nature genetics

Integrative approaches for large-scale transcriptome-wide association studies

Alexander Gusev[1–3], Arthur Ko[4,5], Huwenbo Shi[6], Gaurav Bhatia[1–3], Wonil Chung[1], Brenda W J H Penninx[7], Rick Jansen[7], Eco J C de Geus[8], Dorret I Boomsma[8], Fred A Wright[9], Patrick F Sullivan[10–12], Elina Nikkola[4], Marcus Alvarez[4], Mete Civelek[13], Aldons J Lusis[4,13], Terho Lehtimäki[14], Emma Raitoharju[14], Mika Kähönen[15], Ilkka Seppälä[14], Olli T Raitakari[16,17], Johanna Kuusisto[18], Markku Laakso[18], Alkes L Price[1–3], Päivi Pajukanta[4,5] & Bogdan Pasaniuc[4,6,19]

Many genetic variants influence complex traits by modulating gene expression, thus altering the abundance of one or multiple proteins. Here we introduce a powerful strategy that integrates gene expression measurements with summary association statistics from large-scale genome-wide association studies (GWAS) to identify genes whose *cis*-regulated expression is associated with complex traits. We leverage expression imputation from genetic data to perform a transcriptome-wide association study (TWAS) to identify significant expression-trait associations. We applied our approaches to expression data from blood and adipose tissue measured in ~3,000 individuals overall. We imputed gene expression into GWAS data from over 900,000 phenotype measurements to identify 69 new genes significantly associated with obesity-related traits (BMI, lipids and height). Many of these genes are associated with relevant phenotypes in the Hybrid Mouse Diversity Panel. Our results showcase the power of integrating genotype, gene expression and phenotype to gain insights into the genetic basis of complex traits.

Although a large proportion of variability in complex human traits is due to genetic variation, the mechanistic steps between genetic variation and traits are generally not understood[1–7]. Many genetic variants influence complex traits by modulating gene expression, thus altering the abundance of one or multiple proteins[8–12]. Such relationships between expression and traits could be investigated through association scans in individuals for whom both measurements are available[8,13,14]. Unfortunately, studies that measure gene expression have been hampered by specimen availability and cost, with the few published studies of expression and complex traits being orders of magnitude smaller than studies of traits alone. Consequently, many expression-trait associations cannot be detected, especially those with small effects. To mitigate the reduced power from small sample size, alternative approaches have examined the overlap of genetic variants that influence gene expression (expression quantitative trait loci, eQTLs) with trait-associated variants identified in large, independent GWAS[5,6,8,9,11–13,15]. However, this approach is also likely to miss expression-trait associations of small effect.

We developed a new approach to identify genes whose expression is significantly associated with complex traits in individuals without directly measured expression levels (Online Methods). We leveraged a relatively small set of reference individuals for whom both gene expression and genetic variation (SNPs) were measured to impute the *cis* genetic component of expression into a much larger set of phenotyped individuals using their SNP genotype data (**Fig. 1**). The imputed expression data can be viewed as a linear model of genotypes with weights based on the correlation between SNPs and gene expression in the training data while accounting for linkage disequilibrium (LD) among SNPs. We then correlated the imputed gene expression with traits to perform a TWAS and identify significant expression-trait associations (Online Methods). Work in parallel to ours has also proposed to find expression-trait associations through imputation

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. [4]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [5]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California, USA. [6]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, California, USA. [7]Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands. [8]Department of Biological Psychology, VU University, Amsterdam, the Netherlands. [9]Bioinformatics Research Center, Department of Statistics, Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA. [10]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [11]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA. [12]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [13]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [14]Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland. [15]Department of Clinical Physiology, Pirkanmaa Hospital District and University of Tampere School of Medicine, Tampere, Finland. [16]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [17]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [18]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [19]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. Correspondence should be addressed to A.G. (agusev@hsph.harvard.edu) or B.P. (pasaniuc@ucla.edu).

Received 22 June 2015; accepted 14 January 2016; published online 8 February 2016; doi:10.1038/ng.3506

of gene expression when GWAS data at an individual level are available[16]. However, a critical limitation is that large-scale GWAS data are typically only publicly available at the level of summary association statistics (for example, SNP effect sizes)[2–4]. To capitalize on the largest GWAS studies performed thus far (typically with data available only at the summary level), we extended our approach to impute the expression-trait association statistics directly from GWAS summary statistics (Online Methods). In contrast to expression imputation from individual-level data[16], imputation of expression-trait associations from GWAS summary statistics can exploit publically available data from hundreds of thousands of samples. Linear predictors naturally extend to indirect imputation of the standardized effect of *cis* genetic components of expression on traits starting from only GWAS association statistics[2–4] (Online Methods). This allowed us to increase the effective sample size for expression-trait association testing to hundreds of thousands of individuals. By focusing only on the genetic component of expression, we avoid instances of expression-trait association that are not a consequence of genetic variation but are driven by variation in traits (**Fig. 2**). Our approach can be conceptualized as a test for significant *cis* genetic correlation between expression and traits.

We applied our approaches to expression data from blood and adipose tissue measured in ~3,000 individuals overall. Through extensive simulations and analyses of real data, we show that our proposed approach increases performance over standard GWAS and eQTL-guided GWAS. Furthermore, we reanalyzed a 2010 lipids GWAS[17] to find 25 new expression-trait associations in those data. Among these associations, 19 of 25 contained genome-wide significant SNPs in the more recent and expanded lipids study[5], thus showcasing the power of our approach to find robust associations. We imputed gene expression into GWAS data from over 900,000 phenotype measurements[5–7] to identify 69 new genes significantly associated with obesity-related traits (body mass index (BMI), lipids and height). Many of these genes were associated with relevant phenotypes in the Hybrid Mouse Diversity Panel (HMDP). Overall, our results showcase the power of integrating genotype, gene expression and phenotype to gain insights into the genetic basis of complex traits.

RESULTS

SNP heritability of gene expression

To investigate the potential use of a TWAS based on imputed gene expression, we first estimated the *cis* (1-Mb window around a gene) and *trans* (rest of the genome) SNP heritability (*cis*- and *trans*- h_g^2) for each gene in our data[18,19]. These metrics quantify the maximum possible accuracy (in terms of R^2) of a linear predictor from the corresponding set of SNPs[20,21] (Online Methods). We used 3,234 individuals for whom genome-wide SNP data and expression measurements were available from the Metabolic Syndrome in Men (METSIM; adipose), Young Finns Study (YFS; blood) and Netherlands Twins Registry (NTR; blood) data sets[22–24] (Online Methods and **Supplementary Table 1**). All expression measurements were adjusted for batch confounders, and array probes were merged into a single expression value for each gene, where possible (Online Methods). Consistent with previous work[24,25], we observed significantly nonzero ($P < 1 \times 10^{-16}$) estimates of heritability across all three studies, with mean *cis*-h_g^2 values ranging from 0.01 to 0.07 and mean *trans*-h_g^2 values ranging from 0.04 to 0.06 in genes where estimates converged (**Supplementary Fig. 1** and **Supplementary Table 1**). Although we observed large differences in the average *cis*-h_g^2 estimates between the two blood cohorts, the estimates were strongly correlated across genes (Pearson $\rho = 0.47$ for YFS-NTR, as compared to $\rho = 0.15$ and $\rho = 0.26$ for METSIM-NTR and METSIM-YFS, respectively). This is consistent with a common but not identical genetic architecture. The *cis*-h_g^2 estimate was significantly nonzero (by likelihood-ratio test) for 6,924 genes after accounting for multiple hypotheses (1,985 for METSIM, 3,836 for YFS and 1,103 for NTR) (**Supplementary Fig. 1**), whereas current sample sizes were too small to detect individually significant *trans*-heritable genes. As expected, we also observed a high overlap of genes with significant *cis*-h_g^2 estimates across cohorts (**Supplementary Fig. 2** and **Supplementary Table 2**). We focused subsequent analyses on the 6,924 *cis*-heritable genes, as such genes are typically enriched for trait associations[7,9,13,24–29].

TWAS performance in simulation and cross-validation

We evaluated whether the expression levels of the 6,924 highly heritable genes could be accurately imputed from *cis*-SNP genotype data alone in these three cohorts. In each tissue, we used cross-validation to compare predictions from the best *cis*-eQTL to those from all SNPs at the locus either in a best linear unbiased predictor (BLUP) or a Bayesian model[30,31] (Online Methods). On average, the Bayesian linear mixed model (BSLMM)[31], which uses all *cis*-SNPs and estimates the underlying effect size distribution, attained the best performance, with a 32% gain in prediction R^2 over a prediction computed using only the top *cis*-eQTL (**Fig. 3** and **Supplementary Fig. 3**). BSLMM exhibited a long tail of increased accuracy, more than doubling the prediction R^2 for 25% of genes (**Supplementary Fig. 4**). In contrast to complex traits, where hundreds of thousands of training samples are required for accurate prediction[32,33], a substantial portion of variance in expression can be predicted at current sample sizes because of the much smaller number of independent SNPs in the *cis* region[21]. Furthermore, larger training sizes will continue to increase the total number of genes that can be accurately predicted (**Fig. 4**). We further evaluated cross-cohort prediction of these genes in the YFS and NTR cohorts, which were roughly equally sized and had expression measured in whole blood by microarray but were genotyped on different platforms and were from different Scandinavian populations. After accounting for *cis* heritability in the test cohort,

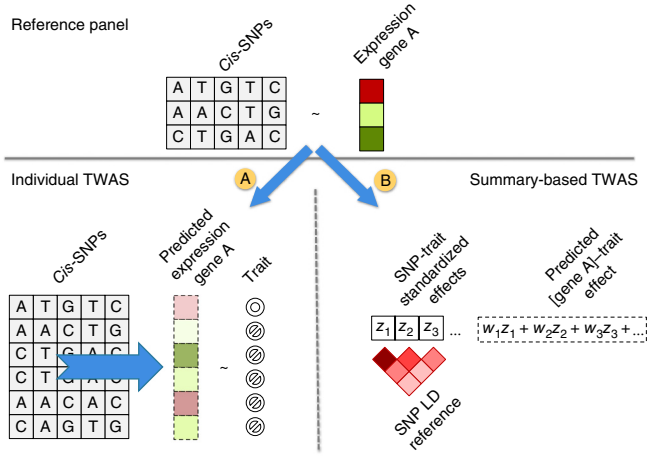

Figure 1 Schematic of the TWAS approach. Top, estimate gene expression effect sizes in the reference panel either directly (eQTL), modeling all SNPs (BLUP), or modeling SNPs and effect sizes (BSLMM). Path A: predict expression directly for genotyped samples using the effect sizes from the reference panel and measure the association between predicted expression and a trait. Path B: indirectly estimate association between predicted expression and a trait as the weighted linear combination of SNP-trait standardized effect sizes while accounting for LD among SNPs.

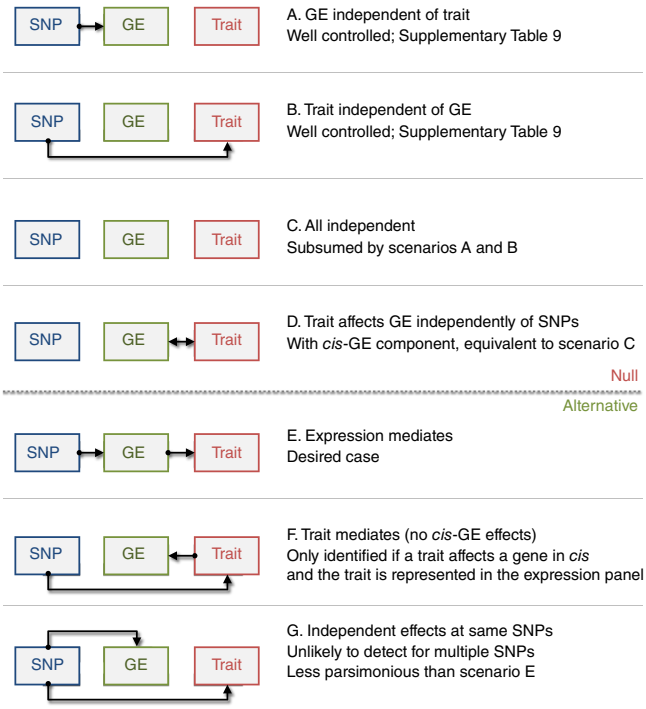

Figure 2 Modes of expression causality. Diagrams are shown for the possible modes of causality for the relationship between genetic markers (SNPs; blue), gene expression (GE; green) and traits (red). Scenarios A–D would be considered null by the TWAS model. Scenarios E–G could be identified as significant.

our cross-cohort standardized accuracy ($R^2/cis\text{-}h^2_g$) was broadly consistent with in-cohort cross-validation accuracy (**Supplementary Table 3**). BSLMM was again the most accurate predictor, with an average cross-cohort $R^2/cis\text{-}h^2_g$) value of 72%, outperforming the best eQTL by an average of 1.17×.

Next, we focused on evaluating the power of the TWAS approach to detect significant expression-trait associations using GWAS summary data from complex traits (equivalent to TWAS from individual-level data; Online Methods and **Supplementary Fig. 5**). For comparison, we also measured power to detect significant SNP-trait associations through standard GWAS (testing each SNP individually) and eQTL-based GWAS (eGWAS; where the best eQTL in each gene is the only variant tested for association), with all three tests corrected for their genome-wide testing burdens. Using real genotype data, we simulated a causal SNP-expression-trait model with realistic effect sizes and measured the power of each strategy to identify genome-wide significant variants (accounting for 1 million SNPs for GWAS and 15,000 expressed genes using family-wise error rate control). Over many diverse disease architectures, TWAS substantially increased power when the expression-causing variants were untyped or poorly tagged by an individual SNP (**Fig. 5** and **Supplementary Figs. 6–11**). The greatest gains in power were observed in the case of multiple causal variants: 92% power for TWAS as compared to 18% and 25% power for GWAS and eGWAS, respectively. This scenario would correspond to expression caused by allelic heterogeneity[9,34,35], or 'apparent' heterogeneity, at common variants (due to tagging of an unobserved causal variant)[36]. TWAS was comparable to the other approaches when a single causal variant was directly typed, in which case combining the effects of neighboring SNPs does not add signal. Under the null hypothesis where expression was completely independent of phenotype (with either being heritable; scenarios A–D in **Fig. 2**), the TWAS false positive rate was well controlled (**Supplementary Table 4**). As expected, all methods were confounded in the case where the same causal variants had independent effects on traits and expression (scenarios F and G in **Fig. 2** and **Supplementary Figs. 8** and **12**).

Our approach can be conceptually viewed as a test for the correlation between the genetic component of expression and the genetic component of a trait (Online Methods). Because several recent methods have been proposed that measure genetic correlation between summary statistics[37], we sought to evaluate this relationship empirically. We compared TWAS to the recently proposed cross-trait LD score regression (LDSC) that estimates genome-wide genetic correlation between traits[37]. Although LDSC is not intended for local analyses because of model assumptions on polygenicity and use of block jackknife across loci for estimating standard errors, we performed the evaluation using expression and phenotype (height) data from the YFS cohort, using the results over individual data as the 'gold standard' (Online Methods). We found that the LDSC estimate of genetic correlation between height and expression from summary data was highly correlated with the gold standard (correlation = 0.7; **Supplementary Fig. 13**), but the relationship was much noisier than that of TWAS (correlation = 0.99; **Supplementary Figs. 5** and **13**). This suggests that TWAS attains more power than LDSC in relating expression to complex traits.

TWAS is also conceptually similar to a test for colocalization of signal between expression and a complex trait[38,39], and we compared it to a recently proposed method, COLOC[38], that evaluates colocalization of expression at known GWAS risk loci. After matching the false discovery rate of the two methods in simulations (Online Methods), TWAS and COLOC had similar power under the scenario with a single typed causal variant (with slightly lower COLOC power at small GWAS sizes), but TWAS had superior performance when the causal variant was untyped or in the presence of allelic heterogeneity (**Supplementary Fig. 10**). This is likely due the fact that TWAS explicitly models LD to better capture untyped variants.

Finally, we investigated the effect of the size of the expression reference panel on performance of TWAS (**Supplementary Fig. 9**). In general, TWAS always outperformed eGWAS when multiple variants were causal. Interestingly, power for either approach did not increase substantially beyond 1,000 expression samples, suggesting that the expression panels analyzed in this manuscript nearly saturate the

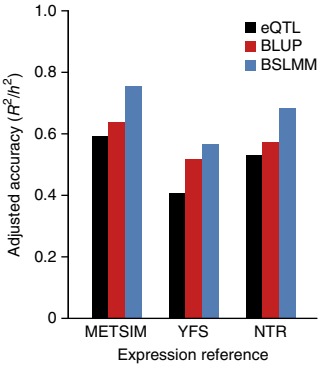

Figure 3 Accuracy of individual-level expression imputation algorithms. Adjusted accuracy was estimated using cross-validation R^2 between predicted and true expression and normalizing by corresponding $cis\text{-}h^2_g$. Bars show the mean estimate across three cohorts and three methods: eQTL, single best *cis*-eQTL in the locus; BLUP, using all SNPs in the locus; and BSLMM, using all SNPs in the locus and noninfinitesimal priors.

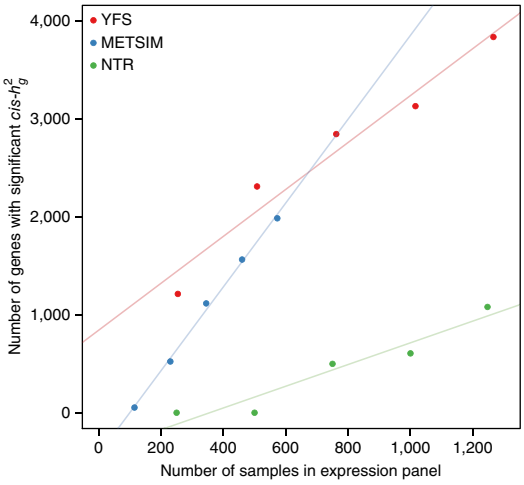

Figure 4 The number of genes with significant *cis* heritability observed at varying sample sizes. The number of genes with significant *cis* heritability was estimated by downsampling each data set (YFS, METSIM and NTR) into quintiles.

available imputation accuracy. This was further reflected in an analysis of real data, where merging expression data sets did not substantially change the distribution of TWAS statistics for the same gene set (**Supplementary Fig. 14**). Although these results come with caveats (for example, standard assumptions of additive effects and normal residuals), they suggest that the main benefit of larger expression reference panels is in increasing the total number of significant *cis*-heritable genes available for imputation (**Fig. 4**).

TWAS performance in GWAS summary data

To further validate our approach, we employed TWAS to identify expression-trait associations at the 697 known GWAS risk loci for height[7] using the YFS data for which height was also measured. At each locus, we considered three strategies for selecting a single causal gene: selection of (i) the gene nearest to the top GWAS SNP; (ii) the gene for which the index SNP was the strongest eQTL in the training data; and (iii) the most significant TWAS gene. For each strategy, we then constructed a risk score using the genetic value of expression for the selected genes and correlated the risk score with height measurements in the YFS individuals (an independent sample from the original height GWAS; **Supplementary Note**). R^2 between the risk score and height was 0.038 (nearest), 0.031 (eQTL) and 0.054 (TWAS), with the TWAS estimate significantly higher than the others in a joint model (Online Methods and **Supplementary Table 5**). When we recomputed the risk scores using TWAS values for expression from the NTR data (which introduces additional noise as a result of heterogeneity between the data sets), the TWAS estimate remained significantly higher than that from the eQTL strategy but was comparable to selecting the nearest gene (**Supplementary Table 5**). This indicates that using expression from a different study to select genes still significantly explains trait variance but is complementary (rather than superior) to selecting the nearest gene. Working from the assumption that genes with a higher *cis* genetic correlation to phenotype are more likely to be causal, these results motivate the use of TWAS to prioritize putative risk genes at known GWAS loci.

Across all known risk loci in our data, 77% of genome-wide significant loci (defined as lead SNP ±500 kb) overlapped at least one gene with significant *cis*-h^2_g and 36% overlapped at least one significant TWAS association (**Supplementary Table 6**). These results suggest that *cis* regulation of expression in blood and adipose tissue is an important mechanism through which genetic variation at known risk loci alters obesity-related traits. We expect that expression studies from other tissues relevant to obesity-related traits will further increase the overlap. Focusing specifically on the 282 TWAS-identified genes that were within 500 kb of the lead SNP, 187 (66%) were not the nearest gene, with many residing more than 100 kb away from the lead GWAS SNP (**Supplementary Fig. 15**). Because GWAS usually report the nearest gene, these 187 genes can be considered new candidates for follow-up at known risk loci. We note that gene-trait associations at known risk loci will not be found by TWAS if the causal mechanism does not involve *cis* expression of the tested genes or if there is insufficient power to identify and detect all *cis*-heritable genes at the locus.

Next, we employed TWAS to identify new expression-trait associations using summary association statistics from a 2010 lipid GWAS[17] (~100,000 samples), that is, associations more than 500 kb away from any genome-wide significant SNPs in that study. We used all three studies (METSIM, YFS and NTR) as separate SNP-expression training panels. We then looked for genome-wide significant SNPs at these loci in the larger 2013 lipid GWAS[5] (expanded to ~189,000 samples). We identified 25 such expression-trait associations in the 2010 study (**Supplementary Table 7**), of which 19 contained genome-wide significant SNPs in the 2013 study ($P = 1 \times 10^{-24}$ by hypergeometric test; Online Methods) and 24 contained a more significant SNP ($P = 1 \times 10^{-4}$), constituting a highly significant validation of the identified loci. The validation remained significant after conservatively accounting for sample overlap across the studies (binomial $P = 3 \times 10^{-16}$; Online Methods and **Supplementary Table 7**). As a sanity check, we compared

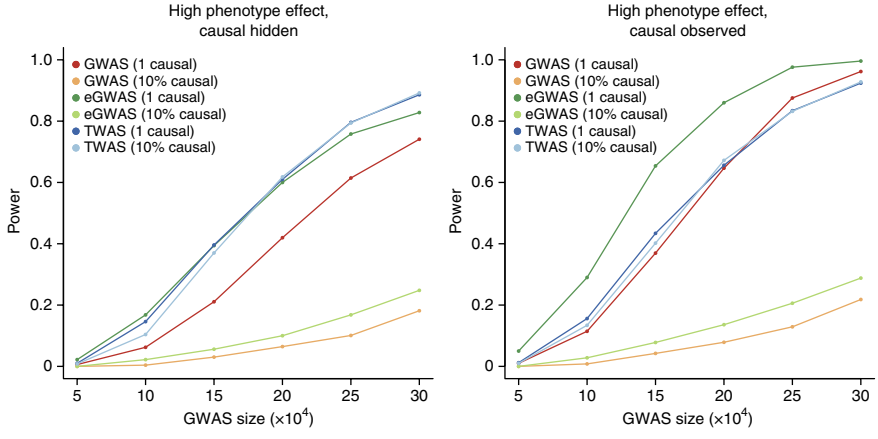

Figure 5 Power of summary-based expression imputation algorithms. Realistic disease architectures were simulated, and power to detect a genome-wide significant association was evaluated across three methods (accounting for 15,000 eGWAS or TWAS tests and 1 million GWAS tests). Colors correspond to the number of causal variants simulated and the methods used: GWAS where every SNP in the locus is tested; eGWAS where only the best *cis*-eQTL is tested; and TWAS computed using summary statistics. The expression reference panel was fixed at 1,000 out-of-sample individuals, and simulated GWAS sample size is designated on the *x* axis. Power was computed as the fraction of 500 simulations where significant association was identified.

direct and summary-level TWAS in the METSIM data and found the two sets of imputed expression-trait z scores to be nearly identical, with summary-level TWAS slightly underestimating the effect (Pearson $\rho = 0.96$; **Supplementary Fig. 16**). Overall, we find the TWAS approach to be highly predictive of robust phenotypic associations.

TWAS identifies new expression-trait associations

Having established the usefulness of TWAS, we applied the approach to identify new expression-trait associations using summary data from three recent GWAS over more than 900,000 phenotype measurements: lipid measures (high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol (TC) and triglycerides (TG))[5], height[7] and BMI[6]. Significantly *cis*-heritable genes across the three expression data sets were tested individually (6,924 tests) and together in an omnibus test that accounts for predictor correlation (1,075 tests; Online Methods), and we conservatively corrected for the 8,000 total tests performed for each trait. Overall, we identified 665 significant gene-trait associations (**Supplementary Table 8**). Of these, 69 gene-trait associations did not overlap a genome-wide significant SNP in the corresponding GWAS, residing in 60 physically non-overlapping *cis* loci (**Table 1** and **Supplementary Table 9**). Averaging over the new genes, the z^2 statistics from TWAS were 1.5× higher than the strongest eQTL SNP for the same gene (although this may be slightly inflated because of winner's curse). Our previous simulations suggest that the substantial gain over testing the *cis*-eQTL is an indication of pervasive allelic heterogeneity[40] at these loci, and analyses of expression showed strong evidence for allelic heterogeneity at the TWAS genes (**Supplementary Fig. 17**).

We further sought to quantify the significance of the expression-trait associations conditional on the SNP-trait effects at the locus with a permutation test (Online Methods). Comparing to this null assesses how much signal is added by the expression given the specific GWAS architecture of the locus. For the 69 genes, this permutation test was significant for 54 (after accounting for 69 tests). After excluding these individually significant genes, the P values were still substantially elevated with λ_{GC} of 19 (ratio of the median χ^2 value to the expected null). For these 54 genes, we can confidently conclude that integration of expression data significantly refined the association with the trait. As before, more evidence of allelic heterogeneity in expression was observed at the loci that passed permutation (**Supplementary Fig. 17**). Our results are consistent with a model of causality where these genes harbor inherited causal variants that modulate expression, which in turn has a complex effect on the cell and downstream impact on complex traits[6].

Next, we evaluated the contribution to heritability of all expression-trait associations, including those that were not genome-wide significant (Online Methods)[29,30]. We estimated the variance in a trait explained by all METSIM and YFS imputed genes (h^2_{GE}) to be 3.4% averaged over six traits (**Supplementary Table 10**). We assumed independence of the two data sets and did not include the NTR genes because of the strong correlation of these data with YFS. Height had the most variance attributable to heritable genes at $h^2_{GE} = 7.1\%$. These combined estimates were consistently higher than those from a corresponding analysis using predictions from permuted expression (**Supplementary Table 10**). For the four traits with individual-level genotype and phenotype data in METSIM (BMI, TG, waist-hip ratio (WHR) and fasting insulin levels (INS)), we estimated h^2_{GE} directly using variance components over the imputed expression values (Online Methods). On average, all significantly heritable genes in adipose and blood explained 4–6% of the trait variance (16–19% of the total trait h^2_g) and were largely orthogonal between the two predictions (**Supplementary Table 11**). The imputed expression consistently explained more trait variance than the best *cis*-eQTL in each gene and did not strongly depend on the size of the *cis* window (**Supplementary Table 12**).

Reevaluation using other expression data sets

To replicate the 69 new expression-trait associations, we reevaluated the GWAS summary statistics with expression data from two external studies: eQTLs from ~900 samples in the MuTHER study[25] of fat, lymphoblastoid cell line (LCL) and skin cells and separate eQTL data from 5,311 samples in whole blood[11] (Online Methods). These expression studies only consist of summary-level associations and are expected to be much noisier as reference. In the relatively smaller MuTHER sample, 20 of 55 available genes replicated significantly in at least one tissue (after accounting for 55 tests; **Supplementary Table 9**). This is substantial given the apparent heterogeneity between expression data sets we previously observed (Online Methods). Notably, the correlations between discovery and replication z scores were strongest for associations found in the corresponding tissue ($\rho = 0.60$, $P = 1.5 \times 10^{-5}$ for blood and LCLs; $\rho = 0.66$, $P = 0.05$ for adipose; **Supplementary Table 13**), constituting significant aggregate replication and further evidence for the tissue-specific nature of our findings. Using the larger but heterogeneous training sample from ref. 11, 24 of 37 available genes replicated significantly (**Supplementary Table 9**). Although these replications are not strictly independent (they use the same GWAS data), they demonstrate that many of the newly identified loci are consistently significant across diverse expression cohorts.

Functional analysis of the new associations

To better understand their functional consequences, we evaluated the 69 new genes in the Hybrid Mouse Diversity Panel (HMDP) for correlation with multiple obesity-related traits. This panel includes 100 inbred mice strains with an extensive collection of obesity-related phenotypes from ~12,000 genes. Of the 69 new TWAS genes previously identified, 40 were present in the panel and could be evaluated for effect on phenotype. Of these, 26 were significantly associated with at least one obesity-related trait (after accounting for genes tested) and 14 remained significant after accounting for 36 phenotypes tested (very conservatively assuming that the phenotypes were independent) (**Supplementary Table 14**). Of the genes, 77% with an association were associated with multiple phenotypes. For example, expression of *Ftsj3* was significantly correlated with fat mass, glucose-to-insulin ratio and body weight in both liver and adipose tissue, with R^2 estimates ranging from 0.20 to 0.28. Another candidate, *Iih4*, was significantly correlated with LDL cholesterol and TC levels in liver. In humans, the corresponding gene is also linked to hypercholesterolemia in Online Mendelian Inheritance in Man (OMIM) and was previously associated with BMI in East Asians[41]. Because of complex correlation of phenotypes, it is difficult to assess whether this gene set is significant in aggregate and genes in the HMDP are typically expected to have strong effects. We could not perform enough random selections of genes to establish significance for this set. However, we consider the 26 individually significant genes to be fruitful targets for follow-up studies.

The BMI and height GWAS evaluated functional enrichment at identified loci, and we performed similar analyses for the new genes that we identified. We tested the ten new BMI-associated genes and 33 new height-associated genes for tissue-specific enrichment using DEPICT[42], a method based on large-scale gene coexpression analyses, following the protocol of the original GWAS[6,7]. Analysis of BMI identified significant enrichment for hypothalamus and neurosecretory systems ($P = 2.6 \times 10^{-4}$, significant at false discovery rate <5%). This enrichment is consistent with the landmark finding in the original

Table 1 TWAS significant genes with no known GWAS risk variants within 500 kb

GWAS	Training expression	Gene	Chr.	Locus start	Locus end	*P* value			
						TWAS	Permuted	Best *cis*-SNP	
LOCKE.BMI	Omnibus	*INO80E*	16	29,506,615	30,517,114	3×10^{-9}	3×10^{-7}	3×10^{-6}	*
LOCKE.BMI	Omnibus	*FTSJ3*	17	61,396,793	62,407,372	1×10^{-7}	5×10^{-6}	9×10^{-7}	*
LOCKE.BMI	Omnibus	*PAM*	5	101,589,685	102,866,809	3×10^{-7}	4×10^{-9}	3×10^{-3}	*
LOCKE.BMI	YFS	*GGNBP2*	17	34,400,737	35,446,278	1×10^{-6}	7×10^{-5}	3×10^{-6}	*
LOCKE.BMI	YFS	*MYO19*	17	34,351,477	35,399,284	3×10^{-6}	4×10^{-5}	3×10^{-6}	*
LOCKE.BMI	Omnibus	*OPRL1*	20	62,211,526	63,231,996	3×10^{-6}	9×10^{-5}	2×10^{-5}	*
LOCKE.BMI	NTR	*RABGAP1*	9	125,203,287	126,367,145	4×10^{-6}	3×10^{-5}	1×10^{-5}	*
LOCKE.BMI	YFS	*SMARCD2*	17	61,409,444	62,420,425	5×10^{-6}	9×10^{-5}	9×10^{-7}	*
LOCKE.BMI	METSIM	*AL049840.1*	14	103,677,607	104,679,149	5×10^{-6}	5×10^{-5}	1×10^{-7}	*
LOCKE.BMI	Omnibus	*LPAR2*	19	19,234,477	20,239,739	6×10^{-6}	3×10^{-5}	4×10^{-6}	*
WILLER.HDL	YFS	*MRPS18B*	6	30,085,486	31,094,172	1×10^{-7}	2×10^{-2}	4×10^{-6}	
WILLER.LDL	Omnibus	*PAM*	5	101,589,685	102,866,809	4×10^{-15}	9×10^{-12}	2×10^{-3}	*
WILLER.LDL	Omnibus	*ITIH4*	3	52,346,991	53,365,495	8×10^{-9}	4×10^{-6}	3×10^{-5}	*
WILLER.LDL	Omnibus	*WARS*	14	100,300,125	101,343,142	1×10^{-8}	6×10^{-7}	3×10^{-5}	*
WILLER.LDL	Omnibus	*MAN2C1*	15	75,148,133	76,160,971	1×10^{-8}	1×10^{-5}	6×10^{-5}	*
WILLER.LDL	YFS	*DHRS13*	17	26,724,799	27,730,089	6×10^{-7}	4×10^{-4}	2×10^{-6}	*
WILLER.LDL	YFS	*ERAL1*	17	26,681,956	27,688,085	8×10^{-7}	9×10^{-4}	2×10^{-6}	
WILLER.LDL	YFS	*HCG27*	6	30,665,537	31,671,745	2×10^{-6}	7×10^{-3}	7×10^{-8}	
WILLER.LDL	YFS	*VARS2*	6	30,376,019	31,394,236	3×10^{-6}	2×10^{-2}	1×10^{-5}	
WILLER.LDL	Omnibus	*PEX6*	6	42,431,608	43,446,958	5×10^{-6}	3×10^{-5}	4×10^{-4}	*
WILLER.LDL	Omnibus	*CSK*	15	74,574,398	75,595,539	6×10^{-6}	1×10^{-4}	1×10^{-5}	*
WILLER.TC	Omnibus	*PAM*	5	101,589,685	102,866,809	9×10^{-15}	3×10^{-13}	5×10^{-3}	*
WILLER.TC	Omnibus	*WARS*	14	100,300,125	101,343,142	2×10^{-8}	4×10^{-6}	2×10^{-5}	*
WILLER.TC	Omnibus	*MAN2C1*	15	75,148,133	76,160,971	3×10^{-7}	7×10^{-5}	2×10^{-6}	*
WILLER.TC	Omnibus	*ITIH4*	3	52,346,991	53,365,495	6×10^{-7}	2×10^{-5}	5×10^{-5}	*
WILLER.TC	NTR	*CDK2AP1*	12	123,245,552	124,256,687	6×10^{-7}	2×10^{-4}	5×10^{-6}	*
WILLER.TC	YFS	*TBKBP1*	17	45,271,447	46,289,416	9×10^{-7}	3×10^{-4}	2×10^{-7}	*
WILLER.TC	METSIM	*RPP25*	15	74,746,757	75,749,805	2×10^{-6}	2×10^{-4}	9×10^{-7}	*
WILLER.TC	Omnibus	*CSK*	15	74,574,398	75,595,539	2×10^{-6}	9×10^{-5}	9×10^{-7}	*
WILLER.TC	YFS	*MPI*	15	74,682,346	75,691,798	2×10^{-6}	2×10^{-4}	5×10^{-6}	*
WILLER.TC	Omnibus	*DAGLB*	7	5,948,757	7,023,821	2×10^{-6}	8×10^{-5}	7×10^{-6}	*
WILLER.TC	NTR	*TOM1*	22	35,195,267	36,243,985	2×10^{-6}	2×10^{-6}	1×10^{-6}	*
WILLER.TC	METSIM	*HMGXB4*	22	35,153,445	36,191,800	3×10^{-6}	4×10^{-5}	4×10^{-7}	*
WILLER.TC	NTR	*C17orf68*	17	7,628,139	8,651,413	3×10^{-6}	6×10^{-6}	2×10^{-7}	*
WILLER.TG	YFS	*PABPC4*	1	39,526,488	40,542,462	1×10^{-8}	1×10^{-4}	8×10^{-8}	*
WILLER.TG	Omnibus	*PACS1*	11	65,337,834	66,512,218	5×10^{-8}	3×10^{-4}	1×10^{-5}	*
WOOD.HEIGHT	Omnibus	*INO80E*	16	29,506,615	30,517,114	2×10^{-10}	2×10^{-5}	1×10^{-7}	*
WOOD.HEIGHT	NTR	*INPP5B*	1	37,825,435	38,912,729	2×10^{-9}	1×10^{-6}	1×10^{-6}	*
WOOD.HEIGHT	Omnibus	*MEGF9*	9	122,863,091	123,976,748	3×10^{-9}	2×10^{-4}	1×10^{-7}	*
WOOD.HEIGHT	Omnibus	*ATF1*	12	50,657,493	51,714,905	6×10^{-9}	4×10^{-5}	1×10^{-6}	*
WOOD.HEIGHT	Omnibus	*PAM*	5	101,589,685	102,866,809	2×10^{-8}	8×10^{-6}	1×10^{-5}	*
WOOD.HEIGHT	Omnibus	*CNIH4*	1	224,044,552	225,067,161	3×10^{-8}	2×10^{-5}	6×10^{-8}	*
WOOD.HEIGHT	Omnibus	*PLEKHA1*	10	123,634,212	124,691,867	1×10^{-7}	3×10^{-5}	6×10^{-7}	*
WOOD.HEIGHT	NTR	*PDXDC1*	16	14,568,832	15,632,186	1×10^{-7}	3×10^{-3}	7×10^{-6}	
WOOD.HEIGHT	YFS	*MSRB2*	10	22,884,435	23,910,942	2×10^{-7}	2×10^{-4}	3×10^{-6}	*
WOOD.HEIGHT	YFS	*ZNF213*	16	2,679,778	3,692,806	2×10^{-7}	1×10^{-3}	6×10^{-7}	
WOOD.HEIGHT	NTR	*YWHAB*	20	43,014,185	44,037,354	5×10^{-7}	3×10^{-4}	4×10^{-6}	*
WOOD.HEIGHT	NTR	*ITM2B*	13	48,307,273	49,336,451	5×10^{-7}	4×10^{-5}	1×10^{-6}	*
WOOD.HEIGHT	Omnibus	*WDSUB1*	2	159,592,304	160,643,310	7×10^{-7}	3×10^{-4}	5×10^{-6}	*
WOOD.HEIGHT	NTR	*STAT6*	12	56,989,190	58,005,129	8×10^{-7}	3×10^{-3}	8×10^{-6}	
WOOD.HEIGHT	Omnibus	*PLCL1*	2	198,169,426	199,937,305	9×10^{-7}	4×10^{-3}	6×10^{-7}	
WOOD.HEIGHT	YFS	*H2AFJ*	12	14,427,270	15,430,936	1×10^{-6}	4×10^{-4}	6×10^{-7}	*
WOOD.HEIGHT	YFS	*FAM8A1*	6	17,100,586	18,111,950	1×10^{-6}	2×10^{-3}	1×10^{-7}	
WOOD.HEIGHT	METSIM	*AC016995.3*	2	38,133,861	39,242,882	1×10^{-6}	3×10^{-4}	4×10^{-5}	*
WOOD.HEIGHT	Omnibus	*CDA*	1	20,415,441	21,445,401	1×10^{-6}	3×10^{-4}	6×10^{-7}	*
WOOD.HEIGHT	YFS	*ECHDC2*	1	52,861,656	53,892,884	1×10^{-6}	3×10^{-3}	1×10^{-5}	
WOOD.HEIGHT	YFS	*NFATC3*	16	67,618,654	68,763,162	2×10^{-6}	4×10^{-3}	4×10^{-7}	
WOOD.HEIGHT	YFS	*SH3YL1*	2	−282,270	766,398	2×10^{-6}	1×10^{-4}	6×10^{-7}	*
WOOD.HEIGHT	Omnibus	*PABPC4*	1	39,526,488	40,542,462	3×10^{-6}	4×10^{-4}	6×10^{-6}	*

(continued)

Table 1 (continued)

GWAS	Training expression	Gene	Chr.	Locus start	Locus end	*P* value			
						TWAS	Permuted	Best *cis*-SNP	
WOOD.HEIGHT	METSIM	*RP11-473M20.14*	16	2,666,043	3,684,883	3×10^{-6}	4×10^{-4}	1×10^{-7}	*
WOOD.HEIGHT	Omnibus	*HEBP1*	12	12,627,798	13,653,207	3×10^{-6}	7×10^{-4}	1×10^{-6}	*
WOOD.HEIGHT	YFS	*KBTBD2*	7	32,407,784	33,433,743	3×10^{-6}	4×10^{-3}	1×10^{-7}	
WOOD.HEIGHT	METSIM	*LRRC69*	8	91,614,060	92,731,464	3×10^{-6}	6×10^{-4}	6×10^{-7}	*
WOOD.HEIGHT	YFS	*RAB23*	6	56,553,607	57,587,078	4×10^{-6}	4×10^{-3}	6×10^{-7}	
WOOD.HEIGHT	YFS	*PPP4C*	16	29,587,299	30,596,698	5×10^{-6}	1×10^{-3}	1×10^{-7}	
WOOD.HEIGHT	NTR	*B3GALNT2*	1	235,110,442	236,167,884	5×10^{-6}	5×10^{-4}	1×10^{-6}	*
WOOD.HEIGHT	YFS	*PSRC1*	1	109,322,178	110,325,808	5×10^{-6}	3×10^{-4}	1×10^{-7}	*
WOOD.HEIGHT	YFS	*ACSS1*	20	24,486,868	25,539,616	5×10^{-6}	6×10^{-4}	1×10^{-6}	*
WOOD.HEIGHT	Omnibus	*GGPS1*	1	234,990,665	236,007,847	5×10^{-6}	4×10^{-4}	3×10^{-7}	*

*Significant after permutation. WILLER, LOCKE and WOOD correspond to GWAS data from refs. 5–7, respectively.

study[6] showing enrichment in these and other central nervous system tissues. Notably, we recapitulated this result using only new genes that did not overlap any genome-wide significant SNPs. In the analysis of height, DEPICT did not identify any tissue-specific enrichment.

DISCUSSION

In this work, we present methods that integrate genetic and transcriptional variation to identify genes with expression associated with complex traits. Using imputed gene expression to guide GWAS has three potential advantages. First, the gene is a more interpretable biological unit than an associated locus, which often contains multiple significant SNPs in LD that may not lie in genes and/or tag variants in multiple genes. Second, the lower total number of genes (or *cis*-heritable genes) means that the multiple-testing burden is substantially reduced relative to all SNPs. Lastly, combining *cis*-SNPs into a single predictor may capture heterogeneous signal better than individual SNPs or *cis*-eQTLs. Focusing prediction on the genetic component of expression also avoids confounding from environmental differences caused by the trait that may influence expression. Our approach builds upon the wealth of GWAS data in massive cohorts to directly implicate the gene-based mechanisms underlying complex traits.

Our proposed method has conceptual similarities with two-sample Mendelian randomization approaches that aim to identify causal relations between traits using genetic variation predictions as a randomizer[43–45]. However, whereas Mendelian randomization is intended to quantify the total causal effect, our method has the less strict goal of identifying significant associations and can operate on summary GWAS data. Notably, our approach maintains the attractive feature of not being confounded by effects on expression and a trait that are independent of the SNPs. Other recent work has proposed to leverage summary statistics to estimate the underlying genetic correlation between traits at the genome-wide level[37] but cannot be applied locally as it requires multiple loci to estimate standard errors (Online Methods). Recent work in parallel to ours also proposes gene expression imputation from individual-level data to find expression-trait associations and observes benefits from a reduced multiple-testing burden and increased interpretability[16]. In contrast, our approach does not require individual-level GWAS data and is applicable directly to GWAS summary data of very large sample sizes, thus increasing discovery power.

Unlike current methods, which focus on individually significant eQTL and SNP associations[5,6,8,9,11,13,26,29], our approach captures the full *cis*-SNP signal and does not require any individual marker to be significant. This is underscored by the fact that TWAS substantially outperformed its *cis*-eQTL analog both in imputing expression and in association with a trait. Our results show that the imputation approach is especially effective when multiple variants influence expression (which in turn influences a trait). The large number of new associations we identified in real data supports this phenomenon and suggests that it may be a strong contributor to common phenotypes[46]. Therefore, our approach can be seen as complementary to GWAS by identifying expression-trait associations that are not well explained by individual tagging SNPs. Future work could leverage the difference in performance of TWAS and GWAS to explicitly detect allelic heterogeneity. We note that it is still possible for some loci to have an independent SNP-phenotype and SNP-expression association driven by the same underlying variant, although we consider this to be an infrequent biological model.

We conclude with several limitations of our approach. First, variants influencing disease that are independent of *cis* expression—in general or in the training data—will not be identified. Second, as with any prediction, the number of genes that can be accurately imputed is still limited by the training cohort size and the quality of the training data. In particular, we found that prediction accuracy did not correspond with theoretical expectations and is likely driven by data quality. The impact of these weaknesses could be better quantified as expression data from larger sample sizes and a more diverse set of tissues become available. Although in this work we used both microarray and RNA sequencing as a measure of gene expression, thus showcasing the applicability of our approach to diverse data sets, the accuracy of our method intrinsically depends on the quality of the expression measurements. For the associated genes, it remains possible that the effect is actually mediated by phenotype (SNP → phenotype → *cis* expression; scenario F in **Fig. 2**). We attempted to quantify this in the YFS data by conditioning the heritability analyses on all the evaluated phenotypes (height, BMI and lipid concentrations) but observed no significant change at individual genes or in the mean *cis*-h_g^2. These results suggest that confounding from phenotype does not substantially affect the tested *cis* expression, although at the current sample size we cannot completely rule out such confounders for individual genes. An alternative confounder arises from independent effects on phenotype and expression at the same SNP or tag (**Fig. 2g** and Online Methods). Such instances could be indistinguishable from the desired causal model (Online Methods) without analyzing individual-level data, although we believe that they are still biologically interesting cases of colocalization. Both types of confounding could potentially be quantified by training the SNP-expression relationships in control individuals where phenotype is fixed or by interrogating the gene experimentally. Lastly, the summary-based TWAS cannot account for rare variants that are poorly captured by the LD reference panel or optimally capture nonlinear relationships between SNPs and expression. Additional sources of information could

potentially be incorporated to improve prediction, including significant *trans* associations[11,28], allele-specific expression[47,48], splice-QTLs affecting individual exons[10], haplotype effects and SNP-specific functional priors[20,49–51].

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The predictor weights computed from the three expression studies as well as software to perform individual- and summary-level prediction are available at http://bogdan.bioinformatics.ucla.edu/software/twas/.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the individuals who participated in the study. We also acknowledge L. Yang for helpful discussions that have improved the quality of this manuscript. We also thank K. Mohlke, M. Boehnke and F. Collins for help with the METSIM data. This work was funded in part by US National Institutes of Health (NIH) grants F32 GM106584 (A.G.), R01 GM053725 (B.P.), R01 GM105857 (A.L.P., A.G. and G.B.), HL-28481 (P.P., A.J.L. and M.C.) and HL-095056 (P.P. and B.P.) and by the US NIH training grant in Genomic Analysis and Interpretation T32 HG002536 (A.K.).

AUTHOR CONTRIBUTIONS

A.G. and B.P. conceived and designed the experiments. A.G., A.K. and H.S. performed the experiments and analyzed the data. G.B., W.C., B.W.J.H.P., R.J., E.J.C.d.G., D.I.B., F.A.W., P.F.S., E.N., M.A., M.C., A.J.L., T.L., E.R., M.K., I.S., O.T.R., J.K. and M.L. generated data, reagents, materials and analysis tools. A.G., A.L.P., P.P. and B.P. wrote the manuscript. All authors reviewed, revised and wrote feedback for the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
2. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375, S1–S3 (2012).
3. Lee, D., Bigdeli, T.B., Riley, B.P., Fanous, A.H. & Bacanu, S.A. DIST: direct imputation of summary statistics for unmeasured SNPs. *Bioinformatics* **29**, 2925–2927 (2013).
4. Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30**, 2906–2914 (2014).
5. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
6. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
7. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
8. Musunuru, K. *et al.* From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
9. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
10. Zhang, X. *et al.* Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.* **47**, 345–352 (2015).
11. Westra, H.J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
12. Albert, F.W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
13. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).
14. Letourneau, A. *et al.* Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature* **508**, 345–350 (2014).
15. Davis, L.K. *et al.* Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genet.* **9**, e1003864 (2013).
16. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
17. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
18. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
19. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
20. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
21. Wray, N.R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
22. Nuotio, J. *et al.* Cardiovascular risk factors in 2011 and secular trends since 2007: the Cardiovascular Risk in Young Finns Study. *Scand. J. Public Health* **42**, 563–571 (2014).
23. Raitakari, O.T. *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* **37**, 1220–1226 (2008).
24. Wright, F.A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–437 (2014).
25. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
26. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
27. Torres, J.M. *et al.* Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *Am. J. Hum. Genet.* **95**, 521–534 (2014).
28. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015).
29. Nica, A.C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
30. Robinson, G.K. That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* **6**, 15–32 (1991).
31. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
32. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
33. Chatterjee, N. *et al.* Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.* **45**, 400–405, e1–e3 (2013).
34. Brown, C.D., Mangravite, L.M. & Engelhardt, B.E. Integrative modeling of eQTLs and *cis*-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* **9**, e1003649 (2013).
35. Wen, X., Luca, F. & Pique-Regi, R. Cross-population joint analysis of eQTLs: fine mapping and functional annotation. *PLoS Genet.* **11**, e1005176 (2015).
36. Wood, A.R. *et al.* Another explanation for apparent epistasis. *Nature* **514**, E3–E5 (2014).
37. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
38. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
39. Lee, D. *et al.* JEPEG: a summary statistics based tool for gene-level joint testing of functional variants. *Bioinformatics* **31**, 1176–1182 (2015).
40. Pritchard, J.K. & Cox, N.J. The allelic architecture of human disease genes: common disease–common variant...or not? *Hum. Mol. Genet.* **11**, 2417–2423 (2002).
41. Wen, W. *et al.* Meta-analysis of genome-wide association studies in East Asian–ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* **23**, 5492–5504 (2014).
42. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
43. Smith, G.D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
44. Pickrell, J. Fulfilling the promise of Mendelian randomization. *bioRxiv* doi:10.1101/018150 (16 April 2015).
45. Pierce, B.L. & Burgess, S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* **178**, 1177–1184 (2013).
46. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLoS Genet.* **9**, e1003993 (2013).
47. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* **325**, 1246–1250 (2009).
48. Montgomery, S.B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
49. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
50. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
51. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

ONLINE METHODS

Data sets. In this study, we included 11,484 participants from two Finnish population cohorts, the Metabolic Syndrome in Men (METSIM; $n = 10{,}197$)[52,53] and the Young Finns Study (YFS; $n = 1{,}414$)[22,23]. 1,400 randomly selected individuals from the 10,197 METSIM participants underwent a subcutaneous abdominal adipose biopsy, of which 600 RNA samples were analyzed using RNA-seq (**Supplementary Note**). BMI, TG, WHR and INS were inverse rank transformed and adjusted for age and age^2. INS was additionally adjusted for T1D and T2D. 1,414 individuals (638 men with a median age of 43 years and 776 women with a median age of 43) with gene expression, phenotype and genotype data available were included in the blood expression analysis. Height, BMI, TG, TC, HDL and LDL were inverse rank transformed and adjusted for age, age^2 and sex. TC was also adjusted for statin intake. The biochemical lipid, glucose and other clinical and metabolic measurements of METSIM and YFS were performed as described previously[22,52,54]. Blood expression array data from the Netherlands Twins Registry (NTR; $n = 1{,}247$)[24,55] was processed as described in the original paper, followed by removal of any individuals with GRM values >0.05. Complete details on the pipeline and quality control procedures can be found in the **Supplementary Note**.

Heritability estimation with individual data. *Cis* and *trans* variance components were estimated using the REML algorithm implemented in GCTA[19]. As in previous studies, estimates were allowed to converge outside the expected 0–1 bound on variance to achieve unbiased mean estimates across all genes[24]. Standard error across gene sets was estimated by dividing the observed standard deviation by the square root of the number of genes that converged (this will lead to underestimation because of correlated genes but is presented for completeness). Genome-wide h^2_g values for the four traits in the GWAS cohort were estimated with GCTA from a single relatedness matrix constructed over all post–quality control SNPs in the strictly unrelated individuals. For estimating expression-wide h^2_{GE}, each predicted expression value was standardized to mean = 0 and variance = 1, and sample covariance across these values was used to define the relatedness matrix. The h^2_{GE} value was then estimated from this component with GCTA, with P values for difference from zero computed using a the likelihood-ratio test. Twenty principal components were always included as fixed effects to account for ancestry. Genetic correlation between traits in the GWAS cohort was estimated from all post–quality control SNPs in the full set of 10,000 individuals with GEMMA[31] (**Supplementary Table 15**). For YFS, we quantified the mediating effects of a trait on *cis* expression by separately re-estimating *cis*-h^2_g with all analyzed traits (height, BMI, TC, TG, HDL cholesterol and LDL cholesterol) included as fixed effects in addition to principal components. We did not observe significant differences in any individual gene (after accounting for 3,836 genes tested) nor in the mean estimate of *cis*-h^2_g.

Heritability estimation with summary data. As shown previously[51,56], for an association study of n independent samples, the expected χ^2 statistic is $E[\chi^2] = 1 + nlh^2_{GE}/m$, where l is the LD score accounting for correlation, m is the number of markers and h^2_{GE} is the variance in the trait explained by imputed expression. We estimated l directly from the genetic values of expression to be close to independence (1.4 and 1.5 for METSIM and YFS, respectively), allowing us to solve for h^2_{GE} from the observed distribution of χ^2 (or, asymptotically equivalent z^2) statistics. We did not compute this value for the BMI GWAS because the conservative multiple-GC correction applied in that study would yield a severe downward bias[6].

Imputing expression into genotyped samples. We evaluated three prediction schemes: (i) *cis*-eQTL, where the single most significantly associated SNP in the training set was used as the predictor; (ii) BLUP[30], which estimates the causal effect sizes of all SNPs in the locus jointly using a single variance component; and (iii) BSLMM[31], which estimates the underlying effect size distribution and then fits all SNPs in the locus jointly. For BLUP and BSLMM, prediction was done over all post–quality control SNPs using GEMMA[31]. We note that BLUP and BSLMM both perform shrinkage of the SNP weights but not variable selection, so all SNPs are included in the predictor. Recent work in parallel to ours also evaluated expression imputation using polygenic risk scores, LASSO and elastic net[16].

Evaluating prediction accuracy. Within-study prediction accuracy was measured by fivefold cross-validation in a random sampling of 1,000 of the highly heritable genes (genes with significant nonzero *cis* heritability) for each study. Cross-study prediction accuracy was measured by merging the YFS and NTR genotyped individuals and predicting from all individuals in one cohort into all individuals in the other cohort. In all instances, the R^2 between predicted and true expression across all predicted folds was used to evaluate accuracy (**Supplementary Fig. 18** and **Supplementary Note**).

Imputing expression into GWAS summary statistics. Summary-based imputation was performed using the ImpG-Summary algorithm[4] extended to train on the *cis* genetic component of expression. Let Z be a vector of standardized effect sizes (z scores) of SNP for a trait at a given *cis* locus (Wald statistics $\beta/\mathrm{se}(\beta)$). We impute the z score of the expression and trait as a linear combination of elements of Z with weights W (these weights are precompiled from the reference panel as $\Sigma_{e,s}\Sigma^{-1}_{s,s}$ for ImpG-Summary or directly from BSLMM). $\Sigma_{e,s}$ is the covariance matrix between all SNPs at the locus and gene expression, and $\Sigma_{s,s}$ is the covariance among all SNPs (LD). Under null data (no association) and a multivariate normal assumption, $Z \sim N(0,\Sigma_{s,s})$. It follows that the imputed z score of expression and trait (WZ) has variance $W\Sigma_{s,s}W^t$; therefore, we use $WZ/(W\Sigma_{s,s}W^t)^{1/2}$ as the imputation Z score of the *cis* genetic effect on the trait. In practice, for each gene, all SNPs within 1 Mb of the gene present in the GWAS were selected, and $\Sigma_{s,s}$ and $\Sigma_{e,s}$ and were computed in the reference panel (expression and SNP data). To account for finite sample size and instances where $\Sigma_{s,s}$ was not invertible, we adjusted the diagonal of the matrix using a technique similar to ridge regression with $\lambda = 0.1$ (as evaluated in Pasaniuc *et al.*[4]). This regularization, as well as noise in the estimation of W, can translate to lower power for association but yield conservative imputed Z statistics.

We used the YFS samples that were assayed for SNPs, phenotype and expression to assess the consistency of individual-level and summary-based TWAS. We first computed GWAS association statistics between phenotype (height) and SNPs and used them in conjunction with the expression data to impute summary-based TWAS statistics. The TWAS statistics were compared to those from the simple regression of (height ~expression) in the YFS data. We observed a correlation of 0.415 (**Supplementary Fig. 5**), consistent with an average *cis*-h^2_g of 0.17 ($\approx 0.415^2$) observed for these genes. When restricting to a regression of (height ~*cis* component of expression), we observed a correlation of 0.998 to the summary-based TWAS, demonstrating the equivalence of the two approaches when using in-sample LD.

Power analysis of the summary-based method. Simulations to evaluate the summary-based method were performed in 6,000 unrelated METSIM GWAS individuals. One hundred genes and the SNPs in the surrounding 1 Mb were randomly selected for testing. For each gene, normally distributed gene expression was simulated as $E = \boldsymbol{X}\beta + \varepsilon$, where $\boldsymbol{X}$ is a matrix of the desired number of causal genotypes, sampled randomly from the locus; β is a vector of normally distributed effect sizes for each causal variant; and ε is a vector of normally distributed noise to achieve a *cis*-h^2_g value of 0.17 (corresponding to the mean observed in our significant gene sets). One thousand individuals with SNPs and simulated expression were then withheld for training the predictors (**Supplementary Fig. 19**). For the remaining 5,000 individuals, normally distributed noise was applied to expression to generate a heritable phenotype where expression explained 0.10/180 or 0.20/180 of the phenotypic variance (with the former corresponding to the average effect sizes for associated genes observed in a large GWAS of height[57] and the latter corresponding to high-effect loci). Association between SNPs and a phenotype was estimated in the 5,000 individuals (standard Z score), and phenotype generation was repeated with different environmental noise (up to 60 iterations) to generate results from multiple GWAS substudies. Association statistics from each run were then subjected to meta-analysis to reach precision corresponding to a larger GWAS of the desired size (up to 300,000) (**Supplementary Note**).

Detecting a locus was defined as follows. The single most significant trait associated SNP was taken as the GWAS association, considered detected if GWAS significance was $< 5 \times 10^{-8}$. The single most significant eQTL in the training set was taken as the eQTL-guided association (eGWAS), and considered detected if GWAS significance was <0.05/15,000. The TWAS association

was measured by training the imputation algorithm on the 1,000 held-out samples with expression and imputing into the GWAS summary statistics, and considered detected if significance was <0.05/15,000. The entire procedure was repeated 500 times (5 per gene) and power was estimated by counting the fraction of instances where each method detected the locus. As in the cross-validation analysis, training on the genetic component of expression instead of the overall expression consistently increased TWAS power by ~10% (**Supplementary Fig. 7**). Two null expression models were tested by generating gene expression for the 1,000 held-out samples that was standard normal as well as heritable expression (*cis*-$h^2_g = 0.17$) with GWAS *Z* scores drawn from the standard normal (**Supplementary Table 4**). See the **Supplementary Note** for detailed simulation setup.

Power comparison to COLOC. COLOC uses summary data from eQTL and GWAS studies and a Bayesian framework to identify the subset of GWAS signals that co-localize with eQTLs. We sought to compare TWAS to the COLOC-estimated posterior probability of association (PPA) being shared for both phenotypes (PP4 in the COLOC implementation). COLOC additionally evaluates the hypothesis of multiple independent associations (PP3), but this is more general than the proposed TWAS model and was not tested. Because COLOC relies on priors of association to produce posterior probabilities of co-localization, we sought to identify a significance threshold that would make a fair comparison to the TWAS *P* value–based threshold. Specifically, we ran both methods on a realistic null expression simulation (with the generative model described previously): the expression was sampled from a null standard normal for 1,000 individuals and eQTLs computed; the trait associations were derived from a simulated 300,000 GWAS with a single typed causal variant that explained 0.001 variance of the trait (high effect). We believe this scenario is both realistic and consistent with the GWAS assumptions of COLOC. We then empirically identified the statistical threshold for COLOC and TWAS that would yield a 5% false discovery rate: co-localization statistic PP4 > 0.17 for COLOC, and $P < 0.05$ for TWAS. We note that this empirical COLOC threshold is much less stringent than PP4 > 0.8 used in the COLOC paper (PP4 > 0.8 would yield lower power for COLOC in our simulations). These thresholds were subsequently used to evaluate the power to detect an expression-trait association in simulations with a true effect (**Supplementary Figs. 10** and **12**). The reported power is for a single locus, and we did not attempt to quantify genome/transcriptome-wide significance.

Individual-level analysis of METSIM GWAS. We imputed the significantly heritable genes into the METSIM GWAS cohort of 5,500 unrelated individuals with individual-level genotypes (and unmeasured expression). We then tested the imputed expression for obesity-related traits: body mass index (BMI), triglycerides (TG), waist-hip-ratio (WHR) and fasting insulin levels (INS). Overall, the evaluated traits exhibited high phenotypic and genetic correlation as well as highly significant genome-wide h^2_g ranging from 23–36% (**Supplementary Table 15**), consistent with common variants having a major contribution to disease risk[1]. Association was assessed using standard regression as well as a mixed-model that accounted for relatedness and phenotypic correlation[31], with similar results. The effective number of tests for each trait was estimated by permuting the phenotypes 10,000 times and, for each permutation, re-running the association analysis on all predicted genes. For each trait, P_{perm}, the *P*-value in the lowest 0.05 of the distribution, was computed and the effective number of tests was $0.05/P_{perm}$ (reported in **Supplementary Tables 16** and **17**). All phenotypes were shuffled together, so any phenotypic correlation was preserved. The effective number of tests corresponded to 88–95% of the total number of genes, indicating a small amount of statistical redundancy (**Supplementary Note**). To evaluate the TWAS approach, we computed phenotype association statistics for the 5,500 unrelated individuals and re-ran the analysis using only these summary statistics and the same expression reference panels. The resulting TWAS associations were nearly identical to the direct TWAS associations across the four traits (Pearson $\rho = 0.96$). Reassuringly, the TWAS was generally more conservative than the direct estimates (**Supplementary Fig. 16**).

Refining trait-associated genes at known loci. We focused on GWAS data from height[7] that identified 697 genome-wide significant variants in 423 loci, and conducted the summary-based TWAS over all genes in these loci using YFS and NTR as expression training data. Because the YFS individuals had been phenotyped for height and not tested in the GWAS, we could directly evaluate whether selected genes were associated with phenotype. At each locus, we considered three strategies for selecting a single causal gene: (i) the gene nearest to the most significantly associated SNP; (ii) the gene for which the index SNP is the strongest eQTL in the training data; (iii) the most significant TWAS gene. For each strategy, we then constructed a risk-score using the genetic value of expression for the selected gene weighted by the corresponding TWAS *Z*-score. The same procedure was then re-evaluated using TWAS values trained in the NTR cohort (which introduces additional noise due to heterogeneity between the cohorts, **Supplementary Table 5**). We separately used GCTA to estimate the heritability of height explained by all of the genes selected by each algorithm by constructing a GRM from the selected genes. In contrast to the risk score, this does not assume predefined weights on each gene but allows them to be fit by the REML model. Results were comparable, with only the TWAS-selected genes explaining significantly nonzero heritability (**Supplementary Table 5**).

Validation analysis in lipid GWAS data. We evaluated the performance of TWAS by identifying significantly associated genes in the 2010 lipid study that did not overlap a genome-wide significant SNP, and looking for newly genome-wide significant SNPs in the expanded 2013 study. The *P*-value for the number of genes with increased significance and genome-wide significance in the 2013 study was computed by a hypergeometric test, with background probabilities estimated from the set of significantly heritable genes. Of the genes not overlapping a significant locus in the 2010 study, 70% had a more significant SNP in the 2013 study, and 3.5% overlapped a genome-wide significant SNP ($P < 5 \times 10^{-8}$).

Meta-analysis of imputed expression from multiple tissues. We proposed a novel omnibus test for significant association across predictions from all three cohorts. Because the imputation is made into the same GWAS cohort, correlation between predictors must be accounted for. For each gene *i*, we estimated a correlation matrix $\mathbf{C}_i$ by predicting from the three tissues into the ~5,500 unrelated METSIM GWAS individuals (though any large panel from the study population could be used). This correlation includes both the genetic correlation of expression as well as any correlated error in the predictors, thus capturing all redundancy. On average, a correlation of 0.01, 0.01 and 0.43 was observed between YFS:METSIM, NTR:METSIM, and YFS:NTR, highlighting the same tissue of origin the last pair. We then used the three-entry vector of TWAS predictions, Z_i, to compute the statistic $\text{omnibus}_i = Z_i' \, \mathbf{C}_i^{-1} \, Z_i$ which is approximately χ^2 (3-dof) distributed and provides an omnibus test for effect in any tissue while accounting for correlation[58,59]. Though the correlation observed in our data was almost entirely driven by the YFS:NTR blood data sets, we expect this to be an especially useful strategy for future studies with many correlated tissues. An alternative approach would be to perform traditional meta-analysis across the three cohorts and then predict the TWAS effect. However, this would lose power when true eQTL effect sizes (or LD) differ across the cohorts, which we have empirically observed to be the case by looking at predictor correlations. The proposed omnibus test aggregates different effects across the studies, at the cost of additional degrees of freedom.

Gene permutation test. The standard TWAS *Z* score is a test against the null of no SNP-trait association; that is, $Z_{\text{TWAS}} = WZ/(W\Sigma_{s,s}W^t)^{1/2}$ is well calibrated (has a mean of 0 and unit variance) only under the null model of $Z \sim N(0,\Sigma_{s,s})$. In the alternate model where *Z* is drawn from a nonzero mean distribution[60,61], Z_{TWAS} has a distribution that depends both on *Z* and the weights *W*. To quantify the impact of the weights on Z_{TWAS} regardless of whether *Z* is null or non-null, we conduct permutations conditional on the observed *Z* vector. For each gene, the expression labels were randomly shuffled, and the summary-based TWAS analysis was trained on the resulting expression to compute a permuted new null for Z_{TWAS}. Testing against this permuted null distribution is equivalent to testing for an expression-trait association (or genetic correlation between expression and a trait) conditional on the observed GWAS statistics at the locus (which may not be drawn from the null of no association).

doi:10.1038/ng.3506

The permutation test empirically computes this distribution of Z_{TWAS} values conditional on the observed Z and asks how extreme the observed Z_{TWAS} is among all possible W values coming from permuted expression data. Note that failing the permutation test may be an indication of lack of power to show that the expression significantly refines the direct SNP-trait signal. In practice, the permutation test was run 1,000 times for each TWAS gene, and a P value was computed by Z test against this null.

Relationship to genetic covariance/correlation. Our tests relate to previously defined estimators of genetic correlation and covariance between traits. We consider two definitions of genetic covariance at a locus: (i) the covariance between the genetic component of expression and the genetic component of a trait and (ii) the covariance between the causal effect sizes for expression and the causal effect sizes for a trait. Under assumptions of independent effect sizes, these definitions yield asymptotically identical quantities[37]. Assuming a substantially large training set where the genetic component of expression can be perfectly predicted, the direct TWAS tests for a significant association between the genetic component of expression and the trait—equivalent to testing definition (i) for a polygenic trait. Likewise, the summary-based TWAS tests for a significant sum of products of the causal expression effect sizes and the causal trait effect sizes—equivalent to definition (ii) up to a scaling factor. The TWAS approach therefore fits naturally with the broader study of shared genetic etiology of multiple phenotypes. At the sample sizes evaluated in this study, the TWAS approach is substantially better powered than an LD-based estimate of local genetic correlation (**Supplementary Note**).

52. Stancáková, A. *et al.* Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* **61**, 1895–1902 (2012).
53. Stancáková, A. *et al.* Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* **58**, 1212–1221 (2009).
54. Turchin, M.C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* **44**, 1015–1019 (2012).
55. Boomsma, D.I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet.* **9**, 849–857 (2006).
56. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
57. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
58. Bolormaa, S. *et al.* A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.* **10**, e1004198 (2014).
59. Zhu, X. *et al.* Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96**, 21–36 (2015).
60. Zaitlen, N., Paşaniuc, B., Gur, T., Ziv, E. & Halperin, E. Leveraging genetic variability across populations for the identification of causal variants. *Am. J. Hum. Genet.* **86**, 23–33 (2010).
61. Han, B., Kang, H.M. & Eskin, E. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS Genet.* **5**, e1000456 (2009).