OXFORD

Human Molecular Genetics, 2021, Vol. 30, No. 8 727–738

doi: 10.1093/hmg/ddab053
Advance Access Publication Date: 22 January 2021
General Article

GENERAL ARTICLE

Imputed gene expression risk scores: a functionally informed component of polygenic risk

Oliver Pain[1,2,*], Kylie P. Glanville[1], Saskia Hagenaars[1], Saskia Selzam[1], Anna Fürtjes[1], Jonathan R. I. Coleman[1], Kaili Rimfeld[1], Gerome Breen[1,2], Lasse Folkersen[3] and Cathryn M. Lewis[1,2,4]

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK, [2]NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, London SE5 8AF, UK, [3]Institute of Biological Psychiatry, Sankt Hans Hospital, Copenhagen 4000 Roskilde, Denmark and [4]Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London WC2R 2LS, UK

*To whom correspondence should be addressed at: Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience - PO80, De Crespigny Park, Denmark Hill, London SE5 8AF, UK. Tel: +44 2078485299; Email: oliver.pain@kcl.ac.uk

Abstract

Integration of functional genomic annotations when estimating polygenic risk scores (PRS) can provide insight into aetiology and improve risk prediction. This study explores the predictive utility of gene expression risk scores (GeRS), calculated using imputed gene expression and transcriptome-wide association study (TWAS) results. The predictive utility of GeRS was evaluated using 12 neuropsychiatric and anthropometric outcomes measured in two target samples: UK Biobank and the Twins Early Development Study. GeRS were calculated based on imputed gene expression levels and TWAS results, using 53 gene expression–genotype panels, termed single nucleotide polymorphism (SNP)-weight sets, capturing expression across a range of tissues. We compare the predictive utility of elastic net models containing GeRS within and across SNP-weight sets, and models containing both GeRS and PRS. We estimate the proportion of SNP-based heritability attributable to *cis*-regulated gene expression. GeRS significantly predicted a range of outcomes, with elastic net models combining GeRS across SNP-weight sets improving prediction. GeRS were less predictive than PRS, but models combining GeRS and PRS improved prediction for several outcomes, with relative improvements ranging from 0.3% for height ($P = 0.023$) to 4% for rheumatoid arthritis ($P = 5.9 \times 10^{-8}$). The proportion of SNP-based heritability attributable to *cis*-regulated expression was modest for most outcomes, even when restricting GeRS to colocalized genes. GeRS represent a component of PRS and could be useful for functional stratification of genetic risk. Only in specific circumstances can GeRS substantially improve prediction over PRS alone. Future research considering functional genomic annotations when estimating genetic risk is warranted.

Received: December 1, 2020. **Revised:** February 8, 2021. **Accepted:** February 15, 2021

Introduction

Polygenic risk scores (PRS) are a useful research tool and a promising opportunity for personalized medicine (1). A PRS indicates an individual's genetic liability for an outcome and is traditionally calculated as the genome-wide association study (GWAS) effect size-weighted sum of alleles (2). The correlation between genetic variants, termed linkage disequilibrium (LD), should be accounted for when estimating PRS. LD-based clumping is often used to obtain LD-independent variants, though more recent methods that estimate the joint effect of variants to account for LD have been shown to improve prediction (3). The predictive utility of PRS can be further increased by incorporating prior probability distributions on causal effect sizes, thereby reducing the signal-to-noise ratio (4). PRS can also be enhanced using prior distributions that incorporate evidence of enrichment for functional categories (5). Further research is required to evaluate alternative strategies that integrate the functional effect of genetic variation when calculating polygenic scores.

A wealth of research has shown enrichment in GWAS of expression quantitative trait loci (eQTLs), variants affecting gene expression (6,7). The eQTL studies have identified many genetic variants associated with differential gene expression (8,9). Integration of eQTL and GWAS summary statistics enables inference of gene expression changes associated with the GWAS phenotype, an approach called transcriptome-wide association study (TWAS) (10,11). TWAS aggregates the effect of genetic associations in a functionally informed manner to highlight associated up-/down-regulated genes within the context of a specific tissue or developmental stage (12). Due to the functionally informed aggregation of individual genetic effects, TWAS can identify novel associations not previously identified as significant in the corresponding GWAS. This approach has been useful for highlighting plausible candidate genes for experimental follow-up (13).

There has been limited research investigating the predictive utility of PRS that considers the effect of each variant on gene expression. One approach is to split genetic variants into high- and low-prior groups based on whether they are eQTLs, and then calculate the PRS using a range of mixing parameters to optimally weight the contribution of high-prior variants (14). This approach of reweighting eQTL variants improved prediction over functionally agnostic PRS in type 2 diabetes. An alternative approach is to calculate gene-expression risk scores (GeRS), which consider the joint effect of variation on each gene's expression (13). GeRS are calculated as the sum of predicted expression for an individual weighted by the TWAS-based effect size, analogous to PRS except using predicted expression instead of individual genotypes, and TWAS effect size instead of GWAS effect size. GeRSs were shown to significantly predict schizophrenia, with GeRS derived using prefrontal cortex eQTL data explaining the most variance compared with other individual tissues, but a model containing GeRS based on multiple tissues providing the largest variance explained. However, whether GeRS can improve prediction in combination with PRS was not investigated. A recent study reports that the genetically regulated transcriptome is a component of broader genetic variation, but modelling these sources of variance separately improved out-of-sample prediction (15). This finding suggests that a GeRS will capture a component of PRS, but modelling GeRS and PRS separately will improve prediction.

Previous research has shown that GeRS can explain significant variance in schizophrenia, and that modelling variance explained by the genetically regulated transcriptome could improve prediction over models considering the genome alone. However, GeRS have only been applied to schizophrenia, and no previous study has combined GeRS with PRS. In this study, we evaluate the predictive utility of GeRS calculated using the TWAS-based approach with eQTL data from a range of tissues. We apply the method to a range of quantitative traits and binary disorders in two samples, UK Biobank (UKB) (16) and the Twins Early Development Study (TEDS) (17). Furthermore, we evaluate whether GeRS provide novel information over PRS and explore the effect of stratifying genes by colocalization estimates of pleiotropy and tissue specificity of eQTL effects.

Results

Predictive utility of GeRS

For the five disorders and seven quantitative traits analyzed in UKB and TEDS, the GeRS calculated were significantly correlated with each phenotype. GeRS were most predictive of height in TEDS using the genotype-tissue expression (GTEx) nerve tibial single nucleotide polymorphism (SNP)-weight set with a correlation between predicted and observed values of 0.22 (SE = 0.01, P-value = 6.8×10^{-61}). The predictive utility of GeRS typically increased as more relaxed P-value thresholds were used to select genes (Fig. 1, Supplementary Material, Figs S1 and S2). The predictive utility of GeRS for outcomes available in both UKB and TEDS, height and body mass index (BMI) were broadly consistent.

Combining GeRS across P-value thresholds in an elastic net model significantly improved prediction over the single best GeRS P-value threshold for all outcomes in UKB except Depression and IBD (Fig. 2A, Supplementary Material, Table S1). The largest improvement in prediction when modelling multiple P-value thresholds was for T2D in UKB (23.6% improvement, P-value = 2.2×10^{-28}). Modelling GeRS across multiple P-value thresholds did not improve prediction for any outcome in the TEDS sample, and led to a significant decrease in prediction for GCSE (6.1% reduction, P-value = 1.9×10^{-3}) (Supplementary Material, Fig. S3A and Table S1).

Modelling GeRS derived using multiple SNP-weights significantly improved prediction over any single SNP-weight set for all outcomes except Depression in UKB, and GCSE and attention deficit hyperactivity disorder (ADHD) symptoms in TEDS (Fig. 2B, Supplementary Material, Fig. S3B and Table S1). Significant relative improvements provided by modelling GeRS from multiple SNP-weight sets varied from 7.1% (P-value = 2.4×10^{-2}) for height in TEDS to 29% (P-value = 3.5×10^{-27}) for RheuArth in UKB.

Comparison of SNP-weight sets

The predictive utility of GeRS derived using each SNP-weight set separately is shown in Supplementary Material, Figures S4 and S5. Often the most predictive GeRS were derived using SNP-weight sets capturing expression in tissues previously implicated for the outcome, such as CMC DLFPC for depression and BMI in UKB. However, the predictive utility of GeRS showed a strong relationship with the size of the sample used to derive the SNP-weights ($r_{\text{Pearson}} = 0.15$ in UKB), and the number of features within the SNP-weight sets ($r_{\text{Pearson}} = 0.29$ in UKB) (Supplementary Material, Figs S6 and S7). When fitting both the SNP-weight set sample size and number of features in a joint model, the effect of sample size was no longer significant. After correcting for the number of features in each SNP-weight set, the most

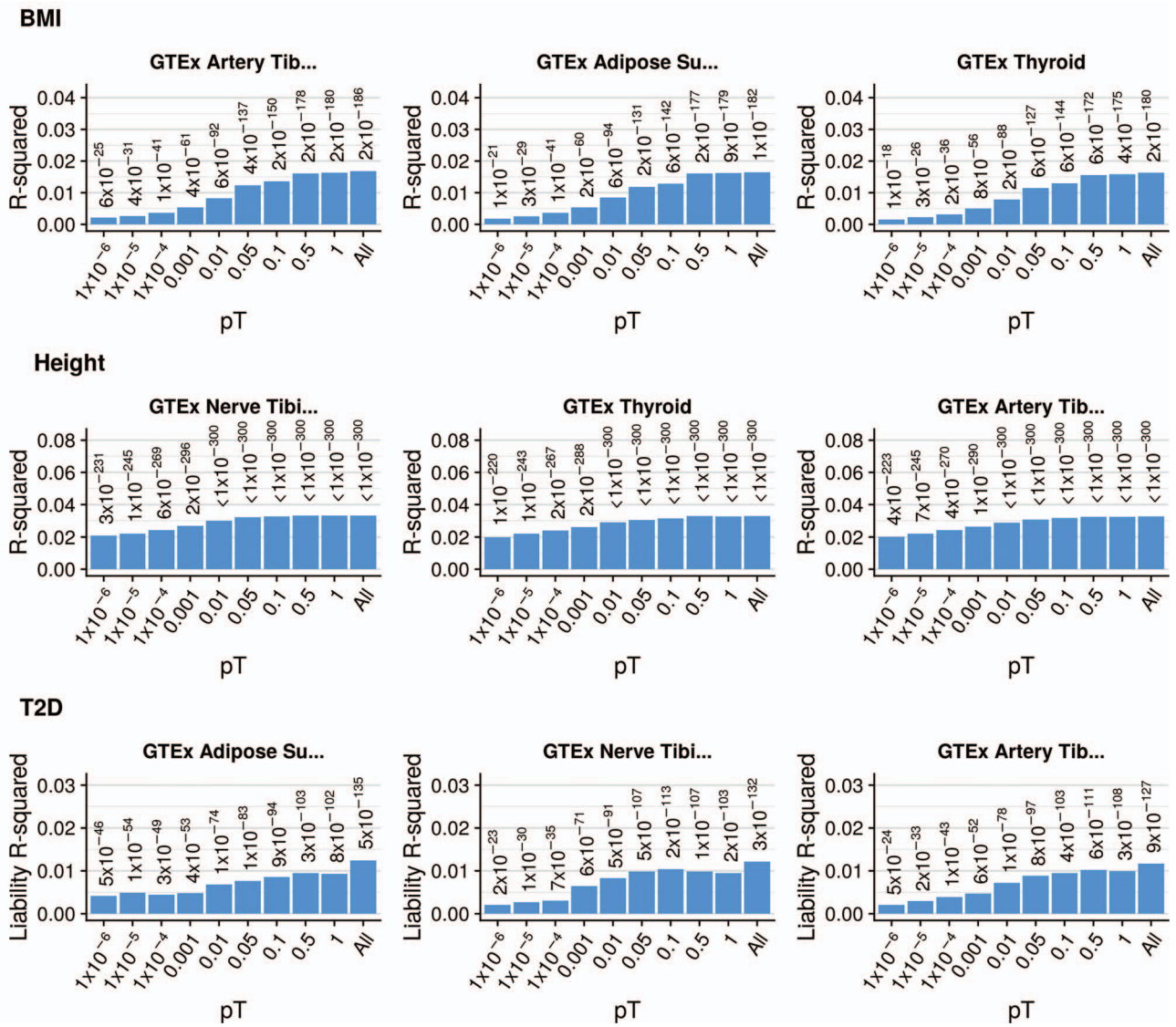

Figure 1. Variance explained (R-squared) by GeRS for three outcomes in UKB across *P*-value thresholds. The all bar indicates the variance explained by an elastic net model including all *P*-value thresholds. GeRS are based on a single SNP-weight set. Figure only shows results for the three SNP-weights with the highest variance explained values. Values above bars are *P*-values indicating whether the variance explained is significantly different from zero.

predictive SNP-weight set varied for most outcomes (Supplementary Material, Figs S8 and S9). For example, the most predictive SNP-weight set for depression was GTEx thyroid but changed to CMC DLPFC after accounting for the number of features within each SNP-weight set. The CMC DLPFC splicing SNP-weight set was often the least predictive after correcting for the number of features due to features often capturing multiple splice variants for a given gene which are therefore highly redundant.

Stratifying by colocalization and tissue specificity

TWAS associations can be driven by the same causal variant driving the association with both gene expression and the phenotype (vertical or horizontal pleiotropy), or the associations can be driven by LD between different causal variants affecting each outcome. As a result, TWAS associations do not necessarily indicate that the observed differential expression of a gene is associated with the outcome. Colocalization estimates of whether both gene expression and the outcome are affected by the same causal variant (PP4), were used to determine whether restricting GeRS to colocalized associations altered the predictive utility of GeRS. We found GeRS restricted to colocalized genes (PP4 > 0.8) had reduced predictive utility compared with unrestricted GeRS (Fig. 3, Supplementary Material, Figs S10–S12).

Cis-eQTL effects are moderately correlated across tissues (8), meaning GeRS for a given SNP-weight set will capture variance attributable to other tissues. To explore the predictive utility of tissue-specific GeRS, we restricted GeRS to genes either not significantly heritable in blood SNP-weight sets, or genes whose predicted expression was uncorrelated with the corresponding feature in the blood SNP-weight sets. We found restricting GeRS to tissue-specific features reduced the predictive utility of GeRS based on individual SNP-weight sets, but the predictive utility of models including all SNP-weight sets did not change substantially (Supplementary Material, Figs S10 and S11).

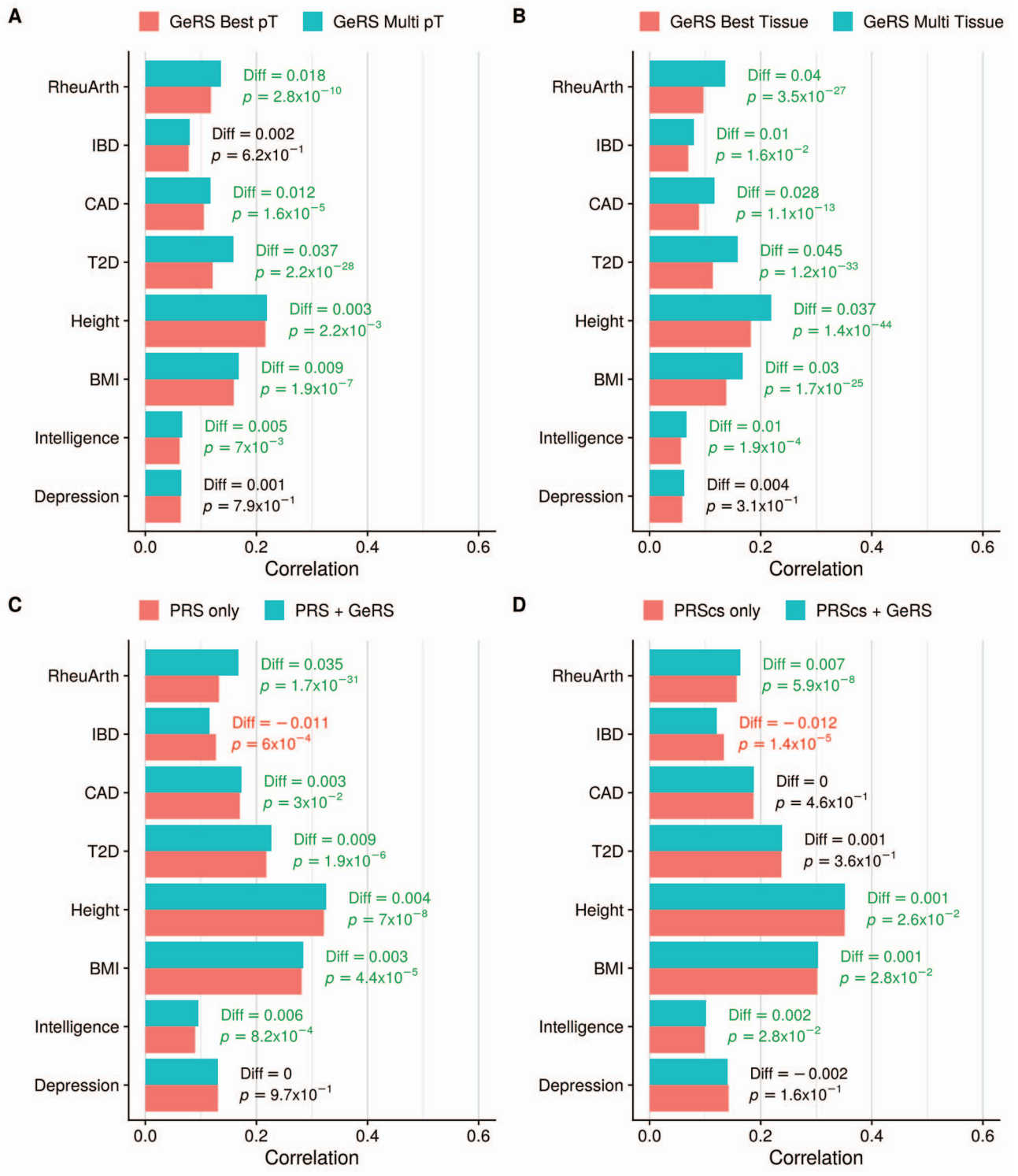

Figure 2. Comparing the predictive utility of GeRS and PRS in UKB. (**A**) Compares the predictive utility of models containing GeRS across SNP-weight sets based on the single best *P*-value threshold and models containing GeRS across all *P*-value thresholds. (**B**) Compares the predictive utility of models containing GeRS across *P*-value thresholds based on the single best SNP-weight set and models containing GeRS based on all SNP-weight sets. (**C**) Compares the predictive utility of models containing PRS and models containing GeRS and PRS. (**D**) Compares the predictive utility of models containing models also containing PRS derived using PRScs, and models also containing GeRS. Values on the right of each bar indicate the absolute difference in predicted–observed correlation between the full and nested model.

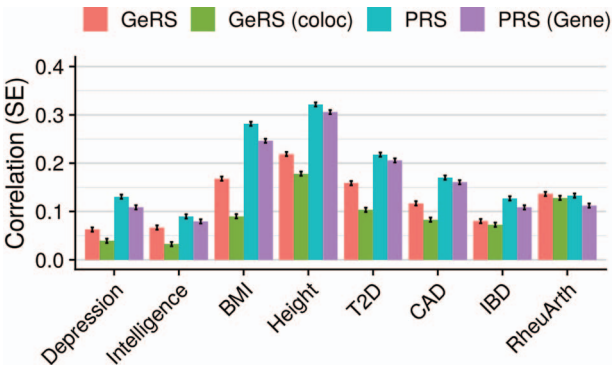

Figure 3. Shows the correlation between predicted and observed values in UKB for models. GeRS = All SNP-weight set GeRS; GeRS (coloc) = All SNP-weight set GeRS restricted to genes with a colocalization PP4 > 0.8; PRS = Genome-wide PRS; PRS (Gene) = PRS restricted to gene regions considered by GeRS.

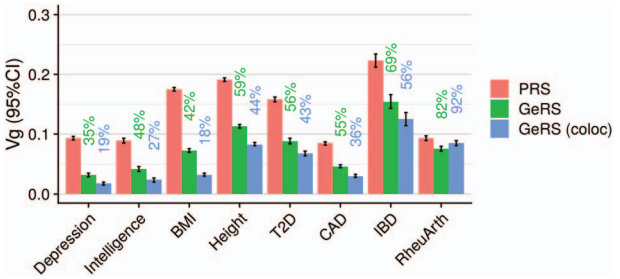

Figure 4. Estimates of SNP-based heritability and GE-based heritability for outcomes in UKB. PRS indicates the SNP-heritability as estimated using PRS association results in AVENGEME. GeRS indicates the GE-based heritability as estimated using GeRS association results in AVENGEME. GeRS (coloc) indicates the GE-based heritability as estimated using GeRS when restricted to genes with colocalization PP4 > 0.8. The value above each bar indicates the proportion of SNP-based heritability accounted for by *cis*-regulated expression.

Comparison of GeRS with PRS

Compared with PRS-only models, models containing PRS and multi-SNP-weight set GeRS provided statistically significant improvements in prediction for all outcomes in UKB except depression and IBD (Fig. 2C, Supplementary Material, Table S1). Inclusion of GeRS did not significantly improve prediction over PRS-only models for any outcome in TEDS (Supplementary Material, Fig. S3C and Table S1). Statistically significant relative improvements varied from 1% (P-value = 4.4×10^{-5}, correlation increased from 0.281 to 0.284) for BMI in UKB to 20.8% for RheuArth in UKB (P-value = 1.7×10^{-31}, correlation increased from 0.133 to 0.168). Inclusion of GeRS significantly decreased the correlation between observed and predicted values for IBD in UKB (−9.6%, P-value = 6×10^{-4}). We then explored whether GeRS improve prediction over PRS derived using PRScs, which models LD to estimate the joint effect of nearby variants, as opposed to LD-based clumping which removes variants in LD. When comparing GeRS with PRScs scores, the improvement in prediction provided by GeRS was attenuated for all outcomes, although statistically significant relative improvements remained when including GeRS for RheuArth (4%), height (0.3%), BMI (0.4%) and intelligence (2.5%) in UKB (Fig. 2C, Supplementary Material, Table S1).

A distinction between the pT + clump PRS and GeRS is how they handle the MHC region. The pT + clump PRS retain a single variant in the MHC region. In contrast, GeRS retain a single gene in the MHC region, which considers information across multiple variants. Given the large genetic effects in the MHC region for RheuArth, we performed a sensitivity analysis to explore whether the approach of retaining only the single variant in the MHC region is responsible for the improved prediction when including GeRS. The analysis showed that inclusion of GeRS still significantly improved prediction of RheuArth over PRS alone (P-value = 6.40×10^{-11}), though the relative improvement was attenuated from 20.8% to 8.4%.

When comparing the predictive utility of all SNP-weight set GeRS to PRS, we found the proportion of PRS–phenotype correlation that GeRS can explain (r_{GeRS}/r_{PRS}) was between 44.6% for BMI in TEDS, and 102.6% for RheuArth in UKB (Fig. 3, Supplementary Material, S12 and Table S2). When restricting GeRS to colocalized genes (PP4 > 0.8), the proportion of PRS–phenotype correlation that GeRS can explain reduced to between −2.8% for ADHD in TEDS and 96.4% for RheuArth in UKB. The predictive utility of PRS stratified to include only variants within gene boundaries was reduced compared with unstratified PRS, but still greater than the GeRS for all outcomes except RheuArth.

Estimating heritability explained by *cis*-heritable expression

AVENGEME estimated the SNP-based heritability of each phenotype based on PRS associations, with values ranging from 7.8% (95%CI = 7.2–8.4%) for CAD in UKB to 27.9% (95%CI = 25.9–30.0%) for height in TEDS (Fig. 4, Supplementary Material, Fig. S13 and Table S3). AVENGEME estimated the phenotypic variance explained by *cis*-heritable expression based on GeRS associations (GE-based heritability), returning estimates between 3.2% (95%CI = 2.9–3.5%) for depression in UKB and 15.4% (95%CI = 14.3–16.6%) for IBD in UKB (Fig. 4 and Supplementary Material, Fig. S13 and Table S3). The proportion of SNP-based heritability explained by *cis*-heritable expression ranged from 26% for BMI in TEDS to 82% for RheuArth in UKB (Fig. 4 and Supplementary Material, Fig. S13 and Table S3). When restricting GeRS to colocalized features, the proportion of SNP-based heritability explained by *cis*-heritable expression ranged from 3% for ADHD in TEDS to 92% for RheuArth in UKB.

Estimates of the proportion of variants with no causal effect on the trait were broadly consistent when using PRS or GeRS, with PRS-based estimates ranging from 76.1% (95%CI = 70.9–80.6%) for GCSE in TEDS and 96.4% (95%CI = 95.9–96.9%) for IBD in UKB (Table S3).

Discussion

This study has characterized the predictive utility of GeRS, an approach that leverages gene expression summary statistics, GWAS summary statistics and target sample genotype data to infer genetic risk conferred via *cis*-regulated gene expression. We investigate factors affecting the predictive utility of GeRS, test whether GeRS can improve prediction over PRS alone, and estimate the proportion of SNP-based heritability that can be accounted for by *cis*-regulated expression. Our findings indicate GeRS represent a component of PRS, with GeRS explaining a substantial proportion of variance explained by PRS, suggesting GeRS may provide a useful approach for stratifying genetic

risk into functional categories. Furthermore, this study finds GeRS generally provide small improvements in prediction over PRS alone, though GeRS can more substantial improvements in specific circumstances.

Prediction using GeRS versus PRS

GeRS typically explained less phenotypic variance than PRS derived using the same GWAS summary statistics. However, for several outcomes, linear models combining GeRS and PRS did improve prediction over PRS alone. GeRS typically provided relative improvements of 1–6% for the correlation between predicted and observed phenotype values, although for rheumatoid arthritis GeRS provided a 20.8% improvement when combined with pT + clump scores. All improvements in prediction provided by inclusion of GeRS were attenuated when using PRScs scores, which models LD as opposed to LD-based clumping, with GeRS only providing a 4% improvement for rheumatoid arthritis.

This pattern of results is likely due to the different method's approaches and ability to jointly model variants in LD. The attenuated improvement for rheumatoid arthritis when using PRScs is particularly pronounced due to the methods ability to model effects within the MHC region as there are well-documented and strong HLA allele effects within the MHC region for rheumatoid arthritis (18). The large effect of HLA alleles is confirmed in the rheumatoid arthritis TWAS results (Supplementary Material, Table S4). The PRScs method models all variation within the MHC region, where as pT + clump PRS considered only the strongest associated variant within the MHC region. In contrast, GeRS jointly models variants integrating their effect of gene expression, and then retains the single lead gene. This explanation is supported by our sensitivity analysis showing the gain in prediction for rheumatoid arthritis was also attenuated when compared with pT + clump PRS that were not clumped to a single variant within the MHC region. Nonetheless, the GeRS approach still provides some advantage over PRS in all cases, indicating that the functionally informed approach used by GeRS for jointly modelling variants better captures the risk in the MHC region than pT + clump or even PRScs can, possibly due to the documented eQTL effects in the locus altering expression of relevant HLA genes (19). Therefore, these results suggest GeRS can provide novel information over PRS alone, albeit in specific circumstances where multiple variants affecting gene expression are the causal risk factor. Given that the gene expression SNP-weights used in GeRS are derived using linear models, further improvement may be provided by using non-linear models that can capture haplotypes more effectively (20).

Inclusion of GeRS did not significantly improve prediction over PRS alone in the TEDS sample for any outcome. GeRS showed a similar correlation with height and BMI as was found in the UKB. These findings indicate the non-significant improvement in prediction when including GeRS is due to the smaller sample size of TEDS compared with UKB as this will reduce the power to detect small increases in prediction between models, and increase the likelihood of overfitting due to the large number of predictors in the model compared with the number of individuals in the sample. The increased likelihood of overfitting may also be responsible for the decrease in prediction accuracy for IBD when including GeRS, as the number of cases for this phenotype was relatively low. Approaches to efficiently integrate transcriptomic data without using many predictors would help alleviate this issue.

Opportunities provided by GeRS

Although GeRS explain less variance than PRS, they may provide novel opportunities over PRS for several reasons. Firstly, GeRS are a gene-based genetic risk score, meaning the GeRS are well suited to stratification by biological pathways or other gene-based characteristics. Gene-based polygenic scores can also be created by restricting the variants considered to those proximal to genes (21). However, genetic variation proximal to a gene may have no effect on the gene's expression or function. Secondly, GeRS are sensitive to the properties of the original gene expression dataset used to derive the expression SNP-weights and can capture gene expression within a range of contexts such as tissues and developmental stages. Therefore, GeRS could serve as a useful predictor for stratifying individuals based on the underlying aetiology of their disorder, addressing the possible that criticism of functionally agnostic PRS, that they are disconnected from aetiological considerations. Complex disorders are heterogenous at the phenotypic level and at the aetiological level. For example, it may be possible to stratify individuals based on the specific tissue underlying their condition.

Factors affecting predictive utility of GeRS

Models containing GeRS derived using multiple tissues improve prediction over the single best tissue, congruent with a previous study (13).

Furthermore, we found the relative predictive utility of GeRS derived using different SNP-weight sets was strongly correlated with the number of genes captured by the SNP-weight set. This is likely due to a multitude of factors including the sample size and quality of the original gene expression dataset. Both of these factors will increase the number of genes captured by the SNP-weight set by detecting more genes with significantly heritable *cis*-regulated expression, and increase the variance in gene expression the SNP-weights explain out-of-sample. It is likely that the relevance of the tissue to the outcome is also an important factor influencing the predictive utility of an outcome, however, the sample size and number of features have a larger effect on the predictive utility of GeRS due to the moderately correlated *cis*-regulated expression across tissues enabling tissues irrelevant to the outcome to act as a proxy for gene expression within relevant but unavailable tissues.

Quantifying heritability accounted for by *cis*-regulated expression

GeRS capture only a small amount of novel phenotypic variance compared with PRS, indicating that GeRS largely represent a component of PRS. These findings are congruent with a previous study modelling the genome and genetically regulated transcriptome using CORE GREML (15). We estimate the proportion of phenotypic variance that can be explained by *cis*-regulated gene expression and compare the results with SNP-based heritability estimates using PRS results. Across the phenotypes, we estimate *cis*-regulated gene expression explains 26–82% of SNP-based heritability. However, due to LD GeRS are likely to capture effects mediated through other mechanisms. To more accurately estimate the proportion of SNP-based heritability accounted for by *cis*-regulated expression, we restricted the analysis to genes that colocalize and their association is therefore unlikely to be driven by linkage. When restricting the analysis to colocalized genes, we estimated 3–92% of SNP-based heritability was accounted for

by *cis*-regulated expression. For most outcomes, this supports previous research showing strong enrichment of eQTLs in GWAS summary statistics (6,7). Our findings suggest that restricting GeRS to colocalized genes will reduce their predictive utility but may provide a more accurate estimate of an individual's risk mediated via *cis*-regulated expression. This raises a further issue for gene-based polygenic scores, as they are more liable to capturing linkage effects and there is no option to restrict analyses to colocalized genes. Even when restricting our analysis to colocalized genes, our estimates of phenotypic variance attributable to *cis*-regulated expression may still be upwardly biased due to GeRS capturing effects driven by horizontal pleiotropy as opposed to vertical pleiotropy (mediation). An example of horizontal pleiotropy would be where a disease-associated variant is an eQTL for a gene, but the variant confers risk for the disease via another mechanistic route, such as trans eQTL effects. A recently developed method called mediated expression score regression (MESC) can be used to identify the variance explained by vertical pleiotropy (mediation) alone (22). Indeed, the results reported by MESC are lower than the estimates based on GeRS in this study.

Opportunities for GeRS based on observed expression

Although the colocalization and tissue specificity of genes did not improve prediction of GeRS when based on *predicted* expression in the target sample, restricting genes by these criteria is likely to improve the predictive utility of GeRS derived using *observed* gene expression in the target sample. This is supported by a previous study which found GeRS derived using GWAS summary statistics and eQTL data, and *observed* gene expression data, could substantially improve prediction over PRS but only when using eQTL data from the relevant tissue and restricting the risk scores to colocalized genes (23). Tissue specificity and colocalization is more important when integrating with observed gene expression as the GeRS must capture genuine differences in expression associated with the outcome. Future research exploring the predictive utility of GeRS derived using TWAS results and observed expression is warranted.

In summary, this study has demonstrated that GeRS explain a substantial proportion of variance for a range of outcomes, with multiple tissue GeRS explaining more variance than the single best tissue. Furthermore, we demonstrate that GeRS can improve prediction of outcomes over PRS alone in specific circumstances, where multiple eQTL effects must be considered to fully capture the genetic risk conferred by a locus. However, the results largely indicate that GeRS capture a component of risk captured by functionally agnostic PRS, and estimates of variance explained by *cis*-regulated expression is 26–82% of total SNP-based heritability, although these estimates likely captures risk not only mediated via *cis*-regulated expression due to horizontal pleiotropy and linkage. In conclusion, GeRS may serve as a useful research tool by providing a novel opportunity to stratify genetic risk by expression within specific tissues, developmental stage and other gene-based characteristics.

Materials and Methods

UK Biobank

UKB is a prospective cohort study that recruited >500 000 individuals aged between 40 and 69 years across the United Kingdom (16). The UKB received ethical approval from the North West – Haydock Research Ethics Committee (reference 16/NW/0274).

Genetic data. UKB released imputed dosage data for 488 377 individuals and ~96 million variants, generated using IMPUTE4 software (16) with the Haplotype Reference Consortium reference panel (24) and the UK10K Consortium reference panel (25). This study retained individuals that were of European ancestry based on four-means clustering on the first two principal components provided by the UKB, had congruent genetic and self-reported sex, passed quality assurance tests by UKB, and removed related individuals (>third degree relative, KING threshold > 0.044) using relatedness kinship (KING) estimates provided by the UKB (16). The imputed dosages were converted to hard-call format for all variants.

Phenotype data. Eight UKB phenotypes were analyzed. Five phenotypes were binary: depression, Type 2 diabetes (T2D), coronary artery disease (CAD), inflammatory bowel disease (IBD) and rheumatoid arthritis (RheuArth). Three phenotypes were continuous: intelligence, height, and BMI. Further information regarding outcome definitions can be found in the Supplementary Material.

Analysis was performed on a subset of ~50 000 UKB participants for each outcome to reduce the computational burden of the analysis whilst maintaining sufficient power to perform downstream analyses. For each continuous trait (intelligence, height, BMI), a random sample was selected. For disease traits, all cases were included, except for depression and CAD where a random sample of 25 000 cases was selected. Controls were randomly selected to obtain a total sample size of 50 000. Sample sizes for each phenotype after genotype data quality control are shown in Table 1.

TEDS

The TEDS is a population-based longitudinal study of twins born in England and Wales between 1994 and 1996 (17). Ethical approval for TEDS has been provided by the King's College London ethics committee (reference: 05/Q0706/228). Parental and/or self-consent was obtained before data collection. For this study, one individual from each twin pair was removed to retain only unrelated individuals.

Genetic data. As previously described (26), TEDS genotype data underwent stringent quality control prior to imputation via the Sanger Imputation server using the Haplotype Reference Consortium reference data (24). Imputed genotype dosages were converted to hard-call format using a hard call threshold of 0.9, with variants for each individual set to missing if no genotype had a probability of >0.9. Variants with an INFO score < 0.4, MAF < 0.001, missingness > 0.05 or Hardy–Weinberg equilibrium *P*-value $<1 \times 10^{-6}$ were removed.

Phenotypic data. This study used four continuous phenotypes within TEDS: height, BMI, educational achievement (GCSE) and ADHD symptom score (Table 1). Further information regarding the phenotype definitions can be found in the supplementary material and a previous study (27).

Genotype-based scoring

GeRS and PRS were calculated within a reference-standardized framework, whereby the resulting PRS and GeRS are not influenced by target sample specific properties including availability of variants and measurements of LD and allele frequency. This

Table 1. Sample size for each target sample phenotype

	Description	Total sample size	No. of controls	No. of cases
UKB phenotype				
Depression	Major depression	49 995	24 999	24 996
Intelligence	Fluid intelligence	49 998	NA	NA
BMI	Body mass index	49 993	NA	NA
Height	Height	49 993	NA	NA
T2D	Type-2 diabetes	49 990	35 102	14 888
CAD	Coronary artery disease	49 991	24 998	24 993
IBD	Inflammatory bowel disease	50 000	46 539	3461
RheuArth	Rheumatoid arthritis	50 000	46 592	3408
TEDS phenotype				
GCSE	Mean GCSE scores	7296	NA	NA
ADHD	ADHD symptoms	7880	NA	NA
BMI21	Body mass index at age 21	5220	NA	NA
Height21	Height at age 21	5455	NA	NA

is achieved by using a common set of typically well imputed variants (HapMap3) and using reference genetic data (European 1KG Phase 3) to estimate LD and allele frequencies. Lastly, all genotype-based scores are scaled and centred based on the mean and standard deviation of scores in the reference sample. This reference-standardized approach and its merits have been described previously (3).

A schematic representation of calculating GeRS is shown in Figure 5.

GWAS summary statistics. GWAS summary statistics were identified for phenotypes the same as or similar as possible to the UKB and TEDS phenotypes (descriptive statistics in Supplementary Material, Table S5), excluding GWAS with documented sample overlap with the target samples. GWAS summary statistics were formatted using the LD-score regression munge_sumstats.py script (see Web Resources) with default settings (listed in the Supplementary Material) except the minimum INFO score was set to 0.6.

Transcriptome-wide association study. FUSION software (10) was used to integrate GWAS summary statistics with precomputed SNP-weights of gene expression to infer differential gene expression associated with the GWAS-phenotype. The term *SNP-weight* refers to a multi-SNP-based predictor of a gene's expression. SNP-weights used in this study were derived using gene expression data from a range of distinct tissues and European-ancestry adulthood samples, downloaded from the FUSION website (see URLs). The weights pertained to five RNA reference samples: (i) the GTEx Consortium (Version 7) (8), measuring gene expression across 48 tissues, including brain regions, blood and peripheral tissues, (ii) The CommonMind Consortium (CMC) (9), measuring expression and differential splicing in the dorsolateral prefrontal cortex, (iii) The Netherlands Twins Register (NTR) (28) and (iv) Young Finns Study (YFS) (10), which both provide information on blood tissue gene expression, and (v) Metabolic Syndrome in Men (METSIM) (10), assessing adipose tissue expression. The SNP-weights obtained from a given sample and tissue (e.g. GTEx thyroid, NTR peripheral blood) are referred to as *SNP-weight sets*. Characteristics for the 53 SNP-weight sets used are available in Supplementary Material, Table S6. The SNP-weights include 260 598 *features* (SNP-weight set and gene pairs), capturing expression of 26 434 unique genes (protein-coding and non-protein coding). The number of features that could be reliably imputed for each GWAS is shown in Supplementary Material, Table S7. TWAS was performed using default settings in FUSION and LD estimates from the European subset of the 1KG Phase 3 reference sample ($N = 503$).

Colocalization analysis tests whether the association between a genetic locus and two or more traits is driven by the same causal variant, or whether the association for each trait is driven by different causal variants that are in LD. Colocalization was performed using the coloc R package (29), implemented within the FUSION software, to estimate the posterior probability that the GWAS phenotype and gene's expression share a single causal variant, termed PP4. A coloc *P*-value threshold of 0.05 was used, to perform colocalization for all features with a TWAS *P*-value < 0.05.

Predicting expression in target samples. The *cis*-heritable component of expression for each gene was imputed in each target sample using the same gene expression SNP-weights described above, and target sample genotype data. Predicted expression levels are calculated as

$$\text{GeneExp}_j = \sum_{j=1}^{m} X_j \times \beta_j \tag{1}$$

where the predicted level of expression (GeneExp) for an individual is the number of effect alleles carried by the individual (X_j) weighted by the effect of each variant on gene expression as estimated from penalized regression model (β), across m variants. This was implemented using the FUSION script 'make_score.R' to convert the TWAS SNP-weights into PLINK score file format, and then using PLINK to carry out the scoring in the target sample. Predicted expression levels are then centred and scaled based on the mean and standard deviation of the predicted expression in the 1KG Phase 3 European reference sample.

Gene expression risk scoring. GeRS were calculated as.

$$\text{GeRS} = \sum_{j=1}^{g} \text{GeneExp}_j \times Z_j \tag{2}$$

where the GeRS of an individual is equal to the TWAS effect size (Z)-weighted sum of the individual's predicted expression

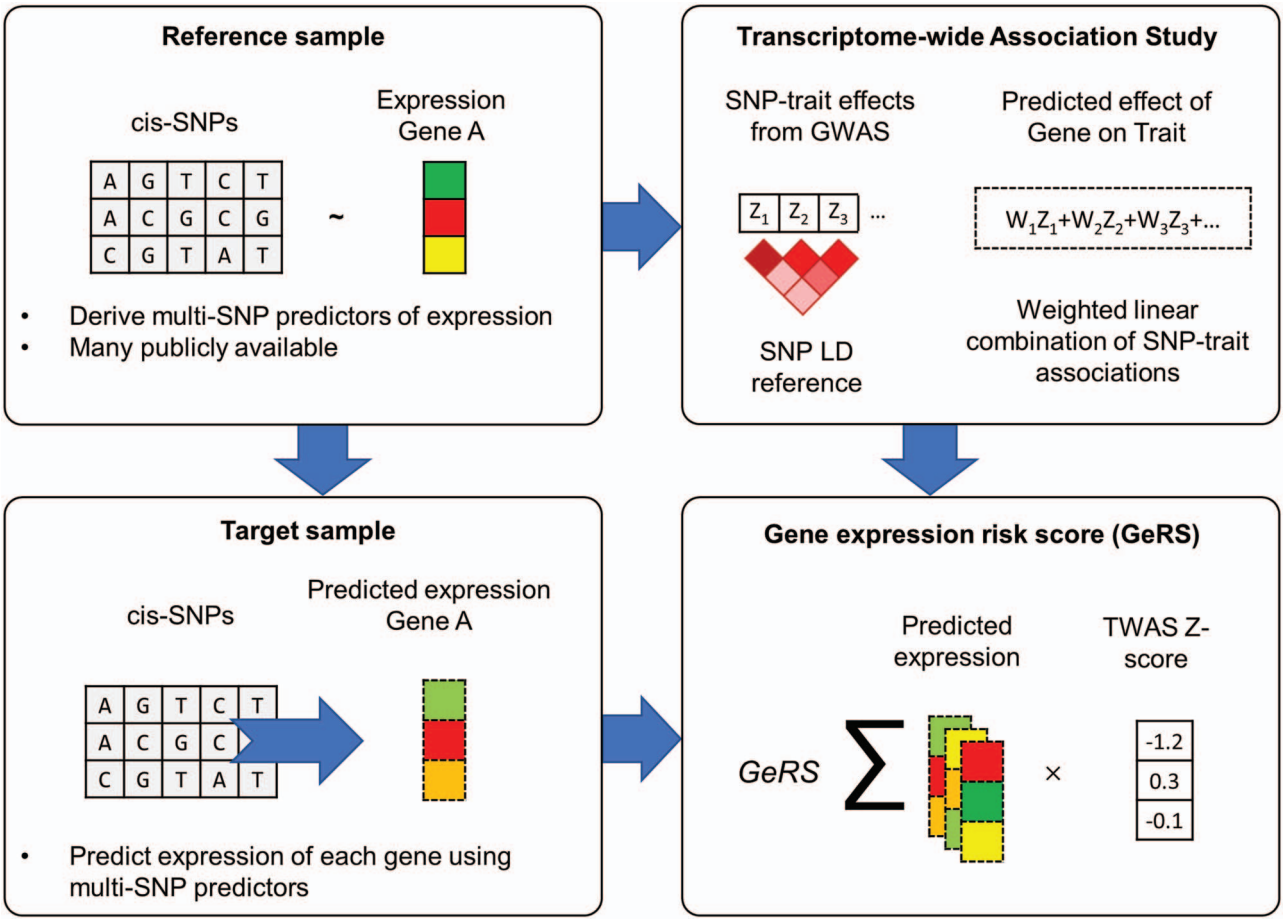

Figure 5. Schematic representation of GeRS Calculation. The top left panel describes the process of deriving SNP-weights predicting gene expression in a sample of individuals with both genotype and gene expression measured (e.g. GTEx consortium). These SNP-weights can be used to perform TWAS, whereby SNP-weights are integrated with GWAS summary statistics to infer gene expression associations with the GWAS trait (upper right panel). The SNP-weights can also be used to predict gene expression levels in a target sample with only genotype data available (lower left panel). Finally, GeRS can be calculated in the target sample by combining the level of predicted expression in each individual weighted by the TWAS effect size (lower right panel). This figure has been adapted from Gusev *et al.* (10).

($GeneExp_j$), at g genes. GeRS were calculated for each SNP-weight set separately, meaning 53 GeRS for each GWAS/TWAS phenotype were generated. To remove genes with highly correlated predicted expression due to LD, genes were ranked by TWAS P-value and clumping was performed to remove genes with a predicted expression $R^2 > 0.9$ within 500 kb of the lead gene boundaries. Within the MHC region, the single most significant gene was retained due to long range and complex LD structures. Predicted expression estimates used for clumping were estimated in the European 1KG Phase 3 reference. A range of nested P-value thresholds were used to select genes considered in the GeRS: 1, 5×10^{-1}, 1×10^{-1}, 5×10^{-2}, 1×10^{-2}, 1×10^{-3}, 1×10^{-4}, 1×10^{-5} and 1×10^{-6}. Scripts used to perform GeRS can be found on the GenoPred website (see URLs).

In addition, we evaluate the predictive utility of GeRS restricted to genes with evidence of colocalization with the outcome (PP4 > 0.8), and GeRS restricted to genes showing tissue specific expression. Tissue-specific GeRS were derived by only considering genes that were either not significantly heritable in blood SNP-weight sets (GTEx Whole blood, YFS or NTR), or genes whose predicted expression was uncorrelated with the corresponding feature in the blood SNP-weight sets ($R^2 < 0.01$). This approach is congruent with a previous study identifying tissue specific eQTL effects prior to risk scoring (23). Blood-specific features were identified using the same criteria but comparing predicted expression across all non-blood SNP-weight sets. The number of tissue-specific features for each SNP-weight set are listed in Table S6.

Polygenic risk scores. Polygenic scoring was carried out using the traditional P-value thresholding and LD-based clumping approach (pT + clump), and a more recent method, PRScs (30), which models LD between genetic variants and applies shrinkage parameters to avoid overfitting. PRScs has been previously reported to out-perform other polygenic scoring methods (3). The pT + clump approach was used as it is analogous to the approach used to remove highly correlated features in the GeRS. pT + clump was performed using an R^2 threshold of 0.1 and window of 250 kb. Within the MHC region (28–34 Mb on chromosome 6), the pT + clump method retains only the single most significant variant due to long range of complex LD in this region. A range of P-value thresholds were used to select variants: 1×10^{-8}, 1×10^{-6}, 1×10^{-4}, 1×10^{-2}, 0.1, 0.2, 0.3, 0.4, 0.5 and 1. PRScs were performed using a range of global shrinkage parameters (1×10^{-6}, 1×10^{-4}, 1×10^{-2} and 1) and the fully Bayesian mode, which estimates the optimal shrinkage parameter. Analogous to the GeRS, only HapMap3 variants were considered during

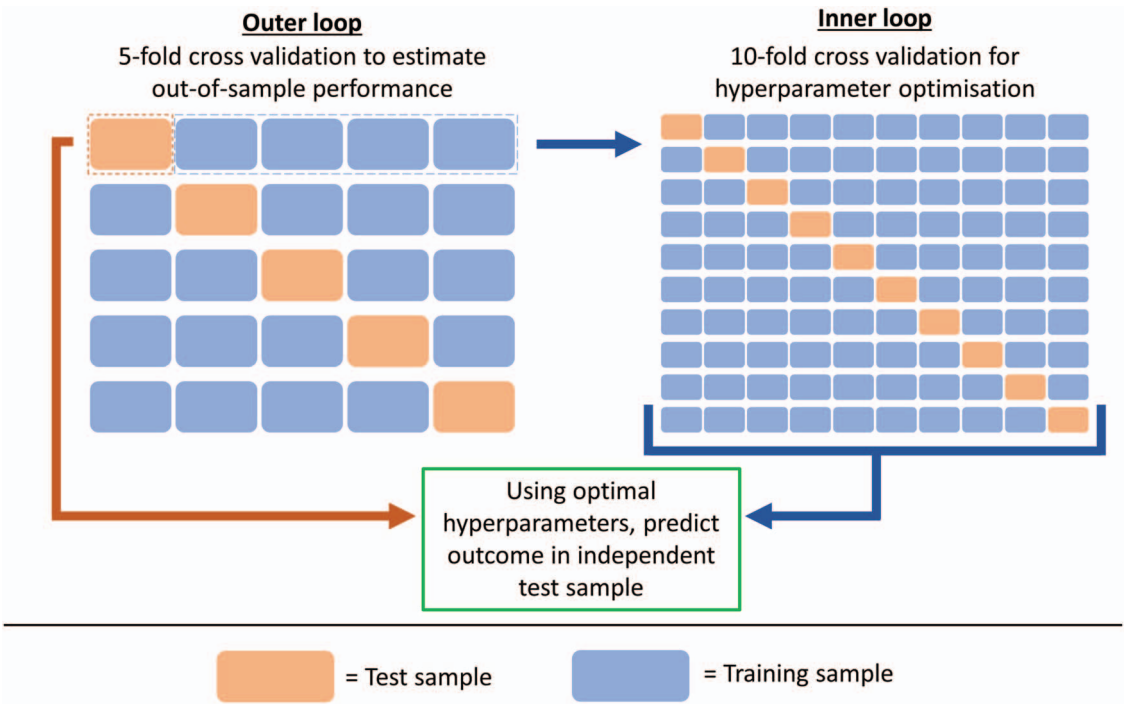

Figure 6. Schematic representation of nested cross-validation procedure. The outer loop splits the sample into five parts, four parts are used as a training sample for hyperparameter optimization, and the resulting model is then used to predict the outcome in the remaining part (test sample). This process is repeated until predictions are available for all parts of the sample. Hyperparameter optimization is carried out within the inner loop, which consists of 10-fold cross validation.

polygenic scoring, and the European subset of the 1KG Phase 3 reference was used to estimate LD.

As a sensitivity analysis, pT + clump PRS were also calculated using only variants within 500 kb of genes used in the TWAS, thereby restricting the PRS to the same variants within the gene expression SNP-weights and highlighting the effect of reweighting genetic variants by their effect on gene expression.

Furthermore, pT + clump PRS for rheumatoid arthritis were also calculated without restricting to a single variant in the MHC region to gain insight into difference between PRS and GeRS prediction for this outcome.

Evaluating predictive utility of GeRS

Prediction accuracy was evaluated as the Pearson's correlation between the observed and predicted phenotype outcomes. Correlation was used as the main test statistic as it is applicable for both binary and continuous outcomes and standard errors are easily computed. Correlations can be easily converted to other test statistics such as R^2 (observed or liability) and area under the curve (equations 8 and 11 in (31)), with relative performance of each method remaining unchanged.

Logistic regression was used for predicting binary outcomes, and linear regression was used for predicting continuous outcomes. If the model contained only one predictor, a generalized linear model was used. If the model contained more than one predictor (e.g. GeRS for each *P*-value threshold), an elastic net model was applied to avoid overfitting due to the inclusion of multiple correlated predictors (32).

Elastic net modelling. Previous research has shown that modelling multiple PRS based on a range of parameters (*P*-value thresholds or shrinkage parameters) optimizes prediction out-of-sample (3). Therefore, elastic net models were derived using multiple pT + clump PRS based on a range of *P*-value thresholds, or multiple PRScs scores based on a range of global shrinkage parameters. Furthermore, elastic net models were derived for GeRSs based on a range of *P*-value thresholds and SNP-weight sets to evaluate the effect of modelling multiple GeRS simultaneously. For example, if 8 *P*-value thresholds were used, and 2 SNP-weight sets were used, the elastic net model would contain 16 GeRS.

A nested cross validation procedure (33) was used to estimate the predictive utility of elastic net models (Fig. 6), where hyperparameter selection is performed using inner 10-fold cross validation, while an outer 5-fold cross validation computes an unbiased estimate of the predictive utility of the model with the inner cross validation based hyperparameter tuning. This approach avoids overfitting whilst providing modelling predictions for the full sample. The inner 10-fold cross validation for hyperparameter optimization was carried out using the 'caret' R package.

The correlation between observed and predicted values of each model were compared using William's test (also known as the Hotelling–Williams test) (34) as implemented by the 'psych' R package's 'paired.r' function, with the correlation between model predictions of each method specified to account for their non-independence. A two-sided test was used when calculating *P*-values.

Estimating variance explained by cis-heritable expression. A schematic representation of this analysis is in Figure S14. To estimate the proportion of SNP-based heritability explained by *cis*-regulated expression, we used the R package AVENGEME (35) to estimate SNP-based heritability of each phenotype in

the target sample based on pT + clump PRS associations across *P*-value thresholds, and the phenotypic variance explained by *cis*-regulated expression (GE-based heritability) based on the GeRS associations across *P*-value thresholds. To estimate the association with GeRS at each *P*-value threshold, we used predictions from elastic net models containing GeRS across all SNP-weight sets for a given *P*-value threshold. The proportion of SNP-based heritability explained by *cis*-heritable expression was then calculated as GE-based heritability divided by the SNP-based heritability. AVENGEME also estimates the fraction of non-causal variants. AVENGEME has been previously used to estimate the proportion of SNP-based heritability attributable to *cis*-regulated gene expression based on GeRS associations, acknowledging that the estimate will be inflated due to LD causing gene expression SNP-weights to tag other causal mechanisms, such as variants affecting protein structure and function (13). As a sensitivity analysis, we estimated the GE-based heritability using GeRS restricted to genes with colocalization PP4 > 0.8 to remove genes which do not colocalize. For the GeRS analysis, the 'nsnp' variable in AVENGEME, indicating the number of independent markers in the score was set to the number of LD independent markers in the TWAS gene stratified PRS.

Supplementary Material

Supplementary Material is available at *HMG* online.

URLS

- LDSC HapMap 3 SNP-list: https://data.broadinstitute.org/alkesgroup/LDSCORE/w_hm3.snplist.bz2
- LDSC Munge Sumstats: https://github.com/bulik/ldsc/blob/master/munge_sumstats.py
- Impute.me: https://impute.me/
- GenoPred website: https://opain.github.io/GenoPred

Acknowledgements

We thank Paul O'Reilly, Sam Choi and Alexander Gusev for useful discussion.

Conflict of Interest statement. Cathryn Lewis sits on the Myriad Neuroscience Scientific Advisory Board. The other authors declare no competing interests.

Funding

UK Medical Research Council (MR/N015746/1 and MR/S0151132); National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. A.F. is funded by the National Institute of Health grant R01AG054628. The authors acknowledge use of the research computing facility at King's College London, Rosalind (https://rosalind.kcl.ac.uk), which is delivered in partnership with the NIHR Maudsley BRC, and part-funded by capital equipment grants from the Maudsley Charity (award 980) and Guy's & St. Thomas' Charity (TR130505). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. We thank the research participants and employees of 23andMe for making the work regarding Depression possible.

UKB: This research was conducted under UK Biobank application 18177.

TEDS: We gratefully acknowledge the ongoing contribution of the participants in TEDS and their families. TEDS is supported by UK Medical Research Council Program Grant MR/M021475/1 (and previously Grant G0901245) (to Robert Plomin (R.P.)), with additional support from National Institutes of Health Grant AG046938. The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC Grant Agreement 295366. R.P. is supported by Medical Research Council Research Professor Award G19/2. K.R. is supported by a Sir Henry Wellcome Postdoctoral Fellowship.

References

1. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A. and Ellinor, P.T. (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.*, **50**, 1219–1224.
2. Choi, S.W., Mak, T.S.-H. and O'Reilly, P.F. (2020) Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.*, **15**, 2759–2772.
3. Pain, O., Glanville, K.P., Hagenaars, S.P., Selzam, S.P., Fürtjes, A.E., Gaspar, H.A., Coleman, J.R.I., Rimfeld, K., Breen, G., Plomin, R. *et al.* (2020) Evaluation of polygenic prediction methodology within a reference-standardized framework. *bioRxiv*. 2020.07.28.224782. doi: https://doi.org/10.1101/2020.07.28.224782.
4. Vilhjálmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G. and Do, R. (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.*, **97**, 576–592.
5. Hu, Y., Lu, Q., Powles, R., Yao, X., Yang, C., Fang, F., Xu, X. and Zhao, H. (2017) Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput. Biol.*, **13**, e1005589.
6. Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M. and Aguet, F. (2018) Using an atlas of gene regulation across 44 human tissues to inform complex disease-and trait-associated variation. *Nat. Genet.*, **50**, 956–967.
7. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E. and Cox, N.J. (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.*, **6(4)**, e1000888.
8. Consortium, Gte (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204.
9. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A. and Shah, H.R. (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.*, **19**, 1442.
10. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., De Geus, E.J.C., Boomsma, D.I. and Wright, F.A. (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, **48**, 245.
11. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T. and Edwards, T.L. (2018) Exploring the phenotypic consequences of tissue specific gene expression variation

inferred from GWAS summary statistics. *Nat. Commun.*, **9**, 1825.
12. Pain, O., Pocklington, A.J., Holmans, P.A., Bray, N.J., O'Brien, H.E., Hall, L.S., Pardiñas, A.F., O'Donovan, M.C., Owen, M.J. and Anney, R. (2019) Novel insight into the aetiology of autism Spectrum disorder gained by integrating expression data with genome-wide association statistics. *Biol. Psychiatry*, **86**, 265–273.
13. Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H.K., Reshef, Y., Song, L., Safi, A., McCarroll, S. and Neale, B.M. (2018) Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.*, **50**, 538.
14. Shi, J., Park, J.-H., Duan, J., Berndt, S.T., Moy, W., Yu, K., Song, L., Wheeler, W., Hua, X. and Silverman, D. (2016) Winner's curse correction and variable thresholding improve performance of polygenic risk modeling based on genome-wide association study summary-level data. *PLoS Genet.*, **12**, e1006493.
15. Zhou, X., Im, H.K. and Lee, S.H. (2020) CORE GREML for estimating covariance between random effects in linear mixed models for complex trait analyses. *Nat. Commun.*, **11**, 1–11.
16. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O. and O'Connell, J. (2018) The UK biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.
17. Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., Dale, P.S., Eley, T.C. and Plomin, R. (2019) Twins early development study: a genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood. *Twin Res. Hum. Genet.*, **22**, 508–513.
18. Busch, R., Kollnberger, S. and Mellins, E.D. (2019) HLA associations in inflammatory arthritis: emerging mechanisms and clinical implications. *Nat. Rev. Rheumatol.*, **15**, 364–381.
19. Raj, P., Rai, E., Song, R., Khan, S., Wakeland, B.E., Viswanathan, K., Arana, C., Liang, C., Zhang, B. and Dozmorov, I. (2016) Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. *elife*, **5**, e12089.
20. Grinberg, N.F. and Wallace, C. (2020) Multi-tissue transcriptome-wide association studies. *bioRxiv*. 1–14.
21. Choi, S.W. and O'Reilly, P.F. (2019) PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience*, **8**, giz082.
22. Yao, D.W., O'Connor, L.J., Price, A.L. and Gusev, A. (2020) Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.*, **52**, 626–633.
23. Marigorta, U.M., Denson, L.A., Hyams, J.S., Mondal, K., Prince, J., Walters, T.D., Griffiths, A., Noe, J.D., Crandall, W.V. and Rosh, J.R. (2017) Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat. Genet.*, **49**, 1517–1521.
24. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P. and Sharp, K. (2016) A reference panel of 64, 976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
25. consortium, U (2015) The UK10K project identifies rare variants in health and disease. *Nature*, **526**, 82–90.
26. Selzam, S., McAdams, T.A., Coleman, J.R.I., Carnell, S., O'Reilly, P.F., Plomin, R. and Llewellyn, C.H. (2018) Evidence for gene-environment correlation in child feeding: links between common genetic variation for BMI in children and parental feeding practices. *PLoS Genet.*, **14(11)**, e1007757.
27. Selzam, S., Ritchie, S.J., Pingault, J.-B., Reynolds, C.A., O'Reilly, P.F. and Plomin, R. (2019) Comparing within-and between-family polygenic score prediction. *Am. J. Hum. Genet.*, **105**, 351–363.
28. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W. and Zhou, Y.-H. (2014) Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.*, **46**, 430–437.
29. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
30. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. and Smoller, J.W. (2019) Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.*, **10**, 1–10.
31. Lee, S.H., Goddard, M.E., Wray, N.R. and Visscher, P.M. (2012) A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.*, **36**, 214–224.
32. Zou, H. and Hastie, T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (statistical Methodol.)*, **67**, 301–320.
33. Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B*, **36**, 111–133.
34. Steiger, J.H. (1980) Tests for comparing elements of a correlation matrix. *Psychol. Bull.*, **87**, 245.
35. Palla, L. and Dudbridge, F. (2015) A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.*, **97**, 250–259.