

## Supplementary Materials for

### **The shared genetic architecture and evolution of human language and musical rhythm**

Gökberk Alagöz<sup>#1</sup>, Else Eising<sup>1</sup>, Yasmina Mekki<sup>2,3</sup>, Giacomo Bignardi<sup>1,4</sup>, Pierre Fontanillas<sup>5</sup>, 23andMe Research Team, Michel G. Nivard<sup>6</sup>, Michelle Luciano<sup>7</sup>, Nancy J. Cox<sup>3</sup>, Simon E. Fisher<sup>##1,8</sup>, Reyna L. Gordon<sup>##2,3,9,10</sup>

<sup>1</sup> Language and Genetics Department, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands

<sup>2</sup> Department of Otolaryngology - Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN, USA.

<sup>3</sup> Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.

<sup>4</sup> Max Planck School of Cognition, Leipzig, Germany

<sup>5</sup> 23andMe, Inc., Sunnyvale, CA, USA

<sup>6</sup> Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands.

<sup>7</sup> Department of Psychology, University of Edinburgh, Edinburgh, UK.

<sup>8</sup> Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 HB Nijmegen, The Netherlands

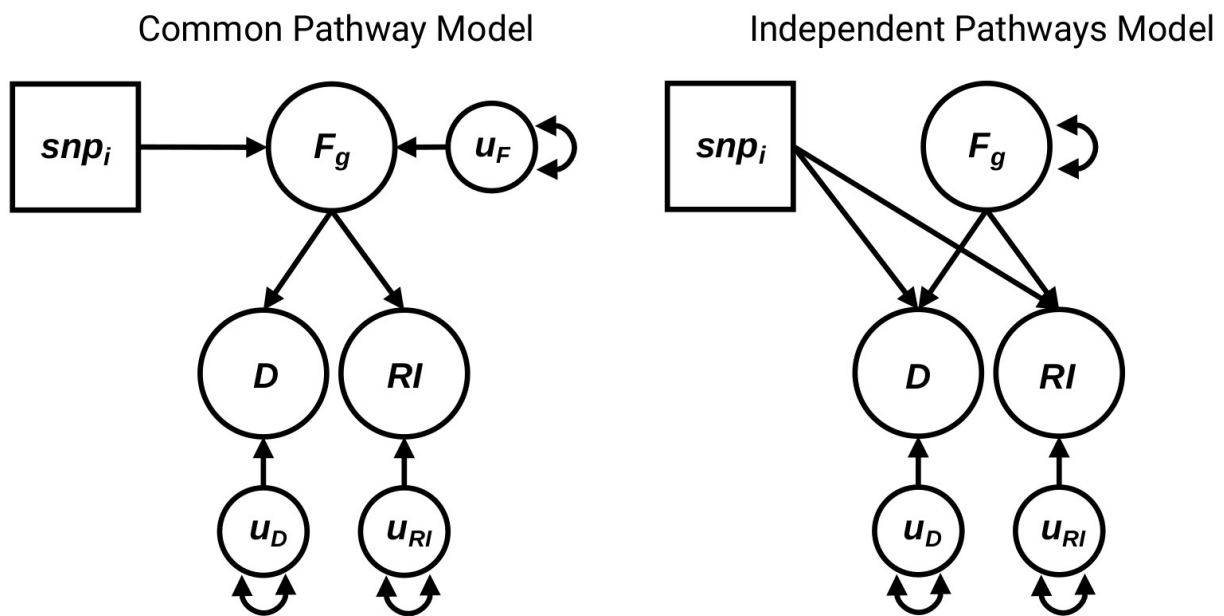
<sup>9</sup> Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA.

<sup>10</sup> The Curb Center, Vanderbilt University, Nashville, TN, USA.

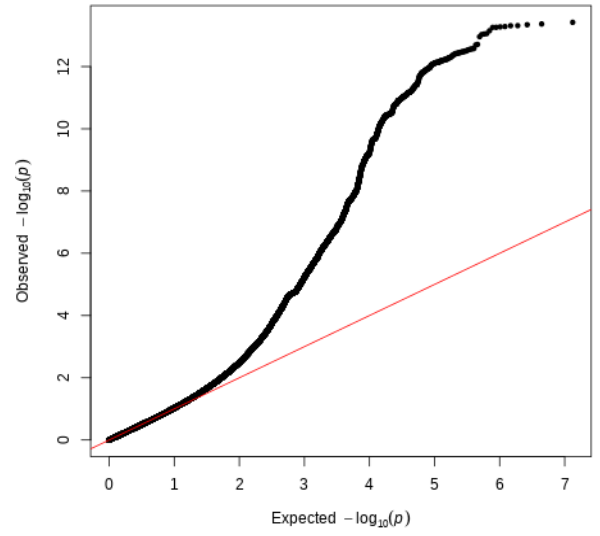
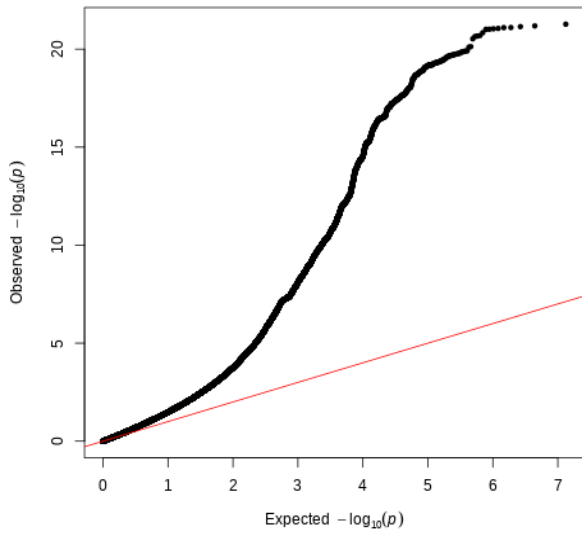
**# Correspondence: [goekberk.alagoez@mpi.nl](mailto:goekberk.alagoez@mpi.nl), [simon.fisher@mpi.nl](mailto:simon.fisher@mpi.nl), [reyna.gordon@alumni.usc.edu](mailto:reyna.gordon@alumni.usc.edu)**

**\* These authors contributed equally to this work.**

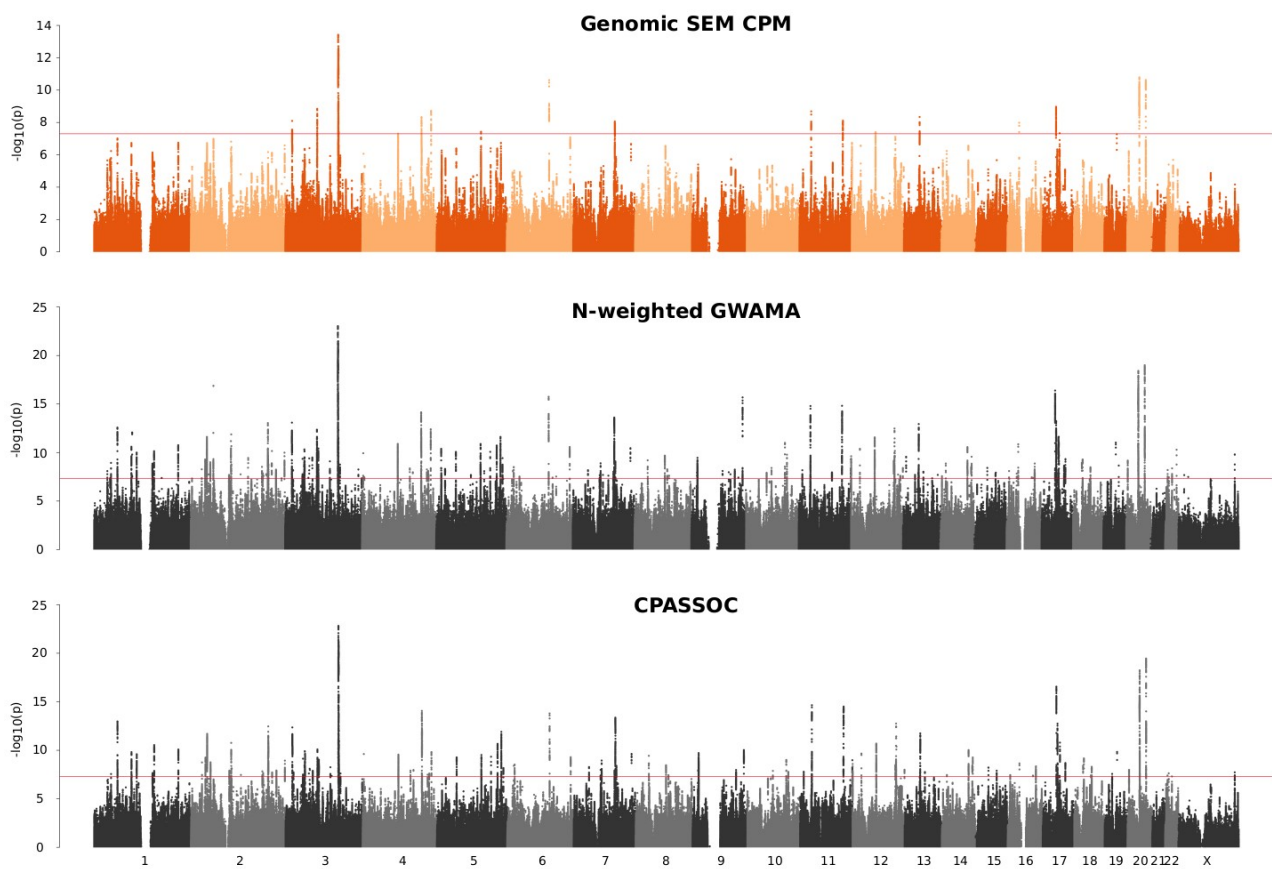
This PDF file includes:  
Supplementary Figures 1 to 8



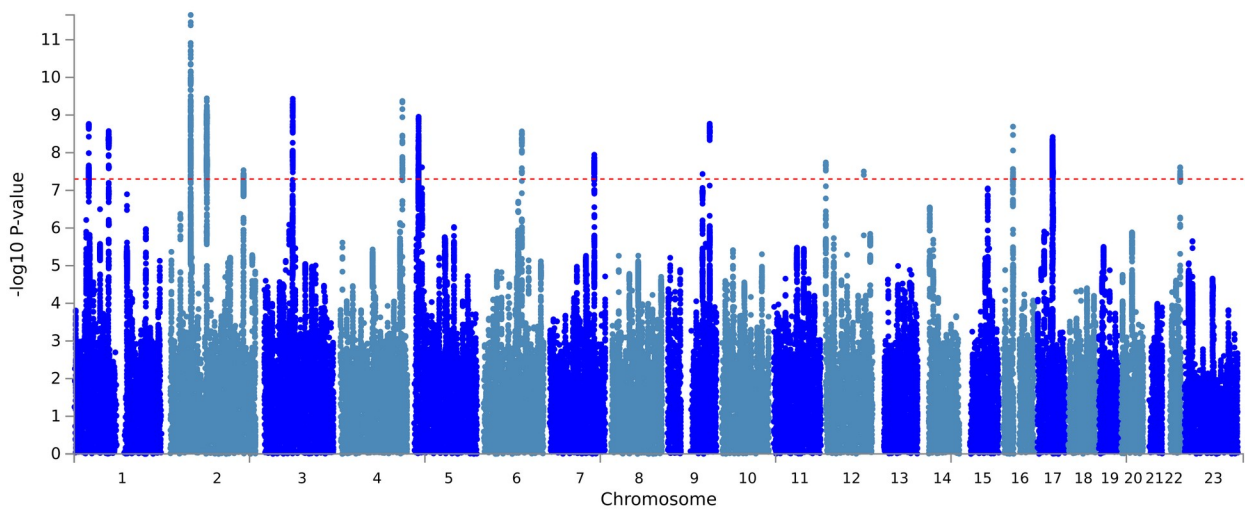
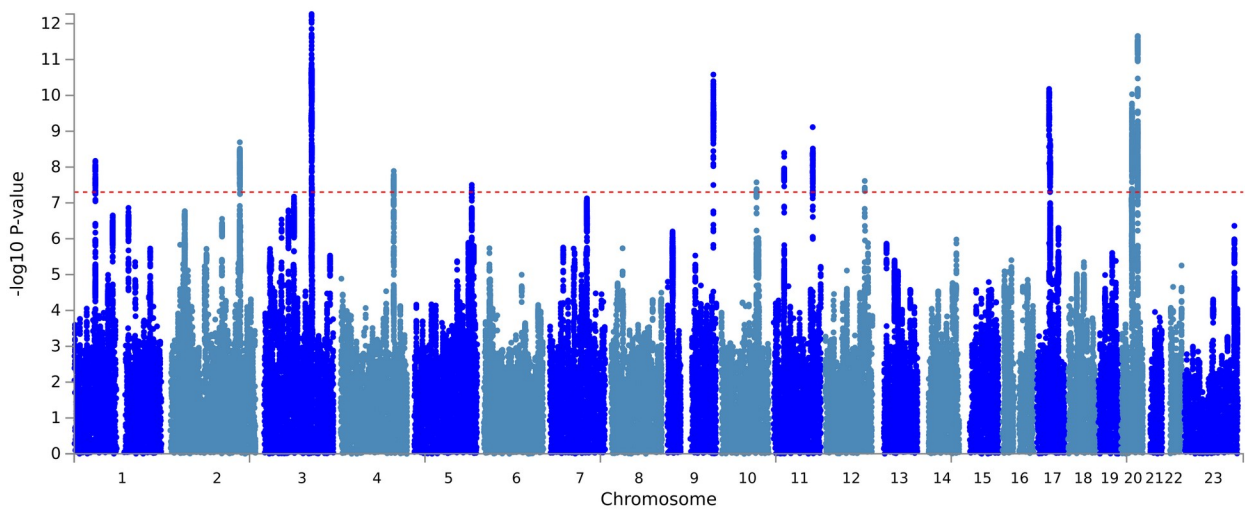
**Supplementary Figure 1:** Genomic SEM model diagrams for common pathway and independent pathways models. The models were used in the mvGWAS for the common factor and independent factor association analyses.  $D$ : dyslexia,  $RI$ : rhythm impairment,  $F_g$ : Common factor,  $u_D$ : residual variance of dyslexia,  $u_{RI}$ : residual variance of rhythm impairment,  $u_F$ : residual variance of the common factor,  $snp_i$ :  $i^{\text{th}}$  SNP regression.



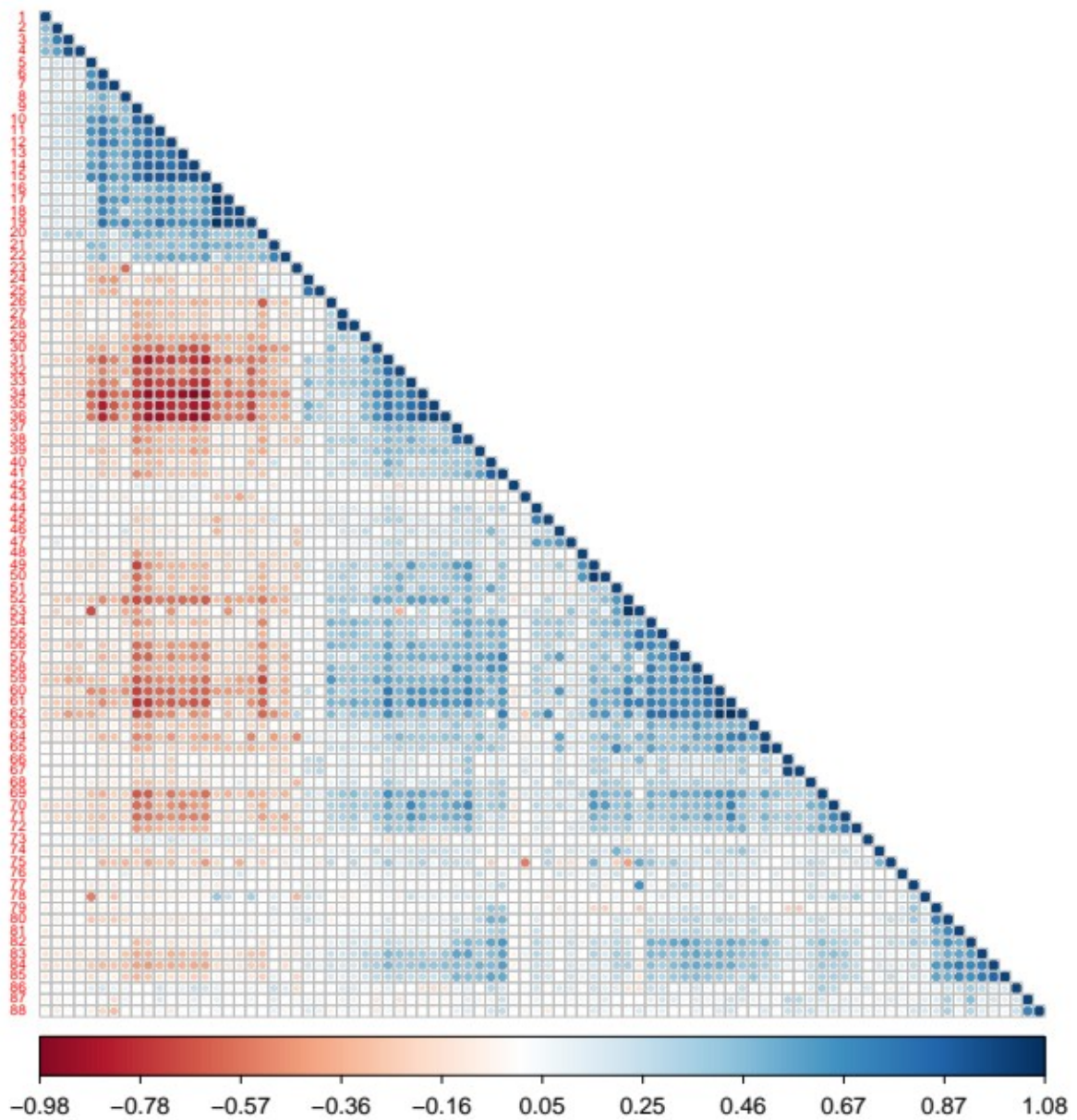
**Supplementary Figure 2:** QQ plots of  $F_{gRI-D}$  mvGWAS summary statistics *prior to* (left) and *after* (right) GC correction.



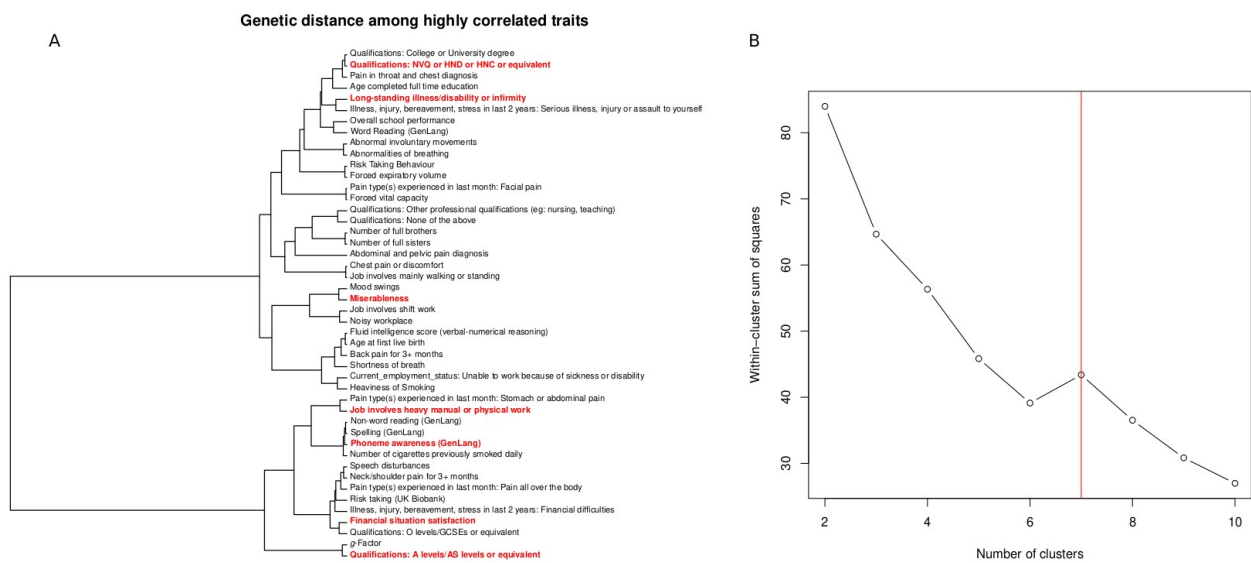
**Supplementary Figure 3:** Manhattan plots for Genomic SEM CPM ( $F_{gRI-D}$ ), N-weighted GWAMA, and CPASSOC. The red lines correspond to the genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ).



**Supplementary Figure 4:** Manhattan plots for the independent factors of dyslexia (top) and rhythm impairment (bottom). The red lines correspond to genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ).

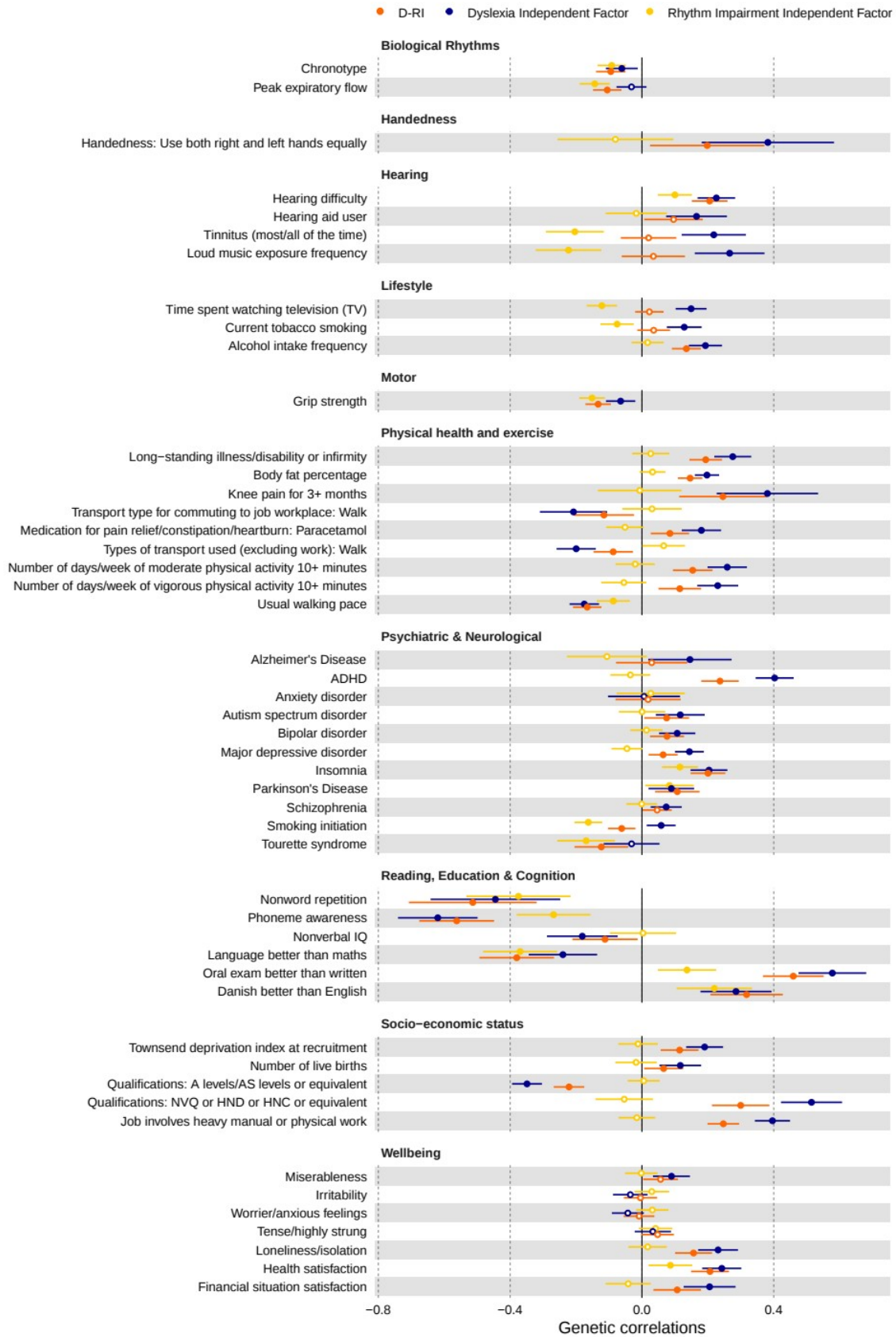


**Supplementary Figure 5:** Genetic correlations ( $r_g$ ) among 88 traits that are significantly correlated either with dyslexia or rhythm. Genetic correlations were estimated using LDSC. Supplementary Table 13 provides the list of traits included.



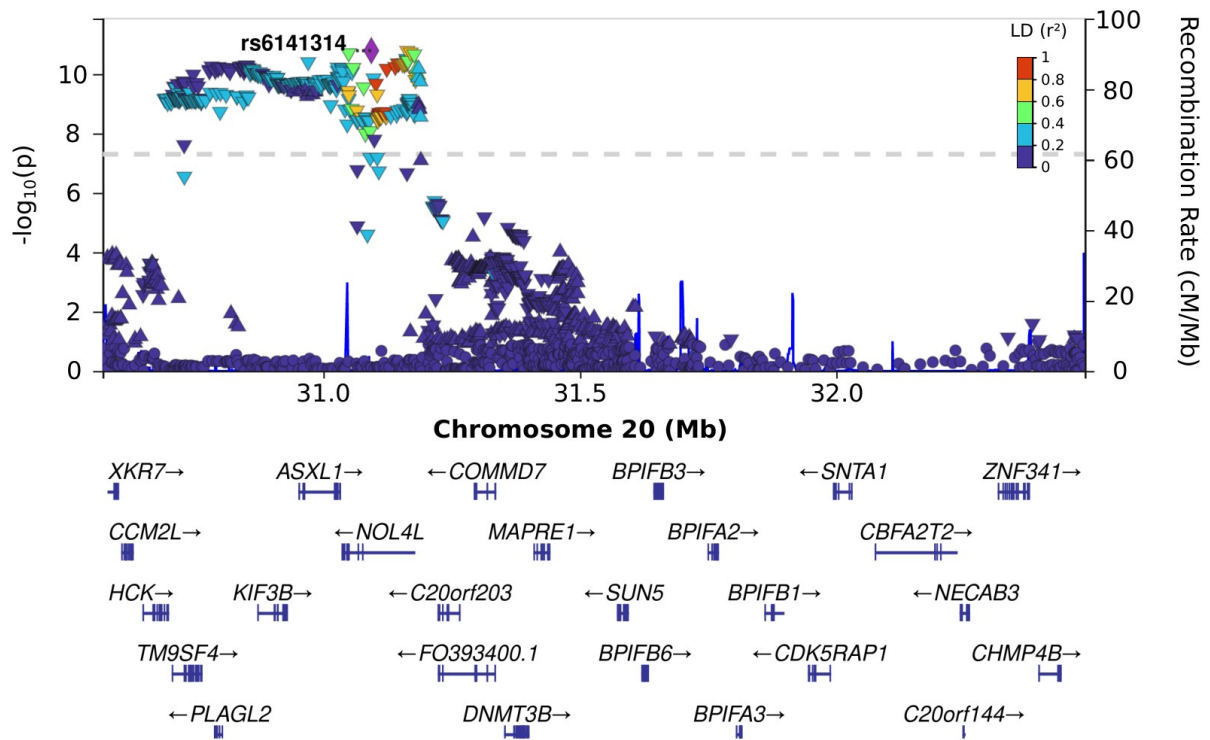
**Supplementary Figure 6:** (A) Dendrogram showing the hierarchical clustering of highly genetically correlated ( $|r_g| > 0.80$ ) traits. (B) Knee-point algorithm identified 7 representative clusters with. For each cluster, one representative trait (shown in bold red) was used in the genetic correlation analysis with the  $F_{gRI-D}$ .





**Supplementary Figure 7:** Genetic correlations ( $r_g$ ) between the selected 49 traits and  $F_{gRI-D}$  (D-RI) (orange) and dyslexia (blue) and rhythm (yellow) independent factors. Genetic correlations were estimated using LDSC. Full circles indicate significant correlations ( $P < 0.05$ ). Error bars represent standard errors.





**Supplementary Figure 8:** LocusZoom plot of chr20: 30,569,660-32,484,506 locus, the region which is identified by local genetic correlation analysis of  $F_{gRI-D}$  and Superior Longitudinal Fasciculus I. P-values represent  $F_{gRI-D}$  mvGWAS significance. Direction of the triangles represent effect directions. LD ( $r^2$ ) levels with rs6141314 are represented in colours. Grey dash-line indicate genome-wide significance level ( $P < 5 \times 10^{-8}$ ).