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

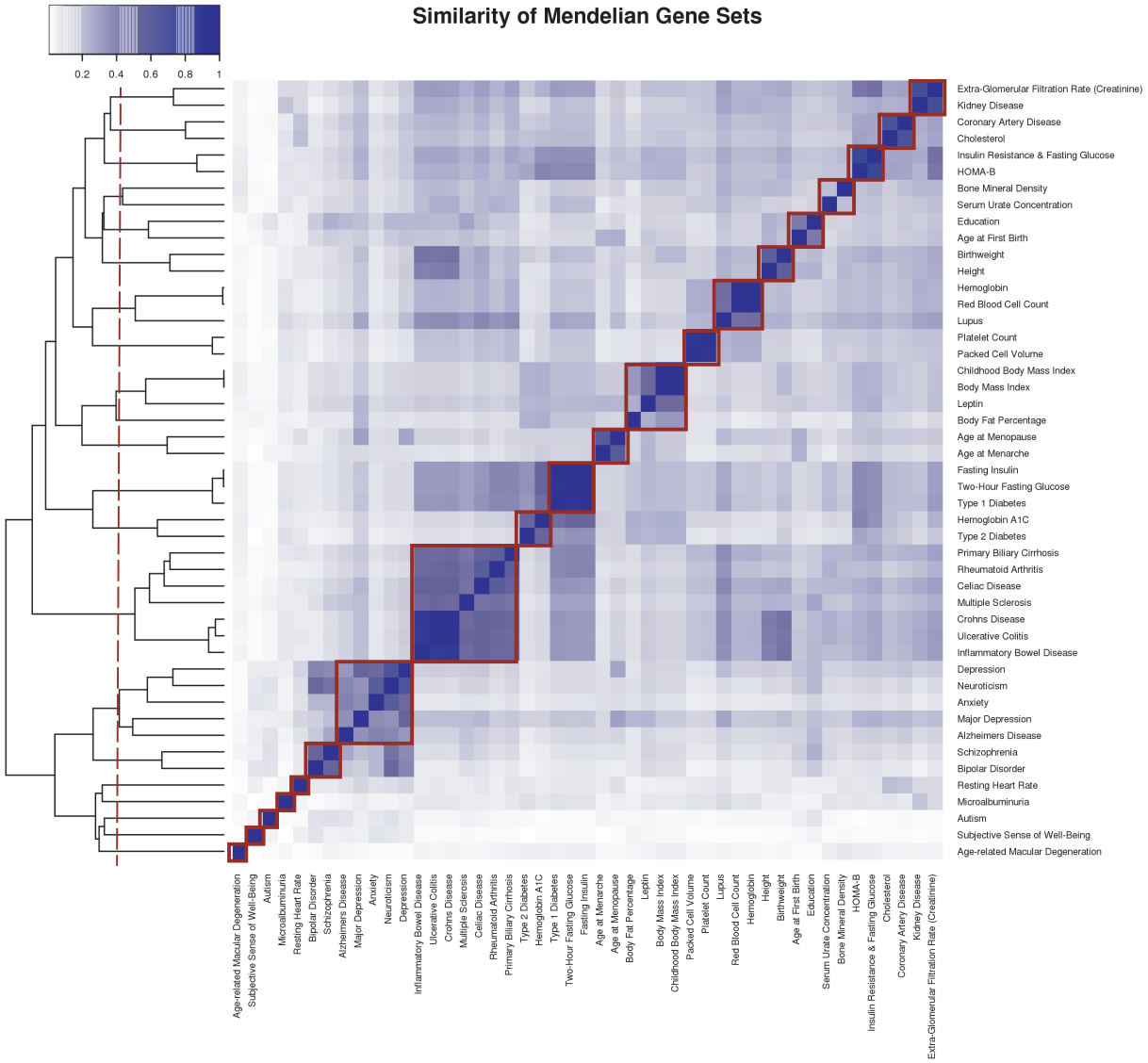

Figure S1: Similarity of Mendelian disorder gene sets

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).

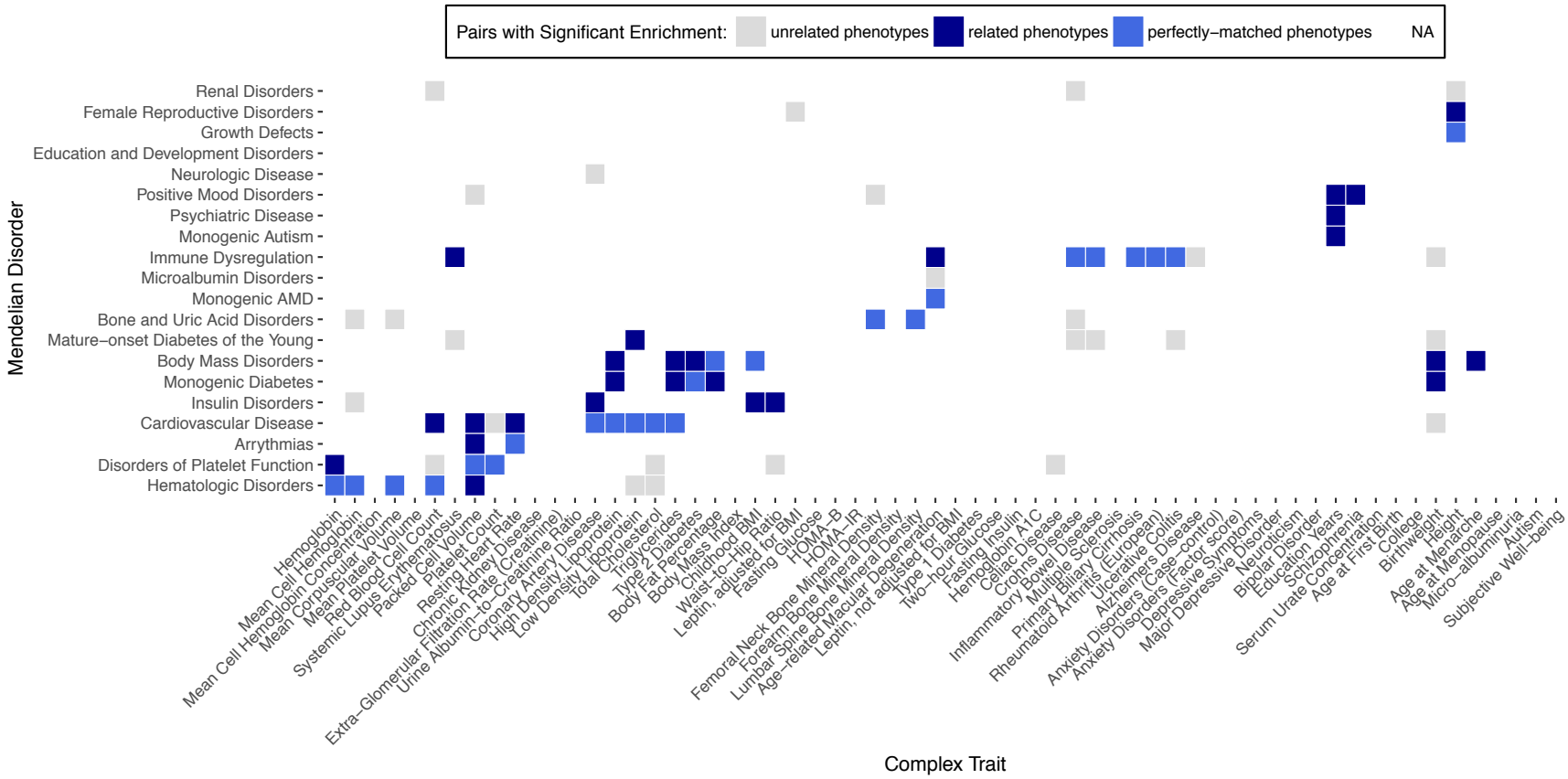

Figure S2: Overlap of GWAS genes with Mendelian disorder genes demonstrates trait-specificity

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.