Integrative analysis of Dupuytren's disease identifies novel risk locus and reveals a shared genetic etiology with BMI

Megan Major,[1] Malika K Freund,[2] Kathryn S Burch,[1] Nicholas Mancuso,[3] Michael Ng,[4] Dominic Furniss,[4,5,6] Bogdan Pasaniuc,[1,2,3,7,9] Roel Ophoff[2,8,9]*

[1] Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, California, 90095, USA
[2] Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA
[3] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA
[4] Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Science, University of Oxford, Oxford, OX3 7HE, UK
[5] Department of Plastic and Reconstructive Surgery, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, OX3 9DU, UK
[6] NIHR Biomedical Research Centre, NDORMS, University of Oxford, Oxford, OX3 7HE, UK
[7] Department of Biomathematics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA
[8] Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, California, 90095, USA

[9] These authors jointly supervised this work.

Correspondence: * ROphoff@mednet.ucla.edu

Abstract

Dupuytren's disease is the common inherited tissue-specific fibrotic disorder. It's characterized by progressive and irreversible fibroblastic proliferation affecting the palmar fascia of the hand, with an onset typically in the sixth decade of life. Although genome-wide association studies (GWAS) have identified 24 genomic regions associated with Dupuytren's risk, the biological mechanisms driving signal at these regions remain elusive. We identify potential biological mechanisms for Dupuytren's disease by integrating the most recent, largest GWAS ($n_{cases} = 3{,}871$, $n_{controls} = 4{,}686$) with eQTLs (47 tissue panels from five consortia, total $n = 3{,}975$) to perform a transcriptome-wide association study (TWAS). We identify 43 tissue-specific gene associations with Dupuytren's risk, one of which resides at least 0.5 Mb away from the 24 risk regions previously identified. We also estimate the genome-wide genetic correlation between Dupuytren's disease and 45 complex traits and find significant genetic correlations between Dupuytren's disease and body mass index ($\hat{r}_g = -0.20$, $P = 1.6 \times 10^{-6}$), type II diabetes ($\hat{r}_g = -0.18$, $P = 1.7 \times 10^{-4}$), triglycerides ($\hat{r}_g = -0.14$, $P = 3.5 \times 10^{-4}$), and high-density lipoprotein ($\hat{r}_g = 0.13$, $P = 4.1 \times 10^{-4}$), which suggests a shared genetic etiology. We further refine the genome-wide genetic correlation signal to identify 8 regions significantly negatively correlated with BMI and 3 regions significantly correlated (1 positively and 2 negatively correlated) with HDL; none of these regions contained the novel gene association identified by TWAS. Our results are consistent with previous epidemiological findings which show that lower BMI increases risk for Dupuytren's disease. These 12 novel risk regions provide new insight into the biological mechanisms of Dupuytren's disease and serve as a starting point for functional validation.

Introduction

Dupuytren's disease (DD [MIM: 126900]) is a common and disabling connective tissue disorder affecting 5-25% of individuals of European ancestry, characterized by progressive and irreversible fibroblastic proliferation affecting the palmar fascia of the hand[1,2]. DD initially manifests as nodules in the palm of the hand, resulting in contraction and ultimately flexion contractures of the digits in a proportion of individuals affected with DD. Recent twin studies estimate the heritability (i.e., proportion of phenotypic variation explained by genetics) of DD to be ~80%[3]. The largest previous genome-wide association study (GWAS) of DD in individuals of European ancestry identified 26 genome-wide significant single-nucleotide polymorphism (SNP) associations in 24 independent risk regions[4], and estimated the proportion of phenotypic variance attributable to additive effects of common variants (i.e., SNP-heritability) to be 0.53[4,5]. The vast majority (23 of 24) of DD associations lie in non-coding genomic regions with only one located in an intron[4], thus the biological implications of these associations are not immediately clear. Investigations into the mechanisms behind the strongest GWAS association, rs16879765 ($P_{GWAS} = 7.2 \times 10^{-41}$), located in the intron of *EPDR1*, revealed an effect on expression and protein secretion of the nearby gene *SFRP4*[4] but implicated *EPDR1* functionally[6]. Overall, the regulatory mechanisms driving signal at the GWAS associations on DD remains unknown.

In this study, we aimed to explore genetic mechanisms at known risk regions for DD, identify complex traits with possible shared genetic etiologies, and find novel risk regions for DD. Recently, transcriptome-wide association studies[7,8] (TWAS) have emerged as a way to identify associations between gene expression and a trait. We performed a multi-tissue TWAS by combining a recent DD GWAS[4] with expression quantitative trait loci[9–14] (eQTL), integrating gene expression from five consortia in 43 unique tissues, to test for association between

expression and DD in 15,198 genes. We identified 43 associations between tissue-specific gene expression and DD, including one novel risk region on chromosome 17. Next, we aimed to understand the genetic relationship between DD and 45 other complex phenotypes by identifying traits that have genetic correlation (i.e., the similarity in genetic effects across two traits) providing etiological insights and plausible causal relationships to investigate[15–18]. We performed genetic correlation analyses through cross-trait linkage disequilibrium (LD) score regression[15] (LDSC), and find that body mass index (BMI), type II diabetes (T2D), triglycerides (TG), and high-density lipoprotein (HDL) levels are significantly genetically correlated with DD. Additionally, we sought to further refine and understand these relationships with DD and BMI, T2D, TG, and HDL by exploring local regions with enrichments of genetic correlation using ρ-HESS[16], and found 8 risk regions significantly correlated with BMI and 3 risk regions significantly correlated with HDL. Finally, we aimed to identify a tissue or cell type to prioritize when studying DD.

Materials and Methods

DD GWAS summary statistics

Results from a GWAS of DD in UK Europeans (3,871 cases and 4,686 controls) were previously reported[4]. This GWAS summary data contained association statistics for 7,218,238 SNPs, with 6,991,033 SNPs that were imputed from individuals of European ancestry in the Haplotype Reference Consortium[19]. We excluded multi-allelic SNPs, SNPs with ambiguous alleles (e.g., A to T or C to G), and SNPs without an rsID defined by dbSNP144, resulting in 6,126,071 SNPs for downstream analyses.

We used PLINK[20] to compute independent risk regions (at least one SNP with $P_{GWAS} \leq 5 \times 10^{-8}$) in the DD GWAS data by clumping SNPs into regions based on LD and distance, using R^2 thresholds of 0.3 and 0.25 for between-block LD and within-block LD, respectively. This resulted in 24 independent risk regions.

TWAS reference panels and details

To find novel risk genes and biologically meaningful associations, we performed a TWAS to test genes expression levels for association with DD. We used FUSION[7] software (see Web Resources) along with prepackaged gene expression weights. Briefly, TWAS identifies candidate risk genes for DD by integrating results from GWAS and reference panels of gene expression measurements from eQTL studies to associate cis-regulated expression with DD, while accounting for LD. Weights for gene expression were from the Genotype-Tissue Expression Project[9] v6 (GTEx; 43 tissues, $n = 449$), the Metabolic Syndrome in Men study[10] (METSIM; adipose, $n = 563$), the Young Finns Study[11,12] (YFS; blood, $n = 1{,}264$), the CommonMind Consortium[13] (CMC; dorsolateral prefrontal cortex, $n = 452$), and the Netherlands Twin Registry[14] (NTR; blood, $n = 1{,}247$) reference panels. This totaled to 47 different reference tissue panels that represent 43 unique tissues (see Supplementary Table 1). Description of quality control procedures for these expression data have been previously described[7,21].

For each reference gene expression panel, FUSION estimates the strength of association between predicted expression of a gene and DD (Z_{TWAS}) as a function of the vector of GWAS summary Z-scores at a given cis-region (Z_{GWAS}) and the LD-adjusted weights vector learned from the gene expression data. A p-value (P_{TWAS}) is obtained using a two-tailed test under $N(0,1)$. This process was repeated for each reference tissue panel and gene, resulting in 98,147

tissue-specific gene models involving 15,189 genes (see Supplementary Table 1). We assessed significance with the family-wise error rate threshold at $P_{TWAS} \leq \frac{0.05}{98,147}$.

Genome-wide genetic correlation with cross-trait LDSC

We estimate genome-wide genetic correlation between DD and 45 complex traits to identify shared genetic risk for DD with other complex traits. To this end, we used cross-trait LDSC[15], a method for estimating genome-wide genetic correlation between two traits that requires only GWAS summary statistics and reference panel LD (European ancestry from the 1000 Genomes (1000G) Project[22]). We defined the genetic correlation ($\hat{r}_g$) between DD and another trait as significant if it passed the Bonferroni-corrected threshold of $P_{T1,T2} \leq \frac{0.05}{45}$.

Local genetic correlation and putative causality using ρ-HESS

For each of the traits with significant genome-wide genetic correlation with DD, we run ρ-HESS[16] to estimate the local genetic correlation between each trait and DD within 1702 approximately independent regions genome-wide[23]. For reference LD, we used European ancestry from the 1000G Project[22]. The local genetic correlation ($\hat{r}_{g,local}$) between two traits at a given region was defined as significant if it passed the threshold of $P_{region} \leq \frac{0.05}{1702}$.

We also aimed to find evidence for putative causal relationships between DD and other genetically correlated traits. We used the implementation in ρ-HESS based on a previously described method[18] to prioritize putative causal models between pairs of complex traits. Essentially, for two complex traits, the local genetic correlation is evaluated at regions harboring genome-wide significant GWAS signals specific to each trait. The local genetic correlation for all trait 1 specific regions are summed ($\hat{r}_{T1,regions}$) and the local genetic correlation for all trait 2

specific regions are summed ($\hat{r}_{T2,regions}$). Confidence intervals are determined by 1.96 times jackknife standard error on each side; significance is determined if the confidence intervals do not overlap. The intuition behind this test is that if trait 1 causally influences trait 2 then trait 1 specific regions would have strong genetic correlation with trait 2 but trait 2 specific regions would not have strong genetic correlation with trait 1. Thus, we can leverage the difference in correlations for trait-specific signal at these regions to see if the correlations are consistent with a suggestive causal model[16,18].

Tissue and cell type prioritization

To identify tissues and/or cell types that are biologically relevant to DD, we used stratified LD score regression to estimate the enrichment of DD SNP-heritability in 205 publicly available specifically expressed gene (SEG) annotations, each of which represents a set of genes that are specifically expressed in a single tissue or cell type (LDSC-SEG)[24]. Briefly, the 205 annotations were originally created from two datasets: RNA-seq gene expression measurements in 53 human tissues from GTEx v6p[9] (average of 161 samples per tissue), and a microarray gene expression dataset comprised of 152 tissues and cell types from either human, mouse, or rat (the "Franke Lab" dataset)[25,26]. For each set of specifically expressed genes, an annotation was created by adding 100-kb windows upstream and downstream from the transcribed region of each gene. In addition, we tested for enrichment of DD SNP-heritability in a set of 489 publicly available tissue- or cell type-specific chromatin annotations[24]. 396 of these annotations were originally created from five activating histone marks (H3K27ac, H3K4me3, H3K4me1, H3K9ac, and H3K36me3) and DNase I hypersensitivity (DHS) regions that were present in a subset of 88 tissues and cell types in the Roadmap Epigenomics Consortium[27]. An additional 93 annotations were created from a set of four activating histone marks (H3K27ac, H3K4me3, H3K4me1, and

H3K36me3) in 27 tissues from EN-TEx[28] that were also present in GTEx. Details on the construction of both the SEG annotations and chromatin-based annotations can be found in the original study[24]. Each annotation was tested individually for enrichment of DD SNP-heritability on top of the baseline-LD model[29] by assessing whether the expected additional per-SNP heritability contribution due to the annotation is significantly nonzero (FDR < 0.1).

We also employed the web application FUMA[30] in the aim of finding tissues or cell types with differentially expressed genes relevant to DD. FUMA maps GWAS results to create a gene set in three ways: (1) physical proximity on the genome, (2) eQTL associations, and (3) chromatin interaction. We used the gene property analyses (implemented from MAGMA[31]) and differentially expressed gene (DEG) analysis to prioritize different tissues or cell types. For the gene property analysis, FUMA tests if expression of the GWAS gene set in a single tissue or cell type is statistically different than the average expression of the GWAS gene set across all tissues or cell types. We perform this gene property analysis in 53 GTEx[9] tissues ($P_{GP,T} \leq \frac{0.05}{53}$) as well as in 5115 study-defined cell types ($P_{GP,CT} \leq \frac{0.05}{5115}$) using single cell RNA-seq data from 28 studies[32–58] as described on the FUMA website (see Web Resources). For the DEG analysis, FUMA defines differentially expressed genes in each tissue by performing a two-sided t-test for that one tissue against all other tissues. Each of the 53 GTEx[9] tissues is tested for up-regulation, down-regulation, and both-sided DEG sets. We removed tissues where DEG sets had less than 30 genes to avoid underpowered correlations; significance was defined by Bonferroni correction for the number of tests ($P_{DEG} \leq \frac{0.05}{153}$).

Results

TWAS identifies 18 risk genes for DD

To explore putative biological mechanisms at known DD risk regions, we performed a multi-tissue TWAS to identify genes (specifically, cis-regulated gene expression), associated with DD (see Materials and Methods). Briefly, TWAS identifies candidate risk genes for DD by integrating results from GWAS and reference panels of gene expression measurements from eQTL studies to associate cis-regulated expression with DD, while accounting for LD. We used tissue reference panels from GTEx[9], METSIM[10], YFS[11,12], CMC[13], and NTR[14] resulting in 47 different reference tissue panels with a combined sample size of 3,975 (see Materials and Methods; Supplementary Table 1). Using these reference panels, we tested 98,147 tissue-specific gene models and found 43 significant tissue-specific gene-trait associations at a Bonferroni-corrected threshold of $P_{TWAS} \leq \frac{0.05}{98{,}147}$ (Table 1, Supplementary Table 2). GWAS SNP association strength and TWAS tissue-specific gene model association strength can be seen in Figure 1. These 43 significant models were composed of 18 genes among 23 tissue panels–7 genes were significant in multiple tissues (Table 1). 36 of the 43 significant tissue-specific gene models were within 0.5Mb of any of the previously identified 24 risk regions.

One region of interest is on chromosome 7, where *EPDR1* was found to be significant in 10 different tissue panels (most significant in lung tissue, $P_{TWAS} = 6.4 \times 10^{-31}$). This region has been previously investigated because of its strong association signal (Odds Ratio 1.93 and $P_{GWAS} = 7.2 \times 10^{-41}$) with DD[4]. The variant with the strongest association in this region, rs16879765, is in an intron of *EPDR1*. Although decreased secretion of the nearby WNT-agonist *SFRP4* was correlated with the high risk genotype[4], genetic and functional evidence point toward

EPDR1 being the disease-relevant gene for this region, which has been functionally validated as contributing to myofibroblast contractility[6]. All three transcripts of *EPDR1* are found in affected DD tissue and knockdown of *EPDR1* attenuates contractility in fibroblast-populated collagen lattice assays[6].

TWAS identifies novel risk region on chromosome 17

To identify possible novel risk regions from TWAS associations, we aimed to see if any tissue-specific gene models were independent of established GWAS associations. After grouping association signal into 1Mb regions, we found 13 regions with only significant GWAS SNP(s), 1 region with only significant TWAS model(s), and 11 regions with both significant GWAS SNP(s) and TWAS model(s). Here we define a region identified through TWAS to be novel if (1) the strongest DD associated SNP in the gene's region is not genome-wide significant (i.e., $P_{GWAS} \geq 5 \times 10^{-8}$) and (2) that the TSS of the TWAS-gene is not within 0.5Mb of the previously known 24 risk regions. With these constraints, we identified one novel risk region for DD (Figure 2). To ensure our result was robust to long-range LD, we expanded our window criteria to include 1Mb and 2Mb and found no change. We found a single tissue-specific gene model, *TMEM106A* ($P_{TWAS} = 1.2 \times 10^{-7}$; GTEx breast mammary tissue), was significantly associated with DD risk at this region (Figure 2). To determine that the *TMEM106A* association was robust to possible LD confounding, we performed a permutation test using GWAS summary statistics and found similar results ($P_{perm} = 8.91 \times 10^{-3}$). There were 12 tissue panels that expression for *TMEM106A* was modeled from (Supplementary Table 3).

Estimates of SNP-heritability in DD are higher than previously proposed

We obtained a SNP-heritability estimate of 0.67 (s.e. = 0.08), using LDSC[59]. We also used Heritability Estimator from Summary Statistics (HESS), a previously described method using similar framework as ρ-HESS that estimates local SNP-heritability, and found the total SNP-heritability to be 0.532 (s.e. = 0.282), similar to the previous estimate of 0.533 using GCTA[4,5]. Because HESS is optimized for GWAS with sample sizes greater than 50,000 (contributing the large standard error), we included the LDSC regression estimate of SNP-heritability when running HESS to obtain more stable estimates of local SNP-heritability[59,60].

Genetic correlation suggests shared genetic etiology with DD

To identify traits that have a shared genetic etiology with DD, we used cross-trait LDSC[15] which estimates the genetic correlation between two traits using GWAS summary statistics (see Materials and Methods). Results for the genetic correlation test between DD and 45 other traits (average sample size 132,115) can be found in Table 2. These 45 traits include a variety of anthropometric, immune, hematological, neurological, and cardiovascular related traits and disorders. Four traits were found to have significant genetic correlation with DD ($P_{T1,T2} \leq \frac{0.05}{45}$, identical results correcting for a FDR < 0.1): body mass index (BMI), $\hat{r}_g = -0.196$; high density lipoprotein (HDL), $\hat{r}_g = 0.133$; triglycerides (TG), $\hat{r}_g = -0.139$; and type II diabetes (T2D), $\hat{r}_g = -0.182$ (Table 2). These results are compatible with previous observational studies[2,61,62]. Notably, the negative genetic correlation between BMI and DD is consistent with a previous epidemiological investigation showing that the risk of DD was inversely proportional to BMI, after correcting for age, race, and sex in 14,844 patients diagnosed with DD[63].

Local genetic correlation analysis yields 11 regions to further study

Having found evidence for BMI, HDL, TG, and T2D sharing genetic factors with DD at a genome-wide scale, we next aimed to locate possible shared genomic regions. We did this by running ρ-HESS[16] to estimate the local genetic correlation between DD and each of the four traits in 1702 approximately independent LD blocks[23] (see Materials and Methods). The genome-wide genetic correlation results from cross-trait LDSC and ρ-HESS are fairly consistent (Pearson's r = 0.94; Supplementary Table 4); differences may have resulted from using metabochip array[64] GWAS statistics with LDSC (HDL and TG), which is discouraged for cross-trait LDSC[15], as well as smaller sample size in the DD GWAS adding noise to estimates from ρ-HESS[16,60]. We found eight regions significantly genetically correlated ($P_{region} \leq \frac{0.05}{1702}$) between DD and BMI, three regions between DD and HDL, and no regions between DD and TG or between DD and T2D (Table 3). Of these 11 regions, three contained a genome-wide significant association in the DD GWAS[4]. Only one of the 11 regions contained significant tissue-specific gene models from TWAS; the 10 models for *EPDR1* were within the DD and HDL genetically correlated 7:37555184-38966703 region (Supplemental Table 2, Table 3).

Genetic correlation patterns of BMI/TG and DD consistent with putative causality

To further elucidate the relationships of these traits with DD, we used ρ-HESS[16] to test for evidence of putative causality through GWAS estimated genetic effects for BMI, HDL, TG, and T2D acting on DD or vice versa (see Materials and Methods). Both BMI and TG showed suggestive patterns that would be consistent with a putative causal relationship with DD, while HDL and T2D did not (Figure 3, Supplemental Figure 1). For example, when considering BMI and DD, the correlation at 399 BMI-specific regions (-0.27, s.e. = 0.044) is seemingly stronger

than the correlation at 19 DD-specific regions (-0.03, s.e. = 0.15), indicating that regions that increase BMI tend to decrease risk of DD; this is consistent with a model where BMI genetic effects decrease risk of DD (Figure 3). The same is true for TG and DD; the correlation at 65 TG-specific regions (-0.3, s.e. = 0.08) is seemingly stronger than the correlation at 22 DD-specific regions (0.018, s.e. = 0.2; Figure 3). Both of these results are not significant (assessed by overlap of confidence intervals, $\hat{r}_{TX,regions} \pm 1.96 \times$s.e.); this is most likely because of the relatively reduced sample size in the DD GWAS study. Nonetheless, there is evidence of a putative causal relationship with BMI affecting TG[16,18] and results from TG and DD may be from a mediated causal relationship of BMI affecting TG, which in turn affects DD (Figure 3).

DD most relevant tissue or cell type unidentifiable with current data

Finally, we aimed to identify relevant tissues or cell types for DD. First, we used S-LDSC[65] to estimate the enrichment of SNP-heritability of DD (controlling for the baseline-LD model[29]) in two sets of publicly available annotations[24]: annotations representing specifically expressed genes (SEG) in 205 tissues or cell types[9,25,26] and 489 annotations representing 6 chromatin features (DHS and 5 histone marks) in 91 tissues or cell types[27,28]. Among the 205 SEG annotations, synovial membrane tissue was the most enriched for DD SNP-heritability on top of the baseline-LD model, but none of the 205 annotations were statistically significant (FDR < 0.1; Supplementary Table 5). Among the 489 chromatin annotations, we found that esophageal-mucosa tissue was the most enriched for DD SNP-heritability, however none of 489 annotations were statistically significant (FDR < 0.1; Supplementary Table 6). Next, we prioritized tissues and cell types using FUMA[30], a platform to visualize and interpret GWAS summary statistics (see Materials and Methods). After using FUMA to create a gene set from the GWAS statistics, we first performed a gene property analysis (which tests if gene expression in a single tissue or

cell type is statistically different than the average gene expression across all tissues or cell types) in 53 tissue types[9]. Although none of the 53 tissues showed a significant effect ($P_{GP,T} \leq \frac{0.05}{53}$), the effect was strongest in cell transformed fibroblasts (Supplementary Table 7). We then assessed whether the GWAS gene set was enriched in any of the differentially expressed gene (DEG) sets for tissues. The up-regulated DEG sets for tibial artery and aorta tissues both demonstrated significant ($P_{DEG} = 5.5 \times 10^{-5}$ and 7.8×10^{-5}, respectively) overlap with the GWAS gene set (Supplementary Table 8). We also performed a gene property analysis using cell type specific expression data for 5115 study-defined cell types from 28 scRNA-seq studies[32–58]. While none of the single cell types were significant ($P_{GP,CT} \leq \frac{0.05}{5115}$), stromal cells and muscle cells were among the top five results (Supplementary Table 9). As a final analysis, we averaged the χ^2-statistic ($Z_{TWAS}{}^2$) for the 43 significant TWAS models within each tissue to determine which tissue had the most enrichment of TWAS signal. We found adipose subcutaneous tissue was most enriched among the 23 tissues with significant TWAS models (Supplementary Figure 2). Because of the lack of consistency between methods and lack of statistical significance in many methods, we are lead to believe that likely the relevant tissue or cell type is not represented in current datasets.

Discussion

In this work, we aimed to better understand the genetic architecture of DD, find plausible biological mechanisms at known risk regions for DD, understand the relationship between DD and a variety of other traits, and identify possible novel risk regions through local genetic correlation with other traits or genetic-mediated gene expression effects. We highlight that the estimated SNP-heritability of DD (0.53-0.67) is relatively close to estimates of heritability from twin studies (0.8). We also note that the strong concentration of DD GWAS signal in a handful

of genomic regions is more consistent with an oligogenic architecture than a polygenic one, suggesting that further functional studies could be particularly fruitful as compared to more polygenic traits and diseases. We also identify a negative genetic correlation between DD and BMI, supporting a previous epidemiological study that observationally showed a negative correlation between the traits[63]; understanding the relationship between DD and BMI as well as that between DD and TG could shed light on shared biologically important pathways. Finally, we identify one novel risk region from TWAS, and identify 11 regions with significant local genetic correlation between DD and BMI or HDL. Overall, our findings highlight the need for more investigation into these regions as a first step.

Additionally, we note a few caveats in our results. First, though the sample size of 8,557 for the DD GWAS is the largest yet, it is possible that additional GWAS regions remain undiscovered due to the limit in power and this also would further reduce power to fully detect associations and relationships with other traits. Second, while the patterns of genetic correlation between BMI and DD as well as TG and DD are somewhat consistent with causal relationships, true causality between these traits cannot be determined without functional experimentation. Third, we emphasize that TWAS may not detect the true mechanism of disease if the gene expression is not mediated through genetics or if disease-relevant tissue is not well-represented in available gene expression reference panels. This may be further illustrated by the fact we were unable to identify a specific tissue or cell type to prioritize for further study in DD. This could also be due to the small sample size of the DD GWAS, the cell-type specificity of enhancer elements, or again the publication bias away from musculoskeletal connective tissues, leading to a gap in the available datasets.

Future work should be taken in multiple directions. First, we provide additional evidence that *EPDR1* may contribute to the pathogenesis of DD; further work should be dedicated to functionally validate and understand this gene in connection with DD, as it may represent an attractive therapeutic target. Second, there is strong evidence for a relationship between BMI/TG and DD–elucidating the mechanism may lead to interesting observations with implications for the treatment of both traits. Third, additional GWAS, with larger sample sizes and in additional populations, will uncover more of the contribution of genetic variation to DD. And fourth, given the putative oligogenic architecture of DD, and that our tissue and cell type analyses lacked consistent results, it might also be rewarding to generate more functional -omics data, such as reference gene expression panels or chromatin accessibility data in the palmar fascia tissue. These resources would offer valuable insight into the underlying mechanisms of DD and opportunity to explore therapeutic avenues.

Supplemental Data

Supplemental Data includes two figures and nine tables.

Acknowledgements

This work was supported by National Institutes of Health (NIH) awards R01HG009120, R01MH115676, R01HG006399, and U01CA194393. M.K.F. is funded by NIH-NCI National Cancer Institute T32LM012424; K.S.B. is funded by NIH Training Grant in Genomic Analysis and Interpretation T32HG002536; D.F. is funded by an Intermediate Fellowship from the Wellcome Trust (097152/Z/11/Z). This work was also supported by the NIHR Biomedical Research Centre, Oxford.

Declaration of Interests

The authors declare no competing interests.

Web Resources

OMIM, https://www.omim.org/

1000 Genomes, http://www.internationalgenome.org/
PLINK, https://www.cog-genomics.org/plink2/
FUSION, http://gusevlab.org/projects/fusion/
GTEx Portal, https://gtexportal.org/home/
LD scores and annotations, https://data.broadinstitute.org/alkesgroup/LDSCORE/
DEPICT, https://data.broadinstitute.org/mpg/depict/depict_download/tissue_expression/
FUMA, http://fuma.ctglab.nl/

References

1. Dupuytren, B. (1834). CLINICAL LECTURES ON SURGERY. Lancet *22*, 222–225.
2. Gudmundsson, K.G., Arngrímsson, R., Sigfússon, N., Björnsson, Á., and Jónsson, T. (2000). Epidemiology of Dupuytren's disease. J. Clin. Epidemiol. *53*, 291–296.
3. Larsen, S., Krogsgaard, D.G., Aagaard Larsen, L., Iachina, M., Skytthe, A., and Frederiksen, H. (2015). Genetic and environmental influences in Dupuytren's disease: a study of 30,330 Danish twin pairs. J. Hand Surg. Eur. Vol. *40*, 171–176.
4. Ng, M., Thakkar, D., Southam, L., Werker, P., Ophoff, R., Becker, K., Nothnagel, M., Franke, A., Nürnberg, P., Espirito-Santo, A.I., et al. (2017). A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. Am. J. Hum. Genet. *101*, 417–427.
5. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
6. Staats, K.A., Wu, T., Gan, B.S., O'Gorman, D.B., and Ophoff, R.A. (2016). Dupuytren's disease susceptibility gene, EPDR1, is involved in myofibroblast contractility. J. Dermatol. Sci. *83*, 131–137.
7. 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., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
8. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., GTEx Consortium, Nicolae, D.L., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
9. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
10. Laakso, M., Kuusisto, J., Stančáková, A., Kuulasmaa, T., Pajukanta, P., Lusis, A.J., Collins, F.S., Mohlke, K.L., and Boehnke, M. (2017). The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. J. Lipid Res. *58*, 481–493.
11. Raitakari, O.T., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., Hutri-Kähönen, N., Taittonen, L., Jokinen, E., Marniemi, J., et al. (2008). Cohort profile: the cardiovascular risk in Young Finns Study. Int. J. Epidemiol. *37*, 1220–1226.
12. Nuotio, J., Oikonen, M., Magnussen, C.G., Jokinen, E., Laitinen, T., Hutri-Kähönen, N., Kähönen, M., Lehtimäki, T., Taittonen, L., Tossavainen, P., et al. (2014). Cardiovascular risk factors in 2011 and secular trends since 2007: the Cardiovascular Risk in Young Finns Study. Scand. J. Public Health *42*, 563–571.
13. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453.
14. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.-H., et al. (2014). Heritability and genomics of gene expression in peripheral blood. Nat. Genet. *46*, 430–437.
15. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen

Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

16. Shi, H., Mancuso, N., Spendlove, S., and Pasaniuc, B. (2017). Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. Am. J. Hum. Genet. *101*, 737–751.
17. Johnson, R., Shi, H., Pasaniuc, B., and Sankararaman, S. (2018). A unifying framework for joint trait analysis under a non-infinitesimal model. Bioinformatics *34*, i195–i201.
18. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. *48*, 709–717.
19. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.
20. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
21. Mancuso, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., and Pasaniuc, B. (2017). Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. Am. J. Hum. Genet. *100*, 473–487.
22. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
23. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics *32*, 283–285.
24. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. *50*, 621–629.
25. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. *6*, 5890.
26. Fehrmann, R.S.N., Karjalainen, J.M., Krajewska, M., Westra, H.-J., Maloney, D., Simeonov, A., Pers, T.H., Hirschhorn, J.N., Jansen, R.C., Schultes, E.A., et al. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat. Genet. *47*, 115–125.
27. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
28. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
29. Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., et al. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. *49*, 1421–1427.

30. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.
31. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.
32. Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell *172*, 1091–1107.e17.
33. Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346.
34. Saunders, A., Macosko, E., Wysoker, A., Goldman, M., Krienen, F., de Rivera, H., Bien, E., Baum, M., Wang, S., Goeva, A., et al. (2018). A Single-Cell Atlas of Cell Types, States, and Other Transcriptional Patterns from Nine Regions of the Adult Mouse Brain.
35. Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods *14*, 955–958.
36. Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999–1014.e22.
37. Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P.V., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153.
38. Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.
39. Zhou, Y., Al-Saaidi, R.A., Fernandez-Guerra, P., Freude, K.K., Olsen, R.K.J., Jensen, U.B., Gregersen, N., Hyttel, P., Bolund, L., Aagaard, L., et al. (2017). Mitochondrial Spare Respiratory Capacity Is Negatively Correlated with Nuclear Reprogramming Efficiency. Stem Cells Dev. *26*, 166–176.
40. Furlan, A., La Manno, G., Lübke, M., Häring, M., Abdo, H., Hochgerner, H., Kupari, J., Usoskin, D., Airaksinen, M.S., Oliver, G., et al. (2016). Visceral motor neuron diversity delineates a cellular basis for nipple- and pilo-erection muscle control. Nat. Neurosci. *19*, 1331–1340.
41. La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566–580.e19.
42. Hochgerner, H., Zeisel, A., Lönnerberg, P., and Linnarsson, S. (2018). Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. Nat. Neurosci. *21*, 290–299.
43. Hochgerner, H., Lönnerberg, P., Hodge, R., Mikes, J., Heskol, A., Hubschle, H., Lin, P., Picelli, S., La Manno, G., Ratz, M., et al. (2017). STRT-seq-2i: dual-index 5' single cell and nucleus RNA-seq on an addressable microwell array. Sci. Rep. *7*, 16327.
44. Romanov, R.A., Zeisel, A., Bakker, J., Girach, F., Hellysaz, A., Tomer, R., Alpár, A., Mulder, J., Clotman, F., Keimpema, E., et al. (2017). Molecular interrogation of

hypothalamic organization reveals distinct dopamine neuronal subtypes. Nat. Neurosci. *20*, 176–188.

45. Joost, S., Zeisel, A., Jacob, T., Sun, X., La Manno, G., Lönnerberg, P., Linnarsson, S., and Kasper, M. (2016). Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. Cell Syst *3*, 221–237.e9.
46. Häring, M., Zeisel, A., Hochgerner, H., Rinwa, P., Jakobsson, J.E.T., Lönnerberg, P., La Manno, G., Sharma, N., Borgius, L., Kiehn, O., et al. (2018). Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types. Nat. Neurosci. *21*, 869–880.
47. Chen, R., Wu, X., Jiang, L., and Zhang, Y. (2017). Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. Cell Rep. *18*, 3227–3241.
48. Vanlandewijck, M., He, L., Mäe, M.A., Andrae, J., Ando, K., Del Gaudio, F., Nahar, K., Lebouvier, T., Laviña, B., Gouveia, L., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. Nature *554*, 475–480.
49. Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., and Quake, S.R. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell *171*, 321–330.e14.
50. Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature *555*, 524–528.
51. Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Südhof, T.C., and Quake, S.R. (2016). Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep. *16*, 1126–1137.
52. Breton, G., Zheng, S., Valieris, R., Tojal da Silva, I., Satija, R., and Nussenzweig, M.C. (2016). Human dendritic cells (DCs) are derived from distinct circulating precursors that are precommitted to become CD1c+ or CD141+ DCs. J. Exp. Med. *213*, 2861–2870.
53. Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. *20*, 1215–1228.
54. Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat. Neurosci. *20*, 484–496.
55. Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333–339.
56. Alles, J., Karaiskos, N., Praktiknjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol. *15*, 44.
57. Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. U. S. A. *112*, 7285–7290.
58. Hu, P., Fabyanic, E., Kwon, D.Y., Tang, S., Zhou, Z., and Wu, H. (2017). Dissecting Cell-Type Composition and Activity-Dependent Transcriptional State in Mammalian Brains by Massively Parallel Single-Nucleus RNA-Seq. Mol. Cell *68*, 1006–1015.e7.

59. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
60. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. *99*, 139–153.
61. Chammas, M., Bousquet, P., Renard, E., Poirier, J.-L., Jaffiol, C., and Allieu, Y. (1995). Dupuytren's disease, carpal tunnel syndrome, trigger finger, and diabetes mellitus. J. Hand Surg. Am. *20*, 109–114.
62. Sanderson, P.L., Morris, M.A., Stanley, J.K., and Fahmy, N.R. (1992). Lipids and Dupuytren's disease. J. Bone Joint Surg. Br. *74-B*, 923–927.
63. Hacquebord, J.H., Chiu, V.Y., and Harness, N.G. (2017). The Risk of Dupuytren Diagnosis in Obese Individuals. J. Hand Surg. Am. *42*, 149–155.
64. Voight, B.F., Kang, H.M., Ding, J., Palmer, C.D., Sidore, C., Chines, P.S., Burtt, N.P., Fuchsberger, C., Li, Y., Erdmann, J., et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. *8*, e1002793.
65. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.
66. Horikoshi, M., Beaumont, R.N., Day, F.R., Warrington, N.M., Kooijman, M.N., Fernandez-Tajes, J., Feenstra, B., van Zuydam, N.R., Gaulton, K.J., Grarup, N., et al. (2016). Genome-wide associations for birth weight and correlations with adult disease. Nature *538*, 248–252.
67. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J. 'an, Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.
68. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.
69. Felix, J.F., Bradfield, J.P., Monnereau, C., van der Valk, R.J.P., Stergiakouli, E., Chesi, A., Gaillard, R., Feenstra, B., Thiering, E., Kreiner-Møller, E., et al. (2016). Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. Hum. Mol. Genet. *25*, 389–403.
70. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. *42*, 105–116.
71. van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. Nature *492*, 369–375.
72. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., et al. (2010). Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic

565 pathways. Diabetes *59*, 3229–3239.
566 73. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic,
567 J., Elling, U., Goodall, A.H., Labrune, Y., et al. (2011). New gene functions in
568 megakaryopoiesis and platelet formation. Nature *480*, 201–208.
569 74. Teumer, A., Tin, A., Sorice, R., Gorski, M., Yeo, N.C., Chu, A.Y., Li, M., Li, Y., Mijatovic,
570 V., Ko, Y.-A., et al. (2016). Genome-wide Association Studies Identify Genetic Loci
571 Associated With Albuminuria in Diabetes. Diabetes *65*, 803–817.
572 75. Pattaro, C., Teumer, A., Gorski, M., Chu, A.Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A.,
573 Sorice, R., Li, Y., et al. (2016). Genetic associations at 53 loci highlight cell types and
574 biological pathways relevant for kidney function. Nat. Commun. *7*, 10023.
575 76. Eppinga, R.N., Hagemeijer, Y., Burgess, S., Hinds, D.A., Stefansson, K., Gudbjartsson, D.F.,
576 van Veldhuisen, D.J., Munroe, P.B., Verweij, N., and van der Harst, P. (2016).
577 Identification of genomic loci associated with resting heart rate and shared genetic
578 predictors with all-cause mortality. Nat. Genet. *48*, 1557–1563.
579 77. Nikpay, M., Goel, A., Won, H.-H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D.,
580 Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000
581 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat.
582 Genet. *47*, 1121–1130.
583 78. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna,
584 A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of loci
585 associated with lipid levels. Nat. Genet. *45*, 1274–1283.
586 79. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee,
587 J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility
588 loci for inflammatory bowel disease and highlight shared genetic risk across populations.
589 Nat. Genet. *47*, 979–986.
590 80. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki,
591 A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and
592 drug discovery. Nature *506*, 376–381.
593 81. Otowa, T., Hek, K., Lee, M., Byrne, E.M., Mirza, S.S., Nivard, M.G., Bigdeli, T., Aggen,
594 S.H., Adkins, D., Wolen, A., et al. (2016). Meta-analysis of genome-wide association
595 studies of anxiety disorders. Mol. Psychiatry *21*, 1485.
596 82. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke, S.,
597 Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M.,
598 Blackwood, D.H.R., Boomsma, D.I., et al. (2013). A mega-analysis of genome-wide
599 association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.
600 83. Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011). Large-scale
601 genome-wide association analysis of bipolar disorder identifies a new susceptibility locus
602 near ODZ4. Nat. Genet. *43*, 977–983.
603 84. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological
604 insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.
605 85. Okbay, A., Baselmans, B.M.L., De Neve, J.-E., Turley, P., Nivard, M.G., Fontana, M.A.,
606 Meddens, S.F.W., Linnér, R.K., Rietveld, C.A., Derringer, J., et al. (2016). Genetic
607 variants associated with subjective well-being, depressive symptoms, and neuroticism
608 identified through genome-wide analyses. Nat. Genet. *48*, 624–633.
609 86. Springelkamp, H., Iglesias, A.I., Mishra, A., Höhn, R., Wojciechowski, R., Khawaja, A.P.,
610 Nag, A., Wang, Y.X., Wang, J.J., Cuellar-Partida, G., et al. (2017). New insights into the

611 genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure
612 and optic disc characteristics. Hum. Mol. Genet. *26*, 438–453.
613 87. Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy,
614 P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new
615 breast cancer risk loci. Nature *551*, 92–94.

FIGURE TITLES AND LEGENDS

Figure 1: DD TWAS and GWAS associations. Shown here are Manhattan plots for TWAS associations (top) and GWAS associations (bottom). For TWAS associations, each point corresponds to an association test between tissue-specific predicted gene expression and DD, with the orange line representing the threshold for significance in log-scale ($P_{TWAS} \leq 5.09 \times 10^{-7}$). The most significant tissue-specific gene model for each peak is labeled by gene. For GWAS associations, each point corresponds to an association test between a SNP and DD, with the orange line representing the traditional genome-wide significance threshold in log-scale ($P_{GWAS} \leq 5 \times 10^{-8}$).

Figure 2: Novel risk region identified on chromosome 17. Shown here is the novel risk region identified through TWAS; the grey points are GWAS SNPs association strength and the blue points are the GWAS SNPs association strength conditioned on the *TMEM106A* expression model (green, significant in GTEx breast mammary tissue). This tissue-specific model was still significant under 1,347 permutations ($P_{perm} = 8.9 \times 10^{-3}$).

Figure 3: Tentative evidence for putative causality with DD. Here we show the genetic correlation for three different pairs of traits (DD/BMI, left; DD/TG, center; and BMI/TG, right) between four groupings of SNPs: (1) GWAS-significant SNPs specific to trait 1, (2) GWAS-significant SNPs specific to trait 2, (3) GWAS-significant SNPs for both trait 1 and trait 2, and (4) all non-significant SNPs shared between studies. In the left plot, since GWAS-significant SNPs specific to BMI have more enrichment of genetic correlation compared to those specific to DD, we can putatively interpret that BMI SNPs are driving the shared genetic etiology. The same can be said for the middle plot with TG. On the right, for completeness, we show the same correlation for BMI and TG, which was significant. Error bars are defined by the genetic correlation ± 1.96 times the s.e. for each grouping of SNPs.

TABLES

Table 1: 43 significant tissue-specific gene expression models from TWAS. These are the 43 significant ($P_{TWAS} \leq 0.05/98{,}147$) tissue-specific gene models across 18 genes.

Gene	Chr	TSS	TES	Best GWAS SNP	Z_{GWAS}	Reference Tissue Panel	cis-h_g^2	P_{TWAS}
PJA2	5	108,670,409	108,745,675	rs414724	-6.4	GTEx.Nerve_Tibial	0.09	1.1E-07
CTD-2587M2.1	5	108,572,821	108,662,070	rs414724	-6.4	METSIM.ADIPOSE.RNASEQ	0.18	3.3E-10
MAN2A1	5	109,025,066	109,205,326	rs414724	-6.4	GTEx.Nerve_Tibial	0.07	2.4E-08
SDK1	7	3,341,079	4,308,631	rs10264803	-6.0	GTEx.Cells_Transformed_fibroblasts	0.17	3.3E-09
						GTEx.Esophagus_Muscularis	0.08	6.4E-08
EPDR1	7	37,960,162	37,991,542	rs17171240	14.7	GTEx.Lung	0.15	6.4E-31
						GTEx.Adipose_Subcutaneous	0.12	5.1E-23
						GTEx.Pancreas	0.19	4.6E-18
						GTEx.Esophagus_Muscularis	0.35	6.6E-14
						YFS.BLOOD.RNAARR	0.18	1.4E-13
						GTEx.Nerve_Tibial	0.30	5.0E-09
						GTEx.Artery_Tibial	0.27	1.5E-08
						GTEx.Thyroid	0.19	5.4E-08
						GTEx.Cells_Transformed_fibroblasts	0.21	1.9E-07
						CMC.BRAIN.RNASEQ	0.24	3.1E-07
TRGC2	7	38,279,181	38,289,173	rs17171240	14.7	GTEx.Prostate	0.37	1.4E-12
SULF1	8	70,378,858	70,573,147	rs542288	11.8	GTEx.Artery_Aorta	0.27	4.0E-25
RSPO2	8	108,911,543	109,095,913	rs612265	-9.3	CMC.BRAIN.RNASEQ	0.12	1.2E-08
EIF3E	8	109,213,971	109,260,959	rs612265	-9.3	CMC.BRAIN.RNASEQ	0.07	7.6E-21
						YFS.BLOOD.RNAARR	0.01	1.4E-15
						GTEx.Brain_Cerebellum	0.16	1.8E-11
EMC2	8	109,455,852	109,499,136	rs612265	-9.3	GTEx.Muscle_Skeletal	0.05	8.9E-12
						GTEx.Esophagus_Gastroesophageal_Junction	0.09	1.3E-09
						GTEx.Brain_Cerebellum	0.29	1.6E-08
MRPL52	14	23,299,091	23,304,246	rs1042704	7.3	YFS.BLOOD.RNAARR	0.44	1.1E-07
NEDD4	15	56,119,116	56,285,944	rs8032158	5.2	GTEx.Artery_Tibial	0.18	2.6E-07
BCAR1	16	75,262,927	75,301,951	rs977987	5.9	GTEx.Artery_Aorta	0.18	2.8E-07
						GTEx.Esophagus_Mucosa	0.16	3.6E-07

CFDP1	16	75,327,607	75,467,387	rs977987	5.9	YFS.BLOOD.RNAARR	0.21	5.6E-08
TMEM170A	16	75,480,922	75,498,584	rs977987	5.9	GTEx.Cells_EBV-transformed_lymphocytes	0.13	1.4E-09
						GTEx.Skin_Sun_Exposed_Lower_leg	0.14	2.7E-08
						GTEx.Skin_Not_Sun_Exposed_Suprapubic	0.09	3.9E-08
TMEM106A	17	41,363,845	41,372,057	rs4793248	4.1	GTEx.Breast_Mammary_Tissue	0.12	1.2E-07
ATXN10	22	46,067,677	46,241,187	rs34088184	13.8	GTEx.Cells_Transformed_fibroblasts	0.17	1.7E-07
LINC00899	22	46,435,786	46,440,748	rs34088184	13.8	GTEx.Adipose_Subcutaneous	0.17	2.8E-32
						GTEx.Muscle_Skeletal	0.24	1.6E-26
						GTEx.Cells_Transformed_fibroblasts	0.18	1.9E-26
						GTEx.Artery_Tibial	0.30	5.6E-23
						GTEx.Esophagus_Muscularis	0.36	2.3E-15
						GTEx.Lung	0.21	5.7E-15
						GTEx.Adrenal_Gland	0.68	9.2E-15
						GTEx.Artery_Coronary	0.56	1.6E-09
						GTEx.Nerve_Tibial	0.35	3.4E-08

Table 2: Genetic correlation results between DD and 45 other traits. We have grouped related traits under the "Type" column. The four traits that were significantly ($P_{T1,T2} \leq 0.05/45$) correlated with DD are shown in italics.

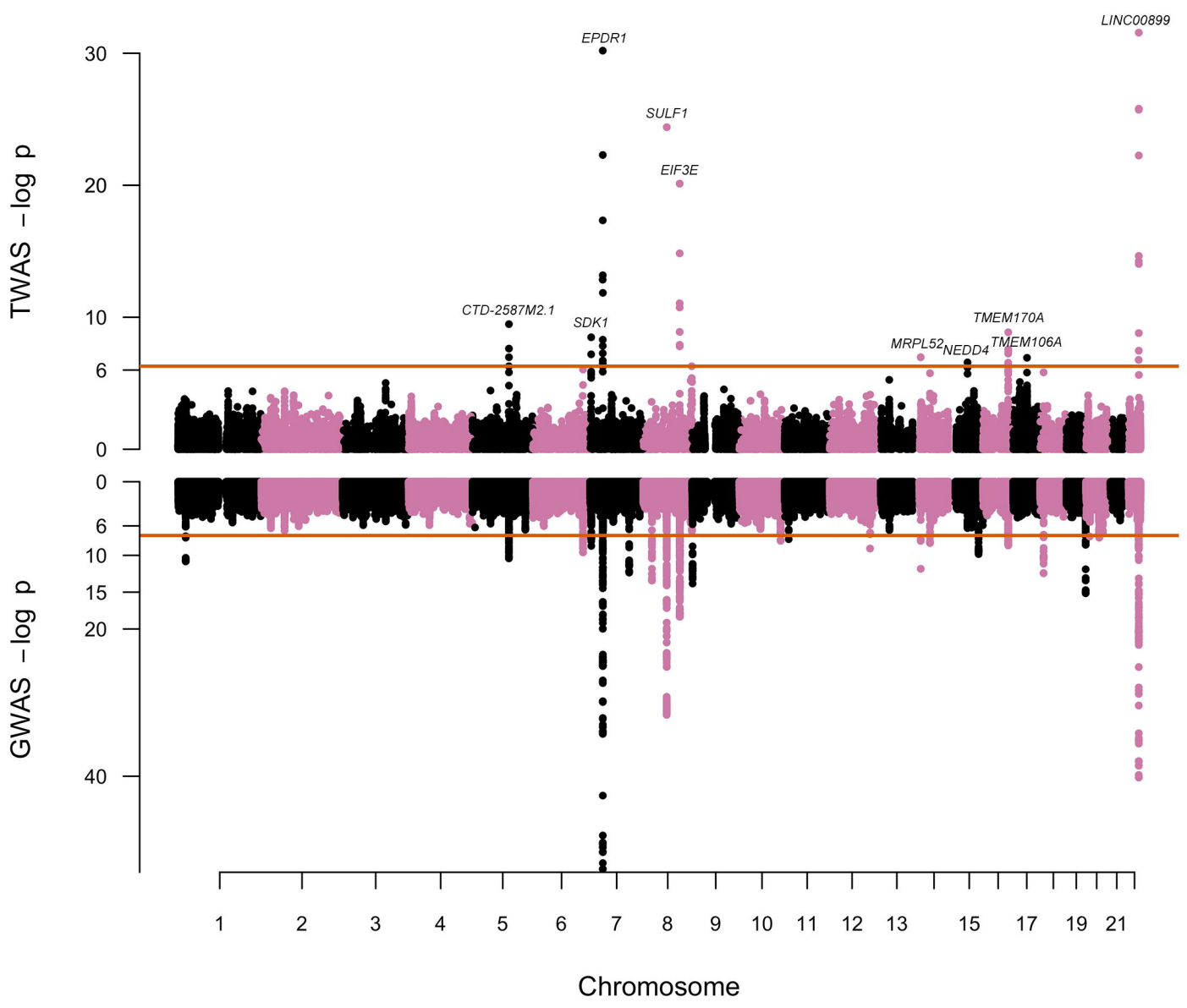
Type	Trait	Sample Size	SNP-$\hat{h}_g^2$ (s.e.)	$\hat{r}_g$ (s.e.)	$P_{T1,T2}$
Skeletal Traits	Birth Weight[66]	153781	0.1 (0.007)	-0.051 (0.05)	3.0E-01
	Height[67]	253288	0.34 (0.017)	0.063 (0.03)	7.3E-02
	Body Mass Index[68]	*336107*	*0.25 (0.009)*	*-0.196 (0.04)*	*1.6E-06*
	Childhood Body Mass Index[69]	35668	0.25 (0.024)	0.005 (0.06)	9.3E-01
	Heel Bone Material Density[68]	194398	0.28 (0.025)	0.057 (0.05)	2.1E-01
Blood and Diabetes Traits	Fasting Glucose[70]	46186	0.08 (0.014)	0.011 (0.09)	9.0E-01
	Fasting Insulin[70]	46186	0.06 (0.01)	-0.092 (0.09)	2.9E-01
	Type II Diabetes[68]	*336473*	*0.04 (0.003)*	*-0.182 (0.05)*	*1.7E-04*
	Hemoglobin[71]	135367	0.09 (0.013)	0.131 (0.08)	1.1E-01
	Hemoglobin A1C[72]	46368	0.06 (0.011)	-0.101 (0.09)	2.8E-01
	Packed Cell Volume[71]	135367	0.08 (0.014)	0.171 (0.09)	5.2E-02
	Mean Cell Hemoglobin[71]	135367	0.22 (0.026)	-0.014 (0.04)	7.4E-01
	Mean Cell Hemoglobin Concentration[71]	172433	0.03 (0.011)	-0.047 (0.11)	6.8E-01
	Mean Corpuscular Volume[71]	172433	0.24 (0.025)	-0.001 (0.04)	9.8E-01

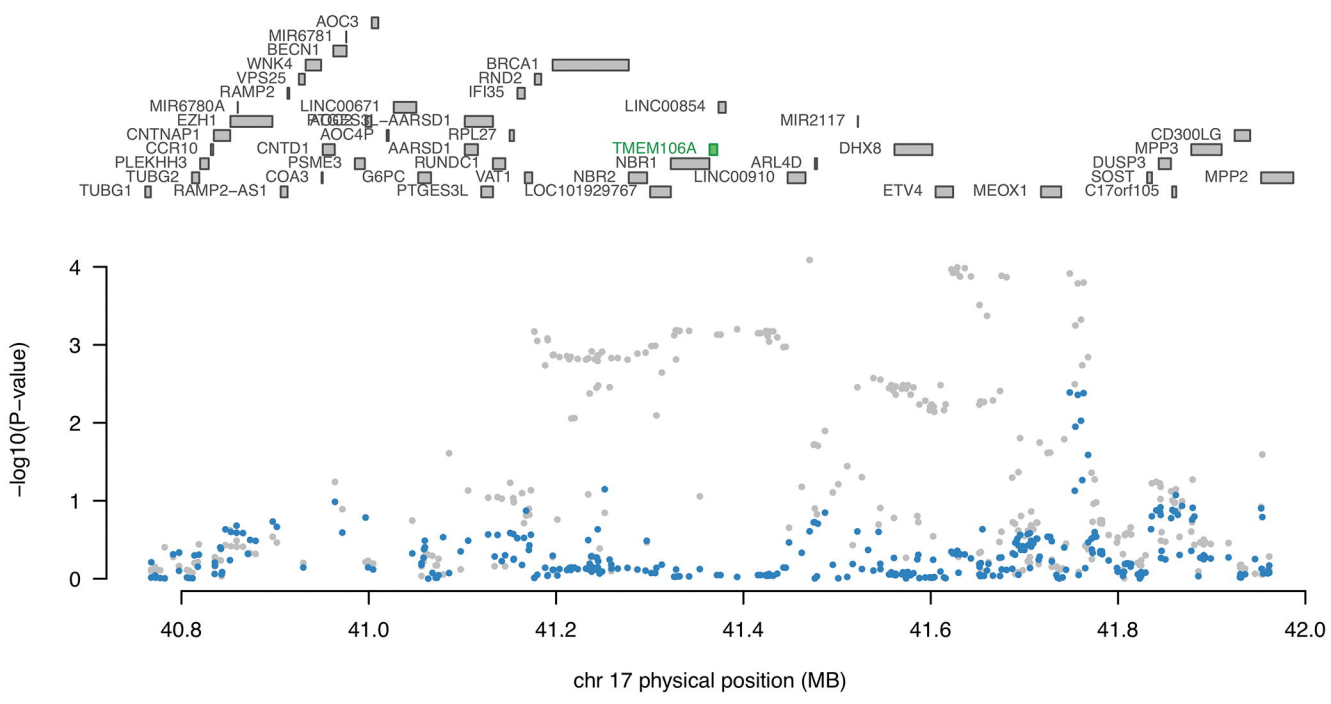
	Red Blood Cell Count[71]	35604	0.13 (0.019)	0.139 (0.08)	6.8E-02
	Platelet Count[73]	66867	0.11 (0.011)	-0.117 (0.06)	4.1E-02
Renal Traits	Chronic Kidney Disease[74,75]	117165	0.02 (0.006)	-0.13 (0.12)	2.7E-01
	Urine Albumin-to-Creatinine Ratio[74]	51886	0.04 (0.009)	0.07 (0.11)	5.3E-01
	Microalbuminuria[74]	51886	0.01 (0.008)	0.04 (0.17)	8.2E-01
Cardiovascular Traits	Resting Heart Rate[76]	134251	0.14 (0.012)	0.045 (0.05)	3.5E-01
	Coronary Artery Disease[77]	184305	0.07 (0.005)	-0.098 (0.05)	6.9E-02
	Triglycerides[78]	*188577*	*0.26 (0.057)*	*-0.139 (0.04)*	*3.5E-04*
	High Density Lipoprotein[78]	*188577*	*0.24 (0.036)*	*0.133 (0.04)*	*4.1E-04*
	Low Density Lipoprotein[78]	188577	0.2 (0.048)	-0.043 (0.04)	2.8E-01
	Total Cholesterol[78]	188577	0.21 (0.046)	-0.033 (0.04)	3.9E-01
Autoimmune Traits	Crohn's Disease[79]	27726	0.38 (0.047)	0.009 (0.07)	9.0E-01
	Inflammatory Bowel Disease[79]	34694	0.32 (0.035)	0.021 (0.07)	7.7E-01
	Ulcerative Colitis[79]	28738	0.22 (0.032)	0.073 (0.08)	3.5E-01
	Rheumatoid Arthritis[80]	58284	0.15 (0.028)	0.124 (0.05)	2.1E-02
Neurological Traits	Anxiety Case-Control[81]	18000	0.07 (0.03)	0.082 (0.17)	6.3E-01
	Major Depressive Disorder[82]	18759	0.15 (0.03)	0.012 (0.11)	9.2E-01
	Bipolar Disorder[83]	16731	0.45 (0.042)	0.101 (0.08)	2.0E-01
	Schizophrenia[84]	150064	0.45 (0.018)	0.064 (0.04)	1.2E-01
	Neuroticism[85]	170911	0.09 (0.006)	-0.067 (0.05)	2.1E-01
Eye Traits	Glaucoma[68]	108817	0.04 (0.005)	0.06 (0.08)	4.6E-01
	Myopia[68]	335700	0.03 (0.002)	0.039 (0.06)	5.4E-01
	Intraocular Pressure[86]	29578	0.13 (0.021)	0.061 (0.09)	4.7E-01
	Cup Area[86]	22489	0.28 (0.037)	0.098 (0.06)	9.9E-02
	Disc Area[86]	22504	0.3 (0.072)	0.035 (0.06)	5.9E-01
	Vertical Cup-Disc Ratio[86]	23899	0.33 (0.045)	0.099 (0.05)	6.1E-02
Other Traits	Subjective Well-being[85]	298420	0.03 (0.002)	0.07 (0.06)	2.7E-01
	Asthma[68]	83529	0.07 (0.01)	0.116 (0.07)	1.1E-01
	Breast Cancer[87]	228951	0.13 (0.011)	0.076 (0.04)	7.3E-02
	Hand Grip Strength (left)[68]	335821	0.1 (0.004)	-0.005 (0.04)	9.1E-01
	Hand Grip Strength (right)[68]	335842	0.1 (0.004)	0.003 (0.04)	9.3E-01

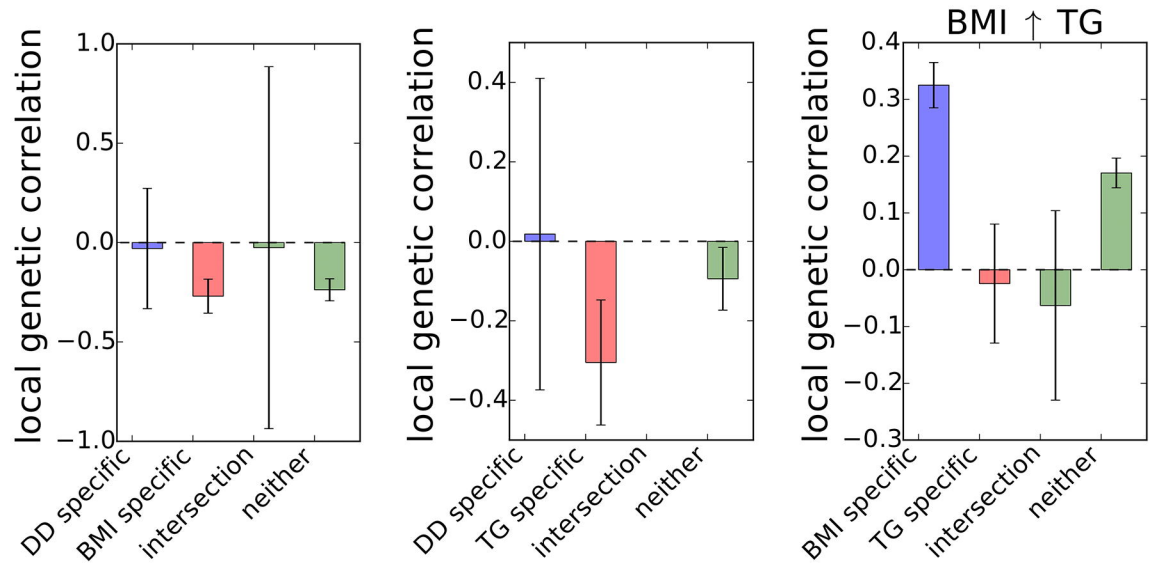
Table 3: Regions with significant genetic correlation between DD and other traits. This table lists the eight regions demonstrating significant genetic correlation between DD and BMI, and the three regions demonstrating significant correlation between DD and HDL; significance was assessed at a Bonferroni-corrected threshold of $P_{region} \leq 0.05/1702$ for each trait. Also

included is the number of SNPs within each region ("# SNPs") as well as the minimum GWAS association p-value for either BMI or HDL ("Min. Trait P_{GWAS}") and DD ("Min. DD P_{GWAS}"). All other regions demonstrated no significant genetic correlation between DD and any trait tested.

Trait	Chr	Start	End	# SNPs	Min. Trait P_{GWAS}	Min. DD P_{GWAS}	$\hat{r}_{g,local}$	s.e.	P_{region}
BMI	1	21736588	23086883	2566	9.6E-05	1.8E-12	-0.00082	0.00017	1.6E-06
BMI	1	189904130	191868930	3313	4.7E-14	5.0E-05	-0.00093	0.0002	2.3E-06
BMI	2	209941529	212379518	1668	1.2E-10	0.00129	-0.00099	0.00023	1.3E-05
BMI	3	49316972	51832015	1872	9.4E-40	0.00230	-0.00146	0.00027	5.6E-08
BMI	3	51832015	54081390	2225	2.6E-10	0.00061	-0.00105	0.00022	2.3E-06
BMI	4	43965045	45189157	2525	9.6E-33	9.2E-05	-0.00101	0.00023	1.5E-05
BMI	4	45189157	47411896	2781	3.2E-12	0.00026	-0.00079	0.00018	1.2E-05
BMI	6	28917608	29737971	60	5.3E-09	0.01434	0.00041	9.00E-05	1.6E-05
HDL	7	37555184	38966703	1221	0.00018	3.4E-49	-0.00131	0.00031	2.0E-05
HDL	9	1079707	1916877	1143	0.00018	2.8E-15	0.00142	0.0003	1.5E-06
HDL	12	39227169	40816185	1246	0.01077	0.00019	0.00116	0.00027	1.5E-05


EPDR1
SULF1
EIF3E
CTD-2587M2.1
SDK1
MRPL52
NEDD4
TMEM170A
TMEM106A
LINC00899
TWAS −log p
GWAS −log p
30
20
10
6
0
0
6
10
15
20
40
1
2
3
4
5
6
7
8
9
10
11
12
13
15
17
19
21
Chromosome


AOC3
MIR6781
BECN1
WNK4
VPS25
RAMP2
MIR6780A
EZH1
CNTNAP1
CCR10
PLEKHH3
TUBG2
TUBG1
RAMP2−AS1
LINC00671
CNTD1
PSME3
COA3
AOC4P
AARSD1
G6PC
PTGES3L
BRCA1
RND2
IFI35
RPL27
RUNDC1
VAT1
PTGES3
LOC101929767
TMEM106A
NBR1
NBR2
LINC00854
LINC00910
MIR2117
ARL4D
DHX8
ETV4
MEOX1
CD300LG
MPP3
DUSP3
SOST
C17orf105
MPP2
−log10(P−value)
0
1
2
3
4
40.8
41.0
41.2
41.4
41.6
41.8
42.0
chr 17 physical position (MB)


local genetic correlation
1.0
0.5
0.0
−0.5
−1.0
DD specific
BMI specific
intersection
neither
local genetic correlation
0.4
0.2
0.0
−0.2
−0.4
DD specific
TG specific
intersection
neither
BMI ↑ TG
local genetic correlation
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
BMI specific
TG specific
intersection
neither