The missing link between genetic association and regulatory function

Noah Connally[1,2,3], Sumaiya Nazeen[1,2,4], Daniel Lee[1,2,3], Huwenbo Shi[3,5], John Stamatoyannopoulos[6], Sung Chun[7], Chris Cotsapas[3,8,9]*, Christopher A. Cassa[2,3]*, Shamil Sunyaev[1,2,3]*

[1] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[2] Brigham and Women's Hospital, Division of Genetics, Harvard Medical School, Boston, MA, USA
[3] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4] Brigham and Women's Hospital, Department of Neurology, Harvard Medical School, Boston, MA, USA
[5] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[6] Altius Institute, Seattle, WA, USA
[7] Division of Pulmonary Medicine, Boston Children's Hospital, Boston, MA, USA
[8] Department of Neurology, Yale Medical School, New Haven, CT, USA
[9] Department of Genetics, Yale Medical School, New Haven, CT, USA
* Co-corresponding authors

The genetic basis of most traits is highly polygenic and dominated by non-coding alleles. It is widely assumed that such alleles exert small regulatory effects on the expression of *cis*-linked genes. However, despite the availability of gene expression and epigenomic data sets, few variant-to-gene links have emerged. It is unclear whether these sparse results are due to limitations in available data and methods, or to deficiencies in the underlying assumed model. To better distinguish between these possibilities, we identified 220 gene-trait pairs in which protein-coding variants influence a complex trait or its Mendelian cognate. Despite the presence of expression quantitative trait loci near most GWAS associations, by applying a gene-based approach we found limited evidence that the baseline expression of trait-related genes explains GWAS associations, whether using colocalization methods (8% of genes implicated), transcription-wide association (2% of genes implicated), or a combination of regulatory annotations and distance (4% of genes implicated). These results contradict the hypothesis that most complex trait-associated variants coincide with homeostatic eQTLs, suggesting that better models are needed. The field must confront this deficit, and pursue this "missing regulation."

Introduction

Modern complex trait genetics has uncovered surprises at every turn, including the paucity of associations between traits and coding variants of large effect, and the "mystery of missing heritability," in which no combination of common and rare variants can explain a large fraction of trait heritability[1]. Further work has revealed unexpectedly high polygenicity for most human traits and very small effect sizes for individual variants. Bulk enrichment analyses have demonstrated that a large fraction of heritability resides in regions with gene regulatory potential, predominantly tissue-specific accessible chromatin and enhancer elements, suggesting that trait-associated variants influence gene regulation[2–4]. Furthermore, genes in trait-associated loci are more likely to have genetic variants that affect their expression levels (expression QTLs, or eQTLs), and the variants with the strongest trait associations are more likely also to be associated with transcript abundance of at least one proximal gene[5]. Combined, these observations have led to the inference that most trait-associated variants are eQTLs, and their effects arise from altering transcript abundance, rather than protein sequence. Equivalent sQTL (splice QTL) analyses of exon usage data have revealed a more modest overlap with trait-associated alleles, suggesting that a fraction of trait-associated variants influence splicing, and hence the relative abundance of different transcript isoforms, rather than overall expression levels. The genetic variant causing expression changes may lie outside the locus and involve a knock-on effect on gene regulation, with the variant altering transcript abundances for genes elsewhere in the genome (a *trans*-eQTL), but the consensus view is that *trans*-eQTLs are typically mediated by the variant influencing a gene in the region (a *cis*-eQTL)[6]. Thus, a model has emerged in which most trait-associated variants influence proximal gene regulation.

Here we argue that this unembellished model—in which GWAS peaks are mediated by the effects on the homeostatic expression in bulk tissues—is the exception rather than the rule. We highlight challenges of current strategies linking GWAS variants to genes and call for a reevaluation of the basic model in favor of more complex models possibly involving context-specificity with respect to cell types, developmental stages, cell states, or the constanstancy of expression effects.

Our argument begins with several observations that challenge the unembellished model. One challenge is the difference between spatial distributions of eQTLs, which are dramatically enriched in close proximity to genes, and GWAS peaks, which are usually farther away[7–9]. Another is that expression levels mediate a minority of complex trait heritability[10]. Finally, many studies have designed tools for colocalization analysis: a test of whether GWAS and eQTL associations are due to the same set of variants, not merely distinct variants in linkage disequilibrium. If the model is correct, most trait

associations should also be eQTLs, but across studies, only 5-40% of trait associations co-localize with eQTLs[11–14].

Despite the doubts raised, the fact that most GWAS peaks do not colocalize with eQTLs cannot disprove the predominant, unembellished model. In a sense, negative colocalization results are confusing because their hypothesis is too broad. If we predict merely that GWAS peaks will colocalize with *some* genes' expression, it is not clear what is meant by a peak's failure to colocalize with *any individual* gene's expression.

Thus, a narrower, more testable hypothesis requires identifying genes we believe *a priori* are biologically relevant to the GWAS trait. If these trait-linked genes have nearby GWAS peaks and eQTLs, failure to colocalize would be a meaningful negative result. Earlier studies tested all GWAS peaks; when a peak has no colocalization, the model is inconclusive. But trait-linked genes that fail to colocalize reveal that our method for detecting non-coding variation is, with current data, incompatible with our model for understanding it.

With this distinction in mind, we created a set of trait-associated genes capable of supporting or contradicting the model of non-coding GWAS associations acting as eQTLs. For this purpose the selection of genes becomes extremely important. Because the model attempts to explain the genetic relationship between traits and gene expression, true positives cannot be selected based on measurements of genetic association to traits (GWAS) or expression (eQTL mapping). With this restriction, one source of true positives is to identify genes that are both in loci associated with a complex trait and are also known to harbor coding mutations tied to a related Mendelian trait or the same complex trait. Using a model not based on expression, Mendelian genes are enriched in common-variant heritability for cognate complex traits[15]. The genes and their coding variants may be detected in familial studies of cognate Mendelian disorders, or by aggregation in a burden test on the same complex phenotypes as GWAS[16]

For genes whose coding variants can cause detectable phenotypic change, the strong expectation is that a variant of small effect influences the gene identified by its rare coding variants. As an example, *APOE* and *LDLR* are both low-density lipoprotein receptor genes[17,18]. Coding variants in *APOE* and *LDLR* can lead to the Mendelian disorder familial hypercholesterolemia[18,19]. Even in the absence of a Mendelian coding variant, experiments in animal models have found that the overexpression of these genes reduces cholesterol levels[20–22]. GWAS on human subjects have found significant associations near *APOE* and *LDLR*, so it seems reasonable to suspect that any noncoding effects in these loci may be mediated by these genes. This general

relationship between Mendelian and complex traits is supported by several lines of evidence summarized in **Supplementary Note 1**.

Results

To test the model that trait-associated variants influence baseline gene expression, we assembled a list of such putatively causative genes. We selected seven polygenic common traits with available large-scale GWAS data, each of which also has an extreme form in which coding variants of large effect alter one or more genes with well-characterized biology (**Table 1**). Our selection included four common diseases: type II diabetes[23], where early onset familial forms are caused by rare coding mutations (insulin-independent MODY; neonatal diabetes; maternally inherited diabetes and deafness; familial partial lipodystrophy); ulcerative colitis and Crohn disease[24,25], which have Mendelian pediatric forms characterized by severity of presentation; and breast cancer[26], where germline coding mutations (e.g., *BRCA1*) or somatic tissue (e.g., *PIK3CA*) are sufficient for disease. We also chose three quantitative traits: low and high density lipoprotein levels (LDL and HDL); and height.

In well-powered GWAS, even relatively rare large-effect coding alleles (mutations in *BRCA1* which cause breast cancer, for instance) may be detectable as an association to common variants, which could make the effect of a coding variant appear to be regulatory instead. To account for this possibility, we computed association statistics in each GWAS locus conditional on coding variants. We applied a direct conditional test to datasets with available individual-level genotype data (height, LDL, HDL); for those studies without available genotype data, we computed conditional associations from summary statistics using COJO[27,28] (**Methods**). With both methods, the resulting GWAS associations should reflect only non-coding variants.

After controlling for coding variation, we examined whether these genes are more likely than chance to be in close proximity to variants associated with the polygenic form of each trait. In agreement with existing literature[29], we observe a significant enrichment for all traits in our combined Mendelian and Backman *et al.* gene sets (**Supp. Fig. 1**).

Of our 220 genes, 147 (67%) fell within 1 Mb of a GWAS locus for the cognate complex trait, over three times as many as the 43 predicted by a random null model (95% confidence interval: 31.5-54.5). Our window of 1 Mb represents roughly the upper bound for distances identified between enhancer-promoter pairs, but most pairs are closer[30], so we would expect enrichment to increase as the window around genes decreases; this proves to be the case. At a distance of 100 kb, we find 104 putatively causative genes (47%), though the null model predicts only 11 (95% CI 4.5-17.0), a order-of-magnitude enrichment (**Supp. Fig. 1**). Given their known causal roles in the

severe forms of each phenotype, these results suggest that the 147 genes near GWAS signals are likely to be the targets of trait-associated non-coding variants. For example, we see a significant GWAS association between breast cancer risk and variants in the estrogen receptor (*ESR1*) locus even after controlling for coding variation; the baseline expression model would thus predict that non-coding risk alleles alter *ESR1* expression to drive breast cancer risk.

We next looked for evidence that the trait-associated variants were also altering the expression of our 147 genes in relevant tissues. Controlling for the number of tests we conducted, 134 of these genes had an eQTL in at least one relevant tissue at a false discovery rate of $Q < 0.05$ (**Methods**). If these variants act through changes in gene expression, phenotypic associations should be driven by the same variants as eQTLs in relevant tissue types. We therefore looked for co-localization between our GWAS signals and eQTLs in relevant tissues (**Table 2**) drawn from the GTEx Project, using three well-documented methods: coloc[11], JLIM[12], and eCAVIAR[14]. We found support for the colocalization of trait and eQTL association for only 7 genes out of 147 (4.8%) for coloc; 10/147 (6.8%) for JLIM; and 8/147 (5.4%) for eCAVIAR. Accounting for overlap, this represents only 18/220 putatively causative genes (8.2%) or 18/147 (12.2%) putatively causative genes near GWAS peaks, even without full multiple-hypothesis testing correction (**Methods**), which is not obviously better than random chance. We note that prior estimates of the fraction of *GWAS associations* colocalizing with eQTLs (25%-40%[11,12,14,31]) do not directly evaluate the ability to find causative genes. By contrast, our estimate of the number of putatively causative *genes* that colocalize with eQTLs tests the consistency of our knowledge, models, and data.

A potential weakness of our approach is the restriction of our search to pre-defined tissues. We believe this is necessary in order to avoid the disadvantages of testing each gene-trait pair in each tissue—either a large number of false positives, or a severe multiple-testing correction that may lead to false negatives. However, restricting to the set of tissues with a known biological role and available expression data almost certainly leaves out tissues with relevance in certain contexts. Some of the tissues we do use have smaller sample sizes, limiting their power to detect eQTLs with smaller effects.

To address potential shortcomings from the available sample of tissue contexts, we incorporated the Multivariate Adaptive Shrinkage Method (MASH)[32]. MASH is a Bayesian method that takes genetic association summary statistics measured across a variety of conditions and, by determining patterns of similarity across conditions, updates the summary statistics of each individual condition. In our case, if an eQTL is difficult to find in a tissue of interest, incorporating information from other tissues may

help us detect it. Unlike meta-analysis, this method generates summary statistics that still correspond to a specific tissue.

We ran MASH on every locus used in our earlier analysis, using data from all non-brain GTEx tissues (**Methods**). Rerunning coloc with these modified statistics increased the number of GWAS-eQTL colocations from 389 to 489. However the 100 new colocations identified only four additional putatively causative genes (**Supp. Fig. 2**). These results indicate that tissue type selection was not the limiting factor in our analysis.

Transcriptome-wide association studies (TWAS)[33–36] are another class of methods applied to identify causative genes under GWAS peaks using gene expression. TWAS measures genetic correlation between traits and is not designed to avoid correlations caused by LD, which gives it higher power in the case of allelic heterogeneity or poorly typed causative variants[37]. However, while sensitive, TWAS analyses typically yield expansive result sets that include many false positives and are sensitive to the number of tissue types[37]. Results from the FUSION implementation of TWAS[35] across all tissues identified our putatively causative genes as likely tied to the GWAS peak in 66/220 loci (30%). However, only 4/220 (1.8%) genes were identified by FUSION when we restricted the analysis to relevant tissues.

Given the paucity of expression-mediated GWAS peaks, we asked whether GWAS variants indeed reside in regulatory sites. Taking the 128 genes in the Mendelian subset of putatively causative genes, we fine-mapped each nearby GWAS association using the SuSiE algorithm[38]. For 37 of these genes, we identified at least one high-confidence fine-mapped SNP (PIP>0.7) within 100kb of the transcription start site. We tested whether these fine-mapped SNPs fall within regulatory DNA marked by chromatin accessibility[39], a narrowly mapped active histone modification feature (H3K27ac, H3K4me1, or H3K4me3[40]), or characterized as an “enhancer” by ChromHMM[41,42] (**Methods**). As many as 32/37 (86%) genes identified this way have a fine-mapped SNP within a regulatory feature across all the tissue types examined, or 25/37 (68%) when restricting to phenotype-relevant tissues (**Fig. 3**; **Supp. Table 1**, **Supp. Table 2**). Despite strong evidence that these GWAS associations are due to regulatory effects, only 5/25 loci (20%) demonstrably correspond to expression effects in our eQTL analysis.

In order to more directly compare our regulatory feature analysis to our eQTL analysis, we measured “activity-by-distance”—a simplification of the “activity-by-contact” method[30,43] (**Methods; Fig. 2**). Taking each locus’s feature with the highest ABD score, we implicate 5/37 (14%) of our Mendelian subset of genes. This reinforces our

observation that, even when a GWAS association and trait-relevant gene are proximal, they are difficult to link, whether using eQTLs or chromatin data.

Discussion

Overall, our results are consistent with the idea that complex traits are governed by non-coding genetic variants whose effects on phenotype are mediated by their contribution to the regulation of nearby genes. However, these results are inconsistent with the model that a common mechanism of this mediation is the effect on baseline expression within tissues. The enrichment of our putatively causative genes—selected based on existing biological knowledge—near GWAS peaks supports their role in complex traits. Additionally, the enrichment of fine-mapped GWAS variants in accessible chromatin regions and regulatory features lends support to the model of GWAS associations being produced by eQTLs. However, the inability of varied statistical methods to actually link GWAS associations and expression contradicts the idea that the causative GWAS variants are homeostatic, bulk-tissue eQTLs of the sort found in broad expression-data collection projects.

Many explanations have been suggested for the limited success of expression methods to explain the mechanisms of GWAS variants. Undirected, broad approaches—including most GWAS-eQTL linking studies—are designed to be largely independent of *a priori* biological knowledge and hypotheses. This unconstrained focus is ideal for discovery, but, though it delivers the largest number of positive findings, it is ill-suited to provide an explanation for negative results—when you don't know what you were looking for, it's hard to explain why you didn't find it. By testing only loci for which there is a strongly suspected contributing gene, we are better able to distinguish which factors prevent us from identifying it using expression.

As a result, we conclude that a number of explanations often considered when evaluating expression-based variant-to-gene methods are not applicable in the context we examined. These include non-expression-mediated mechanisms, lack of statistical power for GWAS, the absence of eQTLs for relevant genes, and underpowered methods for linking expression to GWAS (**Table 3**).

Instead, we believe the "missing regulation" will be found primarily through examining more nuanced models of gene expression. Solving the mystery will require not only identifying the eQTLs behind GWAS peaks, but also explaining the phenotypic irrelevance of our "red herring eQTLs": eQTLs for putatively causative genes that fall near GWAS peaks but do not colocalize with them. Some proposed models involve expression that depends on context—whether cell type, cell state, environment, or developmental stage. Others depend on heterogeneity of expression or the variance of

expression across relatively short time scales. These various models may depend on or be augmented by thresholding or buffering—processes causing a change in gene expression to have a non-linear effect on phenotype. A summary of these models can be found in **Table 4**.

The implications of our results are both conceptual and practical. The inability of current models to identify the expression effects of known trait-associated genes, and to explain the non-effects of their identified eQTLs, calls for new models of the role of gene regulation in complex traits. One long-standing goal of GWAS has been to discover genes contributing to complex traits[1,8], but low rates of positive findings for expression-based variant-to-gene methods have constrained this possibility[12,44]. Among other challenges, this has limited the benefit of GWAS and expression data for disease-gene mapping and drug discovery[44,45]. Another practical question raised is the value of different large-scale datasets. Compared to genotypes, expression data are relatively difficult to collect. If the most relevant models are shown to depend on effects not observable in bulk-tissue, homeostatic eQTL mapping, the field may need to consider prioritizing other forms of expression data.

The introduction to this manuscript includes two examples of suspect genes: *APOE* and *LDLR*. Both genes harbor coding variants causing Mendelian hypercholesterolemia. Both have non-coding variants that GWAS have tied to LDL levels. Both have eQTLs in trait-relevant tissues. For *APOE*, these points cohere into an explanation: the LDL-association is an eQTL for the lipid-binding gene. But for *LDLR*—and for most genes—the association, the mechanism, and the gene cannot be tied together. In the field of complex-trait genetics—both basic and translational—solving this regulatory mystery may prove to be a critical step.

Figures

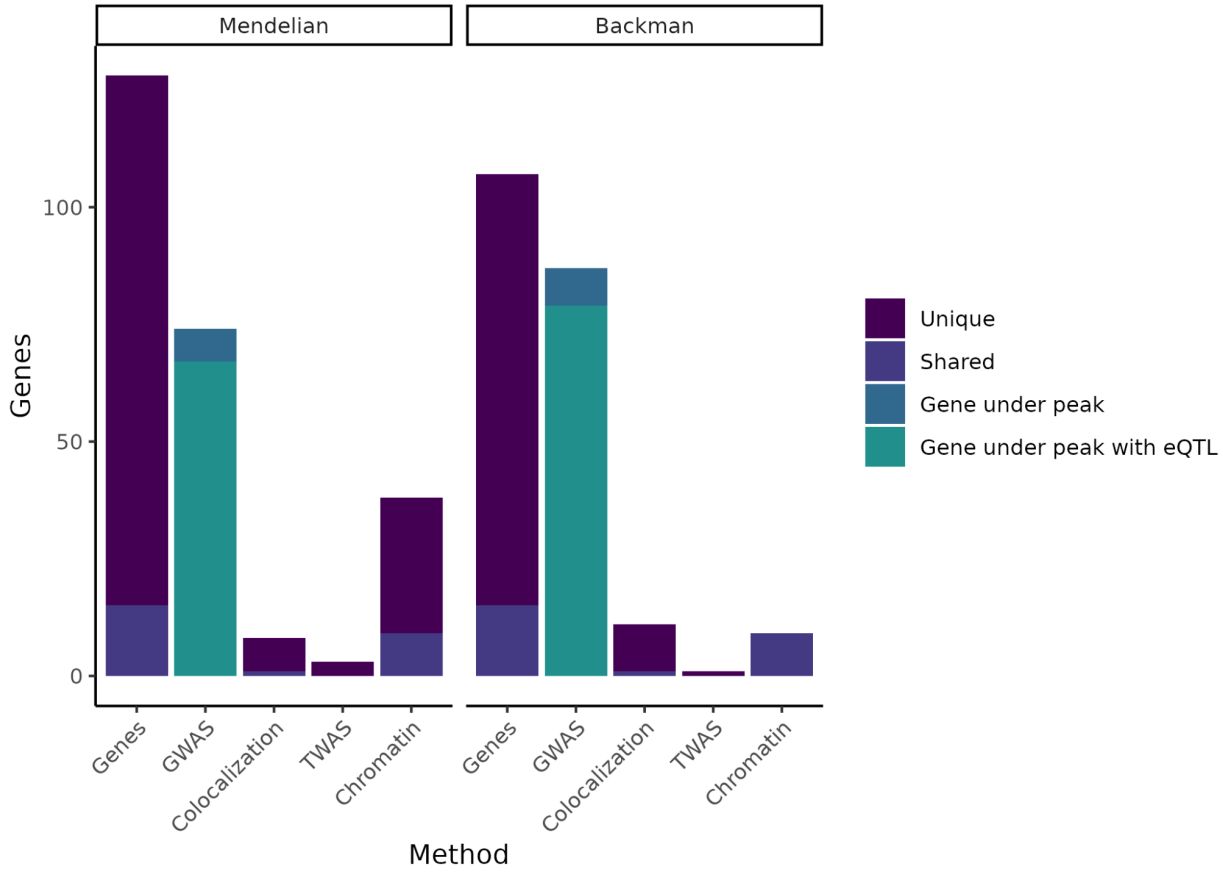


Figure 1. Putatively causative genes identified by each method category.
The leftmost column in each half of the plot displays the entire group of putatively causative genes for our Mendelian set of genes and our Backman *et al.* set of genes respectively, as well as noting how many are unique to each set or shared between the two sets. The second column in each half indicates how many genes from each set have a nearby GWAS peak, or have both a nearby GWAS peak and an eQTL. The remaining columns indicate how many genes were identified through colocalization, TWAS, or chromatin methods, while noting how many of these genes are unique vs. shared between the Mendelian and Backman sets.

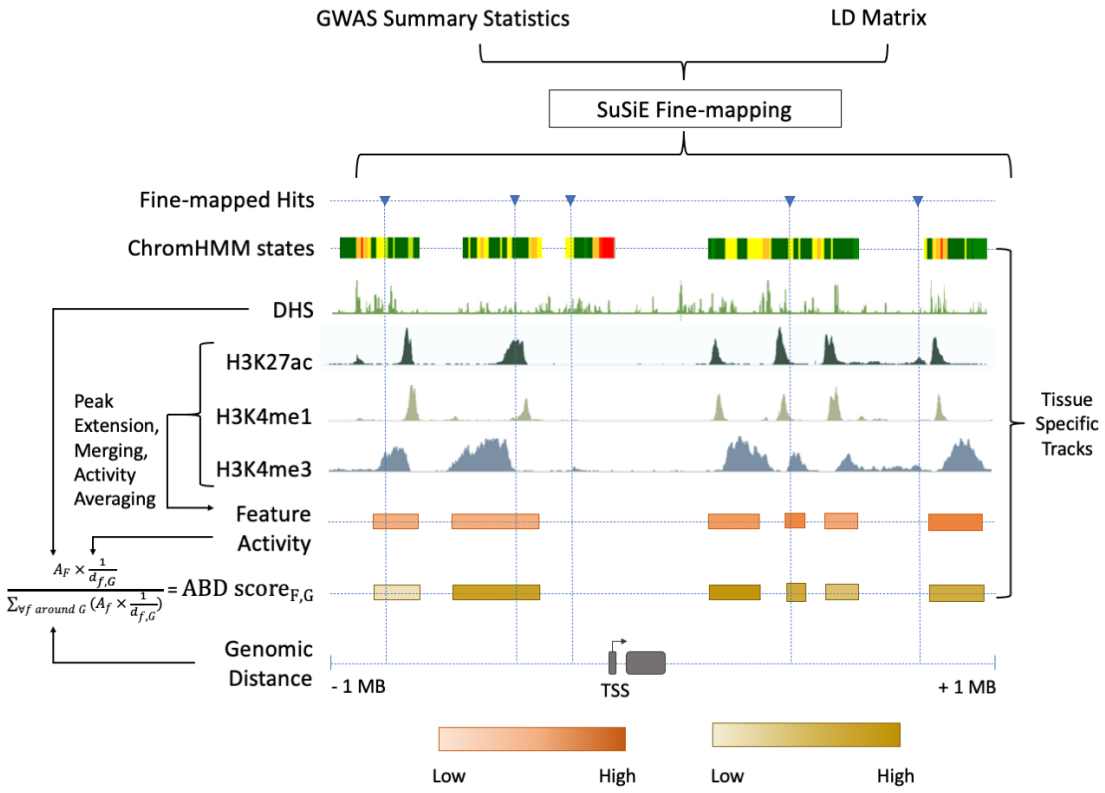


Figure 2. Chromatin-based causative gene identification.
Following the fine-mapping of GWAS variants, three parallel methods were used. The first identified fine-mapped variants falling within regions annotated as enhancers by ChromHMM. The second identified variants within histone modification features, and evaluated their relevance using an ABD score that combined the strength of the feature (i.e., the strength of the acetylation or methylation peak) with its genomic distance to the gene of interest (**Methods**). The third repeated both of these—checking for fine-mapped variants within a region and calculating the ABD score—for DNase I hypersensitivity sites.

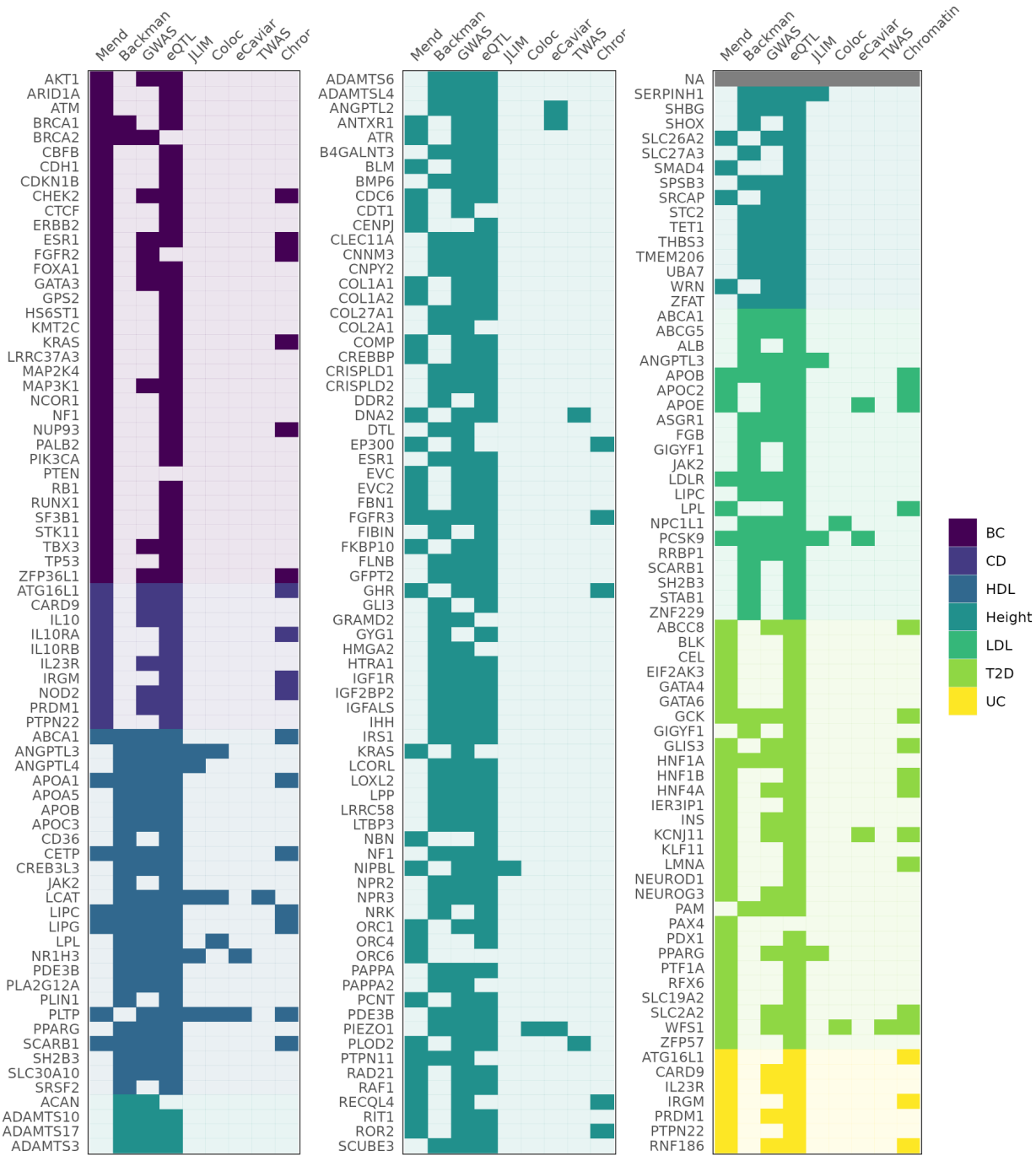


Figure 3. Genes identified as associated with a complex trait by each method. Columns “Mend” and “Backman” indicate whether a gene is from the Mendelian set of putatively causative genes, the Backman et al. set, or both. Subsequent columns indicate whether a gene was identified as a hit using each of our methods: JLIM, coloc, eCaviar, TWAS, and chromatin analysis.

Phenotype	**Genes**
LDL	APOB[46,47] APOC2[48] APOE[49] LDLR[50] LPL[51,52] PCSK9[53]
HDL	ABCA1[54–56] APOA1[57] CETP[58] LIPC[59–61] LIPG[62] PLTP[63] SCARB1[64,65]
Height	ANTXR1[66,67] ATR[68,69] BLM[70,71] CDC6[72] CDT1[72,73] CENPJ[74] COL1A1[75] COL1A2[76,77] COMP[78,79] CREBBP[80–82] DNA2[83] EP300[84,85] EVC[86,87] EVC2[87,88] FBN1[89–92] FGFR3[93–95] FKBP10[96–98] GHR[99–102] KRAS[103–105] NBN[106,107] NIPBL[108,109] ORC1[72,73,110] ORC4[72,73] ORC6L[72,111] PCNT[112–114] PLOD2[115–117] PTPN11[118–120] RAD21[121–123] RAF1[124,125] RECQL4[126–128] RIT1[129–131] ROR2[132–134] SLC26A2[135–137] SMAD4[138–140] SRCAP[141,142] WRN[143–145]
Crohn disease	ATG16L1[146]

	CARD9[147] IL10[148] IL10RA[149,150] IL10RB[151,152] IL23R[153–155] IRGM[156–158] NOD2[159,160] PRDM1[161] PTPN22[162] RNF186
Ulcerative colitis	ATG16L1[163] CARD9[147] IL23R[155,164] IRGM[156] PRDM1[161] PTPN22[162] RNF186[165,166]
Type II diabetes	ABCC8[167] BLK[168] CEL[169,170] EIF2AK3[171–173] GATA4[174] GATA6[175,176] GCK[177] GLIS3[178] HNF1A[179,180] HNF1B[181,182] HNF4A[183,184] IER3IP1[185–187] INS[188] KCNJ11[189,190] KLF11[191] LMNA[192] NEUROD1[193] NEUROG3[194–196] PAX4[197–199] PDX1[200–202] PPARG[203,204] PTF1A[205] RFX6[206,207] SLC19A2[208–210] SLC2A2[211,212] WFS1[213–215] ZFP57[216,217]
Breast cancer (selected using MutPanning[218])	AKT1 ARID1A ATM BRCA1 BRCA2 CBFB CDH1 CDKN1B

	CHEK2 CTCF ERBB2 ESR1 FGFR2 FOXA1 GATA3 GPS2 HS6ST1 KMT2C KRAS LRRC37A3 MAP2K4 MAP3K1 NCOR1 NF1 NUP93 PALB2 PIK3CA PTEN RB1 RUNX1 SF3B1 STK11 TBX3 TP53 ZFP36L1

Table 1. Putatively causative Mendelian genes.
Each gene includes reference(s) to the known biological role of its coding variants, as established in familial studies, *in vitro* experiments, and/or animal models. Genes from Backman *et al.* are not included here, but can be found in **Figure 3**.

Mendelian trait	**GWAS trait**	**Tissues examined**
Breast cancer	Breast cancer	Breast mammary tissue
Crohn disease	Crohn disease	Small intestine terminal ileum Colon Sigmoid Colon Transverse
Ulcerative colitis	Ulcerative colitis	Small intestine terminal ileum Colon Sigmoid Colon Transverse
Dyslipidemia Hyperlipidemia Tangier's disease	HDL	Liver Adipose (subcutaneous) Whole blood
Dyslipidemia Hyperlipidemia	LDL	Liver Adipose (subcutaneous) Whole blood

Mendelian short stature	Height	Skeletal muscle
Monogenic diabetes	Type II diabetes	Pancreas Skeletal muscle Adipose (subcutaneous) Small intestine terminal ileum

Table 2. Tissue-trait pairs
Tissues were selected for each trait based on *a priori* knowledge of disease biology.

Violated assumptions	
Genes implicated via coding variants are irrelevant for non-coding associations	- Our genes are enriched for GWAS associations even after removing the effects of coding variants. - Loss-of-function variants, which underlie many Mendelian-trait genes, can be thought of as large-effect eQTLs. - Genes identified from Backman *et al.* are not based on cognate phenotypes, but the same complex phenotypes as GWAS.
Regulatory mechanisms other than *cis*-eQTLs	- Splice QTLs are consistently found to explain less phenotypic variance than eQTLs, and they cannot explain the many GWAS associations that fall within intergenic regions. - *Trans*-eQTLs are believed to rely on their effects as *cis*-eQTLs for other genes; the few exceptions to this model (e.g., CTCF binding sites) are not broadly applicable.
Insufficient power	
Lack of GWAS power	- GWAS have been shown to have sufficient power to identify small effects even in rare variants. - 2/3 of the genes we used have nearby GWAS associations, reflecting a strong enrichment and indicating that GWAS discovery is not a limiting factor. - Our analysis is conditioned on the presence of GWAS associations.
Lack of eQTL mapping power	- GTEx is well powered for eQTL discovery in bulk tissue[6,33]. - 93% of our genes have a mapped cis-eQTL in a relevant tissue.
Lack of power for colocalization and TWAS methods	- Simulations show that colocalization and TWAS methods are well-powered[12–14,34,219]. - They are robust to levels of LD mismatch higher than what would be expected given our datasets[12,37,219].

	- Some, though not all, of the methods are robust to allelic heterogeneity[12,37].

Table 3. Proposed explanations for negative results under the unembellished model.

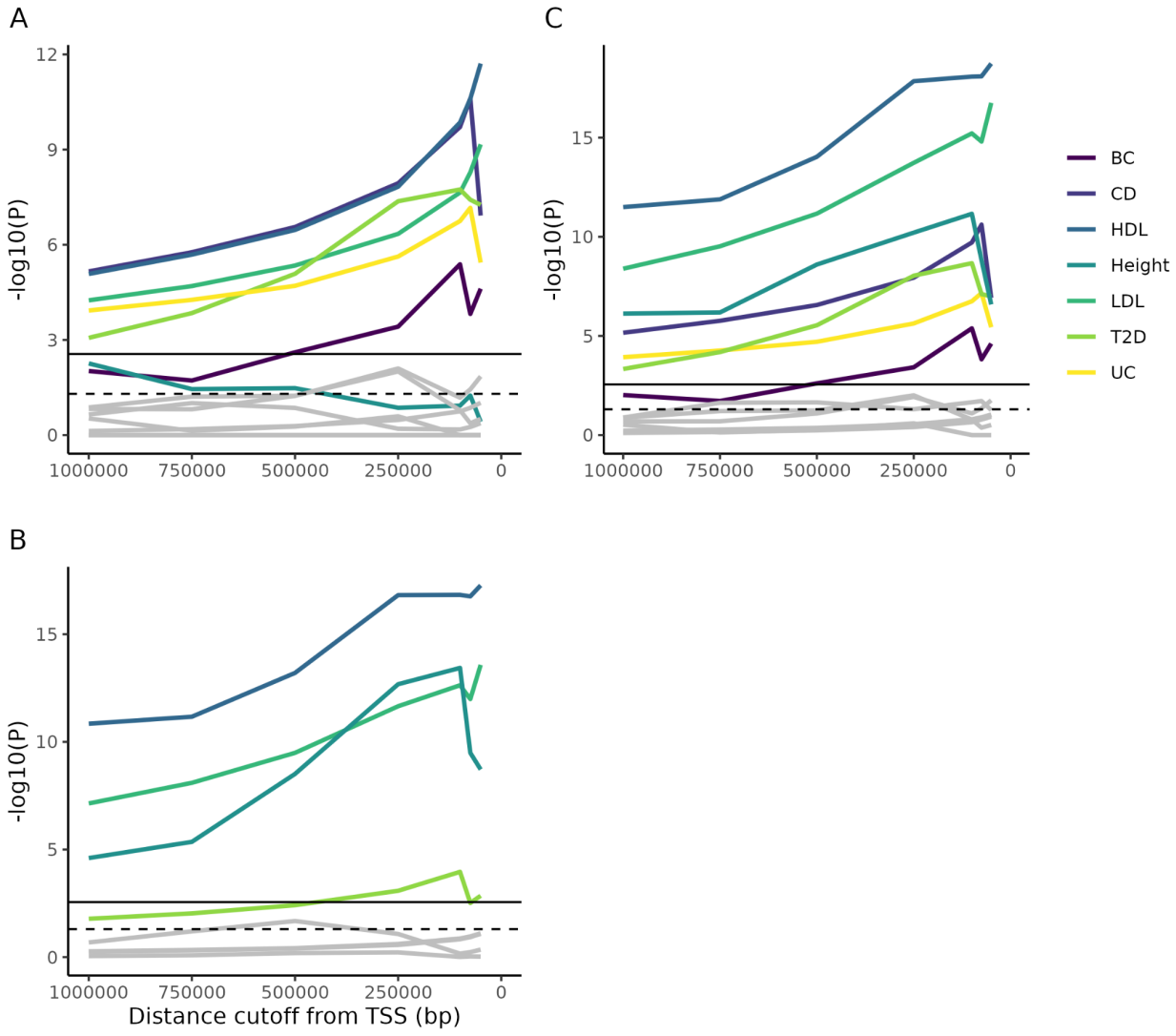
Many explanations have been proposed for GWAS associations that are not explained by *cis*-eQTLs. This table details explanations inconsistent with our results, which are explained in the left column and addressed in the right. Explanations involving more detailed models of gene regulation can be found in **Table 4**. Two of the explanations addressed here involve violations of the assumptions of our and other expression-based complex trait studies. If coding and non-coding variants affect fundamentally different biological pathways, or if trait associations rarely depend on *cis*-eQTLs, our methods of mapping regulation to traits would have nothing to uncover. Even in the presence of eQTL-driven trait associations, insufficient power to detect trait associations, to detect eQTL associations, or to link the two would result in predominantly negative results.

Extended models of gene regulation	
Context dependency: A context-specific eQTL, invisible in bulk tissue, replaces or supplements the bulk tissue homeostatic eQTL	Cell type[220–234] - Only a subset of cell types in the tissue contribute to the GWAS phenotype. - An eQTL specific to such a cell type is causative for the phenotype. - The eQTL cannot be detected in bulk tissue because of the cell type's low prevalence.
	Developmental timing[220,235–240] - The GWAS phenotype is influenced by an earlier point in individual development or cell differentiation. - eQTLs present at the correct moment contribute to phenotype, but eQTLs observed later do not.
	Cell state or environment[225,227,231,233,241–248] - The causative eQTL has effects that are undetectable in steady-state expression under normal conditions. - It may activate only in response to a specific environmental condition, such as infection. - It may depend on the stage of the cell cycle.
Non-linear or non-homeostatic: The relationship between eQTL and genotype is indirect	Nonlinearity[249–269] - There may be buffering that prevents a change in expression from producing a change in protein levels. - Expression below a certain level may not influence phenotype, rendering small eQTLs irrelevant.
	Steady-state expression may be a poor model[270–281] - Phenotype may depend on the kinetics of expression, which

	could be cyclical or follow some other pattern. - Expression may be stochastic, such that only a random subset of cells display the relevant expression pattern at any one time.

Table 4. Explaining negative results with more nuanced models of gene regulation.

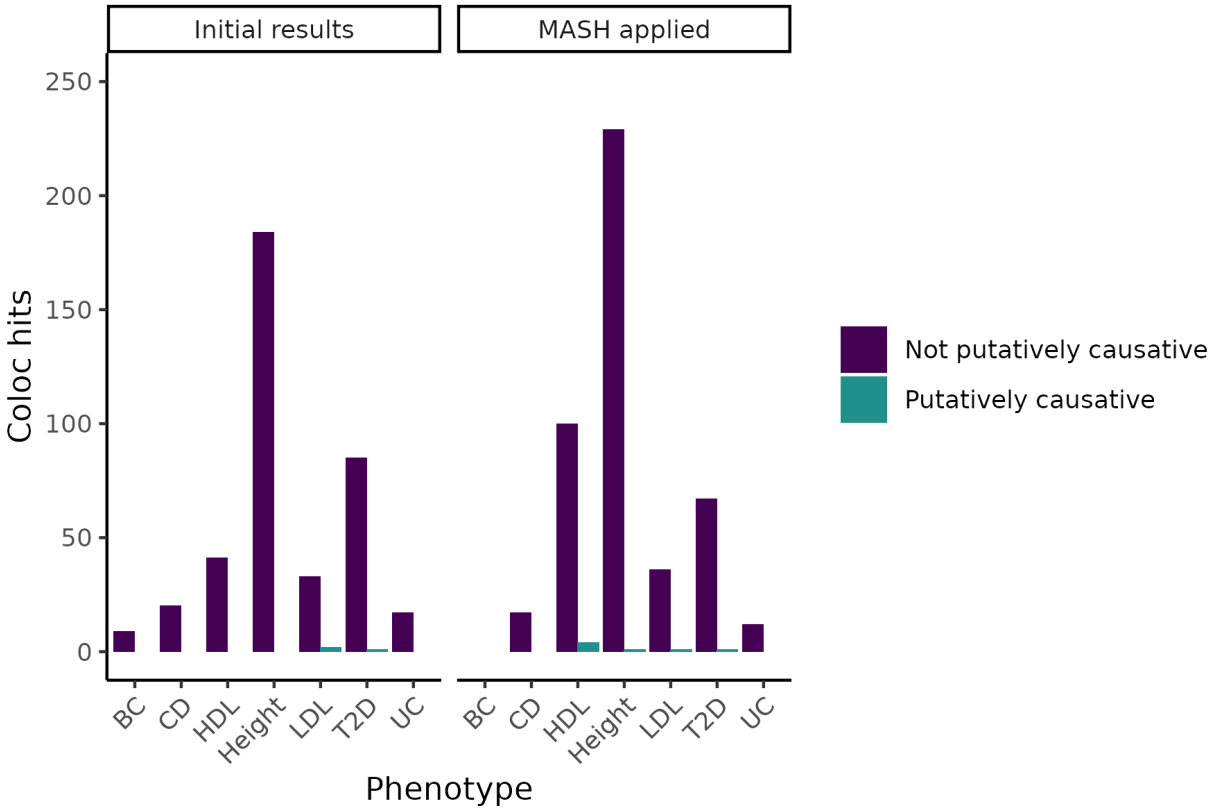
To reconcile an expression-based model with our observations requires us to both explain the absence of trait-linked eQTLs as well as explaining away the inconsequence of eQTLs for trait-linked genes. The left-hand side lists additions or changes to the unembellished model, while the right contains explanations of the models and current relevant research.



Supplementary figure 1. Enrichment of Mendelian genes near GWAS peaks.

A) As the window around GWAS peaks shrinks, the enrichment of Mendelian genes within the window becomes increasingly significant, while the enrichment of non-

matching trait pairs used as controls (gray lines; **Methods**) is not consistently increased. Some controls achieve nominal significance (dotted horizontal line), but none reach significance once multiple-testing is corrected for (solid horizontal line). B) As above, but for genes from Backman et al. (2021)[16]. C) The combined gene lists from parts A and B. Note that, accounting for multiple test correction (based on the total number of tests across all panels) height does not reach significance using the Mendelian gene list, while T2D is barely significant using the Backman list. However, combining the lists increases power and demonstrates significance for all traits.



Supplementary figure 2. Change in coloc hits when adjusting eQTL statistics using MASH.

By using the Bayesian method MASH to update our measurements of eQTLs based on tissues with similar expression patterns, we increased the number of colocalizations found. However, even in tissues in which the number of genes identified increased substantially, we did not meaningfully increase the number of putatively causative genes identified.

Alias	Descriptive Name
E027	Breast myoepithelial primary Cells
E028	Breast luminal epithelial Cells / Human mammary epithelial cells
E062	Primary mononuclear cells from peripheral blood
E063	Adipose nuclei
E066	Liver
E075	Colonic mucosa

E076	Colon smooth muscle
E087	Pancreatic islets
E095	Heart left ventricle
E096	Lung
E104	Heart right atrium
E107	Skeletal Muscle Male
E108	Skeletal Muscle Female
E109	Small Intestine

Supplementary Table 1. Roadmap epigenomics aliases of tissue types used for functional genomic analysis.
Tissue types from the Roadmap Epigenomics Consortium do not perfectly match those from GTEx. However, there is overlap, and as with GTEx, we analyzed trait-relevant tissues.

Tissue	Biosample Name
Adipose	AL
Blood	Namalwa_0, Karpas_422, K562, CD34_T0, CD56, GM12878, CD3, NB4, hTH1, CD34_T17, Dendritic_cells_C51, hTH17, CD34_T18, GM12865, CD14, CD34_T15, CD34_T8, hTR, iTH2, CD19, iTH1, KBM7, hTH2, Oci_Ly_7, Jurkat, Dendritic_cells_C190, CD4, RPMI_8226, fThymus, MM_1S, CD4pos_N, CD34_T6, CD8, CD20, CMK, GM06990, CD34_T4, HAP1, GM12864, Namalwa + Sendai virus_2h, CD34
Bone	fBone_arm_left, fBone_femur, fBone_leg_left, fBone_arm_right, SK_N_MC, MG63, fBone_leg_right, limb, A673, SJSA1
Colon	CACO2, RKO, LoVo, HCT116, CEC, PANC1, SW_480, HT29
Heart	HCF, fLeftVentricle, fHeartFibroblasts, RuES2_cms, H7_hESC_T2, H7_hESC_T9, fHeart, HPAEC, H7_hESC_T14, HCFaa, AoAF, HCM, Heart, fRightVentricle, H7_hESC_T5, fLeftAtrium
Liver	hepatocytes, fLiver, HepG2
Lung	fLung_L, A549, NHBE_RA, WI_38_TAM, fLung_R, NHLF, IMR90, SAEC, AG04450, NCI_H460, fLung, PC9, HMVEC_LBl, WI_38, HPF, HMVEC_LLy
Mammary	MCF10a_ER_SRC_6h_Tam/Washout 18 h, MCF10a_ER_SRC_control, MCF10a_ER_SRC_24h_Tam, MCF7, HMEC, T_47D, MCF7_ER, HMF, vHMEC
Muscle	SkMC, fMuscle_arm, fMuscle_lower_limb, LHCN_M2, fMuscle_upper_back, fMuscle_back, LHCN_M2_D4, Psoas_Muscle, SJCRH30, fMuscle_upper_limb_sk, HSMM_D, HSMM, fMuscle_trunk, fMuscle_leg
Pancreas	Pancreas, ISL1
Small Intestine	fIntestine_Sm, Small_Intestine_Mucosa

Supplementary Table 2. Tissue types and bio-samples from the DNase I hypersensitive sites index used for functional genomic analysis

Meuleman *et al.* assess DNase I hypersensitive sites across 438 cell and tissue types[39]; we selected the above based on their relevance to our complex traits.

Supplementary Note 1. Evidence for the relationship between Mendelian and complex traits.

More generally, this expectation is supported by several lines of evidence. Comorbidity between Mendelian and complex traits has been used to identify common variants associated with the complex traits[282]. Early GWAS found associations near genes identified through familial studies of severe disorders[283,284], and later implicated some of the same genes in complex and Mendelian forms of cardiovascular[285] and neuropsychiatric[286] traits. More recent analyses have found that GWAS associations are enriched in regions near causative genes for cognate Mendelian traits in blood traits[287], lipid traits and diabetes[15], as well as a diverse collection of 62 traits[29]. Another recent method used transcriptomic, proteomic and epigenomic data to prioritize genes and found that, in a selection of nine phenotypes, selected genes were enriched for Mendelian genes causing similar traits[288]. Together, these suggest that genes causing Mendelian traits also influence cognate complex traits, but not through the same coding mechanism.

Genes can also harbor coding variants tied to less severe forms of a trait. These coding variants are more difficult to identify individually, as their effect sizes are much smaller. However, the greater number of variants (in aggregate) and freedom from searching for severe segregating traits, allows the use of large population datasets. Backman *et al.* used burden testing on UK Biobank data to identify genes whose coding variation affects complex traits, finding many genes not identified through familial studies[16].

Supplementary methods

Gene selection

By manual literature search, we selected 128 genes harboring large-effect-size coding variants for one of the seven phenotypes (**Table 1**; specifically, we selected 128 gene-trait pairs, representing 121 unique genes). These genes were identified using familial studies, rare disease exome sequencing analyses, and, for breast cancer, using the MutPanning method[218] (citations for each gene are included in **Table 1**). Review papers, as well as the OMIM database[289], were generally used as starting points, but an examination of the original literature was needed to confirm genes' suitability. For example, though SMC3 is known to cause Cornelia de Lange syndrome, which is characterized in part by short stature, SMC3 mutations lead to a milder form of the syndrome, usually without a marked reduction in stature[290]. Several of these phenotypes —height, HDL, cholesterol, breast cancer, and type II diabetes—were also analyzed in *Backman et al.*, which, through burden testing, identified a total of 110 genes; after accounting for overlaps, this increased our set of putatively causative genes to 220[16].

The inclusion of genes from Backman *et al.* ensures that our results are not dependent on an undetected bias in our selection. The set of genes chosen from familial studies offered the advantage that it was selected based on independent methods and data distinct from the large-scale genotyping studies that have characterized the GWAS era. The tradeoff to this was the impossibility of selecting genes through a fully systematic and non-arbitrary process. Because this work was performed in the UKBB, there is some overlap between their data and ours. However, our work did not use exomes, and most of the variants driving their findings are too rare to influence GWAS results. When this is not the case, our decision to condition on coding variants should make the effects used in our work independent from their findings.

Identifying coding variants

Because GWAS sample sizes are large enough to detect the low-frequency coding variants used to select some of our genes, it is possible that a coding SNP would distort the association signal of nearby eQTLs. To minimize this concern, we removed the effects of coding variants on GWAS. Many variants can fall within coding sequences in rare splice variants, so it is important to remove only those variants that appear commonly as coding. These coding SNPs were selected based on the pext (proportion of expression across transcripts) data[291]. Two filters were used. First, we removed genes whose expression in a trait-relevant tissue was below 50% of their maximum expression across tissues. Second, we removed variants that fell within the coding sequence of less than 25% of splice isoforms in that tissue. The remaining variants were used to correct GWAS signal, as explained below.

GWAS

For height, LDL cholesterol, and HDL cholesterol, GWAS were performed using genotypic and phenotypic data from the UKBB. In order to avoid confounding, we restricted our sample to the 337K unrelated individuals with genetically determined British ancestry identified by Bycroft *et al.*[292] The GWAS were run using Plink 2.0[293], with the covariates age, sex, BMI (for LDL and HDL only), 10 principal components, and coding SNPs.

Conditional analysis

Because UKBB has limited power for breast cancer, Crohn disease, ulcerative colitis, and type II diabetes, we used publicly available summary statistics. The Conditional and Joint Analysis (COJO)[27,28] program can condition summary statistics on selected variants—in our case, coding variants—by using an LD reference panel. For this reference, we used TOPMed subjects of European ancestry[294]. The ancestry of these subjects was confirmed with FastPCA[295,296] and the relevant data were extracted using bcftools[297]. Our conditional GWAS data are available at doi:10.5061/dryad.612jm644q

Enrichment analysis

At each distance, the number of Mendelian and non-Mendelian genes within that window around GWAS peaks are counted. *P*-values are calculated using Fisher's exact test (Supp. Fig. 1). Because Mendelian genes may be unusually important beyond our chosen traits, we conduct a set of controls by measuring the enrichment of non-matching Mendelian and complex traits (CD genes & BC GWAS; BC genes & LDL GWAS; LDL genes & UC GWAS; UC genes & height GWAS; height genes & T2D GWAS; T2D genes & HDL GWAS; HDL genes & CD GWAS).

eQTL detection

eQTL summary statistics were taken from GTEx v7. Some methods detect colocalization with variants that are individually significant, but would not pass a genome-wide threshold[12]. Because we tested only a subset of genes and, we used the Bejamini-Hochberg method[298] to calculate the FDR based on the number of tests we conducted multiplied by a correction factor to account for variants that are tested in combination with a gene but are not reported (a factor of 20 closely matched the genome-wide FDR results for GTEx). With this method 204/220 (93%) of our genes displayed an eQTL, including 134/147 genes with a nearby GWAS peak (91%). Even using the FDR statistics of the GTEx project—which are based on the assumption of testing every gene in every tissue—107/220 (49%) of our genes and 76/147 (52%) of genes near GWAS peaks had an eQTL at $Q < 0.05$.

Colocalization

JLIM[12] was run using GWAS summary statistics and GTEx v7 genotypes and phenotypes, for the tissues listed in **Table 2**. Coloc[11] was run using GWAS and GTEx v7 summary statistics for the same tissues. eCAVIAR[14] was run using GWAS and GTEx v7 summary statistics for these tissues, and a reference dataset of LD from UKBB[299]. MASH was run incorporating data from all non-brain tissues, and coloc was re-run using the adjusted values for the same tissues as before.

MASH

Multivariate adaptive shrinkage (MASH) was applied to all GTEx tissues using the mashr R-package[32]. We restricted this model to non-brain tissues—which include all of our trait-selected tissues—due to the known tendency of brain and non-brait tissues to cluster separately in expression analysis[300–302].

Fusion (TWAS)

We used the FUSION implementation of TWAS, which accounts for the possibility of multiple *cis*-eQTLs linked to the trait-associated variant by jointly calling sets of genes predicted to include the causative gene, to interrogate our 220 loci[35]. FUSION included our putatively causative genes in the set identified as likely relevant to the GWAS peak

in 66/220 loci (30%). However, interpretation of this TWAS result is difficult. For many complex traits, TWAS returns a large number of findings (e.g., over 150 for LDL cholesterol and over 4,800 for height). This is in part due to the multiple genes jointly returned at a locus, and can also be a result of the large number of tissues and cell types included in the implementation of FUSION. Most hits are found in tissues without any clear relevance to the trait, and absent in relevant tissues—LDL, for example, has more TWAS associations between expression and eQTL in prostate adenocarcinoma (24 genes associated), brain pre-frontal cortex (23 genes associated), and transformed fibroblasts (21 genes associated) than it does in adipose (16 genes associated), blood (11 genes associated), or liver (5 genes associated). Individual genes were often identified as hits in multiple tissues, but with an inconsistent direction of effect—that is, increased gene expression correlated with an increase in the quantitative trait or disease risk in some tissues, but a decrease in others, which suggests that the gene in question may not be the one whose expression contributes to the complex trait. Because of this possibility, and the known biological role of many of our genes, we restricted our results to tissues with established relevance to our traits.

Fine-mapping GWAS hits

We fine-mapped the GWAS variants located within +/- 100 kb of our putatively causative genes by applying the SuSiE algorithm[38] on the unconditional summary statistics from the GWAS of breast cancer, Crohn disease, ulcerative colitis, type II diabetes, height, LDL cholesterol, and HDL cholesterol. An LD reference panel from UKBB subjects of European ancestry was used for this analysis. Fine-mapped variants were annotated using snpEff (v4.3t). Only non-coding variants were kept for further analysis.

Functional genomic annotation of fine-mapped hits

We projected fine-mapped GWAS variants onto active regions of the genome, identified using three alternative approaches: (i) histone modification features, (ii) DNase I hypersensitive sites, and (iii) ChromHMM enhancers.

First, we looked at three histone modification marks, namely, acetylation of histone H3 lysine 27 residues (H3K27ac), mono-methylation of histone H3 lysine 4 residues (H3K4me1), and tri-methylation of lysine 4 residues (H3K4me3) from the Roadmap Epigenomics Project[37] to identify functional enhancers which are key contributors of tissue-specific gene regulation. We downloaded imputed narrowPeak sets for H3K27ac, H3K4me1, and H3K4me3 from the Roadmap Epigenomics Project[37] ftp site (https://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidatedImputed/narrowPeak/) for 14 different tissue types (**Supp. Table 1**). For each tissue type, we extracted the narrow peaks that are within +/- 5 Mb of our putatively causative genes. Then following the approach described in Fulco et al.[36], we extended the 150 bp narrow

peaks by 175 bp on both sides to arrive at candidate features of 500 bp in length. All features mapping to blacklisted regions (https://sites.google.com/site/anshulkundaje/projects/blacklists) were removed. Remaining features were re-centered around the peak and overlapping features were merged to give the final set of features per histone modification track. Mean activity/strength of a feature (A_F) was calculated by taking the geometric mean of the corresponding peak strengths from H3K27ac, H3K4me1, and H3K4me3 marks. We then combined these activity measurements with the linear distances between the features and the transcription start sites of causative genes to compute "activity-by-distance" scores (a simplified version of ABC scores[36]) for gene-feature pairs using the following formula.

$$ABD\,score_{F,G} = \frac{A_F \times 1/d_{F,G}}{\sum_{all\,f\ within\ \pm 5\,Mb\ of\ G} A_f \times 1/d_{f,G}}$$

The ABD score can be thought of as a measure of the contribution of a feature, F to the combined regulatory signals acting on gene, G. A high ABD score may serve as a proxy for an increased specificity between a chromatin feature and the gene of interest. We projected the fine-mapped variants onto the chromatin features in different tissue types to assess whether there is an enrichment of likely causal GWAS hits in regulatory features near our putatively causative genes. Both proximity (genomic distance) and specificity (ABD scores) were considered to determine the regulatory contribution of the fine-mapped hits.

Next, we looked at the DNase I hypersensitive sites (DHSs) which are considered to be generic markers of the regulatory DNA and can contain genetic variations associated with traits and diseases[39]. We downloaded the index of human DHS along with biosample metadata from https://www.meuleman.org/research/dhsindex/. The index was in hg38 coordinates which were converted to hg19 coordinates using the online version of the hgLiftOver package (https://genome.ucsc.edu/cgi-bin/hgLiftOver). We created a DHS index for each tissue type relevant to the traits and diseases we analyzed by including all DHS that are present in at least one bio-sample from a certain tissue type (**Supp. Table 2**). We then selected DHS that lie within +/- 100 kb of the TSS of our putatively causative genes. Since DHS are of variable widths, we recentered the summits in a 350 bp window and merged overlapping sites in the same way as we did for other chromatin marks. We calculated the mean activity (A_F) by averaging the strengths of all the merged sites. Next, we calculated the activity by distance score for each DHS and gene pair using the same formula described above. Finally, for each fine-mapped SNP, we identified all DHS that fall within +/- 100 kb of the SNP.

Finally, we used in-silico chromatin state predictions (chromHMM core 15-state model[37]) for relevant tissue types (**Supp. Table 1**) to identify active enhancer regions in the genome. Tissue-specific chromHMM annotations were downloaded from the Roadmap Epigenomics Project[37] ftp site (https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/). We considered a fine-mapped variant to fall in an enhancer region if it was housed within a chromHMM segment described as either *enhancer*, or *bivalent enhancer*, or *genic enhancer*. Since chromHMM annotations are not accompanied by activity measurements, the ABD approach could not be applied here.

References

1. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
2. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).
3. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
4. 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).
5. 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).
6. Consortium, T. Gte. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
7. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
8. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
9. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. *Limited overlap of eQTLs and*

GWAS hits due to systematic differences in discovery. http://biorxiv.org/lookup/doi/10.1101/2022.05.07.491045 (2022) doi:10.1101/2022.05.07.491045.

10. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
11. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, e1004383 (2014).
12. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
13. Giambartolomei, C. *et al.* A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).
14. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
15. Weiner, D. J., Gazal, S., Robinson, E. B. & O'Connor, L. J. Partitioning gene-mediated disease heritability without eQTLs. *Am. J. Hum. Genet.* **0**, (2022).
16. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
17. Schneider, W. J. *et al.* Familial dysbetalipoproteinemia. Abnormal binding of mutant apoprotein E to low density lipoprotein receptors of human fibroblasts and membranes from liver and adrenal of rats, rabbits, and cows. *J. Clin. Invest.* **68**, 1075–1085 (1981).
18. Goldstein, J. L. & Brown, M. S. Familial Hypercholesterolemia: Identification of a Defect in the Regulation of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Activity Associated with Overproduction of Cholesterol. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 2804–2808 (1973).
19. Cenarro, A. *et al.* The p.Leu167del Mutation in APOE Gene Causes Autosomal Dominant Hypercholesterolemia by Down-regulation of LDL Receptor Expression in

Hepatocytes. *J. Clin. Endocrinol. Metab.* **101**, 2113–2121 (2016).

20. Shimano, H. *et al.* Overexpression of apolipoprotein E in transgenic mice: marked reduction in plasma lipoproteins except high density lipoprotein and resistance against diet-induced hypercholesterolemia. *Proc. Natl. Acad. Sci.* **89**, 1750–1754 (1992).

21. Shimano, H. *et al.* Plasma lipoprotein metabolism in transgenic mice overexpressing apolipoprotein E. Accelerated clearance of lipoproteins containing apolipoprotein B. *J. Clin. Invest.* **90**, 2084–2091 (1992).

22. Kawashiri, M. *et al.* Effects of coexpression of the LDL receptor and apoE on cholesterol metabolism and atherosclerosis in LDL receptor-deficient mice. *J. Lipid Res.* **42**, 943–950 (2001).

23. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

24. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).

25. Goyette, P. *et al.* High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat. Genet.* **47**, 172–179 (2015).

26. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **52**, 572–581 (2020).

27. 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).

28. 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 (2012).

29. Freund, M. K. *et al.* Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits. *Am. J. Hum. Genet.* **103**, 535–552 (2018).

30. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).

31. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genet.* **13**, e1006646 (2017).

32. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

33. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).

34. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

35. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017).

36. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).

37. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).

38. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).

39. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).

40. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human

epigenomes. *Nature* **518**, 317–330 (2015).

41. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

42. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).

43. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).

44. Baird, D. A. *et al.* Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLOS Genet.* **17**, e1009224 (2021).

45. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends Genet.* (2020) doi:10.1016/j.tig.2020.08.009.

46. Soria, L. F. *et al.* Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. *Proc. Natl. Acad. Sci.* **86**, 587–591 (1989).

47. Pullinger, C. R. *et al.* Familial ligand-defective apolipoprotein B. Identification of a new mutation that decreases LDL receptor binding affinity. *J. Clin. Invest.* **95**, 1225–1234 (1995).

48. Hegele, R. A. *et al.* An apolipoprotein CII mutation, CIILys19----Thr' identified in patients with hyperlipidemia. *Dis. Markers* **9**, 73–80 (1991).

49. de Knijff, P., van den Maagdenberg, A. M. J. M., Frants, R. R. & Havekes, L. M. Genetic heterogeneity of apolipoprotein E and its influence on plasma lipid and lipoprotein levels. *Hum. Mutat.* **4**, 178–194 (1994).

50. Brown, M. S. & Goldstein, J. L. Analysis of a mutant strain of human fibroblasts with a defect in the internalization of receptor-bound low density lipoprotein. *Cell* **9**, 663–674 (1976).

51. Heizmann, C. *et al.* DNA polymorphism haplotypes of the human lipoprotein lipase gene: possible association with high density lipoprotein levels. *Hum. Genet.* **86**, 578–584 (1991).

52. Clee, S., Loubser, O., Collins, J., Kastelein, J. & Hayden, M. The LPL S447X cSNP is associated with decreased blood pressure and plasma triglycerides, and reduced risk of

coronary artery disease. *Clin. Genet.* **60**, 293–300 (2001).

53. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).

54. Brooks-Wilson, A. *et al.* Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat. Genet.* **22**, 336–345 (1999).

55. Bodzioch, M. *et al.* The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat. Genet.* **22**, 347–351 (1999).

56. Rust, S. *et al.* Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat. Genet.* **22**, 352–355 (1999).

57. Ordovas, J. M. *et al.* Apolipoprotein A-I Gene Polymorphism Associated with Premature Coronary Artery Disease and Familial Hypoalphalipoproteinemia. *http://dx.doi.org/10.1056/NEJM198603133141102* https://www.nejm.org/doi/10.1056/NEJM198603133141102 (1986) doi:10.1056/NEJM198603133141102.

58. Glueck, C. J., Fallat, R. W., Millett, F. & Steiner, P. M. Familial Hyperalphalipoproteinemia. *Arch. Intern. Med.* **135**, 1025–1028 (1975).

59. Isaacs, A., Sayed-Tabatabaei, F. A., Njajou, O. T., Witteman, J. C. M. & van Duijn, C. M. The −514 C→T Hepatic Lipase Promoter Region Polymorphism and Plasma Lipids: A Meta-Analysis. *J. Clin. Endocrinol. Metab.* **89**, 3858–3863 (2004).

60. Grarup, N. *et al.* The −250G>A Promoter Variant in Hepatic Lipase Associates with Elevated Fasting Serum High-Density Lipoprotein Cholesterol Modulated by Interaction with Physical Activity in a Study of 16,156 Danish Subjects. *J. Clin. Endocrinol. Metab.* **93**, 2294–2299 (2008).

61. Iijima, H. *et al.* Association of an intronic haplotype of the LIPC gene with hyperalphalipoproteinemia in two independent populations. *J. Hum. Genet.* **53**, 193–200 (2008).

62. Yamakawa-Kobayashi, K., Yanagi, H., Endo, K., Arinami, T. & Hamaguchi, H. Relationship between serum HDL-C levels and common genetic variants of the endothelial lipase gene in Japanese school-aged children. *Hum. Genet.* **113**, 311–315 (2003).

63. Jiang, X. *et al.* Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. *J. Clin. Invest.* **103**, 907–914 (1999).

64. Tai, E. *et al.* Polymorphisms at the SRBI locus are associated with lipoprotein levels in subjects with heterozygous familial hypercholesterolemia. *Clin. Genet.* **63**, 53–58 (2003).

65. McCarthy, J. J. *et al.* Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. *J. Med. Genet.* **40**, 453–458 (2003).

66. Stránecký, V. *et al.* Mutations in ANTXR1 Cause GAPO Syndrome. *Am. J. Hum. Genet.* **92**, 792–799 (2013).

67. Bayram, Y. *et al.* Whole exome sequencing identifies three novel mutations in ANTXR1 in families with GAPO syndrome. *Am. J. Med. Genet. A.* **164**, 2328–2334 (2014).

68. O'Driscoll, M., Ruiz-Perez, V. L., Woods, C. G., Jeggo, P. A. & Goodship, J. A. A splicing mutation affecting expression of ataxia–telangiectasia and Rad3–related protein (ATR) results in Seckel syndrome. *Nat. Genet.* **33**, 497–501 (2003).

69. Ogi, T. *et al.* Identification of the First ATRIP–Deficient Patient and Novel Mutations in ATR Define a Clinical Spectrum for ATR–ATRIP Seckel Syndrome. *PLOS Genet.* **8**, e1002945 (2012).

70. Ellis, N. A. *et al.* The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**, 655–666 (1995).

71. Foucault, F. *et al.* Characterization of a New BLM Mutation Associated with a Topoisomerase IIα Defect in a Patient with Bloom's Syndrome. *Hum. Mol. Genet.* **6**, 1427–1434 (1997).

72. Bicknell, L. S. *et al.* Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat. Genet.* **43**, 356–359 (2011).

73. Guernsey, D. L. *et al.* Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. *Nat. Genet.* **43**, 360–364 (2011).

74. Al-Dosari, M. S., Shaheen, R., Colak, D. & Alkuraya, F. S. Novel CENPJ mutation causes Seckel syndrome. *J. Med. Genet.* **47**, 411–414 (2010).

75. Wallis, G. A., Starman, B. J., Zinn, A. B. & Byers, P. H. Variable expression of osteogenesis imperfecta in a nuclear family is explained by somatic mosaicism for a lethal point mutation in the alpha 1(I) gene (COL1A1) of type I collagen in a parent. *Am. J. Hum. Genet.* **46**, 1034–1040 (1990).

76. Spotila, L. D., Sereda, L. & Prockop, D. J. Partial isodisomy for maternal chromosome 7 and short stature in an individual with a mutation at the COL1A2 locus. *Am. J. Hum. Genet.* **51**, 1396–1405 (1992).

77. Paepe, A. D., Nuytinck, L., Raes, M. & Fryns, J.-P. Homozygosity by descent for a COL1A2 mutation in two sibs with severe osteogenesis imperfecta and mild clinical expression in the heterozygotes. *Hum. Genet.* **99**, 478–483 (1997).

78. Briggs, M. D. *et al.* Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. *Nat. Genet.* **10**, 330–336 (1995).

79. Mabuchi, A. *et al.* Novel types of COMP mutations and genotype-phenotype association in pseudoachondroplasia and multiple epiphyseal dysplasia. *Hum. Genet.* **112**, 84–90 (2003).

80. Menke, L. A. *et al.* CREBBP mutations in individuals without Rubinstein–Taybi syndrome phenotype. *Am. J. Med. Genet. A.* **170**, 2681–2693 (2016).

81. Menke, L. A. *et al.* Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein–Taybi syndrome. *Am. J. Med. Genet. A.* **176**, 862–876 (2018).

82. Angius, A. *et al.* Confirmation of a new phenotype in an individual with a variant in the last part of exon 30 of CREBBP. *Am. J. Med. Genet. A.* **179**, 634–638 (2019).

83. Shaheen, R. *et al.* Genomic analysis of primordial dwarfism reveals novel disease

genes. *Genome Res.* **24**, 291–299 (2014).

84. Woods, S. A. *et al.* Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap cornelia de lange syndrome. *Am. J. Med. Genet. A.* **164**, 251–258 (2014).

85. Tsai, A. C.-H. *et al.* Exon deletions of the EP300 and CREBBP genes in two children with Rubinstein–Taybi syndrome detected by aCGH. *Eur. J. Hum. Genet.* **19**, 43–49 (2011).

86. Polymeropoulos, M. H. *et al.* The Gene for the Ellis–van Creveld Syndrome Is Located on Chromosome 4p16. *Genomics* **35**, 1–5 (1996).

87. Ruiz-Perez, V. L. *et al.* Mutations in Two Nonhomologous Genes in a Head-to-Head Configuration Cause Ellis-van Creveld Syndrome. *Am. J. Hum. Genet.* **72**, 728–732 (2003).

88. Galdzicka, M. *et al.* A new gene, EVC2, is mutated in Ellis–van Creveld syndrome. *Mol. Genet. Metab.* **77**, 291–295 (2002).

89. Faivre, L. *et al.* In frame fibrillin-1 gene deletion in autosomal dominant Weill-Marchesani syndrome. *J. Med. Genet.* **40**, 34–36 (2003).

90. Le Goff, C. *et al.* Mutations in the TGFβ Binding-Protein-Like Domain 5 of FBN1 Are Responsible for Acromicric and Geleophysic Dysplasias. *Am. J. Hum. Genet.* **89**, 7–14 (2011).

91. Horn, D. & Robinson, P. N. Progeroid facial features and lipodystrophy associated with a novel splice site mutation in the final intron of the FBN1 gene. *Am. J. Med. Genet. A.* **155**, 721–724 (2011).

92. Takenouchi, T. *et al.* Severe congenital lipodystrophy and a progeroid appearance: Mutation in the penultimate exon of FBN1 causing a recognizable phenotype. *Am. J. Med. Genet. A.* **161**, 3057–3062 (2013).

93. Hyland, V. J. *et al.* Somatic and germline mosaicism for a R248C missense mutation in FGFR3, resulting in a skeletal dysplasia distinct from thanatophoric dysplasia. *Am. J. Med. Genet. A.* **120A**, 157–168 (2003).

94. Toydemir, R. M. *et al.* A Novel Mutation in FGFR3 Causes Camptodactyly, Tall Stature, and Hearing Loss (CATSHL) Syndrome. *Am. J. Hum. Genet.* **79**, 935–941 (2006).

95. Makrythanasis, P. *et al.* A Novel Homozygous Mutation in FGFR3 Causes Tall Stature, Severe Lateral Tibial Deviation, Scoliosis, Hearing Impairment, Camptodactyly, and Arachnodactyly. *Hum. Mutat.* **35**, 959–963 (2014).

96. Alanay, Y. *et al.* Mutations in the Gene Encoding the RER Protein FKBP65 Cause Autosomal-Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* **86**, 551–559 (2010).

97. Kelley, B. P. *et al.* Mutations in FKBP10 cause recessive osteogenesis imperfecta and bruck syndrome. *J. Bone Miner. Res.* **26**, 666–672 (2011).

98. Barnes, A. M. *et al.* Kuskokwim Syndrome, a Recessive Congenital Contracture Disorder, Extends the Phenotype of FKBP10 Mutations. *Hum. Mutat.* **34**, 1279–1288 (2013).

99. Berg, M. A. *et al.* Diverse growth hormone receptor gene mutations in Laron syndrome. *Am. J. Hum. Genet.* **52**, 998–1005 (1993).

100. Woods, K. A., Fraser, N. C., Postel-Vinay, M. C., Savage, M. O. & Clark, A. J. A homozygous splice site mutation affecting the intracellular domain of the growth hormone (GH) receptor resulting in Laron syndrome with elevated GH-binding protein. *J. Clin. Endocrinol. Metab.* **81**, 1686–1690 (1996).

101. Goddard, A. D. *et al.* Mutations of the Growth Hormone Receptor in Children with Idiopathic Short Stature. *http://dx.doi.org/10.1056/NEJM199510263331701* https://www.nejm.org/doi/10.1056/NEJM199510263331701 (1995) doi:10.1056/NEJM199510263331701.

102. Ayling, R. M. *et al.* A dominant-negative mutation of the growth hormone receptor causes familial short stature. *Nat. Genet.* **16**, 13–14 (1997).

103. Aoki, Y. *et al.* Germline mutations in HRAS proto-oncogene cause Costello syndrome. *Nat. Genet.* **37**, 1038–1040 (2005).

104. Schubbert, S. *et al.* Germline KRAS mutations cause Noonan syndrome. *Nat. Genet.* **38**,

331–336 (2006).

105. Carta, C. *et al.* Germline Missense Mutations Affecting KRAS Isoform B Are Associated with a Severe Noonan Syndrome Phenotype. *Am. J. Hum. Genet.* **79**, 129–135 (2006).

106. Varon, R. *et al.* Nibrin, a Novel DNA Double-Strand Break Repair Protein, Is Mutated in Nijmegen Breakage Syndrome. *Cell* **93**, 467–476 (1998).

107. Tanzarella, C. *et al.* Chromosome instability and nibrin protein variants in NBS heterozygotes. *Eur. J. Hum. Genet.* **11**, 297–303 (2003).

108. Tonkin, E. T., Wang, T.-J., Lisgo, S., Bamshad, M. J. & Strachan, T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36**, 636–641 (2004).

109. Krantz, I. D. *et al.* Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat. Genet.* **36**, 631–635 (2004).

110. Bicknell, L. S. *et al.* Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat. Genet.* **43**, 350–355 (2011).

111. de Munnik, S. A. *et al.* Meier–Gorlin syndrome genotype–phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. *Eur. J. Hum. Genet.* **20**, 598–606 (2012).

112. Rauch, A. *et al.* Mutations in the Pericentrin (PCNT) Gene Cause Primordial Dwarfism. *Science* **319**, 816–819 (2008).

113. Griffith, E. *et al.* Mutations in pericentrin cause Seckel syndrome with defective ATR-dependent DNA damage signaling. *Nat. Genet.* **40**, 232–236 (2008).

114. Piane, M. *et al.* Majewski osteodysplastic primordial dwarfism type II (MOPD II) syndrome previously diagnosed as Seckel syndrome: Report of a novel mutation of the PCNT gene. *Am. J. Med. Genet. A.* **149A**, 2452–2456 (2009).

115. van der Slot, A. J. *et al.* Identification of PLOD2 as Telopeptide Lysyl Hydroxylase, an

Important Enzyme in Fibrosis*. *J. Biol. Chem.* **278**, 40967–40972 (2003).

116. Ha-Vinh, R. *et al.* Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. *Am. J. Med. Genet. A.* **131A**, 115–120 (2004).

117. Puig-Hervás, M. T. *et al.* Mutations in PLOD2 cause autosomal-recessive connective tissue disorders within the Bruck syndrome—Osteogenesis imperfecta phenotypic spectrum. *Hum. Mutat.* **33**, 1444–1449 (2012).

118. Tartaglia, M. *et al.* Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **29**, 465–468 (2001).

119. Maheshwari, M. *et al.* PTPN11 Mutations in Noonan syndrome type I: detection of recurrent mutations in exons 3 and 13. *Hum. Mutat.* **20**, 298–304 (2002).

120. Kosaki, K. *et al.* PTPN11 (Protein-Tyrosine Phosphatase, Nonreceptor-Type 11) Mutations in Seven Japanese Patients with Noonan Syndrome. *J. Clin. Endocrinol. Metab.* **87**, 3529–3533 (2002).

121. Deardorff, M. A. *et al.* RAD21 Mutations Cause a Human Cohesinopathy. *Am. J. Hum. Genet.* **90**, 1014–1027 (2012).

122. Kruszka, P. *et al.* Cohesin complex-associated holoprosencephaly. *Brain* **142**, 2631–2643 (2019).

123. Goel, H. & Parasivam, G. Another case of holoprosencephaly associated with RAD21 loss-of-function variant. *Brain* **143**, e64 (2020).

124. Pandit, B. *et al.* Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. *Nat. Genet.* **39**, 1007–1012 (2007).

125. Razzaque, M. A. *et al.* Germline gain-of-function mutations in RAF1 cause Noonan syndrome. *Nat. Genet.* **39**, 1013–1017 (2007).

126. Lindor, N. M. *et al.* Rothmund-Thomson syndrome due to RECQ4 helicase mutations: Report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome.

Am. J. Med. Genet. **90**, 223–228 (2000).

127. Beghini, A., Castorina, P., Roversi, G., Modiano, P. & Larizza, L. RNA processing defects of the helicase gene RECQL4 in a compound heterozygous Rothmund–Thomson patient. *Am. J. Med. Genet. A.* **120A**, 395–399 (2003).

128. Wang, L. L. *et al.* Association Between Osteosarcoma and Deleterious Mutations in the RECQL4 Gene in Rothmund–Thomson Syndrome. *JNCI J. Natl. Cancer Inst.* **95**, 669–674 (2003).

129. Aoki, Y. *et al.* Gain-of-Function Mutations in RIT1 Cause Noonan Syndrome, a RAS/MAPK Pathway Syndrome. *Am. J. Hum. Genet.* **93**, 173–180 (2013).

130. Bertola, D. R. *et al.* Further evidence of the importance of RIT1 in Noonan syndrome. *Am. J. Med. Genet. A.* **164**, 2952–2957 (2014).

131. Gos, M. *et al.* Contribution of RIT1 mutations to the pathogenesis of Noonan syndrome: Four new cases and further evidence of heterogeneity. *Am. J. Med. Genet. A.* **164**, 2310–2316 (2014).

132. Afzal, A. R. *et al.* Recessive Robinow syndrome, allelic to dominant brachydactyly type B, is caused by mutation of ROR2. *Nat. Genet.* **25**, 419–422 (2000).

133. van Bokhoven, H. *et al.* Mutation of the gene encoding the ROR2 tyrosine kinase causes autosomal recessive Robinow syndrome. *Nat. Genet.* **25**, 423–426 (2000).

134. Tufan, F. *et al.* Clinical and molecular characterization of two adults with autosomal recessive Robinow syndrome. *Am. J. Med. Genet. A.* **136A**, 185–189 (2005).

135. Hästbacka, J., Salonen, R., Laurila, P., Chapelle, A. de la & Kaitila, I. Prenatal diagnosis of diastrophic dysplasia with polymorphic DNA markers. *J. Med. Genet.* **30**, 265–268 (1993).

136. Rossi, A. & Superti-Furga, A. Mutations in the diastrophic dysplasia sulfate transporter (DTDST) gene (SLC26A2): 22 novel mutations, mutation review, associated skeletal phenotypes, and diagnostic relevance. *Hum. Mutat.* **17**, 159–171 (2001).

137. Barreda-Bonis, A. C. *et al.* Multiple SLC26A2 mutations occurring in a three-generational

family. *Eur. J. Med. Genet.* **61**, 24–28 (2018).

138. Le Goff, C. *et al.* Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nat. Genet.* **44**, 85–88 (2012).

139. Caputo, V. *et al.* A Restricted Spectrum of Mutations in the SMAD4 Tumor-Suppressor Gene Underlies Myhre Syndrome. *Am. J. Hum. Genet.* **90**, 161–169 (2012).

140. Lindor, N. M., Gunawardena, S. R. & Thibodeau, S. N. Mutations of SMAD4 account for both LAPS and Myhre syndromes. *Am. J. Med. Genet. A.* **158A**, 1520–1521 (2012).

141. Hood, R. L. *et al.* Mutations in SRCAP, Encoding SNF2-Related CREBBP Activator Protein, Cause Floating-Harbor Syndrome. *Am. J. Hum. Genet.* **90**, 308–313 (2012).

142. Goff, C. L. *et al.* Not All Floating-Harbor Syndrome Cases are Due to Mutations in Exon 34 of SRCAP. *Hum. Mutat.* **34**, 88–92 (2013).

143. Yu, C.-E. *et al.* Positional Cloning of the Werner's Syndrome Gene. *Science* **272**, 258–262 (1996).

144. Goto, M. *et al.* Analysis of helicase gene mutations in Japanese Werner's syndrome patients. *Hum. Genet.* **99**, 191–193 (1997).

145. Yu, C.-E. *et al.* Mutations in the Consensus Helicase Domains of the Werner Syndrome Gene. *Am J Hum Genet* 12 (1997).

146. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* **39**, 207–211 (2007).

147. Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* **43**, 1066–1073 (2011).

148. Fowler, E. V. *et al.* TNFα and IL10 SNPs act together to predict disease behaviour in Crohn's disease. *J. Med. Genet.* **42**, 523–528 (2005).

149. Gasche, C. *et al.* Novel Variants of the IL-10 Receptor 1 Affect Inhibition of Monocyte TNF-α Production. *J. Immunol.* **170**, 5578–5582 (2003).

150. Mao, H. *et al.* Exome sequencing identifies novel compound heterozygous mutations of

IL-10 receptor 1 in neonatal-onset Crohn's disease. *Genes Immun.* **13**, 437–442 (2012).

151. Glocker, E.-O. *et al.* Inflammatory Bowel Disease and Mutations Affecting the Interleukin-10 Receptor. *http://dx.doi.org/10.1056/NEJMoa0907206* https://www.nejm.org/doi/10.1056/NEJMoa0907206 (2009) doi:10.1056/NEJMoa0907206.

152. Begue, B. *et al.* Defective IL10 Signaling Defining a Subgroup of Patients With Inflammatory Bowel Disease. *Off. J. Am. Coll. Gastroenterol. ACG* **106**, 1544–1555 (2011).

153. Duerr, R. H. *et al.* A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* **314**, 1461–1463 (2006).

154. Libioulle, C. *et al.* Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of PTGER4. *PLOS Genet.* **3**, e58 (2007).

155. Glas, J. *et al.* rs1004819 Is the Main Disease-Associated IL23R Variant in German Crohn's Disease Patients: Combined Analysis of IL23R, CARD15, and OCTN1/2 Variants. *PLOS ONE* **2**, e819 (2007).

156. McCarroll, S. A. *et al.* Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat. Genet.* **40**, 1107–1112 (2008).

157. Craddock, N. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–720 (2010).

158. Prescott, N. J. *et al.* Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. *Hum. Mol. Genet.* **19**, 1828–1839 (2010).

159. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).

160. Hugot, J.-P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).

161. Ellinghaus, D. *et al.* Association Between Variants of PRDM1 and NDP52 and Crohn's Disease, Based on Exome Sequencing and Functional Studies. *Gastroenterology* **145**, 339–

347 (2013).

162. Diaz-Gallo, L.-M. *et al.* Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis. *Inflamm. Bowel Dis.* **17**, 2287–2294 (2011).

163. Fowler, E. V. *et al.* ATG16L1 T300A Shows Strong Associations With Disease Subgroups in a Large Australian IBD Population: Further Support for Significant Disease Heterogeneity. *Off. J. Am. Coll. Gastroenterol. ACG* **103**, 2519–2526 (2008).

164. Fisher, S. A. *et al.* Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat. Genet.* **40**, 710–712 (2008).

165. Beaudoin, M. *et al.* Deep Resequencing of GWAS Loci Identifies Rare Variants in CARD9, IL23R and RNF186 That Are Associated with Ulcerative Colitis. *PLOS Genet.* **9**, e1003723 (2013).

166. Rivas, M. A. *et al.* A protein-truncating R179X variant in RNF186 confers protection against ulcerative colitis. *Nat. Commun.* **7**, 12342 (2016).

167. Reis, A. F. *et al.* Association of a variant in exon 31 of the sulfonylurea receptor 1 (SUR1) gene with type 2 diabetes mellitus in French Caucasians. *Hum. Genet.* **107**, 138–144 (2000).

168. Borowiec, M. *et al.* Mutations at the BLK locus linked to maturity onset diabetes of the young and β-cell dysfunction. *Proc. Natl. Acad. Sci.* **106**, 14460–14465 (2009).

169. Bengtsson-Ellmark, S. H. *et al.* Association between a polymorphism in the carboxyl ester lipase gene and serum cholesterol profile. *Eur. J. Hum. Genet.* **12**, 627–632 (2004).

170. Ræder, H. *et al.* Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat. Genet.* **38**, 54–62 (2006).

171. Harding, H. P. *et al.* Diabetes Mellitus and Exocrine Pancreatic Dysfunction in Perk−/− Mice Reveals a Role for Translational Control in Secretory Cell Survival. *Mol. Cell* **7**, 1153–1163 (2001).

172. Brickwood, S. *et al.* Wolcott-Rallison syndrome: pathogenic insights into neonatal

diabetes from new mutation and expression studies of EIF2AK3. *J. Med. Genet.* **40**, 685–689 (2003).

173. Durocher, F. *et al.* A novel mutation in the EIF2AK3 gene with variable expressivity in two patients with Wolcott–Rallison syndrome. *Clin. Genet.* **70**, 34–38 (2006).

174. Shaw-Smith, C. *et al.* GATA4 Mutations Are a Cause of Neonatal and Childhood-Onset Diabetes. *Diabetes* **63**, 2888–2894 (2014).

175. Yorifuji, T. *et al.* Dominantly inherited diabetes mellitus caused by GATA6 haploinsufficiency: variable intrafamilial presentation. *J. Med. Genet.* **49**, 642–643 (2012).

176. Franco, E. D. *et al.* GATA6 Mutations Cause a Broad Phenotypic Spectrum of Diabetes From Pancreatic Agenesis to Adult-Onset Diabetes Without Exocrine Insufficiency. *Diabetes* **62**, 993–997 (2013).

177. Froguel, P. *et al.* Familial Hyperglycemia Due to Mutations in Glucokinase -- Definition of a Subtype of Diabetes Mellitus. *http://dx.doi.org/10.1056/NEJM199303113281005* https://www.nejm.org/doi/10.1056/NEJM199303113281005 (1993) doi:10.1056/NEJM199303113281005.

178. Senée, V. *et al.* Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat. Genet.* **38**, 682–687 (2006).

179. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-1α gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455–458 (1996).

180. Vaxillaire, M. *et al.* Identification of Nine Novel Mutations in the Hepatocyte Nuclear Factor 1 Alpha Gene Associated with Maturity-Onset Diabetes of the Young (MODY3). *Hum. Mol. Genet.* **6**, 583–586 (1997).

181. Horikawa, Y. *et al.* Mutation in hepatocyte nuclear factor–1β gene (TCF2) associated with MODY. *Nat. Genet.* **17**, 384–385 (1997).

182. Lindner, T. H. *et al.* A Novel Syndrome of Diabetes Mellitus, Renal Dysfunction and Genital Malformation Associated with a Partial Deletion of the Pseudo-POU Domain of

Hepatocyte Nuclear Factor-1β. *Hum. Mol. Genet.* **8**, 2001–2008 (1999).

183. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-4α gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458–460 (1996).

184. Stoffel, M. & Duncan, S. A. The maturity-onset diabetes of the young (MODY1) transcription factor HNF4α regulates expression of genes required for glucose transport and metabolism. *Proc. Natl. Acad. Sci.* **94**, 13209–13214 (1997).

185. Poulton, C. J. *et al.* Microcephaly with Simplified Gyration, Epilepsy, and Infantile Diabetes Linked to Inappropriate Apoptosis of Neural Progenitors. *Am. J. Hum. Genet.* **89**, 265–276 (2011).

186. Abdel-Salam, G. M. H. *et al.* A homozygous IER3IP1 mutation causes microcephaly with simplified gyral pattern, epilepsy, and permanent neonatal diabetes syndrome (MEDS). *Am. J. Med. Genet. A.* **158A**, 2788–2796 (2012).

187. Shalev, S. A. *et al.* Microcephaly, epilepsy, and neonatal diabetes due to compound heterozygous mutations in IER3IP1: insights into the natural history of a rare disorder. *Pediatr. Diabetes* **15**, 252–256 (2014).

188. Støy, J. *et al.* Insulin gene mutations as a cause of permanent neonatal diabetes. *Proc. Natl. Acad. Sci.* **104**, 15040–15044 (2007).

189. Hani, E. H. *et al.* Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* **41**, 1511–1515 (1998).

190. Gloyn, A. L. *et al.* Activating Mutations in the Gene Encoding the ATP-Sensitive Potassium-Channel Subunit Kir6.2 and Permanent Neonatal Diabetes. *http://dx.doi.org/10.1056/NEJMoa032922* https://www.nejm.org/doi/10.1056/NEJMoa032922 (2004) doi:10.1056/NEJMoa032922.

191. Neve, B. *et al.* Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. *Proc. Natl. Acad. Sci.* **102**, 4807–4812 (2005).

192. Cao, H. & Hegele, R. A. Nuclear lamin A/C R482Q mutation in Canadian kindreds with Dunnigan-type familial partial lipodystrophy. *Hum. Mol. Genet.* **9**, 109–112 (2000).

193. Malecki, M. T. *et al.* Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat. Genet.* **23**, 323–328 (1999).

194. Gradwohl, G., Dierich, A., LeMeur, M. & Guillemot, F. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci.* **97**, 1607–1611 (2000).

195. Rubio-Cabezas, O. *et al.* Permanent Neonatal Diabetes and Enteric Anendocrinosis Associated With Biallelic Mutations in NEUROG3. *Diabetes* **60**, 1349–1353 (2011).

196. Pinney, S. E. *et al.* Neonatal Diabetes and Congenital Malabsorptive Diarrhea Attributable to a Novel Mutation in the Human Neurogenin-3 Gene Coding Sequence. *J. Clin. Endocrinol. Metab.* **96**, 1960–1965 (2011).

197. Shimajiri, Y. *et al.* A Missense Mutation of Pax4 Gene (R121W) Is Associated With Type 2 Diabetes in Japanese. *Diabetes* **50**, 2864–2869 (2001).

198. Mauvais-Jarvis, F. *et al.* PAX4 gene variations predispose to ketosis-prone diabetes. *Hum. Mol. Genet.* **13**, 3151–3159 (2004).

199. Plengvidhya, N. *et al.* PAX4 Mutations in Thais with Maturity Onset Diabetes of the Young. *J. Clin. Endocrinol. Metab.* **92**, 2821–2826 (2007).

200. Staffers, D. A., Ferrer, J., Clarke, W. L. & Habener, J. F. Early-onset type-ll diabetes mellitus (MODY4) linked to IPF1. *Nat. Genet.* **17**, 138–139 (1997).

201. Macfarlane, W. M. *et al.* Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. *J. Clin. Invest.* **104**, R33–R39 (1999).

202. Hani, E. H. *et al.* Defective mutations in the insulin promoter factor-1 (*IPF-1*) gene in late-onset type 2 diabetes mellitus. *J. Clin. Invest.* **104**, R41–R48 (1999).

203. Deeb, S. S. *et al.* A Pro12Ala substitution in PPARγ2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat. Genet.* **20**,

284–287 (1998).

204. Savage, D. B. *et al.* Digenic inheritance of severe insulin resistance in a human pedigree. *Nat. Genet.* **31**, 379–384 (2002).

205. Sellick, G. S. *et al.* Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat. Genet.* **36**, 1301–1305 (2004).

206. Smith, S. B. *et al.* Rfx6 directs islet formation and insulin production in mice and humans. *Nature* **463**, 775–780 (2010).

207. Sansbury, F. H. *et al.* Biallelic RFX6 mutations can cause childhood as well as neonatal onset diabetes mellitus. *Eur. J. Hum. Genet.* **23**, 1744–1748 (2015).

208. Labay, V. *et al.* Mutations in SLC19A2 cause thiamine-responsive megaloblastic anaemia associated with diabetes mellitus and deafness. *Nat. Genet.* **22**, 300–304 (1999).

209. Oishi, K. *et al.* Targeted disruption of Slc19a2, the gene encoding the high-affinity thiamin transporter Thtr-1, causes diabetes mellitus, sensorineural deafness and megaloblastosis in mice. *Hum. Mol. Genet.* **11**, 2951–2960 (2002).

210. Shaw-Smith, C. *et al.* Recessive SLC19A2 mutations are a cause of neonatal diabetes mellitus in thiamine-responsive megaloblastic anaemia. *Pediatr. Diabetes* **13**, 314–321 (2012).

211. Laukkanen, O. *et al.* Polymorphisms in the SLC2A2 (GLUT2) Gene Are Associated With the Conversion From Impaired Glucose Tolerance to Type 2 Diabetes : The Finnish Diabetes Prevention Study. *Diabetes* **54**, 2256–2260 (2005).

212. Sansbury, F. H. *et al.* SLC2A2 mutations can cause neonatal diabetes, suggesting GLUT2 may have a role in human insulin secretion. *Diabetologia* **55**, 2381–2385 (2012).

213. Strom, T. M. *et al.* Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy and Deafness (DIDMOAD) Caused by Mutations in a Novel Gene (Wolframin) Coding for a Predicted Transmembrane Protein. *Hum. Mol. Genet.* **7**, 2021–2028 (1998).

214. Hardy, C. *et al.* Clinical and Molecular Genetic Analysis of 19 Wolfram Syndrome

Kindreds Demonstrating a Wide Spectrum of Mutations in WFS1. *Am. J. Hum. Genet.* **65**, 1279–1290 (1999).

215. Khanim, F., Kirk, J., Latif, F. & Barrett, T. G. WFS1/wolframin mutations, Wolfram syndrome, and associated diseases. *Hum. Mutat.* **17**, 357–367 (2001).

216. Mackay, D. J. G. *et al.* Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat. Genet.* **40**, 949–951 (2008).

217. Boonen, S. E. *et al.* Transient Neonatal Diabetes, ZFP57, and Hypomethylation of Multiple Imprinted Loci: A detailed follow-up. *Diabetes Care* **36**, 505–512 (2013).

218. Dietlein, F. *et al.* Identification of cancer driver genes based on nucleotide context. *Nat. Genet.* **52**, 208–218 (2020).

219. Hukku, A. *et al.* Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am. J. Hum. Genet.* **108**, 25–35 (2021).

220. Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am. J. Hum. Genet.* **102**, 1169–1184 (2018).

221. Zhang, T. *et al.* Cell-type–specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Res.* **28**, 1621–1635 (2018).

222. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e16 (2018).

223. Glastonbury, C. A., Alves, A. C., Moustafa, J. S. E.-S. & Small, K. S. Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. *Am. J. Hum. Genet.* **104**, 1013–1024 (2019).

224. Rai, V. *et al.* Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Mol. Metab.* **32**, 109–121 (2020).

225. Findley, A. S. *et al.* Functional dynamic genetic effects on gene regulation are specific to

particular cell types and environmental conditions. *eLife* **10**, e67077 (2021).

226. Neavin, D. *et al.* Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol.* **22**, 76 (2021).

227. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006-3021.e17 (2021).

228. Patel, D. *et al.* Cell-type-specific expression quantitative trait loci associated with Alzheimer disease in blood and brain tissue. *Transl. Psychiatry* **11**, 1–17 (2021).

229. Bryois, J. *et al.* Cell-type specific cis-eQTLs in eight brain cell-types identifies novel risk genes for human brain disorders. 2021.10.09.21264604 Preprint at https://doi.org/10.1101/2021.10.09.21264604 (2021).

230. Arvanitis, M., Tayeb, K., Strober, B. J. & Battle, A. Redefining tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity. *Am. J. Hum. Genet.* **109**, 223–239 (2022).

231. Oelen, R. *et al.* Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. *Nat. Commun.* **13**, 3267 (2022).

232. Perez, R. K. *et al.* Single-cell RNA-seq reveals cell type–specific molecular and genetic associations to lupus. *Science* **376**, eabf1970 (2022).

233. Schmiedel, B. J. *et al.* Single-cell eQTL analysis of activated T cell subsets reveals activation and cell type–dependent effects of disease-risk variants. *Sci. Immunol.* **7**, eabm2508 (2022).

234. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).

235. Strober, B. J. *et al.* Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287–1290 (2019).

236. Cuomo, A. S. E. *et al.* Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).

237. Bonder, M. J. *et al.* Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nat. Genet.* **53**, 313–321 (2021).

238. Jerber, J. *et al.* Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.* **53**, 304–312 (2021).

239. Aygün, N. *et al.* Inferring cell-type-specific causal gene regulatory networks during human neurogenesis. 2022.04.25.488920 Preprint at https://doi.org/10.1101/2022.04.25.488920 (2022).

240. Elorbany, R. *et al.* Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLOS Genet.* **18**, e1009666 (2022).

241. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43**, 95–100 (2011).

242. Knowles, D. A. *et al.* Allele-specific expression reveals interactions between genetic variation and environment. *Nat. Methods* **14**, 699–702 (2017).

243. Kim-Hellmuth, S. *et al.* Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* **8**, 266 (2017).

244. Balliu, B. *et al.* An integrated approach to identify environmental modulators of genetic risk factors for complex traits. *Am. J. Hum. Genet.* **108**, 1866–1879 (2021).

245. Mu, Z. *et al.* The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol.* **22**, 122 (2021).

246. Ward, M. C., Banovich, N. E., Sarkar, A., Stephens, M. & Gilad, Y. Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. *eLife* **10**, e57345 (2021).

247. Nathan, A. *et al.* Single-cell eQTL models reveal dynamic T cell state dependence of

disease loci. *Nature* **606**, 120–128 (2022).

248. Baca, S. C. *et al.* Genetic determinants of chromatin reveal prostate cancer risk mediated by context-dependent gene regulation. *Nat. Genet.* **54**, 1364–1375 (2022).

249. Fu, J. *et al.* System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nat. Genet.* **41**, 166–167 (2009).

250. Dori-Bachash, M., Shema, E. & Tirosh, I. Coupled Evolution of Transcription and mRNA Degradation. *PLOS Biol.* **9**, e1001106 (2011).

251. Ghazalpour, A. *et al.* Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLOS Genet.* **7**, e1001393 (2011).

252. Pai, A. A. *et al.* The Contribution of RNA Decay Quantitative Trait Loci to Inter-Individual Variation in Steady-State Gene Expression Levels. *PLOS Genet.* **8**, e1003000 (2012).

253. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).

254. Khan, Z. *et al.* Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures. *Science* **342**, 1100–1104 (2013).

255. Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79–82 (2013).

256. McManus, C. J., May, G. E., Spealman, P. & Shteyman, A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res.* **24**, 422–430 (2014).

257. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).

258. Bader, D. M. *et al.* Negative feedback buffers effects of regulatory variants. *Mol. Syst. Biol.* **11**, 785 (2015).

259. Battle, A. *et al.* Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2015).

260. Cenik, C. *et al.* Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* **25**, 1610–1621 (2015).

261. McManus, J., Cheng, Z. & Vogel, C. Next-generation analysis of gene expression regulation – comparing the roles of synthesis and degradation. *Mol. Biosyst.* **11**, 2680–2689 (2015).

262. Pai, A. A., Pritchard, J. K. & Gilad, Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLOS Genet.* **11**, e1004857 (2015).

263. Schafer, S. *et al.* Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nat. Commun.* **6**, 7200 (2015).

264. Chick, J. M. *et al.* Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**, 500–505 (2016).

265. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).

266. Schaefke, B., Sun, W., Li, Y.-S., Fang, L. & Chen, W. The evolution of posttranscriptional regulation. *WIREs RNA* **9**, e1485 (2018).

267. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).

268. Wang, Z.-Y. *et al.* Transcriptome and translatome co-evolution in mammals. *Nature* 1–6 (2020) doi:10.1038/s41586-020-2899-z.

269. Kusnadi, E. P., Timpone, C., Topisirovic, I., Larsson, O. & Furic, L. Regulation of gene expression via translational buffering. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1869**, 119140 (2022).

270. Pedraza, J. M. & Paulsson, J. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* **319**, 339–343 (2008).

271. Raj, A. & Oudenaarden, A. van. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).

272. Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci.* **105**, 17256–17261 (2008).

273. Larson, D. R., Singer, R. H. & Zenklusen, D. A single molecule view of gene expression. *Trends Cell Biol.* **19**, 630–637 (2009).

274. Raj, A. & van Oudenaarden, A. Single-Molecule Approaches to Stochastic Gene Expression. *Annu. Rev. Biophys.* **38**, 255–270 (2009).

275. Suter, D. M. *et al.* Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* **332**, 472–474 (2011).

276. Dar, R. D. *et al.* Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci.* **109**, 17454–17459 (2012).

277. Viñuelas, J. *et al.* Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biol.* **11**, 15 (2013).

278. Kumar, N., Singh, A. & Kulkarni, R. V. Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. *PLOS Comput. Biol.* **11**, e1004292 (2015).

279. Nicolas, D., Phillips, N. E. & Naef, F. What shapes eukaryotic transcriptional bursting? *Mol. Biosyst.* **13**, 1280–1290 (2017).

280. Qiu, H., Zhang, B. & Zhou, T. Analytical results for a generalized model of bursty gene expression with molecular memory. *Phys. Rev. E* **100**, 012128 (2019).

281. Wang, Z., Zhang, Z. & Zhou, T. Exact distributions for stochastic models of gene expression with arbitrary regulation. *Sci. China Math.* **63**, 485–500 (2020).

282. Blair, D. R. *et al.* A Nondegenerate Code of Deleterious Variants in Mendelian Loci Contributes to Complex Disease Risk. *Cell* **155**, 70–80 (2013).

283. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).

284. Chan, Y. *et al.* Genome-wide Analysis of Body Proportion Classifies Height-Associated

Variants by Mechanism of Action and Implicates Genes Important for Skeletal Development. *Am. J. Hum. Genet.* **96**, 695–708 (2015).

285. Kathiresan, S. & Srivastava, D. Genetics of Human Cardiovascular Disease. *Cell* **148**, 1242–1257 (2012).

286. Zhu, X., Need, A. C., Petrovski, S. & Goldstein, D. B. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. *Nat. Neurosci.* **17**, 773–781 (2014).

287. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214-1231.e11 (2020).

288. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).

289. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Online Mendelian Inheritance in Man, OMIM®. (2021).

290. Deardorff, M. A. *et al.* Mutations in Cohesin Complex Members SMC3 and SMC1A Cause a Mild Variant of Cornelia de Lange Syndrome with Predominant Mental Retardation. *Am. J. Hum. Genet.* **80**, 485–494 (2007).

291. Cummings, B. B. *et al.* Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452–458 (2020).

292. Bycroft, C. *et al. Genome-wide genetic data on ~500,000 UK Biobank participants*. 166298 https://www.biorxiv.org/content/10.1101/166298v1 (2017) doi:10.1101/166298.

293. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).

294. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

295. Galinsky, K. J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).

296. Galinsky, K. J., Loh, P.-R., Mallick, S., Patterson, N. J. & Price, A. L. Population

Structure of UK Biobank and Ancient Eurasians Reveals Adaptation at Genes Influencing Blood Pressure. *Am. J. Hum. Genet.* **99**, 1130–1139 (2016).

297. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, (2021).

298. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

299. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).

300. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

301. Park, Y., Sarkar, A., Bhutani, K. & Kellis, M. *Multi-tissue polygenic models for transcriptome-wide association studies*. http://biorxiv.org/lookup/doi/10.1101/107623 (2017) doi:10.1101/107623.

302. Gamazon, E. R., Zwinderman, A. H., Cox, N. J., Denys, D. & Derks, E. M. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat. Genet.* **51**, 933–940 (2019).

Acknowledgements

We thank Alkes Price, Alex Bloemendal, Benjamin Neale, Bogdan Pasanuic, Sasha (Alexander) Gusev, and Matt Warman for their helpful discussions. This research was supported by NIH grants R35GM127131, R01HG010372, R01MH101244, and U01HG012009, . N.J.C was supported by NIH training grant T32GM74897. UK Biobank was accessed under projects 14048 and 10438. TOPMed data were used under dbGaP project 28674.