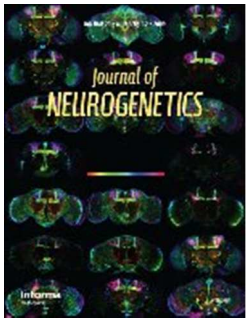

RE Six novel rare non-synonymous mutations for migraine without aura identified by exome sequencing

Journal:	*Journal of Neurogenetics*
Manuscript ID	GNEG-2017-0027.R1
Manuscript Type:	Letter to the Editor
Date Submitted by the Author:	22-Aug-2017
Complete List of Authors:	Albury, Cassie; Queensland University of Technology, Institute of Health and Biomedical Innovation Gerring, Zachary; Queensland University of Technology, Institute of Health and Biomedical Innovation Griffiths, Lyn; Queensland University of Technology, Institute of Health and Biomedical Innovation Nyholt, Dale; Queensland University of Technology, Institute of Health and Biomedical Innovation Rodriguez Acevedo, Astrid; Xavier University Medical School
Keywords:	whole exome sequencing, migraine, rare variants, family genetic studies

Cassie L Albury
Professor Lyn R Griffiths, PhD.
Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

Zachary Gerring
Associate Professor Dale R Nyholt, PhD.
Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

Astrid J Rodriguez-Acevedo, PhD.
Xavier University School of Medicine, Oranjestad, Aruba.

Shortened Title: RE novel rare mutations for migraine without aura

Corresponding author:

Astrid J Rodriguez-Acevedo, PhD.
Xavier University School of Medicine
t: 011-297-588-7766
Santa Helenastraat #23
Oranjestad, Aruba
astrid.rodriguez2409@gmail.com

RE Six novel rare non-synonymous mutations for migraine without aura identified by exome sequencing

Cassie L. Albury[1], Zachary F. Gerring[2], Lyn R. Griffiths[1], Dale R. Nyholt[2] and Astrid J. Rodriguez-Acevedo[3]

1. Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
2. Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
3. Xavier University School of Medicine, Oranjestad, Aruba.

Dear Dr Chun-Fang Wu,

A recent study by Jiang *et al.* published by the *Journal of Neurogenetics* claims the identification of six novel rare mutations involved in the aetiology of migraine without aura (MWO) (Jiang *et al*., 2015). The authors performed whole-exome sequencing (WES) in a Chinese sample of four related cases (father, two sons, and one daughter) and four unrelated controls. The importance of the two variants *UBE2NL* T266G and *EDAR2* G170A, located on the X chromosome is stressed as an endorsement to the observed sexual dimorphism in the presentation of migraine (i.e., females have up to three times higher risk than males) (Bigal & Lipton, 2009). However, upon review of this publication we have some concerns regarding the study's main conclusions. As a team of researchers working on unveiling the genetic causes of migraine, we feel it is necessary to pinpoint the reasons why the results presented in the Jiang *et al.* study may be problematic. In doing so, we hope to promote scientific progress and make suggestions to further strengthen the Jiang *et al.* study.

Firstly, the study design in Jiang *et al.* describes the use of four related cases and four unrelated controls to detect rare novel mutations underlying MWO. The authors correctly state that family-based designs can substantially increase the detection rate of disease-causing mutations (Li, Liu, & Leal, 2013) because they are both robust to population stratification and can increase power to detect low-frequency large-effect risk variants. However, the study design in Jiang *et al*. appears to treat each related case as an unrelated proband and compares them to unrelated controls, thus making the analyses no longer robust to population stratification in addition to failing to adjust for the fact the four cases are not independent. The authors would need to compare the exome data of the four affected family members with the unaffected members within the pedigree (using a within-family test), in order to meet the requirements and benefit from the stipulated advantages of a family-based association study. Such a within-family analysis has potential to identify a novel causal variant segregating with migraine status, which if not typically observed in the general population (i.e., has a very low frequency), has potential to be causal for migraine in the tested family.

Secondly, the authors report the "novel" and "rare" allele frequency (minor allele frequency [MAF] ≤ 0.01) and potential pathogenic effect of the *UBE2NL*-T266G (rs237520), *CLCNKB*-A1225G (rs2275166), *EMR1*-C264G (rs330880), *EDA2R*-G170A (rs1385699), *GBP2*-A907G (rs2230338), and *ARHGAP28*-C413G (rs6506448) variants. However, searching the 1000 Genomes Browser (http://asia.ensembl.org/Homo_sapiens/Info/Index) shows that the MAF of these variants range from 0.194 to 0.499 in the 1000 Genomes Project East Asian (EAS) population (comprising 504 individuals). We suggest this discrepancy reflects the poor sensitivity provided by the authors' use of only four controls to screen mutations. In addition, when each of the six variants were examined using the PredictSNP2 web-tool (Bendl *et al*., 2016), which combines the five best performing prediction methods currently available to assess the tolerability or damaging disposition of a queried single

nucleotide polymorphism (SNP) variant, PredictSNP2 predicted the aforementioned variants to be neutral (i.e. non-pathogenic) (Table 1). Hence, the variants reported by Jiang *et al.* appear to be common in the East Asian population and are not predicted to have a functional impact.

Lastly, genotypes corresponding to *EDA2R*-G170A and *UBE2NL*-T266G variants on the X chromosome are displayed in the manuscripts' results section as homozygous and heterozygous, respectively, for male cases. However, it is expected that in normal males, genotypes corresponding to variants located in X-chromosomal regions other than those in the pseudoautosomal regions (PARs), must be hemizygous. Given neither *EDA2R*-G170A and *UBE2NL*-T266G are located in the PARs, the respective genotypes should be checked and amended.

Considering the above issues, we encourage Jiang and colleagues to revisit their data and consider alternative approaches, with a stronger bioinformatics component, to identify novel causal variants segregating with migraine status. For example, an efficient use of their WES data would be to first search for novel (predicted) pathogenic variants that are shared by the related MWO cases but that are not reported (or at least rare with MAF < 0.01) in large public datasets such as the 1000 Genomes Project data, or preferably the more recent and comprehensive Exome Aggregation Consortium (ExAC) database (http://exac.broadinstitute.org/), and Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/). Such shared rare novel pathogenic variants should then be specifically examined within the entire MWO pedigree to test whether they segregate with migraine status (i.e., are not carried by the non-migraine pedigree members) to support their potential causality and subsequent investigation via direct functional studies.

The analysis of common and rare variants in complex disease requires specific considerations to avoid spurious results. These considerations have been reviewed (Gilissen, Hoischen, Brunner, & Veltman, 2012), and they should be increasingly discussed as the accessibility to technologies like WES continues to grow. This is a call to the scientific community to join efforts in multi-disciplinary teams addressing the statistical and bioinformatics challenges imposed by the new technologies to advance our understanding of migraine aetiology.

Bibliography

Bendl, J., Musil, M., Stourac, J., Zendulka, J., Damborsky, J., & Brezovsky, J. (2016). PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS Comput Biol, 12*(5), e1004962. doi: 10.1371/journal.pcbi.1004962

Bigal, M. E., & Lipton, R. B. (2009). The epidemiology, burden, and comorbidities of migraine. *Neurol Clin, 27*(2), 321-334. doi: 10.1016/j.ncl.2008.11.011

Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet, 20*(5), 490-497. doi: 10.1038/ejhg.2011.258

Jiang, Y., Wu, R., Chen, C., You, Z. F., Luo, X., & Wang, X. P. (2015). Six novel rare non-synonymous mutations for migraine without aura identified by exome sequencing. *J Neurogenet, 29*(4), 188-194. doi: 10.3109/01677063.2015.1122787

Li, B., Liu, D. J., & Leal, S. M. (2013). Identifying rare variants associated with complex traits via sequencing. *Curr Protoc Hum Genet, Chapter 1*, Unit 1 26. doi: 10.1002/0471142905.hg0126s78

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., . . . Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature, 461*(7261), 272-276. doi: 10.1038/nature08250

Table 1. Frequency and SNP prediction effect. From left to right this table shows the mutation, the hg38 chromosome and position (Chr:pos), the frequency of the alternative allele in the East Asian (EAS) population according to 1000 Genomes Browser and the expected accuracy (%) of the SNP prediction effect analysis performed by different prediction tools integrated by PredictSNP2. The PredictSNP2 is a consensus classifier that combines five best performing prediction methods implemented in CADD (Combined Annotation Dependent Depletion) , DANN (Deleterious Annotation of genetic variants using Neural Networks), FATHMM (Functional Analysis through Hidden Markov Models) , FunSeq2 and GWAVA (Genome Wide Annotation of Variants). For further details on the mathematical approach and reliability of these methods, the reader is refereed to Bendl et al, 2016. The Predicted effect is color-coded: neutral variants are in light grey, deleterious variants are in dark grey, unknown/uncertain in white.

Mutation	Chr:pos	Freq	PredictSNP2	CADD	DANN	FATHMM	FunSeq2	GWAVA
CLCNKB-A1225G	1:16053748	0.81 (G)	**89%**	67%	90%	93%	84%	55%
GBP2-A907G	1:89114258	0.78 (C)	**89%**	52%	60%	88%	62%	52%
ARHGAP28-C413G	18:6851059	0.27 (G)	**89%**	92%	94%	89%	62%	56%
EMR1-C264G	19:6897453	0.48 (G)	**89%**	94%	94%	93%	84%	56%
EDA2R-G170A	X:66605144	1.00 (T)	**89%**	56%	89%	77%	62%	?
UBE2NL-T266G	X:143884366	0.64 (G)	**60%**	?	77%	66%	81%	71%