

Was *ADH1B* under Selection in European Populations?

To the Editor: Recently, Galinsky et al.¹ developed the software FastPCA on the basis of principal-component analysis (PCA) to investigate genome-wide population differentiations and selective signals. They applied FastPCA to the allele-frequency data of the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort,² which contains 54,734 European Americans. They detected a genome-wide-significant selection signal at rs1229984 (g.100239319C>T [GenBank: NC_000004.11]) in alcohol dehydrogenase 1B (*ADH1B* [MIM: 103720]) in the GERA cohort on the basis of principal component 1 (PC1). Because the selection for rs1229984*T (p.Arg48His [GenBank: NP_000659.2]) was previously identified in East Asians,³ they proposed that *ADH1B* could serve as an example of convergent evolution in Europe and East Asia.

The geographic frequency of the rs1229984*T (p.Arg48His) allele had two peak distributions in East Asians

(>40%) and West Asians (~30%).^{4,5} In Europe, it was distributed substantially in southern (~10%) and eastern (~5%) regions but was rare (<5%) or even absent in central, western, and northern regions.⁴⁻⁶ The average low frequency raises suspicion for the conclusion that rs1229984*T (p.Arg48His) in *ADH1B* is under selection in Europe. To address this issue, we ran FastPCA for five representative European populations (i.e., Utah residents with ancestry from northern and western Europe from the CEPH collection [CEU], Finnish in Finland [FIN], British in England and Scotland [GBR], Iberian Population in Spain [IBS], and Toscani in Italia [TSI]) from the 1000 Genomes Project⁷ because data on the GERA cohort were not available to us. After quality control as described in Galinsky et al.,¹ a total of 8,539,702 SNPs (with minor allele frequency > 0.01) from 503 European individuals were analyzed. The previously reported signals of selection at *LCT* (MIM: 603202) (rs6754311 [g.136707982T>C] [GenBank: NC_000002.11]) and rs4988235 [g.136608646G>A] [GenBank: NC_000002.11]) and *OCA2* (MIM: 611409) (rs12916300 [g.28165345C>T] [GenBank: NC_000015.10]) could be

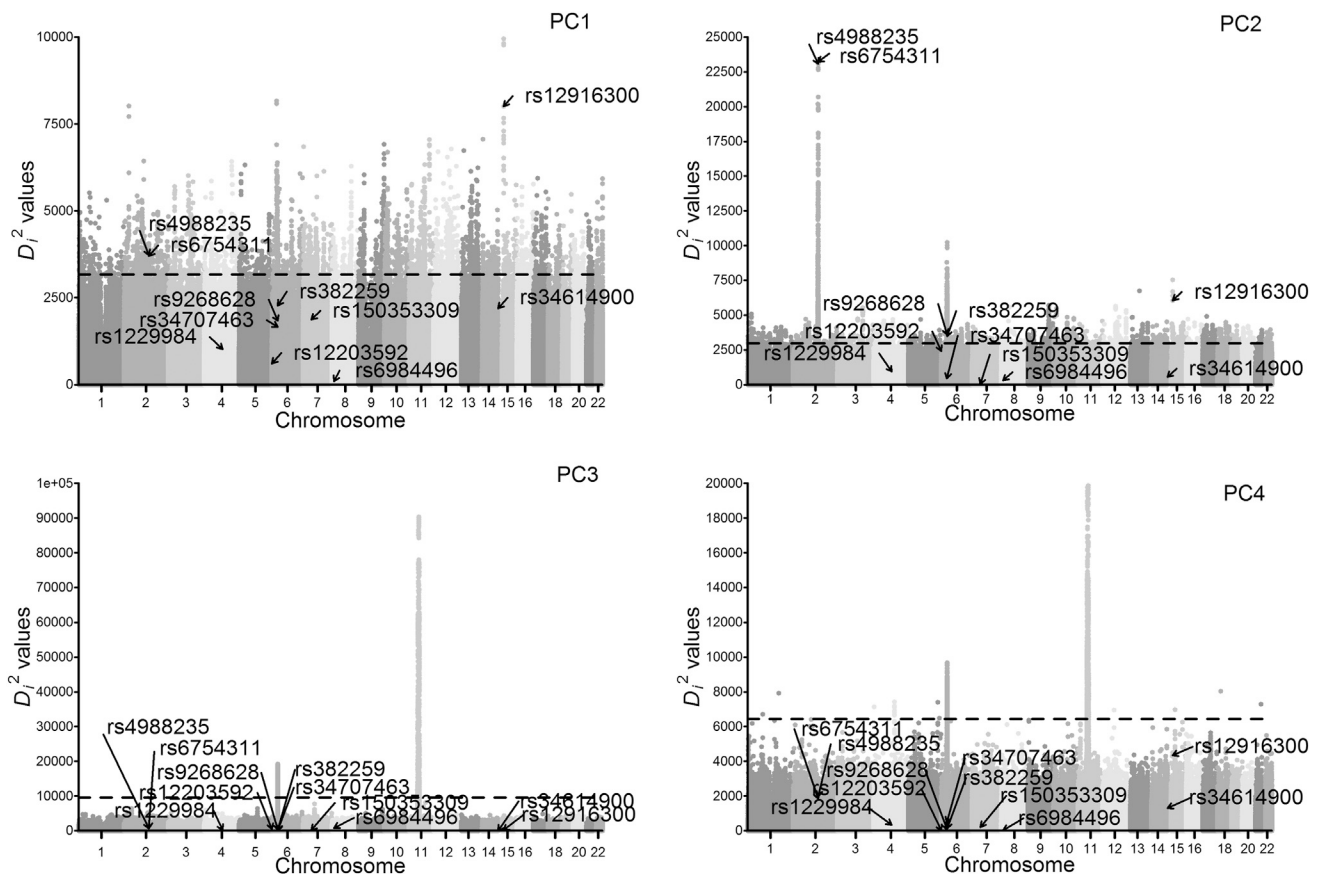


Figure 1. Signals of Selection in the Top Four PCs of European Populations from the 1000 Genomes Project

The vertical axis is the D_i^2 statistic following a chi-square (with 1 df) distribution on the basis of the PCA result. For most large D_i^2 values, it is impossible to calculate the exact corresponding p values (in most cases, $p = 0$), so we adopted the top values to show the results. The black dashed line indicates the top 0.1% threshold.

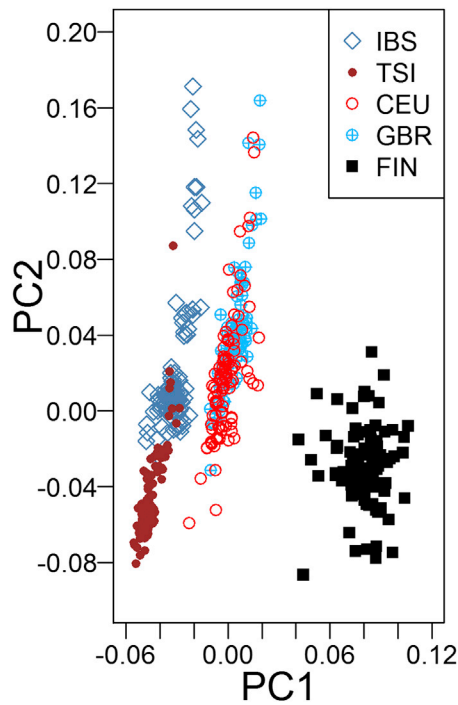


Figure 2. PCA Results on European Populations from the 1000 Genomes Project

We preformed quality control, a Hardy-Weinberg equilibrium test, and linkage-disequilibrium filtering as described in Galinsky et al.¹ by using PLINK 1.9.⁸ After that, a total of 1,063,058 SNPs were analyzed in FastPCA.

detected in both PC1 and PC2,¹ but no selection signal at *ADH1B* (rs1229984) was found in the top four PCs (Figure 1). Although the power of the PC-based selection statistic was decreased as a result of the relatively small sample size ($n = 503$) in comparison with that of the GERA cohort ($n = 54,734$), given the average low frequency (<10%) of rs1229984*T (p.Arg48His) in Europe,⁴⁻⁶ increasing the sample size seems unable to improve the statistic power according to the simulation results (cf. Figure 3 in the original paper¹). In addition, we checked the PCA results based on a total of 1,063,058 SNPs (Figure 2). FIN was separated from other populations by PC1. CEU was clustered with the GBR but diverged from IBS and TSI. The patterns generally agreed with previous results,^{9,10} suggesting that our running of FastPCA was unlikely to be an artifact. Therefore, the significant signal of rs1229984*T (p.Arg48His) as identified by Galinsky et al.¹ might not be the result of regional selection in Europe.

The signal of rs1229984*T (p.Arg48His) was initially inferred from its unusual allele-frequency distribution among different populations in the GERA cohort,¹ which contains 2,750 Ashkenazi Jewish individuals.² Indeed, the allele frequency of rs1229984*T (p.Arg48His) was higher in Ashkenazi Jews (21%) than in the other non-Jewish European populations (i.e., <3% in the Irish and northern Europeans, ~5% in eastern Europeans, and ~10% in southeastern Europeans; Table S6 in the original paper¹). Although Ashkenazi Jews were recruited as European

Americans in the GERA cohort because of their geographic proximity,² many studies revealed genetic differences that distinguished Ashkenazi Jews from non-Jewish Europeans.¹¹ Galinsky et al. also showed this differentiation in their PCA (cf. Figure 4 in the original paper¹). Some genome-wide analyses revealed that Ashkenazi Jews have a common West Asian origin^{12,13} but also suggest that Ashkenazi Jews have European admixture.^{14,15} We recommend re-analyzing the GERA cohort without the samples identified as Ashkenazi Jews to determine whether the signal of selection remains. Previous analysis of *ADH1B* haplotypes had classified most Ashkenazi Jewish individuals with rs1229984*T (p.Arg48His) into haplotype H5, which was mainly distributed in West Asians.¹⁶ Haplotype H5 is generally rare (~1%) in European populations (cf. Table S8 in the original paper¹). Collectively, it is most likely that the high frequency of rs1229984*T (p.Arg48His) in *ADH1B* in Ashkenazi Jews was attributed to their ancestry components of West Asian origin rather than the result of a parallel selection in Europeans.

Interestingly, Jews have lower rates of alcohol dependence (MIM: 103780) and alcohol-related problems than other European populations,¹⁷ and rs1229984*T (p.Arg48His) has been reported to be associated with both unpleasant reactions after alcohol consumption and the frequency of alcohol consumption in Ashkenazi Jews.¹⁸ A previous study identified the selective signal at aldehyde dehydrogenase 2 family (mitochondrial) (*ALDH2* [MIM: 100650]), but not at *ADH1B* in the Ashkenazi Jewish population.¹⁵ More detailed work is required to discern the evolutionary history of *ADH1B* in Ashkenazi Jews as well as in other West Asian populations.

Because of its high computing speed and statistical power, FastPCA will serve as a useful tool in analyzing very large datasets.¹ Nevertheless, population structure, gene flow, and other demographic histories are likely to exist in large populations, especially in meta-populations of multiple ethnic groups with various genetic ancestries (e.g., the GERA cohort). In particular, the evidence of genetic admixture in populations living in the Americas^{19,20} could raise the possibility that African and/or Native American ancestry also exists in some individuals self-reported as European Americans. Those demographic factors can generate “unusual” allele-frequency differences in detecting signals of selection and should be taken into account in the interpretation of the observed selection signals.^{21,22}

Quan-Kuan Shen,^{1,2,3,7} Xierzhatijiang Sulaiman,^{4,7}
Yong-Gang Yao,^{2,3,5} Min-Sheng Peng,^{1,2,3,*}
and Ya-Ping Zhang^{1,2,3,6,*}

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; ²KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming 650223, China; ³Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China; ⁴College of Uyghur Medicine, Xinjiang Medical University, Ürümqi 830011,

China; ⁵Key Laboratory of Animal Models and Human Disease Mechanisms, Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming 650223, China; ⁶State Key Laboratory for Conservation and Utilization of Bio-Resources, Yunnan University, Kunming 650091, China

⁷These authors contributed equally to this work

*Correspondence: pengminsheng@mail.kiz.ac.cn (M.-S.P.), zhangyp@mail.kiz.ac.cn (Y.-P.Z.)

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China and the Bureau of Science and Technology of Yunnan Province. M.-S.P. appreciates the support from the Youth Innovation Promotion Association of the Chinese Academy of Sciences.

Web Resources

1000 Genomes Project, <http://www.1000genomes.org/>
EIGENSOFT 6.1.1, <https://data.broadinstitute.org/alkesgroup/EIGENSOFT/>
PLINK 1.9, <https://www.cog-genomics.org/plink2/>

References

- Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast principal-component analysis reveals convergent evolution of *ADH1B* in Europe and East Asia. *Am. J. Hum. Genet.* *98*, 456–472.
- Banda, Y., Kvale, M.N., Hoffmann, T.J., Hesselton, S.E., Rana-tunga, D., Tang, H., Sabatti, C., Croen, L.A., Dispensa, B.P., Henderson, M., et al. (2015). Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* *200*, 1285–1295.
- Peng, Y., Shi, H., Qi, X.B., Xiao, C.J., Zhong, H., Ma, R.L., and Su, B. (2010). The *ADH1B Arg47His* polymorphism in east Asian populations and expansion of rice domestication in history. *BMC Evol. Biol.* *10*, 15.
- Li, H., Mukherjee, N., Soundararajan, U., Tarnok, Z., Barta, C., Khaliq, S., Mohyuddin, A., Kajuna, S.L., Mehdi, S.Q., Kidd, J.R., and Kidd, K.K. (2007). Geographically separate increases in the frequency of the derived *ADH1B*47His* allele in eastern and western Asia. *Am. J. Hum. Genet.* *81*, 842–846.
- Borinskaya, S., Kal'ina, N., Marusin, A., Faskhutdinova, G., Morozova, I., Kutuev, I., Koshechkin, V., Khusnutdinova, E., Stepanov, V., Puzyrev, V., et al. (2009). Distribution of the alcohol dehydrogenase *ADH1B*47His* allele in Eurasia. *Am. J. Hum. Genet.* *84*, 89–92, author reply 92–94.
- Treutlein, J., Frank, J., Kiefer, F., and Rietschel, M. (2014). *ADH1B Arg48His* allele frequency map: filling in the gap for Central Europe. *Biol. Psychiatry* *75*, e15.
- Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* *526*, 68–74.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* *4*, 7.
- Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. *Nature* *456*, 98–101.
- Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., et al. (2008). Correlation between genetic and geographic structure in Europe. *Curr. Biol.* *18*, 1241–1248.
- Ostrer, H., and Skorecki, K. (2013). The population genetics of the Jewish people. *Hum. Genet.* *132*, 119–127.
- Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. *Nature* *466*, 238–242.
- Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P.F., Morrow, B., Friedman, E., Oddoux, C., Burns, E., and Ostrer, H. (2010). Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am. J. Hum. Genet.* *86*, 850–859.
- Carmi, S., Hui, K.Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., et al. (2014). Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.* *5*, 4835.
- Bray, S.M., Mulle, J.G., Dodd, A.F., Pulver, A.E., Wooding, S., and Warren, S.T. (2010). Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proc. Natl. Acad. Sci. USA* *107*, 16222–16227.
- Li, H., Gu, S., Han, Y., Xu, Z., Pakstis, A.J., Jin, L., Kidd, J.R., and Kidd, K.K. (2011). Diversification of the *ADH1B* gene during expansion of modern humans. *Ann. Hum. Genet.* *75*, 497–507.
- Monteiro, M.G., and Schuckit, M.A. (1989). Alcohol, drug, and mental health problems among Jewish and Christian men at a university. *Am. J. Drug Alcohol Abuse* *15*, 403–412.
- Carr, L.G., Foroud, T., Stewart, T., Castelluccio, P., Edenberg, H.J., and Li, T.K. (2002). Influence of *ADH1B* polymorphism on alcohol use and its subjective effects in a Jewish population. *Am. J. Med. Genet.* *112*, 138–143.
- Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Er-rington, J., Blot, W.J., Bustamante, C.D., Kenny, E.E., Williams, S.M., Aldrich, M.C., and Gravel, S. (2016). The Great Migration and African-American genomic diversity. *PLoS Genet.* *12*, e1006059.
- Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. *Nature* *488*, 370–374.
- Scheinfeldt, L.B., and Tishkoff, S.A. (2013). Recent human adaptation: genomic approaches, interpretation and insights. *Nat. Rev. Genet.* *14*, 692–702.
- Sousa, V., Peischl, S., and Excoffier, L. (2014). Impact of range expansions on current human genomic diversity. *Curr. Opin. Genet. Dev.* *29*, 22–30.

<http://dx.doi.org/10.1016/j.ajhg.2016.09.017>

© 2016 American Society of Human Genetics.