

Methods S1.

1. Genotyping and Quality Control for TEDS Sample

DNA from 8,122 individuals was extracted from saliva and buccal cheek swab samples and hybridized to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. Raw image data were pre-processed in GenomeStudio according to Illumina Exome Chip SOP v1.4. (<http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4>). Prior to genotype calling, 919 multimapping SNPs and 501 samples with call rate < 0.95 were removed. Following initial QC, the program ZCALL was used to augment genotype calling.

DNA from 3,747 individuals was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. From the extracted DNA samples, 3,665 samples were successfully hybridized to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer.

Raw image data were pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (<https://www.wtccc.org.uk/cc2/>). All pre-processing was conducted according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Following initial QC, the program CHIAMO was used for genotype calling (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

After initial quality control, the same quality control was performed on samples from each of the platforms (Illumina and Affymetrix) separately using PLINK (Purcell et al., 2007), R (Team, 2015) and BCFtools (Li, 2011) and EIGENSOFT (Patterson, Price, & Reich, 2006; Price et al., 2006).

DNA samples were excluded from subsequent analyses on the basis of call rate (< 0.98) suspected non-European ancestry, the presence of severe medical or psychiatry problems or severe medical complications during early gestation and relatedness other than dizygotic twin status. SNPs were excluded if the minor allele frequency was < 0.5%, if more than 2% of genotype data were missing, or if the Hardy Weinberg p -value was lower than 10^{-5} . Non-autosomal markers and indels were also removed. Association between SNP and the platform, batch, plate or well on which samples were genotyped was calculated; SNPs with an effect p -value < 10^{-4} were excluded.

A total sample of 10,346 samples, including 7,026 unrelated individuals from which 3,320 individuals had a genotyped dizygotic co-twin remained. Genotype data following quality control was available for 4,776 individuals and 559,772 SNPs from the illumine array and 2,250 individuals and 635,269 SNPs from the Affymetrix array.

Genomewide genotypes from the two arrays were separately phased using EAGLE2 (Loh et al., 2016) and imputed using the Haplotype Reference Consortium (McCarthy et al., 2016) using the Positional Burrows-Wheeler Transform method (Durbin, 2014) and the imputation software Minimac3 1.0.13 (Fuchsberger, Abecasis, & Hinds, 2015), which are available from the Michigan Imputation Server (<https://imputationserver.sph.umich.edu>). A series of quality checks

were performed before merging data from the two arrays and variants with $\text{info} < 0.75$ were excluded and SNPs that were non-overlapping between platforms were removed.

After merging, minor allele frequency differences were tested for between platforms and SNPs with an effect p -value $< 10^{-4}$ were removed. Those SNPs with a Hardy Weinberg p -value $> 10^{-5}$ were also removed. Following these criteria, 7,363,646 genotyped and well-imputed SNPs were retained for analyses. Only unrelated individuals were included in the present analyses. To ease high computational demands by the software LDpred (Vilhjálmsón et al., 2015) for polygenic scoring in large samples, we further excluded SNPs with $\text{info} < 1$, leaving 515,100 SNPs for analysis.

2. Creating polygenic scores using LDpred

Genome wide polygenic scores in TEDS were calculated using the Bayesian approach, LDpred, which has been shown to outperform predictive accuracy of the conventional clumping and p -value thresholding approach (Vilhjálmsón et al., 2015). Here, a posterior effect size is derived for each SNP by re-weighting the original summary statistic coefficient by the relative influence of a SNP given its level of linkage disequilibrium with surrounding SNPs and a prior on the effect size of each SNP. The prior is based on the heritability of the trait and the fraction of markers assumed to casually influence the trait. GPS is then calculated as the sum of the trait-increasing alleles weighted by their posterior effect size estimate. Unlike the conventional clumping and thresholding approach, LDpred retains all SNPs common between GWA summary statistics and genotype data in the target sample.

For the present study we applied a causal fraction of 1, which assumes that all SNPs contribute to the development of the trait. Due to the high computational demand of LDpred, especially in larger sample sizes including many SNPs, we made further restrictions to our analyses, only including the 515,100 SNPs that were perfectly imputed (info score of 1) to reduce analytical load. Only genotypes of unrelated individuals were used to estimate LD structure in our sample because levels of LD are considerably higher in relatives compared to unrelated individuals (Vattikuti, Guo, & Chow, 2012).

The GPS were adjusted for the first ten principal components of the genotype data, chip, batch and plate effects using the regression method, and the standardized residuals were used for all analyses.