Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits

Malika Kumar Freund,[1,*] Kathryn S. Burch,[2] Huwenbo Shi,[2] Nicholas Mancuso,[3] Gleb Kichaev,[2] Kristina M. Garske,[1] David Z. Pan,[2] Zong Miao,[1,2] Karen L. Mohlke,[4] Markku Laakso,[5] Päivi Pajukanta,[1,2] Bogdan Pasaniuc,[1,2,3,6,7] and Valerie A. Arboleda[1,2,3,7,*]

Although recent studies provide evidence for a common genetic basis between complex traits and Mendelian disorders, a thorough quantification of their overlap in a phenotype-specific manner remains elusive. Here, we have quantified the overlap of genes identified through large-scale genome-wide association studies (GWASs) for 62 complex traits and diseases with genes containing mutations known to cause 20 broad categories of Mendelian disorders. We identified a significant enrichment of genes linked to phenotypically matched Mendelian disorders in GWAS gene sets; of the total 1,240 comparisons, a higher proportion of phenotypically matched or related pairs (n = 50 of 92 [54%]) than phenotypically unmatched pairs (n = 27 of 1,148 [2%]) demonstrated significant overlap, confirming a phenotype-specific enrichment pattern. Further, we observed elevated GWAS effect sizes near genes linked to phenotypically matched Mendelian disorders. Finally, we report examples of GWAS variants localized at the transcription start site or physically interacting with the promoters of genes linked to phenotypically matched Mendelian disorders. Our results are consistent with the hypothesis that genes that are disrupted in Mendelian disorders are dysregulated by non-coding variants in complex traits and demonstrate how leveraging findings from related Mendelian disorders and functional genomic datasets can prioritize genes that are putatively dysregulated by local and distal non-coding GWAS variants.

Introduction

Genetic architectures of human traits have traditionally been classified into two major categories. Typically, complex traits demonstrate polygenic architectures arising from many low-effect common variants, whereas rare traits tend to have high-effect monogenic determinants.[1] The underlying and practical distinction between these classes has historically been based on the presence of highly penetrant, rare, single-gene disruptive mutations causing recognizable clinical syndromes and monogenic diseases (e.g., cystic fibrosis [MIM: 219700][2]), and the relative absence of such mutations in complex diseases, such as diabetes and schizophrenia.[3] Accumulating evidence suggests that these two classes of phenotypes might not be as biologically distinct as previously thought.[4] Multiple exceptions to the "common disease, common variant" hypothesis[1] have been identified for complex traits[5–7] and their molecular phenotypes,[8–11] and Mendelian disorders have also been found to be affected by multiple or common genetic variants.[12–15] This suggests that there exists a spectrum of genetic architectures rather than a dichotomous classification. Accordingly, the monogenic forms of complex traits (i.e., phenotypically matched Mendelian disorders) are increasingly used as a starting point for identifying genes relevant to complex traits for further study.[16–18] Furthermore, overlap has been identified between genes, common variants, and copy-number variants linked to Mendelian disorders and genetic determinants of complex diseases and traits such as Parkinson disease (MIM: 68600),[19] obesity,[20] height,[21] ototoxicity,[22] and others.[23] However, the overlap between each of these complex traits and Mendelian disorders has been examined individually with different metrics of overlap. In a large study of patient medical records, Blair et al. identified systematic, significant comorbidities between Mendelian disorders and complex diseases and found that association signals from genome-wide association studies (GWASs) for complex diseases were enriched in genomic regions with known roles in comorbid Mendelian disorders, suggesting a shared genetic basis.[24] However, the study focused on Mendelian disorders comorbid with complex diseases in the same individual rather than examining Mendelian disorders demonstrating phenotypes similar to complex traits. Furthermore, advances in sequencing technology have greatly expanded the phenotypic spectrum in known Mendelian syndromes, allowing for deconstruction of syndromic diseases into component medical phenotypes. As such, it is now possible to identify all the component-phenotype consequences of genes linked to Mendelian disorders, allowing for greater resolution in identifying gene-phenotype relationships. However, to the best of our knowledge, no study has taken

[1]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [2]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA; [3]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; [4]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; [5]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland; [6]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[7]These authors contributed equally to this work
*Correspondence: kumarm@ucla.edu (M.K.F.), varboleda@mednet.ucla.edu (V.A.A.)
https://doi.org/10.1016/j.ajhg.2018.08.017.

advantage of this to identify genes linked to any related component-phenotype regardless of the Mendelian disorder's best-known or primary phenotype. Thus, a thorough quantification of the overlap between genes associated with complex traits and genes linked to Mendelian disorders in a phenotype-specific manner remains elusive.

Given that the majority of GWASs for complex traits and diseases have identified significant associations in non-coding genomic regions,[25] we hypothesize that genes individually involved in Mendelian disease belong to the biological pathway(s) shared by both complex and Mendelian disease. Specifically, we hypothesize that large-effect coding variants disrupt individual genes and thus result in severe phenotypes (i.e., Mendelian disorders), whereas non-coding variants produce complex traits by collectively dysregulating expression of these same genes, allowing for nuanced or tissue-specific phenotypes. On the basis of this hypothesis, we expect to identify an enrichment of GWAS signal for a given complex trait near genes linked to Mendelian disorders demonstrating similar phenotypes but no enrichment near genes linked to Mendelian disorders with phenotypes unrelated to the complex trait of interest. To test this hypothesis, we defined "Mendelian disorder genes" as any genes linked to Mendelian disorders in the Online Mendelian Inheritance in Man (OMIM) database. We used the well-curated phenotypic breakdown of Mendelian disorders in OMIM to identify subsets of these genes linked to particular phenotypes (e.g., growth defects or immune dysregulation) expressed as part of any Mendelian disorder. We then examined publicly available GWASs across 62 complex traits (listed in Table 1 and detailed in Table S1) to identify risk genes (here called GWAS gene sets) for each complex trait and quantified the overlap between each GWAS gene set and 20 other sets of Mendelian disorder genes for particular phenotypes (detailed in Tables 1 and S1). We found a consistent, significant, and specific enrichment between GWAS gene sets for complex traits and Mendelian disorder genes for matched and related phenotypes (50 of 1,240 pairs; e.g., rheumatoid arthritis and immune dysregulation), supporting our hypothesis of a shared genetic basis between complex and Mendelian forms of disease. In addition, we observed instances of enrichment between GWAS gene sets for certain complex traits and Mendelian disorder genes for unrelated phenotypes (27 of 1,240 pairs; e.g., systemic lupus erythematosus [SLE (MIM: 152700)] and mature-onset diabetes of the young), suggestive of shared biological mechanisms yet to be examined. Furthermore, we found an increase in average effect size of GWAS variants near Mendelian disorder genes for matched phenotypes, and we identified examples of associated SNPs found directly at the transcription start sites (TSSs) of these phenotypically matched Mendelian disorder genes as candidates for functional follow-up. Finally, we report examples of significant body mass index (BMI)-associated variants directly interacting with phenotypically related Mendelian disorder genes *CREBBP* (MIM: 180849) and *CYP19A1* (MIM: 139300 and 613546) by using human primary white-adipocyte-specific Hi-C data.[66] Leveraging the growing body of well-curated phenotypic data from studies of Mendelian disorders, we provide a phenotype-driven approach to identifying genetic pathways shared by Mendelian diseases and complex traits.

Material and Methods

Gene Coordinates and Symbols

We downloaded gene-body coordinates (GRCh37/hg19, UCSC Genes track) from the UCSC Table Browser[67] (see Web Resources) and specifically selected the gene symbol from the "knownGene" table, transcription start and end sites for each gene from the "knownCanonical" table, and the longest transcript from the "knownGene" table for genes where no entry or multiple entries were listed in the "knownCanonical" table. We used these coordinates for all analyses in our study. Because many genes have been renamed over time, we standardized gene symbols across all analyses in our study by downloading a table of approved symbols, previous symbols, and locus group for each gene from the HUGO Gene Nomenclature Committee at the European Bioinformatics Institute (HGNC) (see Web Resources) and renaming any genes identified by previous symbols with approved gene symbols. We restricted all analyses in our study to genes classified as protein-coding according to the HGNC locus group from chromosomes 1–22. These processing steps resulted in a final single set of coordinates for 17,695 autosomal protein-coding genes (for data access, see Web Resources).

Mendelian Disorder Genes and LoF-Intolerant Genes

To identify Mendelian disorder genes, we downloaded the OMIM catalog and identified all genes linked to Mendelian disorders satisfying the following criteria: (1) disorder is Mendelian and fully penetrant, therefore excluding susceptibility phenotypes, and (2) molecular basis of the Mendelian disorder is known (i.e., phenotype mapping key = 3). We defined loss-of-function (LoF)-intolerant genes as all genes with greater than 90% probability of being LoF intolerant according to the pLI score (>0.9) from the Exome Aggregation Consortium (ExAC) Browser;[68] this score is derived from the number of observed versus expected LoF variants in a given gene across approximately 60,000 healthy exomes. Following the same restriction and gene symbol standardization criteria described above resulted in a final set of 3,446 Mendelian disorder genes and 2,978 LoF-intolerant genes.

Phenotype-Specific Mendelian Disorder Gene Sets

To identify subsets of Mendelian disorder genes linked to particular phenotypes, for each complex trait we curated a set of standardized clinical phenotype terms to describe the full range of relevant Mendelian phenotypes. We used these terms to search OMIM for all Mendelian disorders demonstrating these phenotypes and then extracted the gene(s) linked to each Mendelian disorder. We restricted gene-phenotype associations to those satisfying criteria (1) and (2) described above and with the following additional criteria: (3) gene-phenotype association description does not contain "genome-wide association study" or other GWAS synonyms unless the description also contains any of the terms "missense," "nonsense," "nonsynonymous," or "frameshift" or the gene contains at least one pathogenic or likely pathogenic allele in ClinVar. We include a full list of

Table 1. Complex Traits and Corresponding Mendelian Disorders

Complex Trait	Abbreviation	No. of GWAS Genes	Matched Mendelian Disorder(s)
Celiac disease[26]	CEL	34	immune dysregulation
Crohn disease[27]	CD	239	
Inflammatory bowel disease[27]	IBD	368	
Ulcerative colitis[28]	UC	202	
Primary biliary cirrhosis[29]	PBC	149	
Rheumatoid arthritis (European)[30]	RA	297	
Multiple sclerosis[31]	MS	160	
Autism[32]	AUT	2	monogenic autism
Hemoglobin[33]	HB	89	hematologic disorders
Mean cell hemoglobin[33]	MCH	164	
Mean cell hemoglobin concentration[33]	MCHC	12	
Mean corpuscular volume[33]	MCV	180	
Mean platelet volume[28]	MPV	102	
Red blood cell count[33]	RBC	107	
Systemic lupus erythematosus[34]	SLE	286	
Birthweight[35]	BW	179	growth defects
Height[36]	HGT	2361	
Femoral neck bone mineral density[37]	FN	58	bone and uric acid disorders
Forearm bone mineral density[37]	FA	8	
Lumbar spine bone mineral density[37]	LS	67	
Serum urate concentration[38]	URT	161	
Packed cell volume[33]	PCV	53	disorders of platelet function
Platelet count[28]	PLT	134	
Coronary artery disease[39]	CAD	132	cardiovascular disease
High-density lipoprotein[40]	HDL	464	
Low-density lipoprotein[40]	LDL	370	
Total cholesterol[40]	TC	500	
Triglycerides[40]	TG	354	
Hemoglobin A1C[41]	HBA	33	monogenic diabetes
Type 2 diabetes[42]	T2D	28	
Age-related macular degeneration[43]	AMD	215	monogenic AMD
Age at menarche[44]	MNR	207	female reproductive disorders
Age at menopause[45]	MNP	316	
Fasting glucose[46]	FG	39	insulin disorders
HOMA-B[46]	HMB	12	
HOMA-IR[46]	HMIR	0	
Micro-albuminuria[47]	MA	2	microalbumin disorders
Fasting insulin[46]	FI	23	mature-onset diabetes of the young
2 hr glucose[48]	2HG	2	
Type 1 diabetes[49]	T1D	144	

(*Continued on next page*)

Table 1. *Continued*

Complex Trait	Abbreviation	No. of GWAS Genes	Matched Mendelian Disorder(s)
Alzheimer disease[50]	ALZ	58	neurologic disease
Anxiety disorders (case-control)[51]	ANXC	2	
Anxiety disorders (factor score)[51]	ANXF	3	
Major depressive disorder[52]	MDD	4	
Depressive symptoms[53]	DS	10	
Neuroticism[53]	NRT	82	
Bipolar disorder[54]	BIP	8	psychiatric disease
Schizophrenia[55]	SCZ	479	
Chronic kidney disease[56]	CKD	16	renal disorders
Glomerular filtration rate (creatinine)[56]	EGFR	162	
Urine albumin-to-creatinine ratio[47]	UACR	2	
Resting heart rate[57]	RHR	304	arrhythmias
Age at first birth[58]	AFB	45	education and development disorders
College[59]	COL	12	
Education years[60]	EY	554	
Subjective well-being[53]	SWB	9	positive mood disorders
Body fat percentage[61]	BFP	22	body mass disorders
Body mass index[62]	BMI	231	
Childhood BMI[63]	CBMI	49	
Leptin (adjusted for BMI)[64]	LEPB	5	
Leptin (not adjusted for BMI)[64]	LEP	0	
Waist-to-hip ratio[65]	WHR	74	

This table lists the phenotypically matched pairs of complex traits (n = 62) and groups of Mendelian disorders (n = 20) examined in our study. More details on these traits, including mean GWAS sample size, number of significant GWAS loci reported from original GWAS publications, and number of significant GWAS SNPs, are included in Table S1. GWAS genes for each complex trait were identified by the mapping approach described in the Material and Methods.

phenotype-specific Mendelian disorder gene sets and clinical phenotype terms used in Table S2.

A comparison of all phenotype-specific Mendelian disorder gene sets revealed a high degree of overlap among the gene sets for clinically related Mendelian phenotypes (Figure S1). Accordingly, we clustered gene sets on the basis of pairwise overlap and intersected gene sets clustering together by visual inspection at a hierarchical clustering threshold to create a single gene set for the representative group of Mendelian disorders. Each complex trait was thus matched with the single Mendelian disorder category in which the original specific Mendelian disorder gene set clustered, which ultimately best exemplified the phenotype.

Because of the systemic and pleiotropic nature of complex traits, some complex traits could conceivably be phenotypically related to more than one Mendelian disorder gene set. For example, we generated the Mendelian disorder gene set for SLE by using clinical keywords for both the driving immunological event and the clinical manifestations associated with SLE autoimmunity across a large number of organ systems (kidney, brain, skin, pleura, joints, etc.), such as "anemia." Although the substantial contribution of Mendelian disorder genes related to anemia resulted in SLE pairing with the "hematological disorders" group, the immunological component of SLE is central to the disease. Thus, we identified immune dysregulation as a "relevant phenotype" for SLE and denoted it as such in Figure 2 and Table 2 (the same occurred with other traits).

After combining similar gene sets, we were left with 20 non-disjoint phenotype-specific Mendelian disorder gene sets with an average of 375 genes per set; we include a description of each cluster in Table S3.

Complex-Trait Gene Sets

We downloaded publicly available summary statistics (per-allele SNP effect sizes, or log odds ratios [OR] for case-control traits, with standard errors[69]) for large-scale GWASs of 62 traits[66] (Tables 1 and S1; average n = 83,170, minimum n = 10,610, maximum n = 298,420; some GWASs were imputed with the 1000 Genomes Project as a reference panel by their respective consortia, whereas others were not). For each trait, we identified a gene set by mapping each autosomal genome-wide-significant SNP ($p < 5 \times 10^{-8}$) to the closest up- and downstream protein-coding genes as defined above, resulting in a total of 62 non-disjoint GWAS gene sets. Because GWAS regions often contain multiple genome-wide-significant SNPs and the relevant gene might not lie adjacent to the lead SNP in a region,[70,71] we defined GWAS gene sets by mapping genes with respect to every genome-wide-significant SNP rather than only the index GWAS SNPs at each genomic risk region.

Table 2. Overlap between GWAS Genes and Phenotypically Matched Mendelian Disorder Genes

Complex Trait (No. of Genes)	Matched or Related Mendelian Disorder (No. of Genes)	Shared Genes	OR for Matched Pair (CI)	Raw p Value	Average OR for Unmatched Pairs (CI)
AMD (215)	monogenic AMD (104)	9	7.99 (3.50, 16.11)	4.94×10^{-6}	1.69 (1.21, 2.16)
	immune dysregulation (550)	17	2.73 (1.55, 4.52)	4.13×10^{-4}	
BFP (22)	body mass disorders (128)	3	22.14 (4.14, 76.70)	5.15×10^{-4}	2.63 (0.75, 4.51)
	monogenic diabetes (182)	3	15.42 (2.90, 53.04)	1.43×10^{-3}	
BW (179)	body mass disorders (128)	6	4.94 (1.76, 11.29)	1.93×10^{-3}	1.93 (1.48, 2.37)
	monogenic diabetes (182)	7	4.03 (1.57, 8.66)	2.56×10^{-3}	
CAD (132)	cardiovascular disease (598)	13	3.17 (1.63, 5.67)	5.36×10^{-4}	2.13 (1.65, 2.61)
	insulin disorders (623)	13	3.04 (1.56, 5.43)	7.83×10^{-4}	
CBMI (49)	body mass disorders (128)	5	16.18 (4.92, 41.63)	2.71×10^{-5}	1.55 (1.13, 1.97)
	insulin disorders (623)	7	4.61 (1.74, 10.41)	1.54×10^{-3}	
CD (239)	immune dysregulation (550)	23	3.42 (2.10, 5.32)	1.70×10^{-6}	2.09 (−0.39, 4.58)
EY (554)	positive mood disorders (69)	9	4.70 (2.04, 9.59)	2.88×10^{-4}	0.94 (0.81, 1.07)
	monogenic autism (111)	10	3.10 (1.44, 5.98)	2.54×10^{-3}	
	psychiatric disease (264)	19	2.45 (1.44, 3.95)	8.96×10^{-4}	
FN (58)	bone and uric acid disorders (220)	5	7.64 (2.36, 19.24)	7.61×10^{-4}	3.79 (2.36, 5.22)
HB (89)	disorders of platelet function (443)	12	6.21 (3.05, 11.59)	2.17×10^{-6}	2.38 (1.68, 3.09)
	hematologic disorders (551)	10	3.99 (1.83, 7.79)	4.42×10^{-4}	
HDL (464)	body mass disorders (128)	12	3.92 (1.95, 7.17)	1.39×10^{-4}	1.18 (1.00, 1.37)
	monogenic diabetes (182)	15	3.41 (1.85, 5.85)	9.36×10^{-5}	
	cardiovascular disease (598)	31	2.10 (1.40, 3.06)	3.45×10^{-4}	
HGT (2361)	female reproductive disorders (288)	61	1.76 (1.30, 2.36)	2.18×10^{-4}	1.31 (1.19, 1.42)
	growth defects (723)	126	1.39 (1.13, 1.70)	1.43×10^{-3}	
IBD (368)	immune dysregulation (550)	34	3.32 (2.23, 4.79)	1.58×10^{-8}	1.38 (1.10, 1.66)
LDL (370)	cardiovascular disease (598)	31	2.70 (1.79, 3.95)	3.45×10^{-6}	1.66 (1.24, 2.08)
	mature-onset diabetes of the young (561)	24	2.17 (1.36, 3.32)	8.54×10^{-4}	
LS (67)	bone and uric acid disorders (220)	6	8.00 (2.80, 18.72)	1.83×10^{-4}	2.53 (1.98, 3.08)
MCH (164)	hematologic disorders (551)	15	3.19 (1.73, 5.48)	1.92×10^{-4}	1.64 (1.16, 2.11)
MCV (180)	hematologic disorders (551)	20	4.00 (2.36, 6.44)	8.90×10^{-7}	1.61 (1.18, 2.04)
MNR (207)	body mass disorders (128)	7	5.02 (1.95, 10.86)	7.76×10^{-4}	1.03 (0.78, 1.27)
PBC (149)	immune dysregulation (550)	13	3.03 (1.56, 5.40)	7.77×10^{-4}	1.07 (0.76, 1.38)
PCV (53)	disorders of platelet function (443)	10	9.24 (4.11, 18.83)	6.50×10^{-7}	4.04 (2.22, 5.87)
	arrhythmias (275)	5	6.70 (2.07, 16.94)	1.36×10^{-3}	
	hematologic disorders (551)	8	5.60 (2.27, 12.08)	2.17×10^{-4}	
	cardiovascular disease (598)	7	4.39 (1.66, 9.84)	1.94×10^{-3}	
PLT (134)	disorders of platelet function (443)	12	3.91 (1.95, 7.15)	1.40×10^{-4}	1.42 (0.98, 1.87)
RA (297)	immune dysregulation (550)	25	2.95 (1.86, 4.50)	6.80×10^{-6}	0.83 (0.65, 1.01)
RBC (107)	hematologic disorders (551)	14	4.78 (2.50, 8.50)	5.82×10^{-6}	2.76 (2.20, 3.33)
	cardiovascular disease (598)	13	4.02 (2.05, 7.26)	6.49×10^{-5}	
RHR (304)	arrhythmias (275)	17	3.93 (2.23, 6.53)	5.87×10^{-6}	1.29 (0.95, 1.64)
	cardiovascular disease (598)	26	2.75 (1.75, 4.16)	1.47×10^{-5}	

(*Continued on next page*)

Table 2. *Continued*

Complex Trait (No. of Genes)	Matched or Related Mendelian Disorder (No. of Genes)	Shared Genes	OR for Matched Pair (CI)	Raw p Value	Average OR for Unmatched Pairs (CI)
SCZ (479)	positive mood disorders (69)	9	5.47 (2.37, 11.19)	9.71×10^{-5}	1.11 (0.94, 1.28)
SLE (286)	immune dysregulation (550)	24	2.94 (1.83, 4.52)	1.09×10^{-5}	1.41 (1.10, 1.72)
T2D (28)	body mass disorders (128)	4	23.55 (5.85, 69.99)	4.67×10^{-5}	2.43 (1.57, 3.30)
	monogenic diabetes (182)	3	11.72 (2.24, 38.93)	2.90×10^{-3}	
TC (500)	cardiovascular disease (598)	38	2.44 (1.69, 3.45)	3.46×10^{-6}	1.40 (1.10, 1.70)
TG (354)	body mass disorders (128)	9	3.77 (1.67, 7.49)	1.10×10^{-3}	1.26 (1.01, 1.51)
	monogenic diabetes (182)	12	3.54 (1.78, 6.43)	3.10×10^{-4}	
	cardiovascular disease (598)	25	2.22 (1.41, 3.38)	5.06×10^{-4}	
UC (202)	immune dysregulation (550)	21	3.72 (2.23, 5.92)	1.39×10^{-6}	1.45 (1.12, 1.78)
WHR (74)	insulin disorders (623)	9	3.83 (1.67, 7.78)	1.13×10^{-3}	2.37 (1.71, 3.04)

For each pair of complex trait and Mendelian disorder, Fisher's exact test was used to quantify the enrichment of shared genes with an OR and p value (see Material and Methods). Pairs with significant enrichment passed the cutoff of FDR < 5% at $p < 0.00310$. This table lists pairs of complex traits and phenotypically matched or related Mendelian disorders with significant overlap. For comparison, the average OR and 95% CI for pairings of each complex trait with all unrelated Mendelian disorder gene sets are included.

Quantifying Overlap between Complex Traits and Mendelian Disorders

For each pair of complex trait and Mendelian disorder, we compared the GWAS gene set and phenotype-specific Mendelian disorder gene set with a 2 × 2 contingency table (indicating the counts of gene membership in the GWAS gene set only, in the Mendelian disorder gene set only, in both, or in neither) whereby the set of autosomal protein-coding genes (n = 17,695) represented the total sample. We used Fisher's exact test[72] to determine significance. We assessed phenotype specificity of overlap significance by comparing the GWAS gene sets for each complex trait (n = 62) with all phenotype-specific Mendelian disorder gene sets (n = 20), a total of 1,240 pairs. Significance was assessed at a false-discovery rate (FDR) threshold < 5% ($p < 0.00310$).

To assess the robustness and stability of our SNP-gene mapping approach for complex traits, we performed an overlap quantification between phenotype-specific Mendelian disorder genes and GWAS gene sets derived from three additional SNP-gene mapping methods: mapping each genome-wide-significant SNP to all genes within a 50 Mb window, mapping each genome-wide-significant SNP to all genes within a 500 Mb window, and mapping all SNPs in the credible set to the closest two genes. Comparison of the ORs produced by Fisher's exact test demonstrated no major difference in outcomes from different mapping methods (Table S4); thus, we found that even more conservative gene sets, such as the GWAS gene sets derived from the credible set for each complex trait, still demonstrate the pattern of trait-specific enrichment.

Estimating Enrichment of GWAS SNP Association Signal

We created genomic annotations to capture the regions spanning 50 kb upstream through 50 kb downstream of gene bodies for four categories of genes: all protein-coding genes (n = 17,695), all Mendelian disorder genes (n = 3,446), all LoF-intolerant genes (n = 2,978), and the phenotype-specific Mendelian disorder gene sets (average n = 609). For each pair of complex trait and gene category, we computed enrichment of GWAS signal within category c with respect to the set of all protein-coding genes according to Equation 1:

$$a_c = \frac{\frac{1}{N_c}\sum_{j=1}^{N_c}\sum_{i=1}^{M_j}\frac{Z_i^2}{M_j}}{\frac{1}{N_p}\sum_{j=1}^{N_p}\sum_{i=1}^{M_j}\frac{Z_i^2}{M_j}} \qquad \text{Equation 1}$$

Here, N_c is the number of genes in category c, M_j is the number of SNPs within 50 kb of gene j, Z_i is the GWAS effect size of SNP i divided by the standard error, and N_p is the total number of protein-coding genes. Thus, a_c is the enrichment of the average SNP effect size (Z^2) per gene in category c (compared with the average Z^2 for any protein-coding gene). The percent increase in average SNP effect size per gene for category c, or $(a_c - 1) \times 100$, is shown in Figure 3. We performed similar comparisons for median SNP effect size per gene for category c and maximum SNP effect size per gene for category c (Table S5).

To ensure that this signal was not driven by linkage disequilibrium (LD), minor allele frequency (MAF), or average gene length per category, we compared these three properties across the gene categories for each complex trait. We calculated LD scores[73] reflecting the amount of LD tagged by each SNP in the HapMap 3 reference panel; then, for each gene category, we averaged the LD scores of SNPs falling within 50 kb of each gene. We performed similar analyses to examine average MAF per gene and average gene length per category across each complex trait (Table S6). For comparison, we additionally performed a permutation test by drawing 100 sets of random genes for each Mendelian disorder gene set (matched for number and length of genes) and computing the average effect size per gene for each phenotypically matched complex trait across all 100 random sets.

Putative Causal Mechanisms at GWAS Risk Regions

We performed statistical fine-mapping of the genome-wide-significant regions ($p < 5 \times 10^{-8}$) for each GWAS by using fgwas[74] with no functional annotations and default parameter settings. For each GWAS, we constructed a 95% credible set (defined as the minimum set of SNPs where 95% of the probability of causation at a region is accumulated) for each region of 500 SNPs containing a significant GWAS association. We achieved this by adding SNPs one at a time with a decreasing posterior probability of causation

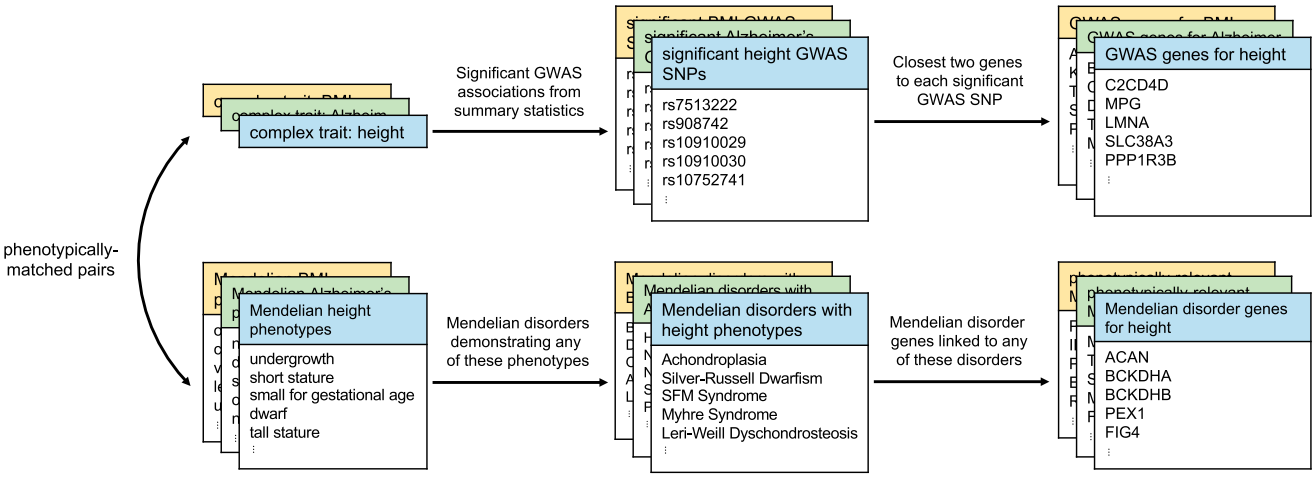

Figure 1. GWAS Gene Sets and Phenotype-Specific Mendelian Disorder Gene Sets
For each complex trait (e.g., height), we first identified matched Mendelian phenotypes (e.g., undergrowth and short stature; Table S2). Using publicly available GWAS data, we defined the "GWAS genes" for a given complex trait to be the closest upstream and closest downstream protein-coding genes for every genome-wide-significant variant in the GWAS. We selected phenotype-matched Mendelian disorder genes by first identifying Mendelian disorders expressing any of the matched Mendelian phenotypes and then identifying all genes linked to any of those disorders.

(posterior probability of association for the SNP, conditioned on there being an association in the region) until a cumulative 95% probability of causation was reached.

Identification of Candidate Regulatory Variants

We intersected credible sets for each complex trait with genomic regions 1 kb upstream of each phenotypically relevant Mendelian disorder gene to identify SNPs localizing at the TSS. To identify genes whose expression the GWAS SNPs might regulate, we queried the UCSC GTEx Combined eQTL Track (version 2017-10-25) and filtered by SNP rsID. This table describes all gene-tissue pairs where a SNP has evidence of regulatory function. We restricted results to phenotype-matched Mendelian genes whose promoter contained a genome-wide-significant SNP in our GWAS fine-mapped results. To identify candidate regulatory variants interacting with promoters of phenotype-matched Mendelian disorder genes, we used interactions from promoter capture Hi-C in human primary white adipocytes[66] for each complex trait and filtered interactions to pairs of interacting regions where at least one region contained a promoter of a phenotype-specific Mendelian disorder gene. We then intersected interaction pairs for each of these regions with credible sets for each complex trait to identify credible SNPs interacting with regions containing promoters of phenotype-specific Mendelian disorder genes.

Estimating the Enrichment of SNP Heritability of Complex Traits within Mendelian Disorder Gene-Set Annotations

We used stratified LD score regression (s-LDSC)[75] to estimate the enrichment of SNP heritability of 47 complex traits and diseases within each of 20 Mendelian disorder gene-set annotations corresponding to the regions spanning 50 kb upstream through 50 kb downstream of gene bodies for each Mendelian disorder gene set. Of the 62 total GWAS traits analyzed in this study, the 47 complex traits and diseases are a subset that meet the criteria for running s-LDSC (i.e., the GWAS did not use custom genotyping arrays). The annotation value for SNP i and gene set k is defined as $a_{ik} = 1$ if SNP i is within 50 kb upstream or 50 kb downstream of any of the gene bodies in gene set k and $a_{ik} = 0$ otherwise. For each of the 20 annotations, we computed LD scores within 1 cM blocks with default parameters and LD estimated from the European individuals in the 1000 Genomes Phase 3 reference panel. For each GWAS-annotation pair, we ran s-LDSC by using the recommended "baseline model"[75] as covariates in the regression for a total of 53 annotations per run (52 "baseline" annotations plus the gene-set annotation of interest).

Results

GWAS Risk Genes Show Specific, Significant Overlap with Phenotypically Matched Mendelian Disorder Genes

We first sought to examine the degree of overlap between phenotype-matched Mendelian disorder genes with risk genes for complex traits as identified through GWASs. For each complex trait, we identified corresponding Mendelian forms, often as familial forms or rare phenotypic extremes, and curated Mendelian disorder gene sets composed of Mendelian disorder genes linked to those specific phenotypes from OMIM (see Material and Methods and Figure 1). We combined similar Mendelian disorder gene sets to create one gene set for the representative Mendelian disorder(s) (for a total of 20 Mendelian disorder gene sets). We separately ascertained GWAS gene sets for each complex trait by identifying the closest up- and downstream genes to each GWAS SNP meeting genome-wide significance (see Material and Methods and Figure 1). Overlap between each phenotype-specific Mendelian disorder gene set (n = 20) and each GWAS gene set (n = 62) was assessed by Fisher's exact test for a total of 1,240 comparisons (Tables 1 and S7). We hypothesized that GWAS gene sets would have a specific

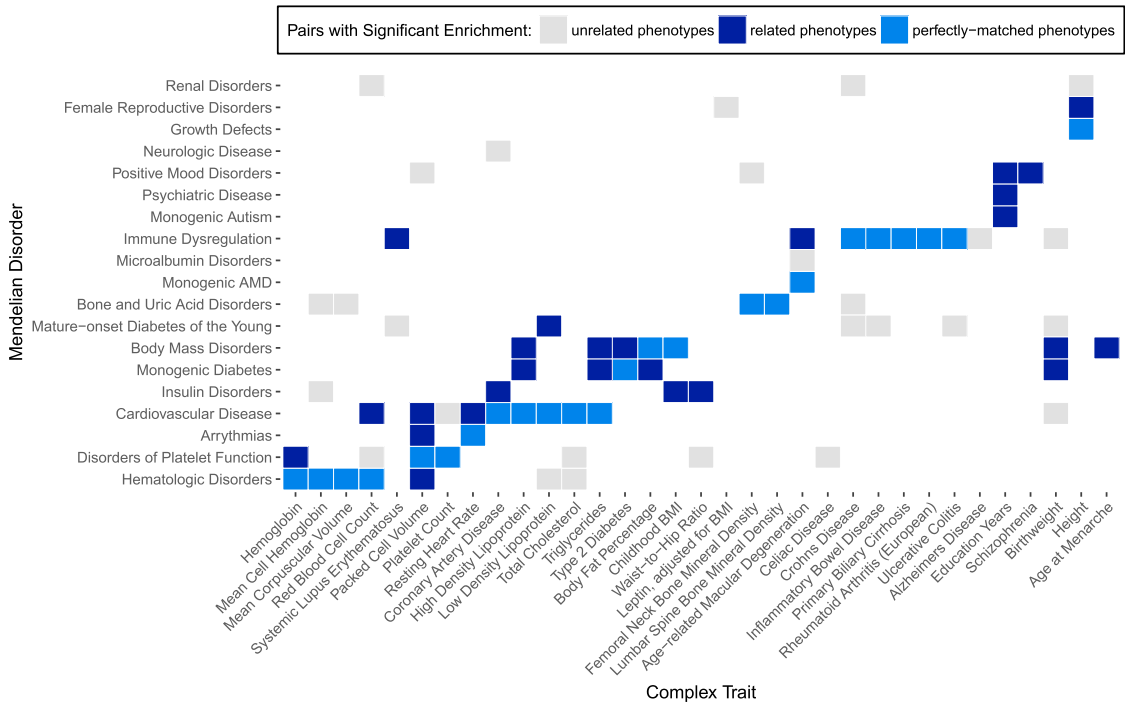

Figure 2. Overlap between GWAS Genes and Mendelian Disorder Genes Demonstrates Trait Specificity
Significant overlaps from phenotypically matched pairs of complex traits and Mendelian disorders (blue) and pairs with unrelated phenotypes (gray) are shown. Phenotypically matched pairs are subdivided into pairs with perfectly matched phenotypes (light blue) and pairs with related phenotypes (dark blue). Complex traits and Mendelian disorders with no significant overlaps are excluded here; results from all traits are presented in Figure S2. We assessed significance by controlling for FDR < 5% at $p < 0.00310$.

significant enrichment of Mendelian disorder genes for perfectly matched Mendelian disorders (as identified in Table 1; 62 of the 1,240 comparisons) or related Mendelian disorders (an additional 30 of the 1,240 comparisons; 92 of 1,240 total) but no enrichment for unrelated Mendelian disorders (the remaining 1,148 of 1,240 comparisons). Among all 1,240 pairs of complex and Mendelian disorder gene sets assessed, we identified 77 pairs with significant overlap crossing an FDR < 5% cutoff at $p < 0.00310$ (Figure 2). An examination of the log ORs for each overlap comparison revealed more extreme enrichments among phenotypically matched pairs than among phenotypically unmatched pairs (Table 2), which is consistent with our hypothesis. 50 out of the 77 significantly overlapping pairs showed perfectly matching phenotypes (as defined in Table 1; see Material and Methods) or reflected known shared biology (identified in dark blue within Figure 2). Specifically, in many of these pairs, monogenic forms of the complex trait have been well established in the genetics literature; examples include age-related macular degeneration and cholesterol traits (high-density lipoprotein [HDL], low-density lipoprotein [LDL], total cholesterol [TC], and triglycerides [TG]).[7,76–79] We confirmed significant enrichment between many of these previously reported pairs; for example, we found significant enrichment between the complex and monogenic forms of height[80] (OR = 1.39, $p = 1.43 \times 10^{-3}$) and between HDL and Mendelian forms of cardiovascular disease[81] (OR = 2.10, $p = 3.45 \times 10^{-4}$). We also identified previously unreported enrichments; for example, we found a strong enrichment between inflammatory bowel disease and Mendelian forms of immune dysregulation (OR = 3.32, $p = 1.58 \times 10^{-8}$) and between hemoglobin and Mendelian hematologic disorders (OR = 3.99, $p = 4.42 \times 10^{-4}$). The remaining 27 pairs with significant overlap suggested shared biological mechanisms yet to be established between complex traits and Mendelian disorders (Table 3). For example, we observed an enrichment between height and renal disorders (OR = 1.48, $p = 3.75 \times 10^{-5}$) and enrichment between Crohn disease and mature-onset diabetes of the young (OR = 2.69, $p = 2.32 \times 10^{-4}$). Strikingly, a higher proportion of phenotypically matched or related pairs (n = 50 of 92 [54%]) than phenotypically unmatched pairs (n = 27 of 1,148 [2%]) demonstrated significant overlap, consistent with our hypothesis of a phenotype-specific enrichment pattern (Table S7 and Figure S2).

To investigate whether intra-chromosomal proximal clustering of genes with similar functionality was confounding our results, we pruned our dataset of all protein-coding genes to include only one gene per 33.2 kb window across the genome (determined by the average distance to the next closest gene in our dataset) and re-computed overlap ORs. The results were highly similar to those of our original approach (Pearson r = 0.96). Moreover, we found the average distance to the next closest gene among the sets of genes shared by a phenotypically matched pair of complex trait and Mendelian disorder to be 21.1 Mb (Table S8).

Table 3. Instances of Significant Overlap between GWAS Genes and Unrelated Mendelian Disorder Genes

Complex Trait (No. of Genes)	Matched or Related Mendelian Disorder (No. of Genes)	Shared Genes	OR (CI)	Raw p Value	Average OR for Remaining Unrelated Pairs (CI)
ALZ (58)	immune dysregulation (550)	7	4.32 (1.65, 9.62)	2.06×10^{-3}	1.95 (1.36, 2.53)
AMD (215)	microalbumin disorders (159)	8	4.43 (1.86, 9.13)	7.37×10^{-4}	1.59 (1.03, 2.15)
BW (179)	mature-onset diabetes of the young (561)	16	3.06 (1.69, 5.16)	1.91×10^{-4}	2.35 (1.94, 2.75)
	immune dysregulation (550)	14	2.69 (1.43, 4.68)	1.44×10^{-3}	
	cardiovascular disease (598)	14	2.46 (1.31, 4.28)	3.10×10^{-3}	
CAD (132)	neurologic disease (222)	7	4.52 (1.76, 9.75)	1.39×10^{-3}	2.34 (1.83, 2.86)
CD (239)	bone and uric acid disorders (220)	9	3.20 (1.42, 6.29)	3.10×10^{-3}	2.09 (−0.39, 4.58)
	mature-onset diabetes of the young (561)	19	2.69 (1.58, 4.35)	2.32×10^{-4}	
	renal disorders (838)	23	2.17 (1.34, 3.37)	1.13×10^{-3}	
CEL (34)	disorders of platelet function (443)	5	6.78 (2.04, 17.83)	1.47×10^{-3}	1.90 (1.28, 2.52)
FN (58)	positive mood disorders (69)	3	14.51 (2.83, 46.53)	1.50×10^{-3}	3.79 (2.36, 5.22)
HGT (2361)	renal disorders (838)	153	1.48 (1.23, 1.78)	3.75×10^{-5}	1.31 (1.19, 1.42)
IBD (368)	mature-onset diabetes of the young (561)	30	2.81 (1.85, 4.13)	2.40×10^{-6}	1.38 (1.10, 1.66)
LDL (370)	hematologic disorders (551)	25	2.31 (1.46, 3.51)	3.50×10^{-4}	1.66 (1.24, 2.08)
LEPB (5)	female reproductive disorders (288)	2	40.52 (3.37, 353.67)	2.56×10^{-3}	10.83 (3.75, 17.91)
MCH (164)	bone and uric acid disorders (220)	8	4.19 (1.75, 8.61)	1.05×10^{-3}	1.65 (1.15, 2.16)
	insulin disorders (623)	15	2.80 (1.52, 4.81)	6.99×10^{-4}	
MCV (180)	bone and uric acid disorders (220)	8	3.80 (1.59, 7.79)	1.89×10^{-3}	1.74 (1.25, 2.22)
PCV (53)	positive mood disorders (69)	3	15.97 (3.11, 51.37)	1.16×10^{-3}	4.04 (2.22, 5.87)
PLT (134)	cardiovascular disease (598)	12	2.85 (1.42, 5.20)	1.97×10^{-3}	1.42 (0.98, 1.87)
RBC (107)	disorders of platelet function (443)	12	5.03 (2.49, 9.29)	1.51×10^{-5}	2.90 (2.13, 3.67)
	renal disorders (838)	18	4.14 (2.33, 6.96)	2.60×10^{-6}	
SLE (286)	mature-onset diabetes of the young (561)	22	2.61 (1.59, 4.07)	1.24×10^{-4}	1.59 (1.24, 1.94)
TC (500)	hematologic disorders (551)	32	2.20 (1.47, 3.18)	1.18×10^{-4}	1.35 (1.05, 1.66)
	disorders of platelet function (443)	25	2.11 (1.34, 3.20)	1.13×10^{-3}	
UC (202)	mature-onset diabetes of the young (561)	19	3.25 (1.90, 5.27)	2.44×10^{-5}	1.41 (1.05, 1.77)
WHR (74)	disorders of platelet function (443)	7	4.12 (1.59, 9.03)	2.50×10^{-3}	2.38 (1.58, 3.18)

As in Table 2, Fisher's exact test was used to quantify the enrichment of shared genes between complex traits and Mendelian disorders with an OR and p value (see Material and Methods). Pairs with significant enrichment passed the cutoff of FDR < 5% at $p < 0.00310$. This table lists pairs of complex traits and phenotypically unrelated Mendelian disorders that demonstrated significant overlap. For comparison, the average OR and 95% CI for pairings of each complex trait with all remaining unrelated Mendelian disorder gene sets are included.

SNPs near Phenotypically Matched Mendelian Disorder Genes Show Increased Effect Size on Complex Traits

Because Mendelian disorder genes are linked with severe biological effects when either one or both alleles are disrupted, dysregulation of the gene through changes in expression or other mechanisms might have a more significant effect than dysregulation of another protein-coding gene not linked to any Mendelian disorders. We hypothesized that SNPs near these phenotype-specific Mendelian disorder genes have further increased effects on complex traits as a result of the increased biological relevance of these gene categories. From the publicly available GWAS summary statistics for each complex trait, we computed the average GWAS effect sizes of SNPs falling within each protein-coding gene and compared the average effect sizes per gene across all Mendelian disorder genes and across phenotypically relevant Mendelian disorder genes (see Material and Methods). Across complex traits, we found an increased average effect size per gene for all Mendelian disorder genes and a further increased average effect size per gene for phenotypically relevant Mendelian disorder genes (Figure 3 and Table S5). This suggests that the genomic regions containing the most biologically relevant genes for each trait contribute most significantly to the biology of complex traits. We also confirmed that LoF-intolerant genes (as defined by ExAC's pLI score > 0.9; see Material

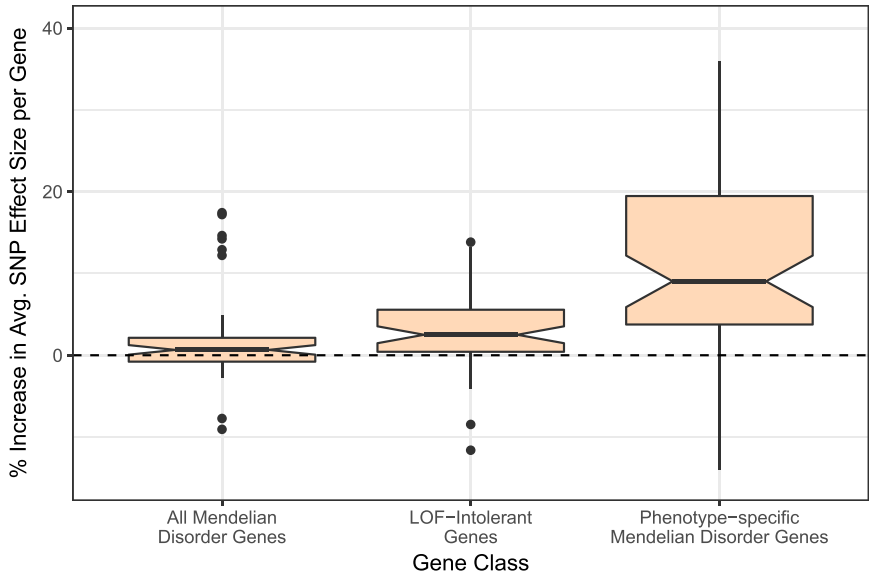

Figure 3. Effect Sizes for SNPs on Complex Traits from GWASs Are Higher for LoF-Intolerant Genes and for Phenotypically Relevant Mendelian Disorder Genes
The increase in average SNP effect size per gene across gene categories, as compared with all protein-coding genes (dashed line), is shown. We averaged effect size (Z^2) across all SNPs falling within 50 kb of a gene to obtain an average SNP effect size per gene and averaged across all genes in each category (all protein-coding genes, all Mendelian disorder genes, all LoF-intolerant genes, and all phenotypically relevant Mendelian disorder genes for each trait). We normalized these averages to the average SNP effect per gene for any protein-coding genes. The boxplots represent the distribution of increases in average effect size per gene across all traits, and notches designate the confidence intervals (CIs). From left to right, CIs read (0.07, 1.24), (1.47, 3.54), and (5.88, 12.19).

and Methods) demonstrate a higher average effect size across most complex traits examined.[68] Given the extreme intolerance of deleterious mutations in these genes, it is possible that LoF-intolerant genes are linked with embryonically lethal mutant phenotypes and are thus undiscovered as Mendelian disorder genes at this time.

We found no significant difference in LD or average MAF between the SNPs within each category and the SNPs within all protein-coding genes (Table S6), suggesting that the observed signal is not driven by any of these potential confounders. In particular, we found the average LD tagged for all protein-coding genes to be 24.38 (95% confidence interval [CI] = (24.35, 24.42)); for none of the other three gene classes did the CIs fall above the upper bound, including the average across all Mendelian disorder gene sets (24.10; 95% CI = (23.47, 24.72)). Similarly, we found the average MAF for all protein-coding genes to be 0.238 (95% CI = (0.237, 0.239)); for none of the other gene classes did the CIs fall below the lower bound, including the average across all Mendelian disorder gene sets (0.238; 95% CI = (0.237, 0.240)). Details for each gene set are included in Table S6. Of note, we did observe an increase in average gene length across gene categories, particularly between all protein-coding genes (159.94 kb; 95% CI = (158.16 kb, 161.71 kb)) and the average across all phenotype-specific Mendelian disorder gene sets (177.87 kb; 95% CI = (168.49 kb, 187.24 kb)) (Table S6). To ensure that our findings of enriched GWAS signal in these gene categories were not due to longer genes' being more likely to tag causal variation, we performed a permutation test comparing the average effect size per gene for phenotype-matched Mendelian disorder genes with the same metric across 100 sets of random genes (matched for number of genes and gene length) (Table S6). We found no evidence that gene length confounded our results given that across 58 of 62 complex traits, the average effect size per gene was higher for phenotype-matched Mendelian disorder genes than for random genes of the same length. Also of note, we did not find evidence of a pervasive phenotype-specific enrichment of SNP heritability within 50 kb of Mendelian disorder genes (Table S9; see Material and Methods). Thus, we can conclude that the average GWAS effect size per gene for phenotypically relevant Mendelian disorder genes is higher than that of all protein-coding genes, all Mendelian disorder genes, and LoF-intolerant genes.

Examples of Credible SNPs for GWAS Regions near Phenotypically Matched Mendelian Disorder Genes

We next sought to identify common non-coding variants that might causally affect complex-trait phenotypes by dysregulating phenotypically relevant Mendelian disorder genes. For each complex trait, we performed statistical fine-mapping of significant GWAS regions to construct 95% credible sets for each region (see Material and Methods) and identified SNPs from the credible set located at the TSS of a gene from the phenotypically relevant Mendelian disorder gene set. We found a total of 786 credible-set SNPs (out of approximately 3.5 million) localizing at the TSS of a phenotypically relevant Mendelian disorder gene (an average of 20 SNPs per trait for 38 traits where at least one such SNP was found; Tables S10 and S11). We identified 25 promising candidate SNPs (attaining genome-wide significance in a GWAS) at TSSs that could be regulating the proximal Mendelian disorder gene (Table 4). We further examined the GTEx Portal to determine whether any of these SNPs were also expression quantitative trail loci (eQTLs) for the corresponding gene; we found 12 variants to be significant eQTLs for the corresponding gene in at least one tissue (Table 4). We highlight two examples: first, we found a significantly associated SNP from the credible set for coronary artery disease (rs1332327, $Z = 6.80$) at the

Table 4. Genome-wide-Significant SNPs Localizing at the TSS of Phenotypically Relevant Mendelian Disorder Genes

Complex Trait	SNP ID	Chromosomal Position	*Z* Score	Gene (MIM)	Maximum eQTL Effect	Maximum −log10 p Value	Tissues
PBC	rs13239597	chr7: 128,695,982	9.85	*TNPO3* (610032)	–	–	–
HGT	rs8028537	chr15: 89,345,946	−9.33	*ACAN* (155760)	–	–	–
HGT	rs10853751	chr19: 41,903,219	8.71	*BCKDHA* (608348)	–	–	–
CD	rs59283234	chr5: 150,225,586	−8.45	*IRGM* (608212)	−0.042	6.733	whole blood
CD	rs751627	chr5: 150,225,112	−8.45	*IRGM* (608212)	−0.042	6.733	whole blood
CD	rs35707106	chr5: 150,225,376	−8.33	*IRGM* (608212)	−0.041	6.412	whole blood
HGT	rs2298307	chr6: 80,816,295	8.28	*BCKDHB* (248611)	–	–	–
HGT	rs12386601	chr7: 92,157,885	8.20	*PEX1* (602136)	–	–	–
BMI	rs17066842	chr18: 58,040,623	−7.54	*MC4R* (155541)	–	–	–
HGT	rs12192268	chr6: 110,011,457	−7.00	*FIG4* (609390)	–	–	–
CAD	rs1332327	chr10: 91,011,680	6.80	*LIPA* (613497)	12.541	5.132	subcutaneous adipose, visceral adipose, adrenal gland, transverse colon, lung, spleen, thyroid, whole blood
RA	rs13239597	chr7: 128,695,982	6.66	*TNPO3* (610032)	–	–	–
IBD	rs59283234	chr5: 150,225,586	6.51	*IRGM* (608212)	−0.042	6.733	whole blood
IBD	rs751627	chr5: 150,225,112	6.51	*IRGM* (608212)	−0.042	6.733	whole blood
HGT	rs7592246	chr2: 219,926,220	6.45	*IHH* (600726)	0.056	6.199	brain cerebellum
IBD	rs34005003	chr5: 150,225,198	6.43	*IRGM* (608212)	−0.043	6.637	whole blood
IBD	rs35707106	chr5: 150,225,376	6.33	*IRGM* (608212)	−0.041	6.412	whole blood
MNR	rs3775971	chr4: 104,641,919	6.20	*TACR3* (162332)	−0.019	8.588	lung
IBD	rs27741	chr16: 28,504,180	6.11	*CLN3* (607042)	–	–	–
HGT	rs4244808	chr11: 2,163,109	6.06	*IGF2* (147470)	–	–	–
RBC	rs1010222	chr19: 13,048,607	−5.96	*CALR* (109091)	17.142	6.851	lung
CD	rs27741	chr16: 28,504,180	−5.87	*CLN3* (607042)	–	–	–
HGT	rs613924	chr11: 65,769,294	−5.86	*BANF1* (603811)	–	–	–
AFB	rs4845357	chr1: 153,896,211	−5.78	*GATAD2B* (614998)	−0.108	4.443	skin not exposed
HGT	rs6591226	chr11: 66,675,989	5.52	*PC* (608786)	–	–	–

GWAS SNPs from the credible set for each complex trait were intersected with TSS regions 1 kb upstream of phenotypically matched Mendelian disorder genes. This table lists all genome-wide-significant SNPs ($p < 5 \times 10^{-8}$ from GWASs) with chromosomal location from all complex traits localizing at the TSS of a phenotypically matched Mendelian disorder gene.

promoter of *LIPA* (MIM: 278000), a Mendelian disorder gene linked to Wolman disease and cholesteryl ester storage disease (both lysosomal acid lipase deficiencies [MIM: 278000]), which involve hypercholesterolemia and hypertriglyceridemia as part of cholesteryl ester- and triglyceride-filled macrophage infiltration syndromes (Figure 4A). Additional analyses identified rs1332327 (along with other SNPs in LD with the variant) as a *cis*-eQTL for *LIPA* in the METSIM adipose RNA-sequencing dataset (Table S13); this finding is consistent with eQTL results reported in the GTEx Portal for these SNPs and *LIPA*. Second, from the credible set for red blood cell count, we found a significantly associated SNP (rs1010222, $Z = -5.97$) at the promoter of *CALR* (MIM: 109091), a Mendelian disorder gene linked to myelofibrosis (MIM: 254450), which involves generalized bone marrow fibrosis, reduced hemopoiesis, a lack of hemophagocytosis, and myeloproliferative disease (Figure 4B). In both cases, the putative causal SNP for the complex trait lies immediately upstream of the TSS of the phenotypically relevant Mendelian disorder gene, in addition to falling within regions containing regulatory epigenetic marks.

Putative Causal SNPs for GWAS Regions Interacting with Promoters of Phenotypically Relevant Mendelian Disorder Genes

Functional genomic datasets, such as chromatin interactions identified through Hi-C, can give us insight into

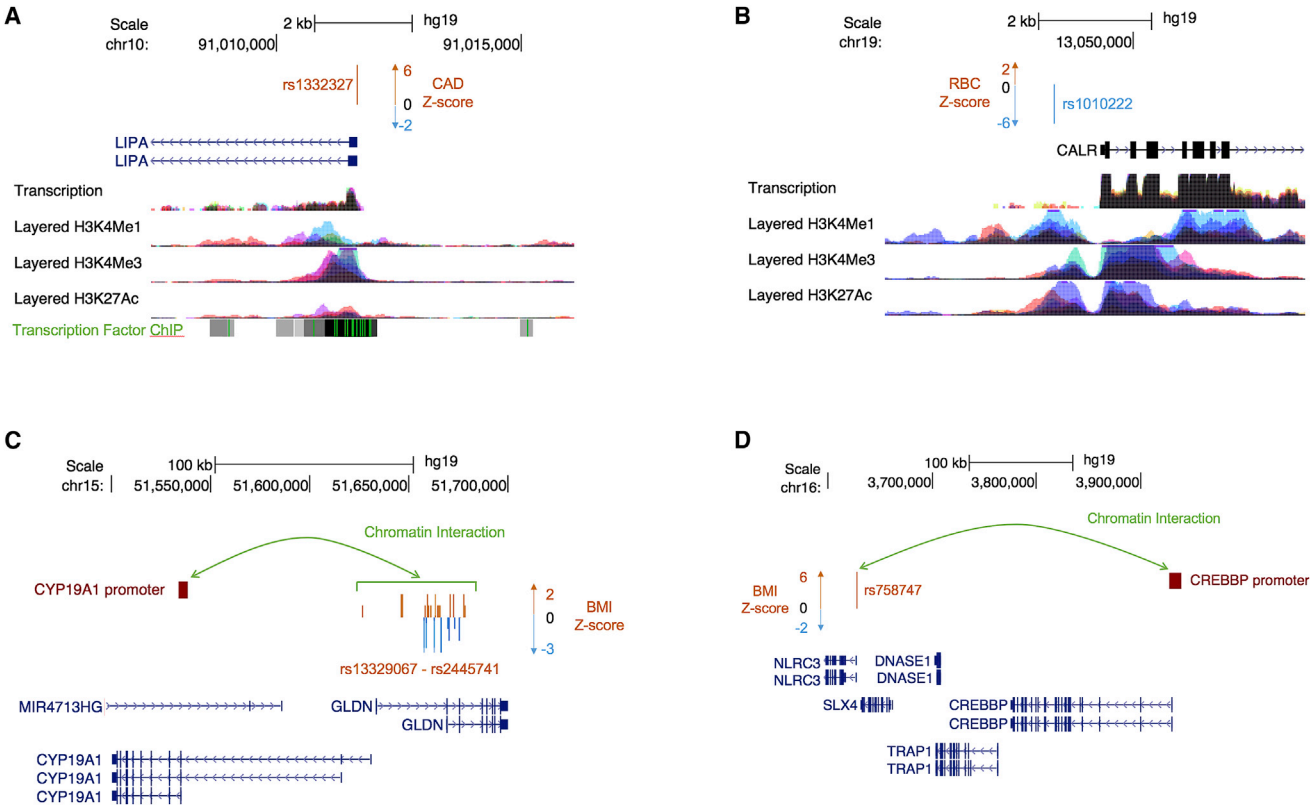

Figure 4. Candidate Regulatory SNPs Fall at TSSs and Long-Range Promoters of Phenotypically Relevant Mendelian Disorder Genes
(A and B) Shown here are two examples of putative causal SNPs localizing at a TSS of a phenotypically relevant Mendelian disorder gene. (A) Putative causal SNP rs1332327, associated with coronary artery disease ($Z = 6.80$), lies at the TSS of *LIPA*. (B) Putative causal SNP rs1010222, associated with red blood cell count ($Z = -5.97$), lies at the TSS of *CALR*.
(C and D) Shown here are two representations of chromatin interactions in white adipose tissue. (C) A cluster of SNPs from the credible set of variants associated with BMI (*Z* score plotted in orange and blue) physically interacts with the promoter of a particular isoform of *CYP19A1*. (D) A single SNP (rs758747) from the credible set, associated with BMI ($Z = 6.08$), physically interacts with the promoter of a distant gene, *CREBBP*.

the functional interpretation of GWAS variants and how they might regulate Mendelian disorder genes. Examination of chromatin interactions in human primary white adipocytes[66] revealed further candidate credible-set, metabolic-trait SNPs physically interacting with promoters of phenotypically relevant Mendelian disorder genes (Table S12). Specifically, we report that a genome-wide-significant SNP for BMI (rs758747, $Z = 6.08$) physically interacts with the promoter of *CREBBP*, a gene linked to Rubinstein-Taybi syndrome 1 (MIM: 180849), in which obesity is one of the syndromic features (Figure 4C). These interactions can also identify the relevant isoforms of genes in disease. From the credible set of variants associated with BMI, we identified a cluster of SNPs that physically interact with the promoter of a specific isoform of *CYP19A1*, a gene linked to aromatase excess syndrome (MIM: 139300), which involves short stature and excess fat storage in the chest (gynecomastia) (Figure 4D). Although longer isoforms of *CYP19A1* are by default chosen to represent the gene, our data suggest that the shorter isoform is likely to be more relevant in obesity. Together, these results demonstrate examples of GWAS variants localizing in regulatory regions for phenotypically relevant Mendelian disorder genes, consistent with the hypothesis that low-effect common variants contribute to complex traits by regulating genes known to cause Mendelian disorders.

Discussion

In this work, we used GWAS summary statistics from 62 complex traits and genes linked to specific phenotypes within 20 Mendelian broad disorders to quantify the shared genetic basis of complex traits and Mendelian disorders. We identified a pervasive, specific enrichment of phenotypically matched and related Mendelian disorder genes in GWAS regions for complex traits; we also identified some pairs of complex traits and phenotypically unmatched Mendelian disorders with significant enrichment. We further found that phenotypically relevant Mendelian disorder genes are enriched with GWAS signal across complex traits in comparison with all Mendelian disorder genes and other protein-coding genes. Finally, we report examples of putative causal SNPs for GWAS regions in potentially regulating phenotypically relevant Mendelian disorder genes. We conclude with four considerations about how our

results contribute to the understanding of genetic architectures and biological mechanisms across complex traits and Mendelian disorders.

First, our finding of a specific enrichment of phenotypically matched and related Mendelian disorder genes in GWAS regions for complex traits suggests that, across complex-trait architectures, many complex traits share a genetic basis (and by extension, biological mechanisms) with their Mendelian forms. This supports our hypothesis that the shared genes contribute to both extreme and common genetic phenotypes and suggests an important role of gene regulation by non-coding variants in complex traits. However, we note that our findings are limited by the power of each GWAS to detect significant associations. As GWASs become better powered, we anticipate being able to identify phenotype-specific enrichments of Mendelian disorder genes in GWAS regions for more complex traits.

Second, we identified a subset of pairs of complex traits and Mendelian disorders that, despite having no known shared biology, still demonstrated significant enrichment of Mendelian disorder genes in GWAS regions. This subset of pairs can offer us insight into the biological mechanisms of complex traits and Mendelian disorders. A high degree of co-morbidity between complex traits and Mendelian disorders has been previously observed regardless of phenotype similarity;[24] these findings together suggest that many complex traits and Mendelian disorders might also be linked by the pleiotropic properties of the underlying genes, in addition to regulatory differences. These observations are also consistent with a multigenic or oligogenic architecture of human disease; the pervasive pleiotropic effects that are seen across complex traits are consistent with the widespread prevalence of multi-system, syndromic phenotypes observed across a majority of Mendelian disorders. We also confirm that LoF-intolerant genes harbor an enrichment of GWAS signal;[68] because genes with pLI $>$ 0.9 exhibit extreme intolerance of deleterious mutation, it is possible that these genes demonstrate embryonically lethal mutant phenotypes and are thus undiscovered as Mendelian disorder genes at this time. Our findings provide further motivation to explore phenotypic consequences of mutations in LoF-intolerant genes (particularly those enriched with GWAS signal for a particular complex trait) for phenotypically relevant Mendelian disorders.

Third, linking Mendelian disorder genes with complex traits can help with characterization of the genetic architecture of complex traits—specifically with genes and pathways that can be functionally characterized for the identification of molecular mechanisms.[7] Identifying causal variants from large-scale GWASs is particularly challenging given that most GWAS loci lie in non-coding regions of the genome; although thousands of genomic loci have been significantly associated with specific diseases, few casual SNPs have been functionally verified.[82,83] Many approaches have been used to tie a particular variant to a relevant gene or genes,[84–86] including newer methods that directly link gene expression to a trait (e.g., transcriptome-wide association studies[70] and PrediXcan[87]); we have found that leveraging GWAS findings with functional data to identify candidate regulatory variants for Mendelian disorder genes can potentially lead to better interpretation of relevant genes and isoforms. Here, we demonstrate the heterogeneity of mechanisms potentially underlying causal variation by showing roles for TSS promoter regions of Mendelian disorder genes and long-range interactions involving significant GWAS regions. We expand on recent work showing that BMI-associated variants interact with genes in GWAS regions to demonstrate similar findings for Mendelian disorder genes.[66] With the appropriate functional data from relevant tissues and cell types, this phenotype-driven approach can identify relevant candidate regulatory variants and their targets. Further, from the perspective of monogenic diseases, identifying common variants that might modify the expressivity of phenotypes can provide insights into gene function in addition to putative drug targets. Many drugs approved by the FDA and developed by pharmaceutical companies are targeted toward the treatment of complex traits and diseases; by identifying underlying links between Mendelian disorders and complex traits through their effects on the same biological genes and pathways, we can systematically and rationally target existing drugs for complex traits and diseases toward those with rare Mendelian disorders that largely do not have any rationally targeted treatments.[88–90]

Fourth and finally, we note that our approach of examining traits and disorders at the component-phenotype level offers us valuable resolution into the specific pathways involved in the overall trait or disorder. In clinical genetics, genome-wide sequencing has expanded the clinical phenotypic spectrum associated with a gene[91,92] through identification of pleiotropic effects due to mutations in specific protein domains.[93,94] These advances have detected a genetic predisposition for diseases previously considered to be due to environment[13] and uncovered variable penetrance for genetic mutations previously thought to be sufficient to cause disease. Furthermore, the findings from clinical genome-wide sequencing have suggested that genetic background influences the phenotypic variability of monogenic diseases.[95,96] The phenotypic characterizations of Mendelian syndromes are deconstructed by expert clinical geneticists into component phenotypes, labeled by standardized clinical terms that identify both the primary phenotypes and phenotypes that have variable penetrance and expressivity.[97] Recent work has demonstrated that incorporating such dense phenotype information into ranking putative disease-causing genetic mutations improves diagnostic rates in clinical exome sequencing tests;[98,99] using component Mendelian phenotypes to identify Mendelian disorders that might be phenotypically relevant to a variety of complex traits can be similarly impactful in identifying biological pathways for complex traits. Ultimately, identification of GWAS-significant regions with biologically relevant genes and pathways will enable effective utilization of GWAS data in medical settings.

Supplemental Data

Supplemental Data include 2 figures and 14 tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.08.017.

Acknowledgments

We would like to thank Ruth Johnson, Megan Major, Megan Roytman, Claudia Giambartolomei, Arunabha Majumdar, Jazlyn Mooney, Brendan Freund, Robert Brown, Tommer Schwarz, and Robert Smith for helpful discussions. This work was funded by National Institutes of Health (NIH) Training Grant in Genomic Analysis and Interpretation T32HG002536 to M.K.F.; NIH Early Independence Award DP5OD024579 to V.A.A.; NIH awards R01HG009120, R01MH115676, R01HG006399, and U01CA194393 to B.P.; NIH National Institute of Mental Health grant T32MH073526 to K.S.B.; NIH grant F31HL142180 to K.M.G.; NIH National Cancer Institute grant T32LM012424 to D.Z.P.; and NIH grants HL-095056 and HL-28481 to P.P. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Interests

The authors declare no competing interests.

Received: May 16, 2018
Accepted: August 28, 2018
Published: October 4, 2018

Web Resources

1000 Genomes Phase 3, http://www.internationalgenome.org/category/phase-3/
ClinVar, ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/gene_specific_summary.txt
ExAC Browser, http://exac.broadinstitute.org/downloads
Gene sets, https://github.com/bogdanlab/gene_sets
GTEx Portal, https://www.gtexportal.org/home/
HapMap 3, https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html
HGNC, http://www.genenames.org/cgi-bin/download
OMIM, https://omim.org/
UCSC GTEx Combined eQTL Track, http://genome.ucsc.edu/cgi-bin/hgGtexTrackSettings?db=hg19&g=gtexEqtlCluster
UCSC Table Browser, https://genome.ucsc.edu/cgi-bin/hgTables

References

1. Schork, N.J., Murray, S.S., Frazer, K.A., and Topol, E.J. (2009). Common vs. rare allele hypotheses for complex diseases. Curr. Opin. Genet. Dev. *19*, 212–219.
2. Cutting, G.R., Kasch, L.M., Rosenstein, B.J., Zielenski, J., Tsui, L.C., Antonarakis, S.E., and Kazazian, H.H., Jr. (1990). A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. Nature *346*, 366–369.
3. Henriksen, M.G., Nordgaard, J., and Jansson, L.B. (2017). Genetics of schizophrenia: Overview of methods, findings and limitations. Front. Hum. Neurosci. *11*, 322.
4. Katsanis, N. (2016). The continuum of causality in human genetic disorders. Genome Biol. *17*, 233.
5. Auer, P.L., Teumer, A., Schick, U., O'Shaughnessy, A., Lo, K.S., Chami, N., Carlson, C., de Denus, S., Dubé, M.P., Haessler, J., et al. (2014). Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat. Genet. *46*, 629–634.
6. Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K., and Hobbs, H.H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet. *37*, 161–165.
7. Tsai, C.W., North, K.E., Tin, A., Haack, K., Franceschini, N., Saroja Voruganti, V., Laston, S., Zhang, Y., Best, L.G., MacCluer, J.W., et al. (2015). Both rare and common variants in PCSK9 influence plasma low-density lipoprotein cholesterol level in American Indians. J. Clin. Endocrinol. Metab. *100*, E345–E349.
8. Han, K., Holder, J.L., Jr., Schaaf, C.P., Lu, H., Chen, H., Kang, H., Tang, J., Wu, Z., Hao, S., Cheung, S.W., et al. (2013). SHANK3 overexpression causes manic-like behaviour with unique pharmacogenetic properties. Nature *503*, 72–77.
9. Sztainberg, Y., and Zoghbi, H.Y. (2016). Lessons learned from studying syndromic autism spectrum disorders. Nat. Neurosci. *19*, 1408–1417.
10. Lin, A., Ching, C.R.K., Vajdi, A., Sun, D., Jonas, R.K., Jalbrzikowski, M., Kushan-Wells, L., Pacheco Hansen, L., Krikorian, E., Gutman, B., et al. (2017). Mapping 22q11.2 gene dosage effects on brain morphometry. J. Neurosci. *37*, 6183–6199.
11. Toro, R., Konyukh, M., Delorme, R., Leblond, C., Chaste, P., Fauchereau, F., Coleman, M., Leboyer, M., Gillberg, C., and Bourgeron, T. (2010). Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet. *26*, 363–372.
12. Posey, J.E., Harel, T., Liu, P., Rosenfeld, J.A., James, R.A., Coban Akdemir, Z.H., Walkiewicz, M., Bi, W., Xiao, R., Ding, Y., et al. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. N. Engl. J. Med. *376*, 21–31.
13. Corvol, H., Blackman, S.M., Boëlle, P.Y., Gallins, P.J., Pace, R.G., Stonebraker, J.R., Accurso, F.J., Clement, A., Collaco, J.M., Dang, H., et al. (2015). Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat. Commun. *6*, 8382.
14. Emond, M.J., Louie, T., Emerson, J., Zhao, W., Mathias, R.A., Knowles, M.R., Wright, F.A., Rieder, M.J., Tabor, H.K., Nickerson, D.A., et al.; National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project; and Lung GO (2012). Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat. Genet. *44*, 886–889.
15. Dorfman, R., Sandford, A., Taylor, C., Huang, B., Frangolias, D., Wang, Y., Sang, R., Pereira, L., Sun, L., Berthiaume, Y., et al. (2008). Complex two-gene modulation of lung disease severity in children with cystic fibrosis. J. Clin. Invest. *118*, 1040–1049.
16. Dron, J.S., and Hegele, R.A. (2017). Genetics of triglycerides and the risk of atherosclerosis. Curr. Atheroscler. Rep. *19*, 31.
17. Hegele, R.A. (2018). Learning from patients with ultrarare conditions: Cholesterol hoof beats. J. Am. Coll. Cardiol. *71*, 289–291.
18. Peltonen, L., Perola, M., Naukkarinen, J., and Palotie, A. (2006). Lessons from studying monogenic disease for common disease. Hum. Mol. Genet. *15*, R67–R74.

19. Hernandez, D.G., Reed, X., and Singleton, A.B. (2016). Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. J. Neurochem. *139* (*Suppl 1*), 59–74.
20. Lim, E.T., Liu, Y.P., Chan, Y., Tiinamaija, T., Käräjämäki, A., Madsen, E., Altshuler, D.M., Raychaudhuri, S., Groop, L., Flannick, J., et al.; Go-T2D Consortium (2014). A novel test for recessive contributions to complex diseases implicates Bardet-Biedl syndrome gene BBS10 in idiopathic type 2 diabetes and obesity. Am. J. Hum. Genet. *95*, 509–520.
21. Chan, Y., Salem, R.M., Hsu, Y.H., McMahon, G., Pers, T.H., Vedantam, S., Esko, T., Guo, M.H., Lim, E.T., Franke, L., et al.; GIANT Consortium (2015). Genome-wide analysis of body proportion classifies height-associated variants by mechanism of action and implicates genes important for skeletal development. Am. J. Hum. Genet. *96*, 695–708.
22. Wheeler, H.E., Gamazon, E.R., Frisina, R.D., Perez-Cervantes, C., El Charif, O., Mapes, B., Fossa, S.D., Feldman, D.R., Hamilton, R.J., Vaughn, D.J., et al. (2017). Variants in *WFS1* and other Mendelian deafness genes are associated with cisplatin-associated ototoxicity. Clin. Cancer Res. *23*, 3325–3333.
23. Amininejad, L., Charloteaux, B., Theatre, E., Liefferinckx, C., Dmitrieva, J., Hayard, P., Muls, V., Maisin, J.M., Schapira, M., Ghislain, J.M., et al.; International IBD Genetics Consortium (2018). Analysis of genes associated with monogenic primary immunodeficiency identifies rare variants in XIAP in patients with Crohn's disease. Gastroenterology *154*, 2165–2177.
24. Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H., Melamed, R., Rabadan, R., Bernstam, E.V., Brunak, S., et al. (2013). A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell *155*, 70–80.
25. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.
26. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adány, R., Aromaa, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. *42*, 295–302.
27. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al.; International Multiple Sclerosis Genetics Consortium; and International IBD Genetics Consortium (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.
28. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y., et al. (2011). New gene functions in megakaryopoiesis and platelet formation. Nature *480*, 201–208.
29. Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfield, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C., et al.; Canadian-US PBC Consortium; Italian PBC Genetics Study Group; and UK-PBC Consortium (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat. Commun. *6*, 8019.
30. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; and GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.
31. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., et al.; International Multiple Sclerosis Genetics Consortium; and Wellcome Trust Case Control Consortium 2 (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature *476*, 214–219.
32. Autism Spectrum Disorder Working Group of the Psychiatry Genomics Consortium (2015). Dataset: PGC-ASD summary statistics from a meta-analysis of 5,305 ASD-diagnosed cases and 5,305 pseudocontrols of European descent (based on similarity to CEPH reference genotypes). March 2015. http://www.med.unc.edu/pgc/results-and-downloads.
33. van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. Nature *492*, 369–375.
34. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tombleson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet. *47*, 1457–1464.
35. Horikoshi, M., Beaumont, R.N., Day, F.R., Warrington, N.M., Kooijman, M.N., Fernandez-Tajes, J., Feenstra, B., van Zuydam, N.R., Gaulton, K.J., Grarup, N., et al.; CHARGE Consortium Hematology Working Group; and Early Growth Genetics (EGG) Consortium (2016). Genome-wide associations for birth weight and correlations with adult disease. Nature *538*, 248–252.
36. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; and LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.
37. Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L., Park-Min, K.H., Tobias, J.H., Kooperberg, C., et al.; AOGC Consortium; and UK10K Consortium (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature *526*, 112–117.
38. Köttgen, A., Albrecht, E., Teumer, A., Vitart, V., Krumsiek, J., Hundertmark, C., Pistis, G., Ruggiero, D., O'Seaghdha, C.M., Haller, T., et al.; LifeLines Cohort Study; CARDIoGRAM Consortium; DIAGRAM Consortium; ICBP Consortium; and MAGIC Consortium (2013). Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat. Genet. *45*, 145–154.
39. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130.
40. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.

41. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., et al.; WTCCC (2010). Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. Diabetes *59*, 3229–3239.
42. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., et al.; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. *44*, 981–990.
43. Fritsche, L.G., Igl, W., Bailey, J.N., Grassmann, F., Sengupta, S., Bragg-Gresham, J.L., Burdon, K.P., Hebbring, S.J., Wen, C., Gorski, M., et al. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat. Genet. *48*, 134–143.
44. Perry, J.R., Day, F., Elks, C.E., Sulem, P., Thompson, D.J., Ferreira, T., He, C., Chasman, D.I., Esko, T., Thorleifsson, G., et al.; Australian Ovarian Cancer Study; GENICA Network; kConFab; LifeLines Cohort Study; InterAct Consortium; and Early Growth Genetics (EGG) Consortium (2014). Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature *514*, 92–97.
45. Day, F.R., Ruth, K.S., Thompson, D.J., Lunetta, K.L., Pervjakova, N., Chasman, D.I., Stolk, L., Finucane, H.K., Sulem, P., Bulik-Sullivan, B., et al.; PRACTICAL consortium; kConFab Investigators; AOCS Investigators; Generation Scotland; EPIC-InterAct Consortium; and LifeLines Cohort Study (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nat. Genet. *47*, 1294–1303.
46. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; and MAGIC investigators (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. *42*, 105–116.
47. Teumer, A., Tin, A., Sorice, R., Gorski, M., Yeo, N.C., Chu, A.Y., Li, M., Li, Y., Mijatovic, V., Ko, Y.A., et al.; DCCT/EDIC (2016). Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. Diabetes *65*, 803–817.
48. Saxena, R., Hivert, M.F., Langenberg, C., Tanaka, T., Pankow, J.S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A.U., et al.; GIANT consortium; and MAGIC investigators (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat. Genet. *42*, 142–148.
49. Onengut-Gumuscu, S., Chen, W.M., Burren, O., Cooper, N.J., Quinlan, A.R., Mychaleckyj, J.C., Farber, E., Bonnie, J.K., Szpak, M., Schofield, E., et al.; Type 1 Diabetes Genetics Consortium (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat. Genet. *47*, 381–386.
50. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al.; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; and Cohorts for Heart and Aging Research in Genomic Epidemiology (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. *45*, 1452–1458.
51. Otowa, T., Hek, K., Lee, M., Byrne, E.M., Mirza, S.S., Nivard, M.G., Bigdeli, T., Aggen, S.H., Adkins, D., Wolen, A., et al. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. Mol. Psychiatry *21*, 1391–1399.
52. Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H., Boomsma, D.I., Cichon, S., et al.; Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol. Psychiatry *18*, 497–511.
53. Okbay, A., Baselmans, B.M., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F., Linnér, R.K., Rietveld, C.A., Derringer, J., et al.; LifeLines Cohort Study (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. *48*, 624–633.
54. Psychiatric, G.C.B.D.W.G.; and Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. *43*, 977–983.
55. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.
56. Pattaro, C., Teumer, A., Gorski, M., Chu, A.Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R., Li, Y., et al.; ICBP Consortium; AGEN Consortium; CARDIOGRAM; CHARGe-Heart Failure Group; and ECHOGen Consortium (2016). Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat. Commun. *7*, 10023.
57. Eppinga, R.N., Hagemeijer, Y., Burgess, S., Hinds, D.A., Stefansson, K., Gudbjartsson, D.F., van Veldhuisen, D.J., Munroe, P.B., Verweij, N., and van der Harst, P. (2016). Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. Nat. Genet. *48*, 1557–1563.
58. Barban, N., Jansen, R., de Vlaming, R., Vaez, A., Mandemakers, J.J., Tropf, F.C., Shen, X., Wilson, J.F., Chasman, D.I., Nolte, I.M., et al.; BIOS Consortium; and LifeLines Cohort Study (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. Nat. Genet. *48*, 1462–1472.
59. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al.; LifeLines Cohort Study (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science *340*, 1467–1471.
60. Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.B., Emilsson, V., Meddens, S.F., et al.; LifeLines Cohort Study (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature *533*, 539–542.
61. Lu, Y., Day, F.R., Gustafsson, S., Buchkovich, M.L., Na, J., Bataille, V., Cousminer, D.L., Dastani, Z., Drong, A.W., Esko, T.,

et al. (2016). New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. Nat. Commun. *7*, 10495.
62. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; and International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.
63. Felix, J.F., Bradfield, J.P., Monnereau, C., van der Valk, R.J., Stergiakouli, E., Chesi, A., Gaillard, R., Feenstra, B., Thiering, E., Kreiner-Møller, E., et al.; Bone Mineral Density in Childhood Study (BMDCS); Early Genetics and Lifecourse Epidemiology (EAGLE) consortium; Early Growth Genetics (EGG) Consortium; and Bone Mineral Density in Childhood Study BMDCS (2016). Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. Hum. Mol. Genet. *25*, 389–403.
64. Kilpeläinen, T.O., Carli, J.F., Skowronski, A.A., Sun, Q., Kriebel, J., Feitosa, M.F., Hedman, A.K., Drong, A.W., Hayes, J.E., Zhao, J., et al. (2016). Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. Nat. Commun. *7*, 10494.
65. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al.; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; and ReproGen Consortium (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature *518*, 187–196.
66. Pan, D.Z., Garske, K.M., Alvarez, M., Bhagat, Y.V., Boocock, J., Nikkola, E., Miao, Z., Raulerson, C.K., Cantor, R.M., Civelek, M., et al. (2018). Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. Nat. Commun. *9*, 1512.
67. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.
68. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
69. Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet. *18*, 117–127.
70. Mancuso, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., and Pasaniuc, B. (2017). Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. Am. J. Hum. Genet. *100*, 473–487.
71. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
72. Fisher, R.A. (1922). On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. J. R. Stat. Soc. *85*, 87–94.
73. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
74. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. *94*, 559–573.
75. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al.; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and RACI Consortium (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.
76. Fritsche, L.G., Chen, W., Schu, M., Yaspan, B.L., Yu, Y., Thorleifsson, G., Zack, D.J., Arakawa, S., Cipriani, V., Ripke, S., et al.; AMD Gene Consortium (2013). Seven new loci associated with age-related macular degeneration. Nat. Genet. *45*, 433–439, e1–e2.
77. Helgason, H., Sulem, P., Duvvari, M.R., Luo, H., Thorleifsson, G., Stefansson, H., Jonsdottir, I., Masson, G., Gudbjartsson, D.F., Walters, G.B., et al. (2013). A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration. Nat. Genet. *45*, 1371–1374.
78. Brunham, L.R., Singaraja, R.R., and Hayden, M.R. (2006). Variations on a gene: rare and common variants in ABCA1 and their impact on HDL cholesterol levels and atherosclerosis. Annu. Rev. Nutr. *26*, 105–129.
79. Kanoni, S., Masca, N.G., Stirrups, K.E., Varga, T.V., Warren, H.R., Scott, R.A., Southam, L., Zhang, W., Yaghootkar, H., Müller-Nurasyid, M., et al.; Wellcome Trust Case Control Consortium (2016). Analysis with the exome array identifies multiple new independent variants in lipid loci. Hum. Mol. Genet. *25*, 4094–4106.
80. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.
81. Rosenson, R.S., Brewer, H.B., Jr., Barter, P.J., Björkegren, J.L.M., Chapman, M.J., Gaudet, D., Kim, D.S., Niesor, E., Rye, K.A., Sacks, F.M., et al. (2018). HDL and atherosclerotic cardiovascular disease: genetic insights into complex biology. Nat. Rev. Cardiol. *15*, 9–19.
82. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530*, 177–183.
83. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375.
84. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants

clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.
85. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.
86. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.
87. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., Im, H.K.; and GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
88. Pascual, V., Allantaz, F., Arce, E., Punaro, M., and Banchereau, J. (2005). Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade. J. Exp. Med. *201*, 1479–1486.
89. Pardoll, D.M. (2012). Immunology beats cancer: a blueprint for successful translation. Nat. Immunol. *13*, 1129–1132.
90. Gerich, M.E., and McGovern, D.P. (2014). Towards personalized care in IBD. Nat. Rev. Gastroenterol. Hepatol. *11*, 287–299.
91. Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., et al.; Centers for Mendelian Genomics (2015). The genetic basis of Mendelian phenotypes: Discoveries, challenges, and opportunities. Am. J. Hum. Genet. *97*, 199–215.
92. Reuter, C.M., Brimble, E., DeFilippo, C., Dries, A.M., Enns, G.M., Ashley, E.A., Bernstein, J.A., Fisher, P.G., Wheeler, M.T.; and Undiagnosed Diseases Network (2018). A new approach to rare diseases of children: The Undiagnosed Diseases Network. J. Pediatr. *196*, 291–297.e2.
93. Goudie, D.R., D'Alessandro, M., Merriman, B., Lee, H., Szeverényi, I., Avery, S., O'Connor, B.D., Nelson, S.F., Coats, S.E., Stewart, A., et al. (2011). Multiple self-healing squamous epithelioma is caused by a disease-specific spectrum of mutations in TGFBR1. Nat. Genet. *43*, 365–369.
94. Arboleda, V.A., Lee, H., Parnaik, R., Fleming, A., Banerjee, A., Ferraz-de-Souza, B., Délot, E.C., Rodriguez-Fernandez, I.A., Braslavsky, D., Bergadá, I., et al. (2012). Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. Nat. Genet. *44*, 788–792.
95. Born, H.A., Dao, A.T., Levine, A.T., Lee, W.L., Mehta, N.M., Mehra, S., Weeber, E.J., and Anderson, A.E. (2017). Strain-dependence of the Angelman Syndrome phenotypes in Ube3a maternal deficiency mice. Sci. Rep. *7*, 8451.
96. Hensman Moss, D.J., Pardiñas, A.F., Langbehn, D., Lo, K., Leavitt, B.R., Roos, R., Durr, A., Mead, S., Holmans, P., Jones, L., Tabrizi, S.J.; TRACK-HD investigators; and REGISTRY investigators (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol. *16*, 701–711.
97. Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. *42*, D966–D974.
98. Yang, H., Robinson, P.N., and Wang, K. (2015). Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Nat. Methods *12*, 841–843.
99. Zemojtel, T., Köhler, S., Mackenroth, L., Jäger, M., Hecht, J., Krawitz, P., Graul-Neumann, L., Doelken, S., Ehmke, N., Spielmann, M., et al. (2014). Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci. Transl. Med. *6*, 252ra123.

The American Journal of Human Genetics, Volume *103*

Supplemental Data

Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits

Malika Kumar Freund, Kathryn S. Burch, Huwenbo Shi, Nicholas Mancuso, Gleb Kichaev, Kristina M. Garske, David Z. Pan, Zong Miao, Karen L. Mohlke, Markku Laakso, Päivi Pajukanta, Bogdan Pasaniuc, and Valerie A. Arboleda

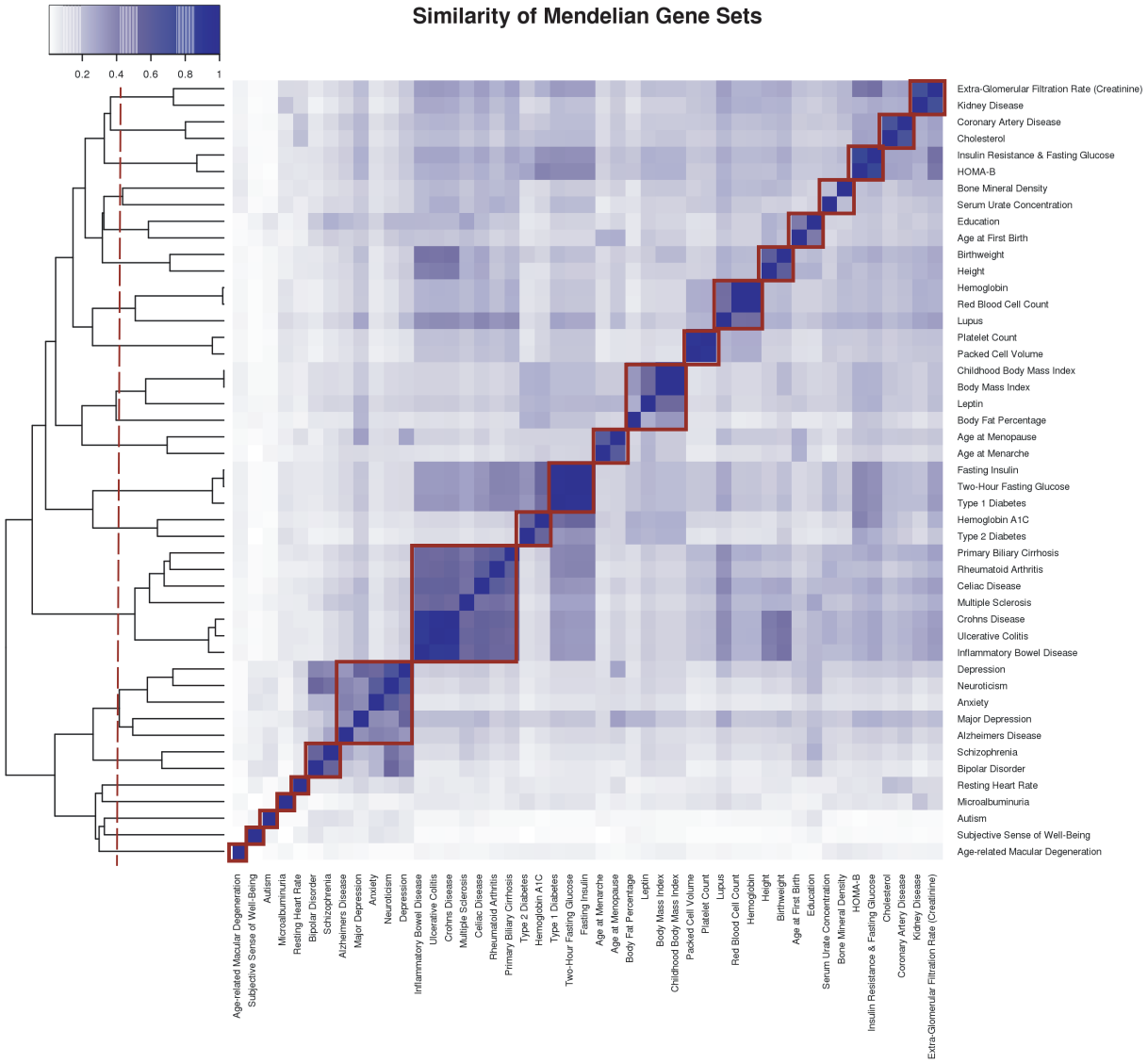

Figure S1: Similarity of Mendelian disorder gene sets

After generation of phenotype-specific Mendelian disorder gene sets, we performed pairwise comparisons of each gene set to determine proportions of genes shared. We performed hierarchical clustering, and gene sets sharing large proportions of genes (identified by visual clusters based on a hierarchical clustering threshold, indicated in red boxes and red dashed line respectively) were intersected to form a single representative Mendelian disorder gene set (see Table S3 for these cluster descriptions).

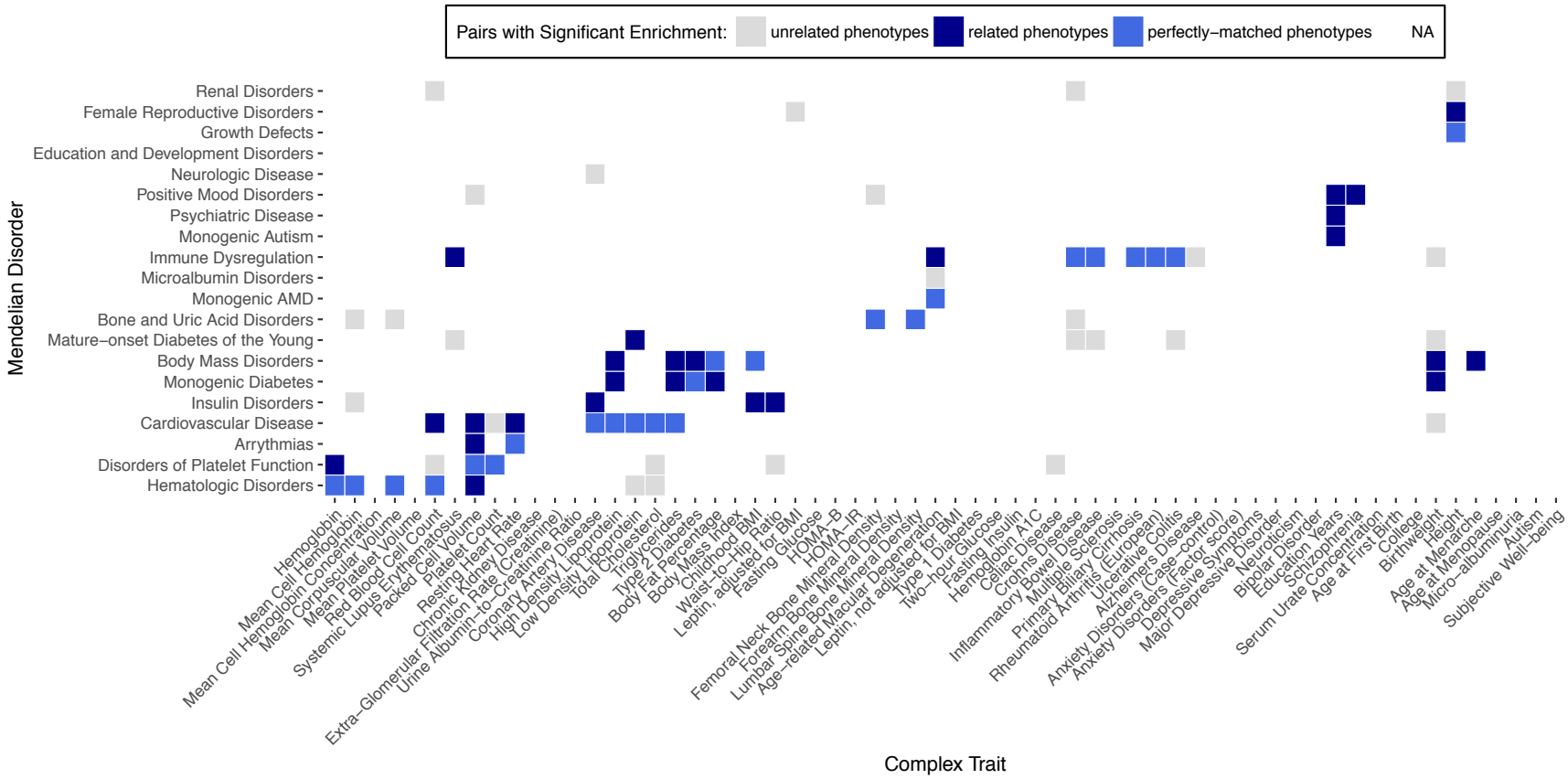

Figure S2: Overlap of GWAS genes with Mendelian disorder genes demonstrates trait-specificity

After generation of phenotype-specific Mendelian disorder gene sets, we performed pairwise comparisons of each gene set to determine proportions of genes shared. We performed hierarchical clustering, and gene sets sharing large proportions of genes (identified by visual clusters based on a hierarchical clustering threshold, indicated in red boxes and red dashed line respectively) were intersected to form a single representative Mendelian disorder gene set (see Table S3 for these cluster descriptions).

Complex Trait	Abbrev.	Mean GWAS Sample Size	Number of Significant GWAS Loci Reported	Number of Significant GWAS SNPs	Number of GWAS Genes	Matched Mendelian Disorder(s)
Celiac Disease[60]	CEL	15283	13	54	34	Immune Dysregulation
Crohn's Disease[61]	CD	27726	231 total	4381	239	
Inflammatory Bowel Disease[61]	IBD	34694		7738	368	
Ulcerative Colitis[67]	UC	28738		4239	202	
Primary Biliary Cirrhosis[62]	PBC	13239	28	704	149	
Rheumatoid Arthritis (European)[63]	RA	58284	101	16502	297	
Multiple Sclerosis[64]	MS	27148	52	487	160	
Autism[65]	AUT	10610	2	2	2	Monogenic Autism
Hemoglobin[66]	HB	51255.8	11	325	89	Hematologic Disorders
Mean Cell Hemoglobin[66]	MCH	43553.6	19	1188	164	
Mean Cell Hemoglobin Concentration[66]	MCHC	46953.9	8	8	12	
Mean Corpuscular Volume[66]	MCV	48472.8	23	1237	180	
Mean Platelet Volume[67]	MPV	16843	25	705	102	
Red Blood Cell Count[66]	RBC	45304.4	10	908	107	
Systemic Lupus Erythematosus[68]	SLE	23210	43	4983	286	
Birthweight[69]	BW	110054.6	60	1978	179	Growth Defects
Height[70]	HGT	239338.3	423	22807	2361	
Femoral Neck Bone Mineral Density[71]	FN	53236	14	788	58	Bone and Uric Acid Disorders
Forearm Bone Mineral Density[71]	FA	53236	3	136	8	
Lumbar Spine Bone Mineral Density[71]	LS	53236	19	998	67	
Serum Urate Concentration[72]	URT	107026	28	1991	161	

Packed Cell Volume[66]	PCV	44925.9	4	141	53	Disorders of Platelet Function
Platelet Count[67]	PLT	66867	43	801	134	
Coronary Artery Disease[73]	CAD	184305	46	1709	132	Cardiovascular Disease
High Density Lipoprotein[74]	HDL	95422	157 total	3131	464	
Low Density Lipoprotein[74]	LDL	90686.4		2796	370	
Total Cholesterol[74]	TC	95651.5		3803	500	
Triglycerides[74]	TG	91882.3		2965	354	
Hemoglobin A1C[75]	HBA	46368	10	174	33	Monogenic Diabetes
Type 2 Diabetes[76]	T2D	61857.4	10	191	28	
Age-related Macular Degeneration[77]	AMD	33975.1	34	4087	215	Monogenic AMD
Age at Menarche[78]	MNR	182416	106	2011	207	Female Reproductive Disorders
Age at Menopause[79]	MNP	70000	44	1656	316	
Fasting Glucose[80]	FG	46186	16 total	231	39	Insulin Disorders
HOMA-B[80]	HMB	46186		95	12	
HOMA-IR[80]	HMIR	46186	1	1	0	
Micro-albuminuria[81]	MA	52988.4	1	4	2	Microalbumin Disorders
Fasting Insulin[80]	FI	108557	1	43	23	Mature-onset Diabetes of the Young
Two-hour Glucose[82]	2HG	15234	3	4	2	
Type 1 Diabetes[83]	T1D	24341	42	1447	144	
Alzheimer's Disease[84]	ALZ	54162	19	813	58	Neurologic Disease
Anxiety Disorders (Case-control)[85]	ANXC	14643.1	1	8	2	
Anxiety Disorders (Factor score)[85]	ANXF	16218.2	1	43	3	
Major Depressive Disorder[86]	MDD	10610	1	5	4	
Depressive Symptoms[87]	DS	161460	2	28	10	
Neuroticism[87]	NRT	170911	11	1933	82	
Bipolar Disorder[88]	BIP	16731	4	39	8	Psychiatric Disease

Schizophrenia[89]	SCZ	150064	108	6808	479	
Chronic Kidney Disease[90]	CKD	117340	4	107	16	Renal Disorders
Glomerular Filtration Rate (CRN)[90]	EGFR	132725.4	43	1401	162	
Urine Albumin-to-Creatinine Ratio[81]	UACR	53343.1	1	4	2	
Resting Heart Rate[91]	RHR	265046	64	4568	304	Arrythmias
Age at First Birth[92]	AFB	251151	12	238	45	Education and Development Disorders
College[93]	COL	126559	3	71	12	
Education Years[94]	EY	293723	74	6537	554	
Subjective Well-being[87]	SWB	298420	3	37	9	Positive Mood Disorders
Body Fat Percentage[95]	BFP	57721.7	12	196	22	Body Mass Disorders
Body Mass Index[96]	BMI	224996	97	1585	231	
Childhood BMI[97]	CBMI	28964.1	15	438	49	
Leptin, adjusted for BMI[98]	LEPB	30202	5 total	8	5	
Leptin, not adjusted for BMI[98]	LEP	30507.6		1	0	
Waist-to-Hip Ratio[99]	WHR	139559.2	49	424	74	

Table S1: Complex Traits and corresponding Mendelian disorders. This table details the phenotypically-matched pairs of complex traits (N=62) and groups of Mendelian disorders (N=20) examined in our study. Mean GWAS Sample Size and Number of Significant GWAS Loci are reported from original GWAS publications for each complex trait. Significant GWAS SNPs are all SNPs from the publicly available summary statistics meeting genome-wide significance at a threshold of $p < 5x10^{-8}$. GWAS genes for each complex trait were identified using the mapping approach described in Methods. References in this table match original article references in Table 1.

Method	Avg. Num. Genes per GWAS Gene Set	Num. Comparisons with Sig. Overlap (% of 1240 Total)	Correlation of Odds Ratios with Closest 2 Genes (r^2)
Closest 2 Genes to each significant SNP	167	77 (6.21%)	1.000
50kb Window	168	48 (3.87%)	0.792
Closest 2 Genes to credible set SNPs	154	34 (2.74%)	0.658
500kb Window	533	7 (0.56%)	0.582

Table S4: Different methods of choosing GWAS genes show similar overlap results

We mapped GWAS SNPs to genes using two additional window approaches and one credible set SNP-mapping approach, and performed overlap comparisons between a subset of complex gene sets and Mendelian order gene sets (see Methods). The correlation of results with results from our chosen approach is shown in the last column.

Gene Class	Avg. LD Tagged (CI)	Avg. MAF (CI)	Avg. Gene Length in bp (CI)
All Protein-Coding Genes	24.38 (24.35, 24.42)	0.238 (0.238, 0.238)	159937 (158161, 161714)
All Mendelian Disorder Genes	24.01 (23.93, 24.08)	0.239 (0.238, 0.239)	174461 (169804, 179117)
LOF-Intolerant Genes	24.19 (24.11, 24.26)	0.237 (0.237, 0.238)	214861 (208823, 220898)
Immune Dysregulation	24.49 (24.26, 24.71)	0.236 (0.235, 0.238)	162527 (154081, 170973)
Monogenic Autism	25.80 (25.47, 26.12)	0.244 (0.242, 0.247)	225164 (175162, 275165)
Hematologic Disorders	24.34 (24.15, 24.53)	0.239 (0.238, 0.240)	162128 (152296, 171959)
Growth Defects	25.54 (25.36, 25.71)	0.236 (0.235, 0.237)	169612 (161098, 178127)
Bone and Uric Acid Disorders	24.10 (23.79, 24.41)	0.242 (0.240, 0.244)	148424 (139958, 156891)
Disorders of Platelet Function	24.92 (24.69, 25.16)	0.240 (0.239, 0.241)	174260 (160545, 187975)
Cardiovascular Disease	24.79 (24.59, 24.98)	0.237 (0.236, 0.238)	168739 (158253, 179226)
Monogenic Diabetes	25.73 (25.29, 26.18)	0.231 (0.229, 0.233)	167935 (153263, 182606)
Monogenic AMD	19.75 (19.38, 20.13)	0.241 (0.238, 0.243)	168410 (149498, 187322)
Female Reproductive Disorders	24.44 (24.17, 24.71)	0.242 (0.241, 0.244)	163127 (152268, 173987)
Insulin Disorders	23.78 (23.58, 23.97)	0.238 (0.237, 0.239)	162444 (154415, 170474)
Microalbumin Disorders	23.38 (23.08, 23.69)	0.239 (0.236, 0.241)	171182 (148727, 193637)
Mature-onset Diabetes of the Young	25.32 (25.08, 25.56)	0.235 (0.234, 0.237)	161166 (152361, 169971)
Neurologic Disease	24.32 (24.05, 24.59)	0.239 (0.237, 0.241)	189466 (169021, 209911)
Psychiatric Disease	24.01 (23.78, 24.24)	0.239 (0.237, 0.241)	213791 (187313, 240270)
Renal Disorders	24.12 (23.97, 24.26)	0.239 (0.238, 0.240)	175981 (166510, 185453)
Arrythmias	22.40 (22.17, 22.63)	0.237 (0.236, 0.239)	179407 (160838, 197976)
Education and Development Disorders	25.16 (24.98, 25.34)	0.239 (0.238, 0.240)	207061 (191501, 222622)
Positive Mood Disorders	23.40 (22.95, 23.85)	0.234 (0.231, 0.238)	217893 (184211, 251575)
Body Mass Disorders	22.17 (21.80, 22.54)	0.238 (0.235, 0.241)	168600 (150261, 186940)

Table S6: Gene classes show no substantial difference in average LD tagged, average MAF, or average gene length

For each gene class as in Supplementary Table ST5, including all phenotype-specific Mendelian disorder gene sets, we identified the average linkage disequilibrium (LD) tagged, average MAF, and average gene length (see Methods).

Complex Trait	Matched Mendelian Disorder	Number of Shared Genes	Average Distance to Closest Gene (kb) (CI)
AFB	Education and Development Disorders	6	238.8 (238.8, 238.8)
AMD	Monogenic AMD	9	1150.6 (-1051.0, 3352.1)
BW	Growth Defects	11	20695.5 (20695.5, 20695.5)
CAD	Cardiovascular Disease	13	25800.4 (11533.2, 40067.6)
CD	Immune Dysregulation	23	30079.3 (7810.1, 52348.4)
EGFR	Renal Disorders	14	37595.2 (7356.9, 67833.5)
EY	Education and Development Disorders	31	23597.5 (6249.6, 40945.4)
HB	Hematologic Disorders	10	36833.0 (-8149.8, 81815.8)
HDL	Cardiovascular Disease	31	36971.5 (13607.7, 60335.3)
HGT	Growth Defects	126	11979.8 (8996.7, 14962.8)
IBD	Immune Dysregulation	34	20056.1 (6878.5, 33233.6)
LDL	Cardiovascular Disease	31	22265.1 (8710.4, 35819.8)
MCH	Hematologic Disorders	15	27415.7 (-7652.6, 62483.9)
MCV	Hematologic Disorders	20	45984.4 (19152.5, 72816.4)
MNP	Female Reproductive Disorders	8	12709.2 (763.2, 24655.3)
MS	Immune Dysregulation	11	508.1 (-73.3, 1089.4)
PBC	Immune Dysregulation	13	26.3 (-16.9, 69.6)
PCV	Disorders of Platelet Function	10	18526.7 (-6117.0, 43170.4)
PLT	Disorders of Platelet Function	12	28178.6 (-10382.2, 66739.4)
RA	Immune Dysregulation	25	5977.6 (-4798.8, 16754.0)
RBC	Hematologic Disorders	14	18321.6 (-8040.5, 44683.7)
RHR	Arrythmias	17	23161.8 (5088.1, 41235.5)
SCZ	Psychiatric Disease	9	35497.2 (15773.4, 55221.0)
SLE	Hematologic Disorders	10	22618.0 (-2792.1, 48028.1)
T1D	Mature-onset Diabetes of the Young	11	30051.2 (795.3, 59307.2)
TC	Cardiovascular Disease	38	15127.2 (4202.1, 26052.2)
TG	Cardiovascular Disease	25	34214.8 (11785.9, 56643.8)
UC	Immune Dysregulation	21	30688.6 (7802.3, 53574.8)
URT	Bone and Uric Acid Disorders	6	156.6 (156.6, 156.6)

Table S8: Average distance to closest gene, among genes shared by phenotypically-matched complex traits and Mendelian disorders

For each matched pair of complex trait and Mendelian disorder sharing at least two genes, we computed the average distance to the next closest gene within the chromosome, among all shared genes.

Trait	Mendelian Gene Set	Num. Credible SNPs interacting with Promoters	Number of SNPs in 95% Credible Set	Max Effect Size	Mendelian Disorder Genes
2HG	Mature-onset Diabetes of the Young	8	5017	2.00	*ADRB1 PNLIP*
BFP	Body Mass Disorders	36	25172	2.94	*CPT1C*
BMI	Body Mass Disorders	156	122891	6.08	*CREBBP CYP19A1 GCGR MMACHC NEUROD1 PDX1 PRKACA PTEN*
BW	Growth Defects	790	110715	7.89	*ACADVL BLM C5 COG6 GMNN HSPG2 JAG1 MMAA MPDU1 MUSK PNPO PTH1R SETD2 SLC34A1*
CAD	Cardiovascular Disease	353	75678	21.05	*ACTA2 BRIP1 CDKN2A CPT2 FLAD1 MAP2K1 NDUFS3 POLG PPM1D PRKAG2 PTPN11 REST*
CBMI	Body Mass Disorders	84	20137	2.57	*BDNF MTTP NFKB1*
HBA	Monogenic Diabetes	99	11348	2.10	*IRS2*
HDL	Cardiovascular Disease	283	41573	11.82	*APOA2 COQ2 DCAF8 GNAI2 MPZ MTTP NDUFS2 NDUFS3 NSMCE2 NTRK1 PPARG SMPD1 TIMMDC1 UAP1*
LDL	Cardiovascular Disease	234	40203	12.59	*ADD1 ANGPTL3 BCS1L CASP8 COL4A3 CREB1 GLB1 HESX1 IDUA MRPL44 NDUFAF4 NDUFS1 SMARCAL1 SPEG TNNC1 TSG101 WFS1*
LEPB	Body Mass Disorders	23	10038	4.75	*LEP PAX4*
RHR	Arrythmias	400	145978	13.05	*CACNA1C CALM1 GJA1 PTCH1 TTN*
T2D	Monogenic Diabetes	51	15092	4.70	*ADD1 CIDEC LEP OGG1 PAX4 PPARG WFS1*
TC	Cardiovascular Disease	229	31232	20.35	*ANGPTL3 CDKN2A CHKB FAM20A GLB1 NSMCE2 PINK1*
TG	Cardiovascular Disease	367	79786	5.75	*ABCA1 ABCG2 CAV1 FKTN FOXC2 GNPTAB IGF1 ISCU MLYCD MYBPC3 MYH11 MYLK2 PNPLA8 PPP1R3A PTGIS RSPO1 SLC2A2 ZMPSTE24*
WHR	Body Mass Disorders	134	55280	3.89	*CPT1C F5 MC3R POLD1 PPARG RAI1*

Table S12: Summary of credible SNPs interacting with promoters of phenotypically-matched Mendelian disorder genes in metabolic traits
Fine mapping was performed for each complex trait to produce a 95% credible set of SNPs (see Methods). This table summarizes the number of SNPs from the 95% credible set for each complex trait physically interacting with the promoter of a phenotypically-matched Mendelian disorder gene. For reference, the maxium effect size of these SNPs are listed along with the Mendelian disorder genes (italicized) whose promoters participate in the interaction. Physical interactions were determined by promoter capture HI-C in human primary white adipocytes (see Results and Figures 4c and d).

SNP	eQTL Effect Size	FDR P-value
rs1412445	0.323	8.49E-09
rs1412444	0.323	8.49E-09
rs1332329	0.311	4.42E-08
rs2246941	0.311	4.42E-08
rs2246833	0.308	5.89E-08
rs2246828	0.293	2.21E-06
rs1051338	0.28	9.89E-06
rs2243547	0.283	7.97E-06
rs1332328	0.311	4.42E-08

Table S13: LIPA promoter SNP and others in LD are eQTLs LIPA in METSIM.
rs1332327 and 8 other SNPs in LD were identified and tested as cis-eQTLs for *LIPA* .This table lists all eQTL effect sizes (betas) and FDR-adjusted p-values from cis-eQTL analyses (see Methods).

SUPPLEMENTAL MATERIAL AND METHODS

Identification of cis-eQTL and cis-splice-QTL SNPs from METSIM

RNA was collected from abdominal subcutaneous adipose needle biopsy from a subset of unrelated participants (n=335, IBD sharing < 0.2) from the Finnish Metabolic Syndrome in Men (METSIM; n = 10,197) cohort, as described in detail previously[1,2]. The METSIM participants are Finnish males who have a median age of 57 years (range: 45–74 years)[1]. As previously described in Pan et al[3]., the genotype data was generated using the Illumina HumanOmniExpress BeadChip and then imputed with IMPUTE2[4] using phase 1 version 3 of the 1000 Genomes Project as the reference panel. The imputed data was filtered using the threshold of info ≥0.8, MAF ≥5%, and Hardy–Weinberg equilibrium (HWE) p > 0.00001. For the RNA-seq data, we performed a 2-pass alignment against the hg19 human reference genome using STAR[5] and used the uniquely mapped reads for gene expression estimation. We inverse normal transformed the FPKMs and regressed out 22 PEER[6] factors to adjust for hidden confounding factors. The cis-eQTL analysis was performed using Matrix-eQTL[7].To identify the GWAS signals among the cis-eQTLs, the GWAS SNPs were obtained from the NHGRI GWAS Catalog[8].

REFERENCES CITED

1. Laakso, M., Kuusisto, J., Stancakova, A., Kuulasmaa, T., Pajukanta, P., Lusis, A.J., Collins, F.S., Mohlke, K.L., and Boehnke, M. (2017). The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. J Lipid Res 58, 481-493.
2. Stancakova, A., Civelek, M., Saleem, N.K., Soininen, P., Kangas, A.J., Cederberg, H., Paananen, J., Pihlajamaki, J., Bonnycastle, L.L., Morken, M.A., et al. (2012). Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. Diabetes 61, 1895-1902.
3. Pan DZ, e.a. Integration of human adipocyte chromosomal interactions with local adipose gene expression identifies obesity genes beyond GWAS. . Nature Communications, .
4. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5, e1000529.
5. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.
6. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 7, 500-507.
7. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353-1358.
8. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42, D1001-1006.