



Response to Comments by Timpson et al. and Yu et al .

Nitzan Mekel-Bobrov, *et al.*
Science **317**, 1036b (2007);
DOI: 10.1126/science.1143658

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 17, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/317/5841/1036b>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/317/5841/1036b#related-content>

This article **cites 19 articles**, 9 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/317/5841/1036b#otherarticles>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Response to Comments by Timpson *et al.* and Yu *et al.*

Nitzan Mekel-Bobrov^{1,2} and Bruce T. Lahn^{1*}

The lack of association of the *abnormal spindle-like microcephaly associated (ASPM)* and *Microcephalin (MCPHI)* genes with brain size or intelligence described by Timpson *et al.* has been reported previously, including by our own group. Moreover, as in most studies of selection, our analyses were entirely independent of phenotypic association. We also respond to the previously published comment by Yu *et al.*, which argued that *ASPM* has not undergone positive selection.

We previously reported evidence of positive selection on the genes *abnormal spindle-like microcephaly associated (ASPM)* and *Microcephalin (MCPHI)* (1, 2). Timpson *et al.* (3) set out to address speculations regarding the role of cognition in driving the adaptive evolution of these genes. To do so, they tested for statistical association between the alleles of these genes and normal variation in several cognition-related phenotypes. We generally agree that this is a good strategy for examining the phenotypic consequences of a signature of ongoing selection. Consequently, we carried out a similar large-scale study ourselves, looking at the association between intelligence and the *ASPM* and *MCPHI* alleles, with similar findings already reported earlier this year (4). Furthermore, Timpson *et al.*'s finding of no association with brain size has already been reported by Woods *et al.* (5), and a smaller study by Rushton *et al.* further corroborated both our own and Woods *et al.*'s findings (6). Timpson *et al.*'s inclusion of various other phenotypes, such as waist size and weight, does not add any substantive value to these three previous studies, because any random trait among the nearly infinite possible traits one can measure is not likely to show an association with a given gene of interest.

Our greater concern is that Timpson *et al.*'s comment seems to confuse our report of a genetic signature of selection, which has to do with patterns of genetic variation, with claims regarding specific phenotypic adaptations. Studies of positive selection are rarely based on phenotypic data. Although our ultimate goal is to uncover the phenotype driving the adaptive evolution of a gene, an analysis of genetic variation and divergence is a necessary first step to identify a history of selection. Although a selective event is driven by a phenotypic trait, signatures of adaptation occur at the level of a genotype. This is why a signature of selection is by definition a genetic

phenomenon and, by extension, why findings of selection are not in themselves based on phenotype data. Thus, our analyses of positive selection on *ASPM* and *MCPHI* are based entirely on population genetic data and are independent of any claims about the phenotypic effects of these genes. We explicitly stated in our papers that the substrate of selection on these genes might be domains of brain biology other than cognition, or unrelated to the brain all together. Consequently, investigation into the phenotypic basis of a signature of adaptive evolution is an interesting follow-up study to the initial finding, but negative results do not invalidate the original observation of a signature of selection. Similarly, positive results showing an association with a trait would not strengthen the original finding; rather, they would simply give a complementary view into the evolutionary history of a gene. Clearly, the jury is still out regarding the nature of the phenotype that is responsible for the observed signature of selection on *ASPM* and *MCPHI*.

In a previously published comment, Yu *et al.* (7) compared the haplotype structure at *ASPM* with genome-wide data and argued that the signature of selection we found (1) does not depart from neutrality. However, their test of selection is based on data from the International Haplotype Map (HapMap) project's Encyclopedia of DNA Elements (ENCODE) (8), which underrepresent rare alleles (9) because of an ascertainment bias introduced by an underlying two-tier genotyping strategy. In particular, the ENCODE data has been shown to underrepresent variants with a minor allele frequency (MAF) $\leq 5\%$ (10, 11). This becomes particularly problematic when applying the data across different populations, even when they are closely related (12). The importance of this bias is best exemplified by the site-frequency spectrum of the non-D haplogroup in our full resequencing *ASPM* data. Of the 166 polymorphic sites segregating in the non-D haplogroup, 58% have a MAF $\leq 5\%$. This strongly suggests that an underrepresentation of rare alleles in the ENCODE data may significantly elevate the frequency of homozygotes.

Yu *et al.* attempted to circumvent this problem by using 19 kb of *ASPM* that they genotyped with

the same two-tier approach. Although underrepresentation of segregating sites would result in overestimating the frequency of homozygotes in the genome-wide data, in the presence of a signature of selection where segregating sites are already rare or absent, this would have a much smaller effect on the estimate of homozygote frequency. Thus, under the model of selection at *ASPM*, the prediction would be that (i) the frequency of homozygotes in the ENCODE data would be overestimated and (ii) the frequency of homozygotes in the *ASPM* data, although generated using the same methodology, would either be accurately estimated or only marginally overestimated, depending on the number of segregating sites. Consequently, Yu *et al.*'s test of selection may have lower power than the test we used based on full resequencing data.

To test for selection using empirical comparison with resequencing data from other loci, we examined the frequency of long-range homozygotes in the 289 genes of the Seattle single-nucleotide polymorphisms (SNPs) project. Although based on different individuals from the ones we used, its panel of 47 individuals represents global genetic diversity much like our samples. Although this data set is much smaller in scope than the ENCODE regions, our comparison is conservative because the Seattle SNPs data most likely overrepresent loci under positive selection, given the project's focus on inflammatory-response genes. Comparing the number of homozygotes between the Seattle SNPs genes and our resequencing data, we found that only 1.7% of the genes (5 out of 289) had a frequency of homozygotes across a comparable region that was equal to or greater than that of *ASPM*. Furthermore, three of these five genes (TRPV5, TRPV6, and MAPT) are loci for which strong evidence of positive selection has been demonstrated previously (13–15). We also carried out a simulation-based test of selection for the Seattle SNPs data. Here only six genes (~2%), including TRPV5, TRPV6, and MAPT, showed a *P* value smaller than or equal to that of *ASPM*. An issue of major concern is the accuracy of recombination rate estimates. Yu *et al.* argued that their differing conclusions stem in part from allowing for uncertainty in the recombination rate. In our study, we estimated a local recombination rate of 1.9 cM/Mb by a population genetic method, using 13 kb of *ASPM* in 84 individuals from an unstructured population (1). We chose to use this estimate in our simulations because it is in close agreement with direct measurements of recombination in this region by pedigree analysis (16), which is not subject to the confounding factors introduced by population-based estimates (17). In light of Yu *et al.*'s estimate of 0.12 cM/Mb, however, we decided to revisit our test of selection. Using a recombination rate of 0.12 cM/Mb, we nonetheless found a significant departure from neutral expectation at *ASPM* ($P = 0.0002$). We repeated the simulations with a wide range of

¹Howard Hughes Medical Institute, Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA.

²Committee on Genetics, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu

demographic histories (1), all of which produced significant results ($P = 0.00002$ to 0.007). Thus, our evidence for selection at *ASPM* is independent of the estimated recombination rate.

We also sought to evaluate whether Yu *et al.*'s test of homozygote frequency is robust to changes in the recombination rate. We repeated their analysis of the European-American ENCODE data, but rather than matching the window size by genetic distance, we used either a physical distance of 19 kb or 29 segregating sites (as in Yu *et al.*'s *ASPM* data). When we used this sliding-window analysis, which does not assume any recombination rate, the evidence for a homozygote excess in Yu *et al.*'s *ASPM* was significant ($P < 0.001$ and $P < 0.008$, respectively).

Based on their estimated recombination rate and the observed breakdown of LD on either side of *ASPM* in the HapMap data, Yu *et al.* estimated the age of the D haplotype at 269,050 to 328,250 years, in stark contrast to our estimate of 5800 years. In light of the uncertainties about human demographic history and recombination rate, we chose in our analysis to estimate the age of the D haplotype using a mutation-based method that is robust to any specific demographic history or recombination rate (18). Yu *et al.*'s recombination-based method, on the other hand, which makes specific assumptions about the population genetic model, is not robust to departures from this model (19). This is best exemplified by the extreme sensitivity of their estimate to the assumed recombination rate. Thus, although with a recombination rate of 0.11 cM/Mb their method estimates an age of 269,050 to 328,250 years, using our previously estimated recombination rate of 1.9 cM/Mb yields an age of 15,577 to 19,004 years. This is in line with our estimated upper limit of 14,100 years. Furthermore, our mutation-based age estimate of 5800 years is also consistent with the recombination-based estimate, when using the local recombination events in our *ASPM* data and our estimated recombination rate. Consequently, we argue that our method, which is robust to demographic history and recombination rate,

yields a more reliable age estimate for the *ASPM* haplogroup D.

We acknowledge that the high frequency of the derived allele in Papua New Guinea is highly interesting but disagree with Yu *et al.*'s assertion that this contradicts our age estimate. Although Papua New Guinea was originally settled around 40,000 years ago, several waves of migration have occurred since, including the major Austronesian migration ~3500 years ago (20). In general, given the tremendous uncertainty surrounding human demographic history, we would caution against drawing broad conclusions about the age of an allele on the basis of its geographic distribution.

Finally, Yu *et al.* carry out several standard frequency-based tests of selection on their *ASPM* and the ENCODE genotyping data, finding that the *ASPM* values do not depart from genome-wide values. Although the authors suggest that this finding contrasts with our analysis, we believe it is consistent with our own findings. Statistical tests that rely on an excess of rare alleles have low power to detect signatures of recent selection. Thus, it has been demonstrated empirically that these tests do not deviate from neutrality even at loci with well-established histories of recent positive selection, such as *G6PD* and *TNFSF5* (21). This is precisely why we chose in our analysis not to rely on these test statistics. Similarly, Yu *et al.* find that the population differentiation at *ASPM* does not depart significantly from the ENCODE comparison regions. Our test of selection, however, is independent of geographic distribution, which was analyzed only to elucidate the origin of the derived allele and is explicitly stated as such.

In conclusion, Yu *et al.*'s analysis of selection at *ASPM* uses a biased data set for genome-wide comparison. Because tests of selection-driven homozygote excess are highly sensitive to underrepresentation of rare alleles, this ascertainment bias makes their test potentially lower in power than the test we used with full resequencing data. The author's use of genotyping data for *ASPM* as well does not increase the power of their test,

because under selection, rare alleles will be underrepresented primarily in the genome-wide data. As discussed above, several lines of evidence suggest that our differing conclusions stem from this bias rather than erroneous assumptions in our simulated data. Comparison of homozygote excess with genome-wide resequencing data shows that *ASPM* is an outlier with significant departure from neutrality, and repeating our test of selection with Yu *et al.*'s estimated rate of recombination shows that our results are robust to varying recombination rates. In contrast, Yu *et al.*'s analysis is highly sensitive to recombination rate, both in terms of their test of homozygote excess in the ENCODE data and in terms of their estimated age for haplogroup D.

References and Notes

1. N. Mekel-Bobrov *et al.*, *Science* **309**, 1720 (2005).
2. P. D. Evans *et al.*, *Science* **309**, 1717 (2005).
3. N. Timpson, J. Heron, G. D. Smith, W. Enard, *Science* **317**, 1036 (2007); www.sciencemag.org/cgi/content/full/317/5841/1036a.
4. N. Mekel-Bobrov *et al.*, *Hum. Mol. Genet.* **16**, 600 (2007).
5. R. P. Woods *et al.*, *Hum. Mol. Genet.* **15**, 2025 (2006).
6. J. P. Rushton, P. A. Vernon, T. A. Bons, *Biol. Lett.* **3**, 157 (2007).
7. F. Yu *et al.*, *Science* **316**, 370 (2007); www.sciencemag.org/cgi/content/abstract/316/5823/370b.
8. The ENCODE Project Consortium, *Science* **306**, 636 (2004).
9. X. Ke *et al.*, *Hum. Mol. Genet.* **14**, 2757 (2005).
10. M. R. Barnes, *Brief. Bioinform.* **7**, 211 (2006).
11. I. Pe'er *et al.*, *Nat. Genet.* **38**, 663 (2006).
12. A. Montpetit *et al.*, *PLoS Genet.* **2**, e27 (2006).
13. J. M. Akey *et al.*, *PLoS Biol.* **2**, e286 (2004).
14. J. M. Akey, W. J. Swanson, J. Madeoy, M. Eberle, M. D. Shriver, *Hum. Mol. Genet.* **15**, 2106 (2006).
15. H. Stefansson *et al.*, *Nat. Genet.* **37**, 129 (2005).
16. A. Kong *et al.*, *Nat. Genet.* **31**, 241 (2002).
17. M. P. Stumpf, G. A. McVean, *Nat. Rev. Genet.* **4**, 959 (2003).
18. R. Thomson, J. K. Pritchard, P. Shen, P. J. Oefner, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360 (2000).
19. J. C. Stephens *et al.*, *Am. J. Hum. Genet.* **62**, 1507 (1998).
20. T. Tsukahara *et al.*, *J. Hum. Genet.* **51**, 244 (2006).
21. P. C. Sabeti *et al.*, *Nature* **419**, 832 (2002).

1 May 2007; accepted 31 July 2007
10.1126/science.1143658