RESEARCH ARTICLE

The missing link between genetic association and regulatory function

Noah J 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 R Sunyaev[1,2,3]*

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

Abstract 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 datasets, 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 expression QTLs, suggesting that better models are needed. The field must confront this deficit and pursue this 'missing regulation.'

Editor's evaluation

The findings reported here are important because they address the issue of how complex traits arise from their genetic underpinnings. There is an assumption that genetically mediated variation in transcript abundance, usually detected via analysis of expressed quantitative trait loci, is key to this process, but we lack robust evidence in support of that view. This article finds limited evidence that the baseline expression of trait-related genes explains the associations between complex traits and genetic variants (as identified from genome-wide association studies), leading to the view that the field needs to confront a problem of 'missing regulation.'

***For correspondence:** noahconnally@g.harvard.edu (NJC); cotsapas@broadinstitute.org (CC); cassa@mit.edu (CAC); ssunyaev@hms.harvard.edu (SRS)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 17

Preprinted: 11 June 2021
Received: 25 October 2021
Accepted: 02 December 2022
Published: 14 December 2022

Reviewing Editor: Jonathan Flint, University of California, Los Angeles, United States

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 (**Manolio et al., 2009**). Further work has revealed unexpectedly high polygenicity for most human traits and very small effect sizes for individual variants. 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 (**Maurano et al., 2012**; **Trynka et al., 2013**; **Gusev et al., 2014**). Furthermore, genes in trait-associated loci are more likely to have genetic variants that affect their expression levels (expression quantitative trait loci, 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 (**Nicolae et al., 2010**). 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) (**GTEx Consortium, 2020**). Thus, a model has emerged in which most trait-associated variants influence proximal gene regulation.

Here we argue that this unembellished model—in which genome-wide association study (GWAS) peaks are mediated by the effects on the homeostatic expression assayed in tissue samples—is the exception rather than the rule. We highlight the 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 constancy 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 (**Stranger et al., 2007**; **Farh et al., 2015**; **Mostafavi et al., 2022**). Another is that expression levels mediate a minority of complex trait heritability (**Yao et al., 2020**). 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 colocalize with eQTLs (**Giambartolomei et al., 2014**; **Chun et al., 2017**; **Giambartolomei et al., 2018**; **Hormozdiari et al., 2016**).

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 (**Weiner et al., 2022**). 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 (**Backman et al., 2021**).

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 (**Schneider et al., 1981**; **Goldstein and Brown, 1973**). Coding variants in *APOE* and *LDLR* can lead to the Mendelian disorder familial hypercholesterolemia (**Goldstein and Brown, 1973**; **Cenarro et al., 2016**). Even in the absence of a Mendelian coding variant, experiments in animal models have found that the overexpression of these genes reduces cholesterol levels (**Shimano et al., 1992a**; **Shimano et al., 1992b**; **Kawashiri et al., 2001**). GWAS on human subjects have found significant associations near *APOE* and *LDLR*, so it seems reasonable to suspect that any non-coding 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 Appendix 1.

Results

To test the model that trait-associated variants influence baseline gene expression, we assembled a list of putatively causative genes for 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 (T2D) (**Mahajan et al., 2018**), 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 (UC) and Crohn disease (CD) (**Liu et al., 2015**; **Goyette et al., 2015**), which have Mendelian pediatric forms characterized by severity of presentation; and breast cancer (BC) (**Zhang et al., 2020**), 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. Between known Mendelian genes and those from **Backman et al., 2021**, our analysis included 220 unique gene-trait pairs (**Figure 1**).

In well-powered GWAS, even relatively rare large-effect coding alleles (mutations in *BRCA1* that 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 (**Yang et al., 2011**; **Yang et al., 2012**; 'Materials and 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 (**Freund et al., 2018**), we observe a significant enrichment for all traits in our combined Mendelian and **Backman et al., 2021** gene sets (**Figure 1—figure supplement 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 (**Nasser et al., 2021**), 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), an order-of-magnitude enrichment (**Figure 1—figure supplement 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.

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., 2021*** are not included here, but can be found in ***Figure 2***.

Phenotype	Genes
Low-density lipoprotein	APOB ***Soria et al., 1989***; ***Pullinger et al., 1995*** APOC2 ***Hegele et al., 1991*** APOE ***de Knijff et al., 1994*** LDLR ***Brown and Goldstein, 1976*** LPL ***Heizmann et al., 1991***; ***Clee et al., 2001*** PCSK9 ***Abifadel et al., 2003***
High-density lipoprotein	ABCA1 ***Brooks-Wilson et al., 1999***; ***Bodzioch et al., 1999***; ***Rust et al., 1999***; ***Ordovas et al., 1986*** APOA1 ***Ordovas et al., 1986*** CETP ***Glueck et al., 1975*** LIPC ***Isaacs et al., 2004***; ***Grarup et al., 2008***; ***Iijima et al., 2008*** LIPG ***Yamakawa-Kobayashi et al., 2003*** PLTP ***Jiang et al., 1999*** SCARB1 ***Tai et al., 2003***; ***McCarthy et al., 2003***
Height	ANTXR1 ***Stránecký et al., 2013***; ***Bayram et al., 2014*** ATR ***O'Driscoll et al., 2003***; ***Ogi et al., 2012*** BLM ***Ellis et al., 1995***; ***Foucault et al., 1997*** CDC6 ***Bicknell et al., 2011a*** CDT1 ***Bicknell et al., 2011a***; ***Guernsey et al., 2011*** CENPJ ***AlDosari et al., 2010*** COL1A1 ***Wallis et al., 1990*** COL1A2 ***Spotila et al., 1992***; ***De Paepe et al., 1997*** COMP ***Briggs et al., 1995***; ***Mabuchi et al., 2003*** CREBBP ***Menke et al., 2016***; ***Menke et al., 2018***; ***Angius et al., 2019*** DNA2 ***Shaheen et al., 2014*** EP300 ***Woods et al., 2014***; ***Tsai et al., 2011*** EVC ***Polymeropoulos et al., 1996***; ***Ruiz-Perez et al., 2003*** EVC2 ***Ruiz-Perez et al., 2003***; ***Galdzicka et al., 2002*** BN1 ***Faivre et al., 2003***; ***Le Goff et al., 2011***; ***Horn and Robinson, 2011***; ***Takenouchi et al., 2013*** FGFR3 ***Hyland et al., 2003***; ***Toydemir et al., 2006***; ***Makrythanasis et al., 2014*** FKBP10 ***Alanay et al., 2010***; ***Kelley et al., 2011***; ***Barnes et al., 2013*** HR ***Berg et al., 1993***; ***Woods et al., 1996***; ***Goddard et al., 1995***; ***Ayling et al., 1997*** KRAS ***Aoki et al., 2005***; ***Schubbert et al., 2006***; ***Carta et al., 2006*** NBN ***Varon et al., 1998***; ***Tanzanella et al., 2003*** NIPBL ***Tonkin et al., 2004***; ***Krantz et al., 2004*** ORC1 ***Bicknell et al., 2011a***; ***Guernsey et al., 2011***; ***Bicknell et al., 2011b*** RC4 ***Guernsey et al., 2011***; ***Bicknell et al., 2011b*** ORC6L ***Bicknell et al., 2011a***; ***de Munnik et al., 2012*** PCNT ***Rauch et al., 2008***; ***Griffith et al., 2008***; ***Piane et al., 2009*** PLOD2 ***van der Slot et al., 2003***; ***HaVinh et al., 2004***; ***Puig-Hervás et al., 2012*** PTPN11 ***Tartaglia et al., 2001***; ***Maheshwari et al., 2002***; ***Kosaki et al., 2002*** RAD21 ***Deardorff et al., 2012***; ***Kruszka et al., 2019***; ***Goel and Parasivam, 2020*** RAF1 ***Pandit et al., 2007***; ***Razzaque et al., 2007*** RECQL4 ***Lindor et al., 2000***; ***Beghini et al., 2003***; ***Wang et al., 2003*** RIT1 ***Aoki et al., 2013***; ***Bertola et al., 2014***; ***Gos et al., 2014*** ROR2 ***Afzal et al., 2000***; ***van Bokhoven, 2000***; ***Tufan et al., 2005*** SLC26A2 ***Hästbacka et al., 1993***; ***Rossi and Superti-Furga, 2001***; ***Barreda-Bonis et al., 2018*** SMAD4 ***Le Goff et al., 2012***; ***Caputo et al., 2012***; ***Lindor et al., 2012*** SRCAP ***Hood et al., 2012***; ***Le Goff et al., 2013*** WRN ***Yu et al., 1996***; ***Goto et al., 1997***; ***Yu, 1997***

Table 1 continued on next page

Table 1 continued

Phenotype	Genes
Crohn disease	ATG16L1 ***Hampe et al., 2007*** CARD9 ***Rivas et al., 2011*** IL10 ***Fowler, 2005*** IL10RA ***Gasche et al., 2003***; ***Mao et al., 2012*** IL10RB ***Glocker et al., 2009***; ***Begue et al., 2011*** IL23R ***Duerr et al., 2006***; ***Libioulle et al., 2007***; ***Glas et al., 2007*** IRGM ***McCarroll et al., 2008***; ***Craddock et al., 2010***; ***Prescott et al., 2010*** NOD2 ***Ogura et al., 2001***; ***Hugot et al., 2001*** PRDM1 ***Ellinghaus et al., 2013*** PTPN22 ***Diaz-Gallo et al., 2011***
Ulcerative colitis	ATG16L1 ***Fowler et al., 2008*** CARD9 ***Rivas et al., 2011*** IL23R ***Glas et al., 2007***; ***Fisher et al., 2008*** IRGM ***McCarroll et al., 2008*** PRDM1 ***Ellinghaus et al., 2013*** PTPN22 ***Diaz-Gallo et al., 2011*** RNF186 ***Rivas et al., 2016***; ***Beaudoin et al., 2013***
Type II diabetes	ABCC8 ***Reis et al., 2000*** BLK ***Borowiec et al., 2009*** CEL ***Bengtsson-Ellmark et al., 2004***; ***Raeder et al., 2006*** EIF2AK3 ***Harding et al., 2001***; ***Brickwood et al., 2003***; ***Durocher et al., 2006*** GATA4 ***Shaw-Smith, 2014*** GATA6 ***Yorifuji et al., 2012***; ***De Franco et al., 2013*** GCK ***Froguel et al., 1993*** GLIS3 ***Senée et al., 2006*** HNF1A ***Yamagata et al., 1996b***; ***Vaxillaire et al., 1997*** HNF1B ***Horikawa et al., 1997***; ***Lindner et al., 1999*** HNF4A ***Yamagata et al., 1996a***; ***Stoffel and Duncan, 1997*** IER3IP1 ***Poulton et al., 2011***; ***Abdel-Salam et al., 2012***; ***Shalev et al., 2014*** INS ***Støy et al., 2007*** KCNJ11 ***Hani et al., 1998***; ***Gloyn et al., 2004***j KLF11 ***Neve et al., 2005*** LMNA ***Cao and Hegele, 2000*** NEUROD1 ***Malecki et al., 1999*** NEUROG3 ***Gradwohl et al., 2000***; ***Rubio-Cabezas et al., 2011***; ***Pinney et al., 2011*** PAX4 ***Shimajiri et al., 2001***; ***Mauvais-Jarvis et al., 2004***; ***Plengvidhya et al., 2007*** PDX1 ***Stoffers et al., 1997***; ***Macfarlane et al., 1999***; ***Hani et al., 1999*** PPARG ***Deeb et al., 1998***; ***Savage et al., 2002*** PTF1A ***Sellick et al., 2004*** RFX6 ***Smith, 2010***; ***Sansbury et al., 2015*** SLC19A2 ***Labay et al., 1999***, ***Oishi et al., 2002***; ***Shaw-Smith et al., 2012*** SLC2A2 ***Laukkanen et al., 2005***; ***Sansbury et al., 2012*** WFS1 ***Strom et al., 1998***; ***Hardy et al., 1999***; ***Khanim et al., 2001*** ZFP57 ***Mackay et al., 2008***; ***Boonen et al., 2013***

Table 1 continued on next page

Table 1 continued

Phenotype	Genes
Breast cancer (selected using MutPanning; ***Dietlein et al., 2020***)	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

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 (91.1%) of these genes had an eQTL in at least one relevant tissue at a false discovery rate (FDR) of $Q < 0.05$ ('Materials and methods'). If these variants act through changes in gene expression, phenotypic associations should be driven by some of 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 (***Giambartolomei et al., 2014***), JLIM (***Chun et al., 2017***), and eCAVIAR (***Hormozdiari et al., 2016***). 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 ('Materials and 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%; ***Giambartolomei et al., 2014***; ***Chun et al., 2017***; ***Hormozdiari et al., 2016***; ***Wen et al., 2017***) 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 predefined 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.

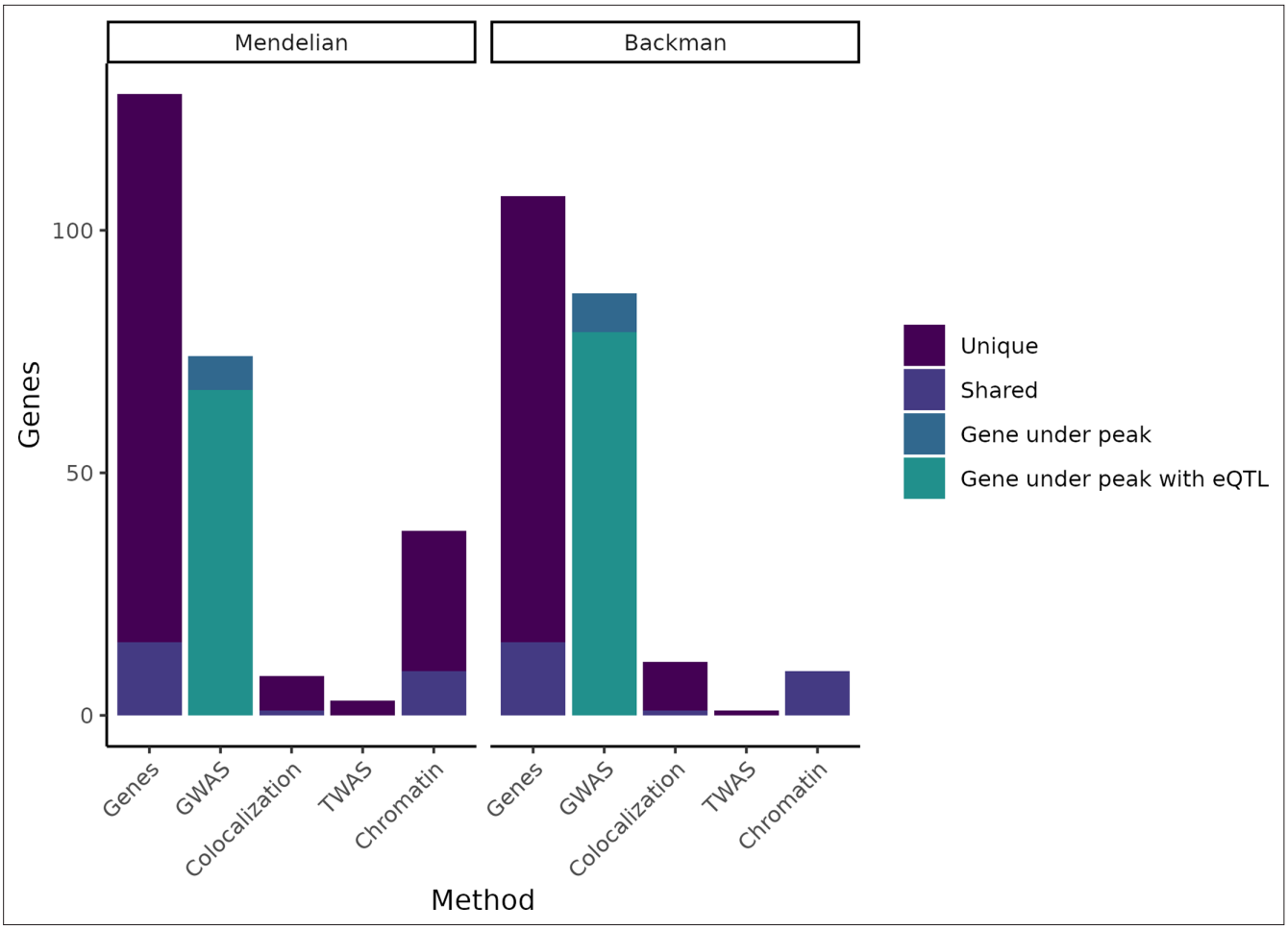

Figure 1. Putatively causative genes identified by each method category. The leftmost column in each half of the plot displays the entire group of putatively causative genes for our Mendelian set of genes and our (***Backman et al., 2021***) set of genes, respectively, as well as noting how many are 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 expression QTL (eQTL). The remaining columns indicate how many genes were identified through colocalization, transcriptome-wide association studies (TWAS), or chromatin methods, while noting how many of these genes are unique vs. shared between the Mendelian and Backman sets.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Enrichment of Mendelian genes near GWAS peaks.

Figure supplement 2. Change in coloc hits when adjusting expression QTL(eQTL) statistics using Multivariate Adaptive Shrinkage Method (MASH).

To address potential shortcomings from the available sample of tissue contexts, we incorporated the Multivariate Adaptive Shrinkage Method (MASH) (***Urbut et al., 2019***). 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 ('Materials and methods'). Rerunning coloc with these modified statistics increased the number of GWAS-eQTL colocalizations across all genes by 26% (from 389 to 489). However, the 100 new colocalizations identified only four additional putatively causative genes (***Figure 1—figure supplement 2***). These results indicate that tissue-type selection was not the limiting factor in our analysis.

Transcriptome-wide association studies (TWAS) (***Gamazon et al., 2015***; ***Gusev et al., 2016***; ***Mancuso et al., 2017***; ***Barbeira et al., 2018***) are another class of methods applied to identify causative

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

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	High-density lipoprotein	Liver Adipose (subcutaneous) Whole blood
Dyslipidemia Hyperlipidemia	Low-density lipoprotein	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

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 (***Wainberg et al., 2019***). However, while sensitive, TWAS analyses typically yield expansive result sets that include many false positives and are sensitive to the number of tissue types (***Wainberg et al., 2019***). Results from the FUSION implementation of TWAS (***Mancuso et al., 2017***) 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 in our study loci reside in likely 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 (***Wang et al., 2020b***). For 37 of these genes, we identified at least one high-confidence fine-mapped variant (defined as a variant with posterior inclusion probability, or PIP, greater than 0.7) within 100 kb of the transcription start site. We tested whether these fine-mapped variants fall within (i) regulatory DNA marked by DNase I hypersensitive sites (DHS) ***Meuleman et al., 2020***; (ii) a narrowly mapped active histone modification feature (H3K27ac, H3K4me1, or H3K4me3; ***Nasser et al., 2021***); or (iii) sites marked as an 'enhancer' by ChromHMM (***Ernst and Kellis, 2012***; ***Ernst and Kellis, 2017***; 'Materials and methods'). As many as 32/37 (86%) genes identified this way have a fine-mapped variants within a candidate regulatory region across all the tissue types examined, or 25/37 (68%) when restricting to trait-relevant tissues (***Figure 2***); (***Supplementary file 1*** and ***Supplementary file 2***). Despite this strong evidence that these GWAS associations arise from effects on regulatory DNA, only 5/25 loci (20%) demonstrably correspond to expression effects in our eQTL analysis.

In order to more analogously compare our regulatory feature analysis to our eQTL analysis, we computed 'activity-by-distance' (ABD)—a simplification of the 'activity-by-contact' method that provides a measure connecting a distal regulatory region to a target gene (***Nasser et al., 2021***; ***Fulco et al., 2019***; 'Materials and methods' ***Figure 3***). Taking each locus's feature with the highest ABD score, we implicate 5/37 (14%) of our Mendelian subset of genes. As such, even when a GWAS association and trait-relevant gene are in the same locus, they are difficult to link, whether using eQTLs or current approaches to integrating chromatin data with target genes.

Discussion

Overall, our results are strongly 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 same results are inconsistent with a model that such effects on gene regulation arise from the genetic influences on baseline gene expression that are captured by eQTLs.

The enrichment of 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 likely regulatory regions marked by DHS and other chromatin features lends support

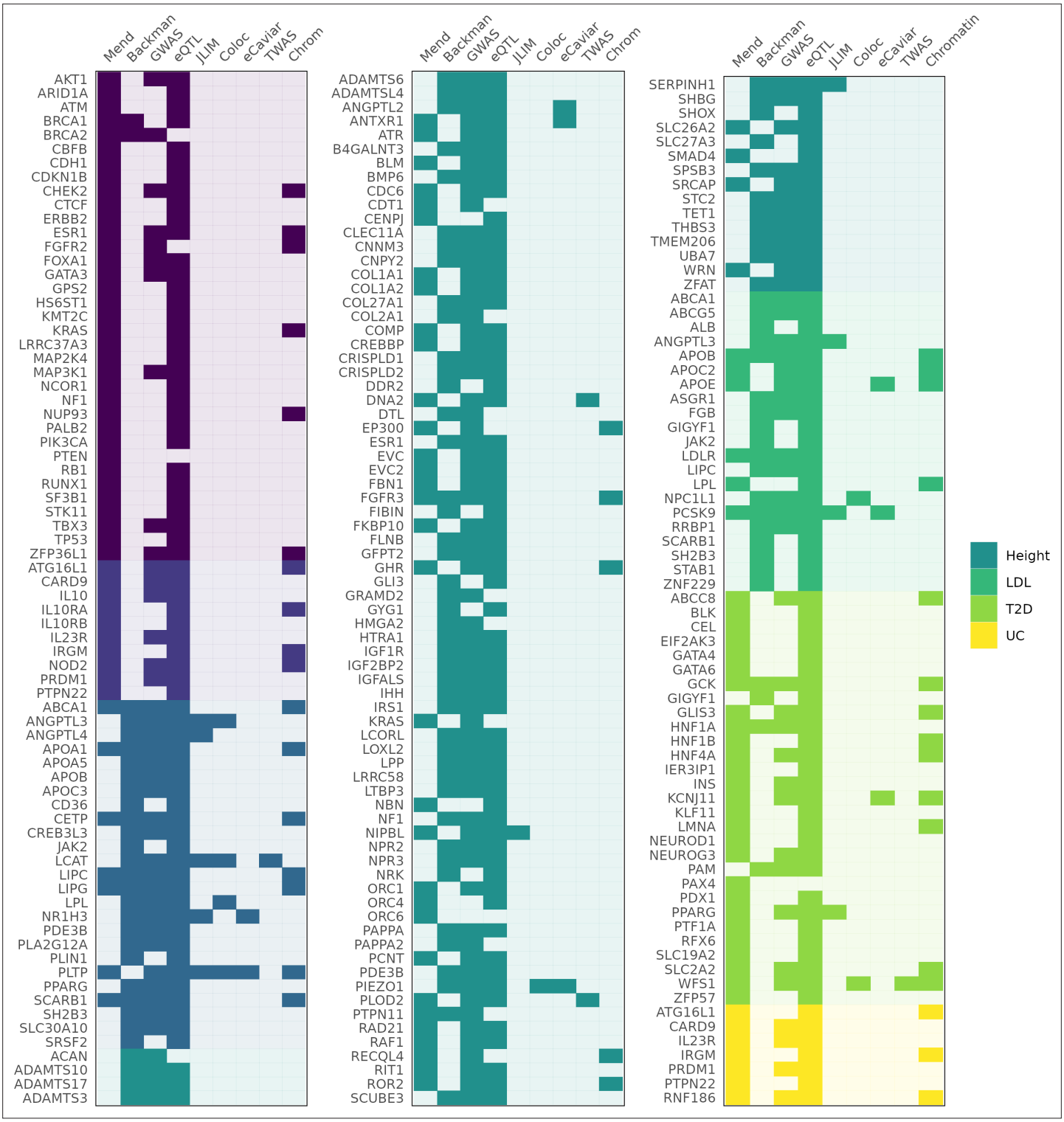

Figure 2. Genes identified as associated with a complex trait by each method. Columns 'Mend' and 'Backman' indicate whether a gene is from the Mendelian set of putatively causative genes, the Backman et al. set, or both. Subsequent columns indicate whether a gene was identified as a hit using each of our methods: JLIM, coloc, eCaviar, transcriptome-wide association studies (TWAS), and chromatin analysis.

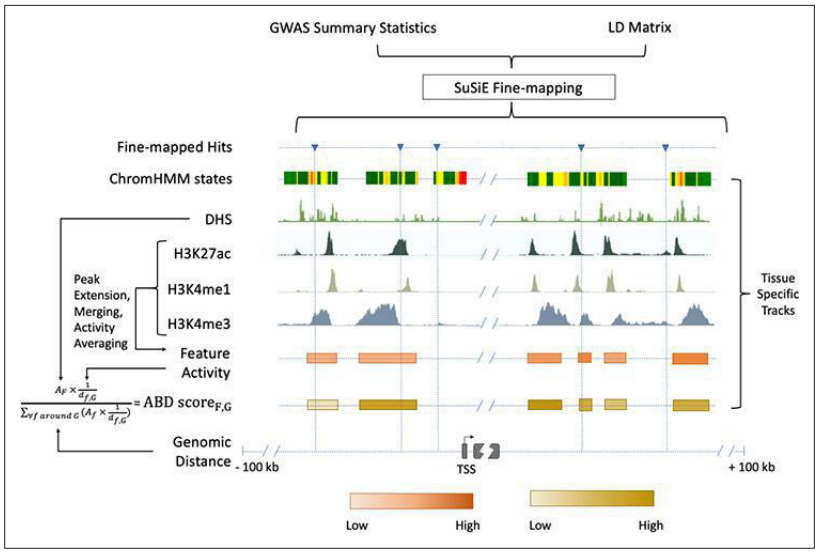

Figure 3. Chromatin-based causative gene identification. Following the fine-mapping of GWAS variants, three parallel methods were used. The first identified fine-mapped variants falling within regions annotated as enhancers by ChromHMM. The second identified variants within histone modification features and evaluated their relevance using an activity-by-distance (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 ('Materials and 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.

The online version of this article includes the following source data for figure 3:

Source data 1. Gene-level results for linked expression and traits.

to the general model of GWAS associations arising from effects on gene regulation and expression. However, the inability of varied statistical methods to connect GWAS associations and expression argues against the idea that the causative GWAS variants exert their effects via the homeostatic or blended effects on gene expression that are captured by bulk-tissue eQTLs of the sort discovered by broad expression-data collection projects.

Many explanations have been suggested for the limited success of expression-centric 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 because it delivers the largest number of positive findings, but it is ill-suited for providing explanations for negative results—when you do not know what you were looking for, it is hard to explain why you did not find it. By testing only loci for which there is a strongly suspected contributing gene, we were better able to distinguish which factors might prevent us from identifying GWAS-to-gene links using expression alone.

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 (i) non-expression-mediated mechanisms, (ii) lack of statistical power for GWAS, (iii) the absence of eQTLs for relevant genes, (iv) and underpowered methods for linking expression to GWAS (***Table 3***).

Our results suggest that there remains to be discovered a major component connecting variants to gene expression, which we term 'missing regulation.' We propose that this currently lacking component could be exposed by considering more nuanced experimental and analytical models of gene expression. One likely model involves context-dependent gene expression—expression whose relevant effects are confined to (i) particular refined cell types or anatomical partitions, or (ii) cell states such as responses to perturbations, whether exogenous (e.g., environmental change) or endogenous (e.g., cell differentiation). A complementary model may incorporate heterogeneity of gene expression or the variance of expression across relatively short time scales. These models and others may depend on or be augmented by thresholding or buffering of variant effects, which may produce changes in

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*-QTLs. This table details the explanations inconsistent with our results, which are explained in the left column and addressed on 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.

	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., 2021*** 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 ***GTEx Consortium, 2020***; ***Gamazon et al., 2015*** • 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 ***Chun et al., 2017***; ***Giambartolomei et al., 2018***; ***Hormozdiari et al., 2016***; ***Gusev et al., 2016***; ***Hukku et al., 2021*** • They are robust to levels of LD mismatch higher than what would be expected given our datasets ***Chun et al., 2017***; ***Wainberg et al., 2019***; ***Hukku et al., 2021*** • Some, though not all, of the methods are robust to allelic heterogeneity ***Chun et al., 2017***; ***Wainberg et al., 2019***

eQTL = expression QTL; TWAS = transcriptome-wide association studies.

gene expression with nonlinear effects on phenotype. A summary of proposed models can be found in ***Table 4***.

We propose that finding the 'missing regulation' will require not only identifying novel eQTLs explaining GWAS peaks, but also explaining the phenotypic irrelevance of 'red herring eQTLs'—that

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-hand side contains explanations of the models and current relevant research.

Extended models of gene regulation	
Context dependency: a context-specific eQTL, invisible in bulk tissues analyzed to date, replaces or supplements the bulk tissue homeostatic eQTL	Cell type ***Dobbyn et al., 2018***; ***Zhang et al., 2018***; ***Schmiedel et al., 2018***; ***Glastonbury et al., 2019***; ***Rai et al., 2020***; ***Findley et al., 2021***; ***Neavin et al., 2021***; ***Ota et al., 2021***; ***Patel et al., 2021***; ***Bryois et al., 2021***; ***Arvanitis et al., 2022***; ***Oelen et al., 2022***; ***Perez et al., 2022***; ***Schmiedel et al., 2022***; ***Yazar et al., 2022*** • 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 either cannot be detected in bulk tissue because of the cell type's low prevalence • The appropriate cellular or anatomical context has not yet been analyzed
	Developmental timing ***Dobbyn et al., 2018***; ***Strober et al., 2019***; ***Cuomo et al., 2020***; ***Bonder et al., 2021***; ***Jerber et al., 2021***; ***Aygün et al., 2022***; ***Elorbany et al., 2022*** • The GWAS phenotype depends on a specific point in cell/tissue development or differentiation • eQTLs present at the correct interval contribute to phenotype, but eQTLs observed at other points do not
	Cell state or environment ***Findley et al., 2021***; ***Ota et al., 2021***; ***Oelen et al., 2022***; ***Schmiedel et al., 2022***; ***Huh and Paulsson, 2011***; ***Knowles et al., 2017***; ***Kim-Hellmuth et al., 2017***; ***Balliu et al., 2021***; ***Mu et al., 2021***; ***Ward et al., 2021***; ***Nathan et al., 2022***; ***Baca et al., 2022*** • 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 immune activation or a metabolic shift

Table 4 continued on next page

Table 4 continued

	Extended models of gene regulation	
Nonlinear or non-homeostatic: the relationship between eQTL and genotype is indirect	Nonlinearity ***Fu et al., 2009***; ***Dori-Bachash et al., 2011***; ***Ghazalpour et al., 2011***; ***Pai et al., 2012***; ***Vogel and Marcotte, 2012***; ***Khan et al., 2013***; ***Wu et al., 2013***; ***McManus et al., 2014***; ***Albert and Kruglyak, 2015***; ***Bader et al., 2015***; ***Battle et al., 2015***; ***Cenik et al., 2015***; ***McManus et al., 2015***; ***Pai et al., 2015***; ***Schafer et al., 2015***; ***Chick et al., 2016***; ***Liu et al., 2016***; ***Schaefke et al., 2018***; ***Buccitelli and Selbach, 2020***; ***Wang et al., 2020a***; ***Kusnadi et al., 2022***	• 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 ***Pedraza and Paulsson, 2008***; ***Raj and van Oudenaarden, 2008***; ***Shahrezaei and Swain, 2008***; ***Larson et al., 2009***; ***Raj and van Oudenaarden, 2009***; ***Suter et al., 2011***; ***Dar et al., 2012***; ***Viñuelas et al., 2013***; ***Kumar et al., 2015***; ***Nicolas et al., 2017***; ***Qiu et al., 2019***; ***Wang et al., 2020c***	• 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

eQTL = expression QTL.

is, eQTLs for putatively causative genes that fall near GWAS peaks but do not colocalize with them. Above, we use the example genes *APOE* and *LDLR*. Both these genes harbor coding variants causing Mendelian hypercholesterolemia, and 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.

Importantly, our results do not diminish the importance or general utility of eQTLs. Rather, they suggest that current models are deficient in two respects: (i) they fail to unify trait-associated non-coding variants with known trait-associated genes, and (ii) they fail to explain the non-effects of identified 'red herring' eQTLs. These deficiencies highlight a need for new approaches to the role of gene regulation in complex traits.

One long-standing goal of GWAS has been to discover genes contributing to complex traits (***Manolio et al., 2009***; ***Farh et al., 2015***), but low rates of positive findings for expression-based variant-to-gene methods have constrained this possibility (***Chun et al., 2017***; ***Baird et al., 2021***). Among other challenges, this has limited the benefit of GWAS and expression data for disease-gene mapping and drug discovery (***Baird et al., 2021***; ***Umans et al., 2021***). Another practical question raised is the value of current large-scale public datasets. Compared to genotypes, expression data are relatively difficult to collect, particularly from specialized cell contexts. If the most relevant models are shown to depend on effects not observable in homeostatic eQTL mapping, the field may need to consider prioritizing other biological contexts and forms of expression data.

Materials and 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 (***Dietlein et al., 2020***) (citations for each gene are included in ***Table 1***). Review papers, as well as the OMIM database (***McKusick-nathans institute of genetic medicine, 2021***), 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 (***Deardorff et al., 2007***). Several of these phenotypes—height, HDL, cholesterol, breast cancer, and type II diabetes—were also analyzed in ***Backman et al., 2021***, which, through burden testing, identified a total of 110 genes; after accounting for overlaps, this increased our set of putatively causative genes to 220.

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 variant 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 variants were selected based on the pext (proportion of expression across transcripts) data (***Cummings et al., 2020***). 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. The code for this analysis, and all other quantitative analyses in this paper, can be accessed at https://github.com/NJC12/missing_link_association_function (***NJC12, 2023***; copy archived at swh:1:rev:46d9072b7cc13f6532203d1494eec4d0f634e092).

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., 2017***. The GWAS were run using Plink 2.0 (***Chang et al., 2015***), with the covariates age, sex, body mass index (for LDL and HDL only), 10 principal components, and coding variants.

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) (***Yang et al., 2011***; ***Yang et al., 2012***) 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 (***Taliun et al., 2021***). The ancestry of these subjects was confirmed with FastPCA (***Galinsky et al., 2016a***; ***Galinsky et al., 2016b***), and the relevant data were extracted using bcftools (***Danecek et al., 2021***). 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 were calculated using Fisher's exact test (***Figure 1***, ***Figure 1—figure supplement 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 and BC GWAS; BC genes and LDL GWAS; LDL genes and UC GWAS; UC genes and height GWAS; height genes and T2D GWAS; T2D genes and HDL GWAS; HDL genes and 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 (***Chun et al., 2017***). Because we tested only a subset of genes, we used the Benjamini–Hochberg method (***Benjamini and Hochberg, 1995***) 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 (***Chun et al., 2017***) was run using GWAS summary statistics and GTEx v7 genotypes and phenotypes for the tissues listed in ***Table 2***. Coloc (***Giambartolomei et al., 2014***) was run using GWAS and GTEx v7 summary statistics for the same tissues. eCAVIAR (***Hormozdiari et al., 2016***) was run using GWAS and GTEx v7 summary statistics for these tissues, and a reference dataset of LD from UKBB (***Weissbrod et al., 2020***). MASH was run incorporating data from all non-brain tissues, and coloc was rerun using the adjusted values for the same tissues as before.

MASH

MASH was applied to all GTEx tissues using the mashr R-package (***Urbut et al., 2019***). 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-brain tissues to cluster separately in expression analysis (***Battle et al., 2017***; ***Park et al., 2017***; ***Gamazon et al., 2019***).

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 (***Mancuso et al., 2017***). 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., >150 for LDL cholesterol and >4800 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 prefrontal 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 (***Wang et al., 2020b***) 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) DHS, 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 (***Wainberg et al., 2019***) 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 (***Wainberg et al., 2019***) ftp site (available here) for 14 different tissue types (***Supplementary file 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. ***Barbeira et al., 2018***, 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/anshul kundaje/projects/blacklists) were removed (***Kundaje et al., 2015***). The remaining features were re-centered around the peak and overlapping features were merged to give the final set of features per histone modification track. The 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 ABD scores (a simplified version of ABC scores ***Barbeira et al., 2018***) for gene–feature pairs using the following formula:

$$ABDscore_{F,G} = \frac{A_F \times 1/d_{F,G}}{\sum_{allfwithin \pm 5MbofG} A_f \times 1/d_{f,G}}$$

The ABD score can be thought of as a measure of the contribution of a feature, F to the combined regulatory 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 DHS that are considered to be generic markers of the regulatory DNA and can contain genetic variations associated with traits and diseases (***Meuleman et al., 2020***). 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 biosample from a certain tissue type (***Supplementary file 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 variant, we identified all DHS that fall within ±100 kb of the variant.

Finally, we used in silico chromatin state predictions (chromHMM core 15-state model ***Wainberg et al., 2019***) for relevant tissue types (***Supplementary file 1***) to identify active enhancer regions in the genome. Tissue-specific chromHMM annotations were downloaded from the Roadmap Epigenomics Project ***Wainberg et al., 2019*** ftp site (available here). 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.

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. NJC was supported by NIH training grant T32GM74897. UK Biobank was accessed under projects 14048 and 10438. TOPMed data were used under dbGaP project 28674.

Additional information

Funding

Funder	Grant reference number	Author
National Institutes of Health	R35GM127131	Shamil R Sunyaev
National Institutes of Health	R01HG010372	Shamil R Sunyaev
National Institutes of Health	R01MH101244	Shamil R Sunyaev
National Institutes of Health	U01HG012009	Chris Cotsapas
National Institutes of Health	T32GM74897	Shamil R Sunyaev

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Noah J Connally, Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Sumaiya Nazeen, Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing – review and editing; Daniel Lee, Data curation, Formal analysis, Investigation, Writing – review and editing; Huwenbo Shi, Resources, Data curation, Software, Formal analysis, Investigation, Methodology, Writing – review and editing; John Stamatoyannopoulos, Resources, Funding acquisition, Writing – review and editing; Sung Chun, Chris Cotsapas, Conceptualization, Supervision, Methodology, Writing – review and editing; Christopher A Cassa, Conceptualization, Resources, Supervision, Methodology, Project administration, Writing – review and editing; Shamil R Sunyaev, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Noah J Connally http://orcid.org/0000-0003-3818-6739
Shamil R Sunyaev http://orcid.org/0000-0001-5715-5677

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.74970.sa1
Author response https://doi.org/10.7554/eLife.74970.sa2

Additional files

Supplementary files

- Supplementary file 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.
- Supplementary file 2. Tissue types and biosamples from the DNase I hypersensitive sites index used for functional genomic analysis.
Meuleman et al., 2020 assess DNase I hypersensitive sites across 438 cell and tissue types; we selected the above based on their relevance to our complex traits.

- Supplementary file 3. TOPMed URLs used.
- Transparent reporting form

Data availability

Numerical data for results is included in Source Data 1. The dataset generated (GWAS summary statistics conditioned on coding variants) can be found at https://doi.org/10.5061/dryad.612jm644q. Code for this project is available at https://github.com/NJC12/missing_link_association_function (copy archived at swh:1:rev:46d9072b7cc13f6532203d1494eec4d0f634e092). UK Biobank data was used and is available by application to the UK Biobank Data Access Committee https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. TOPMed Whole Genome Sequencing Project - Freeze 5b, Phases 1 and 2 data was used and can be accessed at https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-project-freeze-5b-phases-1-and-2, full list of TOPMed URLs available in Supplementary file 3.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Connally NJ	2022	GWAS results conditioned on coding variants	https://dx.doi.org/10.5061/dryad.612jm644q	Dryad Digital Repository, 10.5061/dryad.612jm644q

References

Abdel-Salam GMH, Schaffer AE, Zaki MS, Dixon-Salazar T, Mostafa IS, Afifi HH, Gleeson JG. 2012. A homozygous IER3IP1 mutation causes microcephaly with simplified gyral pattern, epilepsy, and permanent neonatal diabetes syndrome (MEDS). *American Journal of Medical Genetics. Part A* **158A**:2788–2796. DOI: https://doi.org/10.1002/ajmg.a.35583, PMID: 22991235

Abifadel M, Varret M, Rabès J-P, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D, Derré A, Villéger L, Farnier M, Beucler I, Bruckert E, Chambaz J, Chanu B, Lecerf J-M, Luc G, Moulin P, et al. 2003. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature Genetics* **34**:154–156. DOI: https://doi.org/10.1038/ng1161, PMID: 12730697

Afzal AR, Rajab A, Fenske CD, Oldridge M, Elanko N, Ternes-Pereira E, Tüysüz B, Murday VA, Patton MA, Wilkie AO, Jeffery S. 2000. Recessive robinow syndrome, allelic to dominant brachydactyly type B, is caused by mutation of ROR2. *Nature Genetics* **25**:419–422. DOI: https://doi.org/10.1038/78107, PMID: 10932186

Alanay Y, Avaygan H, Camacho N, Utine GE, Boduroglu K, Aktas D, Alikasifoglu M, Tuncbilek E, Orhan D, Bakar FT, Zabel B, Superti-Furga A, Bruckner-Tuderman L, Curry CJR, Pyott S, Byers PH, Eyre DR, Baldridge D, Lee B, Merrill AE, et al. 2010. Mutations in the gene encoding the rER protein FKBP65 cause autosomal-recessive osteogenesis imperfecta. *American Journal of Human Genetics* **86**:551–559. DOI: https://doi.org/10.1016/j.ajhg.2010.02.022, PMID: 20362275

Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nature Reviews. Genetics* **16**:197–212. DOI: https://doi.org/10.1038/nrg3891, PMID: 25707927

AlDosari MS, Shaheen R, Colak D, Alkuraya FS. 2010. Novel CENPJ mutation causes seckel syndrome. *Journal of Medical Genetics* **47**:411–414. DOI: https://doi.org/10.1136/jmg.2009.076646, PMID: 20522431

Angius A, Uva P, Oppo M, Persico I, Onano S, Olla S, Pes V, Perria C, Cuccuru G, Atzeni R, Serra G, Cucca F, Sotgiu S, Hennekam RC, Crisponi L. 2019. Confirmation of a new phenotype in an individual with a variant in the last part of exon 30 of CREBBP. *American Journal of Medical Genetics. Part A* **179**:634–638. DOI: https://doi.org/10.1002/ajmg.a.61052, PMID: 30737887

Aoki Y, Niihori T, Kawame H, Kurosawa K, Ohashi H, Tanaka Y, Filocamo M, Kato K, Suzuki Y, Kure S, Matsubara Y. 2005. Germline mutations in HRAS proto-oncogene cause Costello syndrome. *Nature Genetics* **37**:1038–1040. DOI: https://doi.org/10.1038/ng1641, PMID: 16170316

Aoki Y, Niihori T, Banjo T, Okamoto N, Mizuno S, Kurosawa K, Ogata T, Takada F, Yano M, Ando T, Hoshika T, Barnett C, Ohashi H, Kawame H, Hasegawa T, Okutani T, Nagashima T, Hasegawa S, Funayama R, Nagashima T, et al. 2013. Gain-Of-Function mutations in RIT1 cause Noonan syndrome, a Ras/MAPK pathway syndrome. *American Journal of Human Genetics* **93**:173–180. DOI: https://doi.org/10.1016/j.ajhg.2013.05.021, PMID: 23791108

Arvanitis M, Tayeb K, Strober BJ, Battle A. 2022. Redefining tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity. *American Journal of Human Genetics* **109**:223–239. DOI: https://doi.org/10.1016/j.ajhg.2022.01.002, PMID: 35085493

Aygün N, Liang D, Crouse WL, Keele GR, Love MI, Stein JL. 2022. Inferring Cell-Type-Specific Causal Gene Regulatory Networks during Human Neurogenesis. [bioRxiv]. DOI: https://doi.org/10.1101/2022.04.25.488920, PMID: 36177099

Ayling RM, Ross R, Towner P, Von Laue S, Finidori J, Moutoussamy S, Buchanan CR, Clayton PE, Norman MR. 1997. A dominant-negative mutation of the growth hormone receptor causes familial short stature. *Nature Genetics* **16**:13–14. DOI: https://doi.org/10.1038/ng0597-13, PMID: 9140387

Baca SC, Singler C, Zacharia S, Seo J-H, Morova T, Hach F, Ding Y, Schwarz T, Huang C-CF, Anderson J, Fay AP, Kalita C, Groha S, Pomerantz MM, Wang V, Linder S, Sweeney CJ, Zwart W, Lack NA, Pasaniuc B, et al. 2022. Genetic determinants of chromatin reveal prostate cancer risk mediated by context-dependent gene regulation. *Nature Genetics* **54**:1364–1375. DOI: https://doi.org/10.1038/s41588-022-01168-y, PMID: 36071171

Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, Benner C, Liu D, Locke AE, Balasubramanian S, Yadav A, Banerjee N, Gillies CE, Damask A, Liu S, Bai X, Hawes A, Maxwell E, Gurski L, Watanabe K, et al. 2021. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**:628–634. DOI: https://doi.org/10.1038/s41586-021-04103-z, PMID: 34662886

Bader DM, Wilkening S, Lin G, Tekkedil MM, Dietrich K, Steinmetz LM, Gagneur J. 2015. Negative feedback buffers effects of regulatory variants. *Molecular Systems Biology* **11**:785. DOI: https://doi.org/10.15252/msb.20145844, PMID: 25634765

Baird DA, Liu JZ, Zheng J, Sieberts SK, Perumal T, Elsworth B, Richardson TG, Chen C-Y, Carrasquillo MM, Allen M, Reddy JS, De Jager PL, Ertekin-Taner N, Mangravite LM, Logsdon B, Estrada K, Haycock PC, Hemani G, Runz H, Smith GD, et al. 2021. Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLOS Genetics* **17**:e1009224. DOI: https://doi.org/10.1371/journal.pgen.1009224, PMID: 33417599

Balliu B, Carcamo-Orive I, Gloudemans MJ, Nachun DC, Durrant MG, Gazal S, Park CY, Knowles DA, Wabitsch M, Quertermous T, Knowles JW, Montgomery SB. 2021. An integrated approach to identify environmental modulators of genetic risk factors for complex traits. *American Journal of Human Genetics* **108**:1866–1879. DOI: https://doi.org/10.1016/j.ajhg.2021.08.014, PMID: 34582792

Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, Stahl EA, Huckins LM, Nicolae DL, Cox NJ, Im HK, GTEx Consortium. 2018. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications* **9**:1825. DOI: https://doi.org/10.1038/s41467-018-03621-1, PMID: 29739930

Barnes AM, Duncan G, Weis M, Paton W, Cabral WA, Mertz EL, Makareeva E, Gambello MJ, Lacbawan FL, Leikin S, Fertala A, Eyre DR, Bale SJ, Marini JC. 2013. Kuskokwim syndrome, a recessive congenital contracture disorder, extends the phenotype of FKBP10 mutations. *Human Mutation* **34**:1279–1288. DOI: https://doi.org/10.1002/humu.22362, PMID: 23712425

Barreda-Bonis AC, Barraza-García J, Parrón M, Pastor I, Heath KE, González-Casado I. 2018. Multiple Slc26a2 mutations occurring in a three-generational family. *European Journal of Medical Genetics* **61**:24–28. DOI: https://doi.org/10.1016/j.ejmg.2017.10.007, PMID: 29024831

Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, Gilad Y. 2015. Genomic variation: impact of regulatory variation from RNA to protein. *Science* **347**:664–667. DOI: https://doi.org/10.1126/science.1260793, PMID: 25657249

Battle A, Brown CD, Engelhardt BE, Montgomery SB, GTEx Consortium, Laboratory, Data Analysis &Coordinating Center—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, et al. 2017. Genetic effects on gene expression across human tissues. *Nature* **550**:204–213. DOI: https://doi.org/10.1038/nature24277, PMID: 29022597

Bayram Y, Pehlivan D, Karaca E, Gambin T, Jhangiani SN, Erdin S, Gonzaga-Jauregui C, Wiszniewski W, Muzny D, Baylor-Hopkins Center for Mendelian Genomics, Elcioglu NH, Yildirim MS, Bozkurt B, Zamani AG, Boerwinkle E, Gibbs RA, Lupski JR. 2014. Whole exome sequencing identifies three novel mutations in ANTXR1 in families with GAPO syndrome . *American Journal of Medical Genetics Part A* **164**:2328–2334. DOI: https://doi.org/10.1002/ajmg.a.36678

Beaudoin M, Goyette P, Boucher G, Lo KS, Rivas MA, Stevens C, Alikashani A, Ladouceur M, Ellinghaus D, Törkvist L, Goel G, Lagacé C, Annese V, Bitton A, Begun J, Brant SR, Bresso F, Cho JH, Duerr RH, Halfvarson J, et al. 2013. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLOS Genetics* **9**:e1003723. DOI: https://doi.org/10.1371/journal.pgen.1003723, PMID: 24068945

Beghini A, Castorina P, Roversi G, Modiano P, Larizza L. 2003. Rna processing defects of the helicase gene RecQL4 in a compound heterozygous Rothmund-Thomson patient. *American Journal of Medical Genetics. Part A* **120A**:395–399. DOI: https://doi.org/10.1002/ajmg.a.20154, PMID: 12838562

Begue B, Verdier J, Rieux-Laucat F, Goulet O, Morali A, Canioni D, Hugot J-P, Daussy C, Verkarre V, Pigneur B, Fischer A, Klein C, Cerf-Bensussan N, Ruemmele FM. 2011. Defective IL10 signaling defining a subgroup of patients with inflammatory bowel disease. *The American Journal of Gastroenterology* **106**:1544–1555. DOI: https://doi.org/10.1038/ajg.2011.112, PMID: 21519361

Bengtsson-Ellmark SH, Nilsson J, Orho-Melander M, Dahlenborg K, Groop L, Bjursell G. 2004. Association between a polymorphism in the carboxyl ester lipase gene and serum cholesterol profile. *European Journal of Human Genetics* **12**:627–632. DOI: https://doi.org/10.1038/sj.ejhg.5201204, PMID: 15114370

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**:289–300. DOI: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berg MA, Argente J, Chernausek S, Gracia R, Guevara-Aguirre J, Hopp M, Pérez-Jurado L, Rosenbloom A, Toledo SP, Francke U. 1993. Diverse growth hormone receptor gene mutations in Laron syndrome. *American Journal of Human Genetics* **52**:998–1005 PMID: 8488849.

Bertola DR, Yamamoto GL, Almeida TF, Buscarilli M, Jorge AAL, Malaquias AC, Kim CA, Takahashi VNV, Passos-Bueno MR, Pereira AC. 2014. Further evidence of the importance of RIT1 in Noonan syndrome . *American Journal of Medical Genetics Part A* **164**:2952–2957. DOI: https://doi.org/10.1002/ajmg.a.36722

Bicknell LS, Bongers E, Leitch A, Brown S, Schoots J, Harley ME, Aftimos S, Al-Aama JY, Bober M, Brown PAJ, van Bokhoven H, Dean J, Edrees AY, Feingold M, Fryer A, Hoefsloot LH, Kau N, Knoers N, Mackenzie J, Opitz JM, et al. 2011a. Mutations in the pre-replication complex cause meier-gorlin syndrome. *Nature Genetics* **43**:356–359. DOI: https://doi.org/10.1038/ng.775, PMID: 21358632

Bicknell LS, Walker S, Klingseisen A, Stiff T, Leitch A, Kerzendorfer C, Martin CA, Yeyati P, Al Sanna N, Bober M, Johnson D, Wise C, Jackson AP, O'Driscoll M, Jeggo PA. 2011b. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling meier-gorlin syndrome. *Nature Genetics* **43**:350–355. DOI: https://doi.org/10.1038/ng.776, PMID: 21358633

Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabanian H, Melamed R, Rabadan R, Bernstam EV, Brunak S, Jensen LJ, Nicolae D, Shah NH, Grossman RL, Cox NJ, White KP, Rzhetsky A. 2013. A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. *Cell* **155**:70–80. DOI: https://doi.org/10.1016/j.cell.2013.08.030, PMID: 24074861

Bodzioch M, Orsó E, Klucken J, Langmann T, Böttcher A, Diederich W, Drobnik W, Barlage S, Büchler C, Porsch-Ozcürümez M, Kaminski WE, Hahmann HW, Oette K, Rothe G, Aslanidis C, Lackner KJ, Schmitz G. 1999. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nature Genetics* **22**:347–351. DOI: https://doi.org/10.1038/11914, PMID: 10431237

Bonder MJ, Smail C, Gloudemans MJ, Frésard L, Jakubosky D, D'Antonio M, Li X, Ferraro NM, Carcamo-Orive I, Mirauta B, Seaton DD, Cai N, Vakili D, Horta D, Zhao C, Zastrow DB, Bonner DE, HipSci Consortium, iPSCORE consortium, Undiagnosed Diseases Network, et al. 2021. Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nature Genetics* **53**:313–321. DOI: https://doi.org/10.1038/s41588-021-00800-7, PMID: 33664507

Boonen SE, Mackay DJG, Hahnemann JMD, Docherty L, Grønskov K, Lehmann A, Larsen LG, Haemers AP, Kockaerts Y, Dooms L, Vu DC, Ngoc CTB, Nguyen PB, Kordonouri O, Sundberg F, Dayanikli P, Puthi V, Acerini C, Massoud AF, Tümer Z, et al. 2013. Transient neonatal diabetes, Zfp57, and hypomethylation of multiple imprinted loci: a detailed follow-up. *Diabetes Care* **36**:505–512. DOI: https://doi.org/10.2337/dc12-0700, PMID: 23150280

Borowiec M, Liew CW, Thompson R, Boonyasrisawat W, Hu J, Mlynarski WM, El Khattabi I, Kim SH, Marselli L, Rich SS, Krolewski AS, Bonner-Weir S, Sharma A, Sale M, Mychaleckyj JC, Kulkarni RN, Doria A. 2009. Mutations at the blk locus linked to maturity onset diabetes of the young and β-cell dysfunction. *PNAS* **106**:14460–14465. DOI: https://doi.org/10.1073/pnas.0906474106, PMID: 19667185

Brickwood S, Bonthron DT, Al-Gazali LI, Piper K, Hearn T, Wilson DI, Hanley NA. 2003. Wolcott-rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. *Journal of Medical Genetics* **40**:685–689. DOI: https://doi.org/10.1136/jmg.40.9.685, PMID: 12960215

Briggs MD, Hoffman SM, King LM, Olsen AS, Mohrenweiser H, Leroy JG, Mortier GR, Rimoin DL, Lachman RS, Gaines ES. 1995. Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. *Nature Genetics* **10**:330–336. DOI: https://doi.org/10.1038/ng0795-330, PMID: 7670472

Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, Yu L, Brewer C, Collins JA, Molhuizen HO, Loubser O, Ouelette BF, Fichter K, Ashbourne-Excoffon KJ, Sensen CW, Scherer S, Mott S, Denis M, Martindale D, Frohlich J, et al. 1999. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nature Genetics* **22**:336–345. DOI: https://doi.org/10.1038/11905, PMID: 10431236

Brown MS, Goldstein JL. 1976. Analysis of a mutant strain of human fibroblasts with a defect in the internalization of receptor-bound low density lipoprotein. *Cell* **9**:663–674. DOI: https://doi.org/10.1016/0092-8674(76)90130-6, PMID: 189940

Bryois J, Calini D, Macnair W, Foo L, Urich E, Ortmann W, Iglesias VA, Selvaraj S, Nutma E, Marzin M, Amor S, Williams A, Castelo-Branco G, Menon V, De Jager P, Malhotra D. 2021. Cell-Type Specific Cis-EQTLs in Eight Brain Cell-Types Identifies Novel Risk Genes for Human Brain Disorders. [bioRxiv]. DOI: https://doi.org/10.1101/2021.10.09.21264604

Buccitelli C, Selbach M. 2020. Mrnas, proteins and the emerging principles of gene expression control. *Nature Reviews. Genetics* **21**:630–644. DOI: https://doi.org/10.1038/s41576-020-0258-4, PMID: 32709985

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, McVean G, Leslie S, Donnelly P, Marchini J. 2017. Genome-Wide Genetic Data on ~500,000 UK Biobank Participants. [bioRxiv]. DOI: https://doi.org/10.1101/166298

Cao H, Hegele RA. 2000. Nuclear lamin A/C R482Q mutation in Canadian kindreds with Dunnigan-type familial partial lipodystrophy. *Human Molecular Genetics* **9**:109–112. DOI: https://doi.org/10.1093/hmg/9.1.109, PMID: 10587585

Caputo V, Cianetti L, Niceta M, Carta C, Ciolfi A, Bocchinfuso G, Carrani E, Dentici ML, Biamino E, Belligni E, Garavelli L, Boccone L, Melis D, Andria G, Gelb BD, Stella L, Silengo M, Dallapiccola B, Tartaglia M. 2012. A

restricted spectrum of mutations in the Smad4 tumor-suppressor gene underlies myhre syndrome. *American Journal of Human Genetics* **90**:161–169. DOI: https://doi.org/10.1016/j.ajhg.2011.12.011, PMID: 22243968

Carta C, Pantaleoni F, Bocchinfuso G, Stella L, Vasta I, Sarkozy A, Digilio C, Palleschi A, Pizzuti A, Grammatico P, Zampino G, Dallapiccola B, Gelb BD, Tartaglia M. 2006. Germline missense mutations affecting KRAS isoform B are associated with a severe Noonan syndrome phenotype. *American Journal of Human Genetics* **79**:129–135. DOI: https://doi.org/10.1086/504394, PMID: 16773572

Cenarro A, Etxebarria A, de Castro-Orós I, Stef M, Bea AM, Palacios L, Mateo-Gallego R, Benito-Vicente A, Ostolaza H, Tejedor T, Martín C, Civeira F. 2016. The p.leu167del mutation in ApoE gene causes autosomal dominant hypercholesterolemia by down-regulation of LDL receptor expression in hepatocytes. *The Journal of Clinical Endocrinology and Metabolism* **101**:2113–2121. DOI: https://doi.org/10.1210/jc.2015-3874, PMID: 27014949

Cenik C, Cenik ES, Byeon GW, Grubert F, Candille SI, Spacek D, Alsallakh B, Tilgner H, Araya CL, Tang H, Ricci E, Snyder MP. 2015. Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Research* **25**:1610–1621. DOI: https://doi.org/10.1101/gr.193342.115, PMID: 26297486

Chan Y, Salem RM, Hsu Y-HH, McMahon G, Pers TH, Vedantam S, Esko T, Guo MH, Lim ET, GIANT Consortium, Franke L, Smith GD, Strachan DP, Hirschhorn JN. 2015. Genome-Wide analysis of body proportion classifies height-associated variants by mechanism of action and implicates genes important for skeletal development. *American Journal of Human Genetics* **96**:695–708. DOI: https://doi.org/10.1016/j.ajhg.2015.02.018, PMID: 25865494

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-Generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**:7. DOI: https://doi.org/10.1186/s13742-015-0047-8, PMID: 25722852

Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill GA, Gygi SP. 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**:500–505. DOI: https://doi.org/10.1038/nature18270, PMID: 27309819

Chun S, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, Sunyaev SR, Cotsapas C. 2017. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nature Genetics* **49**:600–605. DOI: https://doi.org/10.1038/ng.3795, PMID: 28218759

Clee SM, Loubser O, Collins J, Kastelein JJ, Hayden MR. 2001. The LPL S447X csnp is associated with decreased blood pressure and plasma triglycerides, and reduced risk of coronary artery disease. *Clinical Genetics* **60**:293–300. DOI: https://doi.org/10.1034/j.1399-0004.2001.600407.x, PMID: 11683775

Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, et al. 2010. Genome-wide association study of cnvs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**:713–720. DOI: https://doi.org/10.1038/nature08979, PMID: 20360734

Cummings BB, Karczewski KJ, Kosmicki JA, Seaby EG, Watts NA, Singer-Berk M, Mudge JM, Karjalainen J, Satterstrom FK, O'Donnell-Luria AH, Poterba T, Seed C, Solomonson M, Alföldi J, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Daly MJ, MacArthur DG. 2020. Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**:452–458. DOI: https://doi.org/10.1038/s41586-020-2329-2, PMID: 32461655

Cuomo ASE, Seaton DD, McCarthy DJ, Martinez I, Bonder MJ, Garcia-Bernardo J, Amatya S, Madrigal P, Isaacson A, Buettner F, Knights A, Natarajan KN, Vallier L, Marioni JC, Chhatriwala M, Stegle O, HipSci Consortium. 2020. Single-cell RNA-sequencing of differentiating ips cells reveals dynamic genetic effects on gene expression. *Nature Communications* **11**:810. DOI: https://doi.org/10.1038/s41467-020-14457-z, PMID: 32041960

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of samtools and bcftools. *GigaScience* **10**:giab008. DOI: https://doi.org/10.1093/gigascience/giab008, PMID: 33590861

Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS. 2012. Transcriptional burst frequency and burst size are equally modulated across the human genome. *PNAS* **109**:17454–17459. DOI: https://doi.org/10.1073/pnas.1213530109, PMID: 23064634

De Franco E, Shaw-Smith C, Flanagan SE, Shepherd MH, International NDM Consortium, Hattersley AT, Ellard S. 2013. Gata6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes* **62**:993–997. DOI: https://doi.org/10.2337/db12-0885, PMID: 23223019

de Knijff P, van den Maagdenberg AM, Frants RR, Havekes LM. 1994. Genetic heterogeneity of apolipoprotein E and its influence on plasma lipid and lipoprotein levels. *Human Mutation* **4**:178–194. DOI: https://doi.org/10.1002/humu.1380040303, PMID: 7833947

de Munnik SA, Bicknell LS, Aftimos S, Al-Aama JY, van Bever Y, Bober MB, Clayton-Smith J, Edrees AY, Feingold M, Fryer A, van Hagen JM, Hennekam RC, Jansweijer MCE, Johnson D, Kant SG, Opitz JM, Ramadevi AR, Reardon W, Ross A, Sarda P, et al. 2012. Meier-Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. *European Journal of Human Genetics* **20**:598–606. DOI: https://doi.org/10.1038/ejhg.2011.269, PMID: 22333897

De Paepe A, Nuytinck L, Raes M, Fryns JP. 1997. Homozygosity by descent for a COL1A2 mutation in two sibs with severe osteogenesis imperfecta and mild clinical expression in the heterozygotes. *Human Genetics* **99**:478–483. DOI: https://doi.org/10.1007/s004390050392, PMID: 9099837

Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, Pie J, Gil-Rodríguez C, Arnedo M, Loeys B, Kline AD, Wilson M, Lillquist K, Siu V, Ramos FJ, Musio A, Jackson LS, Dorsett D, Krantz ID. 2007. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *American Journal of Human Genetics* **80**:485–494. DOI: https://doi.org/10.1086/511888, PMID: 17273969

Deardorff MA, Wilde JJ, Albrecht M, Dickinson E, Tennstedt S, Braunholz D, Mönnich M, Yan Y, Xu W, Gil-Rodríguez MC, Clark D, Hakonarson H, Halbach S, Michelis LD, Rampuria A, Rossier E, Spranger S, Van Maldergem L, Lynch SA, Gillessen-Kaesbach G, et al. 2012. Rad21 mutations cause a human cohesinopathy. *American Journal of Human Genetics* **90**:1014–1027. DOI: https://doi.org/10.1016/j.ajhg.2012.04.019, PMID: 22633399

Deeb SS, Fajas L, Nemoto M, Pihlajamäki J, Mykkänen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J. 1998. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nature Genetics* **20**:284–287. DOI: https://doi.org/10.1038/3099, PMID: 9806549

Diaz-Gallo L-M, Espino-Paisán L, Fransen K, Gómez-García M, van Sommeren S, Cardeña C, Rodrigo L, Mendoza JL, Taxonera C, Nieto A, Alcain G, Cueto I, López-Nevot MA, Bottini N, Barclay ML, Crusius JB, van Bodegraven AA, Wijmenga C, Ponsioen CY, Gearry RB, et al. 2011. Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis. *Inflammatory Bowel Diseases* **17**:2287–2294. DOI: https://doi.org/10.1002/ibd.21630, PMID: 21287672

Dietlein F, Weghorn D, Taylor-Weiner A, Richters A, Reardon B, Liu D, Lander ES, Van Allen EM, Sunyaev SR. 2020. Identification of cancer driver genes based on nucleotide context. *Nature Genetics* **52**:208–218. DOI: https://doi.org/10.1038/s41588-019-0572-y, PMID: 32015527

Dobbyn A, Huckins LM, Boocock J, Sloofman LG, Glicksberg BS, Giambartolomei C, Hoffman GE, Perumal TM, Girdhar K, Jiang Y, Raj T, Ruderfer DM, Kramer RS, Pinto D, CommonMind Consortium, Akbarian S, Roussos P, Domenici E, Devlin B, Sklar P, et al. 2018. Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. *American Journal of Human Genetics* **102**:1169–1184. DOI: https://doi.org/10.1016/j.ajhg.2018.04.011, PMID: 29805045

Dori-Bachash M, Shema E, Tirosh I. 2011. Coupled evolution of transcription and mRNA degradation. *PLOS Biology* **9**:e1001106. DOI: https://doi.org/10.1371/journal.pbio.1001106, PMID: 21811398

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, et al. 2006. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science **314**:1461–1463. DOI: https://doi.org/10.1126/science.1135245, PMID: 17068223

Durocher F, Faure R, Labrie Y, Pelletier L, Bouchard I, Laframboise R. 2006. A novel mutation in the EIF2AK3 gene with variable expressivity in two patients with wolcott-rallison syndrome. *Clinical Genetics* **70**:34–38. DOI: https://doi.org/10.1111/j.1399-0004.2006.00632.x, PMID: 16813601

Ellinghaus D, Zhang H, Zeissig S, Lipinski S, Till A, Jiang T, Stade B, Bromberg Y, Ellinghaus E, Keller A, Rivas MA, Skieceviciene J, Doncheva NT, Liu X, Liu Q, Jiang F, Forster M, Mayr G, Albrecht M, Häsler R, et al. 2013. Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. *Gastroenterology* **145**:339–347. DOI: https://doi.org/10.1053/j.gastro.2013.04.040, PMID: 23624108

Ellis NA, Groden J, Ye TZ, Straughen J, Lennon DJ, Ciocci S, Proytcheva M, German J. 1995. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**:655–666. DOI: https://doi.org/10.1016/0092-8674(95)90105-1, PMID: 7585968

Elorbany R, Popp JM, Rhodes K, Strober BJ, Barr K, Qi G, Gilad Y, Battle A. 2022. Single-Cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLOS Genetics* **18**:e1009666. DOI: https://doi.org/10.1371/journal.pgen.1009666, PMID: 35061661

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**:215–216. DOI: https://doi.org/10.1038/nmeth.1906, PMID: 22373907

Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with chromhmm. *Nature Protocols* **12**:2478–2492. DOI: https://doi.org/10.1038/nprot.2017.124, PMID: 29120462

Faivre L, Gorlin RJ, Wirtz MK, Godfrey M, Dagoneau N, Samples JR, Le Merrer M, Collod-Beroud G, Boileau C, Munnich A, Cormier-Daire V. 2003. In frame fibrillin-1 gene deletion in autosomal dominant weill-marchesani syndrome. *Journal of Medical Genetics* **40**:34–36. DOI: https://doi.org/10.1136/jmg.40.1.34, PMID: 12525539

Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJH, Shishkin AA, Hatan M, Carrasco-Alfonso MJ, Mayer D, Luckey CJ, Patsopoulos NA, De Jager PL, Kuchroo VK, Epstein CB, Daly MJ, Hafler DA, et al. 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**:337–343. DOI: https://doi.org/10.1038/nature13835, PMID: 25363779

Findley AS, Monziani A, Richards AL, Rhodes K, Ward MC, Kalita CA, Alazizi A, Pazokitoroudi A, Sankararaman S, Wen X, Lanfear DE, Pique-Regi R, Gilad Y, Luca F. 2021. Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *eLife* **10**:e67077. DOI: https://doi.org/10.7554/eLife.67077, PMID: 33988505

Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, et al. 2008. Genetic determinants of ulcerative colitis include the ECM1 locus and five

loci implicated in Crohn's disease. *Nature Genetics* **40**:710–712. DOI: https://doi.org/10.1038/ng.145, PMID: 18438406

Foucault F, Vaury C, Barakat A, Thibout D, Planchon P, Jaulin C, Praz F, Amor-Guéret M. 1997. Characterization of a new BLM mutation associated with a topoisomerase II alpha defect in a patient with Bloom's syndrome. *Human Molecular Genetics* **6**:1427–1434. DOI: https://doi.org/10.1093/hmg/6.9.1427, PMID: 9285778

Fowler EV. 2005. Tnf and IL10 SNPs act together to predict disease behaviour in Crohn'' disease. *Journal of Medical Genetics* **42**:523–528. DOI: https://doi.org/10.1136/jmg.2004.027425, PMID: 15937090

Fowler EV, Doecke J, Simms LA, Zhao ZZ, Webb PM, Hayward NK, Whiteman DC, Florin TH, Montgomery GW, Cavanaugh JA, Radford-Smith GL. 2008. Atg16L1 T300A shows strong associations with disease subgroups in a large Australian IBD population: further support for significant disease heterogeneity. *The American Journal of Gastroenterology* **103**:2519–2526. DOI: https://doi.org/10.1111/j.1572-0241.2008.02023.x, PMID: 18671817

Freund MK, Burch KS, Shi H, Mancuso N, Kichaev G, Garske KM, Pan DZ, Miao Z, Mohlke KL, Laakso M, Pajukanta P, Pasaniuc B, Arboleda VA. 2018. Phenotype-Specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. *American Journal of Human Genetics* **103**:535–552. DOI: https://doi.org/10.1016/j.ajhg.2018.08.017, PMID: 30290150

Froguel P, Zouali H, Vionnet N, Velho G, Vaxillaire M, Sun F, Lesage S, Stoffel M, Takeda J, Passa P. 1993. Familial hyperglycemia due to mutations in glucokinase: definition of a subtype of diabetes mellitus. *The New England Journal of Medicine* **328**:697–702. DOI: https://doi.org/10.1056/NEJM199303113281005, PMID: 8433729

Fu J, Keurentjes JJB, Bouwmeester H, America T, Verstappen FWA, Ward JL, Beale MH, de Vos RCH, Dijkstra M, Scheltema RA, Johannes F, Koornneef M, Vreugdenhil D, Breitling R, Jansen RC. 2009. System-Wide molecular evidence for phenotypic buffering in Arabidopsis. *Nature Genetics* **41**:166–167. DOI: https://doi.org/10.1038/ng.308, PMID: 19169256

Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, Nguyen TH, Kane M, Perez EM, Durand NC, Lareau CA, Stamenova EK, Aiden EL, Lander ES, Engreitz JM. 2019. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nature Genetics* **51**:1664–1669. DOI: https://doi.org/10.1038/s41588-019-0538-0, PMID: 31784727

Galdzicka M, Patnala S, Hirshman MG, Cai J-F, Nitowsky H, Egeland JA, Ginns EI. 2002. A new gene, EVC2, is mutated in Ellis-van Creveld syndrome. *Molecular Genetics and Metabolism* **77**:291–295. DOI: https://doi.org/10.1016/s1096-7192(02)00178-6, PMID: 12468274

Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, Price AL. 2016a. Fast principal-component analysis reveals convergent evolution of ADH1B in europe and east asia. *American Journal of Human Genetics* **98**:456–472. DOI: https://doi.org/10.1016/j.ajhg.2015.12.022, PMID: 26924531

Galinsky KJ, Loh PR, Mallick S, Patterson NJ, Price AL. 2016b. Population structure of UK biobank and ancient eurasians reveals adaptation at genes influencing blood pressure. *American Journal of Human Genetics* **99**:1130–1139. DOI: https://doi.org/10.1016/j.ajhg.2016.09.014, PMID: 27773431

Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyler AE, Denny JC, GTEx Consortium, Nicolae DL, Cox NJ, Im HK. 2015. A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* **47**:1091–1098. DOI: https://doi.org/10.1038/ng.3367, PMID: 26258848

Gamazon ER, Zwinderman AH, Cox NJ, Denys D, Derks EM. 2019. Multi-Tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nature Genetics* **51**:933–940. DOI: https://doi.org/10.1038/s41588-019-0409-8, PMID: 31086352

Gasche C, Grundtner P, Zwirn P, Reinisch W, Shaw SH, Zdanov A, Sarma U, Williams LM, Foxwell BM, Gangl A. 2003. Novel variants of the IL-10 receptor 1 affect inhibition of monocyte TNF-alpha production. *Journal of Immunology* **170**:5578–5582. DOI: https://doi.org/10.4049/jimmunol.170.11.5578, PMID: 12759436

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, Park CC, Wen PZ, Brewer H, Weitz K, Camp DG, Pan C, Yordanova R, Neuhaus I, Tilford C, Siemers N, et al. 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLOS Genetics* **7**:e1001393. DOI: https://doi.org/10.1371/journal.pgen.1001393, PMID: 21695224

Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genetics* **10**:e1004383. DOI: https://doi.org/10.1371/journal.pgen.1004383, PMID: 24830394

Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, Pickrell J, Jaffe AE, Pasaniuc B, Roussos P, CommonMind Consortium. 2018. A bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**:2538–2545. DOI: https://doi.org/10.1093/bioinformatics/bty147, PMID: 29579179

Glas J, Seiderer J, Wetzke M, Konrad A, Török H-P, Schmechel S, Tonenchi L, Grassl C, Dambacher J, Pfennig S, Maier K, Griga T, Klein W, Epplen JT, Schiemann U, Folwaczny C, Lohse P, Göke B, Ochsenkühn T, Müller-Myhsok B, et al. 2007. 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. DOI: https://doi.org/10.1371/journal.pone.0000819, PMID: 17786191

Glastonbury CA, Couto Alves A, El-Sayed Moustafa JS, Small KS. 2019. Cell-Type heterogeneity in adipose tissue is associated with complex traits and reveals disease-relevant cell-specific eQTLs. *American Journal of Human Genetics* **104**:1013–1024. DOI: https://doi.org/10.1016/j.ajhg.2019.03.025, PMID: 31130283

Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schäffer AA, Noyan F, Perro M, Diestelhorst J, Allroth A, Murugan D, Hätscher N, Pfeifer D, Sykora KW, Sauer M, Kreipe H, Lacher M, Nustede R, Woellner C,

Baumann U, Salzer U, et al. 2009. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *The New England Journal of Medicine* **361**:2033–2045. DOI: https://doi.org/10.1056/NEJMoa0907206, PMID: 19890111

Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, Howard N, Srinivasan S, Silva J, Molnes J, Edghill EL, Frayling TM, Temple IK, Mackay D, Shield JPH, Sumnik Z, van Rhijn A, Wales JKH, Clark P, Gorman S, et al. 2004. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit kir6.2 and permanent neonatal diabetes. *The New England Journal of Medicine* **350**:1838–1849. DOI: https://doi.org/10.1056/NEJMoa032922, PMID: 15115830

Glueck CJ, Fallat RW, Millett F, Steiner PM. 1975. Familial hyperalphalipoproteinemia. *Archives of Internal Medicine* **135**:1025–1028 PMID: 168823.

Goddard AD, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMS. 1995. Mutations of the growth hormone receptor in children with idiopathic short stature. *New England Journal of Medicine* **333**:1093–1098. DOI: https://doi.org/10.1056/NEJM199510263331701, PMID: 7565946

Goel H, Parasivam G. 2020. Another case of holoprosencephaly associated with Rad21 loss-of-function variant. *Brain* **143**:e64. DOI: https://doi.org/10.1093/brain/awaa173, PMID: 32696056

Goldstein JL, Brown MS. 1973. Familial hypercholesterolemia: identification of A defect in the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity associated with overproduction of cholesterol. *PNAS* **70**:2804–2808. DOI: https://doi.org/10.1073/pnas.70.10.2804, PMID: 4355366

Gos M, Fahiminiya S, Poznański J, Klapecki J, Obersztyn E, Piotrowicz M, Wierzba J, Posmyk R, Bal J, Majewski J. 2014. Contribution of RIT1 mutations to the pathogenesis of Noonan syndrome: four new cases and further evidence of heterogeneity . *American Journal of Medical Genetics Part A* **164**:2310–2316. DOI: https://doi.org/10.1002/ajmg.a.36646

Goto M, Imamura O, Kuromitsu J, Matsumoto T, Yamabe Y, Tokutake Y, Suzuki N, Mason B, Drayna D, Sugawara M, Sugimoto M, Furuichi Y. 1997. Analysis of helicase gene mutations in Japanese Werner's syndrome patients. *Human Genetics* **99**:191–193. DOI: https://doi.org/10.1007/s004390050336, PMID: 9048918

Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, Ripke S, Gusareva ES, Annese V, Hauser SL, Oksenberg JR, Thomsen I, Leslie S, International Inflammatory Bowel Disease Genetics Consortium, Australia and New Zealand IBDGC, Belgium IBD Genetics Consortium, Italian Group for IBD Genetic Consortium, NIDDK Inflammatory Bowel Disease Genetics Consortium, United Kingdom IBDGC, Wellcome Trust Case Control Consortium, et al. 2015. 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. *Nature Genetics* **47**:172–179. DOI: https://doi.org/10.1038/ng.3176, PMID: 25559196

Gradwohl G, Dierich A, LeMeur M, Guillemot F. 2000. Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *PNAS* **97**:1607–1611. DOI: https://doi.org/10.1073/pnas.97.4.1607, PMID: 10677506

Grarup N, Andreasen CH, Andersen MK, Albrechtsen A, Sandbaek A, Lauritzen T, Borch-Johnsen K, Jørgensen T, Schmitz O, Hansen T, Pedersen O. 2008. 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. *The Journal of Clinical Endocrinology and Metabolism* **93**:2294–2299. DOI: https://doi.org/10.1210/jc.2007-2815, PMID: 18364377

Griffith E, Walker S, Martin C-A, Vagnarelli P, Stiff T, Vernay B, Al Sanna N, Saggar A, Hamel B, Earnshaw WC, Jeggo PA, Jackson AP, O'Driscoll M. 2008. Mutations in pericentrin cause seckel syndrome with defective ATR-dependent DNA damage signaling. *Nature Genetics* **40**:232–236. DOI: https://doi.org/10.1038/ng.2007.80, PMID: 18157127

GTEx Consortium. 2020. The gtex Consortium atlas of genetic regulatory effects across human tissues. Science **369**:1318–1330. DOI: https://doi.org/10.1126/science.aaz1776, PMID: 32913098

Guernsey DL, Matsuoka M, Jiang H, Evans S, Macgillivray C, Nightingale M, Perry S, Ferguson M, LeBlanc M, Paquette J, Patry L, Rideout AL, Thomas A, Orr A, McMaster CR, Michaud JL, Deal C, Langlois S, Superneau DW, Parkash S, et al. 2011. Mutations in origin recognition complex gene Orc4 cause Meier-Gorlin syndrome. *Nature Genetics* **43**:360–364. DOI: https://doi.org/10.1038/ng.777, PMID: 21358631

Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E, Schizophrenia Working Group of the Psychiatric Genomics Consortium, SWE-SCZ Consortium, Kähler AK, Hultman CM, Purcell SM, McCarroll SA, Daly M, Pasaniuc B, Sullivan PF, Neale BM, et al. 2014. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *American Journal of Human Genetics* **95**:535–552. DOI: https://doi.org/10.1016/j.ajhg.2014.10.004, PMID: 25439723

Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, Jansen R, de Geus EJC, Boomsma DI, Wright FA, Sullivan PF, Nikkola E, Alvarez M, Civelek M, Lusis AJ, Lehtimäki T, Raitoharju E, Kähönen M, Seppälä I, Raitakari OT, et al. 2016. Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**:245–252. DOI: https://doi.org/10.1038/ng.3506, PMID: 26854917

Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Günther S, Prescott NJ, Onnie CM, Häsler R, Sipos B, Fölsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, et al. 2007. A genome-wide association scan of nonsynonymous snps identifies A susceptibility variant for crohn disease in ATG16L1. *Nature Genetics* **39**:207–211. DOI: https://doi.org/10.1038/ng1954, PMID: 17200669

Hani EH, Boutin P, Durand E, Inoue H, Permutt MA, Velho G, Froguel P. 1998. 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. DOI: https://doi.org/10.1007/s001250051098, PMID: 9867219

Hani EH, Stoffers DA, Chèvre JC, Durand E, Stanojevic V, Dina C, Habener JF, Froguel P. 1999. Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus. *The Journal of Clinical Investigation* **104**:R41–R48. DOI: https://doi.org/10.1172/JCI7469, PMID: 10545531

Harding HP, Zeng H, Zhang Y, Jungries R, Chung P, Plesken H, Sabatini DD, Ron D. 2001. Diabetes mellitus and exocrine pancreatic dysfunction in perk-/- mice reveals a role for translational control in secretory cell survival. *Molecular Cell* **7**:1153–1163. DOI: https://doi.org/10.1016/s1097-2765(01)00264-7, PMID: 11430819

Hardy C, Khanim F, Torres R, Scott-Brown M, Seller A, Poulton J, Collier D, Kirk J, Polymeropoulos M, Latif F, Barrett T. 1999. Clinical and molecular genetic analysis of 19 Wolfram syndrome kindreds demonstrating a wide spectrum of mutations in WFS1. *American Journal of Human Genetics* **65**:1279–1290. DOI: https://doi.org/10.1086/302609, PMID: 10521293

Hästbacka J, Salonen R, Laurila P, de la Chapelle A, Kaitila I. 1993. Prenatal diagnosis of diastrophic dysplasia with polymorphic DNA markers. *Journal of Medical Genetics* **30**:265–268. DOI: https://doi.org/10.1136/jmg.30.4.265, PMID: 8487268

HaVinh R, Alanay Y, Bank RA, Campos-Xavier AB, Zankl A, Superti-Furga A, Bonafé L. 2004. Phenotypic and molecular characterization of bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. *American Journal of Medical Genetics Part A* **131A**:115–120. DOI: https://doi.org/10.1002/ajmg.a.30231

Hegele RA, Connelly PW, Maguire GF, Huff MW, Leiter L, Wolfe BM, Evans AJ, Little JA. 1991. An apolipoprotein CII mutation, ciilys19 -- -- 'hr' identified in patients with hyperlipidemia. *Disease Markers* **9**:73–80 PMID: 1782747.

Heizmann C, Kirchgessner T, Kwiterovich PO, Ladias JA, Derby C, Antonarakis SE, Lusis AJ. 1991. Dna polymorphism haplotypes of the human lipoprotein lipase gene: possible association with high density lipoprotein levels. *Human Genetics* **86**:578–584. DOI: https://doi.org/10.1007/BF00201544, PMID: 1673959

Hood RL, Lines MA, Nikkel SM, Schwartzentruber J, Beaulieu C, Nowaczyk MJM, Allanson J, Kim CA, Wieczorek D, Moilanen JS, Lacombe D, Gillessen-Kaesbach G, Whiteford ML, Quaio CRDC, Gomy I, Bertola DR, Albrecht B, Platzer K, McGillivray G, Zou R, et al. 2012. Mutations in SRCAP, encoding SNF2-related CREBBP activator protein, cause floating-harbor syndrome. *American Journal of Human Genetics* **90**:308–313. DOI: https://doi.org/10.1016/j.ajhg.2011.12.001, PMID: 22265015

Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, Lindner T, Yamagata K, Ogata M, Tomonaga O, Kuroki H, Kasahara T, Iwamoto Y, Bell GI. 1997. Mutation in hepatocyte nuclear factor-1 beta gene (Tcf2) associated with MODY. *Nature Genetics* **17**:384–385. DOI: https://doi.org/10.1038/ng1297-384, PMID: 9398836

Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. 2016. Colocalization of GWAS and eQTL signals detects target genes. *American Journal of Human Genetics* **99**:1245–1260. DOI: https://doi.org/10.1016/j.ajhg.2016.10.003, PMID: 27866706

Horn D, Robinson PN. 2011. Progeroid facial features and lipodystrophy associated with a novel splice site mutation in the final intron of the FBN1 gene. *American Journal of Medical Genetics Part A* **155**:721–724. DOI: https://doi.org/10.1002/ajmg.a.33905

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**:599–603. DOI: https://doi.org/10.1038/35079107, PMID: 11385576

Huh D, Paulsson J. 2011. Non-Genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics* **43**:95–100. DOI: https://doi.org/10.1038/ng.729, PMID: 21186354

Hukku A, Pividori M, Luca F, Pique-Regi R, Im HK, Wen X. 2021. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *American Journal of Human Genetics* **108**:25–35. DOI: https://doi.org/10.1016/j.ajhg.2020.11.012, PMID: 33308443

Hyland VJ, Robertson SP, Flanagan S, Savarirayan R, Roscioli T, Masel J, Hayes M, Glass IA. 2003. Somatic and germline mosaicism for a R248C missense mutation in FGFR3, resulting in a skeletal dysplasia distinct from thanatophoric dysplasia. *American Journal of Medical Genetics. Part A* **120A**:157–168. DOI: https://doi.org/10.1002/ajmg.a.20012, PMID: 12833394

Iijima H, Emi M, Wada M, Daimon M, Toriyama S, Koyano S, Sato H, Hopkins PN, Hunt SC, Kubota I, Kawata S, Kato T. 2008. Association of an intronic haplotype of the LIPC gene with hyperalphalipoproteinemia in two independent populations. *Journal of Human Genetics* **53**:193–200. DOI: https://doi.org/10.1007/s10038-007-0236-0, PMID: 18160998

Isaacs A, Sayed-Tabatabaei FA, Njajou OT, Witteman JCM, van Duijn CM. 2004. The -514 C- > T hepatic lipase promoter region polymorphism and plasma lipids: a meta-analysis. *The Journal of Clinical Endocrinology and Metabolism* **89**:3858–3863. DOI: https://doi.org/10.1210/jc.2004-0188, PMID: 15292318

Jerber J, Seaton DD, Cuomo ASE, Kumasaka N, Haldane J, Steer J, Patel M, Pearce D, Andersson M, Bonder MJ, Mountjoy E, Ghoussaini M, Lancaster MA, HipSci Consortium, Marioni JC, Merkle FT, Gaffney DJ, Stegle O. 2021. Population-Scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nature Genetics* **53**:304–312. DOI: https://doi.org/10.1038/s41588-021-00801-6, PMID: 33664506

Jiang XC, Bruce C, Mar J, Lin M, Ji Y, Francone OL, Tall AR. 1999. Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. *The Journal of Clinical Investigation* **103**:907–914. DOI: https://doi.org/10.1172/JCI5578, PMID: 10079112

Kathiresan S, Srivastava D. 2012. Genetics of human cardiovascular disease. *Cell* **148**:1242–1257. DOI: https://doi.org/10.1016/j.cell.2012.03.001, PMID: 22424232

Kawashiri M, Zhang Y, Usher D, Reilly M, Puré E, Rader DJ. 2001. Effects of coexpression of the LDL receptor and APOE on cholesterol metabolism and atherosclerosis in LDL receptor-deficient mice. *Journal of Lipid Research* **42**:943–950. DOI: https://doi.org/10.1016/S0022-2275(20)31618-7, PMID: 11369802

Kelley BP, Malfait F, Bonafe L, Baldridge D, Homan E, Symoens S, Willaert A, Elcioglu N, Van Maldergem L, Verellen-Dumoulin C, Gillerot Y, Napierala D, Krakow D, Beighton P, Superti-Furga A, De Paepe A, Lee B. 2011. Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. *Journal of Bone and Mineral Research* **26**:666–672. DOI: https://doi.org/10.1002/jbmr.250, PMID: 20839288

Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. 2013. Primate transcript and protein expression levels evolve under compensatory selection pressures. Science **342**:1100–1104. DOI: https://doi.org/10.1126/science.1242379, PMID: 24136357

Khanim F, Kirk J, Latif F, Barrett TG. 2001. Wfs1/Wolframin mutations, Wolfram syndrome, and associated diseases. *Human Mutation* **17**:357–367. DOI: https://doi.org/10.1002/humu.1110, PMID: 11317350

Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, Becker J, Kaiser V, Fricker N, Beier E, Boor P, Castel SE, Nöthen MM, Barreiro LB, Pickrell JK, Müller-Myhsok B, Lappalainen T, Schumacher J, Hornung V. 2017. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nature Communications* **8**:266. DOI: https://doi.org/10.1038/s41467-017-00366-1, PMID: 28814792

Knowles DA, Davis JR, Edgington H, Raj A, Favé M-J, Zhu X, Potash JB, Weissman MM, Shi J, Levinson DF, Awadalla P, Mostafavi S, Montgomery SB, Battle A. 2017. Allele-Specific expression reveals interactions between genetic variation and environment. *Nature Methods* **14**:699–702. DOI: https://doi.org/10.1038/nmeth.4298, PMID: 28530654

Kosaki K, Suzuki T, Muroya K, Hasegawa T, Sato S, Matsuo N, Kosaki R, Nagai T, Hasegawa Y, Ogata T. 2002. Ptpn11 (protein-tyrosine phosphatase, nonreceptor-type 11) mutations in seven Japanese patients with Noonan syndrome. *The Journal of Clinical Endocrinology and Metabolism* **87**:3529–3533. DOI: https://doi.org/10.1210/jcem.87.8.8694, PMID: 12161469

Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, Yaeger D, Jukofsky L, Wasserman N, Bottani A, Morris CA, Nowaczyk MJM, Toriello H, Bamshad MJ, Carey JC, Rappaport E, Kawauchi S, Lander AD, Calof AL, Li H-H, Devoto M, et al. 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature Genetics* **36**:631–635. DOI: https://doi.org/10.1038/ng1364, PMID: 15146186

Kruszka P, Berger SI, Casa V, Dekker MR, Gaesser J, Weiss K, Martinez AF, Murdock DR, Louie RJ, Prijoles EJ, Lichty AW, Brouwer OF, Zonneveld-Huijssoon E, Stephan MJ, Hogue J, Hu P, Tanima-Nagai M, Everson JL, Prasad C, Cereda A, et al. 2019. Cohesin complex-associated holoprosencephaly. *Brain* **142**:2631–2643. DOI: https://doi.org/10.1093/brain/awz210, PMID: 31334757

Kumar N, Singh A, Kulkarni RV. 2015. Transcriptional bursting in gene expression: analytical results for general stochastic models. *PLOS Computational Biology* **11**:e1004292. DOI: https://doi.org/10.1371/journal.pcbi.1004292, PMID: 26474290

Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**:317–330. DOI: https://doi.org/10.1038/nature14248, PMID: 25693563

Kusnadi EP, Timpone C, Topisirovic I, Larsson O, Furic L. 2022. Regulation of gene expression via translational buffering. *Biochimica et Biophysica Acta. Molecular Cell Research* **1869**:119140. DOI: https://doi.org/10.1016/j.bbamcr.2021.119140, PMID: 34599983

Labay V, Raz T, Baron D, Mandel H, Williams H, Barrett T, Szargel R, McDonald L, Shalata A, Nosaka K, Gregory S, Cohen N. 1999. Mutations in SLC19A2 cause thiamine-responsive megaloblastic anaemia associated with diabetes mellitus and deafness. *Nature Genetics* **22**:300–304. DOI: https://doi.org/10.1038/10372, PMID: 10391221

Larson DR, Singer RH, Zenklusen D. 2009. A single molecule view of gene expression. *Trends in Cell Biology* **19**:630–637. DOI: https://doi.org/10.1016/j.tcb.2009.08.008, PMID: 19819144

Laukkanen O, Lindström J, Eriksson J, Valle TT, Hämäläinen H, Ilanne-Parikka P, Keinänen-Kiukaanniemi S, Tuomilehto J, Uusitupa M, Laakso M, Finnish Diabetes Prevention Study. 2005. 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. DOI: https://doi.org/10.2337/diabetes.54.7.2256, PMID: 15983230

Le Goff C, Mahaut C, Wang LW, Allali S, Abhyankar A, Jensen S, Zylberberg L, Collod-Beroud G, Bonnet D, Alanay Y, Brady AF, Cordier MP, Devriendt K, Genevieve D, Kiper PÖS, Kitoh H, Krakow D, Lynch SA, Le Merrer M, Mégarbane A, et al. 2011. Mutations in the TGFβ binding-protein-like domain 5 of FBN1 are responsible for acromicric and geleophysic dysplasias. *American Journal of Human Genetics* **89**:7–14. DOI: https://doi.org/10.1016/j.ajhg.2011.05.012, PMID: 21683322

Le Goff C, Mahaut C, Abhyankar A, Le Goff W, Serre V, Afenjar A, Destrée A, di Rocco M, Héron D, Jacquemont S, Marlin S, Simon M, Tolmie J, Verloes A, Casanova JL, Munnich A, Cormier-Daire V. 2012. Mutations at a single codon in mad homology 2 domain of smad4 cause myhre syndrome. *Nature Genetics* **44**:85–88. DOI: https://doi.org/10.1038/ng.1016, PMID: 22158539

Le Goff C, Mahaut C, Bottani A, Doray B, Goldenberg A, Moncla A, Odent S, Nitschke P, Munnich A, Faivre L, Cormier-Daire V. 2013. Not all floating-harbor syndrome cases are due to mutations in exon 34 of SRCAP. *Human Mutation* **34**:88–92. DOI: https://doi.org/10.1002/humu.22216, PMID: 22965468

Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L, Laukens D, Mni M, Zelenika D, Van Gossum A, Rutgeerts P, et al. 2007. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLOS Genetics* **3**:e58. DOI: https://doi.org/10.1371/journal.pgen.0030058, PMID: 17447842

Lindner TH, Njolstad PR, Horikawa Y, Bostad L, Bell GI, Sovik O. 1999. 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-1beta. *Human Molecular Genetics* **8**:2001–2008. DOI: https://doi.org/10.1093/hmg/8.11.2001, PMID: 10484768

Lindor NM, Furuichi Y, Kitao S, Shimamoto A, Arndt C, Jalal S. 2000. Rothmund-Thomson syndrome due to RecQ4 helicase mutations: report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome. *American Journal of Medical Genetics* **90**:223–228. DOI: https://doi.org/10.1002/(sici)1096-8628(20000131)90:3<223::aid-ajmg7>3.0.co;2-z, PMID: 10678659

Lindor NM, Gunawardena SR, Thibodeau SN. 2012. Mutations of Smad4 account for both LAPS and myhre syndromes. *American Journal of Medical Genetics. Part A* **158A**:1520–1521. DOI: https://doi.org/10.1002/ajmg.a.35374, PMID: 22585601

Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T, Abedian S, Cheon JH, Cho J, Dayani NE, Franke L, Fuyuno Y, Hart A, Juyal RC, Juyal G, Kim WH, et al. 2015. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* **47**:979–986. DOI: https://doi.org/10.1038/ng.3359, PMID: 26192919

Liu Y, Beyer A, Aebersold R. 2016. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**:535–550. DOI: https://doi.org/10.1016/j.cell.2016.03.014, PMID: 27104977

Mabuchi A, Manabe N, Haga N, Kitoh H, Ikeda T, Kawaji H, Tamai K, Hamada J, Nakamura S, Brunetti-Pierri N, Kimizuka M, Takatori Y, Nakamura K, Nishimura G, Ohashi H, Ikegawa S. 2003. Novel types of COMP mutations and genotype-phenotype association in pseudoachondroplasia and multiple epiphyseal dysplasia. *Human Genetics* **112**:84–90. DOI: https://doi.org/10.1007/s00439-002-0845-9, PMID: 12483304

Macfarlane WM, Frayling TM, Ellard S, Evans JC, Allen LI, Bulman MP, Ayres S, Shepherd M, Clark P, Millward A, Demaine A, Wilkin T, Docherty K, Hattersley AT. 1999. Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. *The Journal of Clinical Investigation* **104**:R33–R39. DOI: https://doi.org/10.1172/JCI7449, PMID: 10545530

Mackay DJG, Callaway JLA, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, Hahnemann JMD, Kordonouri O, Masoud AF, Oestergaard E, Storr J, Ellard S, Hattersley AT, Robinson DO, Temple IK. 2008. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in Zfp57. *Nature Genetics* **40**:949–951. DOI: https://doi.org/10.1038/ng.187, PMID: 18622393

Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, et al. 2018. Fine-Mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics* **50**:1505–1513. DOI: https://doi.org/10.1038/s41588-018-0241-6, PMID: 30297969

Maheshwari M, Belmont J, Fernbach S, Ho T, Molinari L, Yakub I, Yu F, Combes A, Towbin J, Craigen WJ, Gibbs R. 2002. Ptpn11 mutations in Noonan syndrome type I: detection of recurrent mutations in exons 3 and 13. *Human Mutation* **20**:298–304. DOI: https://doi.org/10.1002/humu.10129, PMID: 12325025

Makrythanasis P, Temtamy S, Aglan MS, Otaify GA, Hamamy H, Antonarakis SE. 2014. A novel homozygous mutation in FGFR3 causes tall stature, severe lateral tibial deviation, scoliosis, hearing impairment, camptodactyly, and arachnodactyly. *Human Mutation* **35**:959–963. DOI: https://doi.org/10.1002/humu.22597, PMID: 24864036

Malecki MT, Jhala US, Antonellis A, Fields L, Doria A, Orban T, Saad M, Warram JH, Montminy M, Krolewski AS. 1999. Mutations in NeuroD1 are associated with the development of type 2 diabetes mellitus. *Nature Genetics* **23**:323–328. DOI: https://doi.org/10.1038/15500, PMID: 10545951

Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. 2017. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. *American Journal of Human Genetics* **100**:473–487. DOI: https://doi.org/10.1016/j.ajhg.2017.01.031, PMID: 28238358

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**:747–753. DOI: https://doi.org/10.1038/nature08494, PMID: 19812666

Mao H, Yang W, Lee PPW, Ho MH-K, Yang J, Zeng S, Chong C-Y, Lee T-L, Tu W, Lau Y-L. 2012. Exome sequencing identifies novel compound heterozygous mutations of IL-10 receptor 1 in neonatal-onset Crohn's disease. *Genes & Immunity* **13**:437–442. DOI: https://doi.org/10.1038/gene.2012.8, PMID: 22476154

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science **337**:1190–1195. DOI: https://doi.org/10.1126/science.1222794, PMID: 22955828

Mauvais-Jarvis F, Smith SB, Le May C, Leal SM, Gautier J-F, Molokhia M, Riveline J-P, Rajan AS, Kevorkian J-P, Zhang S, Vexiau P, German MS, Vaisse C. 2004. Pax4 gene variations predispose to ketosis-prone diabetes. *Human Molecular Genetics* **13**:3151–3159. DOI: https://doi.org/10.1093/hmg/ddh341, PMID: 15509590

McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, Duerr RH, Silverberg MS, Taylor KD, Rioux JD, Altshuler D, Daly MJ, Xavier RJ. 2008. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nature Genetics* **40**:1107–1112. DOI: https://doi.org/10.1038/ng.215, PMID: 19165925

McCarthy JJ, Lehner T, Reeves C, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ, Genequest investigators. 2003. Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. *Journal of Medical Genetics* **40**:453–458. DOI: https://doi.org/10.1136/jmg.40.6.453, PMID: 12807968

McKusick-nathans institute of genetic medicine. 2021. Online Mendelian Inheritance in Man, OMIM. Baltimore, MD: johns hopkins university.

McManus CJ, May GE, Spealman P, Shteyman A. 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Research* **24**:422–430. DOI: https://doi.org/10.1101/gr.164996.113, PMID: 24318730

McManus J, Cheng Z, Vogel C. 2015. Next-Generation analysis of gene expression regulation -- comparing the roles of synthesis and degradation. *Molecular BioSystems* **11**:2680–2689. DOI: https://doi.org/10.1039/c5mb00310e, PMID: 26259698

Menke LA, van Belzen MJ, Alders M, Cristofoli F, DDD Study, Ehmke N, Fergelot P, Foster A, Gerkes EH, Hoffer MJV, Horn D, Kant SG, Lacombe D, Leon E, Maas SM, Melis D, Muto V, Park S-M, Peeters H, Peters DJM, et al. 2016. CREBBP mutations in individuals without Rubinstein-Taybi syndrome phenotype. *American Journal of Medical Genetics. Part A* **170**:2681–2693. DOI: https://doi.org/10.1002/ajmg.a.37800, PMID: 27311832

Menke LA, DDD study, Gardeitchik T, Hammond P, Heimdal KR, Houge G, Hufnagel SB, Ji J, Johansson S, Kant SG, Kinning E, Leon EL, Newbury-Ecob R, Paolacci S, Pfundt R, Ragge NK, Rinne T, Ruivenkamp C, Saitta SC, Sun Y, et al. 2018. Further delineation of an entity caused by CREBBP and ep300 mutations but not resembling Rubinstein-Taybi syndrome. *American Journal of Medical Genetics. Part A* **176**:862–876. DOI: https://doi.org/10.1002/ajmg.a.38626, PMID: 29460469

Meuleman W, Muratov A, Rynes E, Halow J, Lee K, Bates D, Diegel M, Dunn D, Neri F, Teodosiadis A, Reynolds A, Haugen E, Nelson J, Johnson A, Frerker M, Buckley M, Sandstrom R, Vierstra J, Kaul R, Stamatoyannopoulos J. 2020. Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**:244–251. DOI: https://doi.org/10.1038/s41586-020-2559-3, PMID: 32728217

Mostafavi H, Spence JP, Naqvi S, Pritchard JK. 2022. Limited Overlap of EQTLs and GWAS Hits Due to Systematic Differences in Discovery. [bioRxiv]. DOI: https://doi.org/10.1101/2022.05.07.491045

Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, Fumis L, Hayhurst J, Buniello A, Karim MA, Wright D, Hercules A, Papa E, Fauman EB, Barrett JC, Todd JA, Ochoa D, Dunham I, Ghoussaini M. 2021. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nature Genetics* **53**:1527–1533. DOI: https://doi.org/10.1038/s41588-021-00945-5, PMID: 34711957

Mu Z, Wei W, Fair B, Miao J, Zhu P, Li YI. 2021. The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biology* **22**:122. DOI: https://doi.org/10.1186/s13059-021-02334-x, PMID: 33926512

Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, Mualim K, Natri HM, Weeks EM, Munson G, Kane M, Kang HY, Cui A, Ray JP, Eisenhaure TM, Collins RL, et al. 2021. Genome-Wide enhancer maps link risk variants to disease genes. *Nature* **593**:238–243. DOI: https://doi.org/10.1038/s41586-021-03446-x, PMID: 33828297

Nathan A, Asgari S, Ishigaki K, Valencia C, Amariuta T, Luo Y, Beynor JI, Baglaenko Y, Suliman S, Price AL, Lecca L, Murray MB, Moody DB, Raychaudhuri S. 2022. Single-Cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**:120–128. DOI: https://doi.org/10.1038/s41586-022-04713-1, PMID: 35545678

Neavin D, Nguyen Q, Daniszewski MS, Liang HH, Chiu HS, Wee YK, Senabouth A, Lukowski SW, Crombie DE, Lidgerwood GE, Hernández D, Vickers JC, Cook AL, Palpant NJ, Pébay A, Hewitt AW, Powell JE. 2021. Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biology* **22**:76. DOI: https://doi.org/10.1186/s13059-021-02293-3, PMID: 33673841

Neve B, Fernandez-Zapico ME, Ashkenazi-Katalan V, Dina C, Hamid YH, Joly E, Vaillant E, Benmezroua Y, Durand E, Bakaher N, Delannoy V, Vaxillaire M, Cook T, Dallinga-Thie GM, Jansen H, Charles MA, Clément K, Galan P, Hercberg S, Helbecque N, et al. 2005. Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. *PNAS* **102**:4807–4812. DOI: https://doi.org/10.1073/pnas.0409177102, PMID: 15774581

Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLOS Genetics* **6**:e1000888. DOI: https://doi.org/10.1371/journal.pgen.1000888, PMID: 20369019

Nicolas D, Phillips NE, Naef F. 2017. What shapes eukaryotic transcriptional bursting? *Molecular BioSystems* **13**:1280–1290. DOI: https://doi.org/10.1039/c7mb00154a, PMID: 28573295

NJC12. 2023. Missing_link_association_function. swh:1:rev:46d9072b7cc13f6532203d1494eec4d0f634e092. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:d653f3cd4f7d26dce8c2a13d092f71a0ce2c3f95;origin=https://github.com/NJC12/missing_link_association_function;visit=swh:1:snp:19d7685a753ff743d14a711cf08a068a0d77d690;anchor=swh:1:rev:46d9072b7cc13f6532203d1494eec4d0f634e092

O'Driscoll M, Ruiz-Perez VL, Woods CG, Jeggo PA, Goodship JA. 2003. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in seckel syndrome. *Nature Genetics* **33**:497–501. DOI: https://doi.org/10.1038/ng1129, PMID: 12640452

Oelen R, de Vries DH, Brugge H, Gordon MG, Vochteloo M, Ye CJ, Westra HJ, Franke L, van der Wijst MGP, single-cell eQTLGen consortium, BIOS Consortium. 2022. Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. *Nature Communications* **13**:3267. DOI: https://doi.org/10.1038/s41467-022-30893-5, PMID: 35672358

Ogi T, Walker S, Stiff T, Hobson E, Limsirichaikul S, Carpenter G, Prescott K, Suri M, Byrd PJ, Matsuse M, Mitsutake N, Nakazawa Y, Vasudevan P, Barrow M, Stewart GS, Taylor AMR, O'Driscoll M, Jeggo PA. 2012. Identification of the first ATRIP-deficient patient and novel mutations in ATR define a clinical spectrum for ATR-ATRIP seckel syndrome. *PLOS Genetics* **8**:e1002945. DOI: https://doi.org/10.1371/journal.pgen.1002945, PMID: 23144622

Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nuñez G, Cho JH. 2001. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**:603–606. DOI: https://doi.org/10.1038/35079114, PMID: 11385577

Oishi K, Hofmann S, Diaz GA, Brown T, Manwani D, Ng L, Young R, Vlassara H, Ioannou YA, Forrest D, Gelb BD. 2002. Targeted disruption of slc19a2, the gene encoding the high-affinity thiamin transporter thtr-1, causes diabetes mellitus, sensorineural deafness and megaloblastosis in mice. *Human Molecular Genetics* **11**:2951–2960. DOI: https://doi.org/10.1093/hmg/11.23.2951, PMID: 12393806

Ordovas JM, Schaefer EJ, Salem D, Ward RH, Glueck CJ, Vergani C, Wilson PW, Karathanasis SK. 1986. Apolipoprotein A-I gene polymorphism associated with premature coronary artery disease and familial hypoalphalipoproteinemia. *The New England Journal of Medicine* **314**:671–677. DOI: https://doi.org/10.1056/NEJM198603133141102, PMID: 3081805

Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, Yanaoka H, Kobayashi S, Okubo M, Shirai H, Sugimori Y, Maeda J, Nakano M, Yamada S, Yoshida R, Tsuchiya H, Tsuchida Y, Akizuki S, Yoshifuji H, Ohmura K, et al. 2021. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**:3006–3021.. DOI: https://doi.org/10.1016/j.cell.2021.03.056, PMID: 33930287

Pai AA, Cain CE, Mizrahi-Man O, De Leon S, Lewellen N, Veyrieras J-B, Degner JF, Gaffney DJ, Pickrell JK, Stephens M, Pritchard JK, Gilad Y. 2012. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLOS Genetics* **8**:e1003000. DOI: https://doi.org/10.1371/journal.pgen.1003000, PMID: 23071454

Pai AA, Pritchard JK, Gilad Y. 2015. The genetic and mechanistic basis for variation in gene regulation. *PLOS Genetics* **11**:e1004857. DOI: https://doi.org/10.1371/journal.pgen.1004857, PMID: 25569255

Pandit B, Sarkozy A, Pennacchio LA, Carta C, Oishi K, Martinelli S, Pogna EA, Schackwitz W, Ustaszewska A, Landstrom A, Bos JM, Ommen SR, Esposito G, Lepri F, Faul C, Mundel P, López Siguero JP, Tenconi R, Selicorni A, Rossi C, et al. 2007. Gain-Of-Function Raf1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. *Nature Genetics* **39**:1007–1012. DOI: https://doi.org/10.1038/ng2073, PMID: 17603483

Park Y, Sarkar A, Bhutani K, Kellis M. 2017. Multi-Tissue Polygenic Models for Transcriptome-Wide Association Studies. [bioRxiv]. DOI: https://doi.org/10.1101/107623

Patel D, Zhang X, Farrell JJ, Chung J, Stein TD, Lunetta KL, Farrer LA. 2021. Cell-type-specific expression quantitative trait loci associated with alzheimer disease in blood and brain tissue. *Translational Psychiatry* **11**:250. DOI: https://doi.org/10.1038/s41398-021-01373-z, PMID: 33907181

Pedraza JM, Paulsson J. 2008. Effects of molecular memory and bursting on fluctuations in gene expression. Science **319**:339–343. DOI: https://doi.org/10.1126/science.1144331, PMID: 18202292

Perez RK, Gordon MG, Subramaniam M, Kim MC, Hartoularos GC, Targ S, Sun Y, Ogorodnikov A, Bueno R, Lu A, Thompson M, Rappoport N, Dahl A, Lanata CM, Matloubian M, Maliskova L, Kwek SS, Li T, Slyper M, Waldman J, et al. 2022. Single-Cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. Science **376**:eabf1970. DOI: https://doi.org/10.1126/science.abf1970, PMID: 35389781

Piane M, Della Monica M, Piatelli G, Lulli P, Lonardo F, Chessa L, Scarano G. 2009. Majewski osteodysplastic primordial dwarfism type II (MOPD II) syndrome previously diagnosed as seckel syndrome: report of a novel mutation of the PCNT gene. *American Journal of Medical Genetics. Part A* **149A**:2452–2456. DOI: https://doi.org/10.1002/ajmg.a.33035, PMID: 19839044

Pinney SE, Oliver-Krasinski J, Ernst L, Hughes N, Patel P, Stoffers DA, Russo P, De León DD. 2011. Neonatal diabetes and congenital malabsorptive diarrhea attributable to a novel mutation in the human neurogenin-3 gene coding sequence. *The Journal of Clinical Endocrinology and Metabolism* **96**:1960–1965. DOI: https://doi.org/10.1210/jc.2011-0029, PMID: 21490072

Plengvidhya N, Kooptiwut S, Songtawee N, Doi A, Furuta H, Nishi M, Nanjo K, Tantibhedhyangkul W, Boonyasrisawat W, Yenchitsomanus P, Doria A, Banchuin N. 2007. Pax4 mutations in thais with maturity onset diabetes of the young. *The Journal of Clinical Endocrinology and Metabolism* **92**:2821–2826. DOI: https://doi.org/10.1210/jc.2006-1927, PMID: 17426099

Polymeropoulos MH, Ide SE, Wright M, Goodship J, Weissenbach J, Pyeritz RE, Da Silva EO, Ortiz De Luna RI, Francomano CA. 1996. The gene for the Ellis-van Creveld syndrome is located on chromosome 4p16. *Genomics* **35**:1–5. DOI: https://doi.org/10.1006/geno.1996.0315, PMID: 8661097

Poulton CJ, Schot R, Kia SK, Jones M, Verheijen FW, Venselaar H, de Wit M-CY, de Graaff E, Bertoli-Avella AM, Mancini GMS. 2011. Microcephaly with simplified gyration, epilepsy, and infantile diabetes linked to inappropriate apoptosis of neural progenitors. *American Journal of Human Genetics* **89**:265–276. DOI: https://doi.org/10.1016/j.ajhg.2011.07.006, PMID: 21835305

Prescott NJ, Dominy KM, Kubo M, Lewis CM, Fisher SA, Redon R, Huang N, Stranger BE, Blaszczyk K, Hudspith B, Parkes G, Hosono N, Yamazaki K, Onnie CM, Forbes A, Dermitzakis ET, Nakamura Y, Mansfield JC, Sanderson J, Hurles ME, et al. 2010. Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. *Human Molecular Genetics* **19**:1828–1839. DOI: https://doi.org/10.1093/hmg/ddq041, PMID: 20106866

Puig-Hervás MT, Temtamy S, Aglan M, Valencia M, Martínez-Glez V, Ballesta-Martínez MJ, López-González V, Ashour AM, Amr K, Pulido V, Guillén-Navarro E, Lapunzina P, Caparrós-Martín JA, Ruiz-Perez VL. 2012. Mutations in PLOD2 cause autosomal-recessive connective tissue disorders within the Bruck syndrome -- osteogenesis imperfecta phenotypic spectrum. *Human Mutation* **33**:1444–1449. DOI: https://doi.org/10.1002/humu.22133, PMID: 22689593

Pullinger CR, Hennessy LK, Chatterton JE, Liu W, Love JA, Mendel CM, Frost PH, Malloy MJ, Schumaker VN, Kane JP. 1995. Familial ligand-defective apolipoprotein B. identification of a new mutation that decreases LDL receptor binding affinity. *The Journal of Clinical Investigation* **95**:1225–1234. DOI: https://doi.org/10.1172/JCI117772, PMID: 7883971

Qiu H, Zhang B, Zhou T. 2019. Analytical results for a generalized model of bursty gene expression with molecular memory. *Physical Review. E* **100**:012128. DOI: https://doi.org/10.1103/PhysRevE.100.012128, PMID: 31499786

Raeder H, Johansson S, Holm PI, Haldorsen IS, Mas E, Sbarra V, Nermoen I, Eide SÅ, Grevle L, Bjørkhaug L, Sagen JV, Aksnes L, Søvik O, Lombardo D, Molven A, Njølstad PR. 2006. Mutations in the cel VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nature Genetics* **38**:54–62. DOI: https://doi.org/10.1038/ng1708, PMID: 16369531

Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, Zou LS, Didion JP, Guan Y, Shendure J, Parker SCJ, Collins FS. 2020. Single-Cell ATAC-seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Molecular Metabolism* **32**:109–121. DOI: https://doi.org/10.1016/j.molmet.2019.12.006, PMID: 32029221

Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**:216–226. DOI: https://doi.org/10.1016/j.cell.2008.09.050, PMID: 18957198

Raj A, van Oudenaarden A. 2009. Single-Molecule approaches to stochastic gene expression. *Annual Review of Biophysics* **38**:255–270. DOI: https://doi.org/10.1146/annurev.biophys.37.032807.125928, PMID: 19416069

Rauch A, Thiel CT, Schindler D, Wick U, Crow YJ, Ekici AB, van Essen AJ, Goecke TO, Al-Gazali L, Chrzanowska KH, Zweier C, Brunner HG, Becker K, Curry CJ, Dallapiccola B, Devriendt K, Dörfler A, Kinning E, Megarbane A, Meinecke P, et al. 2008. Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. Science **319**:816–819. DOI: https://doi.org/10.1126/science.1151174, PMID: 18174396

Razzaque MA, Nishizawa T, Komoike Y, Yagi H, Furutani M, Amo R, Kamisago M, Momma K, Katayama H, Nakagawa M, Fujiwara Y, Matsushima M, Mizuno K, Tokuyama M, Hirota H, Muneuchi J, Higashinakagawa T, Matsuoka R. 2007. Germline gain-of-function mutations in RAF1 cause Noonan syndrome. *Nature Genetics* **39**:1013–1017. DOI: https://doi.org/10.1038/ng2078, PMID: 17603482

Reis AF, Ye W-Z, Dubois-Laforgue D, Bellanné-Chantelot C, Timsit J, Velho G. 2000. Association of a variant in exon 31 of the sulfonylurea receptor 1 (SUR1) gene with type 2 diabetes mellitus in French Caucasians. *Human Genetics* **107**:138–144. DOI: https://doi.org/10.1007/s004390000345, PMID: 11030411

Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N, Fennell T, Kirby A, Latiano A, Goyette P, Green T, Halfvarson J, Haritunians T, Korn JM, Kuruvilla F, Lagacé C, et al. 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics* **43**:1066–1073. DOI: https://doi.org/10.1038/ng.952, PMID: 21983784

Rivas MA, Graham D, Sulem P, Stevens C, Desch AN, Goyette P, Gudbjartsson D, Jonsdottir I, Thorsteinsdottir U, Degenhardt F, Mucha S, Kurki MI, Li D, D'Amato M, Annese V, Vermeire S, Weersma RK, Halfvarson J, Paavola-Sakki P, Lappalainen M, et al. 2016. A protein-truncating R179X variant in RNF186 confers protection against ulcerative colitis. *Nature Communications* **7**:12342. DOI: https://doi.org/10.1038/ncomms12342, PMID: 27503255

Rossi A, Superti-Furga A. 2001. Mutations in the diastrophic dysplasia sulfate transporter (DTDST) gene (SLC26A2): 22 novel mutations, mutation review, associated skeletal phenotypes, and diagnostic relevance. *Human Mutation* **17**:159–171. DOI: https://doi.org/10.1002/humu.1, PMID: 11241838

Rubio-Cabezas O, Jensen JN, Hodgson MI, Codner E, Ellard S, Serup P, Hattersley AT. 2011. Permanent neonatal diabetes and enteric anendocrinosis associated with biallelic mutations in Neurog3. *Diabetes* **60**:1349–1353. DOI: https://doi.org/10.2337/db10-1008, PMID: 21378176

Ruiz-Perez VL, Tompson SWJ, Blair HJ, Espinoza-Valdez C, Lapunzina P, Silva EO, Hamel B, Gibbs JL, Young ID, Wright MJ, Goodship JA. 2003. Mutations in two nonhomologous genes in a head-to-head configuration cause Ellis-van Creveld syndrome. *American Journal of Human Genetics* **72**:728–732. DOI: https://doi.org/10.1086/368063, PMID: 12571802

Rust S, Rosier M, Funke H, Real J, Amoura Z, Piette JC, Deleuze JF, Brewer HB, Duverger N, Denèfle P, Assmann G. 1999. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nature Genetics* **22**:352–355. DOI: https://doi.org/10.1038/11921, PMID: 10431238

Sansbury FH, Flanagan SE, Houghton JAL, Shuixian Shen FL, Al-Senani AMS, Habeb AM, Abdullah M, Kariminejad A, Ellard S, Hattersley AT. 2012. Slc2A2 mutations can cause neonatal diabetes, suggesting GLUT2 may have a role in human insulin secretion. *Diabetologia* **55**:2381–2385. DOI: https://doi.org/10.1007/s00125-012-2595-0, PMID: 22660720

Sansbury FH, Kirel B, Caswell R, Lango Allen H, Flanagan SE, Hattersley AT, Ellard S, Shaw-Smith CJ. 2015. Biallelic Rfx6 mutations can cause childhood as well as neonatal onset diabetes mellitus. *European Journal of Human Genetics* **23**:1744–1748. DOI: https://doi.org/10.1038/ejhg.2015.161, PMID: 26264437

Savage DB, Agostini M, Barroso I, Gurnell M, Luan J, Meirhaeghe A, Harding A-H, Ihrke G, Rajanayagam O, Soos MA, George S, Berger D, Thomas EL, Bell JD, Meeran K, Ross RJ, Vidal-Puig A, Wareham NJ, O'Rahilly S, Chatterjee VKK, et al. 2002. Digenic inheritance of severe insulin resistance in a human pedigree. *Nature Genetics* **31**:379–384. DOI: https://doi.org/10.1038/ng926, PMID: 12118251

Schaefke B, Sun W, Li YS, Fang L, Chen W. 2018. The evolution of posttranscriptional regulation. *Wiley Interdisciplinary Reviews. RNA* **9**:e1485. DOI: https://doi.org/10.1002/wrna.1485, PMID: 29851258

Schafer S, Adami E, Heinig M, Rodrigues KEC, Kreuchwig F, Silhavy J, van Heesch S, Simaite D, Rajewsky N, Cuppen E, Pravenec M, Vingron M, Cook SA, Hubner N. 2015. Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nature Communications* **6**:7200. DOI: https://doi.org/10.1038/ncomms8200, PMID: 26007203

Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, Ha B, Altay G, Greenbaum JA, McVicker G, Seumois G, Rao A, Kronenberg M, Peters B, Vijayanand P. 2018. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**:1701–1715.. DOI: https://doi.org/10.1016/j.cell.2018.10.022, PMID: 30449622

Schmiedel BJ, Gonzalez-Colin C, Fajardo V, Rocha J, Madrigal A, Ramírez-Suástegui C, Bhattacharyya S, Simon H, Greenbaum JA, Peters B, Seumois G, Ay F, Chandra V, Vijayanand P. 2022. Single-Cell eQTL analysis of activated T cell subsets reveals activation and cell type-dependent effects of disease-risk variants. *Science Immunology* **7**:eabm2508. DOI: https://doi.org/10.1126/sciimmunol.abm2508, PMID: 35213211

Schneider WJ, Kovanen PT, Brown MS, Goldstein JL, Utermann G, Weber W, Havel RJ, Kotite L, Kane JP, Innerarity TL, Mahley RW. 1981. 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. *The Journal of Clinical Investigation* **68**:1075–1085. DOI: https://doi.org/10.1172/jci110330, PMID: 6270194

Schubbert S, Zenker M, Rowe SL, Böll S, Klein C, Bollag G, van der Burgt I, Musante L, Kalscheuer V, Wehner L-E, Nguyen H, West B, Zhang KYJ, Sistermans E, Rauch A, Niemeyer CM, Shannon K, Kratz CP. 2006. Germline KRAS mutations cause Noonan syndrome. *Nature Genetics* **38**:331–336. DOI: https://doi.org/10.1038/ng1748, PMID: 16474405

Sellick GS, Barker KT, Stolte-Dijkstra I, Fleischmann C, Coleman RJ, Garrett C, Gloyn AL, Edghill EL, Hattersley AT, Wellauer PK, Goodwin G, Houlston RS. 2004. Mutations in Ptf1a cause pancreatic and cerebellar agenesis. *Nature Genetics* **36**:1301–1305. DOI: https://doi.org/10.1038/ng1475, PMID: 15543146

Senée V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec J-C, Charon C, Nicolino M, Boileau P, Cavener DR, Bougnères P, Taha D, Julier C. 2006. Mutations in Glis3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nature Genetics* **38**:682–687. DOI: https://doi.org/10.1038/ng1802, PMID: 16715098

Shaheen R, Faqeih E, Ansari S, Abdel-Salam G, Al-Hassnan ZN, Al-Shidi T, Alomar R, Sogaty S, Alkuraya FS. 2014. Genomic analysis of primordial dwarfism reveals novel disease genes. *Genome Research* **24**:291–299. DOI: https://doi.org/10.1101/gr.160572.113, PMID: 24389050

Shahrezaei V, Swain PS. 2008. Analytical distributions for stochastic gene expression. *PNAS* **105**:17256–17261. DOI: https://doi.org/10.1073/pnas.0803850105, PMID: 18988743

Shalev SA, Tenenbaum-Rakover Y, Horovitz Y, Paz VP, Ye H, Carmody D, Highland HM, Boerwinkle E, Hanis CL, Muzny DM, Gibbs RA, Bell GI, Philipson LH, Greeley SAW. 2014. Microcephaly, epilepsy, and neonatal diabetes due to compound heterozygous mutations in IER3IP1: insights into the natural history of a rare disorder. *Pediatric Diabetes* **15**:252–256. DOI: https://doi.org/10.1111/pedi.12086, PMID: 24138066

Shaw-Smith C, Flanagan SE, Patch AM, Grulich-Henn J, Habeb AM, Hussain K, Pomahacova R, Matyka K, Abdullah M, Hattersley AT, Ellard S. 2012. Recessive slc19a2 mutations are a cause of neonatal diabetes mellitus in thiamine-responsive megaloblastic anaemia. *Pediatric Diabetes* **13**:314–321. DOI: https://doi.org/10.1111/j.1399-5448.2012.00855.x, PMID: 22369132

Shaw-Smith C. 2014. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes* **63**:2888–2894. DOI: https://doi.org/10.2337/db14-0061

Shimajiri Y, Sanke T, Furuta H, Hanabusa T, Nakagawa T, Fujitani Y, Kajimoto Y, Takasu N, Nanjo K. 2001. A missense mutation of Pax4 gene (R121W) is associated with type 2 diabetes in Japanese. *Diabetes* **50**:2864–2869. DOI: https://doi.org/10.2337/diabetes.50.12.2864, PMID: 11723072

Shimano H, Yamada N, Katsuki M, Shimada M, Gotoda T, Harada K, Murase T, Fukazawa C, Takaku F, Yazaki Y. 1992a. Overexpression of apolipoprotein E in transgenic mice: marked reduction in plasma lipoproteins except high density lipoprotein and resistance against diet-induced hypercholesterolemia. *PNAS* **89**:1750–1754. DOI: https://doi.org/10.1073/pnas.89.5.1750, PMID: 1542669

Shimano H, Yamada N, Katsuki M, Yamamoto K, Gotoda T, Harada K, Shimada M, Yazaki Y. 1992b. Plasma lipoprotein metabolism in transgenic mice overexpressing apolipoprotein E. accelerated clearance of lipoproteins containing apolipoprotein B. *The Journal of Clinical Investigation* **90**:2084–2091. DOI: https://doi.org/10.1172/JCI116091, PMID: 1430232

Smith SB. 2010. Rfx6 directs islet formation and insulin production in mice and humans. *Nature* **463**:775–780. DOI: https://doi.org/10.1038/nature08748

Soria LF, Ludwig EH, Clarke HR, Vega GL, Grundy SM, McCarthy BJ. 1989. Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. *PNAS* **86**:587–591. DOI: https://doi.org/10.1073/pnas.86.2.587, PMID: 2563166

Spotila LD, Sereda L, Prockop DJ. 1992. Partial isodisomy for maternal chromosome 7 and short stature in an individual with a mutation at the COL1A2 locus. *American Journal of Human Genetics* **51**:1396–1405 PMID: 1463018.

Stoffel M, Duncan SA. 1997. The maturity-onset diabetes of the young (MODY1) transcription factor hnf4alpha regulates expression of genes required for glucose transport and metabolism. *PNAS* **94**:13209–13214. DOI: https://doi.org/10.1073/pnas.94.24.13209, PMID: 9371825

Stoffers DA, Ferrer J, Clarke WL, Habener JF. 1997. Early-Onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nature Genetics* **17**:138–139. DOI: https://doi.org/10.1038/ng1097-138, PMID: 9326926

Støy J, Edghill EL, Flanagan SE, Ye H, Paz VP, Pluzhnikov A, Below JE, Hayes MG, Cox NJ, Lipkind GM, Lipton RB, Greeley SAW, Patch AM, Ellard S, Steiner DF, Hattersley AT, Philipson LH, Bell GI, Neonatal Diabetes International Collaborative Group. 2007. Insulin gene mutations as a cause of permanent neonatal diabetes. *PNAS* **104**:15040–15044. DOI: https://doi.org/10.1073/pnas.0707291104, PMID: 17855560

Stránecký V, Hoischen A, Hartmannová H, Zaki MS, Chaudhary A, Zudaire E, Nosková L, Barešová V, Přistoupilová A, Hodaňová K, Sovová J, Hůlková H, Piherová L, Hehir-Kwa JY, de Silva D, Senanayake MP, Farrag S, Zeman J, Martásek P, Baxová A, et al. 2013. Mutations in ANTXR1 cause GAPO syndrome. *American Journal of Human Genetics* **92**:792–799. DOI: https://doi.org/10.1016/j.ajhg.2013.03.023, PMID: 23602711

Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis ET. 2007. Population genomics of human gene expression. *Nature Genetics* **39**:1217–1224. DOI: https://doi.org/10.1038/ng2142, PMID: 17873874

Strober BJ, Elorbany R, Rhodes K, Krishnan N, Tayeb K, Battle A, Gilad Y. 2019. Dynamic genetic regulation of gene expression during cellular differentiation. Science **364**:1287–1290. DOI: https://doi.org/10.1126/science.aaw0040, PMID: 31249060

Strom TM, Hörtnagel K, Hofmann S, Gekeler F, Scharfe C, Rabl W, Gerbitz KD, Meitinger T. 1998. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Human Molecular Genetics* **7**:2021–2028. DOI: https://doi.org/10.1093/hmg/7.13.2021, PMID: 9817917

Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. 2011. Mammalian genes are transcribed with widely different bursting kinetics. Science **332**:472–474. DOI: https://doi.org/10.1126/science.1198817, PMID: 21415320

Tai ES, Adiconis X, Ordovas JM, Carmena-Ramon R, Real J, Corella D, Ascaso J, Carmena R. 2003. Polymorphisms at the SRBI locus are associated with lipoprotein levels in subjects with heterozygous familial hypercholesterolemia. *Clinical Genetics* **63**:53–58. DOI: https://doi.org/10.1034/j.1399-0004.2003.630108.x, PMID: 12519372

Takenouchi T, Hida M, Sakamoto Y, Torii C, Kosaki R, Takahashi T, Kosaki K. 2013. Severe congenital lipodystrophy and a progeroid appearance: mutation in the penultimate exon of FBN1 causing a recognizable phenotype . *American Journal of Medical Genetics Part A* **161**:3057–3062. DOI: https://doi.org/10.1002/ajmg.a.36157

Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee S-B, Tian X, Browning BL, Das S, Emde A-K, Clarke WE, Loesch DP, Shetty AC, et al. 2021. Sequencing of 53,831 diverse genomes from the NHLBI topmed program. *Nature* **590**:290–299. DOI: https://doi.org/10.1038/s41586-021-03205-y, PMID: 33568819

Tanzanella C, Antoccia A, Spadoni E, di Masi A, Pecile V, Demori E, Varon R, Marseglia GL, Tiepolo L, Maraschio P. 2003. Chromosome instability and nibrin protein variants in Nbs heterozygotes. *European Journal of Human Genetics* **11**:297–303. DOI: https://doi.org/10.1038/sj.ejhg.5200962, PMID: 12708449

Tartaglia M, Mehler EL, Goldberg R, Zampino G, Brunner HG, Kremer H, van der Burgt I, Crosby AH, Ion A, Jeffery S, Kalidas K, Patton MA, Kucherlapati RS, Gelb BD. 2001. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nature Genetics* **29**:465–468. DOI: https://doi.org/10.1038/ng772, PMID: 11704759

Tonkin ET, Wang TJ, Lisgo S, Bamshad MJ, Strachan T. 2004. Nipbl, encoding a homolog of fungal scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nature Genetics* **36**:636–641. DOI: https://doi.org/10.1038/ng1363, PMID: 15146185

Toydemir RM, Brassington AE, Bayrak-Toydemir P, Krakowiak PA, Jorde LB, Whitby FG, Longo N, Viskochil DH, Carey JC, Bamshad MJ. 2006. A novel mutation in FGFR3 causes camptodactyly, tall stature, and hearing loss (CATSHL) syndrome. *American Journal of Human Genetics* **79**:935–941. DOI: https://doi.org/10.1086/508433, PMID: 17033969

Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. 2013. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genetics* **45**:124–130. DOI: https://doi.org/10.1038/ng.2504, PMID: 23263488

Tsai AC-H, Dossett CJ, Walton CS, Cramer AE, Eng PA, Nowakowska BA, Pursley AN, Stankiewicz P, Wiszniewska J, Cheung SW. 2011. Exon deletions of the ep300 and CREBBP genes in two children with Rubinstein-Taybi syndrome detected by aCGH. *European Journal of Human Genetics* **19**:43–49. DOI: https://doi.org/10.1038/ejhg.2010.121, PMID: 20717166

Tufan F, Cefle K, Türkmen S, Türkmen A, Zorba U, Dursun M, Oztürk S, Palandüz S, Ecder T, Mundlos S, Horn D. 2005. Clinical and molecular characterization of two adults with autosomal recessive robinow syndrome. *American Journal of Medical Genetics Part A* **136A**:185–189. DOI: https://doi.org/10.1002/ajmg.a.30785

Umans BD, Battle A, Gilad Y. 2021. Where are the disease-associated eQTLs? *Trends in Genetics* **37**:109–124. DOI: https://doi.org/10.1016/j.tig.2020.08.009, PMID: 32912663

Urbut SM, Wang G, Carbonetto P, Stephens M. 2019. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics* **51**:187–195. DOI: https://doi.org/10.1038/s41588-018-0268-8, PMID: 30478440

van Bokhoven H. 2000. Mutation of the gene encoding the ROR2 tyrosine kinase causes autosomal recessive robinow syndrome. *Nat. Genet* **25**:423–426. DOI: https://doi.org/10.1038/78113

van der Slot AJ, Zuurmond A-M, Bardoel AFJ, Wijmenga C, Pruijs HEH, Sillence DO, Brinckmann J, Abraham DJ, Black CM, Verzijl N, DeGroot J, Hanemaaijer R, TeKoppele JM, Huizinga TWJ, Bank RA. 2003. Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *The Journal of Biological Chemistry* **278**:40967–40972. DOI: https://doi.org/10.1074/jbc.M307380200, PMID: 12881513

Varon R, Vissinga C, Platzer M, Cerosaletti KM, Chrzanowska KH, Saar K, Beckmann G, Seemanová E, Cooper PR, Nowak NJ, Stumm M, Weemaes CM, Gatti RA, Wilson RK, Digweed M, Rosenthal A, Sperling K, Concannon P, Reis A. 1998. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**:467–476. DOI: https://doi.org/10.1016/s0092-8674(00)81174-5, PMID: 9590180

Vaxillaire M, Rouard M, Yamagata K, Oda N, Kaisaki PJ, Boriraj VV, Chevre JC, Boccio V, Cox RD, Lathrop GM, Dussoix P, Philippe J, Timsit J, Charpentier G, Velho G, Bell GI, Froguel P. 1997. Identification of nine novel mutations in the hepatocyte nuclear factor 1 alpha gene associated with maturity-onset diabetes of the young (MODY3). *Human Molecular Genetics* **6**:583–586. DOI: https://doi.org/10.1093/hmg/6.4.583, PMID: 9097962

Viñuelas J, Kaneko G, Coulon A, Vallin E, Morin V, Mejia-Pous C, Kupiec J-J, Beslon G, Gandrillon O. 2013. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biology* **11**:15. DOI: https://doi.org/10.1186/1741-7007-11-15, PMID: 23442824

Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews. Genetics* **13**:227–232. DOI: https://doi.org/10.1038/nrg3185, PMID: 22411467

Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, et al. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics* **42**:579–589. DOI: https://doi.org/10.1038/ng.609, PMID: 20581827

Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, Chen M-H, Raffield LM, Tardaguila M, Huffman JE, Ritchie SC, Megy K, Ponstingl H, Penkett CJ, Albers PK, Wigdor EM, Sakaue S, Moscati A, Manansala R, Lo KS, et al. 2020. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**:1214–1231.. DOI: https://doi.org/10.1016/j.cell.2020.08.008, PMID: 32888494

Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, Björkegren JLM, Im HK, Pasaniuc B, Rivas MA, Kundaje A. 2019. Opportunities and challenges for transcriptome-wide association studies. *Nature Genetics* **51**:592–599. DOI: https://doi.org/10.1038/s41588-019-0385-z, PMID: 30926968

Wallis GA, Starman BJ, Zinn AB, Byers PH. 1990. 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. *American Journal of Human Genetics* **46**:1034–1040 PMID: 2339700.

Wang LL, Gannavarapu A, Kozinetz CA, Levy ML, Lewis RA, Chintagumpala MM, Ruiz-Maldanado R, Contreras-Ruiz J, Cunniff C, Erickson RP, Lev D, Rogers M, Zackai EH, Plon SE. 2003. Association between osteosarcoma and deleterious mutations in the RecQL4 gene in Rothmund-Thomson syndrome. *JNCI Journal of the National Cancer Institute* **95**:669–674. DOI: https://doi.org/10.1093/jnci/95.9.669

Wang ZY, Leushkin E, Liechti A, Ovchinnikova S, Mößinger K, Brüning T, Rummel C, Grützner F, Cardoso-Moreira M, Janich P, Gatfield D, Diagouraga B, de Massy B, Gill ME, Peters A, Anders S, Kaessmann H. 2020a. Transcriptome and translatome co-evolution in mammals. *Nature* **588**:642–647. DOI: https://doi.org/10.1038/s41586-020-2899-z, PMID: 33177713

Wang G, Sarkar A, Carbonetto P, Stephens M. 2020b. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society* **82**:1273–1300. DOI: https://doi.org/10.1111/rssb.12388

Wang Z, Zhang Z, Zhou T. 2020c. Exact distributions for stochastic models of gene expression with arbitrary regulation. *Science China Mathematics* **63**:485–500. DOI: https://doi.org/10.1007/s11425-019-1622-8

Ward MC, Banovich NE, Sarkar A, Stephens M, Gilad Y. 2021. Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. *eLife* **10**:e57345. DOI: https://doi.org/10.7554/eLife.57345, PMID: 33554857

Weiner DJ, Gazal S, Robinson EB, O'Connor LJ. 2022. Partitioning gene-mediated disease heritability without eQTLs. *The American Journal of Human Genetics* **109**:405–416. DOI: https://doi.org/10.1016/j.ajhg.2022.01.010

Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Márquez-Luna C, O'Connor L, Pirinen M, Finucane HK, Price AL. 2020. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics* **52**:1355–1363. DOI: https://doi.org/10.1038/s41588-020-00735-5, PMID: 33199916

Wen X, Pique-Regi R, Luca F. 2017. Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. *PLOS Genetics* **13**:e1006646. DOI: https://doi.org/10.1371/journal.pgen.1006646, PMID: 28278150

Woods KA, Fraser NC, Postel-Vinay MC, Savage MO, Clark AJ. 1996. A homozygous splice site mutation affecting the intracellular domain of the growth hormone (GH) receptor resulting in Laron syndrome with elevated GH-binding protein. *The Journal of Clinical Endocrinology and Metabolism* **81**:1686–1690. DOI: https://doi.org/10.1210/jcem.81.5.8626815, PMID: 8626815

Woods SA, Robinson HB, Kohler LJ, Agamanolis D, Sterbenz G, Khalifa M. 2014. Exome sequencing identifies a novel ep300 frame shift mutation in a patient with features that overlap cornelia de lange syndrome. *American Journal of Medical Genetics Part A* **164**:251–258. DOI: https://doi.org/10.1002/ajmg.a.36237

Wu L, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, Tang H, Snyder M. 2013. Variation and genetic control of protein abundance in humans. *Nature* **499**:79–82. DOI: https://doi.org/10.1038/nature12223, PMID: 23676674

Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, Fajans SS, Signorini S, Stoffel M, Bell GI. 1996a. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**:458–460. DOI: https://doi.org/10.1038/384458a0, PMID: 8945471

Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, Southam L, Cox RD, Lathrop GM, Boriraj VV, Chen X, Cox NJ, Oda Y, Yano H, Le Beau MM, Yamada S, Nishigori H, Takeda J, Fajans SS, Hattersley AT, et al. 1996b. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**:455–458. DOI: https://doi.org/10.1038/384455a0, PMID: 8945470

Yamakawa-Kobayashi K, Yanagi H, Endo K, Arinami T, Hamaguchi H. 2003. Relationship between serum HDL-C levels and common genetic variants of the endothelial lipase gene in Japanese school-aged children. *Human Genetics* **113**:311–315. DOI: https://doi.org/10.1007/s00439-003-0985-6, PMID: 12884003

Yang J., Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**:76–82. DOI: https://doi.org/10.1016/j.ajhg.2010.11.011, PMID: 21167468

Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM, Genetic Investigation of ANthropometric Traits Consortium, DIAbetes Genetics Replication And Meta-analysis Consortium. 2012. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**:369–375. DOI: https://doi.org/10.1038/ng.2213, PMID: 22426310

Yao DW, O'Connor LJ, Price AL, Gusev A. 2020. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics* **52**:626–633. DOI: https://doi.org/10.1038/s41588-020-0625-2, PMID: 32424349

Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, Lu Q, Rowson A, Taylor TRP, Clarke L, Maccora K, Chen C, Cook AL, Ye CJ, Fairfax KA, Hewitt AW, Powell JE. 2022. Single-Cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science **376**:eabf3041. DOI: https://doi.org/10.1126/science.abf3041, PMID: 35389779

Yorifuji T, Kawakita R, Hosokawa Y, Fujimaru R, Yamaguchi E, Tamagawa N. 2012. Dominantly inherited diabetes mellitus caused by GATA6 haploinsufficiency: variable intrafamilial presentation. *Journal of Medical Genetics* **49**:642–643. DOI: https://doi.org/10.1136/jmedgenet-2012-101161, PMID: 22962692

Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellenberg GD. 1996. Positional cloning of the Werner's syndrome gene. Science **272**:258–262. DOI: https://doi.org/10.1126/science.272.5259.258, PMID: 8602509

Yu CE. 1997. Mutations in the consensus helicase domains of the werner syndrome gene. *American Journal of Human Genetics* **12**:1.

Zhang T, Choi J, Kovacs MA, Shi J, Xu M, NISC Comparative Sequencing Program, Melanoma Meta-Analysis Consortium, Goldstein AM, Trower AJ, Bishop DT, Iles MM, Duffy DL, MacGregor S, Amundadottir LT, Law MH, Loftus SK, Pavan WJ, Brown KM. 2018. Cell-Type-Specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Research* **28**:1621–1635. DOI: https://doi.org/10.1101/gr.233304.117, PMID: 30333196

Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, Jiang X, O'Mara TA, Zhao N, Bolla MK, Dunning AM, Dennis J, Wang Q, Ful ZA, Aittomäki K, Andrulis IL, Anton-Culver H, Arndt V, Aronson KJ, Arun BK, et al. 2020. Genome-Wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature Genetics* **52**:572–581. DOI: https://doi.org/10.1038/s41588-020-0609-2, PMID: 32424353

Zhu X, Need AC, Petrovski S, Goldstein DB. 2014. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. *Nature Neuroscience* **17**:773–781. DOI: https://doi.org/10.1038/nn.3713, PMID: 24866043

Appendix 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 (***Blair et al., 2013***). Early GWAS found associations near genes identified through familial studies of severe disorders (***Voight et al., 2010***; ***Chan et al., 2015***), and later implicated some of the same genes in complex and Mendelian forms of cardiovascular (***Kathiresan and Srivastava, 2012***) and neuropsychiatric (***Zhu et al., 2014***) traits. More recent analyses have found that GWAS associations are enriched in regions near causative genes for cognate Mendelian traits in blood traits (***Vuckovic et al., 2020***), lipid traits and diabetes (***Weiner et al., 2022***), as well as a diverse collection of 62 traits (***Freund et al., 2018***). Another recent method used transcriptomic, proteomic, and epigenomic data to prioritize genes and found that, in a selection of nine phenotypes, the selected genes were enriched for Mendelian genes causing similar traits (***Mountjoy et al., 2021***). 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., 2021*** used burden testing on UK Biobank data to identify genes whose coding variation affects complex traits, finding many genes not identified through familial studies.