ANNALS of Clinical and Translational Neurology
Open Access
ANA American Neurological Association since 1875

REPLY TO LETTER

Reply to: Is a microRNA-328 binding site in PAX6 associated with Rolandic epilepsy?

Lisa J. Strug[1,2] & Deb K. Pal[3,4]

[1]The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada
[2]Division of Biostatistics and Department of Statistics, The University of Toronto, Toronto, Canada
[3]Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
[4]King's College Hospital, London, United Kingdom

Correspondence
Lisa J. Strug, The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada. Tel: 416-813-7654 ext. 301762; Fax: (416) 813-8421; E-mail: lisa.strug@utoronto.ca

doi: 10.1002/acn3.403

We thank McGlade et al. for their interest in our recent publication,[1] where we identify an association between a *PAX6 3′ UTR* variant rs662702 - whose T allele disrupts the seed region of microRNA-328[2,3] - and EEG centrotemporal spikes (CTS) in Rolandic Epilepsy (RE). In their letter to the editor[4] they question the evidence for association with RE [sic], they conjecture that the association is due to inadequately accounted for multiple hypothesis testing and population stratification, and they present data from both public control populations and their own case study. While the control data demonstrate the importance of case–control matching as highlighted in Panjwani et al., their case data provide additional evidence supporting our original conclusion.

To summarize our findings reported in Panjwani et al.,[1] we analyzed the 11p13 CTS linkage locus[5] and provided association evidence between the T allele of rs662702 and CTS under an additive model, from two independent case–control ancestrally matched samples: each displaying similar estimated effect sizes with OR = 1.92 (95% CI: 1.08–3.43) in a North American sample, and OR = 2.20 (95% CI: 1.17–4.13) in a European sample. The findings showed that the T allele frequency is enriched in cases, and furthermore that homozygous individuals with the TT genotype display 12-fold increase in odds of CTS compared to individuals with the CC genotype.

McGlade et al. first remark on the significance threshold used in Panjwani et al., suggesting it falls short of *conventional genome-wide significance thresholds*. Based on the work by Dudbridge and Gusnanto,[6] we agree that a threshold of 5×10^{-8} is the convention for significance for whole-genome scans. However, Panjwani et al. conducted a region-specific study focusing on the 11p13 linkage locus, and McGlade et al. did not state a theoretical basis for using a genome-wide significance threshold for a region-specific analysis. In our previous whole-genome linkage scan,[5] we demonstrated genome-wide significance for linkage of CTS at the 11p13 locus, with a LOD score of 4.3. In Panjwani et al.[1] we aimed to fine-map the 11p13 locus to identify putatively causative SNPs for CTS, using a threshold of $P = 0.05/(\#$ independent SNPs tested in the region). The reported rs662702 passed that regional significance threshold for association with CTS.

McGlade et al. then propose the rs662702 association is due to *inadequately accounted for confounding*. The selection and matching of controls to cases is a vitally important aspect of this study and for genetic studies in general. For each of the two independent American and European case–control analyses, Panjwani et al. determined the ethnic distribution of the control sample using principal component analysis, then matched it to the case set to ensure their similarity in genetic ancestry. Given the variation in allele frequency at this marker across populations, as also observed by McGlade et al., supplemental Figure 1 of Panjwani et al.[1] shows principal component plots spanning 10 principal components, with the T allele carriers noted. We took several other steps to guard against population structure as an explanation for the observed association, some outlined in the publication, others not. As stated in the publication, we removed all participants who clustered with ancestral groups other than Europeans, and we adjusted for genetic ancestry using multiple principal components in the association regression model. Not included in the publication methods was the genomic control λ, estimated as 1.049 suggesting that there is no global case–control confounding. We also explored whether any principal component was associated with case-control status by regressing every

principal component on the CTS status ($P > 0.05$), which again indicated that the observed case–control difference could not be explained by global differences in genetic ancestry. Although no analysis ensures complete protection from confounding, we tested a comprehensive set of alternative explanations. Given the supporting evidence for a well-matched control group, the variation in rs662702 T allele frequency across ancestral groups suggests that the variant's role on CTS may be population-specific for Europeans.

In their letter, McGlade et al. reiterate our report of allele frequency differences across ancestries, and conjecture that *population-specific disease risk is unlikely because it has not been seen before in epilepsy*. Our finding is one of the first genetic associations of a sub-clinical trait in an epilepsy of complex inheritance, and little research has touched on this field to date. However, population-specific genetic risk has been reported for several traits including Alzheimer's,[7] renal disease,[8] risk for carbamazepine-induced Stevens-Johnson syndrome at the HLA locus,[9] and specifically for microRNA variation.[10]

McGlade et al. also present data on the rs662702 genotype distribution in their own sample ($n = 61$ with a diagnosis of childhood epilepsy with CTS or atypical childhood epilepsy with CTS in their Table 1) from Australia, Israel and New Zealand, of which 10 have the CT and one has the TT genotype with an estimated T frequency of 9.8%. Assuming the ancestry of these cases is similar to that of the two European control samples also presented in their Table 1, where the estimated T frequency is 5.7% in non-Finnish and 8.2% in Finnish, their data in fact support an increase in T allele frequency in cases. However, we note that details of genetic ancestry analyses of these groups were not provided, which limit what can be confidently interpreted from the data.

McGlade et al. further remark on their expectation that this variant contributing to CTS (population prevalence of 2–4%)[11] *should be out of Hardy–Weinberg equilibrium* (HWE) due to evolutionary selection pressures. From our perspective, we would not expect a variant contributing to CTS, a subclinical EEG trait, to be under the influence of any severe selection pressure and result in detectable departure from HWE. HWE can also be disrupted by natural selection or by non-random mating among other causes, but we would not predict any of these to apply to CTS.

McGlade et al. also report HWE p-values for publicly available population samples and for their case cohort in their Table 1. They report that none show departures from HWE. HWE analysis can, under certain circumstances, and particularly in a case sample, provide complementary information to association evidence. However, in the context of complex traits there are many situations where HW disequilibrium is never expected.[12] Here we also derive analytically how a variant contributing to a trait as one varies minor allele frequency of the risk allele (p) and penetrances (f_0, f_1, f_2 for $G = 0$, 1, 2 copies of the risk allele) need not be out of detectable HWE in either cases or controls.

Assume HWE holds in the population, the population prevalence of the trait (CTS in our case) is

$$k = P(\text{CTS} = 1) = \sum_{i=0}^{2} P(\text{CTS} = 1|G = i)P(G = i) = f_0(1-p)^2 + 2f_1p(1-p) + f_2p^2$$

Let $P(G|\text{CTS} = 1)$ represent the distribution of the genotypes in the case population, then

$$P(G|CTS = 1) = \frac{P(\text{CTS} = 1|G)P(G)}{P(\text{CTS} = 1)} = \frac{f_G P(G)}{k}$$

Specifically, genotype probabilities in the CTS case group for CC, CT and TT are, respectively,

$$\frac{f_0}{k}(1-p)^2, \frac{f_1}{k}2p(1-p) \text{ and } \frac{f_2}{k}p^2$$

Thus, one can determine if HWE holds for different combinations of p, f_0, f_1 and f_2 in cases and controls by assessing whether

$$\sqrt{\frac{f_2}{k}p^2} + \sqrt{\frac{f_0}{k}(1-p)^2} \approx 1$$

Applying this formulation in R[13] using a minor allele frequency of 8% and penetrance values of 0.2, 0.3 and 0.5 shows no departure in HWE in either cases or controls. As one varies these parameters, cases or controls can be made to display departures assuming adequate sample size.

We welcome the spotlight on research methods in the complex genetics of common epilepsies, since these have been understudied during the exome epoch. Very large patient cohorts are undoubtedly necessary for the coming era of genome-wide syndrome-specific studies.[14] We invite the epilepsy research community to join an international consortium to assemble a clinical-genetic dataset of 3000 RE individuals (REGAIN study, opening 2017 at www.childhood-epilepsy.org). This would allow us to confirm the role of *PAX6* regulatory variants in CTS, and to identify genetic contributors to other components of RE including seizures and neurodevelopmental traits. We would also be able to explore hitherto uncharted territory in epilepsy genetics, such as potential ancestry-specific genetic risk factors and the genetic architecture of a complex epilepsy.

Conflict of Interest

D.K.P is a scientific advisor to Amplexa Genetics. L.J.S declares no conflicts of interest.

References

1. Panjwani N, Wilson MD, Addis L, et al. A microRNA-328 binding site in PAX6 is associated with centrotemporal spikes of rolandic epilepsy. Ann Clin Transl Neurol 2016;3:512–522.
2. Chen KC, Hsi E, Hu CY, et al. MicroRNA-328 may influence myopia development by mediating the PAX6 gene. Invest Ophthalmol Vis Sci 2012;53: 2732–2739.
3. Liang CL, Hsi E, Chen KC, et al. A functional polymorphism at 3′UTR of the PAX6 gene may confer risk for extreme myopia in the Chinese. Invest Ophthalmol Vis Sci 2011;52:3500–3505.
4. McGlade A, Myers KA, Berkovic SF, et al. Is a microRNA-328 binding site in *PAX6* associated with Rolandic epilepsy?. Ann Clin Transl Neurol 2017;doi: 10. 1002/acn3.401.
5. Strug LJ, Clarke T, Chiang T, et al. Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4). Eur J Hum Genet 2009;17:1171–1181.
6. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol 2008;32:227–234.
7. Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nat Rev Neurol 2013;9:106–118.
8. Kao WH, Klag MJ, Meoni LA, et al. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. Nat Genet 2008;40:1185–1192.
9. McCormack M, Alfirevic A, Bourgeois S, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. N Engl J Med 2011;364:1134–1143.
10. Rawlings-Goss RA, Campbell MC, Tishkoff SA. Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. BMC Med Genomics 2014;28:53.
11. Eeg-Olofsson O, Petersen I, Sellden U. The development of the electroencephalogram in normal children from the age of 1 through 15 years. Paroxysmal activity. Neuropadiatrie 1971;2:375–404.
12. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:967–986.
13. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015; http://www.R-project.org/.
14. Pal DK, Strug LJ. The genetics of common epilepsies: common or distinct? Lancet Neurol 2014;13:859–860.

