CORRESPONDENCE

Issues with data transformation in genome-wide association studies for phenotypic variability [version 1; peer review: 2 approved]

Xia Shen¹, Lars Rönneård²,³

¹Division of Computational Genetics, Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, SE-750 07, Sweden
²Statistics, School of Technology and Business Studies, Dalarna University, Falun, SE-791 88, Sweden
³Division of Quantitative Genetics, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, SE-750 07, Sweden

Abstract
The purpose of this correspondence is to discuss and clarify a few points about data transformation used in genome-wide association studies, especially for phenotypic variability. By commenting on the recent publication by Sun et al. in the American Journal of Human Genetics, we emphasize the importance of statistical power in detecting functional loci and the real meaning of the scale of the phenotype in practice.
Correspondence

Recently, Sun et al.\textsuperscript{1} raised an interesting suggestion concerning the use of variance-stabilization transformations in genome-wide association studies (GWAS) for phenotypic variability. Specifically, Sun et al. revisited Yang et al.\textsuperscript{2}’s results on the variability-controlling locus \textit{FTO} for human body mass index (BMI) and claimed that the underlying variability across genotypes might not be as large as Yang et al. had seen. Although it was an important point that Sun et al. discussed, especially when quantitatively studying phenotypic variability has become such a hot topic, it is our opinion that there are some issues with the transformation approach that Sun et al. proposed.

First of all, if we take Sun et al.’s transformation according to Yang et al.’s phenotypic mean and variance per \textit{FTO} genotype class, \textit{i.e.} a one-to-one map through an inverse hyperbolic sine function, the BMI scale will become rather different compared with the ordinary measurement that we normally use (Figure 1). On the transformed scale of BMI, the difference between two persons who have a BMI of 24 and 25 kg/m\textsuperscript{2} is much larger than that between two BMIs of 20 and 21 kg/m\textsuperscript{2}, which is strange in reality since the original BMI scale is what we commonly use and also what we care about.

Sun et al.’s main argument here is that nearly all the measurement units are manmade. However, considering one of the traits of most interest, \textit{e.g.} height, why should we regard the difference between 160cm and 170cm different from 170cm and 180cm? Although the definitions of most units can be arbitrary, some measurement scales do have meaning in real life.

Secondly, a key problem with Sun et al.’s transformation in practice is that such a transformation is marker-specific. Namely, when performing a GWAS, one needs to transform the phenotypic records differently for different markers, according to the phenotypic distribution across the genotypes per marker. This does not make much sense in practical analyses, because if there is a “best” scale of the phenotype, it should be used for all the markers across the genome, before testing the association between the phenotype and the markers. Using the tested marker to determine the transformation of the phenotype is strange. If a marker-specific transformation can be estimated, one should estimate a genome-specific transformation for GWAS, instead of doing different transformations marker-by-marker.

Thirdly, if the transformation of the phenotype is determined by one marker showing a significant effect on the phenotypic variability before testing the other markers, another significant effect on the phenotypic variability might be created due to such a transformation. In such a situation, it is problematic to decide which phenotypic scale we should choose.

Fourthly, several recent studies discussed that gene-gene or gene-environment interactions could cause significant variance heterogeneity across genotypes\textsuperscript{3–6}, which makes testing variance-controlling loci a powerful tool to reveal potential interaction effects. Reducing the difference in variance across genotypes using a marker-specific variance-stabilization transformation would dramatically reduce such power. Regarding the biological sense of genetically regulated variance heterogeneity, empirical evidence has shown that a single causal locus could show a much higher significant effect on variance compared with the mean\textsuperscript{7}. In a particular population, such a locus may only be mappable through testing the variability rather than the magnitude of the phenotype.

The above issues cause us to question Sun et al.’s transformation in practice. The scale of the phenotype is certainly an important concern when interpreting an effect on phenotypic variability\textsuperscript{7}. However, one needs to be careful for the points above before applying any transformation on the data. In particular, the statistical power in detecting functional loci and the real meaning of the scale used should be emphasized.

Author contributions
XS and LR initiated the study. XS performed the analysis. Both authors contributed to writing the report.

Competing interests
The authors declare no competing interest.

Grant information
XS is funded by a Future Research Leaders grant from Swedish Foundation for Strategic Research (SSF) to Örjan Carlborg. LR is funded by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
References


Open Peer Review

Current Peer Review Status: ✅ ✅

Version 1

Reviewer Report 01 November 2013

https://doi.org/10.5256/f1000research.2505.r1948

© 2013 Aulchenko Y et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yurii Aulchenko
Institute of Cytology and Genetics, Siberian Division of the Russian Academy of Sciences, Novosibirsk, Russian Federation

Yakov Tsepilov
Novosibirsk State University, Novosibirsk, Russian Federation

Sodbo Sharapov
Novosibirsk State University, Novosibirsk, Russian Federation

We agree with criticism raised by Shen and Ronnegard in their points 2 and 3 concerning the application of the transformation of Sun et al. in the context of whole-genome scans. Indeed, applying this transformation in SNP-specific manner is difficult to adopt conceptually. Sun et al. rightly suggest that “the scales on which we measure interval-scale quantitative traits are man-made and have little intrinsic biological relevance”, but the underlying intrinsic scale, and the function reflecting this scale into the observed, is likely to be unique and does not change with SNP. In that, the transformation applied to a trait should not change through the markers studied. Practically, this is not very difficult to implement, and as a simplest option one could think of the estimation of Sun’s transformation parameters from upper, middle and lower terciles of the total phenotypic distribution. A more general approach (without restricting the data into three groups, but modelling the variance as a function of the mean) should be straightforward to implement.

We also understand the reasoning behind the Shen and Ronnegard’s points 1 and 4, but here we are less certain that the problem raised could be easily addressed. Specifically, one could argue with point 1 (“why should we regard the difference between 160cm and 170cm different from 170cm and 180cm?”): it is not that hard to imagine a biologically relevant model in which same changes of an “intrinsic scale” lead to different changes on the observed scale as the mean advances (an example would be Michaelis–Menten kinetics). Also both points 1 and 4 (losing power after transformation) relate not only to Sun et al.’s transformation, but to almost any transformation in wide use (e.g. Log, Box-Cox, Gausszenization/inverse-normal). While it is true that analysis of transformed trait may lead to reduced power (and specifically in case of Sun’s transformation applied in marker-specific manner to the analysis of variance heterogeneity it should), we have a feeling that one still would like to check whether the variance heterogeneity found can be modeled as a function of the mean (in which case any SNP affecting
the mean is likely to show “control” of the variance as well).

Finally, we fully agree with comment of William Hill and Ian White who criticize Sun et al.’s statement that “In the absence of genotypic mean differences, we can hardly infer that differences in variances are per se of biological interest”. We think that the differences in variance per se are biologically and genetically plausible and interesting.

**Competing Interests:** No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

Reviewer Report 07 October 2013

https://doi.org/10.5256/f1000research.2505.r1949

© 2013 Hill et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

William G. Hill
Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Ian White
University of Edinburgh

Shen and Rönnegård (SR) comment critically and succinctly on the paper by Sun et al. published in AJHG which advocates that, before any claim of differences in variance among genotypes in a GWAS or similar study, a check should first be made whether these can be removed by a monotonic transformation. Each of SR’s four criticisms seems well justified.

As 105 or more SNPs may be fitted in a GWAS study, what biological interpretation could be given to that number of different transformations or even on a limited subset of loci showing possible variance differences? If some loci give signals of mean but not variance difference, should these then be transformed to eliminate the scale effect on mean and perhaps reveal variance differences? Any concept of an original scale of measurement is lost, as SR point out. It is not obvious why the mere existence of a transformation designed to minimise differences in variance should prevent discussion of variance heterogeneity on the chosen scale. Equivalently, if we considered means of the three genotypes at the locus rather than just average effects, would our ability to transform the data at each locus such that heterozygotes were intermediate imply there was no dominance, or only that it was on a particular scale?

On a further point. Sun et al. (p395) comment: ‘In the absence of genotypic mean differences, we can hardly infer that differences in variances are per se of biological interest.’ That is to take too narrow a view: the mean and phenotypic variance (or CV) of a quantitative trait in any species take typical values, e.g. the CV for adult human height is ca. 4% and for BMI ca. 16%. There is direct evidence of genetic differences within species in environmental variance, with GWAS and other single gene studies, that cannot be removed by scale, so the level of the environmental variance is subject to evolutionary forces (e.g. Hill & Mulder 2010 Genet. Res. 92:381). To view variance as a biological phenomenon which is just some adjunct to the mean seems simplistic, as SR argue. Indeed one has to ask whether scale
transformations have value unless there is a biological basis, such as a log transformation to account for multiplicative genetic effects; but that must then apply across all loci.

**Competing Interests:** No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com