Human Genetics
https://doi.org/10.1007/s00439-022-02502-4

ORIGINAL INVESTIGATION

Identifying shared genetic factors underlying epilepsy and congenital heart disease in Europeans

Yiming Wu[1] · Cigdem Sevim Bayrak[2] · Bosi Dong[1] · Shixu He[1] · Peter D. Stenson[3] · David N. Cooper[3] · Yuval Itan[2,4] · Lei Chen[1]

Received: 16 May 2022 / Accepted: 24 October 2022

Abstract

Epilepsy (EP) and congenital heart disease (CHD) are two apparently unrelated diseases that nevertheless display substantial mutual comorbidity. Thus, while congenital heart defects are associated with an elevated risk of developing epilepsy, the incidence of epilepsy in CHD patients correlates with CHD severity. Although genetic determinants have been postulated to underlie the comorbidity of EP and CHD, the precise genetic etiology is unknown. We performed variant and gene association analyses on EP and CHD patients separately, using whole exomes of genetically identified Europeans from the UK Biobank and Mount Sinai BioMe Biobank. We prioritized biologically plausible candidate genes and investigated the enriched pathways and other identified comorbidities by biological proximity calculation, pathway analyses, and gene-level phenome-wide association studies. Our variant- and gene-level results point to the Voltage-Gated Calcium Channels (VGCC) pathway as being a unifying framework for EP and CHD comorbidity. Additionally, pathway-level analyses indicated that the functions of disease-associated genes partially overlap between the two disease entities. Finally, phenome-wide association analyses of prioritized candidate genes revealed that cerebral blood flow and ulcerative colitis constitute the two main traits associated with both EP and CHD.

Abbreviations

EP	Epilepsy
CHD	Congenital heart disease
NDD	Neurodevelopmental disorders
VEP	Variant effect predictor
MSC	Mutation significance cutoff
CADD	Combined annotation dependent depletion
PCA	Principal component analysis
IPA	Ingenuity pathway analysis
HGMD	The human gene mutation database
HGC	The human gene connectome

Yuval Itan and Lei Chen have contributed equally to this work.

✉ Yuval Itan
yuval.itan@mssm.edu

✉ Lei Chen
leilei_25@126.com

[1] Department of Neurology, West China Hospital of Sichuan University, Chengdu, Sichuan, People's Republic of China

[2] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3] Institute of Medical Genetics, Cardiff University, Cardiff, UK

[4] Icahn School of Medicine at Mount Sinai, The Charles Bronfman Institute for Personalized Medicine, New York, NY, USA

Introduction

It has been extensively reported that various neurodevelopmental disorders (NDD), in particular epilepsy (EP), are comorbid with congenital heart disease (CHD). The incidence of EP in CHD patients increases with the severity of the CHD, such that 10% of all children with CHD and > 50% of children with severe CHD develop NDD, including disorders of cognition, social behavior, speech impairment, autism spectrum disorder, stroke, epilepsy, and cerebral palsy (Argyraki et al. 2019; Homsy et al. 2015; Marino et al. 2012). The overall increased risk of EP among CHD patients compared with a general population cohort has been reported to be 3.7-fold (95% confidence interval [CI] 3.2–4.2) before the age of 5- and 2.3-fold (95% CI 2.1–2.7) among those aged 5–32 years (Leisner et al. 2016). However, the shared pathophysiological mechanisms of neurological disorders and congenital heart disease are still largely unknown. It has been postulated that vasoactive substances may be discharged or metabolized through the pulmonary

Published online: 09 November 2022

circulation where, due to congenital heart malformations, venous blood can mix with arterial blood by shunting without circulating in the lungs. Chemicals and hormones, such as serotonin, may therefore bypass the pulmonary circulation and traverse the blood–brain barrier, thereby contributing to neurological disorders. This may partially explain the increased risk of EP in CHD patients, as reported in previous studies (Billett et al. 2008; Desnous et al. 2019; Leisner et al. 2016; Massin et al. 2007). Although additional unexplained causes of the EP-CHD comorbidity have been partially attributed to genetic factors, the genetics underlying the comorbidity remain largely unknown (Gonzalez-Sulser 2020). Previous studies have proposed a number of genetic factors to be associated with EP-CHD comorbidity (Tan et al. 2014) either by outcome observations or through genetic analyses of patients. Thus, for example, genes encoding potassium channels (KCN) and Zinc Finger E-Box Binding Homeobox 2 (ZEB2) have been suggested to cause channelopathies of the heart and brain (Omichi et al. 2010; Yamada et al. 2014). Another associated genetic factor is the 15q11.2 BP1–BP2 micro-deletion causing Burnside–Butler syndrome, where the maternal deletions tend to be associated with epilepsy, whereas the paternal deletions are mostly associated with CHD (Davis et al. 2019; Vanlerberghe et al. 2015). Moreover, EP and CHD are frequently seen in patients with chromosomal structure variations, such as trisomy 21 and Turner syndrome (Matricardi et al. 2016; Morales-Demori 2017). However, while studies performed on sporadic cases have indicated that variants or genes may contribute to both EP and CHD, cohort-based studies are required to investigate EP and CHD comorbidity genetics at the general population level.

To further identify novel variants and genes underlying the comorbidity of EP and CHD, and to better understand the underlying biological mechanisms, we performed variant-level association and gene burden analyses using 333 CHD and 1599 epilepsy samples from 200 K whole exomes in the UK Biobank. Additionally, we conducted pathway analyses to assess whether there are overlapping functions between EP genes and CHD genes. Finally, we performed gene-level phenome-wide association analyses (PheWAS) to identify phenotypes that correlate with both EP and CHD, and report novel pleiotropies for these genes.

Materials and methods

Sample collection

Participant data were acquired from the UK Biobank, a prospective cohort study with deep genetic and phenotypic data collected on individuals from across the United Kingdom, aged between 40 and 69 at recruitment; 200,000 of the participants have supporting whole exome sequencing data. The cases and controls were identified from the main and secondary diagnosis forms (UK Biobank field ID: 41202 l, 41,204) where the diagnosis codes of participants were acquired from their hospital inpatient records. The phenotypes were coded according to the International Classification of Diseases version-10 (ICD-10). Those samples with WES data were retained for case–control studies. As a result, we obtained 1776 epilepsy cases with G40 and G41 (excluding secondary epilepsy with G40.5), and 375 cases of congenital heart disease with Q20-Q26. To minimize the bias in controls with closely related phenotypes, samples with congenital malformations, deformations, chromosomal abnormalities, diseases of the nervous system, and diseases of the circulatory system were excluded from the WES samples. The remaining samples were deemed to represent controls without any congenital or neurological disease. A total of 125,708 samples qualified as controls.

Genotype calling and quality control

The reads mapping and genotype calling process has been described elsewhere (Szustakowski et al. 2021; Yun et al. 2021). Further quality control was applied so as to ensure high-quality samples and variants in downstream case–control tests. Individuals were excluded/retained according to the following rules: missing genotype rate greater than 5%; discordance between inferred gender from genotype and self-reported gender; duplicates identified with KING (Manichaikul et al. 2010). Only one sample was retained from pairs of identified first- or second-degree relatives, the cases having higher priority than controls to be retained from each pair of related samples. Variants were removed on the basis of the following criteria: variants with gnomAD European MAF > 1% were excluded from SKAT analyses, genotype rate < 95% across samples, and extreme deviation from Hardy–Weinberg equilibrium ($p < 1 \times 10^{-5}$) in controls. All quality control filtering was performed using PLINK1.9 and R. QC-passed variants were used to perform variant-level association analyses.

Genetic ancestry estimates

Admixture analyses were performed to identify Europeans from UK Biobank participants. A principal component analysis (PCA) plot was employed to validate genetically identified samples. The PCA and population structure analyses were performed, based on the same set of variants filtered by the following process: merging all UK Biobank WES samples with all reference panels by Plink, then reducing linkage disequilibrium (LD) between markers (--indep-pairwise 50 5 0.2) by removing all markers with

$r^2 > 0.2$ (window size 50, step size 5), as well as markers in known high LD regions. Variants with MAF > 0.02 and genotyping rate > 95% across the dataset (excluding A/T, C/G mutations). In population structure analyses, the participants were compared to Utah residents with northern and western European ancestry (CEU), east Asian (EAS), and Yoruba in Ibadan (YRI) from the 1,000 Genomes Project database; the parameter K was set to 3 in admixture analysis; samples with a European fraction greater than or equal to 0.8 were deemed to correspond to genetically identified Europeans. We repeated the aforementioned process to identify Europeans from 40 K Mount Sinai BioMe BioBank whole exomes. As a result, 10,796 Europeans were obtained from BioMe Biobank.

Variant annotation

The variants derived from the UK Biobank 200 K exomes were annotated by Variant Effect Predictor (VEP, v90) (McLaren et al. 2016). We used a Python script to manage a parallel running annotation method and merging CADD scores (v1.6) (Rentzsch et al. 2018) into final results. All annotation processes were conducted based on GRCh38 genome coordinates. The most severe category was picked if a variant had multiple annotation results due to multiple gene isoforms.

High impact rare variant filtration

We retained rare and high-impact genetic variants for the SKAT-O test using the following criteria: (1) Utilized Variant Effect Predictor (VEP) to determine the effect of all variants. Variants were filtered by 'consequence' of VEP annotations, high-impact variants being retained by virtue of their impact on genome functions: 'missense variant', 'start lost', 'stop lost', 'stop gained', 'splice_acceptor_variant', 'splice_donor_variant', 'inframe_insertion', 'inframe_deletion', 'protein_altering_variant', 'start_retained_variant', 'stop_retained_variant', 'stop_gained', 'stop_lost', 'regulatory_region_variant', and 'frameshift_variant'. (2) Removed variants with MAF > 0.01 according to gnomAD European allele frequency. When gnomAD allele frequency was missing for a given variant, the allele frequency from UK Biobank exomes was used instead. (3) Employed Mutation Significance Cutoff (MSC) (Itan et al. 2016) to control the false-negative rate of predicted deleterious mutations by well-established predictors, including CADD, SIFT, and Polyphen-2. Here, we retained variants with a CADD score larger than the lower boundary of 95% confidence interval of the corresponding gene's pathogenic mutation's CADD score. (4) Variants that were frequent in a given exome cohort, but absent or rare in public databases, and had also been reported as non-pathogenic variants (NPV). We removed all variants that had been described in the precalculated 'blacklist' (Maffucci et al. 2019). The remaining variants were used for further analyses.

Association analyses and statistics

Variant-level association tests were implemented in plink1.9 with a model-based approach, based on an additive genetic model. All variants, not limited to the high-impact variants, were checked in relation to their associations with EP and CHD using logistic regression. Gender information was utilized as covariates in a regression model. The gene-level association analyses were conducted using an optimized Sequence Kernel Association Test (SKAT-O) with R package SKAT (version 2.0.1). High impact rare variants were aggregated into the genes harboring them for the purpose of running SKAT-O for EP and CHD. We used a genome-wide significance of $p = 5 \times 10^{-8}$ as the variant-level significance. The gene-level significances were adjusted by reference to the number of tests applied following Bonferroni correction. For the gene-level PheWAS analyses, the phenome-wide significance was adjusted by the number of phenotypes having more than 50 cases. P values obtained from EP and CHD variant- and gene-level associations were combined into merged p values by means of a Cauchy combination test running in R package ACAT (Liu et al. 2019). To merge p values at the variant level, the original p values of P_{EP} and P_{CHD} were weighted by OR_{EP} and OR_{CHD}, respectively. For the gene p values, the merging process was unweighted, since genetic effect size differences between EP and CHD are currently unknown.

Phenome-wide association analysis

To evaluate the indirect associations between EP genes and CHD genes, we performed gene-level PheWAS using whole exome sequencing data and diagnostic information for 200 K whole exome sequenced samples. The cases were collected according to their ICD-10-CM codes; phenotypes with at least 50 cases being retained for PheWAS. To minimize the bias from controls which could have included similar or relevant phenotypes as cases, we removed controls having the same ICD-10 codes as cases, and/or having codes belonging to the same super-phenotype categories of codes in cases. Meanwhile, given that the number of cases is relatively small compared to the control set, to minimize inflation due to the extremely unbalanced numbers of cases versus controls, for each phenotype, we randomly selected a subset from the pool of controls to ensure that the ratio of cases to controls was 1:5. Overlapping individuals were removed from the controls dataset. Principal component

analysis (PCA) had been applied to all exomes prior to the PheWAS analyses, where the first two components were used to adjust population structure in the association tests. An R script was written to perform the gene-level PheWAS by combining the 'PheWAS' package (Carroll et al. 2014) and the 'SKAT' package. For each candidate gene, the high-impact rare variants from the UK Biobank whole exomes were used to repeat the SKAT-O tests to identify associations with each phenotype.

At the variant level, to ensure adequate power for the variant-level association analysis, phenotypes having more than 100 cases were included in the variant-level PheWAS. The controls were randomly sampled from the non-cases as the five times the number of cases. There were 316 phenotypes having at least 100 cases, which yielded a phenome-wide significance of $p = 1.58 \times 10^{-4}$; information on gender and the first 10 PCs were employed as the covariates in running PheWAS.

Pathway analyses

The candidate genes for EP and CHD were determined based on SKAT test results ($p < 0.01$) and were investigated for possible shared pathways, GO terms, and biological functions. Enrichment analyses were implemented using QIAGEN Ingenuity Pathway Analysis (http://www.qiagen.com/ingenuity) and the InnateDB analysis tool (https://www.innatedb.com/). InnateDB performs a hypergeometric distribution test to identify over-represented GO terms and pathways (imported from KEGG, NetPath, PID NCI, Reactome, INOH, and PID BioCarta) that are represented more than would be expected by chance alone. IPA software then determined the statistical significance of canonical pathways, diseases, biological functions, and networks that were most relevant to the input genes. IPA analysis included the following parameters: (i) Ingenuity Knowledge Base (genes only) was used as the reference set; both direct and indirect relationships were considered; (ii) endogenous chemicals were included in the networks interaction, the number of molecules per network was selected as 70, and the number of networks was selected as 10; (iii) all node types and all data sources were used; (iv) only experimentally observed information was considered; (v) molecules and interactions were limited to mammals; (vi) molecules and relationships were selected from all tissues and cell lines; and (vii) all mutation findings were used.

The significance of the overlapping pathways between EP and CHD was determined using the Fisher's exact test. The size of the universal set was considered as 43,558 for the GO terms and 734 for the IPA canonical pathways based on the July 2022 releases of both databases.

Gene prioritization

All candidate genes were prioritized by measuring their biological distance to the known EP and CHD genes. The candidate genes of EP and CHD were collected from SKAT-O analyses in the discovery phase using UK Biobank exomes and the replication phase using BioMe Biobank exomes; genes with $p < 0.05$ were deemed to be candidate genes. The known EP and CHD genes were summarized from the HGMD Professional version (HGMD_PRO_2021.2) (Stenson et al. 2020); genes harboring sites known to be associated with EP or CHD were taken as known genes for the condition. The Human Gene Connectome (HGC) (Itan et al. 2013) was employed to calculate the average distance between each candidate gene to known genes.

Phenotype enrichment analysis

We collected 2400 (200 times the number of cases) age, gender, and race-matched controls, and then identified the phenotypes enriched in the cases. The phenotypes for each individual were obtained from the main diagnosis form from UK Biobank. Then, the chi-squared test was performed to compare the distributions of affected samples in cases and controls for each phenotype.

Results

Identifying European cases from UK Biobank whole exomes

Samples in this study were collected from the UK Biobank whole exome sequencing (WES) data (October 2020 release), that include genetic, lifestyle, and medical information from 203,643 volunteer participants. The process of sample recruitment, exome sequencing, and genotype calling has been described in the UK Biobank release study (Sudlow et al. 2015). We applied additional quality control and population structure analyses to the remaining samples, obtaining high-quality samples for analysis (Methods) (Fig. 1). We genetically identified 188,113 Europeans having a European component larger or equal to 0.8 in the admixture analyses (Fig. 2A). We further excluded 8,548 samples that did not pass our additional QC procedures (Methods), leaving 179,565 WES European individuals for the selection of cases and controls (Methods). The EP and CHD cases were selected from the aforementioned QC-passed individuals according to the ICD-10 diagnostic information in UK Biobank (Methods). To minimize the bias from controls displaying phenotypes close to EP and CHD, some phenotypes

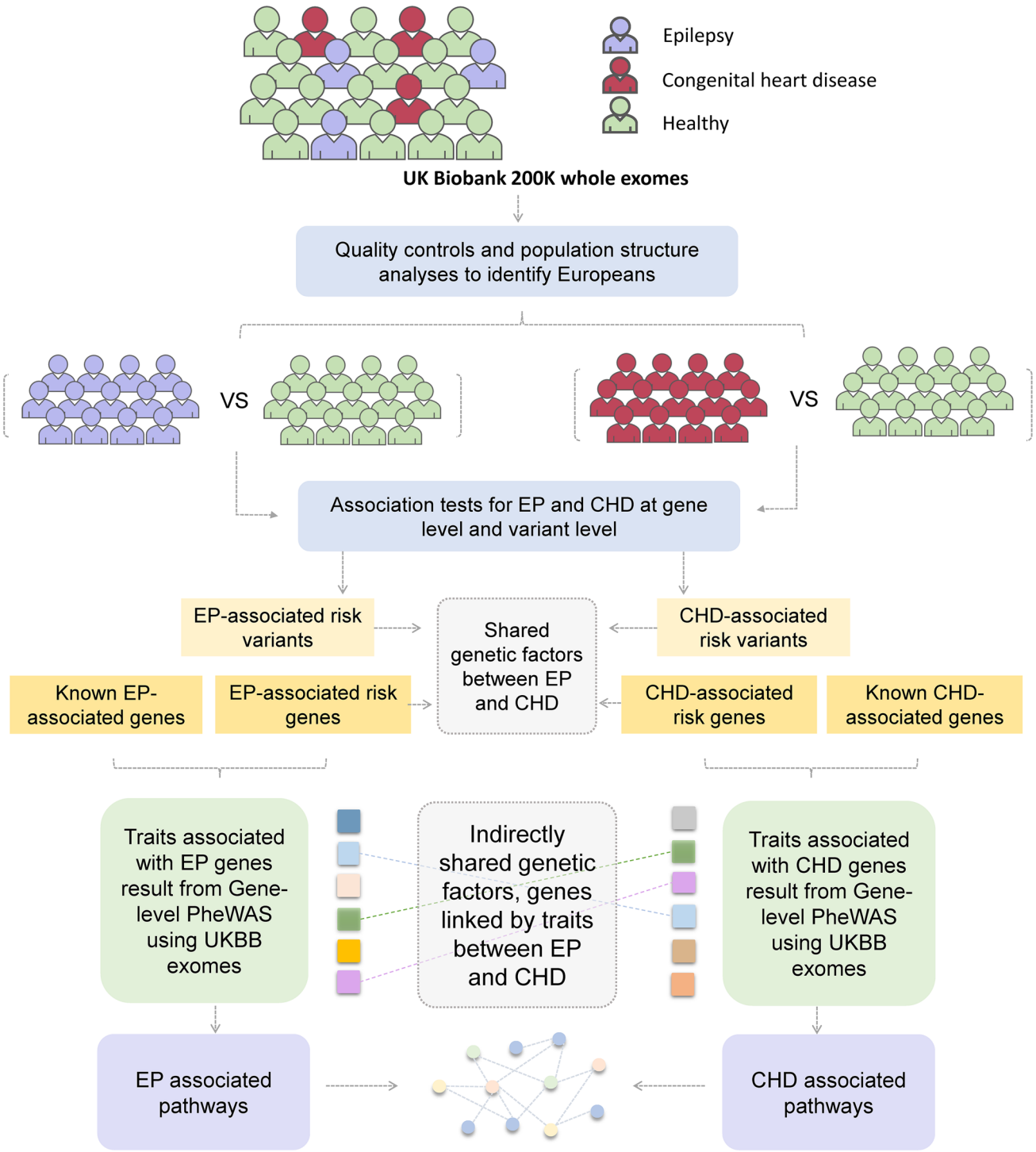

Fig. 1 Study workflow. Quality control estimates were performed on whole exomes of 200 K UK Biobank samples, and further identified Europeans from QC-passed individuals. EP and CHD samples were collected based on diagnostic information (ICD-10 codes), and then compared with European controls unaffected by both diseases to perform variant-level and gene-level associations. Pathway analyses and gene-level PheWAS were performed to identify overlapping gene functions and co-existing traits between the two diseases

were excluded from the controls. Specifically, samples exhibiting congenital malformations, deformations and chromosomal abnormalities (Q00–Q99), diseases of the nervous system (G00-G99), diseases of the circulatory

system (I00–I99), and epileptic seizures related to external causes (G40.5) were excluded from this study. Finally, 1599 European EP cases and 333 European CHD cases were retained for variant- and gene-level association tests, while the remaining 112,732 European WES samples were employed as potential controls (Methods, see Table 1 for sample summary). To reduce potential statistical inflation due to an extremely unbalanced ratio of cases to controls, we randomly picked as controls 2000

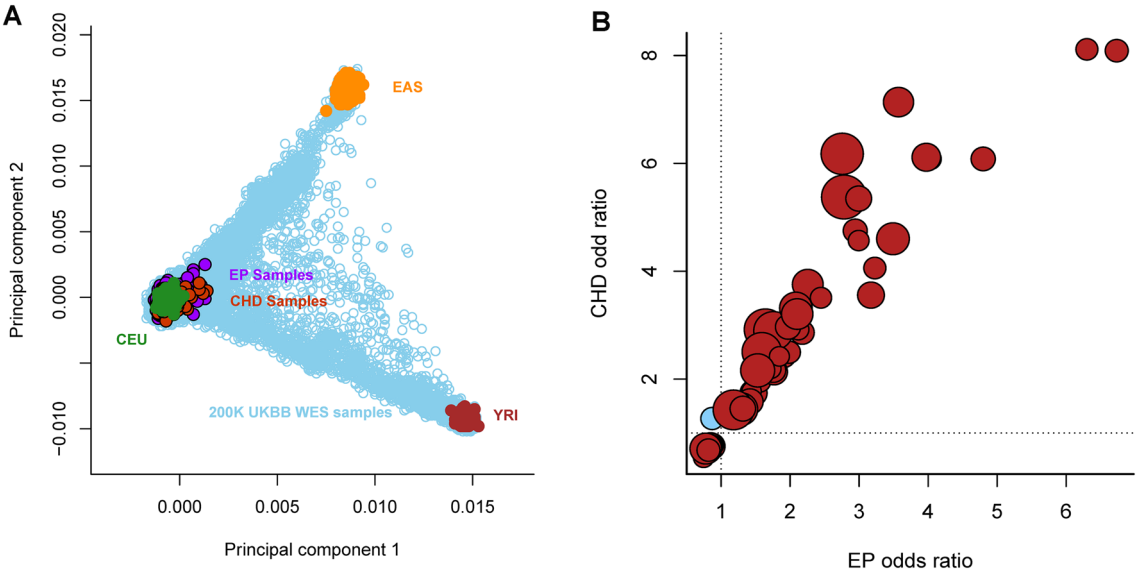

Fig. 2 **A** Principal component analysis plot displaying the filtered European cases resulting from population structure analyses for epilepsy and congenital heart disease. Europeans having a European fraction larger than 0.8 were deemed to be genetically identified Europeans. **B** Odds ratios (OR) of variants: 67 of 68 variants (86%) exhibited effects in the same direction for both diseases ($p < 2.2 \times 10^{-16}$, by binomial test). Red indicates variants for which both diseases have the same direction of effect; blue indicates opposite-direction effects. Only the variants passing nominal significance ($p < 0.01$) in both EP and CHD analyses are shown. Circle sizes are proportional to the *P* value of each variant

Table 1 Summary of case and control data

Phenotypes	ICD10	Total number	No. of Europeans	No. of QC-passed Europeans
Epilepsy and recurrent seizures	G40, G41 (excluding G40.5)	1776	1686	1599
Congenital malformations	Q20-Q26	375	362	333
EP + CHD	(Q20-Q26) & (G40, G41)	12	12	12
Controls*	Excluding Q00-Q99, G00-G99, I00-I99	125,708	117,896	112,732

*2000 randomly sampled from the pool of controls

Table 2 Variants ($p \leq 0.001$) that have risk effects in both EP and CHD

Variants	RsID	Gene	PFAM	Epilepsy		Congenital heart disease		Merged *P* value
				OR (95% CI)	*P* value	OR (95% CI)	*P* value	
22:39,647,872:C:A	rs57732048	*CACNA1I*	PF00520	3.49 (3.42–3.57)	2.24×10^{-4}	4.60 (4.48–4.73)	6.03×10^{-4}	3.48×10^{-4}
12:14,790,617:G:A	rs148994715	*WBP11*	PF12622	2.78 (2.73–2.84)	9.35×10^{-4}	5.38 (5.25–5.51)	1.17×10^{-5}	1.75×10^{-5}
7:132,972,920:A:T	rs7787607	*CHCHD3*	PF05300	1.29 (1.28–1.29)	1.53×10^{-4}	1.44 (1.43–1.45)	9.70×10^{-4}	2.75×10^{-4}
7:132,973,324:C:A	rs2058945	*CHCHD3*	PF05300	1.29 (1.28–1.29)	1.53×10^{-4}	1.44 (1.43–1.45)	9.70×10^{-4}	2.75×10^{-4}
9:132,537,526:G:A	rs13286286	*CFAP77*	PF14825	1.18 (1.17–1.18)	6.66×10^{-4}	1.43 (1.42–1.44)	2.85×10^{-4}	5.01×10^{-5}

non-CHD and non-EP individuals, approximately the sum of EP and CHD patients.

Exome-wide variant-level associations

We performed single-variant association analyses for EP and CHD, respectively (Table 2). Although none of the variants reached genome-wide significance ($p = 5 \times 10^{-8}$) in either CHD or EP association tests (Fig. S1), the overlapping significant variants exhibited very consistent effects for both diseases (Fig. 2B). In the CHD variant-level association analysis, the most significant locus was *MUC17*, which has not yet been reported to be associated with CHD. From the association analysis in EP, rs9919234 (1:245540688:G:T, OR = 0.81, $p = 1.03 \times 10^{-5}$) was the most significant variant, located within the gene *KIF26B*, a member of the kinesin gene family that includes 46 genes. Some members of the kinesin gene family (*KIF5C*, *KIF1A*, *KIF14*, *KIF11*, *KIF10*, *KIF15*, *KIF2A*, and *KIF26B)* are known to be associated with microcephaly, which is characterized by seizures and developmental delay (Kalantari and Filges 2020; Wojcik et al. 2018).

To identify potential risk variants shared between EP and CHD, we cross-checked the lists of CHD-associated variants and EP-associated variants (with a relaxed cutoff of $p \leq 0.001$ to allow inclusion of additional candidates). As a result, we identified 5 SNPs (rs148994715, rs57732048, rs7787607, rs2058945, rs13286286) which exhibited associations with both traits, all of which had the same direction of risk (i.e., risk or protective) for EP and CHD (Table 2). Of these variants, rs148994715 (p.Thr383Ile, $OR_{EP} = 2.78$, $p_{EP} = 9.3 \times 10^{-4}$, $OR_{CHD} = 5.37$, $p_{CHD} = 1.1 \times 10^{-5}$, $p_{merge} = 1.75 \times 10^{-5}$, CADD: 21.5) and rs57732048 (p.His505Asn, $OR_{EP} = 3.49$, $p_{EP} = 2.2 \times 10^{-4}$, $OR_{CHD} = 4.60$, $p_{CHD} = 6.0 \times 10^{-4}$, $p_{merge} = 3.48 \times 10^{-4}$, CADD: 0.049) had the highest effect size with both traits compared to the other three variants. The variant rs148994715, located within the gene *WBP11*, has an OR = 5.376 in CHD patients, which is almost twice the OR = 2.78 in EP patients. It has been reported that *WBP11* may be involved in multiple congenital defects, with loss-of-function and deleterious missense variants in this gene causing cardiac malformations (Martin et al. 2020). Further, Amentoflavone, a neuroprotective drug, is known to significantly downregulate the expression of *WBP11* (Liu et al. 2020), consistent with *WBP11* being implicated in the pathogenesis of both EP and CHD. The other significant variant, rs57732048, located within the gene *CACNA1I*, has high ORs in both EP and CHD patients in relation to controls ($OR_{EP} = 3.49$, $OR_{CHD} = 4.60$). *CACNA1I* encodes the pore-forming alpha subunit of a voltage-gated calcium channel involved in calcium signaling; mutations in this gene can cause congenital heart disease and neurodevelopmental disorders, including seizures and epilepsy (El Ghaleb et al. 2021; Izarzugaza et al. 2020). Both *WBP11* and *CACNA1I* have pathogenic variants recorded in the Human Gene Mutation Database (HGMD) Professional version (HGMD_PRO_2021.2) (Stenson et al. 2020). Mutations in *WBP11* are associated with multiple congenital defects, whereas *CACNA1I* has been implicated in congenital heart disease and neurodevelopmental disorders (Schizophrenia and Autism Spectrum Disorder). The *CHCHD3* gene harbors two variants having an identical effect size on EP and CHD; we therefore performed linkage disequilibrium analysis for the variants rs7787607 (7:132972920:A:T) and rs2058945 (7:132973324:C:A) using 189,448 QC-passed exomes in UK biobank. The results indicate that these variants indeed have strong LD with each other ($R^2 = 0.992$ in LD analysis using Plink1.90). The frequencies of haplotype 'TA' are 0.169 and 0.186 for epilepsy and congenital heart disease, respectively, while the frequency for controls is 0.137. CHCHD3 is a scaffolding protein that stabilizes protein complexes involved in maintaining the architecture of mitochondrial cristae and protein import. Disruption of the cristae has been implicated in a variety of cardiovascular and neurodegenerative diseases (Darshi et al. 2011; Lionel et al. 2011).

To discover the phenotypes associated with the risk variants for both EP and CHD, we additionally performed variant-level phenome-wide association analysis for the variants given in Table 2 by inspecting 316 phenotypes having at least 100 cases in the UK Biobank. Interestingly, in the results, we found that the most significant associated phenotype was cerebral infarction (ICD10: I63.9; $p = 1.70 \times 10^{-3}$; OR = 1.36, Fig. S2). Cerebral infarction is also known as ischemic stroke, which links EP to CHD. Previous studies have shown that CHD patients have a fivefold increased risk of ischemic stroke compared to controls (Mandalenakis et al. 2016). Ischemic stroke is commonly known as a risk factor for epileptic seizures. This is further evidence to support the view that genetic factors may underline the co-occurrence of EP and CHD.

Exome-wide gene-level associations

We additionally performed gene-based association analyses by variant aggregation, using the optimal sequence kernel association test (SKAT-O) to examine shared risk genes between EP and CHD. High impact rare variants meeting the variants' inclusion criteria (236,651 variants, see Methods) were aggregated into 17,453 genes for the gene-level association tests (Methods), which yielded a Bonferroni-adjusted significance p value = $0.05/17453 = 2.86 \times 10^{-6}$. We did not identify any genes passing Bonferroni-adjusted significance among either CHD- or EP-associated genes (Table S1 and S2). However, we found that the top genes from each gene-level results were plausible disease genes for

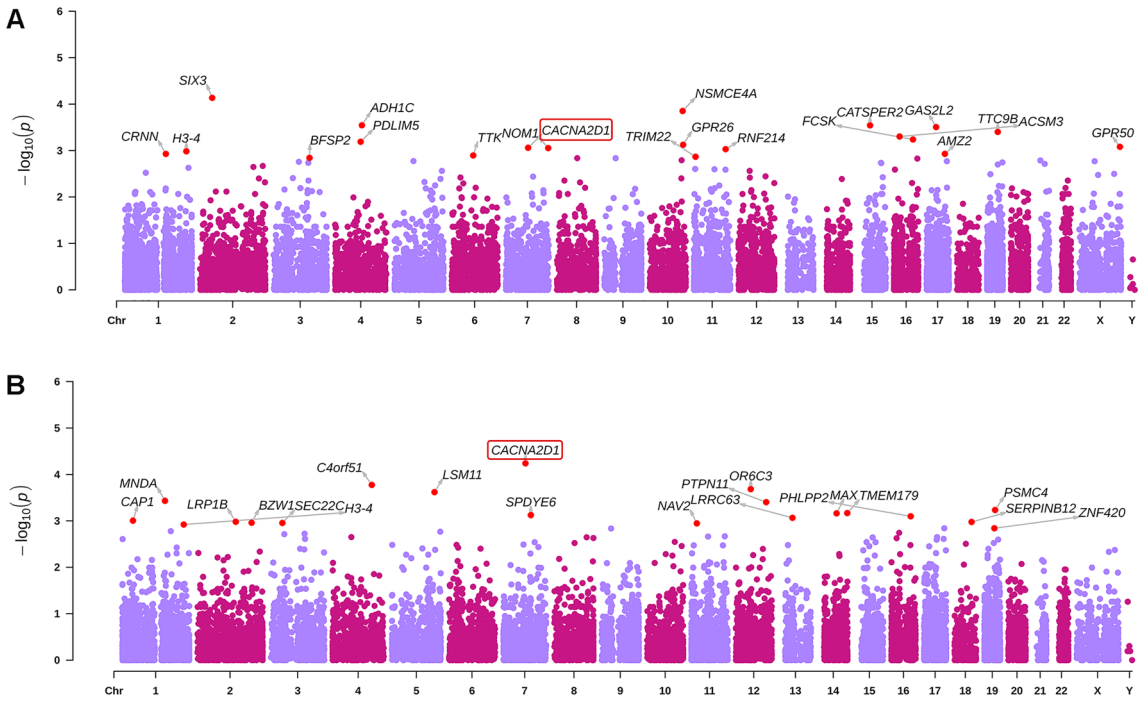

Fig. 3 Manhattan plot displaying gene-level association analysis results from SKAT-O tests for EP (**A**) and CHD (**B**), respectively. The top 20 genes are highlighted in red. *CACNA2D1* is the most significant gene associated with both diseases according to merged *P* values of the genes associated with both conditions

EP and CHD, respectively. In the gene-level association test results for EP, *SIX3* was the most significant gene associated with EP ($p = 4.87 \times 10^{-5}$, Fig. 3A). *SIX3* is associated with schizencephaly, a brain malformation characterized by various neurological symptoms including epileptic seizures (Hehr et al. 2010). For the associations with CHD, *CACNA2D1* was the top ranked gene with $p = 7.15 \times 10^{-5}$ (Fig. 3B), representing a second calcium channel gene identified in this study.

To identify potential risk genes for both EP and CHD, we again checked the overlapping disease-associated genes between EP and CHD ($p < 0.01$), identifying five genes: *CACNA2D1*, *H3-4*, *SCYL2*, *CATSPER2*, and *AHNAK*. In addition to the association with CHD, *CACNA2D1* is also associated with EP, indicating that calcium signaling channel variants may partially explain the higher prevalence of epilepsy in CHD patients and vice versa. Among the genes associated with both traits, *CATSPER2* has previously been reported as a gene in which mutations cause male infertility, and carriers of a micro-deletion in this gene have developed severe symptoms of epilepsy and various combinations of intellectual disability, developmental delay, and autism spectrum disorders (Nicholl et al. 2013). *SCYL2* and *AHNAK* have been implicated in the pathophysiology of tetralogy of Fallot, where *SCYL2* has lower expression in cyanotic as compared to acyanotic hearts (Sacks et al. 2018). Interestingly, *AHNAK* (a calcium signaling gene) exhibits higher expression levels in tetralogy of Fallot as compared to ventricular septal defect (Bond et al. 2018). *H3-4* has not yet been reported as being associated with EP or CHD.

Given that the 12 samples with EP and CHD have limited power for running association analysis, we performed phenotype enrichment for the 12 samples. Other than the phenotypes which were subjectively selected for this study, G81.9 (Hemiplegia, $p = 5.0 \times 10^{-39}$), G81.8 (Other specified paralytic syndromes, $p = 3.35 \times 10^{-34}$), I69.4 (Sequelae of stroke, $p = 8.39 \times 10^{-26}$), I67.8 (Other specified cerebrovascular diseases, $p = 1.05 \times 10^{-23}$), and I51.7 (Cardiomegaly, $p = 6.40 \times 10^{-23}$) were among the top 10 phenotypes highly enriched in the samples with both conditions. The results demonstrate that epilepsy and congenital heart disease may have some side effects, which predispose individuals to a high risk of stroke, hemiplegia, and other cerebrovascular diseases (Table S3).

Replication study using Mount Sinai BioMe BioBank whole exomes

We replicated the gene-level association analyses using high-impact rare variants filtered from genetically identified Europeans in the Mount Sinai BioMe Biobank. Following the same process described above for the UK Biobank samples, we obtained 184 European CHD cases, 286 European EP cases, and 2000 controls from the Mount

Sinai BioMe Biobank (Methods). Then, we ran gene-level SKAT-O analyses on aggregated variant sets for EP and CHD. Of the 813 genes with nominal significance of $p = 0.05$ for EP association using the UK Biobank, 38 genes were replicated in an EP association test using BioMe Biobank samples. According to the sorted combined P values by Cauchy combination test (Table S4, S5 and S6), the gene *SIX3* remained the top gene associated with EP in both studies. The second most significant gene was *EPHB4,* mutations of which give rise to arteriovenous malformations in the central nervous system, a causal factor for EP (Vivanti et al. 2018). BioMe replicated 25 of 660 CHD-associated genes ($p < 0.05$) originally identified by means of the UK Biobank samples. The most significant CHD-associated gene replicated was *DENND2A* (Table S6). In HGMD Professional, the phenotypes listed as being associated with *DENND2A* were 'congenital heart disease' and 'intellectual disability'. It has been reported that mutations in *DENND2A* are associated with multiple phenotypes, including ischemic stroke, Parkinson's disease, and non-syndromic intellectual disability (Lang et al. 2019; Mosallaei et al. 2022). However, we did not find any overlapping genes between the UK Biobank and BioMe BioBank results for genes associated with both EP and CHD (likely due to phenotyping and genotyping differences between the biobanks, as well as sample sizes; see Methods). According to the combined P values from the BioMe BioBank results, the top gene associated with both EP and CHD is *KIAA1755*, a locus that has been found to be strongly associated with heart rate in a large-scale meta-analysis (den Hoed et al. 2013) as well as in other association studies (van den Berg et al. 2017).

Pathway analyses of top association genes

We performed pathway analyses on the top genes identified by the gene-level analyses above ($p < 0.01$) for EP and CHD. There were 5 genes in common between the two candidate gene lists (*CACNA2D1*, *H3-4*, *SCYL2*, *CATSPER2*, and *AHNAK*). We investigated the enrichment of candidate genes in both biological terms and Gene Ontology (GO) terms using Ingenuity Pathway Analysis (IPA) software (http://www.qiagen.com/ingenuity) and the InnateDB data analysis tool (Breuer et al. 2013). The overlap between the canonical pathways and the GO terms of EP and CHD was found to be highly significant ($p < 2.2e\text{-}16$) by Fisher's exact test. Using IPA, we identified four canonical pathways ($p < 0.05$) that were statistically significant for both EP and CHD (Fig. 4), namely the dilated cardiomyopathy signaling pathway, the role of NFAT (Nuclear factor of activated T-cells) in cardiac hypertrophy, GPCR-mediated nutrient sensing in enteroendocrine cells, and nNOS signaling in skeletal muscle cells (Table S7 and S8). These pathways

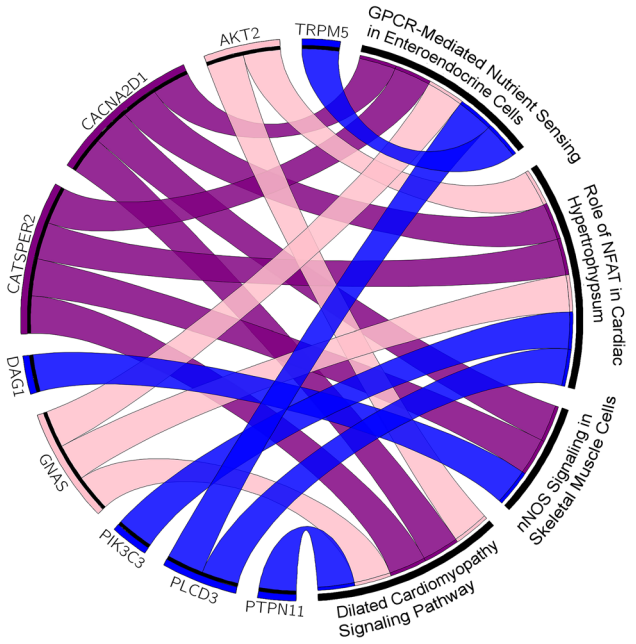

Fig. 4 Shared pathways between EP and CHD candidate genes. Circos map illustrating 4 biological pathways, identified by the IPA tool, which were common to both EP and CHD candidate genes. Among the 9 genes depicted, 2 (*CACNA2D1* and *CATSPER2*) were found to be candidate genes for both EP and CHD (shown in purple), and 2 (*AKT2* and *GNAS*) were candidate genes for EP only (shown in pink), whereas the remaining 5 were found to be candidate genes for CHD only (shown in blue)

involved some of the common candidate genes. For example, the dilated cardiomyopathy signaling pathway was found to be enriched in CHD by virtue of the genes *CACNA2D1*, *CATSPER2*, and *PTPN11*, and enriched in EP through the genes *AKT2*, *CACNA2D1*, *CATSPER2*, and *GNAS*. The role of NFAT in the cardiac hypertrophy pathway involved the candidate genes *CACNA2D1*, *CATSPER2*, *PIK3C3*, and *PLCD3* in CHD, and the candidate genes *AKT2*, *CACNA2D1*, *CATSPER2*, and *GNAS* in EP. Between EP and CHD, there were 12 over-represented ($p < 0.05$) GO terms in common (Table S9 and S10).

Voltage-gated calcium channel activity was enriched in both EP and CHD and involved two shared candidate genes: *CACNA2D1* and *CATSPER2*. To further analyze possible shared biological pathways between EP and CHD, we repeated the enrichment analyses on 272 candidate genes, acquired by virtue of their significant association with both EP and CHD ($p < 0.01$). InnateDB pathway analysis revealed statistical enrichment in negative regulation of the PI3K/AKT network (adjusted $p = 0.04$) (Table S11). This pathway involved the EP candidate genes *AKT2* and *TRIB3*, and the CHD candidate gene *PHLPP2*. The PI3K/AKT pathway has previously been associated with both epilepsy (Jansen et al. 2015) and CHD (Yu et al. 2022). The

top 3 associated diseases and functions of the EP candidate genes were cancer, organismal injury and abnormalities, and neurological disease (Fig. S3), whereas the top 3 associated diseases and functions of the CHD candidate genes were cancer, organismal injury and abnormalities, and gastrointestinal disease; neurological disease ranked 4th in CHD-associated disease. Owing to the locus heterogeneity and the broad disease spectrum covering almost all human tissues, we found that EP and CHD candidate genes are significantly involved in these two disease categories. Moreover, the results of the IPA analyses indicated that the EP genes identified from the SKAT-O analyses are related to neurological disease.

Prioritizing candidate genes using the human gene connectome

To ascertain the relevance of our gene association candidate genes to EP and CHD, we calculated their biological distance to known CHD genes and EP genes obtained from HGMD Professional (Stenson et al. 2020). The biological distance was measured by the Human Gene Connectome (HGC) (Itan et al. 2013) which provides a metric for biological relatedness between any two human genes based on protein–protein interaction data. The candidate gene biologically closest to known EP genes was *MAPK3*, with an average distance of 9.68, while the mean distance between random genes to known EP genes was 22.24. The closest candidate gene to known CHD genes was *RPS27A* with a distance of 7.35, whereas the average distance between random genes to known CHD genes was 20.50. The gene having the shortest distance to both EP and CHD genes on average was *MED1*, with a distance of 10.47 (Table S12). *MED1*, also known as TRAP220, is known to play critical roles in both cardiac development and epilepsy. Thus, a decreased level of TRAP220 protein is known to be involved in the pathophysiology of epilepsy and may be associated with neurological impairment as a consequence of frequent seizures (Li et al. 2006). Moreover, the TRAP220 protein has been shown to be required for cardiac development (Landles et al. 2003).

Gene-level PheWAS detection of traits associated with both EP and CHD genes

Both EP and CHD patients exhibit an increased risk of other diseases (Lui et al. 2017; Van Den Broek et al. 2004). We performed gene-level PheWAS on EP and CHD candidate genes to identify shared clinical symptoms, as well as related genes using the genes identified in our SKAT-O tests (Methods). All gene–trait pairs reaching the threshold of $p < 10^{-5}$ in gene-level PheWAS were employed to identify the common traits associated with both EP genes and CHD genes. The significance of the associations with both EP and CHD was measured by the mean value of log10 converted *P* values (-log10(*P* value)) of gene–trait pairs. From the gene-level association analyses, 'Femoral hernia', 'Occlusion and stenosis of precerebral arteries', and 'Ulcerative colitis' were the first three clinical phenotypes associated with EP genes and CHD genes in descending order of their average P values. Specifically, 'Femoral hernia' (ICD-10: K41) exhibited the highest average level of significance associated with both CHD genes and EP genes ($p = 3.7 \times 10^{-6}$ for *MTUS2* as a CHD gene, $p = 7.6 \times 10^{-5}$ for *PDE8A* as an EP gene). 'Occlusion and stenosis of precerebral arteries' (ICD-10: I65) was the second most significant trait associated with both EP and CHD ($p = 3.7 \times 10^{-6}$ *SULT1E1* as a CHD gene; $p = 7.6 \times 10^{-5}$ for *AURKC* as an EP gene). The third most significant trait was 'Ulcerative colitis' (ICD-10: K51) which has a $p = 7.8 \times 10^{-5}$ associated with the EP gene *CRISPLD1* and $p = 9.6 \times 10^{-5}$ associated with the CHD gene, *MUC13*. Few previous studies have linked hernia to either congenital heart disease or epilepsy; although some studies have reported that there is an increased incidence of congenital cardiac abnormalities among patients with inguinal hernia, the exploration of the underlying connections require further study (Oztürk et al. 2005). Ulcerative colitis is a known comorbidity of EP, with growing evidence, indicating that inflammatory bowel disease is an increased risk factor for neurological complications, epilepsy, seizures, stroke, headache, etc. (Akhan et al. 2002; Morís 2014). The occlusion and stenosis of precerebral arteries may affect the oxygen level of brain tissue by reducing brain oxygen delivery. It is also a proxy indicator of atherosclerosis, and a risk factor for cerebrovascular disease, which is also known to cause epilepsy.

Discussion

In this study, using the whole exomes of genetically identified Europeans from the UK Biobank, we performed variant- and gene-level association analyses for epilepsy and congenital heart disease to explore the potential shared genetic risk factors attributed to high-impact common and rare variants. First, we identified the 95% of the UK Biobank whole exomes that are of European origin, having a CEU fraction ≥ 0.8. Since our case sample size was not huge, especially in relation to congenital heart disease, we excluded non-Europeans from our study to minimize the effect of population stratification on case–control tests. At the variant level, we identified five variants associated with EP and CHD pathogenesis, two of which had high odds ratios (rs148994715, $OR_{EP} = 2.78$, $OR_{CHD} = 5.38$; rs57732048, $OR_{EP} = 3.49$, $OR_{CHD} = 4.60$), whereas the others had modest effect sizes. Both variants had higher

allele frequencies among EP patients, CHD patients, or patients having both compared to the controls without either EP or CHD. Using the precalculated LD information results from the 1,000 Genomes Project's European samples, we did not find any other variants in strong LD with either of the two variants, indicating that they are likely to exert independent effects on EP and CHD (Fig. S4 and S5). The candidate genes *WBP11* and *CACNA1I* have been previously implicated in cardiac malformation and response to neuroprotective drugs, indicating *WBP11* and *CACNA1I* to be potential therapeutic drug targets. Moreover, *CACNA1I* is involved in calcium signaling, a critical pathway where dysfunction can lead to both CHD and NDD. At a gene level, we performed SKAT-O analyses on the SNP sets aggregated from the high-impact rare variants. We identified 5 common genes associated with both EP and CHD. One of these five genes, *CACNA2D1*, is also the most significant gene among the CHD-associated genes. Interestingly, consistent with previous variant-level results that calcium signaling genes may play crucial roles in CHD- and EP-related disorders, *CACNA2D1* encodes a voltage-gated calcium channel. From the replication study using Mount Sinai BioMe Biobank whole exome samples, *SIX3* and *DENND2A* were identified as the most significant genes associated with EP and CHD, respectively, assessed by combining P values from the discovery and replication phases. The genes associated with both EP and CHD have not been replicated, likely due to the phenotypic variance between the two datasets (different phenotyping standards operate in the United States and the United Kingdom), different ages between the biobanks (average ages of cases in datasets: EP UK Biobank 57.08, EP BioMe Biobank 38.98; CHD UK Biobank 57.45, CHD BioMe Biobank 52.48), different sequencing kits resulting in variable genotype coverage, the heterogeneity of sequencing data across the two biobanks, and smaller sample sizes in the BioMe BioBank resulting in reduced association power. Although the variant-level and gene-level association analyses indicate that EP and CHD are likely to share some disease genes, particularly those that encode voltage-dependent calcium channels, we did not obtain any significant results that passed Bonferroni-adjusted cutoffs. Therefore, we additionally performed pathway analyses on the top genes from gene-level association analyses. Our pathway analyses found that *CACNA2D1* and *CATSPER2* were both plausible candidates, since both have roles in pathways shared between EP and CHD. Genes in the PI3K/AKT pathway were found to be related to the comorbidity of EP and CHD. These results indicate that either the observed EP genes or CHD genes are related to neurological disease, thereby providing further supporting evidence for the shared risk factors between EP and CHD.

We performed gene-level PheWAS to estimate the extent of EP and CHD potential comorbidity. Of the three phenotypes which are simultaneously linked to EP and CHD, we found that 'Occlusion and stenosis of precerebral arteries' and 'ulcerative colitis' are relevant to both disease entities. Occlusion and stenosis of precerebral arteries can decrease the oxygen supply to the brain, which is a known trigger for epilepsy. Ulcerative colitis, as a major form of inflammatory bowel disease, is even more interesting. It is consistent with our pathway analyses, which indicate that gastrointestinal disease shares some risk genes with EP and CHD. It also emphasizes the function of the 'brain-gut axis' which may be implicated in the comorbidity of the two disease entities.

Our study of EP and CHD shared genetics has several limitations. First, we have a relatively small number of cases for each disease entity, particularly in relation to CHD (333 samples) and those samples from individuals having both conditions (12 samples). Second, since EP and CHD are well-known disease states having a broad spectrum of variable sub-phenotypes, it is wholly possible that each sub-phenotype has independent risk variants/genes. However, the UK Biobank diagnosis code only provides major phenotypes. Therefore, we analyzed the aggregations of sub-phenotypes (e.g., since both "intractable epilepsy" (G40.40) and "not intractable epilepsy" (G40.41) were labeled as G40.4, we were unable to analyze the genetics specific to samples of "intractable epilepsy"). Third, the recruitment of UK Biobank samples was largely confined to people aged 40–69 years, implying that some extremely deleterious variants, which may cause life-threatening/life-shortening conditions, will not have been represented in our current study owing to the early death of the carriers. Nevertheless, we still identified relatively strong genetic associations for both EP and CHD as well as the common genetic factors between them.

Although shared EP and CHD loci were identified at the variant and the gene level, they may contribute to EP and CHD via different mechanisms within the cardiovascular system and the brain system, respectively. Therefore, further future validation with larger sample sizes and functional experimentation will be necessary to enhance our current study.

Conclusions

In summary, to explore the genetic factors underlying the comorbidity of epilepsy and congenital heart disease, we conducted variant-, gene-, and pathway-level analyses on the whole exomes of European patients in relation to EP and CHD in the UK Biobank. We determined that voltage-gated calcium channels and their genes are likely to be involved in the co-occurrence of epilepsy and congenital heart disease.

This study may attract the attention of both researchers and clinicians with a view to reconsidering the relationships of the pathogenicity in, as well as therapies for, EP and CHD. Future studies allowing larger sample sizes of more accurate sub-phenotypes are now required for the new discovery, association refinements and elucidation of the comorbidities.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-022-02502-4.

Acknowledgements The authors would like to thank Ron Do and Ghislain Rocheleau for kindly providing guidance on statistical methods for pleiotropic variants.

Author contributions Conceptualization: LC, YW, BD, and SH. Methodology: YW, CSB, and YI. Data management and investigation: YW, YI, and PDS. Visualization: YW and CSB. Writing—original draft: YW, BD, and CSB. Writing—review and editing: LC, YI, and DNC.

Funding This study was supported by the fund of the Sichuan University Innovation Spark under Project No. 2018SCUH0059.

Availability of data and materials This study was conducted using the UK Biobank Resource under Application 53074. The whole exome datasets are accessible upon approval of application. The BioMe biobank whole exome datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Akhan G, Andermann F, Gotman MJ (2002) Ulcerative colitis, status epilepticus and intractable temporal seizures. Epileptic Disord 4:135–137

Argyraki M, Damdimopoulou P, Chatzimeletiou K, Grimbizis GF, Tarlatzis BC, Syrrou M, Lambropoulos A (2019) In-utero stress and mode of conception: impact on regulation of imprinted genes, fetal development and future health. Hum Reprod Update 25:777–801. https://doi.org/10.1093/humupd/dmz025

Billett J, Cowie MR, Gatzoulis MA, Vonder Muhll IF, Majeed A (2008) Comorbidity, healthcare utilisation and process of care measures in patients with congenital heart disease in the UK: cross-sectional, population-based study with case-control analysis. Heart 94:1194–1199. https://doi.org/10.1136/hrt.2007.122671

Bond AR, Iacobazzi D, Abdul-Ghani S, Ghorbel M, Heesom K, Wilson M, Gillett C et al (2018) Changes in contractile protein expression are linked to ventricular stiffness in infants with pulmonary hypertension or right ventricular hypertrophy due to congenital heart disease. Open Heart 5:e000716. https://doi.org/10.1136/openhrt-2017-000716

Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL et al (2013) InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res 41:D1228–D1233. https://doi.org/10.1093/nar/gks1147

Carroll RJ, Bastarache L, Denny JC (2014) R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics 30:2375–2376. https://doi.org/10.1093/bioinformatics/btu197

Darshi M, Mendiola VL, Mackey MR, Murphy AN, Koller A, Perkins GA, Ellisman MH et al (2011) ChChd3, an inner mitochondrial membrane protein, is essential for maintaining crista integrity and mitochondrial function. J Biol Chem 286:2918–2932. https://doi.org/10.1074/jbc.M110.171975

Davis KW, Serrano M, Loddo S, Robinson C, Alesi V, Dallapiccola B, Novelli A et al (2019) Parent-of-origin effects in 15q11.2 BP1-BP2 microdeletion (Burnside-Butler) syndrome. Int J Mol Sci. https://doi.org/10.3390/ijms20061459

den Hoed M, Eijgelsheim M, Esko T, Brundel BJ, Peal DS, Evans DM, Nolte IM et al (2013) Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. Nat Genet 45:621–631. https://doi.org/10.1038/ng.2610

Desnous B, Lenoir M, Doussau A, Marandyuk B, Beaulieu-Genest L, Poirier N, Carmant L et al (2019) Epilepsy and seizures in children with congenital heart disease: a prospective study. Seizure 64:50–53. https://doi.org/10.1016/j.seizure.2018.11.011

El Ghaleb Y, Schneeberger PE, Fernández-Quintero ML, Geisler SM, Pelizzari S, Polstra AM, van Hagen JM et al (2021) CACNA1I gain-of-function mutations differentially affect channel gating and cause neurodevelopmental disorders. Brain 144:2092–2106. https://doi.org/10.1093/brain/awab101

Gonzalez-Sulser A (2020) Rodent genetic models of neurodevelopmental disorders and epilepsy. Eur J Paediatr Neurol 24:66–69. https://doi.org/10.1016/j.ejpn.2019.12.012

Hehr U, Pineda-Alvarez DE, Uyanik G, Hu P, Zhou N, Hehr A, Schell-Apacik C et al (2010) Heterozygous mutations in *SIX3* and *SHH* are associated with schizencephaly and further expand the clinical spectrum of holoprosencephaly. Hum Genet 127:555–561. https://doi.org/10.1007/s00439-010-0797-4

Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, DePalma SR et al (2015) De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science 350:1262–1266. https://doi.org/10.1126/science.aac9396

Itan Y, Zhang SY, Vogt G, Abhyankar A, Herman M, Nitschke P, Fried D et al (2013) The human gene connectome as a map of short cuts for morbid allele discovery. Proc Natl Acad Sci U S A 110:5558–5563. https://doi.org/10.1073/pnas.1218167110

Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, Scott E et al (2016) The mutation significance cutoff: gene-level thresholds for variant predictions. Nat Methods 13:109–110

Izarzugaza JMG, Ellesøe SG, Doganli C, Ehlers NS, Dalgaard MD, Audain E, Dombrowsky G et al (2020) Systems genetics analysis identifies calcium-signaling defects as novel cause of congenital heart disease. Genome Medicine 12:76. https://doi.org/10.1186/s13073-020-00772-z

Jansen LA, Mirzaa GM, Ishak GE, O'Roak BJ, Hiatt JB, Roden WH, Gunter SA et al (2015) PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia. Brain 138:1613–1628. https://doi.org/10.1093/brain/awv045

Kalantari S, Filges I (2020) "Kinesinopathies": emerging role of the kinesin family member genes in birth defects. J Med Genet 57:797–807. https://doi.org/10.1136/jmedgenet-2019-106769

Landles C, Chalk S, Steel JH, Rosewell I, Spencer-Dene B, el Lalani N, Parker MG (2003) The thyroid hormone receptor-associated protein TRAP220 is required at distinct embryonic stages in placental, cardiac, and hepatic development. Mol Endocrinol 17:2418–2435. https://doi.org/10.1210/me.2003-0097

Lang W, Wang J, Ma X, Zhang N, Li H, Cui P, Hao J (2019) Identification of shared genes between ischemic stroke and parkinson's disease using genome-wide association studies. Front Neurol 10:297. https://doi.org/10.3389/fneur.2019.00297

Leisner MZ, Madsen NL, Ostergaard JR, Woo JG, Marino BS, Olsen MS (2016) Congenital heart defects and risk of epilepsy: a population-based cohort study. Circulation 134:1689–1691. https://doi.org/10.1161/circulationaha.116.024538

Li JM, Wang XF, Xi ZQ, Gong Y, Liu FY, Sun JJ, Wu Y et al (2006) Decreased expression of thyroid receptor-associated protein 220 in temporal lobe tissue of patients with refractory epilepsy. Biochem Biophys Res Commun 348:1389–1397. https://doi.org/10.1016/j.bbrc.2006.08.010

Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, Gazzellone M et al (2011) Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med 3:95ra75. https://doi.org/10.1126/scitranslmed.3002464

Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X (2019) ACAT: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. Am J Hum Genet 104:410–421. https://doi.org/10.1016/j.ajhg.2019.01.002

Lui GK, Saidi A, Bhatt AB, Burchill LJ, Deen JF, Earing MG, Gewitz M et al (2017) Diagnosis and management of noncardiac complications in adults with congenital heart disease: a scientific statement from the american heart association. Circulation 136:e348–e392. https://doi.org/10.1161/CIR.0000000000000535

Liu Z, Wang F, Ma H, Xia H, Tian J, Sun T (2020) Amentoflavone induces cell cycle arrest, apoptosis, and autophagy in BV-2 cells. Front Biosci 25:798–816. https://doi.org/10.2741/4835

Maffucci P, Bigio B, Rapaport F, Cobat A, Borghesi A, Lopez M, Patin E et al (2019) Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis. Proc Natl Acad Sci U S A 116:950–959. https://doi.org/10.1073/pnas.1808403116

Mandalenakis Z, Rosengren A, Lappas G, Eriksson P, Hansson PO, Dellborg M (2016) Ischemic stroke in children and young adults with congenital heart disease. J Am Heart Assoc. https://doi.org/10.1161/jaha.115.003071

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Marino BS, Lipkin PH, Newburger JW, Peacock G, Gerdes M, Gaynor JW, Mussatto KA et al (2012) Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. Circulation 126:1143–1172. https://doi.org/10.1161/CIR.0b013e318265ee8a

Martin E, Enriquez A, Sparrow DB, Humphreys DT, McInerney-Leo AM, Leo PJ, Duncan EL et al (2020) Heterozygous loss of WBP11 function causes multiple congenital defects in humans and mice. Hum Mol Genet 29:3662–3678. https://doi.org/10.1093/hmg/ddaa258

Massin MM, Astadicko I, Dessy H (2007) Noncardiac comorbidities of congenital heart disease in children. Acta Paediatr 96:753–755. https://doi.org/10.1111/j.1651-2227.2007.00275.x

Matricardi S, Spalice A, Salpietro V, Di Rosa G, Balistreri MC, Grosso S, Parisi P et al (2016) Epilepsy in the setting of full trisomy 18: a multicenter study on 18 affected children with and without structural brain abnormalities. Am J Med Genet C Semin Med Genet 172:288–295. https://doi.org/10.1002/ajmg.c.31513

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P et al (2016) The ensembl variant effect predictor. Genome Biol 17:122. https://doi.org/10.1186/s13059-016-0974-4

Morales-Demori R (2017) Congenital heart disease and cardiac procedural outcomes in patients with trisomy 21 and Turner syndrome. Congenit Heart Dis 12:820–827. https://doi.org/10.1111/chd.12521

Morís G (2014) Inflammatory bowel disease: an increased risk factor for neurologic complications. World J Gastroenterol 20:1228–1237. https://doi.org/10.3748/wjg.v20.i5.1228

Mosallaei M, Ehtesham N, Beheshtian M, Khoshbakht S, Davarnia B, Kahrizi K, Najmabadi H (2022) Phenotype and genotype spectrum of variants in guanine nucleotide exchange factor genes in a broad cohort of Iranian patients. Mol Genet Genomic Med. https://doi.org/10.1002/mgg3.1894

Nicholl J, Waters W, Suwalski S, Brown S, Hull Y, Harbord MG, Entwistle J et al (2013) Epilepsy with cognitive deficit and autism spectrum disorders: prospective diagnosis by array CGH. Am J Med Genet B Neuropsychiatr Genet 162:24–35. https://doi.org/10.1002/ajmg.b.32114

Omichi C, Momose Y, Kitahara S (2010) Congenital long QT syndrome presenting with a history of epilepsy: misdiagnosis or relationship between channelopathies of the heart and brain? Epilepsia 51:289–292. https://doi.org/10.1111/j.1528-1167.2009.02267.x

Oztürk F, Tander B, Baysal K, Bernay F (2005) High association of congenital heart disease with indirect inguinal hernia. Pediatr Cardiol 26:80–82. https://doi.org/10.1007/s00246-004-0700-y

Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2018) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47:D886–D894. https://doi.org/10.1093/nar/gky1016

Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, Dippel D, Eesa M et al (2018) Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. Int J Stroke 13:612–632. https://doi.org/10.1177/1747493018778713

Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M et al (2020) The human gene mutation database (HGMD(®)): optimizing its use in a clinical diagnostic or research setting. Hum Genet 139:1197–1207. https://doi.org/10.1007/s00439-020-02199-3

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P et al (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12:e1001779. https://doi.org/10.1371/journal.pmed.1001779

Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, Sasson A, Wong E et al (2021) Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat Genet 53:942–948. https://doi.org/10.1038/s41588-021-00885-0

Tan ZP, Xie L, Deng Y, Chen JL, Zhang WZ, Wang J, Yang JF et al (2014) Whole-exome sequencing identifies Y1495X of *SCN5A* to be associated with familial conduction disease and sudden death. Sci Rep 4:5616. https://doi.org/10.1038/srep05616

van den Berg ME, Warren HR, Cabrera CP, Verweij N, Mifsud B, Haessler J, Bihlmeyer NA et al (2017) Discovery of novel heart rate-associated loci using the Exome Chip. Hum Mol Genet 26:2346–2363. https://doi.org/10.1093/hmg/ddx113

Van Den Broek M, Beghi E, Group R- (2004) Accidents in patients with epilepsy: types, circumstances, and complications: a European cohort study. Epilepsia 45:667–672. https://doi.org/10.1111/j.0013-9580.2004.33903.x

Vanlerberghe C, Petit F, Malan V, Vincent-Delorme C, Bouquillon S, Boute O, Holder-Espinasse M et al (2015) 15q11.2 microdeletion (BP1-BP2) and developmental delay, behaviour issues, epilepsy and congenital heart disease: a series of 52 patients. Eur J Med Genet 58:140–147. https://doi.org/10.1016/j.ejmg.2015.01.002

Vivanti A, Ozanne A, Grondin C, Saliou G, Quevarec L, Maurey H, Aubourg P et al (2018) Loss of function mutations in *EPHB4* are responsible for vein of Galen aneurysmal malformation. Brain 141:979–988. https://doi.org/10.1093/brain/awy020

Wojcik MH, Okada K, Prabhu SP, Nowakowski DW, Ramsey K, Balak C, Rangasamy S et al (2018) De novo variant in *KIF26B* is associated with pontocerebellar hypoplasia with infantile spinal muscular atrophy. Am J Med Genet A 176:2623–2629. https://doi.org/10.1002/ajmg.a.40493

Yamada Y, Nomura N, Yamada K, Matsuo M, Suzuki Y, Sameshima K, Kimura R et al (2014) The spectrum of *ZEB2* mutations causing the Mowat-Wilson syndrome in Japanese populations. Am J Med Genet A 164:1899–1908. https://doi.org/10.1002/ajmg.a.36551

Yu B, Yao S, Liu L, Li H, Zhu J, Li M, Han S et al (2022) The role of polypeptide PDTLN1 in suppression of PI3K/AKT signaling causes cardiogenetic disorders in vitro and *in vivo*. Life Sci 289:120244. https://doi.org/10.1016/j.lfs.2021.120244

Yun T, Li H, Chang P-C, Lin MF, Carroll A, McLean CY (2021) Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics 36:5582–5589. https://doi.org/10.1093/bioinformatics/btaa1081