A founder event causing a dominant childhood epilepsy survives 800 years through weak selective pressure

Bronwyn E. Grinton,[1] Erandee Robertson,[2,3] Liam G. Fearnley,[2,3] Ingrid E. Scheffer,[1,4-6] Anthony G. Marson,[7] Terence J. O'Brien,[8-11] W. Owen Pickrell,[12,13] Mark I. Rees,[12,14] Sanjay M. Sisodiya,[15,16] David J. Balding,[17] Mark F. Bennett,[1-3] Melanie Bahlo,[2,3] Samuel F. Berkovic,[1,*] Karen L. Oliver[1-3]

1. Epilepsy Research Centre, Department of Medicine, Austin Health, University of Melbourne, Heidelberg, Victoria, 3084, Australia.
2. Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, 3052, Australia.
3. Department of Medical Biology, University of Melbourne, Victoria, 3010, Australia
4. Department of Paediatrics, The University of Melbourne, Royal Children's Hospital, Parkville, Victoria, 3052, Australia.
5. Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, 3052, Australia.
6. Florey Institute of Neuroscience and Mental Health, Heidelberg, Victoria, 3084, Australia
7. Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, L69 3BX, United Kingdom.
8. Department of Neuroscience, Central Clinical School, Monash University, Melbourne, Victoria, 3004, Australia.

9. Department of Neurology, The Royal Melbourne Hospital, Parkville, Victoria, 3052, Australia.

10. Department of Neurology, Alfred Health, Melbourne, Victoria, 3004, Australia.

11. Department of Medicine, The University of Melbourne, Victoria, 3010, Australia.

12. Swansea University Medical School, Swansea University, Swansea, SA2 8PP, United Kingdom.

13. Department of Neurology, Morriston Hospital, Swansea Bay University Health Board, Swansea, SA2 8PP, United Kingdom.

14. Faculty of Medicine & Health, University of Sydney, New South Wales, 2006, Australia.

15. Chalfont Centre for Epilepsy, Buckinghamshire, HP11 2FZ, United Kingdom.

16. Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology, London, WC1N 3BG, United Kingdom.

17. Melbourne Integrative Genomics, School of BioSciences and School of Mathematics & Statistics, University of Melbourne, Parkville, Victoria, 3010, Australia.

* Correspondence: s.berkovic@unimelb.edu.au

Abstract

Genetic epilepsy with febrile seizures plus (GEFS+) is an autosomal dominant familial epilepsy syndrome characterised by distinctive phenotypic heterogeneity within families. The *SCN1B* c.363C>G (p.Cys121Trp) variant has been identified in independent, multi-generational families with GEFS+. Despite the variant being present in population databases (at very low frequency), there is strong clinical, genetic and functional evidence to support pathogenicity. Recurrent variants may be due to a founder event in which the variant has been inherited from a common ancestor. Here, we report evidence of a single founder event giving rise to the *SCN1B* c.363C>G variant in 14 independent families with epilepsy. A common haplotype was observed in all families, with the age of the most recent common ancestor estimated to be approximately 800 years ago. Analysis of UK Biobank whole exome sequencing data identified 74 individuals with the same variant. All individuals carried haplotypes matching the epilepsy families, suggesting all instances of the variant derive from a single mutational event. This unusual finding of a variant causing an autosomal dominant, early-onset disease in an outbred population that has persisted over many generations can be attributed to the relatively mild phenotype in most carriers and incomplete penetrance. Founder events are well established in autosomal recessive and late-onset disorders, but are rarely observed in early-onset, autosomal dominant diseases. These findings suggest variants present in the population at low frequencies should be considered potentially pathogenic in mild phenotypes with incomplete penetrance and may be more important contributors to the genetic landscape than previously thought.

Report

Genetic epilepsy with febrile seizures plus (GEFS+ [MIM: 604233]) is an autosomal dominant familial epilepsy syndrome characterised by distinctive phenotypic heterogeneity. Within families, a spectrum of phenotypes beginning in infancy or childhood is seen, ranging from mild (febrile seizures) to severe (epilepsy with myoclonic-atonic seizures and Dravet syndrome) disorders [1,2]. Phenocopies are frequently observed, particularly individuals with febrile seizures, which are a common childhood condition [3]. GEFS+ is genetically heterogeneous with pathogenic variants identified in several genes. The sodium channel beta-1 subunit gene, *SCN1B* (MIM: 600235), was the first gene associated with GEFS+ [4], with pathogenic variants in *SCN1B* accounting for approximately 8% of GEFS+ families [2]. Penetrance of *SCN1B* variants has been estimated at 62-76% based on large family studies [4,5].

The *SCN1B* c.363C>G (p.Cys121Trp) variant has been associated with epilepsy, and specifically GEFS+, in at least six independent, multi-generational families [2,4-7]. Despite being present in control databases, albeit at low frequency ($1.4e^{-5}$ (4 carriers) in gnomAD v2.1.1 [8], $1.1e^{-5}$ (6 carriers) in TopMed (https://bravo.sph.umich.edu/freeze8/hg38/variant/snv/19-35033654-C-G, accessed 07/07/2022)), the variant meets ACMG guidelines [9] for pathogenic/likely pathogenic classification with strong support coming from *in vitro*, *in vivo* and *in silico* data. ClinGen Gene Validity Evaluation determined there is moderate evidence to support the *SCN1B*-GEFS+ gene-disease relationship (https://search.clinicalgenome.org/kb/gene-validity/CGGV:assertion_91fb4b50-478b-4e3a-ac8a-a5cb7634a492-2022-01-04T211619.605Z, accessed 08/15/2022). Multiple studies in NaV β1 c.387C>G (p.Cys121Trp) heterozygous mice provide robust evidence of a deleterious gain of function effect [10-12].

A recurrent variant observed in unrelated individuals or families may result from the variant arising independently, or may be due to at least one founder event where individuals have inherited the variant from a common ancestor. Observation of the variant alone does not inform that distinction: if the variant is *de novo* in a family, or is present on independent haplotypes in different families, recurrent mutational events can be assumed. If the variant is present on a shared haplotype across different families, then a founder effect is likely. Whilst founder effects are well known for recessive disorders [13,14], and late-onset dominant conditions (e.g. Huntington's Disease [15], amyotrophic lateral sclerosis [16], breast/ovarian cancer [17]), pathogenic variants associated with *early-onset* dominant diseases rarely show founder effects, as they undergo purifying selection [18]. A practical clinical corollary of this is that such variants are characteristically absent from 'control' databases [19]. Indeed, this metric has been powerfully used as a filter to facilitate discovery of novel disease-causing variants [20,21].

Here we analyse a cohort of 14 independent families with the *SCN1B* c.363C>G variant for evidence of a common ancestor to determine whether this recurrent variant arose in a single common ancestor or by independent mutational events. We demonstrate evidence of a founder effect in an autosomal dominant, mild childhood onset epilepsy syndrome caused by the *SCN1B* c.363C>G variant in an outbred population.

Nine Australian families segregating the heterozygous *SCN1B* c.363C>G variant were identified from the Epilepsy Research Centre database (Table 1, Supplemental Figure S1). The families were recruited to a long-term study of the genetic basis of epilepsy and were originally identified because of the family history of seizures. Six families have been previously

reported [2,4-7]. An additional five individuals from the UK, the USA and Australia carrying the same variant were identified through the Epi25 Collaborative [20].

All individuals with the *SCN1B* c.363C>G variant included in this study with clinical data available had epilepsy phenotypes consistent with GEFS+ [1,2]. The family history of epilepsy, if present, was also consistent with GEFS+. The phenotypes of family members carrying the *SCN1B* variant ranged from unaffected at the mildest end of the spectrum to Dravet Syndrome at the most severe end (Table 1, Supplemental Figure S1). The most common phenotypes observed in the families were febrile seizures and febrile seizures plus. Twelve individuals known to carry the *SCN1B* c.363C>G variant had no known history of seizures. Limited clinical data was available for families M and N: The proband of family M was recorded as having non-acquired focal epilepsy, and the proband of family N was recorded only as “epilepsy”.

No genealogical link could be made between the families despite being able to trace five family pedigrees back to at least the early 1800s, although a common haplotype had previously been noted between families A and F [5]. Of the nine Australian families, six originated from the British Isles prior to immigrating to Australia (Table 1). The countries of origin of the remaining families were unknown but all were self-reported to be of white European origin.

Forty-four individuals were genotyped for the same 573,453 genome-wide SNPs (Supplemental methods). Genotyping quality was high with all samples achieving SNP

genotyping call rates >98%. Datasets were harmonised to the GRCh37/hg19 HRC reference genome and merged for quality control checks, ancestry and identity by descent analyses.

Principal component analysis performed with the 1000 Genomes Project Data set as the reference showed that all individuals clustered with European population samples, with the closest sub-populations being the British (GBR) and Utah residents (CEPH) with Northern and Western European ancestry (CEU) populations (Supplemental Figure S2).

Relatedness between all sample pairs both within and between families was estimated by Identity by descent (IBD) analyses using KING [22] and TRIBES [23]. We did not identify any cryptic relationships between individuals from different families, with up to 7th degree relatives reportedly detectable by TRIBES. Genetically inferred sex and within family relationships verified the accuracy of clinical and pedigree records.

To explore whether the variant may still have arisen from a common ancestor (i.e. a founder effect) despite lack of evidence to support interfamilial relationships, we reconstructed the chromosome 19 haplotype using phased SNP data either side of the *SCN1B* c.363C>G variant for each family. A core ancestral haplotype spanning approximately 260 kilobase pairs (kb), or 0.5 centiMorgans (cM), was shared by all 14 families at chr19q13.11 (chr19: 35,298,500 – 35,559,474 (hg19), Figure 1A and Supplemental Figure S3). Beyond the core 0.5 cM region, shared haplotypes between any two families ranged from approximately 630 kb to 6.9 megabase pairs (Mb) (Figure 1A). No alternate rare variants (i.e. putatively pathogenic) were detected along the ancestral haplotype and shared across families.

Haplotype age was estimated using an algorithm that predicts the age of the most recent common ancestor based on the lengths of shared regions utilizing recombination rates [24]. The age of the most recent common ancestor of these 14 families was estimated to be 31.2 generations (95% confidence interval 14.2-70.4 generations), or approximately 800 years (95% confidence interval 355 years-1760 years) assuming 25 years per generation.

Twenty-four SNPs representing the core 0.5 cM ancestral haplotype that are not in linkage disequilibrium ($r^2 \leq 0.1$; Supplemental Table S1) were used to determine the population frequency of the core ancestral haplotype in the 1000 Genomes Project European populations [25] via LDLink [26]. This established the haplotype frequency to be 0.7% in all European cohort samples and 1.3% in samples of British and Western European origin (GBR and CEU 1000 Genomes populations combined). This haplotype was not present in any of the non-European populations represented in the 1000 Genomes Project.

To further investigate the frequency of the *SCN1B* c.363C>G variant in the British population, whole exome sequencing and imputed SNP data from the UK Biobank [27] were accessed through project ID 36610. The c.363C>G variant was present in 0.039% of individuals in the European cohort of the UK Biobank with WES data available (74/188,663 unrelated individuals). There were no homozygous carriers. The 74 variant carriers were unrelated (kinship coefficient < $1/2^{(9/2)}$) as reported by the UK Biobank [27]. All 74 individuals self-reported their ethnic background as British, Irish or "any other white background". This equates to an allele frequency of $2.0e^{-4}$ (compared with the allele frequency of $1.4e^{-5}$ observed in gnomAD (v2.1.1) [8], Odds ratio 13.9, 95% CI 5.1-37.9). This difference in frequency may reflect the different distribution of ethnicities represented in the two databases, and suggest that the

variant is more concentrated in the British population. We note that the frequency of the variant observed in the UK Biobank data (0.039%) is far less than the frequency of the haplotype observed in the 1000 Genomes Project data (1.3%). This is consistent with the mutation event occurring on a haplotype already present multiple times within the population.

We evaluated the genetic risk of the variant by using clinical data from hospital and GP records from the UK Biobank. General practitioner (GP) and hospital records were available for 34 and 66 of the 74 variant carriers respectively and were examined for enrichment of seizure-related codes. All ICD and READ codes associated with any seizure phenotype were captured. The mean age of first hospital and GP records available were 54 and 28 years respectively. Five of the 74 individuals had relevant codes recorded: two referred to childhood epilepsy or convulsions (ascertained from GP records), while the remaining three indicated epilepsy or convulsions in the 6^{th}-8^{th} decades of life with limited data available (ascertained from hospital records).

The odds ratio (OR) estimate for the UK Biobank was low (OR 2.8, 95% CI 1.1-7.0) suggesting the variant was a weak/moderate risk allele but still higher than the current highest reported OR in GWAS for febrile seizures (OR 2.09 [28]). However, we have low confidence in this calculation and think it is a gross under-estimate, as childhood seizures in these individuals, aged over 40 years at the time of recruitment [27], will be greatly under-reported due to childhood seizures having been missed, forgotten or not captured in the medical records [29,30] . In addition, the frequency of any type of seizure captured in the UK Biobank records was 2.8%, compared with population estimates of convulsive disorders of 10% [31] and febrile seizures

were noted in only 0.05% (compared with repeated epidemiological estimates of 2-5% [3,32]). Further, the penetrance for seizures in our multigenerational families is approximately 70%. Thus, the absence of childhood data available for UK Biobank participants limits conclusions regarding the penetrance or effect size of the variant being drawn from this data.

Sixty-six UK Biobank variant carriers (89%) shared the full 0.5 cM core ancestral haplotype found in the 14 epilepsy families (Figure 1B, Supplemental methods). The remaining eight UK Biobank variant carriers shared smaller regions of the core haplotype ranging from 0.19 cM to 0.48 cM. All 74 individuals also shared segments of the extended haplotype with the original 14 families outside the core haplotype region (Figure 1B).

The identification of a shared chromosomal haplotype in 14 epilepsy families and 74 UK Biobank samples is consistent with a single common ancestral origin for the *SCN1B* c.363C>G variant. The most recent common ancestor of the 14 epilepsy families is estimated to have lived approximately 31 generations, or 800 years, ago. The age of this common founder suggests the variant arose prior to white settlement of Australia in the late 1700s. The additional finding that this variant is present in the European cohort of the UK Biobank at 14 times the frequency seen in the gnomAD database, together with the ancestral origins of the epilepsy families, supports a British origin of the variant, which subsequently spread through settlement of Australia and the USA following the migration of individuals from Britain.

While there are a small number of entries for this variant in public repositories (four entries in gnomAD [8] and 12 entries in ClinVar [33] (accessed 04/11/2022)), there are only a small number of other reports in the literature [34-36]. Limited clinical details and ethnicities regarding

these individuals are provided or not known. Given the frequency of this variant observed in the UK Biobank, it may be somewhat surprising that there are not more individuals or families reported. However, as carriers of this variant may be unaffected or have predominantly mild phenotypes, these individuals may be less likely to access genetic testing than those with more severe epilepsies and therefore remain undetected, or variants detected through clinical testing may not be reported in the literature.

There are no reported homozygous carriers of the *SCN1B* c.363C>G variant, and no homozygous haplotype carriers were observed in our epilepsy families or the UK Biobank variant carriers. However there are a growing number of reports of other *SCN1B* variants in the homozygous state [37]. The phenotype in these patients is usually more severe than the phenotypes seen in the c.363C>G families. Most of the homozygous patients were the product of consanguineous unions with two of the variants being recurrent within the same ethnic population, suggesting another possible founder effect. *SCN1B* is now recognized as one of a growing list of genes that can cause epilepsy under both autosomal dominant and recessive inheritance models. Similarly, the homologous NaV β1 variant modelled in mice (c.387C>G) has a febrile seizure phenotype in the heterozygous state [10] but an epileptic encephalopathy phenotype in homozygous animals [38].

The incomplete penetrance of the c.363C>G variant and its presence in population databases raises the interesting question of whether this is a risk allele rather than a Mendelian variant with incomplete penetrance. These two concepts lie at different ends of a spectrum of genetic variant effect sizes with limited guidance regarding how these terms should be applied and where the line should be drawn [39]. The ClinGen Low Penetrance/Risk Allele Working Group

currently propose using risk allele to refer to “variants with very low penetrance that do not manifest in a Mendelian pattern of inheritance”, while low or incomplete penetrance variants are conceptualized as those “that may have evidence of segregating in a Mendelian pattern” [40]. Large family studies facilitate this distinction, and indeed there are many examples of autosomal dominant epilepsies with incomplete penetrance [41-43]. With the most reliable penetrance estimates for the *SCN1B* c.363C>G variant being approximately 70%, coupled with the functional study evidence, the classification as a Mendelian “variant with incomplete penetrance” stands, however, we note that this is an evolving spectrum.

Founder variants have typically been identified in specific ethnic populations, populations with high rates of consanguinity or populations that have undergone significant bottleneck events. The majority of diseases caused by such variants are autosomal recessive or late-onset autosomal dominant disorders. While there are some examples of founder events in early-onset, dominant disorders in outbred populations [44-47], the relative rarity of such examples is presumably due to their negative effect on reproductive fitness. The finding here of a variant arising from a single founder event causing an early-onset disease in a highly outbred population across multiple continents is therefore of particular interest.

The characteristics of the *SCN1B* c.363C>G variant and the associated phenotype may explain why the variant has persisted in the population over many generations despite causing an autosomal dominant, childhood-onset disease. Although most new, non-synonymous variants that are deleterious to human health are rapidly eliminated from the population through purifying selection [48,49], variants associated with mild phenotypes (e.g. febrile seizures), those with incomplete penetrance or causing late-onset disease may be subject to

weaker selection and therefore reach relatively high population frequencies [18]. Conversely, variants leading to haploinsufficiency are subject to stronger selective pressure than variants with gain of function or dominant negative effect [49]. It is well established that *SCN1B* c.363C>G causes a gain of function [11,12] and most carriers present with the mild phenotype of febrile seizures, which typically resolve by six years of age, or are unaffected. These factors likely explain its continued presence in the population even after ~800 years.

Our results provide evidence of a common founder for the *SCN1B* c.363C>G (p.Cys121Trp) variant and estimate the time since the most recent common ancestor. These findings suggest variants present in the population at low frequencies should be considered potentially pathogenic in mild phenotypes with incomplete penetrance and may be more important than previously thought. Such variants may be erroneously dismissed in cohort studies or by clinical laboratories due to their population frequency.

Figure titles and legends

Figure 1: Representation of shared haplotype regions around the *SCN1B* locus

A) Shared haplotype regions identified in the 14 epilepsy families. Family identifiers from top to bottom are: M, D, C, I, L, E, N, A, K, J, B, H, G, F. Location of the *SCN1B* c.363C>G variant is shown by the dotted line. Dark orange represents the 0.5 cM core ancestral region shared by all 14 families. Yellow represents the regions shared by at least two epilepsy families.

B) Haplotype regions in 74 UK Biobank samples carrying the *SCN1B* c.363C>G variant that are shared with the 14 epilepsy families. Red arrows indicate that some individuals may share longer haplotypes than displayed in the plots.

Table 1: Clinical and demographic details of epilepsy families

Family ID (reference)	Number of individuals genotyped	Country: State/Region(s) of residence	Countries of origin (prior to immigration to Australia)	Individual phenotypes within families
A [1]	17	Australia: Tasmania	England, Ireland	FS, FS+, FS+ & Abs, MAE
B [2]	5	Australia: Tasmania	England, Ireland	FS, FS+
C	1	Australia: Victoria, New South Wales, Queensland	England, Ireland, Scotland, Switzerland	FS
D	1	Australia: Victoria	Unknown	FS, TLE
E [4]	7	Australia: Victoria	England	FS, FS+, TLE
F [5]	4	Australia: South Australia, Victoria, Queensland	England	FS, FS+, TLE
G [4]	2	Australia: Queensland, Northern Territory	England, Germany	FS, TLE

H [6]	1	Australia: New South Wales	Unknown	FS, DS
I	1	Australia: Victoria	Unknown	FS, FS+
J	1	Australia: Victoria	Unknown	FS+, Abs
K	1	UK: Wales	Not applicable	FS, FIA (possible temporal origin), TLE, NAFE
L	1	UK	Not applicable	FS, Myo, Abs, BCS Generalised and focal discharges on EEG
M	1	UK: North East	Not applicable	NAFE
N	1	USA: Texas	Not applicable	Unknown[#]

Abs: Absence; BCS: Bilateral convulsive seizures; DS: Dravet Syndrome; FIA: Focal impaired awareness; FAS: Focal aware; FS: Febrile seizures; FS+: Febrile seizures plus; MAE: Myoclonic-astatic Epilepsy; Myo: Myoclonic; NAFE: Non-acquired focal epilepsy; TLE: Temporal Lobe Epilepsy

[#]no ethical approval to obtain additional patient data

Acknowledgements

We thank the patients and their families for participating in this study. We are grateful for research support from Dr Alica M Goldman and Dr Slavé Petrovski.

Funding support was provided by the Epilepsy Society Tasmania and the Estate of Kathleen Beulah Grace; National Health and Medical Research Council (NHMRC) Program Grant 1091593 (SFB, IES); NHMRC Investigator Grants 1195236 (MB), 1196637 (SFB) and 1172897 (IES); Australian Government Research Training Program Scholarship APP533086 provided by the Australian Commonwealth Government and the University of Melbourne (KLO); DHB Foundation Centenary Postdoctoral Fellowship in Neurogenetic Systems Biology (LGF); Taking Flight Award from CURE Epilepsy (MFB); Epilepsy Society, UK (SMS). Funding for the UK-Wales samples was provided by the Health and Care Research Wales funding for the Wales Epilepsy Research Network (WOP/MIR).

This work was also supported by the Victorian Government's Operational Infrastructure Support Program and the NHMRC Independent Research Institute Infrastructure Support Scheme, and partly supported by the Brain Repair and Intracranial Neurotherapeutics Unit. Part of this work was undertaken at University College London Hospitals, which received a proportion of funding from the UK NIHR Biomedical Research Centres funding scheme.

We thank the Epi25 principal investigators, local staff from individual cohorts, and the patients with epilepsy who participated in Epi25 for making possible this global collaboration and resource to advance epilepsy genetics research.

This research was conducted using data from UK Biobank (www.ukbiobank.ac.uk), a major biomedical database, under data use agreement 36610 (PI Bahlo).

Additional acknowledgments are included in Supplemental data.

Declaration of interests

Ingrid Scheffer has served on scientific advisory boards for BioMarin, Chiesi, Eisai, Encoded Therapeutics, GlaxoSmithKline, Knopp Biosciences, Nutricia, Rogcon, Takeda Pharmaceuticals, UCB, Xenon Pharmaceuticals; has received speaker honoraria from GlaxoSmithKline, UCB, BioMarin, Biocodex, Chiesi, Liva Nova and Eisai; has received funding for travel from UCB, Biocodex, GlaxoSmithKline, Biomarin and Eisai; has served as an investigator for Anavex Life Sciences, Cerecin Inc, Cereval Therapeutics, Eisai, Encoded Therapeutics, EpiMinder Inc, Epygenyx, ES-Therapeutics, GW Pharma, Marinus, Neurocrine BioSciences, Ovid Therapeutics, Takeda Pharmaceuticals, UCB, Ultragenyx, Xenon Pharmaceutical, Zogenix and Zynerba; and has consulted for Atheneum Partners, Care Beyond Diagnosis, Epilepsy Consortium, Ovid Therapeutics, UCB and Zynerba Pharmaceuticals; and is a Non-Executive Director of Bellberry Ltd and a Director of the Australian Academy of Health and Medical Sciences and the Australian Council of Learned Academies Limited. She may accrue future revenue on pending patent WO61/010176 (filed: 2008): Therapeutic Compound; has a patent for SCN1A testing held by Bionomics Inc and licensed to various diagnostic companies; has a patent molecular diagnostic/therapeutic target for benign familial infantile epilepsy (BFIE) [PRRT2] 2011904493 & 2012900190 and PCT/AU2012/001321 (TECH ID:2012-009).

Web resources

GCTA, https://yanglab.westlake.edu.cn/software/gcta/

Genotype Harmonizer, https://github.com/molgenis/systemsgenetics/wiki/Genotype-Harmonizer

gnomAD, https://gnomad.broadinstitute.org/

LDLink, https://ldlink.nci.nih.gov/

Michigan Imputation Server, https://imputationserver.sph.umich.edu/

Mutation dating, https://shiny.wehi.edu.au/rafehi.h/mutation-dating/

OMIM, http://www.omim.org/

Plink 1.9, https://www.cog-genomics.org/plink/1.9/

SHAPEIT4, https://odelaneau.github.io/shapeit4/

TRIBES, https://github.com/aehrc/TRIBES

UK Biobank, https://www.ukbiobank.ac.uk/

Data and code availability

This research has been conducted using data from UK Biobank, a major biomedical database. The UK Biobank is an open access resource. To access the UKBB datasets, you need to register as a UKBB researcher (https://www.ukbiobank.ac.uk/enable-your-research/register). Additional genetic data used in this study is not available due to patient privacy and ethical restrictions. This study did not generate any code.

References

1. Scheffer, I.E., and Berkovic, S.F. (1997). Generalized epilepsy with febrile seizures plus. A genetic disorder with heterogeneous clinical phenotypes. Brain *120*, 479-490. 10.1093/brain/120.3.479.

2. Zhang, Y.H., Burgess, R., Malone, J.P., Glubb, G.C., Helbig, K.L., Vadlamudi, L., Kivity, S., Afawi, Z., Bleasel, A., Grattan-Smith, P., et al. (2017). Genetic epilepsy with febrile seizures plus: Refining the spectrum. Neurology *89*, 1210-1219. 10.1212/WNL.0000000000004384.

3. Verity, C.M., Butler, N.R., and Golding, J. (1985). Febrile convulsions in a national cohort followed up from birth. I--Prevalence and recurrence in the first five years of life. Br Med J (Clin Res Ed) *290*, 1307-1310. 10.1136/bmj.290.6478.1307.

4. Wallace, R.H., Wang, D.W., Singh, R., Scheffer, I.E., George, A.L., Jr., Phillips, H.A., Saar, K., Reis, A., Johnson, E.W., Sutherland, G.R., et al. (1998). Febrile seizures and generalized epilepsy associated with a mutation in the Na+-channel beta1 subunit gene *SCN1B*. Nat Genet *19*, 366-370. 10.1038/1252.

5. Wallace, R.H., Scheffer, I.E., Parasivam, G., Barnett, S., Wallace, G.B., Sutherland, G.R., Berkovic, S.F., and Mulley, J.C. (2002). Generalized epilepsy with febrile seizures plus: mutation of the sodium channel subunit *SCN1B*. Neurology *58*, 1426-1429. 10.1212/wnl.58.9.1426.

6. Carvill, G.L., Weckhuysen, S., McMahon, J.M., Hartmann, C., Moller, R.S., Hjalgrim, H., Cook, J., Geraghty, E., O'Roak, B.J., Petrou, S., et al. (2014). *GABRA1* and *STXBP1*: novel genetic causes of Dravet syndrome. Neurology *82*, 1245-1253. 10.1212/WNL.0000000000000291.

7. Scheffer, I.E., Harkin, L.A., Grinton, B.E., Dibbens, L.M., Turner, S.J., Zielinski, M.A., Xu, R., Jackson, G., Adams, J., Connellan, M., et al. (2007). Temporal lobe epilepsy and GEFS+ phenotypes associated with *SCN1B* mutations. Brain *130*, 100-109. 10.1093/brain/awl272.

8. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434-443. 10.1038/s41586-020-2308-7.

9. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405-424. 10.1038/gim.2015.30.

10. Wimmer, V.C., Reid, C.A., Mitchell, S., Richards, K.L., Scaf, B.B., Leaw, B.T., Hill, E.L., Royeck, M., Horstmann, M.T., Cromer, B.A., et al. (2010). Axon initial segment dysfunction in a mouse model of genetic epilepsy with febrile seizures plus. J Clin Invest *120*, 2661-2671. 10.1172/JCI42219.

11. Kruger, L.C., O'Malley, H.A., Hull, J.M., Kleeman, A., Patino, G.A., and Isom, L.L. (2016). beta1-C121W Is Down But Not Out: Epilepsy-Associated Scn1b-C121W Results in a Deleterious Gain-of-Function. J Neurosci *36*, 6213-6224. 10.1523/JNEUROSCI.0405-16.2016.

12. Hatch, R.J., Reid, C.A., and Petrou, S. (2014). Enhanced in vitro CA1 network activity in a sodium channel beta1(C121W) subunit model of genetic epilepsy. Epilepsia *55*, 601-608. 10.1111/epi.12568.

13. Morral, N., Bertranpetit, J., Estivill, X., Nunes, V., Casals, T., Gimenez, J., Reis, A., Varon-Mateeva, R., Macek, M., Jr., Kalaydjieva, L., and et al. (1994). The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nat Genet *7*, 169-175. 10.1038/ng0694-169.

14. Corbett, M.A., Schwake, M., Bahlo, M., Dibbens, L.M., Lin, M., Gandolfo, L.C., Vears, D.F., O'Sullivan, J.D., Robertson, T., Bayly, M.A., et al. (2011). A mutation in the Golgi Qb-SNARE gene *GOSR2* causes progressive myoclonus epilepsy with early ataxia. Am J Hum Genet *88*, 657-663. 10.1016/j.ajhg.2011.04.011.

15. Lee, J.M., Gillis, T., Mysore, J.S., Ramos, E.M., Myers, R.H., Hayden, M.R., Morrison, P.J., Nance, M., Ross, C.A., Margolis, R.L., et al. (2012). Common SNP-based haplotype analysis of the 4p16.3 Huntington disease gene region. Am J Hum Genet *90*, 434-444. 10.1016/j.ajhg.2012.01.005.

16. Henden, L., Twine, N.A., Szul, P., McCann, E.P., Nicholson, G.A., Rowe, D.B., Kiernan, M.C., Bauer, D.C., Blair, I.P., and Williams, K.L. (2020). Identity by descent analysis identifies founder events and links *SOD1* familial and sporadic ALS cases. NPJ Genom Med *5*, 32. 10.1038/s41525-020-00139-8.

17. Ferla, R., Calo, V., Cascio, S., Rinaldi, G., Badalamenti, G., Carreca, I., Surmacz, E., Colucci, G., Bazan, V., and Russo, A. (2007). Founder mutations in *BRCA1* and *BRCA2* genes. Ann Oncol *18 Suppl 6*, vi93-98. 10.1093/annonc/mdm234.

18. Blekhman, R., Man, O., Herrmann, L., Boyko, A.R., Indap, A., Kosiol, C., Bustamante, C.D., Teshima, K.M., and Przeworski, M. (2008). Natural selection on genes that underlie human disease susceptibility. Curr Biol *18*, 883-889. 10.1016/j.cub.2008.04.074.

19. Bennett, C.A., Petrovski, S., Oliver, K.L., and Berkovic, S.F. (2017). ExACtly zero or once: A clinically helpful guide to assessing genetic variants in mild epilepsies. Neurol Genet *3*, e163. 10.1212/NXG.0000000000000163.

20. Epi25 Collaborative (2019). Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. Am J Hum Genet *105*, 267-282. 10.1016/j.ajhg.2019.05.020.

21. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet *49*, 504-510. 10.1038/ng.3789.

22. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867-2873. 10.1093/bioinformatics/btq559.

23. Twine, N.A., Szul, P., Henden, L., McCann, E.P., Blair, I.P., Williams, K.L., and Bauer, D.C. (2019). TRIBES: A user-friendly pipeline for relatedness detection and disease gene discovery. bioRxiv, 686253. 10.1101/686253.

24. Gandolfo, L.C., Bahlo, M., and Speed, T.P. (2014). Dating rare mutations from small samples with dense marker data. Genetics *197*, 1315-1327. 10.1534/genetics.114.164616.

25. The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68-74. 10.1038/nature15393.

26. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of

possible functional variants. Bioinformatics *31*, 3555-3557. 10.1093/bioinformatics/btv402.

27. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203-209. 10.1038/s41586-018-0579-z.

28. Skotte, L., Fadista, J., Bybjerg-Grauholm, J., Appadurai, V., Hildebrand, M.S., Hansen, T.F., Banasik, K., Grove, J., Albinana, C., Geller, F., et al. (2022). Genome-wide association study of febrile seizures implicates fever response and neuronal excitability genes. Brain *145*, 555-568. 10.1093/brain/awab260.

29. Sillanpaa, M., Camfield, P.R., Camfield, C.S., Aromaa, M., Helenius, H., Rautava, P., and Hauser, W.A. (2008). Inconsistency between prospectively and retrospectively reported febrile seizures. Dev Med Child Neurol *50*, 25-28. 10.1111/j.1469-8749.2007.02006.x.

30. von Wirth, E., Mandler, J., Breuer, D., and Dopfner, M. (2021). The Accuracy of Retrospective Recall of Childhood ADHD: Results from a Longitudinal Study. J Psychopathol Behav *43*, 413-426. 10.1007/s10862-020-09852-1.

31. Hauser, W.A., Annegers, J.F., and Rocca, W.A. (1996). Descriptive epidemiology of epilepsy: contributions of population-based studies from Rochester, Minnesota. Mayo Clin Proc *71*, 576-586. 10.4065/71.6.576.

32. Nelson, K.B., and Ellenberg, J.H. (1976). Predictors of epilepsy in children who have experienced febrile seizures. N Engl J Med *295*, 1029-1033. 10.1056/NEJM197611042951901.

33. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res *44*, D862-868. 10.1093/nar/gkv1222.

34. Butler, K.M., da Silva, C., Alexander, J.J., Hegde, M., and Escayg, A. (2017). Diagnostic Yield From 339 Epilepsy Patients Screened on a Clinical Gene Panel. Pediatr Neurol *77*, 61-66. 10.1016/j.pediatrneurol.2017.09.003.

35. Haider, M.Z., Habeeb, Y., Al-Nakkas, E., Al-Anzi, H., Zaki, M., Al-Tawari, A., and Al-Bloushi, M. (2005). Lack of an association between candidate gene loci and idiopathic generalized epilepsy in Kuwaiti Arab children. J Biomed Sci *12*, 815-818. 10.1007/s11373-005-9009-y.

36. Maksemous, N., Sutherland, H.G., Smith, R.A., Haupt, L.M., and Griffiths, L.R. (2020). Comprehensive Exonic Sequencing of Known Ataxia Genes in Episodic Ataxia. Biomedicines *8*. 10.3390/biomedicines8050134.

37. Scala, M., Efthymiou, S., Sultan, T., De Waele, J., Panciroli, M., Salpietro, V., Maroofian, R., Striano, P., Van Petegem, F., Houlden, H., and Bosmans, F. (2021). Homozygous *SCN1B* variants causing early infantile epileptic encephalopathy 52 affect voltage-gated sodium channel function. Epilepsia *62*, e82-e87. 10.1111/epi.16913.

38. Reid, C.A., Leaw, B., Richards, K.L., Richardson, R., Wimmer, V., Yu, C., Hill-Yardin, E.L., Lerche, H., Scheffer, I.E., Berkovic, S.F., and Petrou, S. (2014). Reduced dendritic arborization and hyperexcitability of pyramidal neurons in a Scn1b-based model of Dravet syndrome. Brain *137*, 1701-1715. 10.1093/brain/awu077.

39. Senol-Cosar, O., Schmidt, R.J., Qian, E., Hoskinson, D., Mason-Suares, H., Funke, B., and Lebo, M.S. (2019). Considerations for clinical curation, classification, and reporting of low-penetrance and low effect size variants associated with disease risk. Genet Med *21*, 2765-2773. 10.1038/s41436-019-0560-8.

40. ClinGen Low Penetrance/Risk Allele Working Group (2020). Recommended terminology when describing variants with decreased penetrance for Mendelian conditions: A proposal by the ClinGen Low Penetrance/Risk Allele Working Group. 18 February 2020.

41. Chen, Y., Chen, D., Zhao, S., Liu, G., Li, H., and Wu, Z.Y. (2021). Penetrance estimation of *PRRT2* variants in paroxysmal kinesigenic dyskinesia and infantile convulsions. Front Med *15*, 877-886. 10.1007/s11684-021-0863-4.

42. Lal, D., Reinthaler, E.M., Schubert, J., Muhle, H., Riesch, E., Kluger, G., Jabbari, K., Kawalia, A., Baumel, C., Holthausen, H., et al. (2014). *DEPDC5* mutations in genetic focal epilepsies of childhood. Ann Neurol *75*, 788-792. 10.1002/ana.24127.

43. Wallace, R.H., Marini, C., Petrou, S., Harkin, L.A., Bowser, D.N., Panchal, R.G., Williams, D.A., Sutherland, G.R., Mulley, J.C., Scheffer, I.E., and Berkovic, S.F. (2001). Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet *28*, 49-52. 10.1038/ng0501-49.

44. Naito, T., Nishio, S.Y., Iwasa, Y., Yano, T., Kumakawa, K., Abe, S., Ishikawa, K., Kojima, H., Namba, A., Oshikawa, C., and Usami, S. (2013). Comprehensive genetic screening of *KCNQ4* in a large autosomal dominant nonsyndromic hearing loss cohort: genotype-phenotype correlations and a founder mutation. PLoS One *8*, e63231. 10.1371/journal.pone.0063231.

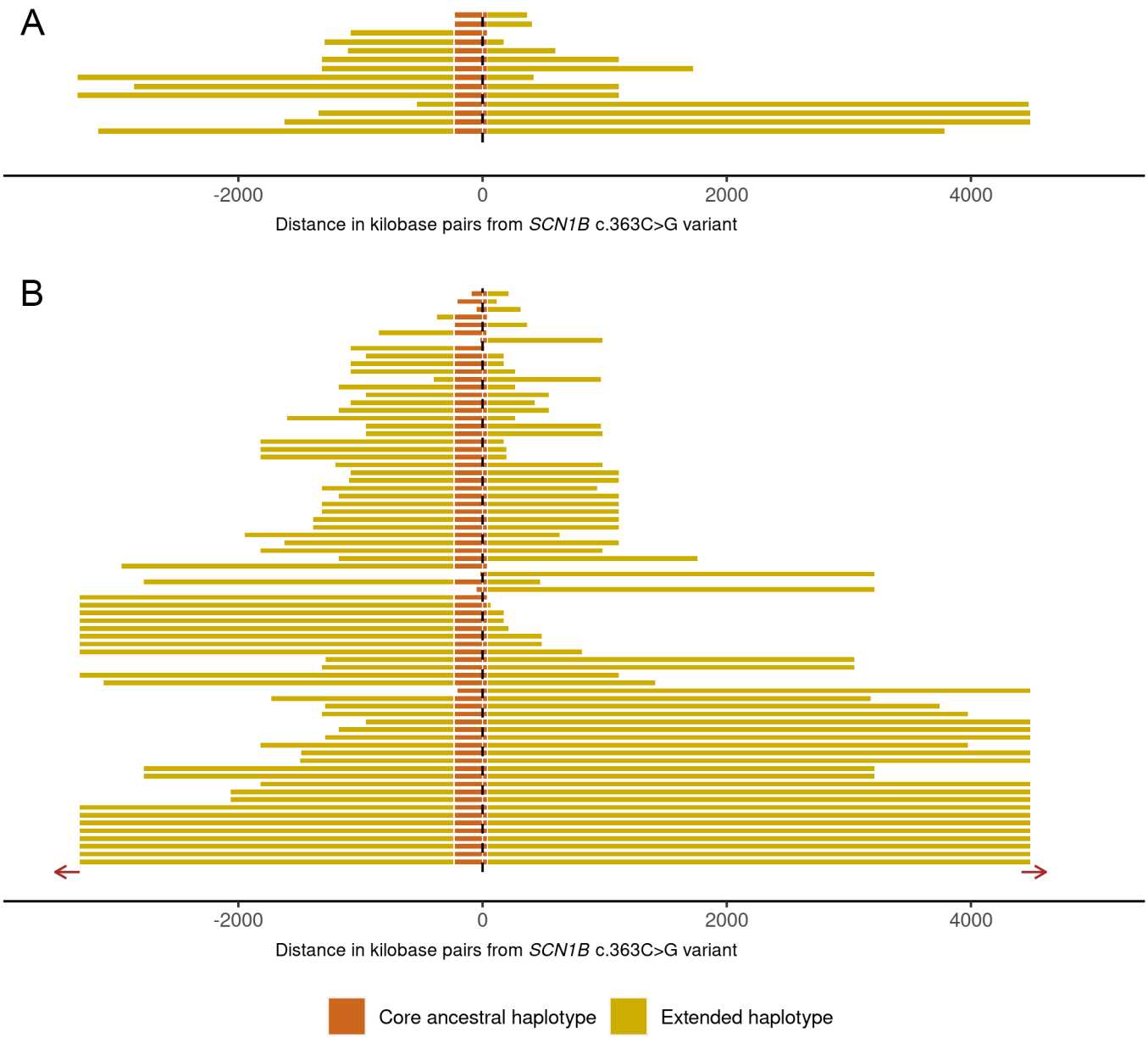
45. Nishio, A., Noguchi, Y., Sato, T., Naruse, T.K., Kimura, A., Takagi, A., and Kitamura, K. (2014). A *DFNA5* mutation identified in Japanese families with autosomal dominant hereditary hearing loss. Ann Hum Genet *78*, 83-91. 10.1111/ahg.12053.

46. Sambuughin, N., Yau, K.S., Olive, M., Duff, R.M., Bayarsaikhan, M., Lu, S., Gonzalez-Mera, L., Sivadorai, P., Nowak, K.J., Ravenscroft, G., et al. (2010). Dominant mutations in *KBTBD13*, a member of the BTB/Kelch family, cause nemaline myopathy with cores. Am J Hum Genet *87*, 842-847. 10.1016/j.ajhg.2010.10.020.

47. Salvatori, R., Radian, S., Diekmann, Y., Iacovazzo, D., David, A., Gabrovska, P., Grassi, G., Bussell, A.M., Stals, K., Weber, A., et al. (2017). In-frame seven amino-acid duplication in AIP arose over the last 3000 years, disrupts protein interaction and stability and is associated with gigantism. Eur J Endocrinol *177*, 257-266. 10.1530/EJE-17-0293.

48. Quintana-Murci, L. (2016). Understanding rare and common diseases in the context of human evolution. Genome Biol *17*, 225. 10.1186/s13059-016-1093-y.

49. Rapaport, F., Boisson, B., Gregor, A., Beziat, V., Boisson-Dupuis, S., Bustamante, J., Jouanguy, E., Puel, A., Rosain, J., Zhang, Q., et al. (2021). Negative selection on human genes underlying inborn errors depends on disease outcome and both the mode and mechanism of inheritance. Proc Natl Acad Sci U S A *118*. 10.1073/pnas.2001248118.


A
-2000
0
2000
4000
Distance in kilobase pairs from SCN1B c.363C>G variant
B
-2000
0
2000
4000
Distance in kilobase pairs from SCN1B c.363C>G variant
Core ancestral haplotype
Extended haplotype

Supplemental Figures

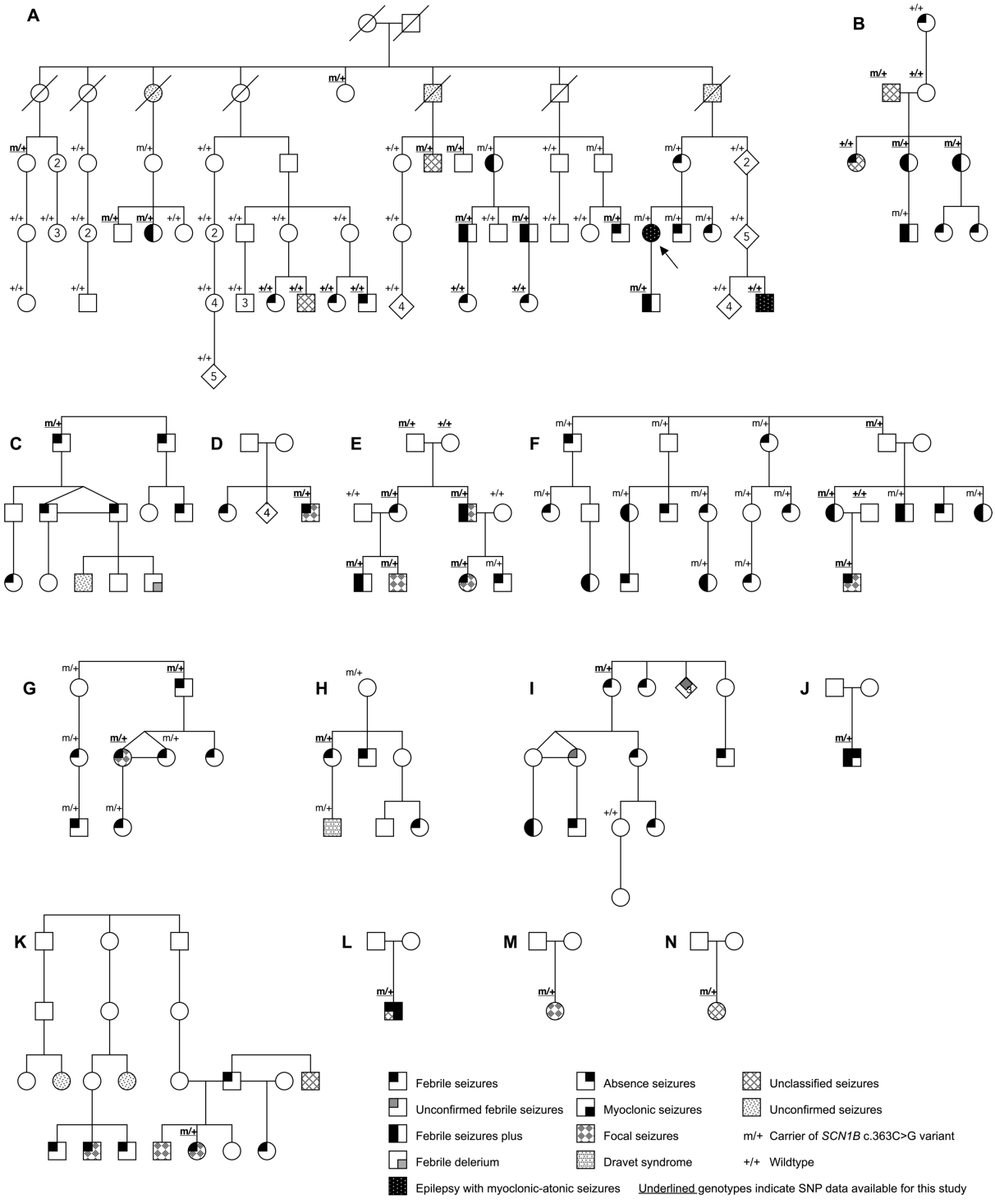


Figure S1: Pedigrees of multigenerational epilepsy families. See Table 1 for references for previously published families.

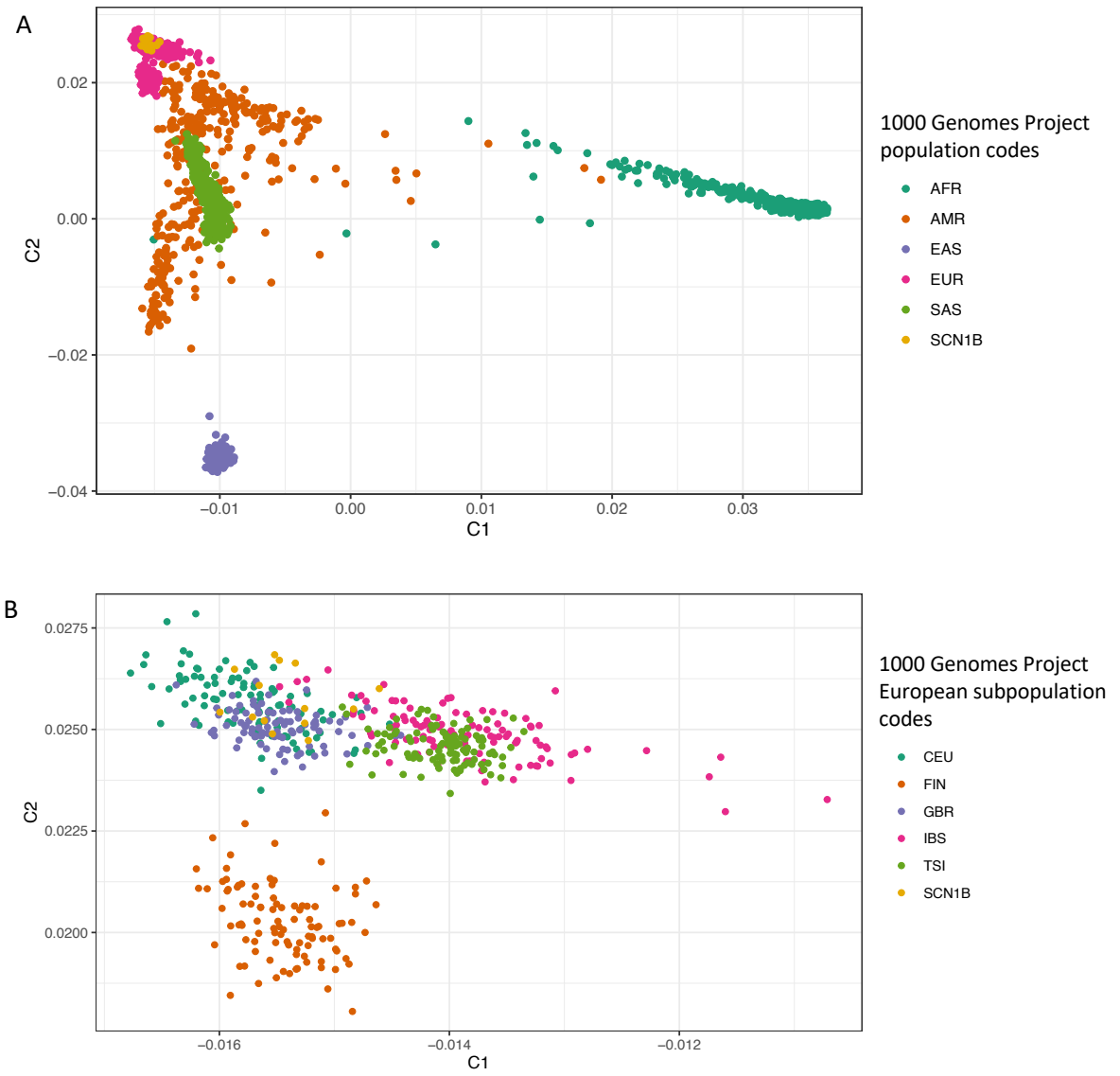


Figure S2: Principal component analysis comparing 14 epilepsy family cohort samples with the 1000 Genomes Project populations. A) *SCN1B* epilepsy family cohort samples (yellow) cluster with the European (EUR) population (pink). B) Within the European cohort, the *SCN1B* Epilepsy family cohort samples cluster most closely with the GBR and the CEU subpopulations.

AFR: African, AMR: Admixed American, CEU: Utah residents with Northern and Western European ancestry, EAS: East Asian, EUR: European, GBR: British in England and Scotland, IBS: Iberian populations in Spain, SAS: South Asian, TSI: Toscani in Italia, Fin: Finnish in Finland.

#CHROM	Physmap (hg19)	SNP	REF	ALT	Core ancestral haplotype
19	35,298,500	rs16969747	T	C	T
19	35,298,961	rs3746236	C	T	C
19	35,310,791	rs10414962	C	T	C
19	35,315,264	rs10439133	A	C	A
19	35,321,775	rs755733	G	T	G
19	35,328,604	rs77351121	T	C	T
19	35,329,265	rs10500276	T	G	T
19	35,329,919	rs17704752	T	C	T
19	35,337,427	rs71351794	T	C	T
19	35,339,966	rs749981	G	A	G
19	35,342,922	rs8106116	T	C	T
19	35,344,974	rs2546048	T	C	T
19	35,357,338	rs4806060	T	C	T
19	35,368,639	rs75107343	T	C	T
19	35,388,152	rs8113651	A	G	G
19	35,400,147	rs117738748	G	A	G
19	35,418,092	rs2651125	G	A	A
19	35,418,450	rs76499006	G	A	G
19	35,420,841	rs2546001	G	A	G
19	35,424,444	rs74413818	C	T	C
19	35,434,448	rs62122088	A	G	G
19	35,434,648	rs145945527	T	C	T
19	35,435,001	rs78688282	C	T	C
19	35,435,006	rs1345658	A	G	A
19	35,435,776	rs1053213	G	A	G
19	35,449,486	rs34687332	T	C	T
19	35,451,124	rs77771822	C	T	C
19	35,454,215	rs2546027	T	G	T
19	35,454,430	rs78090888	G	A	G
19	35,459,133	rs77981981	C	T	C
19	35,462,191	rs2546043	G	A	G
19	35,463,646	rs77396367	C	T	C
19	35,470,408	rs11084794	G	A	A
19	35,472,417	rs295771	C	T	C
19	35,473,576	rs7248217	A	G	A
19	35,479,913	rs786506	G	A	G
19	35,480,974	rs55930889	T	G	T
19	35,482,114	rs7247980	C	T	C
19	35,486,475	rs12610307	T	G	T
19	35,496,231	rs117868756	G	A	G
19	35,503,025	rs67813969	C	T	C
19	35,504,627	rs7251988	C	A	C
19	35,506,729	rs2290647	G	A	G
19	35,510,304	rs73038384	C	T	C
19	35,520,962	rs67242970	T	C	T
19	35,524,558	*SCN1B* c.363C>G			G
19	35,527,030	rs111437296	C	T	C
19	35,530,073	rs16969930	T	C	T
19	35,531,222	rs41275828	T	C	T
19	35,536,364	rs8107142	G	A	G
19	35,545,385	rs73597479	C	T	C
19	35,556,123	rs2305744	A	G	A
19	35,557,440	rs1688029	T	C	T
19	35,559,474	rs11671010	T	C	T

Figure S3: Core ancestral haplotype identified in 14 epilepsy families

Supplemental Tables

rs16969747	rs3746236	rs10414962	rs17704752	rs71351794	rs749981
rs75107343	rs117738748	rs76499006	rs2546001	rs145945527	rs78688282
rs1053213	rs78090888	rs2546043	rs11084794	rs55930889	rs117868756
rs67813969	rs73038384	rs67242970	rs111437296	rs41275828	rs2305744

Table S1: SNPs representing the core 0.5cM ancestral haplotype that are not in linkage disequilibrium ($r^2 \leq 0.1$)

Supplemental Materials and Methods

Genotyping and quality control

Individuals from the 14 epilepsy families were genotyped for the same 573,453 SNPs using the Illumina Global Screening Array-24 (San Diego, CA, USA) at three different centres (Broad Institute, USA; Australian Genome Research Facility, Australia; and Erasmus University Medical Centre, The Netherlands). Datasets were harmonised to the GRCh37/hg19 HRC reference genome using GenotypeHarmonizer [1] prior to merging using Plink v.1.9.3 [2] for quality control checks. The sex of each sample was genetically inferred and assessed for concordance with reported sex from clinical records. Principal components analysis to assess ancestry was performed with the 1000 Genomes Project Data set as the reference using the GCTA software (version 1.93.2) [3].

Chromosome 19 SNP data was phased using Eagle2 [4] with the Michigan Imputation server (Minimac4) [5]. Haplotypes were compared across families using graphical and analytical methods with the statistical programming software R (version 4.1.1).

Whole exome sequencing data filtering

Whole exome sequencing data was available from seven families. Variants within the core ancestral haplotype region that passed quality filtering and were present in gnomAD exomes less than 100 times were extracted and filtered for variants shared by all individuals.

UK Biobank variant and haplotype analysis

Whole exome sequencing data from the European cohort of the UK Biobank was scrutinized for the *SCN1B* c.363C>G variant. 1,263 imputed SNPs (version 3 of the UK Biobank imputed

SNP data) with imputation accuracy ≥0.9 across a 10.4cM region spanning the chromosome 19 *SCN1B* locus were phased using SHAPEIT4-4.2.2 to infer haplotypes in variant carriers. These haplotypes were compared with the haplotypes seen in the epilepsy families.

Ethics statement

This study was approved by the Austin Health Human Research Ethics Committee. Informed consent was obtained and archived from all participants or their legal guardian.

Supplemental Acknowledgements

The Epi25 Collaborative is part of the Centers for Common Disease Genomics (CCDG) program, funded by the National Human Genome Research Institute (NHGRI) and the National Heart, Lung, and Blood Institute (NHLBI). CCDG-funded Epi25 research activities at the Broad Institute, including genomic data generation in the Broad Genomics Platform, are supported by NHGRI grant UM1 HG008895 (PIs: Eric Lander, Stacey Gabriel, Mark Daly, Sekar Kathiresan). The Genome Sequencing Program efforts were also supported by NHGRI grant 5U01HG009088-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank the Stanley Center for Psychiatric Research at the Broad Institute for supporting the genomic data generation efforts.

Supplemental References

1. Deelen, P., Bonder, M.J., van der Velde, K.J., Westra, H.J., Winder, E., Hendriksen, D., Franke, L., and Swertz, M.A. (2014). Genotype harmonizer: automatic strand

alignment and format conversion for genotype data integration. BMC Res Notes *7*, 901.

2. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

3. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet *88*, 76-82.

4. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y, A.R., H, K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet *48*, 1443-1448.

5. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat Genet *48*, 1284-1287.