Application of Rare Variant Transmission-Disequilibrium Tests to Epileptic Encephalopathy Trio Sequence Data

Running title: Epileptic encephalopathies and recessive genotypes

Authors: Epi4K Consortium, EuroEPINOMICS-RES Consortium, and Epilepsy Phenome Genome Project

Corresponding author: Epi4K Consortium

701 West 168th St., New York, NY 10032

Epi4K@columbia.edu

Funding Support:

This work was supported by grants from the National Institute of Neurological Disorders and Stroke (The Epilepsy Phenome/Genome Project NS053998; Epi4K NS077364, NS077274, NS077303, and NS077276), The Andrew's Foundation, Finding a Cure for Epilepsy and Seizures, the Richard Thalheimer Philanthropic Fund, and the Eurocores program EuroEPINOMICS-RES of the European Science Foundation. The project received further support through grants from the Fund for Scientific Research Flanders (FWO); the Academy of Finland (141549); the Folkhälsan Research Foundation; the program "Investissements d'avenir" ANR-10-IAIHU-06; the Federal Ministry for Education and Research (IonNeurONet: 01GM1105), the German Research Foundation (DFG: HE5415/3-

1; Le1030/11-1; RO3396/2-1), the German Society for Epileptology (DGfE), the Foundation noepilep.; the Swiss National Science Foundation (SNF: 32EP30_136042/1); the Wellcome Trust (09805); intramural funds of the University of Kiel; the Popgen 2.0 network (P2N) through the German Ministry for Education and Research (01EY1103); the European Union through Seventh Framework Programme (FP7) under the project DESIRE (N602531). The project also received infrastructural support through the Institute of Clinical Molecular Biology in Kiel, supported in part by DFG Cluster of Excellence ''Inflammation at Interfaces'' and ''Future Ocean''. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. Orrin Devinsky, David B. Goldstein, Steve Petrou and Slavé Petrovski have interests in companies related to epilepsy precision medicine.

Abstract

The classic epileptic encephalopathies, including Infantile Spasms (IS) and Lennox-Gastaut Syndrome (LGS), are severe seizure disorders that usually arise sporadically. *De novo* mutations in genes mainly encoding ion channel and synaptic proteins have been found to account for over 15% of patients with IS or LGS. The contribution of autosomal recessive genetic variation, however, is less well understood. We implemented a rare variant transmission-disequilibrium test (TDT) to search for autosomal recessive epileptic encephalopathy genes in a cohort of 320 outbred patient-parent trios that were generally pre-screened for rare metabolic disorders. In the current sample, our rare variant transmission-disequilibrium test did not identify individual genes with significantly distorted transmission over expectation after correcting for the multiple tests. While the rare variant transmission-disequilibrium test did not find evidence of a role for autosomal recessive genes, our current sample is insufficiently powered to assess the overall role of autosomal recessive genotypes in an outbred epileptic encephalopathy population.

Keywords: Epileptic encephalpathy; infantile spasms; transmission-disequilibrium test (TDT); recessive genotypes

INTRODUCTION

Epileptic encephalopathies are severe and therapy-resistant epilepsies of childhood, which frequently lead to developmental delay and multiple associated medical issues. Infantile Spasms (IS) and Lennox-Gastaut Syndrome (LGS) represent two of the more common broad subtypes of epileptic encephalopathies. Many novel genes for epileptic encephalopathies have been discovered in the last five years, fueled by the access to whole-exome sequencing. In particular, exome sequencing has highlighted the important role of *de novo* mutations with current estimates suggesting that over 15% of classical epileptic encephalopathy cases are explained by a *de novo* mutation in an established epileptic encephalopathy gene.[1; 2] Up to a further 3% have been reported to be explained by likely pathogenic *de novo* copy number variants (CNVs).[3]

While the role of *de novo* genetic variation in epileptic encephalopathies is increasingly understood, the role of recessive genetic variation, outside of recessive neurometabolic disorders such as lysosomal disorders, amino acid or organic acid imbalances, congenital disorders of glycosylation, and some mitochondrial diseases, remains unclear. In our current study we systematically assessed autosomal recessive inheritance in 320 IS or LGS patient-parent trios who did not have a likely disease-causing *de novo* mutation among one of the established dominant epileptic encephalopathy genes.[1; 2] In general, the 320 cases studied here had already been intensively studied for neurometabolic disorders using biochemical assessments.

SUBJECTS and METHODS

Cohort

Three-hundred and twenty epileptic encephalopathy trios were recruited through multiple international consortia, including 57 IS or LGS trios unpublished in our earlier studies.[1; 2] Patients did not have a clearly identified metabolic or genetic cause for their epilepsy based on clinically available testing, which varied across institutions. This collection of 320 trios did not include: a) patients previously found to have a *de novo* mutation in an established dominant epileptic encephalopathy gene, and b) trios where exome sequencing was based on a lymphoblastoid cell line (LCL) source for at least one of the three family members. The overall cohort was not enriched for consanguineous parents. Only two parent pairs showed an identity-by-descent (IBD) > 0.125, both < 0.15, which is approximately equivalent to 3rd degree relatives.[4]

Among the 320 trios; two families reported multiple affected children. For one of these families both the proband and affected sibling were investigated through exome sequencing while for the second family only the proband and parents were studied. Sequencing methods used to generate the sequence data have been previously described.[1; 2]

Transmission Disequilibrium Tests

For the transmission test we used two approaches that we have previously introduced.[5; 6] First, we tested for an autosomal homozygous or compound heterozygous effect using coreTDT.[6] In computing the test, we selected loss-of-function and missense single nucleotide substitution variants (SNVs) found at a global population minor allele frequency

below five percent (MAF<0.05). The loss-of-function variants were defined as stop gain, stop lost, start lost and canonical splice acceptor and donor site variants. For the missense variants we used our in-house Analysis Tool for Annotated Variants (ATAV) platform to identify the possibly and probably damaging variants based on a maximum Polyphen-2 HumDiv and HumVar prediction score[7] of greater than 0.4333. This test was applied to each autosomal gene individually as well as collectively across a set of 99 autosomal recessive neurometabolic genes published by van Karnebeek and colleagues.[8]

Second, we tested for a general effect of inherited autosomal variation by using a rare variant TDT that uses information from an independent collection of population controls (6503 EVS[9] plus 1,303 IGM sequenced controls) to weight the contribution of variants to the final test statistic.[5] In this analysis, qualifying variants were defined using the same PolyPhen-2 thresholds as above and were again required to have a global MAF less than 5%. Given that population stratification can impact the power of the test but not the type I error, we restricted this second analysis to trios with European ancestry (n=286 trios).

RESULTS

We assessed the role of inherited rare variation using the population control-weighted rare-variant TDT.[5] This test was applied to each autosomal gene across 320 eligible trios. No gene reached exome-wide significance after correcting for the 17,816 consensus coding sequence (CCDS release 14) autosomal genes (adjusted $\alpha = 2.81 \times 10^{-6}$, Table 1). Though population stratification cannot affect the false positive rate of the test, it can affect the

power.[5] We also conducted an analysis that was restricted to the 286 trios of European ancestry. Again, no gene reached the exome-wide significance level (Table 1).

We then tested for the presence of a recessive effect in each autosomal gene across the 320 trios. After quality control, only 3,472 autosomal genes were found to have at least one informative family, i.e., contain qualifying variants within the gene and that could, potentially, lead to homozygous or compound heterozygous offspring. None of these 3,472 genes achieved significance after correcting for the number of genes tested (adjusted $\alpha = 1.44 \times 10^{-5}$). The 10 most significant genes are listed in Table 2.

To investigate whether there is any evidence of recessive neurometabolic involvement in this sample, we also applied the coreTDT to the 99 autosomal recessive neurometabolic genes,[8] looking for an enrichment of homozygous or compound heterozygous offspring across the entire gene set. No enrichment was found ($p = 0.51$).

A power simulation was conducted to evaluate the types of effects that we could exclude based on this analysis. In these simulations, we conditioned on the parental genotype information contained in this IS/LGS population sample and characterized the distribution of offspring genotypes given this information and the fact that the offspring is affected. This distribution is a function of the number of causal genes, for which the family is informative, which is related to the density of causal genes within the actual gene set, and the relative risk of the offspring developing disease given that they have two affected gene copies (Supplementary Methods). Offspring are randomly sampled from this distribution and the resulting dataset is analyzed via coreTDT.[6] Since only 54 families are informative for at least one of the 99 autosomal recessive neurometabolic genes, and only 20 genes

have at least one informative family, our analyses are effectively restricted to these 54 families and 20 genes. We vary the proportion of informative genes that are actually disease causal and the relative risk and identify combinations of these parameters that attain at least 80% power (Figure 1). As can be seen, even when the compound heterozygous or homozygous qualifying variants are fully penetrant, the causal gene proportion must be larger than 40% to attain 80% power. When the proportion of causal genes is larger, e.g., 80%, we will have high power to detect an effect even with a relatively low relative risk.

Using established standards to identify pathogenic recessive genotypes[10; 11], one trio was found to have inherited two *SPATA5* pathogenic variants in a compound heterozygous manner.[12] The proband's phenotype is consistent with the *SPATA5* disease literature, and both pathogenic variants (p.Tyr559* and p.Arg84Gln) have previously been described as pathogenic among patients with *SPATA5* encephalopathy.[12]

DISCUSSION

A number of rare recessive disorders can present with an epileptic encephalopathy, particularly neurometabolic disorders; the latter are generally identified by biochemical analyses of blood, urine or CSF. We performed a global, hypothesis-free test to assess the role of autosomal recessive genetic variation in 320 patients with classic epileptic encephalopathies undiagnosed with standard clinical workups. Our sample of patient-parent trios did not identify a genome-wide significant departure in the observed number of offspring with recessive genotypes from that expected for any specific gene, or among 99 genes compiled for autosomal recessive neurometabolic disorders.

Many classical recessive metabolic disorders are routinely identified through biochemical screening prior to research study enrollment. Within our sample of 320 trios we did not find any genetic neurometabolic disorders that were missed through the conventional biochemical screening. From a clinical perspective, we emphasize that evaluation for these treatable causes should continue to be pursued. We did, however, identify a single case among the 320 with a known pathogenic recessive genotype in *SPATA5*,[12] a recently described gene for a recessive condition characterized by seizures, microcephaly, intellectual disability, and hearing loss.

The role of various dominant epilepsy genes including *ALG13*, *CDKL5*, *DNM1, GABRB3*, *SCN1A*, *SCN2A*, and *STXBP1*, for epileptic encephalopathies was securely established through exome sequencing of 356 trios and subsequent genome-wide assessments for excess *de novo* mutations identified in individual genes.[1; 2] No single gene passes a comparable threshold among the 320 trios studied here when assessing autosomal recessive genotypes. We demonstrate that the current sample of 320 trios is insufficiently powered to appropriately estimate what overall contribution autosomal recessive epilepsy genes have on the epileptic encephalopathies. Using a similar approach, a recent study on 4,125 patient-parent trios with various developmental disorders identified two novel autosomal recessive disease genes exceeding genome-wide significance,[13] emphasizing the importance of acquiring larger numbers to more confidently interpret the current lack of signal for very rare genetic epilepsies with recessive inheritance. Large-scale collaborative initiatives like the Epilepsy Genetic Initiative (EGI) and the Epi25 effort will aid the efforts to analyze genomic data on this scale.

Acknowledgments

We are deeply grateful to the probands, their families, clinical research coordinators, and referring physicians for their participation and provision of phenotype data and DNA samples used in this study.

We thank the EPGP Administrative (C. Freyer, K. Fox, R. Fahlstrom, S. Cristofaro, and K. McGovern), and Bioinformatics Core (G. Nesbitt, K. McKenna, and V. Mays), staff at Coriell Institute – NINDS Genetics Repository, and members of the Institute for Genomic Medicine, Columbia University (P. Cansler, J. Charoensri, B. Copeland, S. Kamalakaran, J. Keebler, B. Krueger, C. Malone, C. Mebane, and M. Cook) for their dedication and commitment to this work. We also thank R. Stewart, K. Gwinn, R. Corriveau, B. Fureman and V. Whittemore from the National Institute of Neurological Disorders and Stroke for their careful oversight and guidance of both EPGP and Epi4K.

We thank the following organizations for assistance in publicizing EPGP; enabling us to recruit participants effectively: AED Pregnancy Registry, American Epilepsy Society, Association of Child Neurology Nurses, California School Nurses Organization, Child Neurology Society, Citizens United for Research in Epilepsy, Dravet Syndrome Foundation, Epilepsy Alliance of Orange County, Epilepsy Foundation, Epilepsy Therapy Project, Finding a Cure for Epilepsy and Seizures, IDEA League, InfantileSpasms.com, Lennox-Gastaut Syndrome Foundation, PatientsLikeMe, People Against Childhood Epilepsy, PVNH Support & Awareness, and Seizures & Epilepsy Education.

We would like to acknowledge the following individuals or groups for the contributions of control samples: D. Daskalakis; P. Lugar; J. Milner; T. Young and K. Whisenhunt; Z.

Farfel, D. Lancet, and E. Pras; W. Lowe; R. Gbadegesin and M. Winn; K. Schmader, S. McDonald, H. K. White and M. Yanamadala; A. Holden; E. Behr; C. Moylan; A. M. Diehl and M. Abdelmalek; S. Palmer; G. Nestadt; J Samuels; Y. Wang; M. Carrington; M. Harms; T. Miller; A. Pestronk; R. Bedlack; R. Brown; N. Shneider; S. Gibson; J. Ravits; A. Gilter; J. Glass; F. Baas; E. Simpson; and G. Rouleau; K. Welsh-Bomer, C. Hulette, J. Burke; The ALS Sequencing Consortium; The Murdock Study Community Registry and Biorepository; M. Connors, L. Morris and the CHAVI investigators; the Carol Woods and Crosdaile Retirement Communities; and DUHS (Duke University Health System) Nonalcoholic Fatty Liver Disease Research Database and Specimen Repository. The collection of control samples and data was funded in part by: Biogen Idec.; The Duke Chancellor's Discovery Program Research Fund 2014; Bill and Melinda Gates Foundation; The Division of Intramural Research; B57 SAIC-Fredrick Inc M11-074; Bryan ADRC NIA P30 AG028377; The Ellison Medical Foundation New Scholar award AG-NS-0441-08; National Institute of Mental Health (K01MH098126, R01MH097993); National Institute of Allergy and Infectious Diseases (1R56AI098588-01A1); and National Institute of Allergy and Infectious Diseases Center (U19-AI067854, UM1-AI100645).

The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010).

Conflict of Interest / Funding Support

This work was supported by grants from the National Institute of Neurological Disorders and Stroke (The Epilepsy Phenome/Genome Project NS053998; Epi4K NS077364, NS077274, NS077303, and NS077276), The Andrew's Foundation, Finding a Cure for Epilepsy and Seizures, the Richard Thalheimer Philanthropic Fund, and the Eurocores program EuroEPINOMICS-RES of the European Science Foundation. The project received further support through grants from the Fund for Scientific Research Flanders (FWO); the Academy of Finland (141549); the Folkhälsan Research Foundation; the program "Investissements d'avenir" ANR-10-IAIHU-06; the Federal Ministry for Education and Research (IonNeurONet: 01GM1105), the German Research Foundation (DFG: HE5415/3-1; Le1030/11-1; RO3396/2-1), the German Society for Epileptology (DGfE), the Foundation noepilep.; the Swiss National Science Foundation (SNF: 32EP30_136042/1); the Wellcome Trust (09805); intramural funds of the University of Kiel; the Popgen 2.0 network (P2N) through the German Ministry for Education and Research (01EY1103); the European Union through Seventh Framework Programme (FP7) under the project DESIRE (N602531). The project also received infrastructural support through the Institute of Clinical Molecular Biology in Kiel, supported in part by DFG Cluster of Excellence ''Inflammation at Interfaces'' and ''Future Ocean''. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. Orrin Devinsky, David B. Goldstein, Steve Petrou and Slavé Petrovski have interests in companies related to epilepsy precision medicine.

References

1. Epi4K Consortium, Epilepsy Phenome/Genome Project (2013). *De novo* mutations in epileptic encephalopathies. *Nature* 501, 217-221.

2. EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, and Epi4K Consortium (2014). *De novo* mutations in synaptic transmission genes including *DNM1* cause epileptic encephalopathies. *Am J Hum Genet* 95, 360-370.

3. Epilepsy Phenome/Genome Project, Epi4K Consortium (2015). Copy number variant analysis from exome data in 349 patients with epileptic encephalopathy. *Ann Neurol* 78, 323-328.

4. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575.

5. Jiang, Y., Satten, G.A., Han, Y., Epstein, M.P., Heinzen, E.L., Goldstein, D.B., and Allen, A.S. (2014). Utilizing population controls in rare-variant case-parent association tests. *Am J Hum Genet* 94, 845-853.

6. Jiang, Y., McCarthy, J.M., and Allen, A.S. (2015). Testing the effect of rare compound-heterozygous and recessive mutations in case--parent sequencing studies. *Genet Epidemiol* 39, 166-172.

7. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249.

8. van Karnebeek, C.D., Shevell, M., Zschocke, J., Moeschler, J.B., and Stockler, S. (2014). The metabolic evaluation of the child with an intellectual developmental disorder: diagnostic algorithm for identification of treatable causes and new digital resource. *Mol Genet Metab* 111, 428-438.

9. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/).

10. Zhu, X., Petrovski, S., Xie, P., Ruzzo, E.K., Lu, Y.F., McSweeney, K.M., Ben-Zeev, B., Nissenkorn, A., Anikster, Y., Oz-Levi, D., et al. (2015). Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. *Genet Med* 17, 774-781.

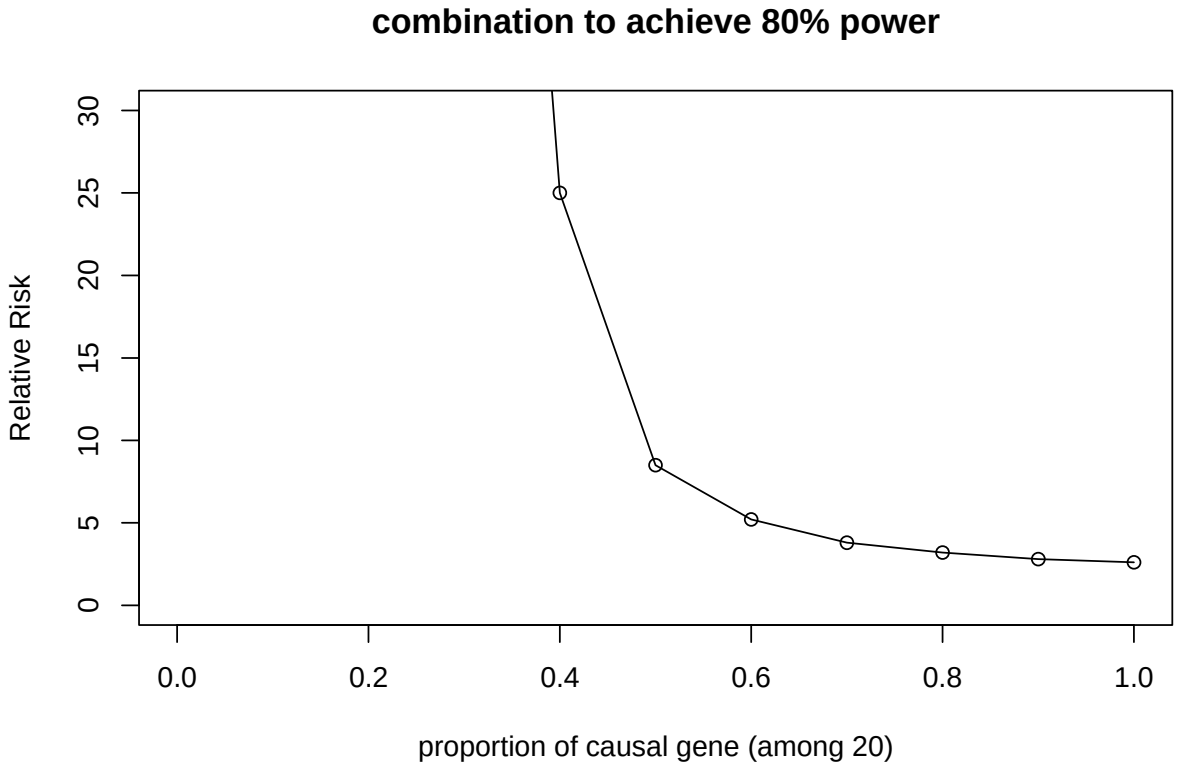
11. Need, A.C., Shashi, V., Hitomi, Y., Schoch, K., Shianna, K.V., McDonald, M.T., Meisler, M.H., and Goldstein, D.B. (2012). Clinical application of exome sequencing in undiagnosed genetic conditions. *J Med Genet* 49, 353-361.

12. Tanaka, A.J., Cho, M.T., Millan, F., Juusola, J., Retterer, K., Joshi, C., Niyazov, D., Garnica, A., Gratz, E., Deardorff, M., et al. (2015). Mutations in *SPATA5* are associated with microcephaly, intellectual disability, seizures, and hearing loss. *Am J Hum Genet* 97, 457-464.

13. Akawi, N., McRae, J., Ansari, M., Balasubramanian, M., Blyth, M., Brady, A.F., Clayton, S., Cole, T., Deshpande, C., Fitzgerald, T.W., et al. (2015). Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat Genet* 47, 1363-1369.

295 **FIGURE LEGEND**

296 **Figure 1.** coreTDT power simulation conditional on the parental genotype of 54
297 informative families and 20 informative genes in the compound heterozygous analysis.
298 Presents the combination of the relative risk and the proportion of disease causal genes
299 among these 20 informative genes, under which the tests can achieve 80% power.



300

Table 1. Top 10 genes from the analysis of rvTDT with 320 and subsequently with the subset of 286 European ancestry trios.

Test	Gene	Number of SNVs[#]	Uncorrected p-value
Analysis of rvTDT with 320 trios	*ABCA13*	52	1.17E-05
	CENPO	4	0.00236
	DST	51	0.00297
	IPPK	5	0.00198
	PCF11	10	0.00032
	SCAPER	5	0.00137
	SLC46A3	6	0.00185
	STEAP4	4	0.00291
	TRAF3IP1	7	0.00292
	ZNF878	7	0.00252
Analysis of rvTDT with 286 trios	*ABCA13*	52	0.00025
	SCAPER	5	0.00046
	ANKZF1	11	0.00079
	DST	51	0.00123
	TRAF3IP1	7	0.00172
	STEAP4	4	0.00247
	SBSN	12	0.00295
	OR2B2	5	0.00298
	SAP130	8	0.00326
	SCARF1	11	0.00344

[#]Representing the number of qualifying variants found in this gene across all families. P-value is based on the linear combination test with population controls. Adjusted α correcting for the number of genes is $p < 2.81 \times 10^{-6}$.

Table 2: Top 10 autosomal genes from the analysis of coreTDT with 320 trios

Gene	Number of SNVs[#]	Uncorrected p-value
PGM2L1	4	0.00195
CEP120	8	0.00391
CR1	25	0.00756
C14ORF177	4	0.01288
CNTRL	24	0.01563
DACT1	8	0.01563
KATNB1	10	0.01563
SYNJ1	17	0.01563
ZNF677	11	0.01563
KIAA1614	15	0.01973

[#]Representing the number of qualifying variants found in this gene across all families. Adjusted α correcting for the number of genes is $p < 2.81 \times 10^{-6}$.

Supplementary Methods

Power Simulation

Let G_f, G_m, G_o be the number of gene copies harboring a qualifying mutation in the trio's father, mother and offspring, respectively. We condition our power analysis on the observed parental genotype and study our ability to identify signal given a differing proportion of causal genes (out of the total number of genes considered), γ, and differing relative risks, R, of being diseased given two gene copies (of a causal disease gene) are affected versus less than two copies are affected. Since the analysis is conditional on the observed parental data, only a subset of genes and families are informative. [6] Specifically, only 20 genes across 54 families can have compound genotypes that lead to informative transmissions, i.e., $G_f = G_m = 1$, $G_f = 1, G_m = 2$ or $G_f = 2, G_m = 1$. 46 families are informative for only one gene and eight families are informative for two genes. In each of these eight families, the two genes are located on different chromosomes, so we assume that the transmissions of each gene are independent.

Let $D_o = 1$ indicate the fact that the offspring is affected. Let C be an indicator of whether the gene whose transmission is being considered is among the set of disease causal genes or not. When a family is informative for two genes disease causal indicators are given for each gene by C_1 and C_2. Note, we assume the disease risk for samples with multiple affected disease genes are the same with those with only one affected disease gene.

To simulate trios under the alternative, we first randomly select 20γ genes as disease causal and then generate offspring as follows.

If the family is informative for only one gene, the distribution of both offspring's gene copies being affected is given by

$$\Pr\left(G_o = 2 \middle| G_f = 1, G_m = 1, D_o = 1, C = 1\right) = \frac{R}{R+3}$$

$$\Pr\left(G_o = 2 \middle| G_f + G_m = 3, D_o = 1, C = 1\right) = \frac{R}{R+1}$$

$$\Pr\left(G_o = 2 \middle| G_f = 1, G_m = 1, D_o = 1, C = 0\right) = 0.75$$

$$\Pr\left(G_o = 2 \middle| G_f + G_m = 3, D_o = 1, C = 0\right) = 0.5.$$

If the family is informative for two genes and no more than one of them are disease causal, the compound genotype the two genes can be computed independently of one another using the equation above. When both genes are disease causal, their transmissions are not independent given the offspring is affected. In this case the compound genotypes of the offspring, for the two genes, can be given by,

$$\Pr\left(G_{o1} = 2, G_{o2} = 2 \middle| G_{f1} = G_{m1} = G_{f2} = G_{m2} = 1, D_o = 1, C_1 = C_2 = 1\right) = \frac{R}{7R+9}$$

$$\Pr\left(G_{o1} = 2, G_{o2} \neq 2 \middle| G_{f1} = G_{m1} = G_{f2} = G_{m2} = 1, D_o = 1, C_1 = C_2 = 1\right) = \frac{6R}{7R+9}$$

$$\Pr\left(G_{o1} = 2, G_{o2} = 2 \middle| G_{f1} = G_{m1} = 1, G_{f2} = 1, = G_{m2} = 2, D_o = 1, C_1 = C_2 = 1\right) = \frac{R}{5R+3}$$

$$\Pr\left(G_{o1} = 2, G_{o2} \neq 2 \middle| G_{f1} = G_{m1} = 1, G_{f2} = 1, = G_{m2} = 2, D_o = 1, C_1 = C_2 = 1\right) = \frac{R}{5R+3}$$

$$\Pr\left(G_{o1} \neq 2, G_{o2} = 2 \middle| G_{f1} = G_{m1} = 1, G_{f2} = 1, = G_{m2} = 2, D_o = 1, C_1 = C_2 = 1\right) = \frac{3R}{5R+3},$$

where G_{o1}, G_{m1}, G_{f1} and G_{o2}, G_{m2}, G_{f2} denotes the trio's compound genotypes at the first and second gene, respectively. We apply coreTDT to each simulated dataset and for each combination of γ and R, we use 1000 replicates to estimate the power. The combination of γ and R which obtains 80% power are presented in Figure 1.

Consortium Membership

Epi4K

Andrew S. Allen [1], Samuel F. Berkovic [2], Joshua Bridgers [3], Patrick Cossette [4], Dennis Dlugos [5], Michael P. Epstein [6], Tracy Glauser [8], David B. Goldstein [3], Erin L. Heinzen [3], Yu Jiang [1], Michael R. Johnson [9], Ruben Kuzniecky [10], Daniel H. Lowenstein [11], Anthony G. Marson [12], Heather C. Mefford [13], Terence J. O'Brien [14], Ruth Ottman [15], Steven Petrou[14,16,17], Slavé Petrovski [2,3,14], Annapurna Poduri [18], Zhong Ren [3], Ingrid E. Scheffer [19], Elliott Sherr [20], Quanli Wang [3]

EuroEPINOMICS-RES

Rudi Balling[21], Nina Barisic[22], Stéphanie Baulac[23], Hande Caglayan[24], Dana Craiu[25], Peter De Jonghe[26], Christel Depienne[27], Renzo Guerrini[28], Ingo Helbig[29, 30] Helle Hjalgrim[31], Dorota Hoffman-Zacharska[32], Johanna Jähn[29], Karl Martin Klein[33, 34], Bobby Koeleman[35], Vladimir Komarek[36], Roland Krause[21], Eric Leguern[23], Anna-Elina Lehesjoki[37], Johannes R Lemke[38] Holger Lerche[39], Tarja Linnankivi[40], Carla Marini[28], Patrick May[21], Rikke S. Møller[31] ,Hiltrud Muhle[29] , Deb Pal[41] Aarno Palotie[42] , Felix Rosenow[33, 34], Kaja Selmer[43], Jose M Serratosa[44], Sanjay Sisodiya[45], Ulrich Stephani[29], Katalin Sterbova[36], Pasquale Striano[46], Arvid Suls[26, 47], Tiina Talvik[48], Sarah von Spiczak[29], Yvonne Weber[39], Sarah Weckhuysen[26], Federico Zara[49]

EPGP

Bassel Abou-Khalil [50], Brian K. Alldredge [51], Dina Amrom [52], Eva Andermann [53], Frederick Andermann [54], Jocelyn F. Bautista [55], Samuel F. Berkovic [2], Judith Bluvstein [10], Gregory D. Cascino [56], Damian Consalvo [57], Patricia Crumrine [58], Orrin Devinsky [59], Dennis Dlugos [5], Michael P. Epstein [6], Miguel E. Fiol [60], Nathan B. Fountain [61], Jacqueline French [62], Daniel Friedman [63], Tracy Glauser [8], Kevin Haas [64], Sheryl R. Haut [65], Jean Hayward [66], Sucheta Joshi

[67], Andres Kanner[68] , Heidi E. Kirsch [69], Eric H. Kossoff [70], Rachel Kuperman [71], Ruben Kuzniecky [10], Daniel H. Lowenstein [11], Shannon M. McGuire [72], Paul V. Motika [73], Edward J. Novotny [74], Ruth Ottman [15], Juliann M. Paolicchi [75], Jack Parent [76], Kristen Park [77], Annapurna Poduri [18], Ingrid E. Scheffer [19], Renée A. Shellhaas [78], Elliott Sherr [20], Joseph Sirven [79], Michael C. Smith [80], Joseph Sullivan [11], Liu Lin Thio [81], Anu Venkat [82], Eileen P.G. Vining [83], Gretchen K. Von Allmen [84], Judith L. Weisenberg [85], Peter Widdess-Walsh [86], Melodie R. Winawer [87]

Author contributions

Sequence Analysis and Statistical Interpretation: Y.J. (lead analyst), A.S.A and D.B.G.

Bioinformatics processing: J.Br., E.L.H., Y.J., Sl.P., Z.R. and Q.W.

Clinical expert panel: I.H., H.C.M., A.P., and S.W.

Writing of manuscript: A.S.A., S.F.B., P.D.J., D.B.G., E.L.H., I.H., Y.J., D.H.L., and Sl.P.

Epi4K Steering Committee: A.S.A, S.F.B., P.C., N.D., D.D., E.E.E., M.P.E., T.G., D.B.G., E.L.H., M.R.J., Ru.Ku., D.H.L., A.G.M., H.C.M., T.J.O., R.O., St.P, Sl.P., A.P., I.E.S., E.S.

EuroEPINOMICS-RES consortium leadership and study coordination: I.H., P.D.J., S.W.

EuroEPINOMICS Study design: A-E.L., A.S., B.K, H.L., I.H., P.G., P.D.J., S.W.

EPGP Study Design: B.K.A., O.D., D.D., M.P.E., Ru.Ku., D.H.L., R.O., E.S., M.R.W.

Epi4K Epileptic Encephalopathy Phenotyping: S.F.B., P.C., D.D., Ru.Ku., D.H.L., H.C.M., R.O., A.P., I.E.S., E.S.

EuroEPINOMICS-RES Proband recruitment and phenotyping: A-E.L., A.R., C.M., C.D, D.C., D.P., D.H., E.L., F.Z., F.Ros., H.C, H.H, H.M, H.L., I.H., J.J., J.L., J.S., K.Se., K.M.K., K.St., N.B., P.S., R.G., R.S.M., S.v.S., S.W., S.B., T.L. T.T., U.S., V.K., Y.W. EPGP

EPGP Proband Recruitment and Phenotyping: B.A., E.A., F.A., D.A, J.F.B., S.F.B., G.D.C., D.Co., P.Cr., O.D., D.D., M.E.F., N.B.F., D.F., E.B.G., T.G., S.G., S.R.H., J.H., K.H, S.L.H., H.E.K., R.C.Kn., E.K., Ra.Ku., Ru.Ku, D.H.L., S.M.M., P.V.M., E.J.N., J.M.P., J.P., K.P., A.P., I.E.S., J.J.S., R.A.S., J.Si., L.S., M.S., L.L.T., A.V., E.P.G.V., G.K.V., J.W., P.W.

EPGP Phenotype Data Analysis: B.A., B.K.A., A.B., J.B., G.D.C., O.D., D.D., M.P.E., J.F., T.G., S.J., A.K., R.C.Kn., Ru.Ku., D.H.L., R.O., J.M.P., A.P., I.E.S., R.A.S., R.S., E.S., J.J.S., J.Su., P.W., M.R.W.

Affiliations

[1.] Department of Biostatistics and Bioinformatics, Duke Clinical Research Institute, Durham, North Carolina 27710 USA.

[2.] Epilepsy Research Centre, Department of Medicine, University of Melbourne (Austin Health), Heidelberg, Victoria 3084, Australia.

[3.] Institute for Genomic Medicine, Columbia University Medical Center, New York, New York 10032, USA.

[4.] Department of Neurosciences and CHUM Research Center, Université de Montréal, CHUM-Hôpital Notre-Dam Montréal, Quebec H2L 4M1e, Canada.

[5.] Department of Neurology and Pediatrics, The Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104 USA.

[6.] Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA.

[7.] Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195 USA.

[8.] Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229 USA.

[9.] Centre for Clinical Translation Division of Brain Sciences, Imperial College London, London, SW7 2AZ UK.

[10.] NYU School of Medicine, New York University, New York, New York 10016 USA.

[11.] Department of Neurology, University of California, San Francisco, San Francisco, California 94143 USA.

[12.] Department of Molecular and Clinical Pharmacology, University of Liverpool, Clinical Sciences Centre, Lower Lane, Liverpool, L9 7LJ, United Kingdom.

[13.] Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, Washington 98115 USA.

[14.] Departments of Medicine and Neurology, The Royal Melbourne Hospital, Parkville, Victoria, 3146 Australia.

[15.] Departments of Epidemiology and Neurology, and the G.H. Sergievsky Center, Columbia University; and Division of Epidemiology, New York State Psychiatric Institute, New York, New York 10032 USA.

[16.] Florey Institute for Neuroscience and Mental Health. The University of Melbourne, VIC 3010, Australia.

[17.] Centre for Neural Engineering. The University of Melbourne, VIC 3010, Australia.

[18.] Division of Epilepsy and Clinical Neurophysiology, Department of Neurology Boston Children's Hospital, Boston, Massachusetts 02115 USA.

[19.] Epilepsy Research Centre, Department of Medicine, University of Melbourne (Austin Health), Heidelberg, Victoria 3084, Australia, Florey Institute and Department of Pediatrics, Royal Children' s Hospital, University of Melbourne, Victoria 3052, Australia.

[20.] Departments of Neurology, Pediatrics and Institute of Human Genetics, University of California, San Francisco, San Francisco, California 94158 USA.

[21.] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg

[22] Department of Paediatrics, University of Zagreb, Medical School, University Hospital Centre Zagreb, 10000 Zagreb, Croatia

452 [23.] Sorbonne Universités, UPMC Univ Paris 06, UM 75, F-75013 Paris, France; INSERM,
453 U1127, F-75013 Paris, France; CNRS, UMR 7225, F-75013 Paris, France ; ICM, F-75013 Paris,
454 France; Department of genetic and cytogenetic, AP-HP, Hôpital Pitié-Salpêtrière, F-75013 Paris,
455 France
456 [24] Department of Molecular Biology and Genetics, Bogazici University, 34342 Istanbul, Turkey
457 [25] Pediatric Neurology Clinic II, Department of Neurology, Pediatric Neurology, Psychiatry,
458 Neurosurgery, "Carol Davila" University of Medicine, Bucharest, Sector 4, Romania; Pediatric
459 Neurology Clinic, "Professor Doctor Alexandru Obregia" Clinical Hospital, Sector 4, Bucharest,
460 Romania
461 [26] Neurogenetics group, Department of Molecular Genetics, VIB, 2610 Antwerp, Belgium;
462 Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, 2610 Antwerp,
463 Belgium
464 [27]Département de Médicine translationnelle et Neurogénétique, IGBMC, CNRS UMR
465 7104/INSERM U964/Université de Strasbourg, 67400 Illkirch, France ; Laboratoire de
466 cytogénétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
467 [28] Pediatric Neurology Unit and Laboratories, Children's Hospital A. Meyer – University of
468 Florence, 50132 Florence, Italy
469 [29] Department of Neuropediatrics, University Medical Center Schleswig-Holstein, Christian-
470 Albrechts University, 24105 Kiel, Germany
471 [30] Division of Neurology, The Children's Hospital of Philadephia, 3401 Civic Center Blvd.,
472 Philadelphia, PA 19104-4399, USA
473 [31] Danish Epilepsy Centre, Dianalund, Denmark; Institute for Regional Health research,
474 University of Southern Denmark, Odense, Denmark

[32] Department of Medical Genetics, Institute of Mother and Child, 01211 Warsaw, Poland

[33] Epilepsy Center Rhine-Main, Center of Neurology and Neurosurgery, Goethe-University Frankfurt, Schleusenweg 2-16, Haus 95, 60528 Frankfurt a.M.

[34] Epilepsy Center Hessen, Philipps-University Marburg, Marburg, Germany

[35] Department of Medical Genetics, University Medical Center, Utrecht, The Netherlands

[36] Department of Child Neurology, Second Medical Faculty and University Hospital Motol, 150 06 Praha, Czech Republic

[37] Folkhälsan Institute of Genetics; Research Programs Unit, Molecular Neurology and Neuroscience Center, University of Helsinki, 00290 Helsinki, Finland

[38] Institute of Human Genetics, University of Leipzig Hospitals and Clinics, Leipzig, Germany

[39] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany

[40] Children´s Hospital, Pediatric Neurology, University of Helsinki and Helsinki University Central Hospital, FIN-00029 HUS, Helsinki, Finland

[41] Department of Clinical Neuroscience, Institute of Psychiatry, King's College London, London, UK

[42] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, United Kingdom; Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00290 Helsinki, Finland; Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[43] Department of Medical Genetics, Oslo University Hospital, Oslo, Norway

[44] Epilepsy Unit, Department of Neurology, Institute of Medical Research at Fundación Jiménez Díaz University Hospital and Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER) Unit 744, Madrid, Spain

[45] Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, London, UK; Epilepsy Society, Buckinghamshire, UK

[46] Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophtalmology, Genetics, Maternal and Child Health, University of Genova and Gaslini Institute, 16147 Genova, Italy

[47] GENOMED, Center of Medical Genetics, University of Antwerp & Antwerp University Hospital, Antwerp, Belgium.

[48] Department of Pediatrics, University of Tartu, 51014 Tartu, Estonia; Department of Neurology and Neurorehabilitation, Children's Clinic, Tartu University Hospital, 50406 Tartu, Estonia

[49] Laboratory of Neurogenetics, Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, University of Genova and Gaslini Institute, 16147 Genova, Italy

[50.] Department of Neurology, Vanderbilt University Medical Center, Nashville, Tennessee 37232 USA.

[51.] Department of Clinical Pharmacy, UCSF School of Pharmacy, Department of Neurology, UCSF School of Medicine 94143 USA.

[52.] Neurogenetics Unit, Montreal Neurological Hospital and Institute; Departments of Neurology & Neurosurgery, McGill University, Montreal QC H3A 2B4 Canada

[53.] Neurogenetics Unit and Epilepsy Research Group, Montreal Neurological Hospital and Institute; Departments of Neurology & Neurosurgery and Human Genetics, McGill University, Montreal QC H3A 2B4 Canada

519 [54.] Epilepsy Research Group, Montreal Neurological Hospital and Institute; Departments of
520 Neurology & Neurosurgery and Pediatrics, McGill University, Montreal QC H3A 2B4 Canada
521 [55.] Department of Neurology, Cleveland Clinic Lerner College of Medicine & Epilepsy Center of
522 the Cleveland Clinic Neurological Institute, Cleveland, Ohio 44195 USA.
523 [56.] Divison of Epilepsy, Mayo Clinic, Rochester, Minnesota 55905 USA.
524 [57.] Epilepsy Center, Neurology Division, Ramos Mejía Hospital, Buenos Aires, 1221, Argentina.
525 [58.] Medical Epilepsy Program & EEG & Child Neurology, Children's Hospital of Pittsburgh of
526 UPMC, Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15224
527 USA.
528 [59.] NYU and Saint Barnabas Epilepsy Centers, NYU School of Medicine, New York, New York
529 10016 USA.
530 [60.] Department of Neurology, Epilepsy Care Center, University of Minnesota Medical School,
531 Minneapolis, Minnesota 55414 USA.
532 [61.] FE Dreifuss Comprehensive Epilepsy Program, University of Virginia, Charlottesville,
533 Virginia 22908 USA.
534 [62.] NYU Comprehensive Epilepsy Center, New York, New York 10016 USA.
535 [63.] Department of Neurology, NYU School of Medicine, New York, New York, 10016 USA.
536 [64.] Vanderbilt University Medical Center, Nashville, Tennessee 37232 USA.
537 [65.] Comprehensive Epilepsy Center, Montefiore Medical Center, Bronx, New York 10467 USA.
538 [66.] The Kaiser Permanente Group, Oakland, California 94618 USA.
539 [67.] Pediatrics & Communicable Diseases, University of Michigan, Ann Arbor, Michigan 48109
540 USA.

[68.] Comprehensive Epilepsy Center, Department of Neurology, Miller School of Medicine, University of Miami, Miami USA.

[69.] Departments of Neurology and Radiology, University of California, San Francisco, California 94143 USA.

[70.] Neurology and Pediatrics, Child Neurology, Pediatric Neurology Residency Program, Johns Hopkins Hospital, Baltimore, Maryland 21287 USA.

[71.] Epilepsy Program, Children's Hospital & Research Center Oakland, Oakland, California 94609 USA.

[72.] Clinical Neurology, Children's Hospital Epilepsy Center of New Orleans, New Orleans, Louisiana 70118 USA.

[73.] Comprehensive Epilepsy Center, Oregon Health and Science University, Portland, Oregon 97239 USA.

[74.] Departments of Neurology and Pediatrics, University of Washington School of Medicine, Seattle Children's Hospital, Seattle, Washington 98105 USA.

[75.] Weill Cornell Medical Center, New York, New York 10065 USA.

[76.] Department of Neurology and Neuroscience Graduate Program, University of Michigan Medical Center, Ann Arbor, MI 49108 and Ann Arbor Veterans Administration Healthcare System, Ann Arbor, Michigan 48105 USA.

[77.] University of Colorado School of Medicine, Aurora, Colorado 80045, U.S.A.; Division of Neurology, Department of Pediatrics, Children's Hospital Colorado, Aurora, Colorado 80045 USA.

[78.] University of Michigan, Pediatric Neurology, Ann Arbor, Michigan 48109 USA.

[79.] Department of Neurology, Mayo Clinic, Scottsdale, Arizona 85259 USA.

564 [80.] Department of Neurological Sciences, Rush Epilepsy Center, Rush University Medical
565 Center, Chicago, Illinois 60612 USA.
566 [81.] Department of Neurology, Washington University School of Medicine, St. Louis, Missouri
567 63110.
568 [82.] Department of Pediatrics, Saint Peter's University Hospital, New Brunswick, New Jersey,
569 08901
570 [83.] Department of Neurology, Johns Hopkins Hospital, Baltimore, Maryland 21287 USA.
571 [84.] Division of Child & Adolescent Neurology, Departments of Pediatrics , University of Texas
572 Medical School, Houston, Texas 77030 USA.
573 [85.] Department of Neurology, Division of Pediatric Neurology, Washington University School of
574 Medicine, St. Louis, Missouri 63110 USA.
575 [86.] Department of Neurology, Beaumont Hospital and Royal College of Surgeons, Dublin 9
576 Ireland.
577 [87.] Department of Neurology and the G.H. Sergievsky Center, Columbia University, New York,
578 New York, 10032 USA.
579