a similar claim as does the rejection of the hypothesis $H_0 : \mu_{A_1} = \mu_{A_2}$; yet, the variance contrast might be substantial when the usual mean difference is undetectable. In either case, one can claim that the locus *A* is either directly associated with the trait or that it is a marker associated via correlation with an unobserved factor, *B*. When potential confounding due to population stratification is not an issue, the latter case leads to a standard claim that there is a nearby causal locus *B* correlated with the marker *A* via LD.

A practical question remains: How do we distinguish a genuine flip-flop from a statistical artifact? Our analysis shows that the underlying mechanism of a flip-flop is a change in the *AB* haplotype frequencies or, in the case of a zero-LD flip-flop, in the allele frequencies of *B* between populations. Examples can be constructed where both the allele frequency of the observed variant as well as the population prevalence of the trait ($\mathbf{M} \cdot \mathbf{P}$) remain the same across populations, despite the flip-flop. Nevertheless, these are contrived situations that take place only at specific values of the four haplotype frequencies. Thus, a flip-flop is usually accompanied by a change in the population prevalence and in the case of a nonzero LD, by a change in the frequency of the observed variant as well. There would be a higher confidence that the flip-flop is genuine in those cases where studied populations are of distinct ancestry, with evidence of allele-frequency differences at many loci. In addition, we suggest that in the case of a quantitative trait, the allelic-variance contrast can be examined. This contrast can be informative even at the flip-flop point, where no allelic effect can be detected. If normality of the trait can be assumed, the variance contrast provides an independent evidence that the studied variant has a genetic involvement, either as a LD proxy for causal variation or as a part of a functional unit. A significant allelic-variance contrast in both samples that exhibit a flip-flop may serve as an additional evidence for a genuine genetic association. Statistical tests for comparison of allelic and haplotypic variances will be detailed in a subsequent paper.

Dmitri V. Zaykin[1,*] and Kyoko Shibata[1]
[1]National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA
*Correspondence: zaykind@niehs.nih.gov

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. Luda Diatchenko, Norman Kaplan, Clare Weinberg, David Umbach, and Rongheng Lin provided stimulating discussion. We thank three anonymous reviewers for valuable comments.

References

1. Lin, P.I., Vance, J.M., Pericak-Vance, M.A., and Martin, E.R. (2007). No gene is an island: The flip-flop phenomenon. Am. J. Hum. Genet. *80*, 531–538.
2. Terwilliger, J., and Hiekkalinna, T. (2006). An utter refutation of the 'Fundamental Theorem of the HapMap'. Eur. J. Hum. Genet. *14*, 426–437.
3. Weir, B.S. (1996). Genetic Data Analysis II (Sunderland, MA: Sinauer Associates).
4. Lewontin, R.C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. Genetics *49*, 49–67.
5. Zaykin, D.V., Meng, Z., and Ehm, M.G. (2006). Contrasting linkage-disequilibrium patterns between cases and controls as a novel association-mapping method. Am. J. Hum. Genet. *78*, 737–746.
6. Zhao, J., Jin, L., and Xiong, M. (2006). Test for interaction between two unlinked loci. Am. J. Hum. Genet. *79*, 831–845.
7. Schrodi, S.J., Garcia, V.E., Rowland, C., and Jones, H.B. (2007). Pairwise linkage disequilibrium under disease models. Eur. J. Hum. Genet. *15*, 212–220.
8. Wang, T., Zhu, X., and Elston, R.C. (2007). Improving power in contrasting linkage-disequilibrium patterns between cases and controls. Am. J. Hum. Genet. *80*, 911–920.

DOI 10.1016/j.ajhg.2008.02.001.

Response to Zaykin and Shibata

Opposite directions of association of the same allele with disease in different populations (i.e., the flip-flop phenomenon) complicate the interpretation of association findings. We recently reported that variation in linkage disequilibrium (LD) or interlocus correlation in the context of multilocus effects may lead to flip-flop associations.[1] In the current issue of the *Journal* Zaykin and Shibata report that the flip-flop phenomenon may also be observed when there is constant LD, even without interactive multilocus effects, or when there is no LD for certain interactive disease models.

Zaykin and Shibata show how a flip-flop can occur in the case of constant LD with an example in which the frequencies of two haplotypes (i.e., A_1B_2 and A_2B_1) are switched in two populations, resulting in the same level of LD, but a reversal of the effect of allele A_1 in the two populations. This occurs because the effect of A_1 is a weighted sum of the haplotype effects over alleles at the B locus. The weights change in the two populations with different haplotype-frequency configurations. This example represents a special case in which haplotype frequencies differ significantly but LD remains the same. This may be the exception rather than the rule when haplotype frequencies diverge. Nevertheless, this example correctly demonstrates that it is differences in haplotype-frequency configuration, not

necessarily LD itself, that give conditions in which a flip of allelic effects can occur. It is important, even if estimates of LD measures are the same, to examine the distribution of haplotype frequencies in different samples with apparent flip-flop effects.

As a second case, Zaykin and Shibata consider loci in linkage equilibrium. They show how certain configurations of haplotypes penetrances can give rise to a flip-flop when there is an unobserved variant whose allele frequency varies in different populations. This results when the effects at the observed locus (*A*) and unobserved locus (*B*) interact such that the effect of A_1 may be revessed depending on whether it is on the B_1 or B_2 background. This example highlights our point that failure to account for other interacting variants can produce ambiguous association results at the observed locus under question,[1] and it shows that this can happen even without LD.

Zaykin and Shibata's study and our study have given evidence-based explanations for the controversial phenomenon of flip-flop associations. They demonstrate that failure to account for multilocus differences in samples can lead to legitimate flip-flops in a variety of scenarios. However, neither of these two studies has attempted to provide a definitive explanation for the flip-flops because such a phenomenon can stem from various reasons, ranging from genotyping errors to genomic complexity. Still, the lesson is consistent: Genomic context is important. We need to interpret associations in the context of differences in haplotype structure that occur in different populations or as a result of sample heterogeneity. Furthermore, the effect of one locus on disease risk may be inconsistent or missed completely if we fail to examine it jointly in the context of other known disease variants. These examples help to emphasize the key point that "no gene is an island."

Ping-I Lin,[1,2] Jeffery M. Vance,[3]
Margaret A. Pericak-Vance,[3] and Eden R. Martin[3,*]
[1]Department of Medicine, Program in Genetics and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA; [2]Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Catonsville, MD 21228, USA; [3]Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL 33101, USA
*Correspondence: emartin1@med.miami.edu

References

1. Lin, P.I., Vance, J.M., Pericak-Vance, M.A., and Martin, E.R. (2007). No gene is an island: The flip-flop phenomenon. Am. J. Hum. Genet. *80*, 531–538.

DOI 10.1016/j.ajhg.2008.02.004.

Optimal Two-Stage Testing for Family-Based Genome-wide Association Studies

To the Editor: A recent paper[1] in the *Journal* addressed the important issue of hypothesis testing for family-based genome-wide association studies of quantitative traits. The authors discuss the optimal use of the two sources of information (between and within[2,3]) available with family-based samples and recommend the use of a "screening" step, followed by a "testing" step.[1,4,5] By drawing an analogy with two-stage studies, in which independent samples are used rather than between and within components, we show here that statistical power is always greater with a single ("total" or "joint") test than with a "screening" approach. Furthermore, Ionita-Laza et al.[1] propose a rank-based weighting scheme for use with the "screening" approach, but such an approach fails to take into account the magnitude of the evidence for association in the between-component test. An approach based on the total test (with the between component controlled for population stratification) should provide greater power than an approach simply based on ranks.

Ionita-Laza et al.[1] focus on the "conditional power," a statistic derived from simulations that use the parental genotypes and the offspring phenotypes but not the offspring genotypes.[4,5] It is worthwhile clarifying that the "conditional power" uses the same information as the between-family test—for the between component, the parental genotypes are used for calculating a coding that summarizes the information contained in the parents. In the simplest case, association is tested by regression of offspring quantitative trait on this coding. In Abecasis et al.,[3] the coding is based on a "genotype score," where for genotype 11, 12, or 22, the genotype score is −1, 0, or 1, respectively. The between coding, b_i, where i indexes each family in the data, equals the average of the genotype score of the parents. If the parents are unknown, coding based on the offspring can be used. The within component is based on the deviation of each offspring from the between component and by construction is orthogonal (independent) to the between component. Specifically, the within coding, w_{ij}, equals $g_{ij} - b_i$ where g_{ij} is the genotype score of offspring j in family i. The information used for the within-component test is the offspring phenotype and the offspring genotype conditional on the parents genotype. Programs such as QTDT[3] and PLINK[6] offer a within-only test of association, as well as a total test of association (i.e., between plus within). An explicit between-only test is offered in PLINK.

Because the between and within components are independent, the question is then how best to use these two