**Supplementary Information for “Phenome-wide Heritability Analysis of the UK Biobank”**

Tian Ge1,2,3, Chia-Yen Chen2,3,4, Benjamin M. Neale2,3,4, Mert R. Sabuncu1,5,\*, and Jordan W. Smoller2,3,\*

1Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital / Harvard Medical School, Charlestown, MA 02129, USA;

2Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA;

3Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02138, USA;

4Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA;

5School of Electrical and Computer Engineering and School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA.

\*MRS and JWS contributed equally to this work.

Correspondence to: [tge1@mgh.harvard.edu](mailto:tge1@mgh.harvard.edu) (TG) or [jsmoller@hms.harvard.edu](mailto:jsmoller@hms.harvard.edu) (JWS)

**The Moment-matching Method for SNP Heritability Estimation**

We consider the linear random effect model , where an -dimensional trait is partitioned into the sum of additive genetic effects and unique (subject-specific) environmental effects . The covariance structure of is , where is the empirical genetic similarity matrix for each pair of individuals estimated from genome-wide SNP data, is an identity matrix, and are the total additive genetic variance captured by genotyped common SNPs and the variance of unique environmental factors across individuals, respectively.

To obtain unbiased estimates of and , we regress the empirical estimate of the phenotypic covariance onto the matrices and : , where is the matrix vectorization operator that converts a matrix into a vector by stacking its columns, and is the residual of the regression. The ordinary least squares (OLS) estimator of this multiple regression problem can be obtained by solving the linear system:

In the presence of covariates, i.e., , where is an covariate matrix and is a vector of fixed effects, an matrix always exists, which satisfies , , , and . Applying to both sides of the model removes the covariate matrix and gives . The covariance structure of the transformed trait is , and the linear system becomes

We note that for large sample size , the genetic similarity matrix and the residual forming matrix can be very large, making the computation of ,,, and memory intensive. To reduce the memory demand, we note that (1) the number of covariates is typically orders of magnitude smaller than ; (2) the quantities ,, , and can be computed by iteratively reading columns (or block columns) of into the memory; and (3) we have the following derivations:

Therefore, it can be seen that once ,, , and have been computed by iteratively loading columns (or block columns) of into the memory, all quantities in the linear system can be computed without manipulating any matrix. In particular, the residual forming matrix does not need to be explicitly computed. This makes the moment-matching algorithm computationally and memory efficient even if the sample size is very large.