Multivariate GWAS With Life Satisfaction And Positive Affect

Bart M. L. Baselmans, Social Science Genetic Association Consortium, M.G. Nivard & M. Bartels

Motivation

Recently, the first genetic variants for SWB were found in a genome-wide association meta-analysis (GWAMA; Okbay et al., 2016). Although the majority of the cohorts provided a single measurement of life satisfaction (LS) or positive affect (PA) measurement, a substantial part (N=25) ran a GWAS for both LS and PA bringing the total sample size to ~347K. In conventional GWAMA's however, sample independency is of crucial importance as non-independency will lead to an inflation of type-I errors. Because of this, the pooled meta-analysis required the omission of either LS or PA (N = 48,083) to ensure sample-independence, bringing the effective sample-size to 298,420.

This study

We re-analyzed the SWB GWAMA, including all available (N = 346,503) data, while accounting for violation of independence samples. Doing so, we obtained five genome-wide significant variants ($p < 5 \times 10^{-8}$) associated with SWB. Importantly, estimation of the linkage disequilibrium score intercept (LDSC) did not differ from 1, indicating that these variants are true genetic signal and not due to an inflation of type I errors caused by using a non-independent sample.

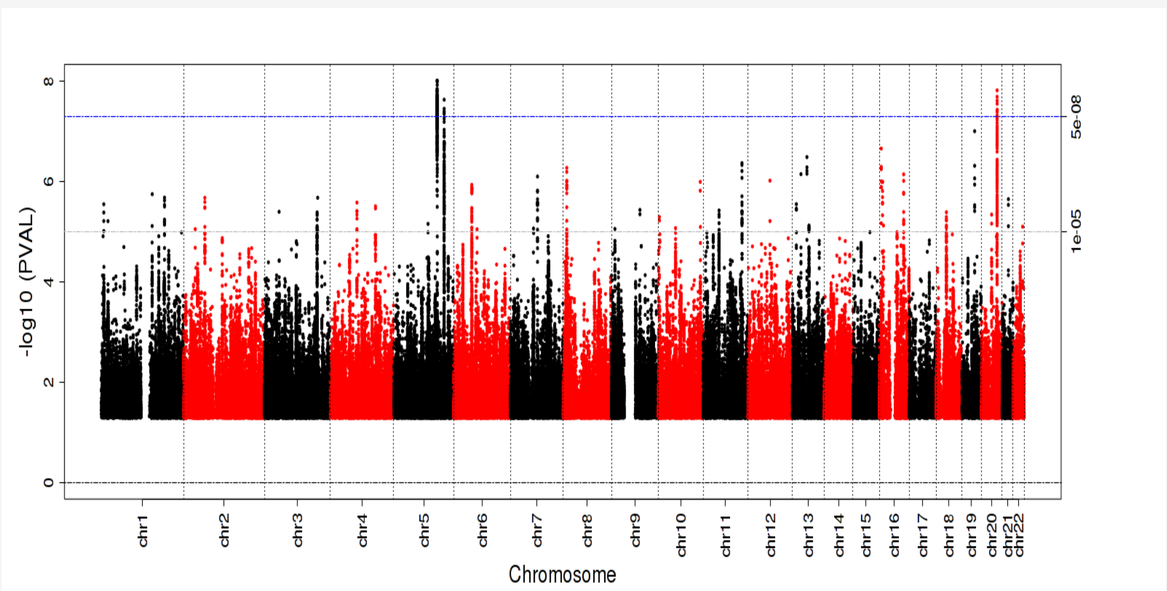
Conventional GWAS (N = 298,420)

In the original SWB GWAMA, test statistics from the regression of the two phenotypes (LS and PA) on individual SNPs are combined using:

$$Z_{LSPA} = \frac{Z_{LS} + Z_{PA}}{\sqrt{Var(Z_{LS}) + Var(Z_{PA}) + 2COV(Z_{LS}, Z_{PA})}}$$

$COV(Z_{LS}, Z_{PA}) = 0$

This aggregation included the aggregation of the test statistic itself, together with the aggregation of the variance in the estimate (or standard error) and resulted in the following Manhattan plot:



Multivariate GWAS (N = 346,503)

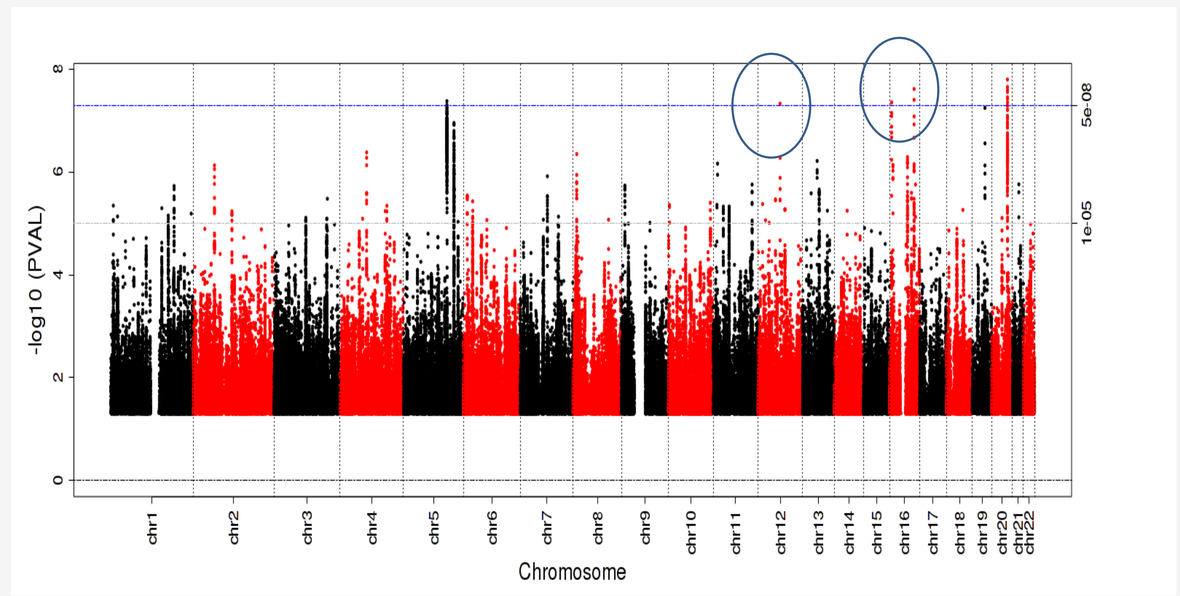
If the samples are not independent, information will be needed on the overlap between samples and the correlation between the traits to determine the genetic covariance. If this information cannot be directly obtained (usually the case), we can use linkage disequilibrium score regression (LDSC) to obtain the magnitude of inflation that is due to sample overlap using:

$$E[z_{1jz2j}] = \frac{\sqrt{N_1 N_2}\, r_g}{M} \ell_j + \frac{rNs}{\sqrt{N_1 N_2}}$$

Slope Cross-Trait Intercept

Using the full GWAS summary statistics from both LS (N = 166,290) and PA (N = 180,293), the cross-trait-regression intercept (CTI) can be calculated , representing the magnitude of inflation due to the sample-overlap between both included phenotypes. **Therefore, using this cross-trait-regression intercept, we can estimate the test-statistics of two correlated traits with sample dependency.**

$$Z_{LSPA} = \frac{Z_{LS} + Z_{PA}}{\sqrt{Var(Z_{LS}) + Var(Z_{PA}) + 2(\text{CTILSPA})}}$$



	Okbay et al. (2016)	Cross-trait-Intercept	No Correction
Lambda GC	1.206	1.207	1.351
Intercept	0.998	0.998	1.112

Conclusion

Our method showed that when two genetically correlated traits are meta-analyzed, violation of sample independence can be overcome by using the CTI obtained from LDSC regression as a correction factor. As a result, , it becomes easier to meta-analyze a spectrum of correlated traits of which only summary statistics are available and independency of samples is more difficult to guarantee.