Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits

Malika Kumar Freund,[1*] Kathryn Burch,[2] Huwenbo Shi,[2] Nicholas Mancuso,[3] Gleb Kichaev,[2] Kristina M. Garske,[1] David Z. Pan,[2] Päivi Pajukanta,[1,2] Bogdan Pasaniuc,[1,2,3,4,5] and Valerie A. Arboleda[3,5**]

[1] Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA
[2] Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles California, 90095, USA
[3] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA
[4] Department of Biomathematics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA

[5]These authors jointly supervised this work.

Correspondence: * kumarm@ucla.edu (MKF), ** varboleda@mednet.ucla.edu (VAA)

ABSTRACT

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 quantify the overlap of genes identified through large-scale genome-wide association studies (GWAS) for 62 complex traits and diseases with genes known to cause 20 broad categories of Mendelian disorders. We identify a significant enrichment of phenotypically-matched Mendelian disorder genes in GWAS gene sets. Further, we observe elevated GWAS effect sizes near phenotypically-matched Mendelian disorder genes. Finally, we report examples of GWAS variants localized at the transcription start site or physically interacting with the promoters of phenotypically-matched Mendelian disorder genes. Our results are consistent with the hypothesis that genes that are disrupted in Mendelian disorders are dysregulated by noncoding 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 monogenic diseases (e.g., cystic fibrosis[2]), and the relative absence of such mutations in complex diseases such as diabetes and schizophrenia[3]. Evidence is accumulating that these two classes of phenotypes may 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[5-8] and their molecular phenotypes[9-12], and Mendelian disorders have also been found to be affected by multiple or common genetic variants [13-16]. 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 to identify genes relevant to complex traits for further study[17-19]. Furthermore, overlap has been identified between genes linked with Mendelian disorders and genetic determinants of complex traits and diseases such as Parkinson's disease[20; 21], obesity[22], height[23], ototoxicity[24], and others[25]. However, the overlap of each of these complex traits with 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 that association signals from genome-wide association studies (GWAS) for complex diseases were enriched in genomic regions with known roles in comorbid Mendelian disorders, suggesting a shared genetic basis[26]. However, the study focuses on Mendelian disorders comorbid with complex diseases in the same individual, rather than Mendelian disorders demonstrating similar phenotypes 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 Mendelian disorder genes, allowing for greater resolution in identifying gene-phenotype relationships. However, to the best of our knowledge, no study has taken advantage of this to identify genes causing 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 genome-wide association studies for complex traits and diseases have identified significant associations in non-coding genomic regions[27], 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, resulting in severe phenotypes (i.e., Mendelian disorders), while non-coding variants produce complex traits by collectively dysregulating expression of these same genes, allowing for nuanced or tissue-specific phenotypes. Based on 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 define "Mendelian disorder genes" as any genes linked to Mendelian disorders in the Online Mendelian Inheritance in Man (OMIM) database[28], and use the well-curated phenotypic breakdown of Mendelian disorders 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 GWAS across 62 complex traits (detailed in **Table 1**) to identify risk genes (here called "GWAS gene sets") for each complex trait, and quantified the overlap of each GWAS gene set

with 20 other sets of Mendelian disorder genes for particular phenotypes (detailed in **Table 1**). We find a consistent, significant, and specific enrichment between GWAS gene sets for complex traits and Mendelian disorder genes for matched and related phenotypes (51/1,240 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 observe instances of enrichments between GWAS gene sets for certain complex traits and Mendelian disorder genes for unrelated phenotypes (20/1,240 pairs; e.g., systemic lupus erythematosus and mature-onset diabetes of the young), suggestive of shared biological mechanisms yet to be examined. Furthermore, we find an increase in average effect size of GWAS variants near Mendelian disorder genes for matched phenotypes, and identify 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 novel examples of significant body mass index (BMI)-associated variants directly interacting with phenotypically-related Mendelian disorder genes *CREBBP* and *CYP19A1*, using human primary white adipocyte-specific Hi-C data[29]. 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 (NCBI build 37/hg19, UCSC Genes track) from the UCSC Table Browser[30] (see Web Resources) using 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. Since 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 HUGO Gene Nomenclature Committee at the European Bioinformatics Institute (HGNC)[31] (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 loss-of-function (LOF) intolerant genes

To identify Mendelian disorder genes, we downloaded the Online Mendelian Inheritance in Man (OMIM) catalogue database[28] 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 any gene with greater than 90% probability of being loss-of-function intolerant, according to the pLI score (pLI > 0.9) from the Exome Aggregation Consortium (ExAC)[32]; 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 the OMIM database via API for all Mendelian disorders demonstrating these phenotypes, then extracted the gene(s)

linked to each Mendelian disorder. We restricted gene-phenotype associations to those satisfying the same criteria (1) and (2) as 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 the ClinVar database[33]. We include a full list of phenotype-specific Mendelian disorder gene sets and clinical phenotype terms used in **Table S1**.

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 based on pairwise overlap, and intersected gene sets visually clustering together to create a single gene set for the representative group of Mendelian disorders. After combining similar gene sets, a total of 20 non-disjoint phenotype-specific Mendelian disorder gene sets remained with an average of 375 genes per set; we include a description of each cluster in **Table S2**.

Complex trait gene sets

We downloaded publicly available summary statistics (per-allele SNP effect sizes, or log-odds ratios for case–control traits, with standard errors[34]) for large-scale GWAS of 62 traits[26] (**Table 1**; average N=83,170; some GWAS were imputed using the 1000 Genomes Project as a reference panel by their respective consortia while others were not.) For each trait, we identified a gene set by mapping each autosomal genome-wide significant SNP ($p < 5\text{x}10^{-8}$) to the closest up- and downstream protein-coding genes as defined above, resulting in a total of 62 non-disjoint GWAS gene sets. As GWAS regions often contain multiple genome-wide significant SNPs, and the true causal gene may not lie adjacent to the lead SNP in a region[35; 36],

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.

Quantifying overlap between complex trait and Mendelian disorder

For each complex trait-Mendelian disorder pair, we compared the GWAS gene set and phenotype-specific Mendelian disorder gene set using a 2x2 contingency table (counting whether each gene was in the GWAS gene set or not, and in the Mendelian disorder gene set or not), with the set of autosomal protein-coding genes (n=17,695) representing the total sample. We used Fisher's Exact Test[37] to determine significance. Phenotype-specificity of overlap significance was assessed by comparing the GWAS gene sets for each complex trait (n=62) to all phenotype-specific Mendelian disorder gene sets (n=20), a total of 1,240 pairs. Significance was assessed at a trait-specific Bonferroni-corrected threshold ($p < 0.05/20$) based on the number of phenotype-specific Mendelian disorder gene sets (n=20) being compared to each GWAS gene set. Since not all autosomal protein-coding genes are Mendelian, we also assessed significance in permutations; we drew 10,000 random sets of Mendelian disorder genes and 10,000 random sets of any protein-coding genes with replacement (matching the size of each random gene set to the size of the phenotype-specific Mendelian disorder gene set), and counted the number of times more genes were shared between the GWAS gene set and a random set.

To assess the robustness and stability of our SNP-gene mapping approach for complex traits, we performed an overlap quantification with phenotype-specific Mendelian disorder genes using GWAS gene sets derived from two additional SNP-gene mapping methods: by mapping each SNP to all genes within a 50Mb window, and to all genes within a 500Mb window. Comparison of the odds ratios produced by Fisher's Exact Test for the comparisons of GWAS gene sets (derived by each mapping method) and phenotype-specific Mendelian disorder gene sets demonstrates no major difference in outcomes from different mapping methods (**Table S3**).

Estimating enrichment of GWAS SNP association signal

We created genomic annotations to capture the regions spanning 50kb upstream through 50kb 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 complex trait-gene category pair, we computed enrichment of GWAS signal within the category c with respect to the set of all protein-coding genes as

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

where N_c is the number of genes in category c, M_j is the number of SNPs within 50kb of gene j, Z_i = GWAS effect size of SNP i divided by standard error, with total number of protein-coding genes N_p. Thus, a_c is the enrichment in average SNP effect size (Z^2) per gene in category (compared to average Z^2 for any protein-coding gene). The percent increase in average SNP effect size per gene for category c, or $(a_c - 1) * 100$, 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 S4**).

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[38] reflecting the amount of LD tagged by each SNP in the HapMap 3 reference panel[39]. For each gene category, we averaged the LD scores of SNPs falling within 50kb of each gene, and found no consistent difference across gene sets. Similar analyses were performed to examine average MAF per gene and average gene length per category across each complex trait (**Table S5**); we again found no consistent difference across gene sets.

193

194 *Putative causal mechanisms at GWAS risk regions*

195 We performed statistical fine-mapping of the genome-wide significant regions ($p < 5 \times 10^{-8}$) for
196 each GWAS using fgwas[40] with no functional annotations and default parameter settings. For
197 each GWAS, we constructed a 95% credible set (defined as the minimum set of SNPs where
198 95% of the probability of causation at a region is accumulated) for each 5kb region (as default
199 assigned by fgwas) containing a significant GWAS association. We achieved this by adding
200 SNPs one at a time with a decreasing posterior probability of causation (posterior probability of
201 association for the SNP, conditioned on there being an association in the region) until a
202 cumulative 95% probability of causation is reached.

203

204 *Identification of candidate regulatory variants*

205 We intersected credible sets for each complex trait with genomic regions 1kb upstream of each
206 phenotypically-relevant Mendelian disorder gene to identify SNPs localizing at the TSS. To
207 identify candidate regulatory variants interacting with promoters of phenotype-matched
208 Mendelian disorder genes, we used interactions from promoter capture Hi-C in human primary
209 white adipocytes [29] for each complex trait, and filtered interactions to pairs of interacting regions
210 where at least one region contained a promoter of a phenotype-specific Mendelian disorder
211 gene. We then intersected interaction pairs for each of these regions with credible sets for each
212 complex trait to identify credible SNPs interacting with regions containing promoters of
213 phenotype-specific Mendelian disorder genes.

214

215 **RESULTS**

216 *GWAS risk genes show specific, significant overlap with phenotypically-matched Mendelian*
217 *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 GWAS. 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 known to cause those specific phenotypes from the OMIM database[28] (see 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 Methods, and **Figure 1**). Overlap between each phenotype-specific Mendelian disorder gene set (n=20) and each GWAS gene set (n=62) was assessed using Fisher's Exact Test, for a total of 1,240 comparisons (**Table 1** and **Table S6**). We hypothesized that GWAS gene sets would have a specific significant enrichment of Mendelian disorder genes for matched Mendelian disorders, but no enrichment for unrelated Mendelian disorders. Among all 1,240 pairs of complex and phenotype-specific Mendelian disorder gene sets assessed, we identified 71 pairs with significant overlap at a $p < 0.05/20$ (**Figure 2**). An examination of the log-odds ratios for each overlap comparison revealed more extreme enrichments among phenotypically-matched pairs compared to phenotypically-unmatched pairs (**Table 2**), which is consistent with our hypothesis. 51 out of the 71 significantly overlapping pairs showed perfectly matching phenotypes or reflected known shared biology. 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 (AMD) and cholesterol traits (high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC), and triglycerides (TG))[6; 41-44]. We confirmed significant enrichment after trait-specific Bonferroni adjustment (20 comparisons, see Methods) between many of these previously reported pairs such as the complex and monogenic forms of height[45] (OR=1.39, p=0.029) and HDL and Mendelian forms of cardiovascular disease[46]

(OR=2.10, p=0.006). We also identified previously unreported enrichments; for example, we find a strong enrichment between inflammatory bowel disease (IBD) and Mendelian forms of immune dysregulation (OR=3.32, p=0.01) and between hemoglobin (HB) and Mendelian hematologic disorders (OR=3.99, p=0.009). The remaining 20 pairs with significant overlap suggested novel shared biological mechanisms between complex traits and Mendelian disorders yet to be established (**Table 3**). For example, we observed an enrichment between height and renal disorders (OR=1.48, p=0.001), and enrichment between Crohn's Disease and mature-onset diabetes of the young (OR=2.69, p=0.005). Of note, the vast majority of complex and Mendelian disorder gene set pairs without perfectly-matched phenotypes (n=1,158 of 1,178) did not demonstrate any significant overlap (**Table S6**).

SNPs near phenotypically-matched Mendelian disorder genes show increased effect size on complex traits

Because Mendelian disorder genes exhibit 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 a non-Mendelian disease gene. We hypothesized that SNPs near these phenotype-specific Mendelian disorder genes have further increased effects on complex traits due to 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 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 S4**). This suggests that the genomic regions containing the most biologically-relevant genes for each trait contribute most significantly to

complex trait biology. We also confirmed that loss-of-function (LOF) intolerant genes (as defined by ExAC's pLI score > 0.9, see Methods) demonstrate a higher average effect size across most complex traits examined[32]. Given the extreme intolerance of deleterious mutations in these genes, it is possible that LOF-intolerant genes demonstrate embryonic lethal mutant phenotypes, and are thus undiscovered as Mendelian disorder genes at this time. We found no significant increase in linkage disequilibrium or decrease in average minor allele frequency (MAF) of the SNPs within each category compared to the SNPs within all protein-coding genes (**Table S5)**, suggesting that the observed signal is not driven by any of these confounders. Of note, we did observe a respective increase in average gene length between all protein-coding genes, all Mendelian disorder genes, and LOF-intolerant genes (**Table S5**). Therefore, it is possible that our findings of enriched GWAS signal in these gene categories is due instead to longer genes being more likely to tag causal variation. However, in general we did not observe a significant increase in average gene length for the phenotype-specific Mendelian disorder gene sets as compared to all Mendelian disorder genes (**Table S5**), but still found an increase in enrichment of GWAS signal (**Figure 3**), suggesting that gene length is not significantly confounding our results. The only exceptions to this are the phenotype-specific Mendelian disorder gene sets for neurological phenotypes, for which the average gene length was increased compared to all Mendelian disorder genes; this is consistent with other reported findings about gene length in neurological traits[47].

Examples of credible SNPs for GWAS regions near phenotypically-matched Mendelian disorder genes

We next sought to identify common non-coding variants that may causally impact 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 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 S7 and S8**), and identified 25 promising candidate SNPs (attaining genome-wide significance in GWAS) at TSSs that could be regulating the proximal Mendelian disorder gene (**Table 4**). We highlight two examples: first, we found a significantly associated SNP from the credible set for coronary artery disease (rs1332327, Z=6.798) at the 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) causing hypercholesterolemia and hypertriglyceridemia as part of cholesteryl ester- and triglyceride-filled macrophage infiltration syndromes (**Figure 4A**). Second, from the credible set for red blood cell count, we found a significantly associated SNP (rs1010222, Z= -5.961) at the promoter of *CALR* (MIM# 109091), a Mendelian disorder gene known to cause Myelofibrosis (MIM# 254450) involving generalized bone marrow fibrosis, reduced hemopoiesis, no 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 by 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 the functional interpretation of GWAS variants and how they might regulate Mendelian disorder genes. Examination of chromatin interactions in human primary white adipocytes[29] revealed further candidate credible set SNPs for metabolic traits physically interacting with promoters of phenotypically-relevant Mendelian disorder genes (**Table S9)**.

Specifically, we report that a genome-wide significant SNP for BMI (rs758747, Z=6.081) physically interacts with the promoter of *CREBBP*, a gene known to cause Rubinstein-Taybi Syndrome 1 (MIM# 180849) in which obesity is one of the syndromic features[28] (**Figure 4C**). These interactions can also identify the relevant isoforms of genes in disease. We identified a cluster of SNPs from the credible set of variants associated with BMI that physically interact with the promoter of a specific isoform of *CYP19A1*, a gene known to cause Aromatase Excess Syndrome (MIM# 139300) involving short stature and excess fat storage in the chest (gynecomastia)[28] **(Figure 4D)**. Although longer isoforms of *CYP19A1* are by default chosen to represent the gene, our data suggests that the shorter isoform is likely to be more relevant in obesity. Taken 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 specific enrichment of phenotypically-matched and related Mendelian disorder genes in GWAS regions for complex traits; we also identified fewer pairs of complex traits and phenotypically-unmatched Mendelian disorders with similar significant enrichment. We further found that phenotypically-relevant Mendelian disorder genes are enriched for GWAS signal across complex traits, compared to 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 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 the genetic bases (and by extension, biological mechanisms) with their Mendelian forms. This supports our hypothesis that the same set of genes generally underlie 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 GWAS become better-powered, we anticipate being able to identify phenotype-specific enrichments of Mendelian disorder genes in GWAS regions for more complex traits.

Second, the subset of complex trait-Mendelian disorder pairs with no known shared biology that still demonstrated significant enrichment of Mendelian disorder genes in GWAS regions can offer us novel 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[26]; these findings together suggest that many complex traits and Mendelian disorders may 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 observed across complex traits are consistent with the wide-spread 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[32]; because genes with pLI > 0.9 exhibit extreme intolerance of deleterious mutation, it is possible that these genes demonstrate embryonic 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

375 genes (particularly those enriched for GWAS signal for a particular complex trait) for
376 phenotypically-relevant Mendelian disorders.

377 Third, linking Mendelian disorder genes with complex traits can help with
378 characterization of the genetic architecture of complex traits – specifically, with genes and
379 pathways that can be functionally characterized to identify molecular mechanisms[6]. Identifying
380 causal variants from large-scale GWAS studies is particularly challenging given that most
381 GWAS loci lie in non-coding regions of the genome; though thousands of genomic loci have
382 been significantly associated with specific diseases, few casual SNPs have been functionally
383 verified[48; 49]. Although many approaches have been used to tie a particular variant to a causal
384 gene or genes[50-52], including newer methods that directly link gene expression to a trait (e.g.,
385 TWAS[35], PrediXcan[53]), we find that leveraging GWAS findings with functional data to identify
386 candidate regulatory variants for Mendelian disorder genes can potentially lead to better
387 interpretation of relevant genes and isoforms. Here, we demonstrate the heterogeneity of
388 mechanisms potentially underlying causal variation, showing roles for TSS promoter regions of
389 Mendelian disorder genes and long-range interactions involving significant GWAS regions. We
390 expand on recent work showing that BMI-associated variants interact with genes in GWAS
391 regions to demonstrate similar findings for Mendelian disorder genes[29]. With the appropriate
392 functional data from relevant tissues and cell types, this phenotype-driven approach can identify
393 relevant candidate regulatory variants and their targets. Further, from the perspective of
394 monogenic diseases, identifying common variants that might modify the expressivity of
395 phenotypes can provide novel insights into gene function in addition to putative drug targets.
396 Many drugs approved by the FDA and developed by pharmaceutical companies are targeted
397 towards the treatment of complex traits and diseases; by identifying underlying links between
398 Mendelian disorders and complex traits through their effects on the same biological genes and
399 pathways, we can systematically and rationally target existing drugs for complex traits and

diseases towards those with rare Mendelian disorders which largely do not have any rationally targeted treatments[54-56].

Last, we note that our approach of examining traits and disorders at the component-phenotype level offers us unprecedented resolution into the specific pathways involved the overall trait or disorder. In clinical medicine, genome-wide sequencing has expanded the clinical phenotypic spectrum associated with a gene [57; 58] through identification of pleiotropic effects due to mutations in specific protein domains[59; 60], detected a genetic predisposition for diseases previously considered to be due to environment[13], uncovered variable penetrance for genetic mutations previously thought to be sufficient to cause disease, and has suggested that genetic background influences the phenotypic variability of monogenic diseases[61; 62]. 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[28; 63]. Recent work has demonstrated that incorporation of such dense phenotype information to rank putative disease-causing genetic mutations improves diagnostic rates in clinical exome sequencing tests[64; 65]. However, to our knowledge no studies as of yet have taken advantage of component Mendelian phenotypes to identify Mendelian disorders that may be phenotypically-relevant to a variety of 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

The supplement contains one figure and ten tables.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENTS

We would like to thank Ruth Johnson, Megan Major, Megan Roytman, Claudia Giambartolomei, Arunabha Majumdar, Jazlyn Mooney, Brendan Freund, Robert Brown, 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 MKF; NIH Early Independence Award

DP5OD024579 to VAA; National Institute of Mental Health of the NIH grant T32MH073526 to KB; NIH grant F31HL142180 to KMG; NIH-NCI National Cancer Institute grant T32LM012424 to DZP; and NIH grants HL-095056 and HL-28481 to PP.

WEB RESOURCES

UCSC Table Browser: https://genome.ucsc.edu/cgi-bin/hgTables
HGNC: http://www.genenames.org/cgi-bin/download
Gene sets: https://github.com/bogdanlab/gene_sets
OMIM: https://omim.org/downloads/
ExAC: http://exac.broadinstitute.org/downloads
ClinVar: ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/gene_specific_summary.txt

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. 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.
6. 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, E345-349.
7. Auer, P.L., Teumer, A., Schick, U., O'Shaughnessy, A., Lo, K.S., Chami, N., Carlson, C., de Denus, S., Dube, 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.
8. Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al. (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43, 1066-1073.
9. 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.
10. Sztainberg, Y., and Zoghbi, H.Y. (2016). Lessons learned from studying syndromic autism spectrum disorders. Nat Neurosci 19, 1408-1417.
11. 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.
12. 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.
13. 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. The New England journal of medicine 376, 21-31.
14. Corvol, H., Blackman, S.M., Boelle, 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.
15. 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. (2012). Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet 44, 886-889.

16. 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.
17. Dron, J.S., and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. Curr Atheroscler Rep 19, 31.
18. Hegele, R.A. (2018). Learning From Patients With Ultrarare Conditions: Cholesterol Hoof Beats. J Am Coll Cardiol 71, 289-291.
19. Peltonen, L., Perola, M., Naukkarinen, J., and Palotie, A. (2006). Lessons from studying monogenic disease for common disease. Hum Mol Genet 15 Spec No 1, R67-74.
20. Petrucci, S., Consoli, F., and Valente, E.M. (2014). Parkinson Disease Genetics: A "Continuum" from Mendelian to Multifactorial Inheritance. Curr Mol Med 14, 1079-1088.
21. 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.
22. Lim, E.T., Liu, Y.P., Chan, Y., Tiinamaija, T., Karajamaki, A., Madsen, E., Go, T.D.C., Altshuler, D.M., Raychaudhuri, S., Groop, L., et al. (2014). A novel test for recessive contributions to complex diseases implicates Bardet-Biedl syndrome gene BBS10 in idiopathic type 2 diabetes and obesity. American journal of human genetics 95, 509-520.
23. Chan, Y., Salem, R.M., Hsu, Y.H., McMahon, G., Pers, T.H., Vedantam, S., Esko, T., Guo, M.H., Lim, E.T., Consortium, G., et al. (2015). Genome-wide Analysis of Body Proportion Classifies Height-Associated Variants by Mechanism of Action and Implicates Genes Important for Skeletal Development. American journal of human genetics 96, 695-708.
24. 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.
25. Amininejad, L., Charloteaux, B., Theatre, E., Liefferinckx, C., Dmitrieva, J., Hayard, P., Muls, V., Maisin, J.M., Schapira, M., Ghislain, J.M., et al. (2018). Analysis of Genes Associated with Monogenic Primary Immunodeficiency Identifies Rare Variants in XIAP in Patients With Crohn's disease. Gastroenterology.
26. 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.
27. 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.
28. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM
Number: {180849}: {01/18/2018}: Retrieved from http://www.ncbi.nlm.nih.gov/omim/
. In. (
29. 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.
30. 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, D493-496.
31. HUGO Gene Nomenclature Committee at the European Bioinformatics Institute. In. (
32. 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. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.

33. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46, D1062-D1067.
34. Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet 18, 117-127.
35. 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. American journal of human genetics 100, 473-487.
36. 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.
37. Fisher, R.A. (1922). On the interpretation of χ2 from contingency tables, and the calculation of P. Journal of the Royal Statistical Society 85.
38. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.
39. International HapMap, C., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature 467, 52-58.
40. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. American journal of human genetics 94, 559-573.
41. 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. (2013). Seven new loci associated with age-related macular degeneration. Nat Genet 45, 433-439, 439e431-432.
42. 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.
43. 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.
44. Kanoni, S., Masca, N.G., Stirrups, K.E., Varga, T.V., Warren, H.R., Scott, R.A., Southam, L., Zhang, W., Yaghootkar, H., Muller-Nurasyid, M., et al. (2016). Analysis with the exome array identifies multiple new independent variants in lipid loci. Hum Mol Genet 25, 4094-4106.
45. 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.
46. Rosenson, R.S., Brewer, H.B., Jr., Barter, P.J., Bjorkegren, 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.
47. Zylka, M.J., Simon, J.M., and Philpot, B.D. (2015). Gene length matters in neurons. Neuron 86, 353-355.
48. 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. (2016). Schizophrenia risk from complex variation of complement component 4. Nature 530, 177-183.
49. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, 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.

50. 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.
51. 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.
52. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8, 1826.
53. Wang, J., Gamazon, E.R., Pierce, B.L., Stranger, B.E., Im, H.K., Gibbons, R.D., Cox, N.J., Nicolae, D.L., and Chen, L.S. (2016). Imputing Gene Expression in Uncollected Tissues Within and Beyond GTEx. American journal of human genetics 98, 697-708.
54. 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.
55. Pardoll, D.M. (2012). Immunology beats cancer: a blueprint for successful translation. Nat Immunol 13, 1129-1132.
56. Gerich, M.E., and McGovern, D.P. (2014). Towards personalized care in IBD. Nat Rev Gastroenterol Hepatol 11, 287-299.
57. 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. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. American journal of human genetics 97, 199-215.
58. Reuter, C.M., Brimble, E., DeFilippo, C., Dries, A.M., Undiagnosed Diseases, N., Enns, G.M., Ashley, E.A., Bernstein, J.A., Fisher, P.G., and Wheeler, M.T. (2018). A New Approach to Rare Diseases of Children: The Undiagnosed Diseases Network. J Pediatr.
59. Goudie, D.R., D'Alessandro, M., Merriman, B., Lee, H., Szeverenyi, 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.
60. Arboleda, V.A., Lee, H., Parnaik, R., Fleming, A., Banerjee, A., Ferraz-de-Souza, B., Delot, E.C., Rodriguez-Fernandez, I.A., Braslavsky, D., Bergada, I., et al. (2012). Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. Nat Genet 44, 788-792.
61. 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.
62. Hensman Moss, D.J., Pardinas, A.F., Langbehn, D., Lo, K., Leavitt, B.R., Roos, R., Durr, A., Mead, S., investigators, T.-H., investigators, R., et al. (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol 16, 701-711.
63. Kohler, 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, D966-974.
64. Yang, H., Robinson, P.N., and Wang, K. (2015). Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Nat Methods 12, 841-843.
65. Zemojtel, T., Kohler, S., Mackenroth, L., Jager, 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.

643 66. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A.,
644 Heap, G.A., Adany, R., Aromaa, A., et al. (2010). Multiple common variants for celiac
645 disease influencing immune gene expression. Nat Genet 42, 295-302.
646 67. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee,
647 J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility
648 loci for inflammatory bowel disease and highlight shared genetic risk across populations.
649 Nat Genet 47, 979-986.
650 68. Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfield, G.M., Greene, C.S., Xie, G., Juran,
651 B.D., Zhu, D., Qian, D.C., et al. (2015). International genome-wide meta-analysis
652 identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat
653 Commun 6, 8019.
654 69. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki,
655 A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and
656 drug discovery. Nature 506, 376-381.
657 70. International Multiple Sclerosis Genetics, C., Wellcome Trust Case Control, C., Sawcer, S.,
658 Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A.,
659 Su, Z., et al. (2011). Genetic risk and a primary role for cell-mediated immune
660 mechanisms in multiple sclerosis. Nature 476, 214-219.
661 71. Consortium., A.S.D.W.G.o.t.P.G. (2015). PGC-ASD summary statistics from a meta-analysis
662 of 5,305 ASD-diagnosed cases and 5,305 pseudocontrols of European descent (based
663 on similarity to CEPH reference genotypes). In. (http://www.med.unc.edu/pgc/results-
664 anddownloads).
665 72. van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul,
666 D.S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the
667 human red blood cell. Nature 492, 369-375.
668 73. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic,
669 J., Elling, U., Goodall, A.H., Labrune, Y., et al. (2011). New gene functions in
670 megakaryopoiesis and platelet formation. Nature 480, 201-208.
671 74. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tombleson, P., Behrens, T.W.,
672 Martin, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association
673 analyses implicate aberrant regulation of innate and adaptive immunity genes in the
674 pathogenesis of systemic lupus erythematosus. Nat Genet 47, 1457-1464.
675 75. Horikoshi, M., Beaumont, R.N., Day, F.R., Warrington, N.M., Kooijman, M.N., Fernandez-
676 Tajes, J., Feenstra, B., van Zuydam, N.R., Gaulton, K.J., Grarup, N., et al. (2016).
677 Genome-wide associations for birth weight and correlations with adult disease. Nature
678 538, 248-252.
679 76. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y.,
680 Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in
681 the genomic and biological architecture of adult human height. Nat Genet 46, 1173-
682 1186.
683 77. Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L.,
684 Park-Min, K.H., Tobias, J.H., Kooperberg, C., et al. (2015). Whole-genome sequencing
685 identifies EN1 as a determinant of bone density and fracture. Nature 526, 112-117.
686 78. Kottgen, A., Albrecht, E., Teumer, A., Vitart, V., Krumsiek, J., Hundertmark, C., Pistis, G.,
687 Ruggiero, D., O'Seaghdha, C.M., Haller, T., et al. (2013). Genome-wide association
688 analyses identify 18 new loci associated with serum urate concentrations. Nat Genet 45,
689 145-154.
690 79. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D.,
691 Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000
692 Genomes-based genome-wide association meta-analysis of coronary artery disease.
693 Nat Genet 47, 1121-1130.

80. 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. (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet 45, 1274-1283.
81. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., et al. (2010). Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes 59, 3229-3239.
82. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44, 981-990.
83. 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.
84. 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. (2014). Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature 514, 92-97.
85. 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. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nat Genet 47, 1294-1303.
86. 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. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42, 105-116.
87. 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. (2016). Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. Diabetes 65, 803-817.
88. 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. (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet 42, 142-148.
89. 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. (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.
90. 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. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45, 1452-1458.
91. 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.
92. Major Depressive Disorder Working Group of the Psychiatric, G.C., 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., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 18, 497-511.
93. Okbay, A., Baselmans, B.M., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F., Linner, R.K., Rietveld, C.A., Derringer, J., et al. (2016). Genetic variants

744 associated with subjective well-being, depressive symptoms, and neuroticism identified
745 through genome-wide analyses. Nat Genet 48, 624-633.
746 94. Psychiatric, G.C.B.D.W.G. (2011). Large-scale genome-wide association analysis of bipolar
747 disorder identifies a new susceptibility locus near ODZ4. Nat Genet 43, 977-983.
748 95. Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights
749 from 108 schizophrenia-associated genetic loci. Nature 511, 421-427.
750 96. Pattaro, C., Teumer, A., Gorski, M., Chu, A.Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A.,
751 Sorice, R., Li, Y., et al. (2016). Genetic associations at 53 loci highlight cell types and
752 biological pathways relevant for kidney function. Nat Commun 7, 10023.
753 97. Eppinga, R.N., Hagemeijer, Y., Burgess, S., Hinds, D.A., Stefansson, K., Gudbjartsson,
754 D.F., van Veldhuisen, D.J., Munroe, P.B., Verweij, N., and van der Harst, P. (2016).
755 Identification of genomic loci associated with resting heart rate and shared genetic
756 predictors with all-cause mortality. Nat Genet 48, 1557-1563.
757 98. Barban, N., Jansen, R., de Vlaming, R., Vaez, A., Mandemakers, J.J., Tropf, F.C., Shen, X.,
758 Wilson, J.F., Chasman, D.I., Nolte, I.M., et al. (2016). Genome-wide analysis identifies
759 12 loci influencing human reproductive behavior. Nat Genet 48, 1462-1472.
760 99. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J.,
761 Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). GWAS of 126,559 individuals
762 identifies genetic variants associated with educational attainment. Science 340, 1467-
763 1471.
764 100. Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley,
765 P., Chen, G.B., Emilsson, V., Meddens, S.F., et al. (2016). Genome-wide association
766 study identifies 74 loci associated with educational attainment. Nature 533, 539-542.
767 101. Lu, Y., Day, F.R., Gustafsson, S., Buchkovich, M.L., Na, J., Bataille, V., Cousminer, D.L.,
768 Dastani, Z., Drong, A.W., Esko, T., et al. (2016). New loci for body fat percentage reveal
769 link between adiposity and cardiometabolic disease risk. Nat Commun 7, 10495.
770 102. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C.,
771 Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass
772 index yield new insights for obesity biology. Nature 518, 197-206.
773 103. Felix, J.F., Bradfield, J.P., Monnereau, C., van der Valk, R.J., Stergiakouli, E., Chesi, A.,
774 Gaillard, R., Feenstra, B., Thiering, E., Kreiner-Moller, E., et al. (2016). Genome-wide
775 association analysis identifies three new susceptibility loci for childhood body mass
776 index. Hum Mol Genet 25, 389-403.
777 104. Kilpelainen, T.O., Carli, J.F., Skowronski, A.A., Sun, Q., Kriebel, J., Feitosa, M.F., Hedman,
778 A.K., Drong, A.W., Hayes, J.E., Zhao, J., et al. (2016). Genome-wide meta-analysis
779 uncovers novel loci influencing circulating leptin levels. Nat Commun 7, 10494.
780 105. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R.,
781 Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al. (2015). New genetic loci
782 link adipose and insulin biology to body fat distribution. Nature 518, 187-196.
783
784

FIGURE TITLES AND LEGENDS:

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, short stature; **Table S9**). 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 gene 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 causing any of those disorders.

Figure 2: Overlap of GWAS genes with unrelated Mendelian disorder genes demonstrates trait-specificity. Significant overlaps from phenotypically-matched pairs of complex traits and Mendelian disorders (blue) and pairs with unrelated phenotypes (grey) are shown. Phenotypically-matched pairs are subdivided into pairs with perfectly-matched phenotypes (light blue) and pairs with related phenotypes (dark blue). Traits with no significant overlaps are excluded. Significance was assessed at a threshold for each complex trait ($p < 0.05/20$) based on the number of Mendelian disorder gene sets (n=20) compared to each GWAS gene set.

Figure 3: Effect sizes for SNPs on complex traits from GWAS are increased for genes that are loss-of-function intolerant and for phenotypically-relevant Mendelian disorder genes.

The increase in average SNP effect size per gene across four gene categories. We averaged effect size (Z^2) across all SNPs falling within 50kb 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 box plots represent the distribution of increase in average effect size per gene across all traits.

Figure 4: Candidate regulatory SNPs fall at transcription start sites and long-range promoters of phenotypically-relevant Mendelian disorder genes
A, 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 = -5.961), lies at the TSS of *LIPA*. **B**) Putative causal SNP rs1010222, associated with red blood cell count with a Z-score of -5.961, lies at the TSS of *CALR*. **C, 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 = 6.081), physically interacts with the promoter of a distant gene *CREBBP*.

827 **TABLES:**

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

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[66]	CEL	15283	13	54	34	Immune Dysregulation
Crohn's Disease[67]	CD	27726	231 total	4381	239	
Inflammatory Bowel Disease[67]	IBD	34694		7738	368	
Ulcerative Colitis[67]	UC	28738		4239	202	
Primary Biliary Cirrhosis[68]	PBC	13239	28	704	149	
Rheumatoid Arthritis (European)[69]	RA	58284	101	16502	297	
Multiple Sclerosis[70]	MS	27148	52	487	160	
Autism[71]	AUT	10610	2	2	2	Monogenic Autism
Hemoglobin[72]	HB	51255.8	11	325	89	Hematologic Disorders
Mean Cell Hemoglobin[72]	MCH	43553.6	19	1188	164	
Mean Cell Hemoglobin Concentration[72]	MCHC	46953.9	8	8	12	
Mean Corpuscular Volume[72]	MCV	48472.8	23	1237	180	
Mean Platelet Volume[73]	MPV	16843	25	705	102	
Red Blood Cell Count[72]	RBC	45304.4	10	908	107	
Systemic Lupus Erythematosus[74]	SLE	23210	43	4983	286	
Birthweight[75]	BW	110054.6	60	1978	179	Growth Defects
Height[76]	HGT	239338.3	423	22807	2361	
Femoral Neck Bone Mineral	FN	53236	14	788	58	Bone and Uric Acid Disorders

Density[77]						
Forearm Bone Mineral Density[77]	FA	53236	3	136	8	
Lumbar Spine Bone Mineral Density[77]	LS	53236	19	998	67	
Serum Urate Concentration[78]	URT	107026	28	1991	161	
Packed Cell Volume[72]	PCV	44925.9	4	141	53	Disorders of Platelet Function
Platelet Count[73]	PLT	66867	43	801	134	
Coronary Artery Disease[79]	CAD	184305	46	1709	132	Cardiovascular Disease
High Density Lipoprotein[80]	HDL	95422	157 total	3131	464	
Low Density Lipoprotein[80]	LDL	90686.4		2796	370	
Total Cholesterol[80]	TC	95651.5		3803	500	
Triglycerides[80]	TG	91882.3		2965	354	
Hemoglobin A1C[81]	HBA	46368	10	174	33	Monogenic Diabetes
Type 2 Diabetes[82]	T2D	61857.4	10	191	28	
Age-related Macular Degeneration[83]	AMD	33975.1	34	4087	215	Monogenic AMD
Age at Menarche[84]	MNR	182416	106	2011	207	Female Reproductive Disorders
Age at Menopause[85]	MNP	70000	44	1656	316	
Fasting Glucose[86]	FG	46186	16 total	231	39	Insulin Disorders
HOMA-B[86]	HMB	46186		95	12	
HOMA-IR[86]	HMIR	46186	1	1	0	
Micro-albuminuria[87]	MA	52988.4	1	4	2	Microalbumin Disorders
Fasting Insulin[86]	FI	108557	1	43	23	Mature-onset Diabetes of the Young
Two-hour Glucose[88]	2HG	15234	3	4	2	
Type 1 Diabetes[89]	T1D	24341	42	1447	144	
Alzheimer's Disease[90]	ALZ	54162	19	813	58	Neurologic Disease
Anxiety Disorders (Case-control)[91]	ANXC	14643.1	1	8	2	
Anxiety Disorders	ANXF	16218.2	1	43	3	

(Factor score)[91]						
Major Depressive Disorder[92]	MDD	10610	1	5	4	
Depressive Symptoms[93]	DS	161460	2	28	10	
Neuroticism[93]	NRT	170911	11	1933	82	
Bipolar Disorder[94]	BIP	16731	4	39	8	Psychiatric Disease
Schizophrenia[95]	SCZ	150064	108	6808	479	
Chronic Kidney Disease[96]	CKD	117340	4	107	16	Renal Disorders
Glomerular Filtration Rate (CRN)[96]	EGFR	132725.4	43	1401	162	
Urine Albumin-to-Creatinine Ratio[87]	UACR	53343.1	1	4	2	
Resting Heart Rate[97]	RHR	265046	64	4568	304	Arrythmias
Age at First Birth[98]	AFB	251151	12	238	45	Education and Development Disorders
College[99]	COL	126559	3	71	12	
Education Years[100]	EY	293723	74	6537	554	
Subjective Well-being[93]	SWB	298420	3	37	9	Positive Mood Disorders
Body Fat Percentage[101]	BFP	57721.7	12	196	22	Body Mass Disorders
Body Mass Index[102]	BMI	224996	97	1585	231	
Childhood BMI[103]	CBMI	28964.1	15	438	49	
Leptin, adjusted for BMI[104]	LEPB	30202	5 total	8	5	
Leptin, not adjusted for BMI[104]	LEP	30507.6		1	0	
Waist-to-Hip Ratio[105]	WHR	139559.2	49	424	74	

Table 2: Overlap of GWAS genes and phenotypically-matched Mendelian disorder genes. For each pair of complex trait and Mendelian disorder, Fisher's exact test was used to quantify the enrichment of shared genes with an odds ratio and p-value (see Methods). Significance was assessed at a threshold of $p < (0.05/20)$ correcting for the number of Mendelian disorder gene sets compared to each complex trait gene set. This table lists pairs of complex traits and phenotypically-matched or related Mendelian disorders with significant overlap. For comparison, the average odds ratio for pairings of each complex trait with all unrelated Mendelian disorder gene sets is included.

Complex	Matched or Related	Number of	Odds Ratio	P-value	Average Odds

Trait	Mendelian Disorder(s)	Shared Genes	for Matched Pair (CI)			Ratio across Unmatched Pairs
AMD	Monogenic AMD	9	7.99 (3.50, 16.11)	9.88E-05	*	2.06
AMD	Immune Dysregulation	17	2.73 (1.55, 4.52)	8.27E-03	*	
BFP	Body Mass Disorders	3	22.14 (4.14, 76.70)	1.03E-02	*	4.24
BFP	Monogenic Diabetes	3	15.42 (2.90, 53.04)	2.85E-02	*	
BW	Mature-onset Diabetes of the Young	16	3.06 (1.69, 5.16)	3.81E-03	*	1.99
BW	Body Mass Disorders	6	4.94 (1.76, 11.29)	3.86E-02	*	
CAD	Cardiovascular Disease	13	3.17 (1.63, 5.67)	1.07E-02	*	2.23
CAD	Insulin Disorders	13	3.04 (1.56, 5.43)	1.57E-02	*	
CBMI	Body Mass Disorders	5	16.18 (4.92, 41.63)	5.41E-04	*	2.65
CBMI	Insulin Disorders	7	4.61 (1.74, 10.41)	3.08E-02	*	
CD	Immune Dysregulation	23	3.42 (2.10, 5.32)	3.39E-05	*	1.65
EY	Positive Mood Disorders	9	4.70 (2.04, 9.59)	5.76E-03	*	1.07
EY	Psychiatric Disease	19	2.45 (1.44, 3.95)	1.79E-02	*	
FN	Bone and Uric Acid Disorders	5	7.64 (2.36, 19.24)	1.52E-02	*	3.98
HB	Disorders of Platelet Function	12	6.21 (3.05, 11.59)	4.34E-05	*	2.66
HB	Hematologic Disorders	10	3.99 (1.83, 7.79)	8.84E-03	*	
HDL	Monogenic Diabetes	15	3.41 (1.85, 5.85)	1.87E-03	*	1.48
HDL	Body Mass Disorders	12	3.92 (1.95, 7.17)	2.78E-03	*	
HDL	Cardiovascular Disease	31	2.10 (1.40, 3.06)	6.89E-03	*	
HGT	Female Reproductive Disorders	61	1.76 (1.30, 2.36)	4.36E-03	*	1.33
HGT	Growth Defects	126	1.39 (1.13, 1.70)	2.85E-02	*	
IBD	Immune Dysregulation	34	3.32 (2.23, 4.79)	3.15E-07	*	1.47
LDL	Cardiovascular Disease	31	2.70 (1.79, 3.95)	6.90E-05	*	1.74

LDL	Mature-onset Diabetes of the Young	24	2.17 (1.36, 3.32)	1.71E-02	*	
LS	Bone and Uric Acid Disorders	6	8.00 (2.80, 18.72)	3.66E-03	*	2.80
MCH	Hematologic Disorders	15	3.19 (1.73, 5.48)	3.85E-03	*	1.71
MCV	Hematologic Disorders	20	4.00 (2.36, 6.44)	1.78E-05	*	1.73
MNR	Body Mass Disorders	7	5.02 (1.95, 10.86)	1.55E-02	*	0.93
PBC	Immune Dysregulation	13	3.03 (1.56, 5.40)	1.55E-02	*	1.17
PCV	Disorders of Platelet Function	10	9.24 (4.11, 18.83)	1.30E-05	*	4.53
PCV	Hematologic Disorders	8	5.60 (2.27, 12.08)	4.33E-03	*	
PCV	Arrythmias	5	6.70 (2.07, 16.94)	2.73E-02	*	
PCV	Cardiovascular Disease	7	4.39 (1.66, 9.84)	3.89E-02	*	
PLT	Disorders of Platelet Function	12	3.91 (1.95, 7.15)	2.80E-03	*	1.55
PLT	Cardiovascular Disease	12	2.85 (1.42, 5.20)	3.95E-02	*	
RA	Immune Dysregulation	25	2.95 (1.86, 4.50)	1.36E-04	*	0.94
RBC	Hematologic Disorders	14	4.78 (2.50, 8.50)	1.16E-04	*	2.93
RBC	Disorders of Platelet Function	12	5.03 (2.49, 9.29)	3.02E-04	*	
RBC	Cardiovascular Disease	13	4.02 (2.05, 7.26)	1.30E-03	*	
RHR	Arrythmias	17	3.93 (2.23, 6.53)	1.17E-04	*	1.50
RHR	Cardiovascular Disease	26	2.75 (1.75, 4.16)	2.94E-04	*	
SCZ	Positive Mood Disorders	9	5.47 (2.37, 11.19)	1.94E-03	*	0.99
SLE	Immune Dysregulation	24	2.94 (1.83, 4.52)	2.18E-04	*	1.41
T2D	Body Mass Disorders	4	23.55 (5.85, 69.99)	9.35E-04	*	1.72
TC	Cardiovascular Disease	38	2.44 (1.69, 3.45)	6.91E-05	*	1.45
TC	Hematologic Disorders	32	2.20 (1.47, 3.18)	2.36E-03	*	
TG	Monogenic Diabetes	12	3.54 (1.78, 6.43)	6.20E-03	*	1.55
TG	Cardiovascular	25	2.22 (1.41,	1.01E-	*	

	Disease		3.38)	02		
TG	Body Mass Disorders	9	3.77 (1.67, 7.49)	2.19E-02	*	
UC	Immune Dysregulation	21	3.72 (2.23, 5.92)	2.77E-05	*	1.56
WHR	Insulin Disorders	9	3.83 (1.67, 7.78)	2.25E-02	*	2.12

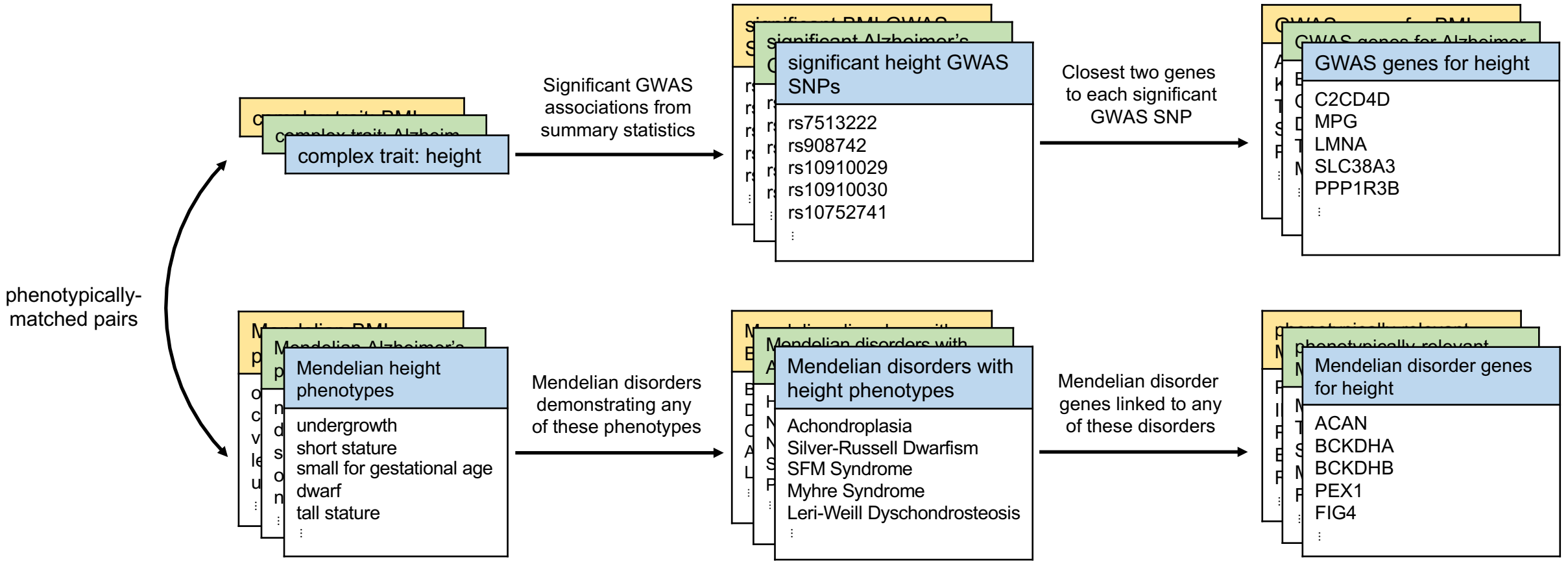
Table 3: Instances of significant overlap of GWAS genes and unrelated Mendelian disorder genes. As in Table 2, Fisher's exact test was used to quantify the enrichment of shared genes between complex traits and Mendelian disorders with an odds ratio and p-value (see Methods). Significance was assessed at a threshold of $p < (0.05/20)$ correcting for the number of Mendelian disorder gene sets compared to each complex trait gene set. This table lists pairs of complex traits and phenotypically-unrelated Mendelian disorders that demonstrated significant overlap.

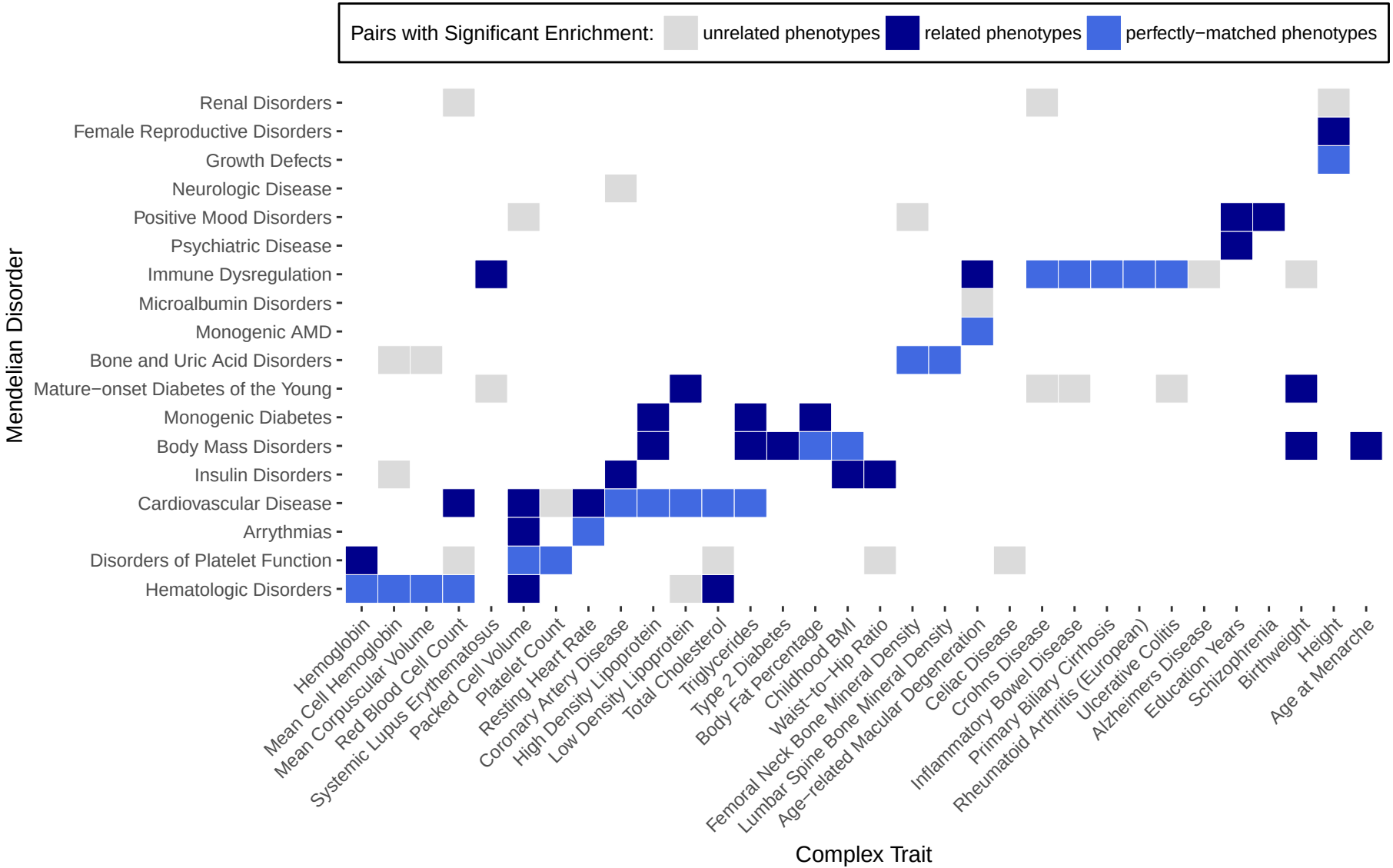
Complex Trait	Mendelian Disorder(s)	Number of Shared Genes	Odds Ratio (CI)	P-value	
IBD	Mature-onset Diabetes of the Young	30	2.81 (1.85, 4.13)	4.81E-05	*
RBC	Renal Disorders	18	4.14 (2.33, 6.96)	5.21E-05	*
UC	Mature-onset Diabetes of the Young	19	3.25 (1.90, 5.27)	4.88E-04	*
HGT	Renal Disorders	153	1.48 (1.23, 1.78)	7.51E-04	*
SLE	Mature-onset Diabetes of the Young	22	2.61 (1.59, 4.07)	2.47E-03	*
CD	Mature-onset Diabetes of the Young	19	2.69 (1.58, 4.35)	4.64E-03	*
LDL	Hematologic Disorders	25	2.31 (1.46, 3.51)	7.01E-03	*
MCH	Insulin Disorders	15	2.80 (1.52, 4.81)	1.40E-02	*
AMD	Microalbumin Disorders	8	4.43 (1.86, 9.13)	1.47E-02	*
MCH	Bone and Uric Acid Disorders	8	4.19 (1.75, 8.61)	2.09E-02	*
TC	Disorders of Platelet Function	25	2.11 (1.34, 3.20)	2.25E-02	*
CD	Renal Disorders	23	2.17 (1.34, 3.37)	2.26E-02	*
PCV	Positive Mood Disorders	3	15.97 (3.11, 51.37)	2.31E-02	*
CAD	Neurologic Disease	7	4.52 (1.76, 9.75)	2.79E-02	*
BW	Immune Dysregulation	14	2.69 (1.43, 4.68)	2.89E-02	*
CEL	Disorders of Platelet Function	5	6.78 (2.04, 17.83)	2.94E-02	*
FN	Positive Mood Disorders	3	14.51 (2.83,	3.00E-	*

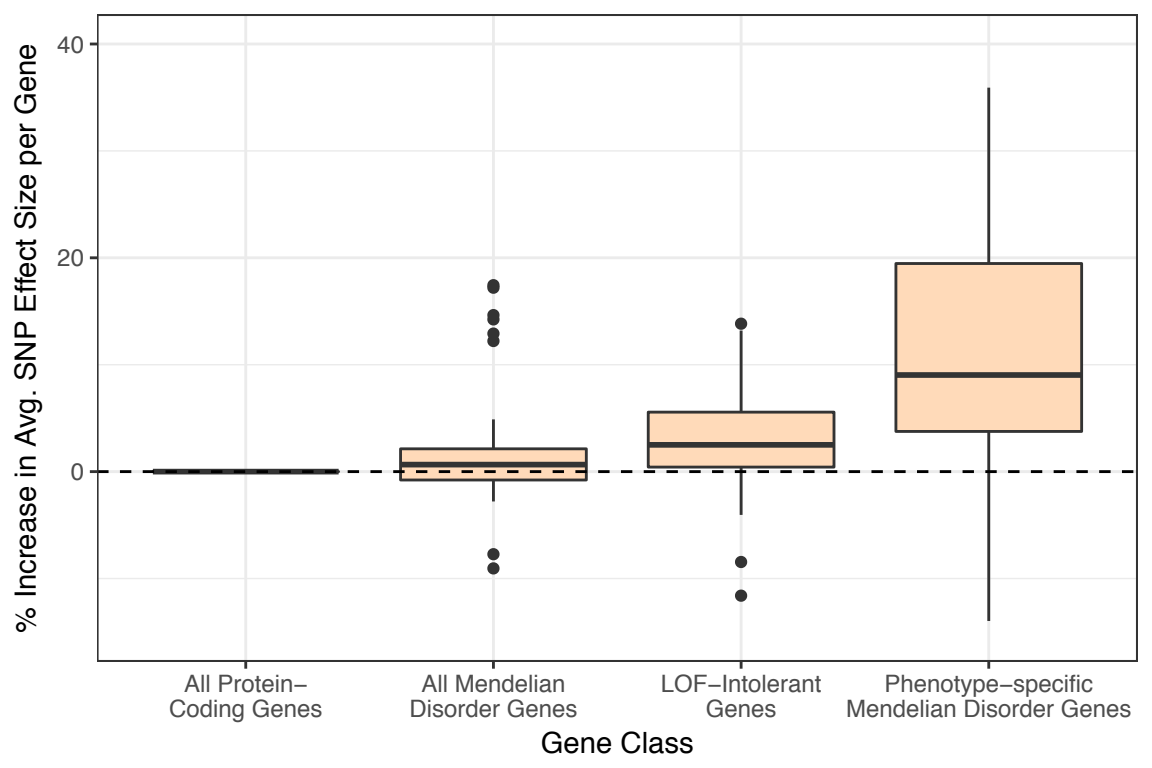
			46.53)	02	
MCV	Bone and Uric Acid Disorders	8	3.80 (1.59, 7.79)	3.78E-02	*
ALZ	Immune Dysregulation	7	4.32 (1.65, 9.62)	4.11E-02	*
WHR	Disorders of Platelet Function	7	4.12 (1.59, 9.03)	4.99E-02	*

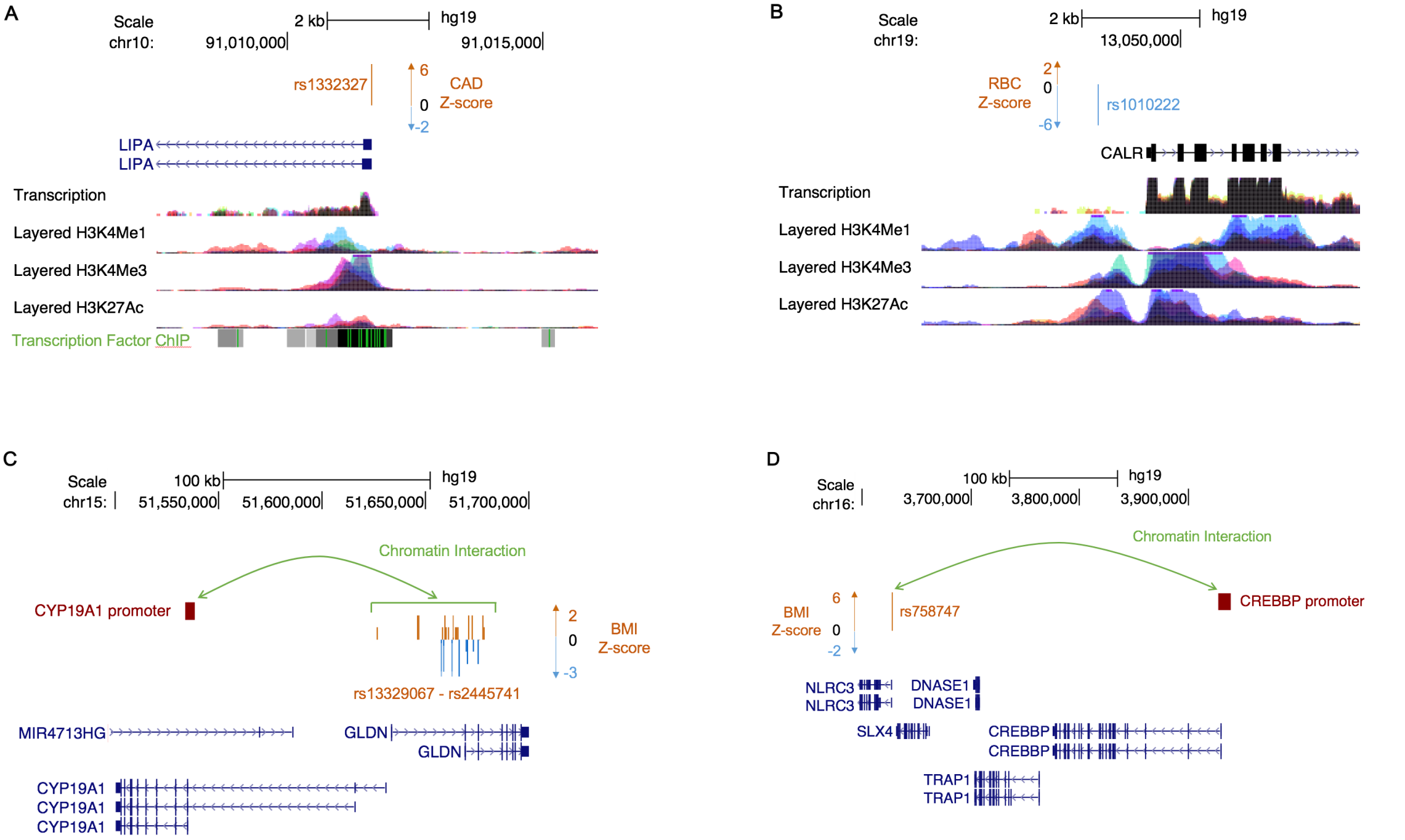
Table 4: Genome-wide significant SNPs localizing at TSS of phenotypically-relevant Mendelian disorder genes. GWAS SNPs from the credible set for each complex trait were intersected with transcription start site (TSS) regions 1kb upstream of phenotypically-matched Mendelian disorder genes. This table lists all genome-wide significant SNPs ($p < 5x10^{-8}$ from GWAS, with chromosomal location) from all complex traits localizing at the TSS of a phenotypically-matched Mendelian disorder gene (italicized).

Complex Trait	SNP ID	Chr	Position	Z score	Gene	Strand
PBC	rs13239597	chr7	128695983	9.85309	*TNPO3*	-
HGT	rs8028537	chr15	89345947	-9.333	*ACAN*	+
HGT	rs10853751	chr19	41903220	8.71	*BCKDHA*	+
CD	rs59283234	chr5	150225587	-8.454	*IRGM*	+
CD	rs751627	chr5	150225113	-8.451	*IRGM*	+
CD	rs35707106	chr5	150225377	-8.332	*IRGM*	+
HGT	rs2298307	chr6	80816296	8.276	*BCKDHB*	+
HGT	rs12386601	chr7	92157886	8.2	*PEX1*	-
BMI	rs17066842	chr18	58040624	-7.542	*MC4R*	-
HGT	rs12192268	chr6	110011458	-7	*FIG4*	+
CAD	rs1332327	chr10	91011681	6.798	*LIPA*	-
RA	rs13239597	chr7	128695983	6.65672	*TNPO3*	-
IBD	rs59283234	chr5	150225587	6.51	*IRGM*	+
IBD	rs751627	chr5	150225113	6.507	*IRGM*	+
HGT	rs7592246	chr2	219926221	6.452	*IHH*	-
IBD	rs34005003	chr5	150225199	6.427	*IRGM*	+
IBD	rs35707106	chr5	150225377	6.326	*IRGM*	+
MNR	rs3775971	chr4	104641920	6.20413	*TACR3*	-
IBD	rs27741	chr16	28504181	6.109	*CLN3*	-
HGT	rs4244808	chr11	2163110	6.061	*IGF2*	-
RBC	rs1010222	chr19	13048608	-5.96154	*CALR*	+
CD	rs27741	chr16	28504181	-5.866	*CLN3*	-
HGT	rs613924	chr11	65769295	-5.862	*BANF1*	+
AFB	rs4845357	chr1	153896212	-5.775	*GATAD2B*	-
HGT	rs6591226	chr11	66675990	5.517	*PC*	-


phenotypically-matched pairs
complex trait: height
Significant GWAS associations from summary statistics
significant height GWAS SNPs
rs7513222
rs908742
rs10910029
rs10910030
rs10752741
Closest two genes to each significant GWAS SNP
GWAS genes for height
C2CD4D
MPG
LMNA
SLC38A3
PPP1R3B
Mendelian height phenotypes
undergrowth
short stature
small for gestational age
dwarf
tall stature
Mendelian disorders demonstrating any of these phenotypes
Mendelian disorders with height phenotypes
Achondroplasia
Silver-Russell Dwarfism
SFM Syndrome
Myhre Syndrome
Leri-Weill Dyschondrosteosis
Mendelian disorder genes linked to any of these disorders
Mendelian disorder genes for height
ACAN
BCKDHA
BCKDHB
PEX1
FIG4


Pairs with Significant Enrichment:
unrelated phenotypes
related phenotypes
perfectly−matched phenotypes
Mendelian Disorder
Renal Disorders
Female Reproductive Disorders
Growth Defects
Neurologic Disease
Positive Mood Disorders
Psychiatric Disease
Immune Dysregulation
Microalbumin Disorders
Monogenic AMD
Bone and Uric Acid Disorders
Mature−onset Diabetes of the Young
Monogenic Diabetes
Body Mass Disorders
Insulin Disorders
Cardiovascular Disease
Arrythmias
Disorders of Platelet Function
Hematologic Disorders
Hemoglobin
Mean Cell Hemoglobin
Mean Corpuscular Volume
Red Blood Cell Count
Systemic Lupus Erythematosus
Packed Cell Volume
Platelet Count
Resting Heart Rate
Coronary Artery Disease
High Density Lipoprotein
Low Density Lipoprotein
Total Cholesterol
Triglycerides
Type 2 Diabetes
Body Fat Percentage
Childhood BMI
Waist−to−Hip Ratio
Femoral Neck Bone Mineral Density
Lumbar Spine Bone Mineral Density
Age−related Macular Degeneration
Celiac Disease
Crohns Disease
Inflammatory Bowel Disease
Primary Biliary Cirrhosis
Rheumatoid Arthritis (European)
Ulcerative Colitis
Alzheimers Disease
Education Years
Schizophrenia
Birthweight
Height
Age at Menarche
Complex Trait


% Increase in Avg. SNP Effect Size per Gene
40
20
0
All Protein−
Coding Genes
All Mendelian
Disorder Genes
LOF−Intolerant
Genes
Phenotype−specific
Mendelian Disorder Genes
Gene Class


A
Scale
chr10:
2 kb
hg19
91,010,000
91,015,000
rs1332327
6
0
-2
CAD
Z-score
LIPA
LIPA
Transcription
Layered H3K4Me1
Layered H3K4Me3
Layered H3K27Ac
Transcription Factor ChIP
B
Scale
chr19:
2 kb
hg19
13,050,000
RBC
Z-score
2
0
-6
rs1010222
CALR
Transcription
Layered H3K4Me1
Layered H3K4Me3
Layered H3K27Ac
C
Scale
chr15:
100 kb
hg19
51,550,000
51,600,000
51,650,000
51,700,000
Chromatin Interaction
CYP19A1 promoter
2
0
-3
BMI
Z-score
rs13329067 - rs2445741
MIR4713HG
GLDN
GLDN
CYP19A1
CYP19A1
CYP19A1
D
Scale
chr16:
100 kb
hg19
3,700,000
3,800,000
3,900,000
Chromatin Interaction
BMI
Z-score
6
0
-2
rs758747
CREBBP promoter
NLRC3
NLRC3
DNASE1
DNASE1
SLX4
CREBBP
CREBBP
TRAP1
TRAP1