



CORRESPONDENCE

## Sun et al.'s study led to the underperformance of EigenGWAS

Feng Lin <sup>1</sup> · Guo-Bo Chen <sup>1</sup>

Received: 6 September 2018 / Revised: 30 January 2019 / Accepted: 31 January 2019 / Published online: 26 February 2019  
© The Genetics Society 2019

Sun et al. recently proposed a partial least square (PLS) method to explore selection signatures between populations (Sun et al., 2018). In their publication, EigenGWAS (Chen et al., 2016) was chosen as a benchmark, and aspects (e.g. power) of the methods were compared. Similar methods resembling EigenGWAS have been independently proposed and evaluated by others (Duforet-Frebourg et al., 2016; Galinsky et al., 2016). To our surprise, Sun et al.'s comparison found that EigenGWAS had nearly zero statistical power, whereas the power of PLS could be as high as 91%. This is contrary to the findings of many researchers, who have demonstrated that a properly conducted EigenGWAS is powerful in finding selection signatures (Bosse et al., 2017). We therefore investigated what led to the poor performance of EigenGWAS in Sun et al.'s study.

### Sun et al.'s PLS

Sun et al.'s study used a continuous phenotype  $z$ . In their simulation, upon the mean of  $z$ , they defined a binary vector  $y$  (known as a threshold trait in quantitative genetics), and then had 2 subpopulations. Their analysis used  $\mathbf{b} = \mathbf{X}^T \mathbf{y}$ , in which  $\mathbf{X}$  was the genotype matrix. Of note,  $\mathbf{b}$  was statistically equivalent to the regression coefficient of the simple linear regression. According to their description, Sun et al. put  $\mathbf{b}^T$  into singular value decomposition (SVD). It could be found in their online example (<http://klab.sjtu.edu.cn/PLS/>) that the output of their SVD was completely linear to  $\mathbf{b}$

(Supplementary Notes). Furthermore, as expected, when the original continuous phenotype had been converted to a threshold trait the statistical power was compromised (Supplementary Notes).

### Power comparison for supervised and unsupervised methods

We agree that supervised and unsupervised methods can supplement each other; however, they are not always comparable if their application scenarios are not well clarified. Because PLS is a supervised method and EigenGWAS is an unsupervised method, the simulation scenarios used by Sun et al., which highly preferred supervised methods, resulted in an inappropriate power comparison. Sun et al. assigned an  $h^2$  close to 5–10% for each of the 10 functional loci, generated a quantitative variable and then classified this underlying variable to form 2 subpopulations. The EigenGWAS method did not detect any loci because the simulated subpopulation information could not have been captured by EigenGWAS. This is the computational reason for the finding of up to 91% power for PLS and nearly zero for EigenGWAS (shown in Sun et al.'s Table 2). Of note, Sun et al.'s power comparison was incapable of balancing type I and type II error rates (Lynch and Walsh, 1998). In their Table 2 the signals had been simply picked by the top 1% (or 0.1%) ranked SNPs, ignoring the issue of false positives due to multiple testing.

### Different thresholds for real data analyses

Furthermore, in their real data analysis, where PLS and EigenGWAS could potentially perform similarly, more stringent thresholds were applied for EigenGWAS and comparatively less stringent thresholds were applied for PLS. In their Figure 2Sa for EigenGWAS the threshold was  $-\log_{10}\left(\frac{0.05}{363251}\right) = 6.86$ –Bonferroni correction  $p$ -value at 5% type I error rate given 363,251 SNPs. However, for the

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41437-019-0199-x>) contains supplementary material, which is available to authorised users.

✉ Guo-Bo Chen  
chenguoobo@gmail.com

<sup>1</sup> Clinical Research Institute, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, 310014 Hangzhou, Zhejiang Province, China

same data in applying PLS, the threshold was reduced to  $-\log_{10}(0.0001) = 4$ , a far lower threshold. If 6.86 was fairly applied to PLS, neither *HERC2* (in their Figure 2A,  $p$ -value of  $2.13 \times 10^{-6}$ ) nor *OCA2* (in their Figure 2B,  $p$ -value not given but was about 0.0001) was significant. No justification was given why such different thresholds had been applied for EigenGWAS versus PLS.

## Conclusion

As analysed above, EigenGWAS and PLS should find their respective application, but Sun et al. did not appropriately compare for these two methods. Consequently, their simulations were highly preferentially biased towards the supervised methods. When it was applied to analyse real data, a more stringent threshold was adopted for EigenGWAS compared with PLS. These comparisons led to the underperformance of EigenGWAS.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Bosse M, Spurgin LG, Laine VN, Cole EF, Firth JA, Gienapp P, et al. (2017) Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science* 358:365–8
- Chen G-B, Lee SH, Zhu Z-X, Benyamin B, Robinson MR (2016) EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity* 117:51–61
- Duforet-Frebourg N, Luu K, Laval G, Bazin E, Blum MGB (2016) Detecting genomic signatures of natural selection with principal component analysis: Application to the 1000 genomes data. *Mol Biol Evol* 33:1082–93
- Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, et al. (2016) Fast principal components analysis reveals independent evolution of ADH1B gene in Europe and East Asia. *Am J Hum Genet* 98:456–72
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland, MA, USA
- Sun H, Zhang Z, Olasege BS, Xu Z, Zhao Q, Ma P, et al. (2018) Application of partial least squares in exploring the genome selection signatures between populations. *Heredity* 122:288–93