



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Nontrivial Replication of Loci Detected by Multi-Trait Methods

Citation for published version:

Ning, Z, Tsepilov, YA, Sharapov, SZ, Wang, Z, Grishenko, AK, Feng, X, Shirali, M, Joshi, PK, Wilson, JF, Pawitan, Y, Haley, CS, Aulchenko, YS & Shen, X 2021, 'Nontrivial Replication of Loci Detected by Multi-Trait Methods', *Frontiers in genetics*, vol. 12. <https://doi.org/10.3389/fgene.2021.627989>

Digital Object Identifier (DOI):

[10.3389/fgene.2021.627989](https://doi.org/10.3389/fgene.2021.627989)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Frontiers in genetics

Publisher Rights Statement:

Copyright © 2021 Ning, Tsepilov, Sharapov, Wang, Grishenko, Feng, Shirali, Joshi, Wilson, Pawitan, Haley, Aulchenko and Shen.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Nontrivial Replication of Loci Detected by Multi-Trait Methods

Zheng Ning^{1,2†}, Yakov A. Tsepilov^{3,4†}, Sodbo Zh. Sharapov⁴, Zhipeng Wang^{2,5,6}, Alexander K. Grishenko⁴, Xiao Feng¹, Masoud Shirali⁷, Peter K. Joshi⁸, James F. Wilson^{7,8}, Yudi Pawitan², Chris S. Haley⁷, Yurii S. Aulchenko^{9,10*} and Xia Shen^{1,2,7,8*}

¹ Biostatistics Group, School of Life Sciences and School of Ecology, Sun Yat-sen University, Guangzhou, China,

² Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, ³ Division of Biology, Novosibirsk State University, Novosibirsk, Russia, ⁴ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, ⁵ College of Animal Science and Technology, Northeast Agricultural University, Harbin, China, ⁶ Bioinformatics Center, Northeast Agricultural University, Harbin, China, ⁷ MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom, ⁸ Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom, ⁹ Kurchatov Genomics Center, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, ¹⁰ PolyOmica, 's-Hertogenbosch, Netherlands

OPEN ACCESS

Edited by:

Can Yang,

Hong Kong University of Science and Technology, Hong Kong

Reviewed by:

Zhonghua Liu,

The University of Hong Kong,

Hong Kong

Qiongshi Lu,

University of Wisconsin-Madison,

United States

Lin Hou,

Tsinghua University, China

*Correspondence:

Yurii S. Aulchenko

yurii@bionet.nsc.ru

Xia Shen

xia.shen@ed.ac.uk

†These authors have contributed equally to this work

Specialty section:

This article was submitted to Statistical Genetics and Methodology, a section of the journal *Frontiers in Genetics*

Received: 10 November 2020

Accepted: 04 January 2021

Published: 03 February 2021

Citation:

Ning Z, Tsepilov YA, Sharapov SZ, Wang Z, Grishenko AK, Feng X, Shirali M, Joshi PK, Wilson JF, Pawitan Y, Haley CS, Aulchenko YS and Shen X (2021) Nontrivial Replication of Loci Detected by Multi-Trait Methods. *Front. Genet.* 12:627989. doi: 10.3389/fgene.2021.627989

The ever-growing genome-wide association studies (GWAS) have revealed widespread pleiotropy. To exploit this, various methods that jointly consider associations of a genetic variant with multiple traits have been developed. Most efforts have been made concerning improving GWAS discovery power. However, how to replicate these discovered pleiotropic loci has yet to be discussed thoroughly. Unlike a single-trait scenario, multi-trait replication is not trivial considering the underlying genotype-multi-phenotype map of the associations. Here, we evaluate four methods for replicating multi-trait associations, corresponding to four levels of replication strength. Weak replication cannot justify pleiotropic genetic effects, whereas strong replication using our developed correlation methods can inform consistent pleiotropic genetic effects across the discovery and replication samples. We provide a protocol for replicating multi-trait genetic associations in practice. The described methods are implemented in the free and open-source R package `MultiABEL`.

Keywords: pleiotropy, multivariate analysis, genome-wide association study, cross-phenotype association, replication, genotype-phenotype map

INTRODUCTION

During the past decade, single-trait genome-wide association studies (GWAS) have successfully identified a multitude of genetic variants underlying complex traits (Visscher et al., 2017). However, the effects of the genetic variants, such as single-nucleotide polymorphisms (SNPs), on complex traits are usually very small. This directly limits the discovery power in most GWASs. Given widespread pleiotropy (Watanabe et al., 2019) and aiming to improve power, many multi-trait analysis methods have been developed in recent years to jointly analyze multiple correlated phenotypes. At the early stage, most multi-trait tools were based on individual-level data. For example, the `-multivariate` module of PLINK implements canonical correlation analysis (CCA) to identify the association between each SNP and linear combinations of phenotypes (Ferreira and Purcell, 2008); Combined-PC (Aschard et al., 2014) performed a principal components analysis on the phenotype data to improve statistical power. By combining a linear mixed model and multivariate analysis of variance (MANOVA), a multi-trait analysis was also shown to be effective for omics measurements (Shen et al., 2017).

As many GWAS have been performed in different study cohorts or populations, given the difficulty of sharing and combining individual-level data, multi-trait methods based on GWAS summary-level data became popular. Many such methods have been developed and demonstrated their benefits in boosting discovery power (Cotsapas et al., 2011; Solovieff et al., 2013; van der Sluis et al., 2013; Kim et al., 2015). For example, Stephens (2013) outlined a unified multivariate analysis framework based on Bayesian model comparisons; Zhu et al. (2015) introduced two test statistics S_{Hom} and S_{Het} to improve statistical power under different assumptions of effect sizes in which seven additional loci were suggested by jointly analyzing the summary statistics of three traits from the GIANT consortium (Park et al., 2016). A simulation study (Porter and O'Reilly, 2017) demonstrated that the statistical powers of most methods are similar to the power of the standard MANOVA. There are multi-trait methods where, although the information of multiple traits is used, the discoveries are yield from a univariate test. For example, GenomicSEM investigates the association between an SNP and a latent factor defined by several phenotypes (Grotzinger et al., 2019). MTAG (Multi-trait analysis of GWAS) was developed to integrate GWAS summary results of multiple related traits and improve the inference in each single-trait GWAS (Turley et al., 2018). In this perspective paper, we limit the term “multi-trait methods” to the methods where a multivariate test statistic is applied.

For any scientific discovery, replication—the ability to reproduce the findings of an original study in an independent experiment—plays a key role in the evaluation of credibility (Randall and Welser, 2018). Although many multi-trait methods were proposed and were demonstrated to boost genetic association discovery power, the multi-trait replication strategy has yet to be agreed upon. When an SNP is discovered in multi-trait analysis, a commonly used approach for replication is to replicate the associations trait-by-trait in a replication sample (Liang et al., 2017; Gialluisi et al., 2019). However, there are at least two disadvantages of such “univariate” replication: firstly, when the number of tested traits is large, multiple testing arises when determining the replication significance threshold; secondly, univariate replication does not account for phenotypic correlations between the tested traits, which generates conservative significance threshold after correction for multiple testing. Another straightforward way for replication is to directly perform the multi-trait test in a replication sample and see whether the overall association (omnibus p -value) is significant (Karnes et al., 2017; Liang et al., 2017; Luo et al., 2020). Although this strategy provides a unified test statistic, it does not reveal whether the effects that locus exhibits on traits in the discovery are the same (or similar) as those observed in replication. Thus, the consistency of the multiple genetic effects between the discovery and replication samples is usually overlooked. Even if the genetic effect sizes and directions are distinct between the discovery and replication samples, the multivariate replication test may still show significance.

Aiming to develop a stronger replication criterion, Shen et al. (2017) suggested a phenotype score replication strategy. For a discovered SNP, the strategy in the discovery sample is as follows:

construct a phenotype score that best fits the SNP genotype data; then, in the replication sample, compute the corresponding phenotype score and test the association between the phenotype score and the SNP genotype in the replication sample. In this way, the phenotype score replication provides a signed test statistic, which conceptually replicates the overall direction of the cross-phenotype effects. However, this replication strategy has not been implemented using GWAS summary statistics, and its performance has not been systematically examined.

To verify a locus detected by multi-trait methods, the first task is to replicate the association between the locus and the set of traits tested in the discovery phase. The next task is to evaluate the consistency of multi-trait genetic associations between discovery and replication samples across the traits in question, which calls for stronger replication strategies. To accomplish these two tasks in the replication of multi-trait signals, in this perspective article, we investigated and compared the performance of four replication methods. We began by reviewing the MANOVA and phenotype score, which we suggest as methods for locus replication. Then, we evaluated the consistency of effects that the locus exhibits onto multiple traits in question, i.e., replication of multivariate effects. A discovery study may have enough power to establish the overall association between a locus and a set of traits but insufficient power to robustly estimate all the effects the locus exhibits on multiple traits. In this case, one would expect that a locus would pass a locus-level replication but would fail to pass the multivariate genetic effects replication. To evaluate the similarity between estimates of multivariate genetic effects in discovery and replication, we introduced a Monte-Carlo (MC)-based correlation method. Through simulations of different scenarios, we have illustrated the strength and complementarity of four methods that cover different aspects of multivariate replication: (i) MANOVA, as a representative of multivariate methods providing unsigned omnibus p -values; (ii) phenotype score replication; (iii) Pearson correlation method; and (iv) Kendall correlation method. The implementation of these methods only requires summary association statistics. We suggest a four-level replication strategy where the above four replication methods are applied sequentially. To demonstrate the application of the four-level replication strategy, we studied the SNPs discovered by MANOVA, using the GWAS summary statistics from the GIANT consortium as an example, and try to replicate them in the UK Biobank (UKB).

SUMMARY OF THE METHODS

MANOVA

It has been shown that MANOVA can be performed using summary statistics (Stephens, 2013; Zhu et al., 2015). To simplify the formulae, we assume the phenotypes are standardized to have mean zero and variance one, and genotypes are centered to have mean zero. k traits Y_1, \dots, Y_k are dependent variables. If we denote the true marginal effects of a SNP on the k traits by β , then the null hypothesis in MANOVA is $H_0: \beta = \mathbf{0}$. Let $\mathbf{t} = [t_1, \dots, t_k]'$ be the vector of single-trait t-test statistics from association tests between the genotypes \mathbf{g} of a single SNP and the k phenotypes, and $\mathbf{R}^* \equiv \text{Cor}(\mathbf{t} \mid \beta) = \text{Var}(\mathbf{t} \mid \beta)$. If \mathbf{R}^* is available, the

test statistic

$$T^2 = \mathbf{t}'\mathbf{R}^{*-1}\mathbf{t}, \tag{1}$$

asymptotically follows a χ^2 distribution with k degrees of freedom under the null hypothesis. In practice, an unbiased estimate of \mathbf{R}^* can be obtained by selecting a large number of independent variants from the meta-GWAS summary statistics and calculating their correlation coefficients (Stephens, 2013; Zhu et al., 2015). More details can be found in Material and Methods and **Supplementary Note**.

Phenotype Score

If individual-level data are available, given an SNP, we can use CCA to get its most associated linear combination of traits. This linear combination of traits, which we name as phenotype-score, can be treated as a new phenotype. It has been shown (O'Reilly et al., 2012) that the coefficients in CCA are equivalent to the estimate of $\mathbf{b}_{k \times 1}$ in this reversed multiple regression

$$\mathbf{g} = \mathbf{Y}\mathbf{b} + \boldsymbol{\epsilon},$$

where $\mathbf{g}_{n \times 1}$ and $\mathbf{Y}_{n \times k}$ represent genotypes and phenotypes respectively. Assuming Hardy-Weinberg equilibrium (HWE), $\hat{\mathbf{b}}$, which is the estimate of \mathbf{b} , can be obtained by

$$\hat{\mathbf{b}} = 2f(1 - f) \cdot \hat{\mathbf{R}}^{*-1}\hat{\boldsymbol{\beta}},$$

where f is the coding allele frequency of the SNP (**Supplementary Note**). Therefore, we can get $\hat{\mathbf{b}}$ based on summary statistics and construct the phenotype-score $\mathbf{S} = \mathbf{Y}\hat{\mathbf{b}}$. Taking \mathbf{S} as a new phenotype and denoting the effect of the SNP on \mathbf{S} as β_s , we can estimate and test β_s in the discovery and replication populations (**Supplementary Note**). If $\hat{\beta}_s$ is significantly different from 0 and has the same sign in both populations, then we consider the association between the SNP and the phenotype-score is replicated.

Correlation Methods

Aiming to evaluate the similarity between estimates of multivariate genetic effects from discovery and those from replication samples across multiple traits, we developed this MC-based correlation method. The key idea is to evaluate the similarity of marginal effects with the uncertainty in estimates taken into account. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$ be the vector of estimated marginal effects across the k phenotypes. Then $\text{Cor}(\hat{\boldsymbol{\beta}} | \boldsymbol{\beta}) \approx \mathbf{R}^*$ (**Supplementary Note**). Because the variances of $\{\hat{\beta}_i\}$ are available, we can obtain $\hat{\Sigma}$, which is an estimate of $\Sigma = \text{Cov}(\hat{\boldsymbol{\beta}} | \boldsymbol{\beta})$. This allows us to draw $\boldsymbol{\beta}_{disc}^{MC}$ from $\mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\Sigma}_{disc})$ for the discovery sample, and $\boldsymbol{\beta}_{rep}^{MC}$ from $\mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\Sigma}_{rep})$ for the replication sample. Then we can compute their correlation coefficients. Here, we compute (i) Pearson's correlation coefficient:

$$\rho_{\beta}^{MC} = \frac{\sum_{j=1}^k (\beta_{j, disc}^{MC} - \bar{\beta}_{disc}^{MC})(\beta_{j, rep}^{MC} - \bar{\beta}_{rep}^{MC})}{\sqrt{\sum_{j=1}^k (\beta_{j, disc}^{MC} - \bar{\beta}_{disc}^{MC})^2} \sqrt{\sum_{j=1}^k (\beta_{j, rep}^{MC} - \bar{\beta}_{rep}^{MC})^2}}$$

and (ii) Kendall's rank correlation coefficient

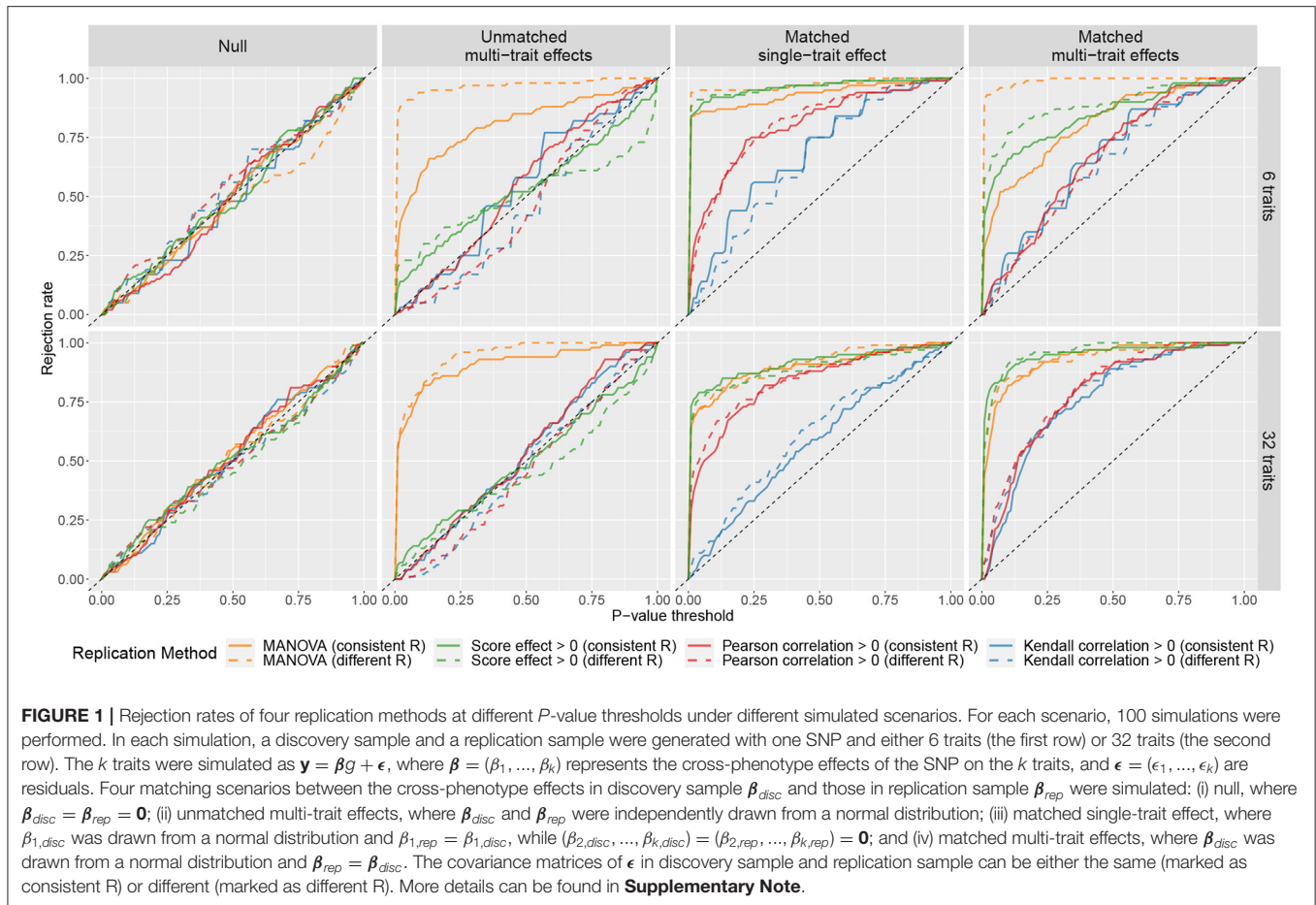
$$\tau_{\beta}^{MC} = \frac{2}{k(k-1)} \sum_{j < j'} \text{sgn}(\beta_{j, disc}^{MC} - \beta_{j', disc}^{MC}) \cdot \text{sgn}(\beta_{j, rep}^{MC} - \beta_{j', rep}^{MC}),$$

which measures the ordinal association between $\boldsymbol{\beta}_{disc}^{MC}$ and $\boldsymbol{\beta}_{rep}^{MC}$. Both correlation coefficients measure the similarity of estimates of genetic effects between discovery and replication samples across multiple traits. However, compared to Kendall's correlation, Pearson's correlation gives more weight to those associations with large effect sizes. In other words, if an SNP only affects few traits but those with large effect sizes, Kendall's correlation is likely to be close to 0, but Pearson's correlation may not be. By performing parametric bootstrap simulations, we can obtain estimated distributions of ρ_{β} and τ_{β} . Based on this distribution, the parametric bootstrap confidence intervals (CI) are attained, which can be used to evaluate the similarity between estimates of multivariate genetic effects.

SIMULATIONS

We performed a series of simulations to compare the replication results of four methods: MANOVA, phenotype score, Pearson correlation, and Kendall correlation. Four different scenarios of cross-phenotype effects were simulated: (i) null, where the SNP did not affect any trait; (ii) unmatched multi-trait effects, where the SNP exhibited effects on all traits, but the effects were independently simulated for discovery and replication samples; (iii) matched single-trait effect, where the SNP affected only one trait, and the true effect was the same in the discovery and replication samples; and (iv) matched multi-trait effects, where the SNP had effects on all the traits, and the true effects were the same in the discovery and replication samples. Two different numbers of traits, 6 or 32, were simulated. In practice, the phenotypic correlation matrices between the discovery and replication samples can be consistent or inconsistent, e.g., when estimated from two samples with different sample overlaps among the phenotypes. Its potential impact on replication was also evaluated. In the simulations, the phenotypic correlation matrices were based on the true observed correlation matrices in the GIANT and UKB data. More details can be found in **Supplementary Note**.

The results of the simulations are summarized in **Figure 1**, where the curves showed the rejection rates at different P -value thresholds. The curves of MANOVA were above the diagonal in scenarios (ii), (iii), and (iv). This means with MANOVA, an SNP is considered replicated as long as it affects at least one trait in the replication sample no matter how the cross-phenotype effect compares to that in the discovery sample. The curves of phenotype score replication did not deviate much from the diagonal in scenarios (ii). It shows when there is a large discrepancy in the cross-phenotype effect between the samples; the phenotype score does not consider it as replicated. However, the curves of phenotype score replication were above the diagonal in scenario (iii). So even when the SNP affects only one trait, the association between the SNP



and the phenotype score can still be replicated. When there were 32 traits, the curves of the Kendall correlation obviously deviated from the diagonal in scenario (iv) but only slightly deviated from the diagonal in scenario (iii). This indicates that Kendall correlation replication is more specific for confirming the cross-phenotype effect where an SNP affects multiple traits. The performance of the Pearson correlation is similar to that of the Kendall correlation; the exception is in scenario (iii), where the curves of the Pearson correlation were between those of the Kendall correlation and those of the other two methods. Although the correlation methods were more specific to evaluate the pattern of the multi-trait effects, their curves were under those of MANOVA and phenotype score, which indicates lower sensitivity.

FOUR-LEVEL REPLICATION STRATEGY FOR ASSOCIATIONS DETECTED BY MULTI-TRAIT METHODS

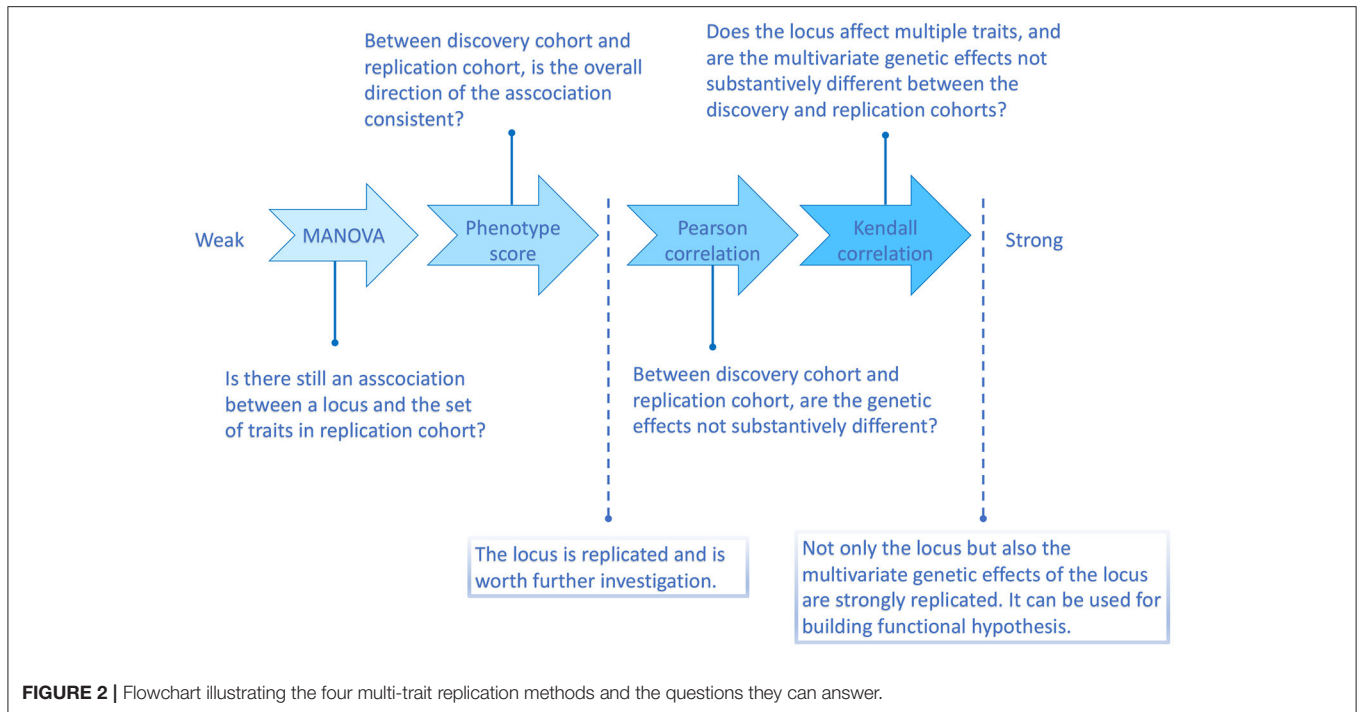
According to the above simulation results, the four methods are with different replication strength. Therefore, we suggest the following four-level replication strategy, summarized in **Figure 2**.

Level 1: Omnibus p -Value Test for Locus Replication

Omnibus p -value test (including MANOVA and many other multivariate methods) verifies whether there is indeed an association between the discovered locus and the set of traits in a replication sample. As long as the association is present in the replication sample, the locus is considered as replicated regardless of whether the multivariate genetic effects are similar between discovery and replication or not. This is the weakest type of multivariate replication.

Level 2: Phenotype Score for Locus Replication

The phenotype score method (Shen et al., 2017) can filter out some inconsistent cross-phenotype associations, such as scenario (ii) in the simulation, where the directions of the associations were distinct in discovery and replication samples. The score replication is stronger than MANOVA replication and is similar to a classical univariate replication in that it asks for both experiment-wise significance and consistency of effects between discovery and replication. However, in a univariate case, consistency is defined as and justified by the same sign of the effect of an associated allele. In contrast, the phenotype score operates in many dimensions and can be dominated by a small



number of traits, which precludes it from evaluating the overall similarity of estimates of genetic effects a locus exhibits on the tested traits. Thus, if a locus has passed the score replication, we can consider it an established locus. However, we can not be sure about the quality and stability of the locus-phenotype map suggested by the discovery study. The next steps aim to judge how well we can replicate the vector of effects.

Level 3: Pearson Correlation Method for Replication of Multivariate Genetic Effects

Pearson correlation method is a bridge between the phenotype score and the Kendall correlation method. As a correlation test strategy, each tested trait contributes more than it does in the phenotype score method. However, Pearson's correlation is more responsive to large genetic effects than Kendall's correlation. Therefore, when an SNP only affects few traits with similarly large effect sizes in both discovery and replication samples, such as scenario (iii) in the simulation, the Pearson correlation method would consider the multivariate genetic effects similar.

Level 4: Kendall Correlation Method for Replication of Multivariate Genetic Effects

Kendall correlation method is the strongest one among the four methods. Since Kendall's correlation is based on the ranks of estimated marginal genetic effects, any factor disturbing the ranks would weaken the correlation. For example, if an SNP does not affect all the tested traits, then its estimated marginal effects around zero on those irrelevant traits will be randomly ranked, which reduces Kendall's correlation (**Supplementary Figure 1**). An extreme case is when an SNP associates with only one trait. In this case, there will be almost no similarity of the estimated

effect sizes rank between samples. Therefore, if an SNP can be replicated by the Kendall correlation method, it is more likely to be associated with multiple traits, and the effect sizes (ranks) are consistent.

EXAMPLE: UKB REPLICATION OF ANTHROPOMETRIC LOCI DISCOVERED BY MANOVA IN THE GIANT CONSORTIUM

In order to compare the performance of different replication methods using real data, we investigated the loci discovered in the GIANT consortium and replicated them in UKB by MANOVA. We used six anthropometric traits for multi-trait analysis: BMI, height, weight, hip circumference (denoted here as HIP), waist circumference (WC), and waist-to-hip ratio (WHR). Among the 2,348,642 SNPs in the GIANT data, 21,240 SNPs had MANOVA p -value $< 5 \times 10^{-8}$. Considering SNPs < 1 Mb distant from most significant SNP as one locus, the 21,240 SNPs were clumped into 449 loci. For each locus, we defined its top SNP as the SNP with the smallest MANOVA p -value in GIANT. After, we used the top SNPs to represent loci.

A total of 317 out of the 449 top SNPs could be replicated in UKB with a MANOVA p -value threshold of 0.05/449. To illustrate the performance of different replication methods, among the 317 loci, we looked into 24 "multivariate-only" loci as an example. These 24 loci could be discovered and replicated by MANOVA, but none of them could reach p -value $< 5 \times 10^{-8}$ in univariate GWAS for any of the six anthropometric traits. The replication results were summarized in **Supplementary Table 1**.

According to the results from the phenotype score method, all except two SNPs had a p -value < 0.0021 (experiment-wise

$P < 0.05$ after Bonferroni correction for 24 tests) in UKB. Additionally, all of their $\hat{\beta}_s$ had the same sign in both GIANT and UKB. The results suggest that, for any of the 22 SNPs, there is no large discrepancy in its multi-trait association between GIANT and UKB.

Results from correlation methods indicated different levels of similarity of multivariate genetic effects between discovery and replication across the 22 loci. Both the two correlation methods identified the same 19 SNPs for which the pattern of multi-trait association was significantly similar between the discovery and replication samples. For some loci, the lower bounds of their 95% confidence intervals (CI) were higher than the others, suggesting a higher level of similarity of multivariate genetic effects between discovery and replication. **Supplementary Figure 2** visualized loci with different levels of similarity. For instance, although both rs2138275 and rs4968164 could be replicated by MANOVA, the 95% CI of τ_β was [0.73, 1] for rs2138275, and [-0.07, 1] for rs4968164. This means the multi-trait marginal genetic effects across samples were similar for rs2138275 but not for rs4968164. The contrast indicated that the underlying six-trait cross-phenotype association was more plausible at the locus led by rs2138275. For the locus led by rs4968164, although it could be detected and replicated by MANOVA, the SNP might be associated with only a small subset of the six traits.

Because the association of a locus with multiple traits is necessary but not sufficient for pleiotropy (Solovieff et al., 2013), it is more promising to discover pleiotropy from loci with reliable associations to multiple traits. We hypothesized that the loci with significant results in the Kendall correlation method were more likely to be associated with more traits in general, even those not limited to anthropometric measurements. According to the results based on the records in PhenoScanner (Staley et al., 2016), indeed the loci supported by the Kendall correlation method tended to be associated with more traits than the others (**Supplementary Figure 3**).

DISCUSSION

For signals discovered in multivariate analysis, a more systematic replication strategy is necessary to better evaluate the credibility of these signals. In this perspective, we introduced a four-level replication strategy for replication of loci detected by multi-trait methods. The replication strategy only requires summary association statistics, which are straightforward to apply to multi-trait GWAS analyses. Our results show that the four methods are complementary and provide different degrees of replication strength.

In the strategy proposed (**Figure 2**), to verify a locus detected by multi-trait methods, in the first phase we replicate the association between the locus and the set of traits tested in the discovery phase (MANOVA and phenotype score); in the next phase, we evaluate the consistency of multi-trait genetic associations between discovery and replication across the traits in question (Pearson and Kendall correlations). When only the

MANOVA test is passed, this is the weakest type of replication, under which we have formally reached the p -value low enough (usually, experiment-wise $p < 0.05$) to claim replication of an association. This criterion is, however, weaker than that conventionally used in single-trait GWAS in which we ask not only for significance but also for consistency. The requirement of overall consistency is reached with the phenotype score replication step. At this stage, we can consider the locus to be established, that is, we can claim consistent and significant association of the locus to (at least some of the) investigated traits. The question of whether the discovery phase was informative enough to establish the genotype-multi-phenotype map of the associations of the locus is addressed in the second phase. In principle, given score replication was passed, one does expect that the Pearson correlation between effects estimated in discovery and replication will be positive. However, it may be weak, meaning that a researcher can not trust individual estimates of effects. In such a situation, any attempt to utilize observed associations of individual traits, e.g., the benefit of functional interpretation, is likely doomed. However, the Pearson correlation may be high because the locus exhibits a strong effect on just a few—or even only one—traits in the set or because it exhibits effects of many traits in the set. The fourth method, the Kendall correlation, allows one to distinguish between the two scenarios (see **Figure 2**).

To detect which traits are more likely to be irrelevant for an SNP, we can compare the ordinal discordance generated by each trait in Kendall's correlation. For an SNP, if the effects of the locus on a trait introduce much discordance, then the association between the SNP and the trait would be more suspicious. For example, the marginal effect of height generated much discordance for rs11231693, which indicated that height was more likely not contributing to the multivariate association between rs11231693 and the group of traits. This could be verified by the non-significant association in height GWAS (p -value was 0.91 in GIANT2015 and 0.96 in UKB).

In the simulation, as we aim to replicate multi-trait genetic effects, for a good replication method, the performance should not be driven by whether the phenotypic correlation matrices are the same between the discovery and replication cohort. However, we found that for MANOVA, replication samples with a different phenotypic correlation structure ("different **R**") could perform better than the "consistent **R**" scenario in the six traits settings. The phenomenon can be attributed to the choice of **R** for the replication cohort. In the six traits setting, **R** was from GIANT for the discovery cohort. Therefore, for "consistent **R**," **R** for the replication cohort was also from GIANT; whereas for "different **R**," **R** was from UKB instead. The T^2 statistic for MANOVA based on **R** from UKB could be on average larger than those based on **R** from GIANT, as for MANOVA, the replication results are fully determined by the replication cohort without comparing with the multi-trait genetic effects in the discovery cohort.

With our results, we emphasize the value of various replication strategies for multi-trait analysis. Proper replicated results from the multivariate analysis may substantially enhance our understanding of shared genetic architecture between complex traits and disease and reveal interesting biological knowledge.

The developed methods are implemented and freely available in the R package **MultiABEL**: <https://github.com/xiashen/MultiABEL>).

MATERIALS AND METHODS

Estimation of \mathbf{R}^*

Let \mathbf{R} represent the phenotypic correlation matrix of the k phenotypes. According to Zhu et al. (2015), $\mathbf{R}^* \approx \mathbf{R}$ when the phenotypes are measured on the same set of individuals. If the individuals for trait j and those for trait j' partially overlap, we denote the number of overlapping individuals as n_0 , those with trait j but not trait j' as n_1 , and those with trait j' but not trait j as n_2 . We then have

$$R_{j,j'}^* = \text{Cor}(t_j, t_{j'} | \beta) \approx \frac{n_0}{\sqrt{(n_0 + n_1)(n_0 + n_2)}} R_{j,j'} \quad (2)$$

(**Supplementary Note**). Therefore, the correlation of t-statistics is a shrinkage version of the phenotypic correlation, with a factor determined by the level of overlap. Because (2) holds for all SNPs, an unbiased estimate of the correlation matrix \mathbf{R}^* can be obtained by selecting a large number of independent variants from the meta-GWAS summary statistics and calculating their correlation coefficients (Stephens, 2013; Zhu et al., 2015).

For correlation matrix estimation, we used previously proposed approach based on correlation of Z-statistics between independent unassociated SNPs (Zhu et al., 2015). We filtered SNPs based on given criteria: $\text{MAF} > 0.1$; high imputations quality as indicated either by $\text{INFO} > 0.99$ (for UKB) or by $N_e/N > 0.9$, where $N_e = 1/(2pq(\text{se})^2)$, for GIANT; $|Z| < 2$; the sample of 200,000 independent (LD pruned) SNPs to compute correlation matrix (the list of SNPs was obtained by—"prune" option for PLINK using 1,000 Genomes data). In total, 128,670 independent SNPs were used to estimate the correlation matrix for GIANT and 166,000 SNPs for UKB. All estimated correlation matrices can be found in **Supplementary Table 2**.

Discovery Sample: GIANT

We downloaded the summary association statistics of six anthropometric traits meta-GWAS by the GIANT consortium from: https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. We used six anthropometric traits: BMI, height, weight, hip circumference (denoted here as HIP), waist circumference (WC), and waist-to-hip ratio (WHR). BMI data are from Locke et al. (2015); height data are from Wood et al. (2014); weight data are from Randall et al. (2013); HIP, WC, and WHR data are from Shungin et al. (2015).

As HapMap II allele frequencies were reported in the meta-GWAS instead of pooled allele frequencies across all the cohorts, we excluded SNPs with sample size $< 70,000$ and $\text{MAF} < 0.01$. SNPs with missing allele frequencies were also excluded. All SNPs were merged with genome positions (GRCh37) and filtered for autosomes only and position missings. In total, we ended with 2,348,642 SNPs.

Replication Sample: UK Biobank

UKB participants were recruited from the general UK population across 22 centers between 2006 and 2010. Subjects were aged 40–69 at baseline, underwent extensive phenotyping by questionnaire and clinic measurements, and provided a blood sample. Genotyping has been done with the first wave released in July 2015 and the second wave released in July 2018. Because the objective of this study is to demonstrate the replication methods, we use the first wave data of UKB as an example dataset. Data access to UKB was granted under Project No. 14302 and No. 19655. Phenotypes and genotypes were downloaded directly from UKB. In total 502,664 subjects had phenotypic information available, of whom 152,732 had been genotyped, of these 120,286 were identified as genetically British by UKB, of which 118,182 (55,842 men) had completed phenotyping. These subjects were taken forward for analysis.

Participants provided full informed consent to participate in UKB. This study was covered by the generic ethical approval for UKB studies from the NHS National Research Ethics Service (approval letter dated 17th June 2011, Ref 11/NW/0382). The authors in this study were completed blinded to individual-level data collection and preparation. The phenotypes involved in this study were adjusted for age, sex, and batch before being standardized to have mean zero and variance one.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XS and YA initiated the study. XS, YA, YP, and CH coordinated the study. ZN, XS, YP, and YA developed statistical methods. ZN and YT performed the data analysis. ZW contributed to the method development. ZW, SS, AG, XF, MS, PJ, and JW contributed to the data analysis. ZN and XS drafted the manuscript. YT, PJ, JW, YP, CH, and YA contributed to the manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

XS received the Swedish Research Council grants (no. 2014-00371 and 2017-02543). ZN was supported by the Swedish Research Council grant (No. 2017-02543) to XS. The work of SS and YA was funded by a grant from the Russian Science Foundation (RSF) No. 19-15-00115. The work of YT was supported by the Russian Ministry of Education and Science under 5-100 Excellence Programme, by the Ministry of Education and Science of the RF via the Institute of Cytology and Genetics SB RAS (project number 0259-2021-0009/AAAA-A17-117092070032-4), and by the Russian Foundation for Basic

Research (project 19-015-00151). CH and MS were supported by the UK Medical Research Council and the UK Biotechnology and Biological Sciences Research Council (Grant BB/J002844/1).

ACKNOWLEDGMENTS

We thank the global consortium GIANT for making their reported summary-level data freely available online. We thank

the UK Biobank (Project No. 14302 and No. 19655) for the data resource used in our replication analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.627989/full#supplementary-material>

REFERENCES

- Aschard, H., Vilhjálmsson, B. J., Greliche, N., Morange, P.-E., Trégouët, D.-A., and Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am. J. Hum. Genet.* 94, 662–676. doi: 10.1016/j.ajhg.2014.03.016
- Cotsapas, C., Voight, B. F., Rossin, E., Lage, K., Neale, B. M., Wallace, C., et al. (2011). Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* 7:e1002254. doi: 10.1371/journal.pgen.1002254
- Ferreira, M. A., and Purcell, S. M. (2008). A multivariate test of association. *Bioinformatics* 25, 132–133. doi: 10.1093/bioinformatics/btn563
- Gialluisi, A., Andlauer, T. F., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., et al. (2019). Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. *Transl. Psychiatry* 9, 1–15. doi: 10.1038/s41398-019-0402-0
- Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., et al. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* 3, 513–525. doi: 10.1038/s41562-019-0566-x
- Karnes, J. H., Bastarache, L., Shaffer, C. M., Gaudieri, S., Xu, Y., Glazer, A. M., et al. (2017). Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci. Transl. Med.* 9:eaa18708. doi: 10.1126/scitranslmed.aai8708
- Kim, J., Bai, Y., and Pan, W. (2015). An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genet. Epidemiol.* 39, 651–663. doi: 10.1002/gepi.21931
- Liang, J., Le, T. H., Edwards, D. R. V., Tayo, B. O., Gaulton, K. J., Smith, J. A., et al. (2017). Single-trait and multi-trait genome-wide association analyses identify novel loci for blood pressure in African-ancestry populations. *PLoS Genet.* 13:e1006728. doi: 10.1371/journal.pgen.1006728
- Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206. doi: 10.1038/nature14177
- Luo, L., Shen, J., Zhang, H., Chhibber, A., Mehrotra, D. V., and Tang, Z.-Z. (2020). Multi-trait analysis of rare-variant association summary statistics using MTAR. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-16591-0
- O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C., Elliott, P., Jarvelin, M.-R., et al. (2012). Multiphen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE* 7:e34861. doi: 10.1371/journal.pone.0034861
- Park, H., Li, X., Song, Y. E., He, K. Y., and Zhu, X. (2016). Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from giant consortium. *PLoS ONE* 11:e0163912. doi: 10.1371/journal.pone.0163912
- Porter, H. F., and O'Reilly, P. F. (2017). Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci. Rep.* 7:38837. doi: 10.1038/srep38837
- Randall, D., and Welsch, C. (2018). *The Irreproducibility Crisis of Modern Science*. National Association of Scholars.
- Randall, J. C., Winkler, T. W., Kutalik, Z., Berndt, S. I., Jackson, A. U., Monda, K. L., et al. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* 9:e1003500. doi: 10.1371/journal.pgen.1003500
- Shen, X., Klarić, L., Sharapov, S., Mangino, M., Ning, Z., Wu, D., et al. (2017). Multivariate discovery and replication of five novel loci associated with immunoglobulin GN-glycosylation. *Nat. Commun.* 8:447. doi: 10.1038/s41467-017-00453-3
- Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518:187. doi: 10.1038/nature14132
- Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., and Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* 14:483. doi: 10.1038/nrg3461
- Staley, J. R., Blackshaw, J., Kamat, M. A., Ellis, S., Surendran, P., Sun, B. B., et al. (2016). Phenoscanner: a database of human genotype-phenotype associations. *Bioinformatics* 32, 3207–3209. doi: 10.1093/bioinformatics/btw373
- Stephens, M. (2013). A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* 8:e65245. doi: 10.1371/journal.pone.0065245
- Turley, P., Walters, R. K., Maghzi, O., Okbay, A., Lee, J. J., Fontana, M. A., et al. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50, 229–237. doi: 10.1038/s41588-017-0009-4
- van der Sluis, S., Posthuma, D., and Dolan, C. V. (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* 9:e1003235. doi: 10.1371/journal.pgen.1003235
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005
- Watanabe, K., Stringer, S., Frei, O., Umičević Mirkov, M., de Leeuw, C., et al. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* 51, 1339–1348. doi: 10.1038/s41588-019-0481-0
- Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1186. doi: 10.1038/ng.3097
- Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., et al. (2015). Meta-analysis of correlated traits via summary statistics from GWAS with an application in hypertension. *Am. J. Hum. Genet.* 96, 21–36. doi: 10.1016/j.ajhg.2014.11.011

Conflict of Interest: YA was a co-founder and co-owner of private research organizations PolyOmica and PolyKnomics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ning, Tsepilov, Sharapov, Wang, Grishenko, Feng, Shirali, Joshi, Wilson, Pawitan, Haley, Aulchenko and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.