University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Oliver, KL;Lukic, V;Freytag, S;Scheffer, IE;Berkovic, SF;Bahlo, M

Title:
In silico prioritization based on coexpression can aid epileptic encephalopathy gene discovery

Date:
2016-02-01

Citation:
Oliver, K. L., Lukic, V., Freytag, S., Scheffer, I. E., Berkovic, S. F. & Bahlo, M. (2016). In silico prioritization based on coexpression can aid epileptic encephalopathy gene discovery. Neurology Genetics, 2 (1), https://doi.org/10.1212/NXG.0000000000000051.

Persistent Link:
https://hdl.handle.net/11343/262366

License:
CC BY-NC-ND

In silico prioritization based on coexpression can aid epileptic encephalopathy gene discovery

OPEN

Karen L. Oliver, MSc
Vesna Lukic, MSc
Saskia Freytag, PhD
Ingrid E. Scheffer, MBBS, PhD, FRACP
Samuel F. Berkovic, MD, FRS
Melanie Bahlo, PhD

Correspondence to A/Prof. Bahlo: bahlo@wehi.edu.au

ABSTRACT

Objective: To evaluate the performance of an in silico prioritization approach that was applied to 179 epileptic encephalopathy candidate genes in 2013 and to expand the application of this approach to the whole genome based on expression data from the Allen Human Brain Atlas.

Methods: PubMed searches determined which of the 179 epileptic encephalopathy candidate genes had been validated. For validated genes, it was noted whether they were 1 of the 19 of 179 candidates prioritized in 2013. The in silico prioritization approach was applied genome-wide; all genes were ranked according to their coexpression strength with a reference set (i.e., 51 established epileptic encephalopathy genes) in both adult and developing human brain expression data sets. Candidate genes ranked in the top 10% for both data sets were cross-referenced with genes previously implicated in the epileptic encephalopathies due to a de novo variant.

Results: Five of 6 validated epileptic encephalopathy candidate genes were among the 19 prioritized in 2013 (odds ratio = 54, 95% confidence interval [7, ∞], $p = 4.5 \times 10^{-5}$, Fisher exact test); one gene was false negative. A total of 297 genes ranked in the top 10% for both the adult and developing brain data sets based on coexpression with the reference set. Of these, 9 had been previously implicated in the epileptic encephalopathies (*FBXO41*, *PLXNA1*, *ACOT4*, *PAK6*, *GABBR2*, *YWHAG*, *NBEA*, *KNDC1*, and *SELRC1*).

Conclusions: We conclude that brain gene coexpression data can be used to assist epileptic encephalopathy gene discovery and propose 9 genes as strong epileptic encephalopathy candidates worthy of further investigation. ***Neurol Genet* 2016;2:e51; doi: 10.1212/NXG.0000000000000051**

Currently, the genetic diagnostic yield for epileptic encephalopathies using high-throughput sequencing technologies is 25%–30%.[1] Although whole-exome sequencing has entered the clinical arena, data interpretation remains a considerable challenge for the majority of patients, who remain unsolved. When trios are studied, the presence of a de novo mutation in an established disease gene is usually diagnostic. However, the interpretation of de novo mutations in candidate genes remains difficult because healthy controls have 0–3 (median 1) de novo exonic variants.[2] There is now a growing list of candidate epileptic encephalopathy genes that harbor a plausible (e.g., novel and likely functional) de novo variant in a single patient.

In 2013, the Epi4K/EPGP Consortia performed whole-exome sequencing on 264 epileptic encephalopathy trios.[3] The Consortia identified >300 de novo variants, with the majority representing "single hits" in genes not previously implicated in epilepsy. We developed and applied an in silico prioritization approach[4] to a subset of these candidate epileptic encephalopathy genes (n = 179). Those candidate genes with de novo variants deemed most likely to be pathogenic (e.g., nonsynonymous or splice-site) were chosen. Our in silico approach used data

Supplemental data at Neurology.org/ng

From the Epilepsy Research Centre (K.L.O., I.E.S., S.F.B.), Department of Medicine, Austin Health, University of Melbourne, Heidelberg, Australia; Population Health and Immunity Division (V.L., S.F., M.B.), The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia; Florey Institute (I.E.S.), Melbourne, Australia; Department of Paediatrics (I.E.S.), University of Melbourne, Royal Children's Hospital, Melbourne, Australia; and Department of Mathematics and Statistics (M.B.) and Department of Medical Biology (M.B.), University of Melbourne, Australia.

Funding information and disclosures are provided at the end of the article. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was paid by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially.

© 2016 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.

from the Allen Human Brain Atlas.[5] We prioritized 19 of 179 candidate genes in 2013 because of high brain coexpression with established epileptic encephalopathy genes, based on an empirical false discovery rate of 0.25.[4]

New epileptic encephalopathy genes have since been confirmed. This provides an opportunity to validate the performance of our prioritization approach based on gene coexpression data (BrainGEP: http://bioinf.wehi.edu.au/software/BrainGEP/) and to expand its application to the wider genome.

METHODS The original reference set of 29 established epileptic encephalopathy genes (table e-1 at Neurology.org/ng) was identified by PubMed searches using the keywords "epilepsy," "epileptic encephalopathy," and "genetics" in June 2013.[4] Using the same search terms, we formed an updated list of epileptic encephalopathy genes published between June 2013 and August 2015. To be established as a causal epileptic encephalopathy gene, we required that variants in the same gene and similar epileptic encephalopathy[6] clinical presentation be confidently implicated in multiple individuals.[7] To be confidently implicated, the reported variants were required to meet the American Medical Genetics Genomics guidelines for "pathogenic" or "likely pathogenic" classification (table e-2).[8]

Performance evaluation. Newly established epileptic encephalopathy genes were cross-referenced for overlap with the list of 179 candidate genes used in our original study.[4] For those genes present in the candidate gene list, it was noted whether they were one of the 19 prioritized genes by BrainGEP, thus being validated.

Genome-wide prioritization. The updated list of established epileptic encephalopathy genes was used to form a new reference set. This reference set (n = 51; table e-1) was used to prioritize the 13,157 and 12,365 genes represented in the adult and developing brain expression data sets, respectively, using BrainGEP. Genome-wide candidates that ranked in the top 10% for both data sets were cross-referenced to genes reported with a Sanger-validated de novo variant, typically in a single case, by the EuroEPINOMICS-RES and Epi4K Consortia.[9]

RESULTS **Performance evaluation.** Since June 2013, of the 179 Epi4K/EPGP candidate genes with "single hits," 6 have been established as epileptic encephalopathy genes: *GNAO1*,[10] *GRIN2B*,[11] *DNM1*,[9] *SLC35A2*,[12] *KCNB1*,[13] and *GRIN1*.[14] Five of the 6 now-validated candidates were prioritized in 2013, representing true positives (odds ratio = 54, 95% confidence interval $[7, \infty]$, $p = 4.5 \times 10^{-5}$, one-sided Fisher exact test) (table 1). *SLC35A2* on the X chromosome represents the single false-negative finding. This gene ranked in the top 40% of the 179 candidate genes; genes in the top 10% were prioritized.[4]

Genome-wide prioritization. A total of 297 genes ranked in the top 10% of genome-wide candidates based on their coexpression, in both the adult and developing human brain, with the 51 reference epileptic encephalopathy genes. Of these top-ranked genome-wide candidates (table e-3), 9 were reported by the EuroEPINOMICS-RES and Epi4K Consortia[9] and therefore have already been implicated in the epileptic encephalopathies with a de novo variant, typically in a single case (table 2).

Table 1 Summary of prioritized vs validated candidate epileptic encephalopathy genes from original study[4]

	Prioritized candidates		
	Yes	No	Total
Validated	5 (true positive)	1 (false negative)	6
Not yet validated	14	159	173
Total	19	160	179

DISCUSSION Genetic research has been revolutionized by high-throughput sequencing technology; no longer is the rate-limiting step data generation but rather the interpretation of these data. This can be particularly challenging for diseases with appreciable genetic heterogeneity, such as the epileptic encephalopathies,[15] where a common challenge is the interpretation of novel genes with a plausible de novo variant in a single case. Here we have demonstrated the merit of incorporating brain-specific gene coexpression data to add a further layer of information for or against candidates by way of in silico gene prioritization. In addition, we used this information to identify a small number of the most promising epileptic encephalopathy candidate genes from the whole genome.

We systematically analyzed the performance of our in silico approach that prioritized 19 candidate epileptic encephalopathy genes as those most likely to be pathogenic from a list of 179 in 2013.[4] Since then, 6 of the 179 candidates have been confirmed as new epileptic encephalopathy genes, 5 of which had been prioritized, demonstrating noteworthy success. This reinforces the remaining 14 prioritized genes[4] as strong epileptic encephalopathy candidates; it is expected that future publications will result in a number of them being validated.

The one validated epileptic encephalopathy candidate gene that was not prioritized by our approach, *SLC35A2*, is located on the X chromosome. Complex mechanisms of dosage compensation balance X-linked and autosomal gene expression levels; however, substantial variability can be seen between individuals and tissue types.[16] It may be that this complexity somewhat compromised the result for

© 2016 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.

Table 2 Nine previously implicated epileptic encephalopathy candidate genes[9] prioritized in the top 10% of the genome based on adult and developing brain gene coexpression with 51 established causative genes

Gene	Adult rank percentile	Developing rank percentile	De novo variant type[9]
FBXO41	0.21	1.13	Stop gain (n = 1)
PLXNA1	0.84	0.52	Missense (n = 1)
ACOT4	1.00	9.50	Missense (n = 1)
PAK6	1.79	1.56	Synonymous (n = 1)
GABBR2	2.21	0.05	Missense (n = 2)
YWHAG	3.75	5.08	Missense (n = 1)
NBEA	6.92	9.48	Stop gain (n = 1)
KNDC1	8.76	8.51	Synonymous (n = 1)
SELRC1	9.49	5.00	Missense (n = 1)

SLC35A2; however, *IQSEC2* is also located on the X chromosome and this candidate gene was one of the 19 prioritized. *IQSEC2* is a well-established intellectual disability gene, and although rare cases have been reported with seizures,[17] it did not meet our criteria for an established epileptic encephalopathy gene.

Having demonstrated the validity of our approach, we applied BrainGEP to the whole genome and prioritized candidates according to their coexpression with an updated reference set of 51 established epileptic encephalopathy genes. Of the 297 top-ranked candidate genes, 9 had been previously implicated in the epileptic encephalopathies due to the presence of de novo mutation but had not been statistically confirmed.[9] The prioritization of these genes (table 2) provides an additional layer of support for their role in the pathogenesis of the epileptic encephalopathies, particularly because the prioritization is based on coexpression data from relevant tissue (i.e., brain). We suggest that these 9 candidates are those most likely to validate and thus are excellent targets for candidate gene resequencing approaches.[18] In fact, the prioritization of *GABBR2* as one of the 9 candidate genes further reinforces this, as evidence for this gene is already quite strong. The EuroEPINOMICS-RES and Epi4K Consortia reported de novo mutations in *GABBR2* in 2 unrelated individuals with epileptic encephalopathy.[9] However, this did not reach statistical significance, so the evidence for *GABBR2* being causative was classified as only "suggestive" by the authors.[9]

In silico prioritization results are predictions based on the quantitative interpretation of biological networks captured by the data; results should not be interpreted as strong or independent lines of evidence for pathogenicity. Specific limitations to the approach include the assumption that similar syndromes are caused by mutations in genes that form part of the same biological pathway(s) as established disease genes (i.e., the reference set). This means that genes representing novel biological pathways are disadvantaged, as predicted gene-gene associations with the reference set are unlikely. The ability of in silico prioritization approaches to predict these gene-gene associations is, in turn, directly related to the quality of data sources used. An advantage of our approach is that it targets the disease of interest by using gene coexpression data from the brain. However, other data sources, such as text mining and protein-protein interactions, may have detected additional gene-gene associations not captured by expression data.

Despite the limitations, this work has highlighted how brain gene coexpression data can be harnessed to uncover important biological networks for the epileptic encephalopathies. This approach has the potential to frame future research strategies and therapeutic development. Our in silico prioritization work continues to evolve and now incorporates a new methodologic approach (RUVcorr) that denoises large gene expression data resources with an emphasis on extracting gene coexpression signals.[19] By using expression data from the brain, the application of this work is not limited to patients with epileptic encephalopathy but can be used to target the broader epilepsies and other neurologic diseases as well. We propose this as a valuable starting point for selecting the most promising candidate genes to target in resequencing experiments or to focus on when reanalyzing the exome data of "unsolved" patients (e.g., the Epilepsy Genetics Initiative) and when faced with a long list of novel putative causative genes.

AUTHOR CONTRIBUTIONS

Ms. Oliver drafted the manuscript and performed data analysis/interpretation. Ms. Lukic revised the manuscript and performed data analysis. Dr. Freytag revised the manuscript and contributed to study concept/design. Dr. Scheffer and Dr. Berkovic revised the manuscript and contributed to study concept/design and data interpretation. Dr. Bahlo revised the manuscript, contributed to study concept/design and data interpretation, and provided study supervision.

STUDY FUNDING

This work was supported by Victorian State Government Operational Infrastructure Support and Australian Government National Health and Medical Research Council (NHMRC) IRIISS funding to the Walter and Eliza Hall Institute of Medical Research (M.B. and V.L.). The NHMRC of Australia provided further support to S.F.B, I.E.S., and M.B. (Program Grant 628952 to S.F.B. and I.E.S.; Australia Fellowship 466671 to S.F.B.; Practitioner Fellowship 1006110 to I.E.S.; Senior Research Fellowship 1002098 and Program Grant 1054618 to M.B.). M.B. was also supported by an Australian Research Council Future Fellowship (FT100100764).

DISCLOSURE

Ms. Oliver, Ms. Lukic, and Dr. Freytag report no disclosures. Dr. Scheffer has received travel support/speaker honoraria from GSK, AOCCN, the Weizmann Institute, the American Academy of Neurology, IRCCS Oasi Maria SS, Sanofi China, QBI, University of Queensland,

© 2016 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.

the International League Against Epilepsy, the Australian Academy of Science, the Commonwealth Department of Industry (Australia), Westmead Hospital, Perpetual, the University of California, Matthew's Friends, SBS (Australia), PTS for NFLE conference, the Turkish Child Neurology Association, the European Congress on Epileptology, the International Child Neurology Association, UCB, the Movement Disorder Society, the International Epilepsy Congress, the University of Auckland, the World Congress of Neurology, Epileptic Disorders, Eisai, Athena Diagnostics, and Transgenomic; was funded by an NHMRC Practitioner Fellowship; serves on the editorial boards of the *Annals of Neurology*, *Neurology*, *Epilepsy Currents*, *Progress in Epileptic Disorders* series, *Virtual Neuro Centre*, and *Epileptic Disorders*; may accrue future revenue on a pending patent WO61/010176: Therapeutic Compound that relates to discovery of *PCDH19* gene as the cause of familial epilepsy with mental retardation limited to females; is one of the inventors listed on a patent held by Bionomics Inc. on Diagnostic testing using the *SCN1A* gene, WO2006/133508; has received revenue for patents on Diagnostic and therapeutic methods for EFMR, A Diagnostic Method for Epilepsy (also published as Methods for the Diagnosis and Treatment of Epilepsy), Mutations in Ion Channels, Diagnostic and Treatment Methods Relating to Autosomal Dominant Nocturnal Frontal Lobe Epilepsy (pending), and Gene and mutations thereof associated with seizure disorders; and receives/has received research support from the National Health and Medical Research Council of Australia, NINDS, the Epilepsy Association of Tasmania, the Melbourne Neurosciences Institute, the Weizmann Institute, the CURE SUDEP Award, and Perpetual Philanthropic Services. Dr. Berkovic has served on scientific advisory boards for UCB and Janssen-Cilag; serves on the editorial boards of *Lancet Neurology* and *Epileptic Disorders* and the Advisory Board of *Brain*; may accrue future revenue on pending patent WO61/010176: Therapeutic Compound that relates to discovery of *PCDH19* gene as the cause of familial epilepsy with mental retardation limited to females; is one of the inventors listed on a patent held by Bionomics Inc. on diagnostic testing using the *SCN1A* gene, WO2006/133508; has received speaker honoraria from UCB; has received unrestricted educational grants from UCB, Janssen-Cilag, and Sanofi-Aventis; and receives/has received research support from the National Health and Medical Research Council of Australia and NINDS. Dr. Bahlo receives/has received research support from the National Health and Medical Research Council of Australia, the Human Genetics Society of Australia, the Epilepsy Society of Australia, the Royal Statistical Society, the Institute of Mathematical Statistics, the Biometrics Society of Australasia, the Australian Statistics Society, the American Society of Human Genetics, and the International Genetic Epidemiology Society; has served on the scientific advisory board of Macular Telangiectasia (MacTel) Consortium and the Department of Health (Australian Government); has received travel funding/speaker honoraria from the 7th International Conference on Genomics & Bio-IT APAC, the International Mathematics Institute conference, the Australian Statistical Society Conference, the International Stroke Genetics Meeting, the Human Genome Meeting, and AGTA; has served on the editorial board of *Statistical Applications in Genetics and Molecular Biology*; holds patents for Method of determining response to treatment with immunomodulatory composition and Details genetic markers that predict response to Hepatitis C treatment; and has been a consultant for the MacTel Consortium Meeting, the Human Genetics Society of Australia Conference, the Australian Statistical Society-Institute of Mathematical Statistics, the International Conference on Systems Biology, the Sixth Barossa meeting, and the 8th International Conference on Genomics & Bio-IT APAC. Go to Neurology.org/ng for full disclosure forms.

Received September 28, 2015. Accepted in final form December 10, 2015.

REFERENCES

1. Mercimek-Mahmutoglu S, Patel J, Cordeiro D, et al. Diagnostic yield of genetic testing in epileptic encephalopathy in childhood. Epilepsia 2015;56:707–716.
2. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012;485:242–245.
3. Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen AS, Berkovic SF, Cossette P, et al. De novo mutations in epileptic encephalopathies. Nature 2013;501:217–221.
4. Oliver KL, Lukic V, Thorne NP, Berkovic SF, Scheffer IE, Bahlo M. Harnessing gene expression networks to prioritize candidate epileptic encephalopathy genes. PLoS One 2014;9:e102079.
5. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 2012;489:391–399.
6. Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia 2010;51:676–685.
7. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. Nature 2014;508:469–476.
8. Richards S, Aziz N, Bale S, et al. 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 2015;17:405–424.
9. EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am J Hum Genet 2014;95:360–370.
10. Nakamura K, Kodera H, Akita T, et al. De Novo mutations in GNAO1, encoding a Gαo subunit of heterotrimeric G proteins, cause epileptic encephalopathy. Am J Hum Genet 2013;93:496–505.
11. Lemke JR, Hendrickx R, Geider K, et al. GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann Neurol 2014;75:147–154.
12. Kodera H, Nakamura K, Osaka H, et al. De novo mutations in SLC35A2 encoding a UDP-galactose transporter cause early-onset epileptic encephalopathy. Hum Mut 2013;34:1708–1714.
13. Torkamani A, Bersell K, Jorge BS, et al. De novo KCNB1 mutations in epileptic encephalopathy. Ann Neurol 2014;76:529–540.
14. Ohba C, Shiina M, Tohyama J, et al. GRIN1 mutations cause encephalopathy with infantile-onset epilepsy, and hyperkinetic and stereotyped movement disorders. Epilepsia 2015;56:841–848.
15. Mastrangelo M. Novel genes of early-onset epileptic encephalopathies: from genotype to phenotypes. Pediatr Neurol 2015;53:119–129.
16. Deng X, Berletch JB, Nguyen DK, Disteche CM. X chromosome regulation: diverse patterns in development, tissues and disease. Nat Rev Genet 2014;15:367–378.
17. Tran Mau-Them F, Willems M, Albrecht B, et al. Expanding the phenotype of IQSEC2 mutations: truncating mutations in severe intellectual disability. Eur J Hum Genet 2014;22:289–292.
18. Carvill GL, Heavin SB, Yendle SC, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet 2013;45:825–830.
19. Freytag S, Gagnon-Bartsch J, Speed TP, Bahlo M. Systematic noise degrades gene co-expression signals but can be corrected. BMC Bioinform 2015;16:309.

© 2016 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.