

On the value of haplotype-based genotype–phenotype analysis and on data transformation in pharmacogenetics and -genomics

Stefan Viktor Vormfelde and Jürgen Brockmüller

In his excellent Review, David Balding describes the statistical approaches to population association studies as well as the constraints that apply to each type of data analysis¹. We would like to comment on two points that were raised. First, we believe that exploring haplotype–phenotype associations may be of greater value than estimated in the Review. Second, the logarithmic data transformation that the author recommends be carried out before linear regression analysis could lead to false conclusions about genotype–phenotype associations. The reasoning behind our statements is explained below.

Concerning haplotype-based phenotype analysis, take the example of two loci (A and B); each has two alleles, A_A' and B_B', with minor allele frequencies of 0.31 and 0.4, respectively (FIG. 1). Let us assume that allele A' has one-third the activity of A and that B' has 60% higher activity than B. The activity of the haplotypes AB, AB', A'B and A'B' would thus be 100%, 160%, 33.3% and 53.3%, respectively, as depicted on the right-hand side of FIG. 1 for individuals who are homozygous for each haplotype. Let us also

assume that there is moderate linkage disequilibrium, resulting in haplotype frequencies of AB, AB', A'B and A'B' of 0.54, 0.15, 0.06 and 0.25, respectively. On the basis of polymorphism analysis alone, the calculated activity of B'/B' homozygotes would be identical to that of B/B homozygotes (B/B: $(100\% \cdot 0.54 + 33.3\% \cdot 0.06) / (0.54 + 0.06) = 93.3\%$ versus B'/B': $(160\% \cdot 0.15 + 53.3\% \cdot 0.25) / (0.15 + 0.25) = 93.3\%$). In A'/A' homozygotes, activity would be slightly overestimated as 44% of that in A/A homozygotes (A/A: $(100\% \cdot 0.54 + 160\% \cdot 0.15) / (0.54 + 0.15) = 113\%$, A'/A': $(33.3\% \cdot 0.06 + 53.3\% \cdot 0.25) / (0.06 + 0.25) = 49.5\%$).

Although this case is hypothetical, it illustrates a documented example: the polymorphism analysis of *SLCO1B1* (solute carrier organic anion transporter family, member 1B1) resulted in the year-long assumption that a specific polymorphism in this gene was non-functional². The situation was clarified only on haplotype-based phenotypic analysis³. Therefore, we do not completely agree with David Balding that haplotype–phenotype analysis provides no advantage over the polymorphism-wise analysis¹. In fact, we

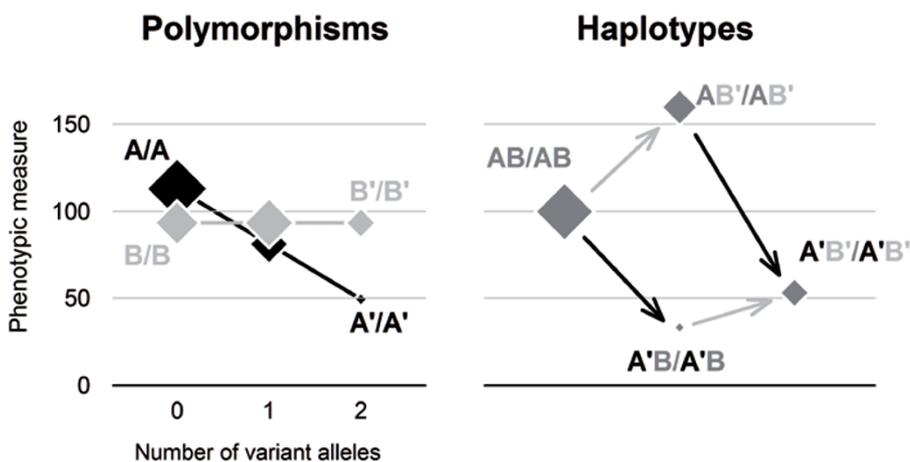


Figure 1 | **Genotype–phenotype association based on either polymorphisms or haplotypes.** The figure illustrates the outcome of polymorphism-based genotype–phenotype association on the left, and of haplotype-based genotype–phenotype association on the right, as described in the example given in the text. Data are given without spread parameter. Diamond sizes represent the number of respective alleles.

CORRESPONDENCE

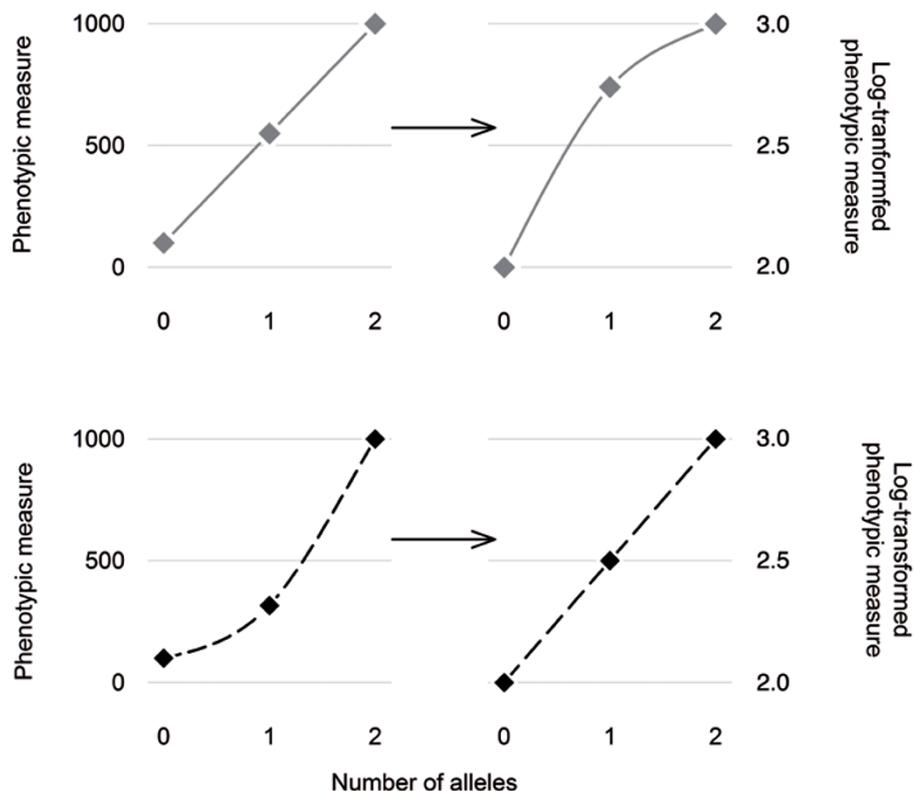


Figure 2 | Effect of log-transformation of data on the shape of genotype–phenotype associations. The left-hand side of the figure depicts a linear (top) and a non-linear (bottom) relationship between the number of allelic copies and a phenotypic measure; the right-hand side shows the same data after logarithmic transformation.

believe that haplotype-based phenotype association analysis should routinely complement polymorphism analysis.

The second point concerns data transformation. Following standard practice in statistics, David Balding recommends that phenotypic data that are not normally distributed should be logarithmically transformed before linear regression for genotype–phenotype analysis¹. However, such data transformation alters the shape of a relationship between the number of certain alleles and the phenotype (FIG. 2). Linear regression analysis might fail to detect some linear gene–dose–phenotype relationships (FIG. 2, top) or, conversely, some non-linear relationships will appear to be significant

(FIG. 2, bottom). The extent of both effects depends on the particular data set and the kind of data transformation used. Thus, we would usually discourage data transformation before applying linearization methods in genetics and pharmacogenetics.

Department of Clinical Pharmacology, University Medical Center, Robert-Koch-Straße 40, 37075 Göttingen, Germany.

Correspondence to S.V.V. e-mail: svormfe@gwdg.de

1. Balding, D. J. A tutorial on statistical methods for population association studies. *Nature Rev. Genet.* **7**, 781–791 (2006).
2. Niemi, M. Role of OATP transporters in the disposition of drugs. *Pharmacogenomics* **8**, 787–802 (2007).
3. Vormfelde, S. V. *et al.* The polymorphisms Asn130Asp and Val174Ala in OATP1B1 and the CYP2C9 allele *3 independently affect torsemide pharmacokinetics and dynamics. *Clin. Pharmacol. Ther.* (in the press).