Journal of Human Genetics (2012) 57, 403–404

www.nature.com/jhg

COMMENTARY

An abundance of population-specific monomorphic SNPs may or may not be meaningful: a commentary on differences in allele frequencies of familial hypercholesterolemia SNPs in the Malaysian population

Wanna Thongnoppakhun, Anunchai Assawamakin and Sissades Tongsima

Journal of Human Genetics (2012) **57,** 403–404; doi:10.1038/jhg.2012.52; published online 31 May 2012

In previous issue of the *Journal of Human Genetics*, Alex *et al.*[1] report the differences in allele frequencies of familial hypercholesterolemia (FH) single-nucleotide polymorphisms (SNPs) in the Malaysian population. As one of the most common inherited disorders in humans, autosomal dominant hypercholesterolemia (ADH) (MIM #143890) or its more common name, FH is caused predominantly by mutations in *LDLR* in the majority of patients. Additionally, less strong but important associations were also identified for *ApoE* and *PCSK9*. Either heterozygous or homozygous mutations in these genes undoubtedly led to hypercholesterolemia ranging from mild-to-severe phenotypes. However, these findings underplay several previous studies highlighting the importance of SNPs as genetic modifying factors in ADH, which affect cholesterol regulation differently across populations (reviewed in Fahed and Nemer[2]).

The article by Alex *et al.* aimed to identify ADH variants in self-identified multiethnic Malaysians including 140 clinically diagnosed FH patients and 111 healthy controls by evaluating 310 previously reported ADH-associated SNPs mainly on the three ADH-causing genes and a few genes involved in the cholesterol metabolic pathways. These SNPs under interrogation comprised either causative SNPs or merely linked to ADH SNPs (as a signpost), whereas multiple-nucleotide polymorphisms, as well as structural variations, were excluded from this study. Furthermore, the genotyping platform was a custom-designed microarray in which the designed probes were based on common SNPs from three databases, namely BHF, dbSNP and SNPedia.

The data presented by the authors were interesting, that is, in both subject groups 44.1% (137/310) monomorphic (mono-allelic) SNPs were found. Their report is very similar in terms of number on other genes in African Americans or European Americans[3] and the Danish population[4]. However, the high proportion of invariable SNPs observed could be caused either by a rare haplotype presented in the Human Genome Project or possibly by genotyping errors, frequently found in the SNP database.[5] Recently, 8.32% of biallelic coding SNPs in dbSNP were classified as artifacts, called 'single nucleotide differences' (SNDs), which result from highly similar (paralogous) genes.[6] *LDLR*, which has several paralogs including *VLDLR*, *LRP*, *LRP1B*, *LRP8* and *LRP1* (www.genecards.org), is the only paralogous gene reported in Alex *et al*'s study. Given that the paralogs of *LDLR* potentially introduce SNDs, but only small number of monomorphic SNPs (28 out of 137) were reported, when comparing with other reported non-paralogous genes, *APOB* (73 SNPs) and *PCSK9* (36 SNPs).

The authors, however, claimed that 23 monomorphic SNPs in Malaysians are truly unique as these SNPs in other populations published elsewhere were reported as polymorphic. For example, the authors reported rs61318752, which is a SNP in *LDLR*, as a monomorphic. Considering that some Malaysians are of Chinese descent, on the contrary, this rs61318752 SNP was previously reported for Han Chinese with MAF of 11.4%. Even though population-specific monomorphic sites could be explained by either low mutation rate or positive natural selection, before making such a conclusion, it is imperative to thoroughly consider the study populations, patient recruitment, SNP selection criteria, reliability of genotyping methods applied and publicly available SNP databases used for comparison.

The research population studied by Alex *et al.* was 251 Malaysians consisting of Malay (61.8%), Indian (22.7%) and Chinese (15.5%). Such samples are of Southeast Asian ancestry in which the data could provide a better understanding of genetic factors influencing ADH by the race than those reported with European ancestry (reviewed in Fahed and Nemer[2]). However, using larger samples that are ethnically homogeneous would be more appropriate as the allele frequency estimated from small groups of controls deviates more from the true frequency than those from the larger ones.[7] Alex *et al.*

W Thongnoppakhun is at the Department of Research and Development, Division of Molecular Genetics, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand and A Assawamakin is at the National Center for Genetic Engineering and Biotechnology (BIOTEC), Pathum Thani, Thailand and S Tongsima is at the National Center for Genetic Engineering and Biotechnology (BIOTEC), Pathum Thani, Thailand.
E-mail: wanna.tho@mahidol.ac.th.

recruited the ADH patients on the basis of the commonly used Dutch Lipid Clinic Network criteria.[8] The SNP selection criteria were comprehensive and also based on effects of SNPs to the disease information collected and curated in SNPedia (http://www.SNPedia.com). Nonetheless, the variant causality in the analysis can be frequently misclassified by either inclusion of non-causal variants or exclusion of causal variants[9]. Genotyping results by Alex *et al.* were validated by direct sequencing of randomly selected samples. The authors compared their results with multiple SNP databases including dbSNP, HAPMAP, mutDB and BHF to assess the uniqueness of the SNPs uncovered across worldwide populations.[10] It would be better if a recent database of Pan-Asian population, PanSNPdb[11], was also taken into account.

The presence of monomorphic SNPs was rarely mentioned and analyzed in previous reports. This is possibly due to publication bias; positive findings were published, whereas negative ones were ignored. Alex *et al.* presented the first evaluation of population differentiation in allele frequencies for ADH-associated variants in Malaysians (as indicated by high Fst levels of 0.1285–0.3552), contributing to ADH molecular research. The authors should not prematurely draw the conclusion that these SNPs could be the targets of positive selection using Fst values.[12] However, further studies of the disease in Malaysians and possibly other Southeast Asian populations could utilize these data to select informative variants for either disease-associated studies or molecular screening for preventive medicine in the future.

ACKNOWLEDGEMENTS

WT is supported by 'Chalermphrakiat' Grant, Faculty of Medicine Siriraj Hospital, Mahidol University. ST is supported in part by the Thailand Research Fund Grant number RSA5480026.

1 Alex, L., Chahil, J. K., Lye, S. H., Bagali, P. & Ler, L. W. Differences in allele frequencies of autosomal dominant hypercholesterolemia SNPs in the Malaysian population. *J. Hum. Genet.* **57,** 358–362 (2012).
2 Fahed, A. C. & Nemer, G. M. Familial hypercholesterolemia: the lipids or the genes? *Nutr. Metab.* **8,** 23 (2011).
3 Carlson, C. S., Eberle, M. A, Rieder, M. J., Smith, J. D., Kruglyak, L. & Nickerson, D. A. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat. Genet.* **33,** 518–521 (2003).
4 Gaustadnes, M., Ørntoft, T. F., Jensen, J. L. & Torring, N. Validation of the use of DNA pools and primer extension in association studies of sporadic colorectal cancer for selection of candidate SNPs. *Hum. Mutat.* **27,** 187–194 (2006).
5 Mitchell, A. A., Zwick, M. E., Chakravarti, A. & Cutler, D. J. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. *Bioinformatics* **20,** 1022–1032 (2004).
6 Musumeci, L., Arthur, J.W., Cheung, F. S., Hoque, A., Lippman, S. & Reichardt, J. K. Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. *Hum. Mutat.* **31,** 67–73 (2010).
7 Taioli, E., Pedotti, P. & Garte, S. Importance of allele frequency estimates in epidemiological studies. *Mutat. Res.* **567,** 63–70 (2004).
8 Civeira, F. Guidelines for the diagnosis and management of heterozygous familial hypercholesterolaemia. *Atherosclerosis* **173,** 55–68 (2004).
9 Liu, D. J. & Leal, S. M. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS. Genet.* **6,** e1001156 (2010).
10 Doron, S & Shweiki, D. SNP Uniqueness Problem: a proof-of-principle in HapMap SNPs. *Hum. Mutat.* **32,** 355–357 (2011).
11 Ngamphiw, C., Assawamakin, A., Xu, S., Shaw, P. J., Yang, J. O., Ghang, H. *et al.* PanSNPdb: the Pan-Asian SNP genotyping database. *PLoS. One* **6,** e21451 (2011).
12 Gardner, M., Williamson, S., Casals, F., Bosch, E., Navarro, A., Calafell, F. *et al.* Extreme individual marker FST values do not imply population-specific selection in humans: the NRG1 example. *Hum. Genet.* **121,** 759–762 (2007).