Supplementary Methods

Exon sequence data. Quality control and filtering of the genotype data were performed using the VCFtools software (Danecek et al., 2011) and R (RCoreTeam, 2013). Sites were included into the study based on read quality (Phred-scaled quality > 20), the quality of the bases surrounding the variant (median minimum base quality > 0), strand bias (binomial p-value for strand bias test > .1), mapping quality (root mean square of mapping qualities of reads at the variant position > 20), average variant quality (variant quality / read depth for the variant > 5), probability that the variant segregates in the data (Phred-scaled posterior probability > 20), and proportion of missing data per site. Sites were retained if the sample missingness per site did not exceed 5% (i.e. if at least 95% of samples had at least one read at the site), and sites with sample missingness between 5% and 20% were retained only if their mean depth per sample exceeded 20x. The application of these criteria reduced the number of variants to 2900 (including 972 singletons and 61 doubletons) and the number of genes to 168 (from an initial pool of 175). The frequency distribution of the 2900 variants, and their distribution over the 168 genes, are given in Figures 1 and 2. The mean depth was ~212x, with an average of ~2% sites missing per individual.

SNP data. The SNP data for the polygenic prediction were obtained from a larger NTR dataset (N=14,003; see Lin et al., 2014; Nivard et al., 2013). Buccal or blood samples for DNA extraction were collected as part of multiple projects within NTR. DNA extraction and purification were performed at various points in time, following several manufacturer-specific protocols. Genotyping of several partly overlapping subsets was performed on multiple platforms. Chronologically, the following

platforms were used: Affymetrix Perlegen 5.0, Illumina 370, Illumina 660, Illumina Omni Express 1M, and Affymetrix 6.0. Genotype calls were made using platform-specific software (APT Genotyper, Beadstudio (Illumina)).

Quality control was performed within and between platforms and subsets. For each platform, the individual SNPs were lifted over to build 37 (HG19) of the Human reference genome using the LiftOver tool (Kuhn et al., 2007). SNPs that did not map at all, had ambiguous locations, or did not have matching (or strand-opposite) alleles were removed. Subsequently, the data were strand-aligned with the 1000 Genomes phase 1 release v3 panel. SNPs were removed based on mismatches of alleles with those in the reference set, differences of allele frequencies (larger than .2) from those in the reference set, minor allele frequency (<1%), absence of Hardy-Weinberg equilibrium ($p<.00001$), and call rate (<95%). Samples were excluded from the dataset based on mismatch of expected sex and the sex derived from the genotype data, genotype missing rate (>10%), and the coefficient of inbreeding ($F>0.10$ or < -0.10).

Following these steps, the data from the individual arrays were merged into a single dataset using PLINK 1.07 (Purcell, et al., 2007). Within the merged set, identity by state (IBS) sharing was calculated between all possible pairs of individuals; subsequently, IBS was compared to the known family structure within the NTR. Samples that did not display the expected IBS sharing were removed. DNA samples that were typed on multiple platforms and displayed discordance between the overlapping SNPs (concordance rate <99%) were removed. Subsequently, a single DNA sample was selected from each MZ twin pair, resulting in a total of 12,240 unique DNA samples to be imputed. Hardy-Weinberg equilibrium, minor allele frequency, and allele frequency (>.2) filters were re-applied to the merged data. To

avoid erroneous strand alignment, SNPs with the allele combinations C/G and A/T and a minor allele frequency between .35 and .5 were removed.

Phasing of the samples and the imputation of the SNPs missing across some platforms were performed using MACH 1.0 (Li & Abecasis, 2006). The phased data were imputed using Minimac (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012), in batches of ~500 individuals for 561 chromosome chunks obtained using the CHUNKCHROMOSOME software (Liu, Li, Wang, & Li, 2013). After imputation, the data on DNA-confirmed MZ twins were duplicated back into the dataset, resulting in a dataset containing a total of 14,003 individuals. The mean imputation R^2 was 0.38. The imputed dataset contained 30,051,533 autosomal SNPs. Post-imputation, SNPs were filtered based on Mendelian error rate in families (>2%), Hardy-Weinberg equilibrium (<0.00001), imputation quality R^2 (<.3), minor allele frequency (<.005), and a discrepancy in allele frequencies of the imputed SNPs and the 1000 Genomes reference panel (>.15). This resulted in a final dataset containing 7,981,681 autosomal SNPs with a mean R^2 of .86.