imputation at the Michigan Imputation Server (MACH;

v1.0.4) in five batches using the cosmopolitan 1,000 genomes reference panel.[75,84–87] Imputation was run using Minimac3 with HRC r1.1as the reference panel and Eagle v2.3 for phasing with a "Mixed" population setting.

Following imputation, SNPs with $R^2 < 0.3$ were filtered. Non-biallelic SNPs were also filtered. Next, SNPs with a call rate < 0.98 or MAF < 0.05 were removed. This stringent MAF was applied due to the low number of cases in the sample for GWAS. Individuals with a call rate of < 0.98 were also filtered. Logistic regression analysis on genotype and imputation batch was performed using PLINK v1.9 and to find and filter out SNPs showing batch effects. SNPs demonstrating significant inflation in p-value, as visualized on the QQ plot, were removed.

SNPs with low imputation quality ($R^2<0.3$) and low minor allele frequency (MAF<0.05) were filtered post-GWAS, after which this genotyping data included 3,944,055 SNPs, 41,714 (1.06%) of which were not shared among genotyping data from any of the other 5 sites (Table 13).

Statistical Analysis

Variants associated with FS cases compared to controls were identified using the GWAS software SAIGE (Scalable and Accurate Implementation of GEneralized mixed model), which uses a generalized mixed-model and implements saddlepoint approximation to account for case-control imbalance.[88–91] In the GWAS analysis we covaried for the top ten principal components estimated from genetic data to further control for population stratification, sex, number of ICD codes per day, and median age of record (Table 14).

Cleveland Clinic (CC) GWAS

Phenotyping approach

FS cases for GWAS analysis were ascertained through the Epilepsy Biorepository of the Cleveland Clinic Epilepsy Center. Charts of these participants were then reviewed by an epilepsy biorepository research coordinator trained in clinical epilepsy phenotyping (Table 11). [35] All others were considered controls. This resulted in the identification of 109 functional seizure cases and 17,811 controls with genotyping data available (Table 12).

Quality Control of Genetic Data

All samples were genotyped at the Broad Institute of Harvard and MIT (Cambridge, MA, USA) using the Illumina Global Screening Array with Multi-disease drop-in (GSA-MD v1.0). SNP genotypes were called using Illumina's genotyping analysis software Autocall. Rare SNPs (MAF < 0.1) were called with the zCall software 34 and included in the Autocall output.

QC and imputation of these samples was done in parallel with the Epi25 cohort and is detailed elsewhere [92]. Before imputation, genotyped individuals were excluded based on the following criteria: i) genotype call rate <0.95; ii) high (>0.2) or low (<-0.2) inbreeding coefficient estimate of the observed versus expected number of homozygous genotypes; iii) missing, ambiguous, or sex mismatch between X-chromosome genotype and reported gender; iv) population outliers not clustering with the 1000 Genomes Project[93] European samples in a principal component analysis. Then, single-nucleotide polymorphisms (SNPs) were excluded based on the following criteria: i) SNP call rate <0.98 in the combined case/control dataset; ii)

minor allele frequency (MAF) <0.01; iii) deviation from the Hardy-Weinberg equilibrium with $P<1x10^{-6}$. Sample and SNP QC procedures were performed using PLINK v1.9[94] and GCTA.[95] The genotyped dataset was aligned to the imputation reference (variant name, variant position, and strand orientation) using the Genotype Harmonizer.[96] Imputation to the Haplotype Reference Consortium (HRC) reference r1.1[81] was performed using reference-based phasing with Eagle v2.4[97] and Minimac4 (https://github.com/statgen/Minimac4), as implemented on the Michigan Imputation Server.[98]

After QC, SNPs were filtered out (using all imputed samples) with i) Minimac4 imputation quality score, R2 < 0.3; ii) Minimac4 squared correlation value between masked genotypes of genotyped SNPs and the imputed dosages, Emp-R2 < 0.3; iii) logistic regression P-value < 1e-4 between control/control imputation batches; iv) call rate < 0.98; v) Hardy-Weinberg p-value < 1e-5; vi) MAF < 0.01. Individuals were filtered to have max 2nd-degree relationship (kinship coefficient > 0.0884 filtered out) using KING. Post-GWAS filtering imputation R^2>0.3 and MAF>0.05, this genotyping data included 2,458,425 SNPs, 54,681 (2.22%) of which were not shared among genotyping data from any of the other 5 sites (Table 13).

Statistical Analysis

Variants associated with FS cases compared to controls were identified using the GWAS software BoltLMM, which uses a generalized mixed-model for association testing.[99,100] Additionally, sex was covaried for during the association analysis (Table 13).

Massachusetts General BioBank (MGBB)

Phenotyping Approach

The Mass General Brigham (MGB) Biobank [https://biobank.partners.org] is a hospital-based research program launched in 2010 designed to empower genomic and translational research for human health. Participants are patients above age 18 who provided informed consent to join the biobank in the MGB network (previously Partners HealthCare), including Massachusetts General Hospital, Brigham and Women's Hospital, and other affiliated institutions. For each consented subject, a collection of blood samples is obtained (plasma, serum, and DNA), which are then linked to their clinical data in the electronic health records (EHR) as well as survey data on lifestyle, behavioral and environmental factors, and family history.[101] As of December 2019, MGB Biobank has enrolled more than 120,000 participants and released genotyping array data for 36,424 subjects. MGB investigators can access the de-identified datasets from the MGB Biobank under a Data Use Agreement (DUA) without additional study protocols.

Eligible functional seizure (FS) cases from the MGB biobank were defined as those with at least one occurrence of ICD-10 codes F44.5 or R56.9 *and* at least one qualifying CPT code (group 95812-95830 and group 95950-95967), excluding subjects with at least one instance of G40.20, G40.30, G40.00, or G40.10 (Table 11). This resulted in the identification of 1,134 FS cases and 24,562 controls of EUR descent (Table 12).

Quality Control of Genetic Data

The biobank samples are genotyped on Multi-Ethnic Global array (MEGA) from Illumina (Illumina Inc., San Diego, USA) and are released in several batches, covering an average of 1.7 million genetic markers. The dataset uses genome build 37 (hg19).

Batch-specific quality control of genotype data was performed to remove SNPs with genotype missing rate >0.05, samples with genotype missing rate >0.02, and SNPs with differential missing rate >0.01 between any two batches, after which different batches were merged for subsequent QC steps. As MGB Biobank included individuals from diverse populations, genetic ancestry of biobank participants was inferred using 1000 Genomes samples (1KG) as the population reference panel.[75] Specifically, principal components (PCs) for biobank samples and 1KG samples combined were computed and a Random Forest classifier was trained to assign a "super population" label for biobank samples with a prediction probability ≥0.9 using the first 6 PCs of the 1KG samples as the training data. This resulted in 26,677 individuals classified as European (EUR), as well as a small proportion of non-EUR descents (N= 4,248). Within the EUR ancestry, samples with a mismatched reported and genetic sex, outliers of the absolute value of heterozygosity (>5SD from the mean), and one from each pair of related individuals (IBD >0.2) were removed; SNPs that showed significant batch associations at $P < 1 \times 10{-}4$, with a missing rate > 0.02 or HWE test $P < 1 \times 10{-}10$ were also discarded.

Michigan Imputation Server (Minimac4) was used to impute genotype dosages for biobank samples, with the Haplotype Reference Consortium (HRC) as the reference panel for

the EUR ancestry. Markers with imputation quality INFO score <0.8 and minor allele frequency (MAF) <0.05 were removed, resulting in 5,361,986 SNPs for analysis, 489,192 (9.12%) of which were not shared among any other biobank (Table 13).

Statistical Analysis

Variants associated with FS case status were identified by fitting a generalized mixed model to the data, adjusting for sex, current age, race, and the first 10 principal components estimated from genetic data (Table 14). Analysis was performed using the SAIGE GWAS software.[88]

BioMe

Phenotyping Approach

Participants were recruited for the Bio*Me* Biobank throughout the Mount Sinai healthcare system, as per a protocol approved by the local Institutional Review Board (IRB), initiated in 2007. Participants were recruited across age, ancestry, and across clinics throughout the healthcare system, medical and neuropsychiatric. In providing informed consent, Biobank participants authorized access to their de-identified healthcare records and donated a blood sample for extraction of genetic material for research purposes. As per the approved protocol, no disclosure/feedback of genetic results would be provided, as the analyses were for research purposes. The BioMe Biobank clinical data include longitudinal demographics, ICD9 and ICD10 codes, laboratory test results, and clinical notes, with clinical data since 2003, increasing in volume and data entry by progressive year. Participants also provided additional information on

self-reported ancestry, personal and family medical history through questionnaires administered upon enrollment.

FS cases were defined using ICD9 or ICD10 inclusion codes (without the inclusion of CPT codes in case status), identifying n=145 FS cases compared to n=9,462 controls (Table 11, Table 12).

Quality Control of Genetic Data

Bio*Me* biobank participants (n=32,595) were genotyped on the Illumina Global Screening Array (GSA). Samples were blacklisted following genotyping for gender discordance, low sequencing coverage, heterozygosity rates (falling outside of six standard deviations from the mean were excluded), contamination, low call rate, and the discovery of duplicates, yielding n=31,705 individuals for downstream analyses. A random individual from all pairs with apparent relatedness (kinship coefficient > 0.0885) was excluded. Self- reported ancestry information was then used to extract the subset of European individuals for the present analysis (n=9,607). SHAPEIT/IMPUTE2 were used to pre-phase and impute genotypes using the 1000 Genomes Phase 3 reference panel. Variants with MAF<0.05 or imputation quality R2<0.3 were filtered out, resulting in 5,356,909 variants, 62,195 (1.16%) of which were not shared among any other biobank (Table 13).

Statistical Analysis

Variants associated with FS cases compared to controls were identified using BOLT-LMM. The GWAS covaried for the top ten principal components estimated from genetic data to further control for population stratification, sex, and number of ICD codes per day (Table 14).

iPSYCH

Phenotyping approach

Genotyping samples were obtained from the Danish Neonatal Screening Biobank hosted by Statens Serum Institut. All individuals born in Denmark since 1981 have provided dried bloodspots (Guthrie cards). These samples are connected to the Danish register system. DNA is extracted from these bloodspots and genotyped at the Broad Institute. Functional seizure cases from this biobank were collected using ICD-10-based inclusion (code F44.5 for conversion disorder with seizures) and exclusion (ICD-10: G40.20, G40.30, G40.00 and G40.10) criteria (Table 11). This resulted in the identification of 232 cases and 2,660 controls for individuals of European descent (Table 12). This study has been approved by the Danish research ethical committee system.

Quality Control of Genetic Data

Genotyping data were QC'ed using the parameters SNP missingness < 0.05; subject missingness < 0.02; autosomal heterozygosity deviation (|Fhet| < 0.2); SNP missingness < 0.02; difference in SNP missingness between cases and controls < 0.02; and SNP Hardy-Weinberg equilibrium (P > 10−6 in controls or P > 10−10 in cases).[102,103] Population stratification was

controlled by with imputed marker dosages and principal components. Imputation was conducted using IMPUTE2 / SHAPEIT using 1000 genomes reference data.[104] After filtering, this data set included 4,245,406 variants, of which 59,241 (1.4%) were not found in the other biobanks (Table 13).

Statistical Analysis

Variants associated with case status were identified by fitting a logistic regression model to the data with covariates for age, sex, genotyping phase, and ancestry-specific population PC1 to PC10 (Table 14). Analysis was performed using Plink v2.0 GWAS software.

Million Veteran Program (MVP)

Phenotyping approach

The Million Veteran Program is an observational cohort study and mega-biobank in the Department of Veterans Affairs (VA).[105] Participants are active users of the Veterans Health Administration and provide a blood sample (from which DNA is isolated), responses to questionnaires and consent to allow access to clinical data from the VA health records.[105] For the current analysis, we considered 455,789 individuals (MVP v3.0 data release). Functional seizure cases from this biobank were collected using CPT-based inclusion (Group 95812-95830, Group 95950-95967), ICD-based inclusion (ICD-9: 300.11 and 780.39; ICD-10: F44.5) and exclusion (ICD-9: 345.1, 345.10, 345.4, 345.40, 345.5 and 345.50; ICD-10: G40.20, G40.30, G40.00 and G40.10) criteria (Table 11). This resulted in the identification of 7,087 cases and

291,711 controls for individuals of European descent, and 2,252 cases and 78,156 controls for individuals of African descent (Table 12).

Quality Control of Genetic Data

Pre-imputation variant level QC, pre-phasing and genotype imputation from the 1000 Genomes Project phase 3, version 5 reference panel into Million Veteran Program (MVP) participants was performed as previously described.[75,106] After filtering, this data set included 3,150,381 variants, of which 450,416 (14.3%) were not found in the other biobanks (Table 13).

Samples from 432,318 individuals passed individual-level quality control: a) sample call rates > 98.5%, b) sample heterozygosity rates which deviate 3SD or less from the samples' heterozygosity rate mean, c) related samples and samples with cryptic relationship were removed with a kinship coefficient cut-off of >= 0.0884 as measured by KING v2.0 software.[106–108] Two sources of data were used for sex determination, genotype, and core demographic data. We trained a logistic regression model to predict reported (phenotypic) sex from the F score. F scores were obtained with PLINK's --check-sex command applied to the SNPs (MAF > 0.01) located in the X chromosome (after excluding the PARs).[77] We then used the individuals' F scores to assign the sex. Individuals whose predicted sex did not match the reported one (n = 218) were excluded from the analysis.

Nonbiallelic SNPs, SNPs with R2 < 0.4, SNP call rate < 0.05, HWE p < 5x10-8, or MAF < 0.05. Ancestry-specific principal component analysis was performed using the EIGENSOFT v6 software as previously described, to generate the top ten genetic principal components

explaining the greatest variability.[78,106] Individuals of European (EUR) and African (AFR) ancestry were identified by the HARE approach as previously described.[109]

Statistical Analysis

Variants associated with case status were identified by fitting a logistic regression model to the data with covariates for age, sex, density of record, ancestry-specific population PC1 to PC10 (Table 14). Density of record was defined as anyone with more than 5 codes on separate days over the span of at least 3 years. Analysis was performed using Plink v2.0 GWAS software.[76]

Meta-analysis of genome-wide association study summary statistics

We used Metal to conduct an inverse variance-weighted meta-analysis of the summary statistics from the six studies by combining p-values across studies taking into account a study specific weight, the sample size, and direction of effect (Figure 9).[110–112] This resulted in 9,289 cases and 417,818 controls (n=427,107) available for the meta-analysis. The effective size (Neff) was 18,174, and was calculated using the following formula.[74]

$$Neff = \frac{2}{\left(\frac{1}{Ncases}\right) + \left(\frac{1}{Ncontrols}\right)}$$

A total of 1,210,977 variants were genotyped or imputed across all 6 biobanks, an additional 1,505,538 variants genotyped or imputed across 5 of the biobanks, 1,148,506 variants across 4 of the biobanks, 707,792 variants across 3 of the biobanks, and 835,406 variants across 2 of the biobanks. Variants which were only found in 1 of the biobanks

genotyping data were only included for the MVP and MGBB samples, as these are the only two samples with more than 1,000 cases identified. This resulted in 450,416 variants from MVP alone and 489,192 variants from MGBB alone. Genome-wide statistical significance was defined as P < 5.0e-8, with the meta-analysis results from METAL reported.

BioBank	MAF Filter applied	Imputation filter applied	Average imputation score	Pre-QC # SNPs	Post-QC # SNPs
BioVU	0.05	R^2 > 0.3	R^2 = 0.99	9,243,980	3,944,055
CC	0.05	R^2 > 0.3	R^2 = 0.99	3,193,188	2,458,425
MVP	0.05	R^2 > 0.3	R^2 = 0.96	7,830,224	3,150,381
MGBB	0.05	INFO > 0.8	INFO = 0.98	21,694,551	5,361,986
Mt. Sinai	0.05	INFO > 0.8	INFO = 0.96	7,749,766	5,356,909
iPSYCH	0.05	INFO > 0.8	INFO = 0.97	5,026,277	4,245,406

Table 13. Biobank-specific information for post-GWAS, pre-meta-analysis filtering. For each Biobank, we have listed the MAF filter applied, the imputation filter applied, the average imputation score, and number of independent SNPs tested, both before and after this filtering.

BioBank	Sex	Genetic Ancestry	Age	Density of Records	Genotyping phases
BioVU	X	X	X	X	
CC	X	X			
Mt. Sinai	X	X	X		
MVP	X	X	X	X	
MGBB	X	X	X		
iPSYCH	X	X	X		X

Table 14. Confounders accounted for across sites in meta-analysis GWAS. Each row represents a site of genetic FS analysis while each column represents a different confounder that is accounted for. 'X's represent that a particular site accounted for the confounder.

SNP-based Heritability of Functional Seizures

We calculated heritability estimates from the meta-analysis summary statistics using LD Score Regression (LDSC).[113–115] This method evaluates linkage disequilibrium (LD) data for each lead single nucleotide polymorphism (SNP), so that the chi square statistic of each variant includes the effects of all other loci in LD with the lead SNP.[115] The SNP-based heritability was calculated on the liability scale, using a range of prevalence from 0.02% to 0.14%.

Genetic Correlations Between Functional Seizures, PTSD, generalized epilepsy, and focal epilepsy

We calculated the genetic correlation (Rg) between functional seizures and posttraumatic stress disorder (PTSD), generalized epilepsy (GE), and focal epilepsy (FE). The summary statistics for PTSD, GE, and FE used were downloaded directly from online repositories.[36,37,40] LDSC was used, which calculates Rg similar to the heritability calculation described above, but the chi-square statistic is calculated as the product of both z scores from the two phenotypes being compared.[113]

Transcriptome-wide association study (TWAS)

A gene-based analysis was conducted using the results from our meta-GWAS and employing S-MultiXcan to assess the simultaneous effect of multiple genetic variants on gene expression, then to determine the degree of association between those genes and functional seizures.[116] To perform TWAS we leveraged pre-trained prediction models fitted in GTEx v8 data (49 tissues, N=838).[117–120] SNP weights and their respective covariance for 49 available tissues (amygdala, anterior cingulate cortex, caudate basal ganglia, cerebellar hemisphere,

cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia, spinal cord, and substantia nigra, lung, liver, kidney) from 80 to 154 individuals were obtained from predict.db (http://predictdb.org/), which is based on GTEx version 8 eQTL data.[120–122] Gene expression prediction models were then trained on GTEx data using the 49 available GTEx tissues available.[117–120,123] MetaXcan models are built on S-PrediXcan and existing approaches, and integrate eQTL information with GWAS results to map disease-associated genes. Prediction models for each tissue were integrated with FS meta analyzed GWAS data using the software S-PrediXcan.[121,122] To combine association statistics across all tissues while adjusting for tissue–tissue correlation, we used S-MultiXcan.[116] For 21,200 genes, we performed a joint multi-tissue approach using S-MultiXcan which uses multivariate regression to take advantage of the correlation in gene expression regulation between tissues thus increasing statistical power. Here, we applied a single Bonferroni correction of 0.05/21,200 = 2.35E-06.

Miami visualization plots were produced using R.[122] Miami plots show two different P value associations: the meta-analysis GWAS results on the bottom half, and the associations between the genes implicated by S-MultiXcan along each chromosome on the top half.

Results:

Individual Site FS GWAS

GWAS results were plotted for each individual GWAS (Figure 10). The only biobank to have a SNP that exceeded the genome-wide significance line was Mt. Sinai, though this association did not withstand pre-meta-analysis QC and appeared to be a false-positive

association. The associated QQ plots for each biobank's GWAS indicated that there did not appear to be any inflation issues (Figure 11). Heritability estimates for each individual GWAS confirmed that the Lambda GC for each site's analysis did not show any indication of inflation (Table 15).

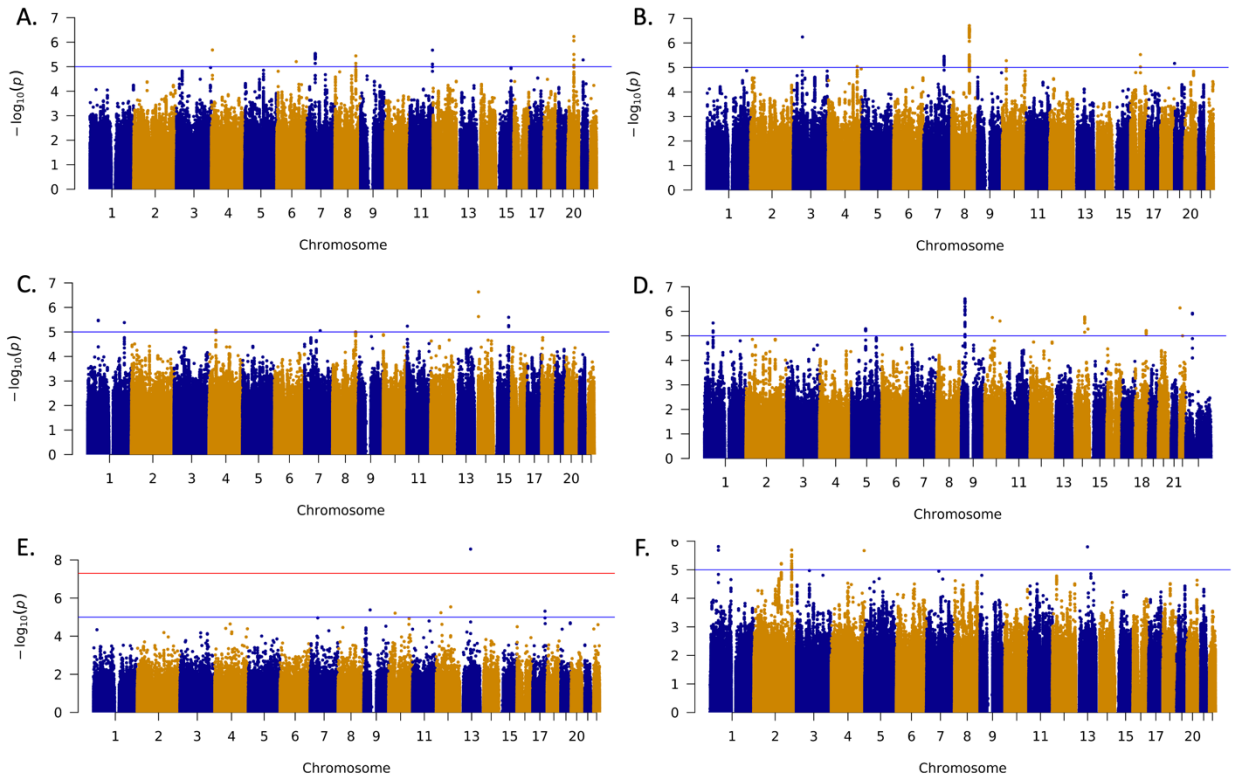


Figure 10. Biobank-specific GWAS result Manhattan plots for A) BioVU, B) MVP, C) MGBB, D) CC, E) Mt. Sinai, and F) iPSYCH. The blue line represents the suggestive p-value, p=1e-05, while the red line represents the genome-wide significance level, p<5e-08.

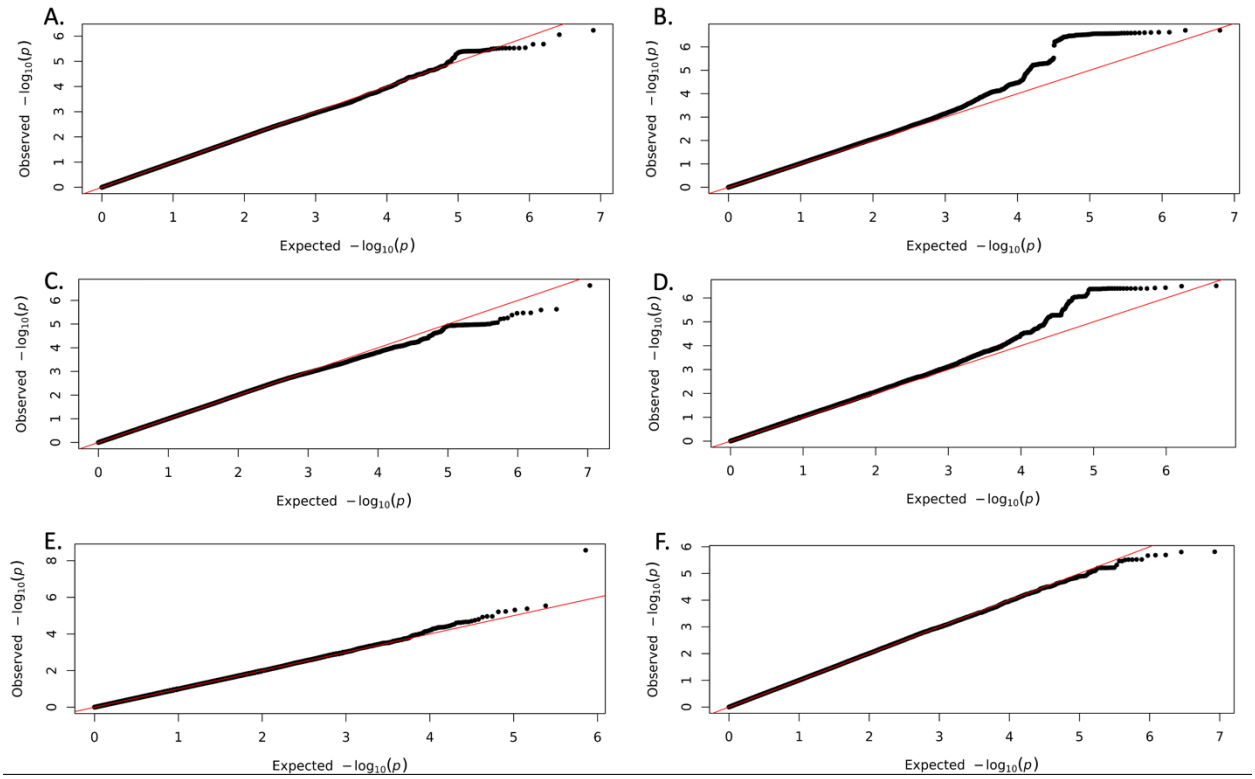


Figure 11. Biobank-specific GWAS result QQ plots for A) BioVU, B) MVP, C) MGBB, D) CC, E) Mt. Sinai, and F) iPSYCH. The red line represents the line that would result if the observed and expected p-values were the same throughout the distribution of results.

BioBank	LDSC Intercept	LDSC Lambda GC	LDSC h^2 obs	LDSC h^2 liability
BioVU	1.007 (0.008)	0.98	-0.01 (0.01)	-0.19
MGBB	1.029 (0.011)	1.047	0.05 (0.04)	0.056
MVP	0.99 (0.009)	1.026	0.01 (0.002)	0.028
CC	0.99 (0.007)	1.005	0.025 (0.019)	0.75
Mt. Sinai	1.008 (0.008)	0.999	-0.056 (0.047)	-0.39
iPSYCH	1.004 (0.009)	1.011	0.051 (0.17)	0.0597
Meta-analysis	**1.000 (0.008)**	**1.029**	**0.005 (0.0015)**	**0.022**

Table 15. Heritability results for individual functional seizure GWAS and meta-analysis. Each site's GWAS heritability estimates are shown as calculated by LDSC, including the intercept, lamda GC, observed heritability

(h^2 obs), and liability-scale heritability (h^2 liability). Meta-analysis estimates are shown on the bottom row in bold.

FS GWAS meta-analysis

Although there were no significant associations in the GWAS meta-analysis, the Manhattan plot for the meta-analysis shows several peaks beginning to emerge that may become significant as the analysis increases in power (Figure 12).

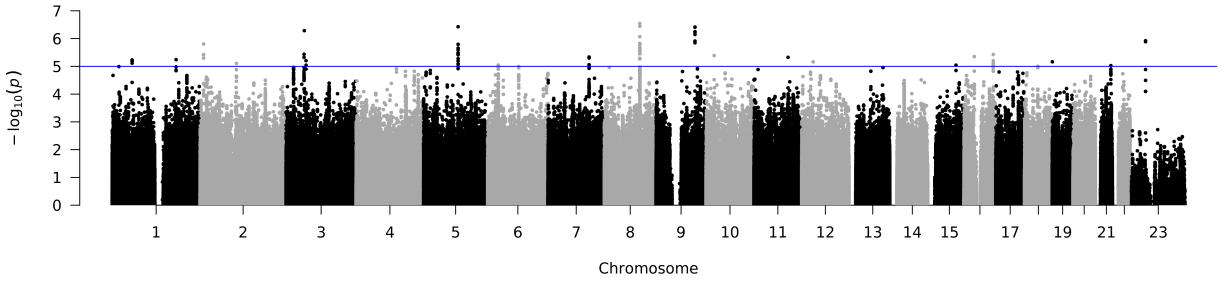


Figure 12. Biobank-specific GWAS result Manhattan plots for a GWAS meta-analysis of results from BioVU, MVP, MGBB, CC, Mt. Sinai, and iPSYCH. The blue line represents the suggestive p-value, p=1e-05, while the red line represents the genome-wide significance level, p<5e-08.

One limitation of this study was that not every biobank was able to access QC'ed X-chromosome data, resulting in FS GWAS meta-analysis X-chromosome data only from Cleveland Clinic and BioVU (Figure 12). Heritability analysis for the GWAS meta-analysis indicated that the heritability of FS in our samples is 2.21% (SE = 0.015%). Additionally, through LDSC analysis, we determined that the lambda GC and intercept of the meta-analysis were 1.03 and 1.00, respectively, suggesting no significant inflation due to population substructure or cryptic

relatedness. The QQ plot, lambda GC, and intercept for the meta-analysis did not show any signs of inflation (Figure 13).

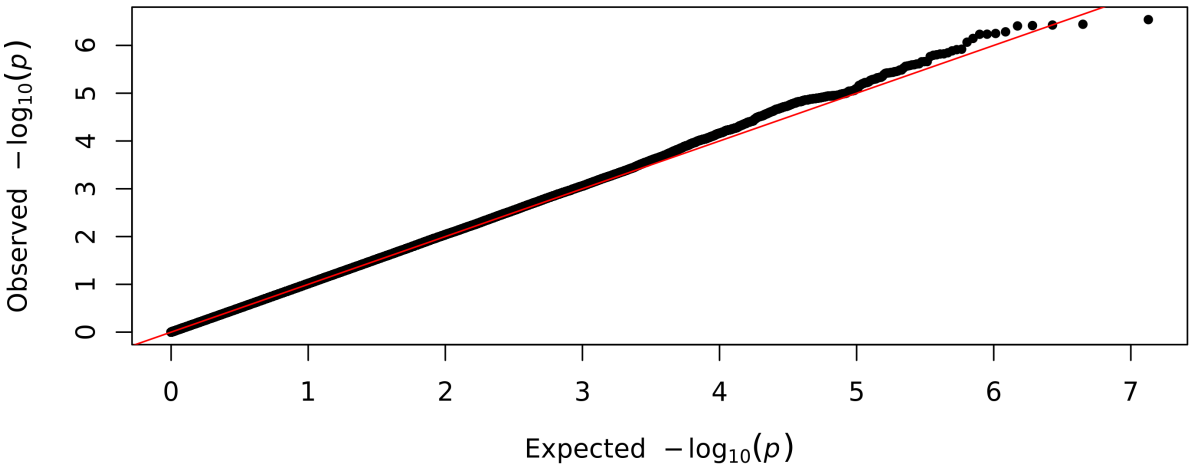


Figure 13. GWAS meta-analysis QQ plot. The red identity line represents the line that would result if the observed and expected p-values were the same throughout the distribution of results.

FS GWAS meta-analysis genetic correlation with related disorders

Genetic correlation (Rg) analysis using LDSC indicated that there was not any significant genetic correlation between FS and generalized epilepsy (Rg = -0.099, SE = 0.085, p = 0.25), focal epilepsy (Rg = -0.56, SE = 0.30, p = 0.067), or PTSD (Rg = 0.092, SE = 0.45, p = 0.65; Table 16).

Phenotype	Genetic Correlation	Z-score	P
Generalized Epilepsy	-0.099 (0.085)	-1.16	0.25
Focal Epilepsy	-0.56 (0.30)	-1.84	0.067
PTSD	0.092 (0.20)	0.45	0.65

Table 16. Genetic correlations with related disorders using LDSC. Rg (SE), Z-score, and p-value are shown for each related phenotype tested with respect to FS.

FS GWAS meta-analysis-based Transcriptome-wide association study

TWAS performed using s-MultiXcan showed 58 genes whose expression is significantly associated with functional seizures (Figure 14). The most significant TWAS associations included the genes Syntaxin Binding Protein 3 (STXBP3, Chr 1; Z = 2.55, SD = 1.79, p = 3.71e-25), TRNA-YW Synthesizing Protein 1 Homolog B (TYW1B, Chr 7; Z = -2.51, SD = 0.26, p = 1.45e-23), Family With Sequence Similarity 186 Member B (FAM186B, Chr 12; Z = -2.75, SD = 2.41, p = 1.19e-21), Methionine Sulfoxide Reductase B2 (MSRB2, Chr 10; Z = 3.38, SD = 1.14, p = 3.60e-19), and cat eye syndrome chromosome region, candidate 2 (CECR2, Chr 22; Z = 3.23, SD = 1.59, p = 6.72e-19).

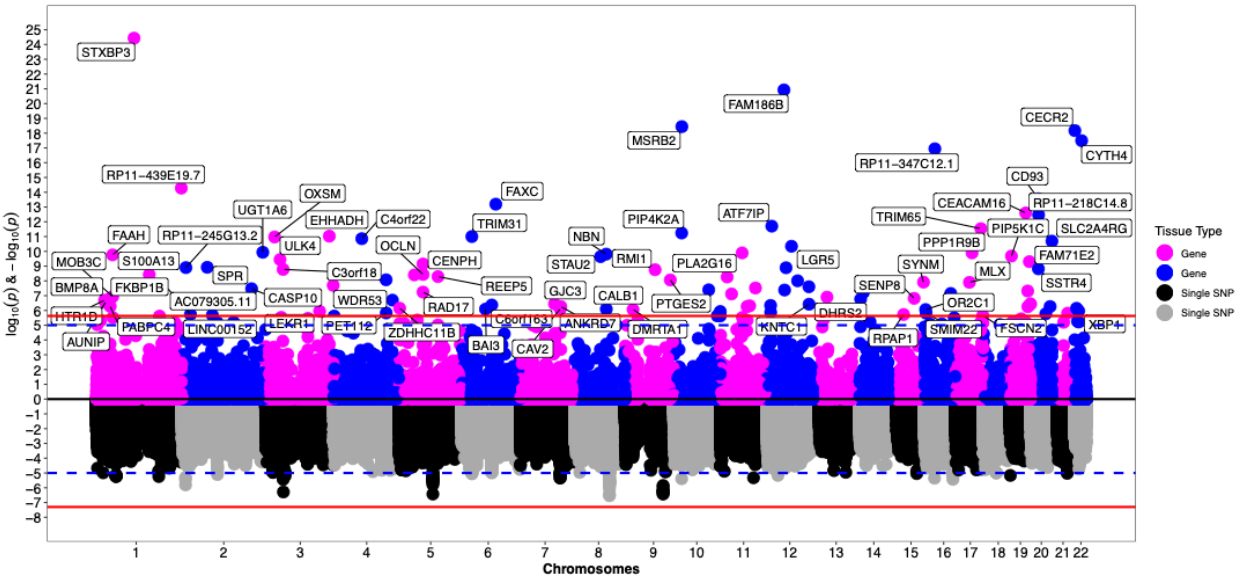


Figure 14. Transcriptome wide analysis with functional seizures cases vs. controls. TWAS results are plotted in green at the top of the figure, while the GWAS results are reflected at the bottom of the figure. Alternating chromosomes are plotted in black and grey in the GWAS and in alternating triangles vs circles in the TWAS. For

each side, the blue dotted line represents the suggestive significance line for the GWAS while the red solid line represents the genome-wide significance threshold.

Discussion:

This study aimed to examine the genetic architecture of FS. First, we identified FS patients across 6 international biobanks, BioVU, BioMe, Cleveland Clinic, MVP, MGBB, and iPSYCH using pragmatic EHR-based phenotyping approaches (Table 11). Each biobank performed a GWAS to determine the genetic associations with FS cases as compared to controls. Here, we have meta-analyzed these GWAS results, and now report the first-to-date GWAS meta-analysis of functional seizures. No individual variants exceeded the genome-wide significance threshold at this time (Figure 12), though we observed several associations that exceeded the suggestive significance threshold, for which further validation and more genetic samples are needed to confirm. We also report a SNP-based heritability for FS of 2.21% (Table 15). Genetic correlation analyses revealed that at this time, FS patients' genotypes were not significantly correlated with the genotypes of PTSD, generalized epilepsy, nor focal epilepsy patients (Table 16). Finally, transcriptome-wide association analysis revealed nearly 60 genes associated with functional seizures, implicating STXBP3 as the top association (Figure 14).

Previous studies have shown that polygenicity is inversely correlated with the discoverability of GWAS, and that sample size is correlated with discoverability. Although a peak on chromosome 8 is beginning to emerge, no variants exceeded genome-wide significance in this meta-analysis. Given the high polygenicity and relatively low sample size, this was not unexpected. Highly polygenic traits such as height, bipolar disorder, and schizophrenia did not

yield substantial discoverability until around 15,000 cases, which appears to be a crucial inflection point. Our FS meta-analysis currently includes approximately 9,000 cases (Table 12).[124] Based on a power analysis using a genetic power calculator, the number of cases needed to discover a variant significantly associated with FS with a relative risk of 1.15 and 80% power would be 28,096.[125] Given the results of our study, we hypothesize that this GWAS study is still in the so-called 'initial dead zone,' where discoverability until the critical inflection point is very low and not linear.[124] Because this dead zone depends on the number of cases and largest effect size, the genetic architecture of FS may be highly polygenic, meaning that the largest effect size will be small relative to other phenotypes, and thus FS GWAS studies may require a relatively large number of cases to get past this dead zone. Despite being underpowered for individual SNP discovery, early GWAS remain useful for SNP-based heritability and as a substrate for secondary analyses, like TWAS.

An additional limitation to the power of this study was that not all the biobanks in the analysis were able to use the same phenotyping algorithm to identify FS cases within their respective EHRs (Table 11). This was due to the variety of data types that must be accessible to produce a gold standard phenotyping approach for FS, including keywords in charts, ICD9, ICD10, and CPT codes. This limitation further underscores the lack of discoverability in this first-ever FS GWAS meta-analysis.

Due to the inherent nature of discoverability in polygenic traits, the substantial environmental risk factors known to contribute to FS (e.g., trauma), and the heterogeneity of the phenotyping approaches across the biobanks in this analysis, we anticipated that statistical power would present a challenge to this study. Nevertheless, despite the inability to identify

single SNP associations, we have revealed significant, albeit low, SNP-based heritability of FS of 2.21% (Table 15). This is the first ever reported heritability estimate of FS and the second ever study to report a heritable component of FS.[35] Although this estimate is low, it establishes a genetic contribution to FS and provides further motivation for genomic research of FS. Given the low sample size, heterogeneity in the phenotyping, and likely highly polygenic nature of the disorder, the 2% estimate is consistent with our pre-study expectations. Additionally, this heritability component is in the general range of similar, highly comorbid, polygenic disorders such as PTSD, for which the low end of the heritability range was estimated at 4%.[37] Through LDSC heritability analysis, we were also able to determine that the lambda GC and intercept of the meta-analysis were 1.03 and 1.00, respectively. This indicates that there was no significant inflation due to population stratification issues.

Genetic correlation analyses revealed that at this time, FS patients' genotypes were not significantly correlated with the genotypes of PTSD, generalized epilepsy, nor focal epilepsy patients (Table 16). Genetic correlation estimates for generalized epilepsy and focal epilepsy were negative, which may be due to the exclusion of generalized epilepsy and focal epilepsy patients from this analysis, which may have resulted in an overrepresentation of epileptic cases in the controls as compared to cases.

Using TWAS approaches, we provided gene-based associations, which are more biologically interpretable. Through this analysis, we also significantly improved power by reducing the number of independent tests, thus reducing the multiple testing burden, and increasing the effect of individual SNPs by combining eQTLs that effect each gene into a model to determine the full effect of underlying variants on gene expression. S-MultiXcan, a variation

on the MetaXcan analysis, further improves power by using information across tissues, rather than analyzing each tissue independently.[116,122] Through this analysis, we discovered 58 genes significantly associated with FS (Figure 14).

The strongest association was with the gene Syntaxin Binding Protein 3 (STXBP3; munc-18). STXBP3 is a vesicular protein expressed in tissues throughout the body.[126] It is implicated in a number of diverse physiological roles, and Munc18c-null mice develop abnormal cerebral cortex morphology.[127] STXBP3 may also play a role in insulin-dependent movement of GLUT4 and in docking/fusion of intracellular GLUT4-containing vesicles with the cell surface in adipocytes.[128] As aforementioned, there is an association between FS and obesity that has been both previously reported and corroborated by our work (Figure 8, Table 10).[8,49] This underlying association with predicted increased STXBP3 gene expression may help to explain this association, and further work will be necessary to disentangle this biological relationship. Additionally, closely-related protein STXBP1 has been associated with rare forms of childhood epilepsy, encephalopathy, and some features of autism spectrum disorder.[129–131]

The second strongest association, methionine sulfoxide reductase B2 (MSRB2) is known to protect cells from oxidative stress. The gene has been shown to play a role that is closely associated with its antioxidant activity in the pathophysiology of other neurological disorders such as Alzheimer's disease.[132]

Although the exact nature of the association between these genes and the pathophysiology of FS is not yet known, this study lays the foundation for molecular biology studies into the underlying biological causes and contributors to FS, which we believe will increase knowledge, decrease stigma, and provide avenues for future therapeutic studies. As

no pharmaceutical treatments for FS are available, we hope this emerging set of genes associated with FS will fuel the discovery process for pharmaceutical FS treatments, including drugs that target these genes which could be repurposed.

In conclusion, here we have conducted the first and only functional seizure GWAS meta-analysis to date. This resulted in the first identification of significant heritability of functional seizures. Additionally, we have identified genes associated with functional seizures, which will power future investigation into the molecular biology of FS, and which could help to identify effective drug targets to treat FS.

REFERENCES

1. Kerr WT, Janio EA, Le JM, et al. Diagnostic delay in psychogenic seizures and the association with anti-seizure medication trials. *Seizure*. 2016;40:123-126. doi:10.1016/j.seizure.2016.06.015

2. Benbadis SR. How Many Patients with Pseudoseizures Receive Antiepileptic Drugs prior to Diagnosis? *Eur Neurol*. 1999;41(2):114-115. doi:10.1159/000008015

3. Vinton A, Carino J, Vogrin S, et al. "Convulsive" Nonepileptic Seizures Have a Characteristic Pattern of Rhythmic Artifact Distinguishing Them from Convulsive Epileptic Seizures. *Epilepsia*. 2004;45(11):1344-1350. doi:10.1111/j.0013-9580.2004.04704.x

4. LaFrance WC, Baker GA, Duncan R, Goldstein LH, Reuber M. Minimum requirements for the diagnosis of psychogenic nonepileptic seizures: A staged approach. *Epilepsia*. 2013;54(11):2005-2018. doi:10.1111/epi.12356

5. Asadi-Pooya AA, Emami M. Demographic and clinical manifestations of psychogenic non-epileptic seizures: The impact of co-existing epilepsy in patients or their family members. *Epilepsy Behav*. 2013;27(1):1-3. doi:10.1016/j.yebeh.2012.12.010

6. Kutlubaev MA, Xu Y, Hackett ML, Stone J. Dual diagnosis of epilepsy and psychogenic nonepileptic seizures: Systematic review and meta-analysis of frequency, correlates, and outcomes. *Epilepsy Behav*. 2018;89:70-78. doi:10.1016/j.yebeh.2018.10.010

7. Rawlings GH, Brown I, Stone B, Reuber M. Written Accounts of Living With Epilepsy or Psychogenic Nonepileptic Seizures: A Thematic Comparison. *Qual Health Res*. 2018;28(6):950-962. doi:10.1177/1049732317748897

8. Goleva SB, Lake AM, Torstenson ES, Haas KF, Davis LK. Epidemiology of Functional Seizures Among Adults Treated at a University Hospital. *JAMA Netw Open*. 2020;3(12). doi:10.1001/jamanetworkopen.2020.27920

9. Diprose W, Sundram F, Menkes DB. Psychiatric comorbidity in psychogenic nonepileptic seizures compared with epilepsy. *Epilepsy Behav*. 2016;56:123-130. doi:10.1016/j.yebeh.2015.12.037

10. Oto M, Conway P, McGonigal A, Russell AJ, Duncan R. Gender differences in psychogenic non-epileptic seizures. *Seizure*. 2005;14(1):33-39. doi:10.1016/j.seizure.2004.02.008

11. Baslet G, Seshadri A, Bermeo-Ovalle A, Willment K, Myers L. Psychogenic Non-epileptic Seizures: An Updated Primer. *Psychosomatics*. 2016;57(1):1-17. doi:10.1016/j.psym.2015.10.004

12. Kanemoto K, LaFrance WC, Duncan R, et al. PNES around the world: Where we are now and how we can close the diagnosis and treatment gaps-an ILAE PNES Task Force report. *Epilepsia Open*. 2017;2(3):307-316. doi:10.1002/epi4.12060

13. LaFrance WC, Devinsky O. The Treatment of Nonepileptic Seizures: Historical Perspectives and Future Directions. *Epilepsia*. 2004;45(s2):15-21. doi:10.1111/j.0013-9580.2004.452002.x

14. Mandeville B. *A Treatise of the Hypochondriack and Hysterick Diseases: In Three Dialogues*. J. Tonson; 1730.

15. Todd RB. *Clinical Lectures on Paralysis, Disease of the Brain, and Other Affections of the Nervous System*. Lindsay & Blakiston; 1855.

16. Temkin O. *The Falling Sickness: A History of Epilepsy from the Greeks to the Beginnings of Modern Neurology*. JHU Press; 1994.

17. Allbutt TC. *On Visceral Neuroses: Being the Gulstonian Lectures on Neuralgia of the Stomach and Allied Disorders. Delivered at the Royal College of Physicians in March, 1884*. Churchill; 1884.

18. M B. *Traité clinique et thérapeutique de l'hystérie*. Nabu Press; 2010.

19. Asadi-Pooya AA, Brigo F, Mildon B, Nicholson TR. Terminology for psychogenic nonepileptic seizures: Making the case for "functional seizures." *Epilepsy Behav*. 2020;104:106895. doi:10.1016/j.yebeh.2019.106895

20. Stone J, Campbell K, Sharma N, Carson A, Warlow CP, Sharpe M. What should we call pseudoseizures? The patient's perspective. :5.

21. Morgan LA, Dvorchik I, Williams KL, Jarrar RG, Buchhalter JR. Parental Ranking of Terms Describing Nonepileptic Events. *Pediatr Neurol*. 2013;48(5):378-382. doi:10.1016/j.pediatrneurol.2012.12.029

22. LaFrance WC, Baird GL, Barry JJ, et al. Multicenter Pilot Treatment Trial for Psychogenic Nonepileptic Seizures: A Randomized Clinical Trial. *JAMA Psychiatry*. 2014;71(9):997-1005. doi:10.1001/jamapsychiatry.2014.817

23. Goldstein LH, Robinson EJ, Mellers JDC, et al. Cognitive behavioural therapy for adults with dissociative seizures (CODES): a pragmatic, multicentre, randomised controlled trial. *Lancet Psychiatry*. 2020;7(6):491-505. doi:10.1016/S2215-0366(20)30128-0

24. Mellers J. The approach to patients with "non-epileptic seizures." *Postgrad Med J*. 2005;81(958):498-504. doi:10.1136/pgmj.2004.029785

25. Benbadis SR. A spell in the epilepsy clinic and a history of “chronic pain” or “fibromyalgia” independently predict a diagnosis of psychogenic seizures. *Epilepsy Behav*. 2005;6(2):264-265. doi:10.1016/j.yebeh.2004.12.007

26. Asadi-Pooya AA, Sperling MR. Epidemiology of psychogenic nonepileptic seizures. *Epilepsy Behav*. 2015;46:60-65. doi:10.1016/j.yebeh.2015.03.015

27. Kerr MP, Mensah S, Besag F, et al. International consensus clinical practice statements for the treatment of neuropsychiatric conditions associated with epilepsy. *Epilepsia*. 2011;52(11):2133-2138. doi:10.1111/j.1528-1167.2011.03276.x

28. LaFrance WC, Reuber M, Goldstein LH. Management of psychogenic nonepileptic seizures. *Epilepsia*. 2013;54(s1):53-67. doi:10.1111/epi.12106

29. Hingray C, El-Hage W, Duncan R, et al. Access to diagnostic and therapeutic facilities for psychogenic nonepileptic seizures: An international survey by the ILAE PNES Task Force. *Epilepsia*. 2018;59(1):203-214. doi:10.1111/epi.13952

30. Benbadis SR, Allen Hauser W. An estimate of the prevalence of psychogenic non-epileptic seizures. *Seizure*. 2000;9(4):280-281. doi:10.1053/seiz.2000.0409

31. Diprose W, Sundram F, Menkes DB. Psychiatric comorbidity in psychogenic nonepileptic seizures compared with epilepsy. *Epilepsy Behav*. 2016;56:123-130. doi:10.1016/j.yebeh.2015.12.037

32. Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. *J Intern Med*. 2013;274(6):547-560. doi:10.1111/joim.12119

33. Cowie MR, Blomster JI, Curtis LH, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. 2017;106(1):1-9. doi:10.1007/s00392-016-1025-6

34. Roden DM, Pulley JM, Basford MA, et al. Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. *Clin Pharmacol Ther*. 2008;84(3):362-369. doi:10.1038/clpt.2008.89

35. Leu C, Bautista JF, Sudarsanam M, et al. Neurological disorder-associated genetic variants in individuals with psychogenic nonepileptic seizures. *Sci Rep*. 2020;10(1):15205. doi:10.1038/s41598-020-72101-8

36. ILAE Consortium. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat Commun*. 2018;9(1):5269. doi:10.1038/s41467-018-07524-z

37. Nievergelt CM, Maihofer AX, Klengel T, et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun*. 2019;10(1):4558. doi:10.1038/s41467-019-12576-w

38. Thomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. *Am J Hum Genet*. 2005;77(3):337-345. doi:10.1086/432962

39. Ikegawa S. A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going. *Genomics Inform*. 2012;10(4):220-225. doi:10.5808/GI.2012.10.4.220

40. Duncan LE, Ratanatharathorn A, Aiello AE, et al. Largest GWAS of PTSD (*N*=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry*. 2018;23(3):666-673. doi:10.1038/mp.2017.77

41. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 Loci Associated with Bipolar Disorder. *bioRxiv*. Published online January 24, 2018:173062. doi:10.1101/173062

42. Otowa T, Hek K, Lee M, et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry*. 2016;21(10):1391-1399. doi:10.1038/mp.2015.197

43. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3

44. Abel T, Nickl-Jockschat T. *The Neurobiology of Schizophrenia*. Academic Press; 2016.

45. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. *Pharmacogenomics*. 2009;10(2):191-201. doi:10.2217/14622416.10.2.191

46. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics*. 2010;26(9):1205-1210. doi:10.1093/bioinformatics/btq126

47. Jette N, Beghi E, Hesdorffer D, et al. ICD coding for epilepsy: Past, present, and future—A report by the International League Against Epilepsy Task Force on ICD codes in epilepsy. *Epilepsia*. 2015;56(3):348-355. doi:10.1111/epi.12895

48. Leu C, Stevelink R, Smith AW, et al. Polygenic burden in focal and generalized epilepsies. *Brain*. 2019;142(11):3473-3481. doi:10.1093/brain/awz292

49. Marquez AV, Farias ST, Apperson M, et al. Psychogenic nonepileptic seizures are associated with an increased risk of obesity. *Epilepsy Behav*. 2004;5(1):88-93. doi:10.1016/j.yebeh.2003.10.019

50. Goodloe R, Farber-Eger E, Boston J, Crawford DC, Bush WS. Reducing Clinical Noise for Body Mass Index Measures Due to Unit and Transcription Errors in the Electronic Health Record. *AMIA Summits Transl Sci Proc*. 2017;2017:102-111.

51. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31(12):1102-1111. doi:10.1038/nbt.2749

52. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. *UCLA Stat Stat Assoc*. Published online August 2014. Accessed March 4, 2020. https://dspace.mit.edu/handle/1721.1/91154

53. Martin J, Khramtsova EA, Goleva SB, et al. Examining Sex-Differentiated Genetic Effects Across Neuropsychiatric and Behavioral Traits. *Biol Psychiatry*. 2021;89(12):1127-1137. doi:10.1016/j.biopsych.2020.12.024

54. Ferrari AJ, Somerville AJ, Baxter AJ, et al. Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychol Med*. 2013;43(3):471-481. doi:10.1017/S0033291712001511

55. Saha S, Chant D, Welham J, McGrath J. A Systematic Review of the Prevalence of Schizophrenia. *PLOS Med*. 2005;2(5):e141. doi:10.1371/journal.pmed.0020141

56. Merikangas KR, Jin R, He J-P, et al. Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative. *Arch Gen Psychiatry*. 2011;68(3):241. doi:10.1001/archgenpsychiatry.2011.12

57. Kilpatrick DG, Resnick HS, Milanak ME, Miller MW, Keyes KM, Friedman MJ. National Estimates of Exposure to Traumatic Events and PTSD Prevalence Using DSM-IV and DSM-5 Criteria. *J Trauma Stress*. 2013;26(5):537-547. doi:10.1002/jts.21848

58. Roth T. Insomnia: Definition, Prevalence, Etiology, and Consequences. *J Clin Sleep Med*. 2007;3(5 suppl):S7-S10. doi:10.5664/jcsm.26929

59. Adam Y, Meinlschmidt G, Gloster AT, Lieb R. Obsessive–compulsive disorder in the community: 12-month prevalence, comorbidity and impairment. *Soc Psychiatry Psychiatr Epidemiol*. 2012;47(3):339-349. doi:10.1007/s00127-010-0337-5

60. Lépine J-P. The Epidemiology of Anxiety Disorders: Prevalence and Societal Costs. *J Clin Psychiatry*.:5.

61. Elliott DM, Mok DS, Briere J. Adult sexual assault: Prevalence, symptomatology, and sex differences in the general population. *J Trauma Stress*. 2004;17(3):203-211. doi:10.1023/B:JOTS.0000029263.11104.23

62. Bora IH, Taskapilioglu O, Seferoglu M, et al. Sociodemographics, clinical features, and psychiatric comorbidities of patients with psychogenic nonepileptic seizures: Experience at a specialized epilepsy center in Turkey. *Seizure*. 2011;20(6):458-461. doi:10.1016/j.seizure.2011.02.007

63. Hauser WA, Annegers JF, Rocca WA. Descriptive Epidemiology of Epilepsy: Contributions of Population-Based Studies From Rochester, Minnesota. *Mayo Clin Proc*. 1996;71(6):576-586. doi:10.4065/71.6.576

64. Lhatoo SD, Sander JWAS. Cause-Specific Mortality in Epilepsy. *Epilepsia*. 2005;46(s11):36-39. doi:10.1111/j.1528-1167.2005.00406.x

65. Herman ST. Epilepsy after brain insult: targeting epileptogenesis. *Neurology*. 2002;59(9 Suppl 5):S21-26. doi:10.1212/wnl.59.9_suppl_5.s21

66. Gibson LM, Hanby MF, Al-Bachari SM, Parkes LM, Allan SM, Emsley HC. Late-Onset Epilepsy and Occult Cerebrovascular Disease. *J Cereb Blood Flow Metab*. 2014;34(4):564-570. doi:10.1038/jcbfm.2014.25

67. Camilo Osvaldo, Goldstein Larry B. Seizures and Epilepsy After Ischemic Stroke. *Stroke*. 2004;35(7):1769-1775. doi:10.1161/01.STR.0000130989.17100.96

68. Seneviratne U, Briggs B, Lowenstern D, D'Souza W. The spectrum of psychogenic non-epileptic seizures and comorbidities seen in an epilepsy monitoring unit. *J Clin Neurosci*. 2011;18(3):361-363. doi:10.1016/j.jocn.2010.07.120

69. Walczak TS, Papacostas S, Williams DT, Scheuer ML, Lebowitz N, Notarfrancesco A. Outcome After Diagnosis of Psychogenic Nonepileptic Seizures. *Epilepsia*. 1995;36(11):1131-1137. doi:10.1111/j.1528-1157.1995.tb00472.x

70. Doss RC, LaFrance WC. Psychogenic non-epileptic seizures. *Epileptic Disord Int Epilepsy J Videotape*. 2016;18(4):337-343. doi:10.1684/epd.2016.0873

71. Bodde NMG, Brooks JL, Baker GA, Boon P a. JM, Hendriksen JGM, Aldenkamp AP. Psychogenic non-epileptic seizures--diagnostic issues: a critical review. *Clin Neurol Neurosurg*. 2009;111(1):1-9. doi:10.1016/j.clineuro.2008.09.028

72. Plioplys S, Doss J, Siddarth P, et al. Risk factors for comorbid psychopathology in youth with psychogenic nonepileptic seizures. *Seizure*. 2016;38:32-37. doi:10.1016/j.seizure.2016.03.012

73. Korucuk M, Gazioglu S, Yildirim A, Karaguzel EO, Velioglu SK. Semiological characteristics of patients with psychogenic nonepileptic seizures: Gender-related differences. *Epilepsy Behav*. 2018;89:130-134. doi:10.1016/j.yebeh.2018.10.032

74. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-1212. doi:10.1038/nprot.2014.071

75. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

76. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4. doi:10.1186/s13742-015-0047-8

77. 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.

78. 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(8):904-909. doi:10.1038/ng1847

79. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLOS Genet*. 2006;2(12):e190. doi:10.1371/journal.pgen.0020190

80. the Haplotype Reference Consortium, McCarthy S, Das S, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

81. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

82. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;9(2):179-181. doi:10.1038/nmeth.1785

83. Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J. Haplotype Estimation Using Sequencing Reads. *Am J Hum Genet*. 2013;93(4):687-696. doi:10.1016/j.ajhg.2013.09.002

84. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816-834. doi:10.1002/gepi.20533

85. Delaneau O, Marchini J, The 1000 Genomes Project Consortium, et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun*. 2014;5:3934. doi:10.1038/ncomms4934

86. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75-81. doi:10.1038/nature15394

87. Birney E, Soranzo N. The end of the start for population sequencing. *Nature*. 2015;526(7571):52-53. doi:10.1038/526052a

88. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335-1341. doi:10.1038/s41588-018-0184-y

89. Kuonen D. Miscellanea. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*. 1999;86(4):929-935. doi:10.1093/biomet/86.4.929

90. Dey R, Schmidt EM, Abecasis GR, Lee S. A Fast and Accurate Algorithm to Test for Binary Phenotypes and Its Application to PheWAS. *Am J Hum Genet*. 2017;101(1):37-49. doi:10.1016/j.ajhg.2017.05.014

91. Imhof JP. Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika*. 1961;48(3/4):419-426. doi:10.2307/2332763

92. Leu C, Stevelink R, Smith AW, et al. Polygenic burden in focal and generalized epilepsies. *Brain J Neurol*. 2019;142(11):3473-3481. doi:10.1093/brain/awz292

93. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

94. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7. doi:10.1186/s13742-015-0047-8

95. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

96. Deelen P, Bonder MJ, van der Velde KJ, et al. Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res Notes*. 2014;7:901. doi:10.1186/1756-0500-7-901

97. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet*. 2016;48(11):1443-1448. doi:10.1038/ng.3679

98. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

99. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284-290. doi:10.1038/ng.3190

100. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet*. 2018;50(7):906-908. doi:10.1038/s41588-018-0144-6

101. Karlson 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(1):2. doi:10.3390/jpm6010002

102. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

103. Robinson EB, St Pourcain B, Anttila V, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet*. 2016;48(5):552-555. doi:10.1038/ng.3529

104. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

105. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016

106. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet*. 2019;51(11):1574-1579. doi:10.1038/s41588-019-0519-3

107. Gelernter J, Sun N, Polimanti R, et al. Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. *Nat Neurosci*. 2019;22(9):1394-1401. doi:10.1038/s41593-019-0447-7

108. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinforma Oxf Engl*. 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559

109. Fang H, Hui Q, Lynch J, et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet*. 2019;105(4):763-772. doi:10.1016/j.ajhg.2019.08.012

110. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma Oxf Engl*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340

111. Sanna S, Jackson AU, Nagaraja R, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet*. 2008;40(2):198-203. doi:10.1038/ng.74

112. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*. 2008;40(2):161-169. doi:10.1038/ng.76

113. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

114. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211

115. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272-279. doi:10.1093/bioinformatics/btw613

116. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genet*. 2019;15(1):e1007889. doi:10.1371/journal.pgen.1007889

117. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660. doi:10.1126/science.1262110

118. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

119. Keen JC, Moore HM. The Genotype-Tissue Expression (GTEx) Project: Linking Clinical Data with Molecular Analysis to Advance Personalized Medicine. *J Pers Med*. 2015;5(1):22-29. doi:10.3390/jpm5010022

120. Barbeira AN, Bonazzola R, Gamazon ER, et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol*. 2021;22(1):49. doi:10.1186/s13059-020-02252-4

121. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091-1098. doi:10.1038/ng.3367

122. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun*. 2018;9(1):1825. doi:10.1038/s41467-018-03621-1

123. Melé M, Ferreira PG, Reverter F, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015;348(6235):660-665. doi:10.1126/science.aaa0355

124. Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry*. 2018;175(1):15-27. doi:10.1176/appi.ajp.2017.17030283

125. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19(1):149-150. doi:10.1093/bioinformatics/19.1.149

126. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419. doi:10.1126/science.1260419

127. Kanda H, Tamori Y, Shinoda H, et al. Adipocytes from Munc18c-null mice show increased sensitivity to insulin-stimulated GLUT4 externalization. *J Clin Invest*. 2005;115(2):291-301. doi:10.1172/JCI22681

128. Schillemans M, Karampini E, Hoogendijk AJ, et al. Interaction networks of Weibel-Palade body regulators syntaxin-3 and syntaxin binding protein 5 in endothelial cells. *J Proteomics*. 2019;205:103417. doi:10.1016/j.jprot.2019.103417

129. Saitsu H, Kato M, Mizuguchi T, et al. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nat Genet*. 2008;40(6):782-788. doi:10.1038/ng.150

130. Stamberger H, Nikanorova M, Willemsen MH, et al. STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology*. 2016;86(10):954-962. doi:10.1212/WNL.0000000000002457

131. López-Rivera JA, Pérez-Palma E, Symonds J, et al. A catalogue of new incidence estimates of monogenic neurodevelopmental disorders caused by de novo variants. *Brain*. 2020;143(4):1099-1105. doi:10.1093/brain/awaa051

132. Xiang X-J, Song L, Deng X-J, et al. Mitochondrial methionine sulfoxide reductase B2 links oxidative stress to Alzheimer's disease-like pathology. *Exp Neurol*. 2019;318:145-156. doi:10.1016/j.expneurol.2019.05.006