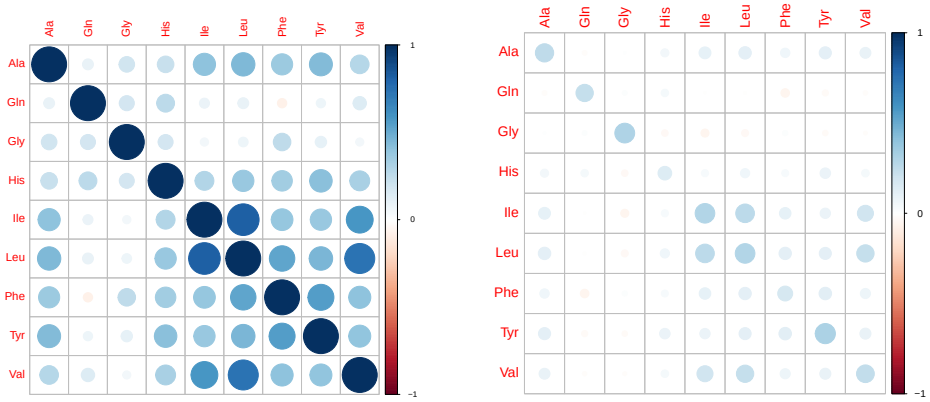
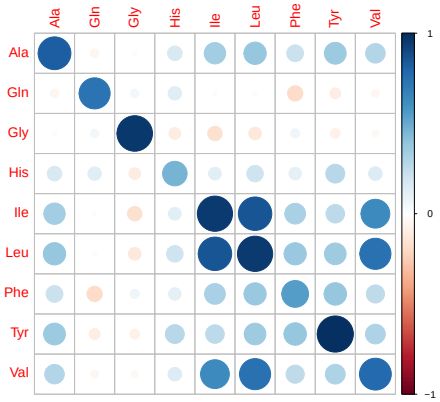
Supplemental figures & tables

Supplemental Figures



(a) **(b)**



(c)

Figure S1: (a): Correlation matrix of the 9 amino acid phenotypes from METSIM study
(b): Co-heritability matrix of the same phenotypes as estimated from PHENIX.
(c): Scaled co-heritability matrix: The elements in the matrix as shown in (b) were divided by the maximum diagonal element.
Ala: Alanine, Gln: Glutamine, Gly: Glycine, His: Histine, Ile: Isoleucine, Leu: Leucine, Phe: Phenylalanine, Tyr: Tyrosine, Val: Valine.

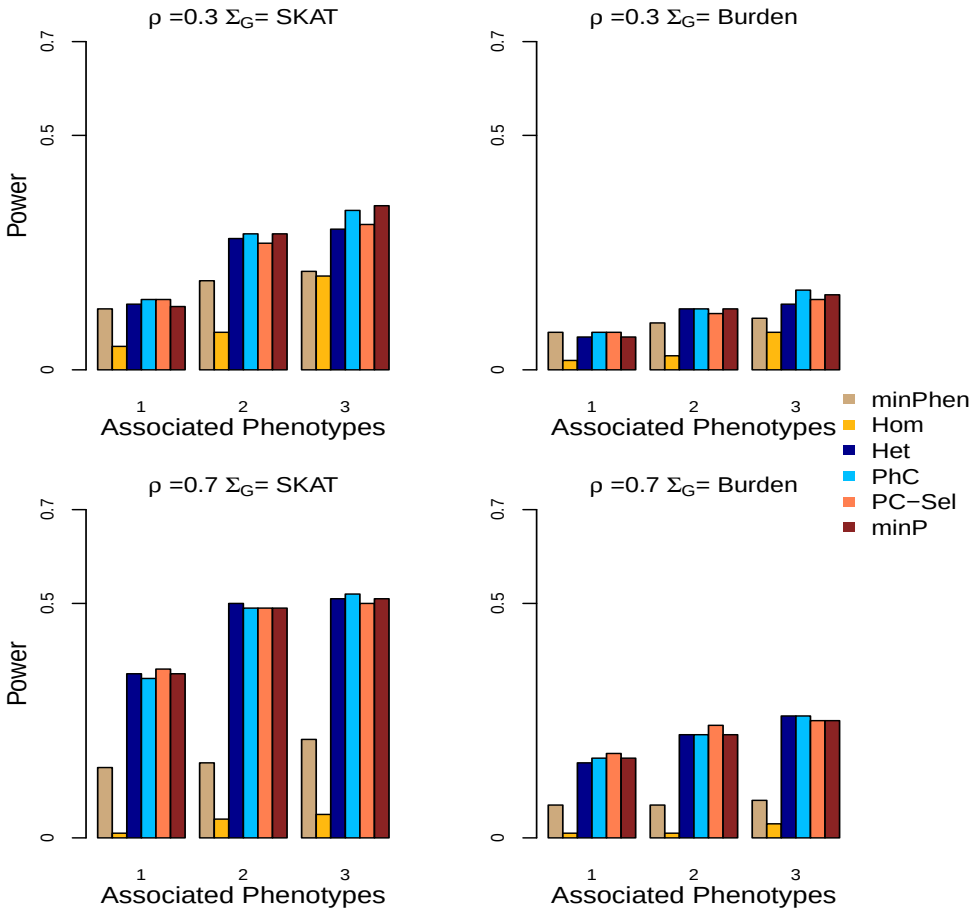


Figure S2: Power for Multi-SKAT tests when phenotypes have compound symmetric correlation structures. Empirical power for minPhen, Hom, Het, PhC, PC-Sel, minP plotted against the number of phenotypes associated with the gene of interest with a total of 5 phenotypes under consideration. Upper row shows the results for $\rho = 0.3$ and lower row for $\rho = 0.7$. Left column shows results with SKAT kernel Σ_G, and right columns shows results with Burden kernel. 80%/20% of the causal variants were trait-increasing/trait-decreasing variants.

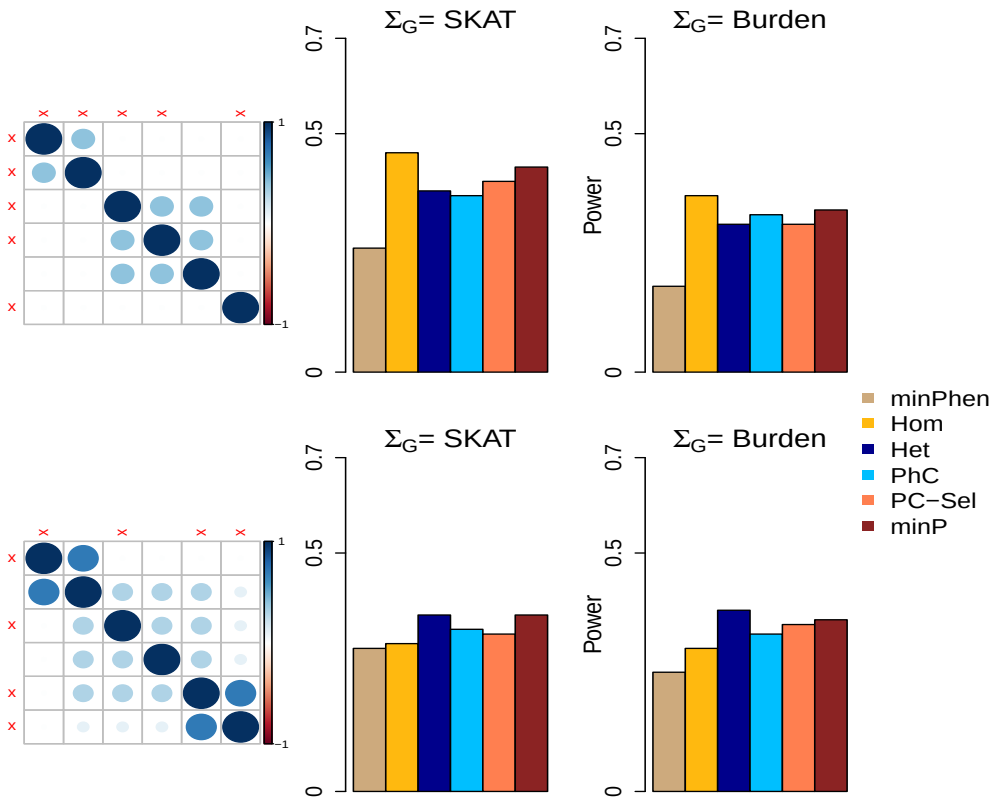


Figure S3: Power for Multi-SKAT tests when phenotypes have clustered correlation structures. Empirical powers for minPhen, Hom, Het, PhC, PC-Sel, minP are plotted under different levels of association with a total of 6 phenotypes and with clustered correlation structures. Middle column shows the empirical powers for different combinations of phenotypes associated with SKAT kernel Σ_G; the rightmost column shows the corresponding results with Burden kernel; left column shows the corresponding correlation matrices for the phenotypes. The associated phenotypes are indicated in red cross marks across the correlation matrices. 80%/20% of the causal variants were trait-increasing/trait-decreasing variants.

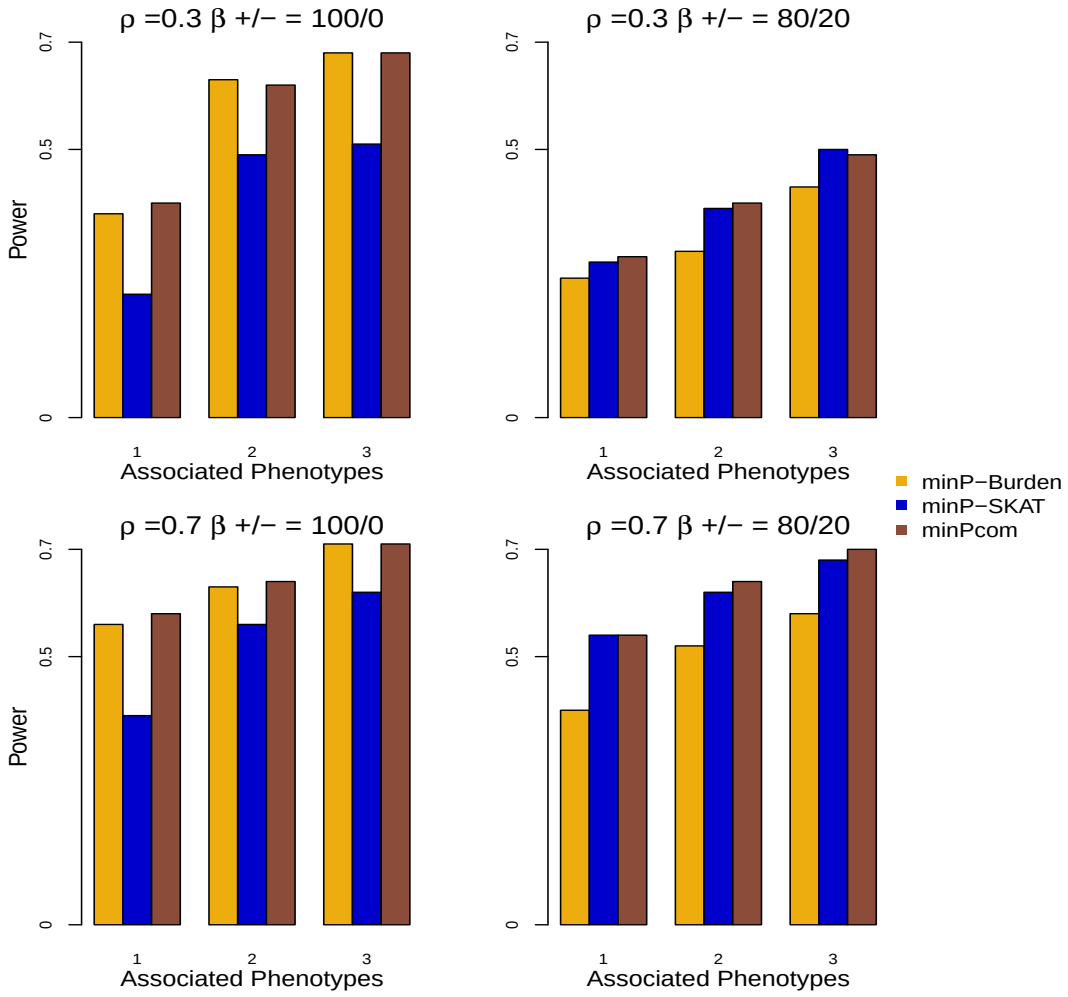


Figure S4: Power for Multi-SKAT by combining tests with Σ_P as Hom, Het, PhC, PC-Sel and Σ_G as SKAT and Burden when phenotypes have compound symmetric correlation structures. Empirical powers for minP-Burden, minP-SKAT and minP_{com} are plotted against the number of phenotypes associated with the gene of interest with a total of 5 phenotypes under consideration and **50%** of the variants in the region are causal. Upper row shows the results for $\rho = 0.3$ and lower row for $\rho = 0.7$. Left column shows results when all the causal variants were trait-increasing variants, and right column shows results when 80%/20% of the causal variants were trait-increasing/trait-decreasing variants.

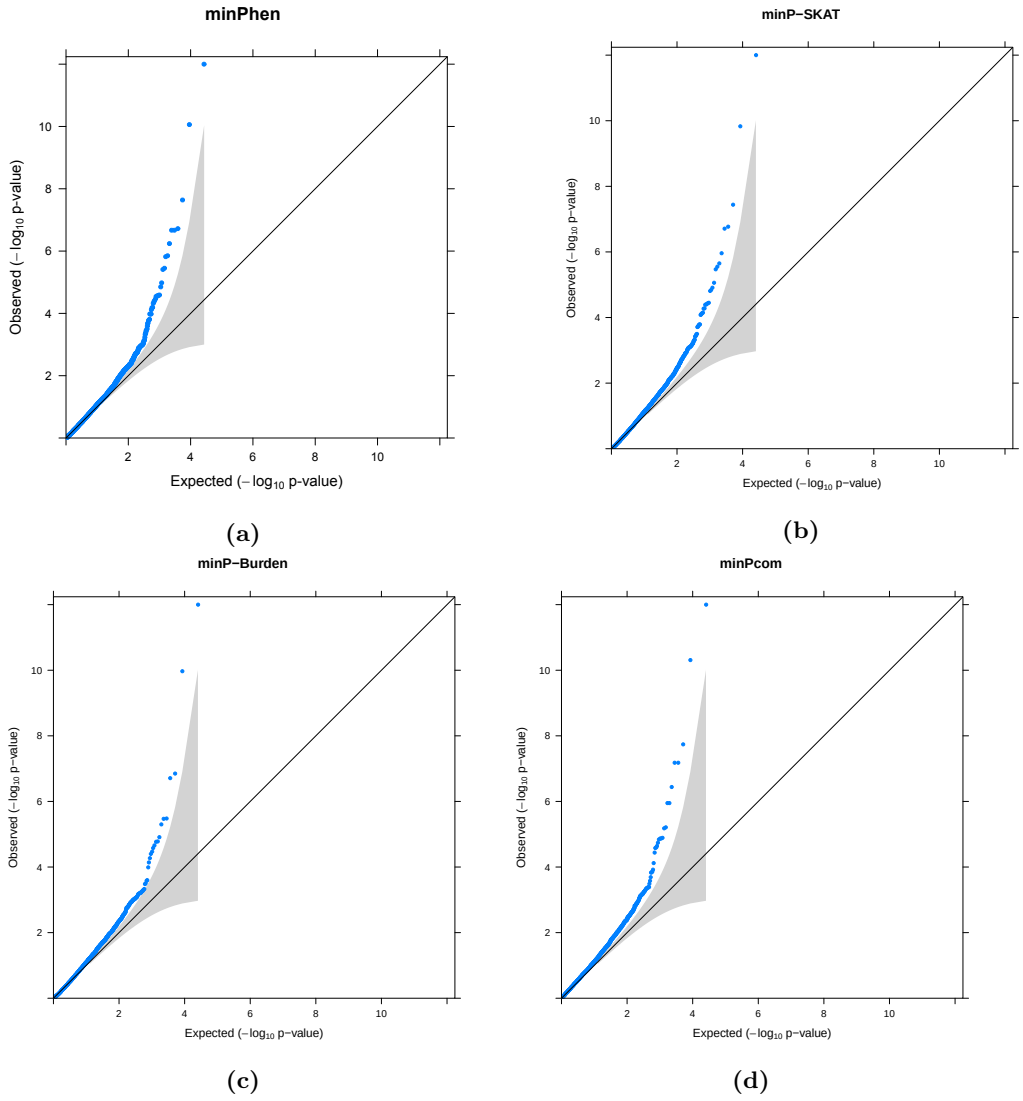


Figure S5: QQplot of the p-values of Multi-SKAT omnibus tests **without kinship adjustment** for the METSIM data (n = 8545). For the ease of viewing, any associations with p-values $< 10^{-12}$ have been collapsed to 10^{-12}

.

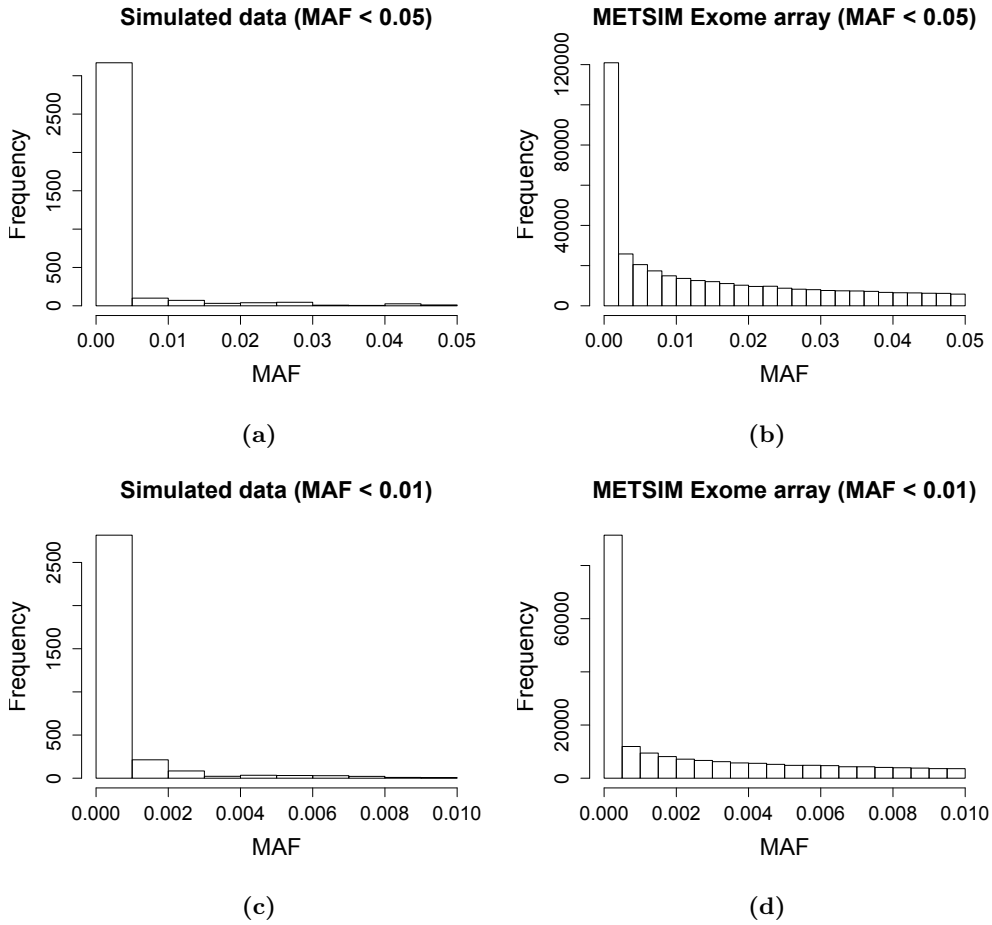
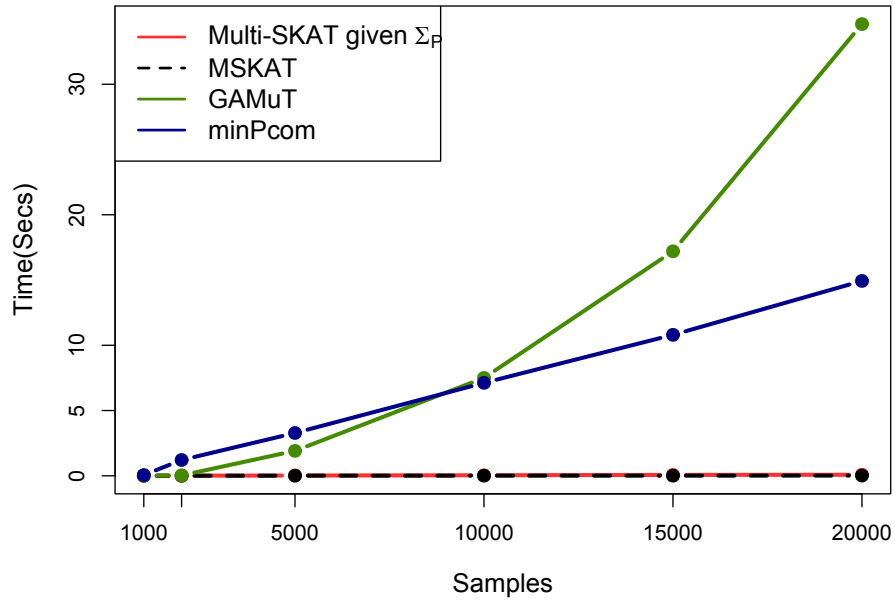
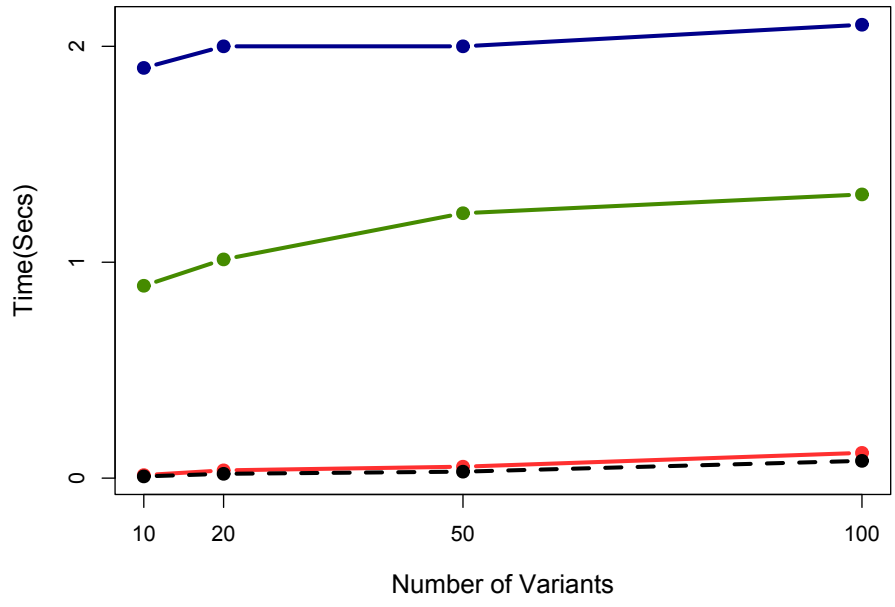


Figure S6: Minor allele frequency (MAF) spectrum for the variants simulated (left panel) and genotyped in METSIM exome array data (right panel). Upper panel shows the MAFs for variants having MAF $< 5\%$. Lower panel zooms in into a region with variants having MAF $< 1\%$



(a)



(b)

Figure S7: Computation time of Multi-SKAT and existing methods with unrelated individuals and 10 phenotypes. (a) Estimated computation time for different sample sizes when the number of variant was 20. (b) Estimated computation time for different number of variants when the sample size was 5000. Each dot represents the average from 100 datasets.

Supplemental Tables

Table S1: Computation time of MultiSKAT tests to analyze a dataset with 5000 individuals, 20 variants and 10 phenotypes. Analysis was done on a 2.80 GHz Intel Xeon CPU.

	Method	CPU sec
Independent samples (Without kinship adjustment)	Multi-SKAT (given Σ_P and Σ_G)	0.014 secs
	minP	2.133 secs
	minP_{com}	3.971 secs
Related samples (With kinship adjustment)	Multi-SKAT (given Σ_P and Σ_G)	2.845 secs
	minP	6.961 secs
	minP_{com}	10.349 secs

Table S2: Backward elimination results for the top 5 genes in Table 2. For a particular gene, each row indicates the phenotype eliminated and the p-value produced correspondingly. The last row indicates the remaining phenotype after backward elimination has been performed. This is the phenotype that drives the signal of association for the particular gene.

	GLDC		HAL		DHODH		PAH		MED1	
Phenotype deleted	p-values	Phenotype deleted	p-values	Phenotype deleted	p-values	Phenotype deleted	p-values	Phenotype deleted	p-values	
	Ile	$\mathbf{2.8 \times 10^{-73}}$	Phe	$\mathbf{9.0 \times 10^{-12}}$	Gln	$\mathbf{3.5 \times 10^{-07}}$	Leu	$\mathbf{6.3 \times 10^{-09}}$	Gly	$\mathbf{9.3 \times 10^{-07}}$
	Val	$\mathbf{8.2 \times 10^{-74}}$	Ile	$\mathbf{3.2 \times 10^{-12}}$	His	$\mathbf{2.4 \times 10^{-08}}$	Ile	$\mathbf{3.2 \times 10^{-10}}$	Ala	$\mathbf{3.3 \times 10^{-06}}$
	Leu	$\mathbf{1.3 \times 10^{-74}}$	Leu	$\mathbf{8.6 \times 10^{-14}}$	Ile	$\mathbf{5.7 \times 10^{-09}}$	Ala	$\mathbf{2.5 \times 10^{-10}}$	His	$\mathbf{2.5 \times 10^{-06}}$
	Tyr	$\mathbf{3.3 \times 10^{-74}}$	Ala	$\mathbf{1.0 \times 10^{-14}}$	Phe	$\mathbf{2.3 \times 10^{-09}}$	His	$\mathbf{7.6 \times 10^{-11}}$	Gln	1.1×10^{-05}
	His	$\mathbf{4.8 \times 10^{-74}}$	Val	$\mathbf{2.6 \times 10^{-14}}$	Val	$\mathbf{2.9 \times 10^{-10}}$	Val	$\mathbf{2.6 \times 10^{-09}}$	Ile	2.3×10^{-05}
	Phe	$\mathbf{2.4 \times 10^{-76}}$	Gly	$\mathbf{2.7 \times 10^{-13}}$	Tyr	$\mathbf{1.2 \times 10^{-10}}$	Gln	$\mathbf{6.3 \times 10^{-07}}$	Phe	1.5×10^{-05}
	Ala	$\mathbf{3.2 \times 10^{-71}}$	Gln	$\mathbf{1.2 \times 10^{-11}}$	Gly	$\mathbf{6.7 \times 10^{-11}}$	Gly	$\mathbf{1.4 \times 10^{-06}}$	Val	1.0×10^{-03}
	Gln	$\mathbf{7.4 \times 10^{-64}}$	Tyr	$\mathbf{3.3 \times 10^{-09}}$	Leu	$\mathbf{1.5 \times 10^{-07}}$	Tyr	6.8×10^{-05}	Leu	4.9×10^{-02}
Remaining phenotype	Gly		His		Ala		Phe		Tyr	

Table S3: Smallest 10 p-values and corresponding genes obtained by PhC($\Sigma_G = SKAT$), GAMuT (Projection and Linear kernel) and MSKAT (Q and Q' statistic). Each method produces the same set of top 10 genes, differing slightly by p-values. The tests were performed on unrelated individuals only (n = 7213).

PhC ($\Sigma_G = SKAT$)		GAMuT (Projection)		MSKAT (Q)		GAMuT (Linear)		MSKAT (Q')	
Gene	p-value	Gene	p-value	Gene	p-value	Gene	p-value	Gene	p-value
GLDC	$\mathbf{8.1 \times 10^{-54}}$	*GLDC*	**0**	*GLDC*	$\mathbf{8.9 \times 10^{-54}}$	*GLDC*	$\mathbf{6.2 \times 10^{-15}}$	*GLDC*	$\mathbf{6.1 \times 10^{-15}}$
DHODH	$\mathbf{1.9 \times 10^{-06}}$	*DHODH*	$\mathbf{2.4 \times 10^{-06}}$	*DHODH*	$\mathbf{2.1 \times 10^{-06}}$	*METTL4*	3.0×10^{-05}	*METTL4*	3.0×10^{-05}
PAH	9.9×10^{-06}	*PAH*	1.0×10^{-05}	*PAH*	9.9×10^{-06}	*ASB10*	4.9×10^{-05}	*ASB10*	4.8×10^{-05}
ALDH1L1	6.0×10^{-05}	*DHODH*	6.1×10^{-05}	*ALDH1L1*	5.9×10^{-05}	*MEOX1*	6.4×10^{-05}	*MEOX1*	6.4×10^{-05}
HAL	9.5×10^{-05}	*HAL*	9.6×10^{-05}	*HAL*	9.5×10^{-05}	*PAH*	1.1×10^{-04}	*PAH*	1.3×10^{-04}
BCAT2	6.2×10^{-04}	*BCAT2*	6.3×10^{-04}	*BCAT2*	6.1×10^{-04}	*ABCC8*	2.5×10^{-04}	*ABCC8*	2.5×10^{-04}
STK33	6.7×10^{-04}	*STK33*	6.7×10^{-04}	*STK33*	6.8×10^{-04}	*OLFML2A*	2.8×10^{-04}	*OLFML2A*	2.8×10^{-04}
TBC1D4	1.7×10^{-04}	*TBC1D4*	1.6×10^{-04}	*TBC1D4*	1.6×10^{-04}	*OGG1*	3.9×10^{-04}	*OGG1*	4.0×10^{-04}
ABCC8	2.1×10^{-04}	*ABCC8*	2.3×10^{-04}	*ABCC8*	2.3×10^{-04}	*CPT1C*	4.7×10^{-04}	*CPT1C*	4.5×10^{-04}
MED1	1.7×10^{-03}	*MED1*	1.7×10^{-03}	*MED1*	1.7×10^{-03}	*DHODH*	2.2×10^{-03}	*DHODH*	2.2×10^{-03}

Table S4: Functions and clinical implications for the significant and suggestive genes

Gene	Function	Clinical Implication / Associations
GLDC	catalyst in glycine cleavage system	glycine encephanlopathy, Autosomal recessive inheritance
HAL	catabolism of Histidine	Histidinemia, vitamin D measurement
DHODH	catalyzing pyrimidine de novo biosynthesis	postaxial acrofacial dysostosis, total cholesterol
PAH	iron containing enzyme	blood metabolite measurements
MED1	coactivator in the transcription of RNA polymerase II-dependent genes	asthma, inflammatory bowel disease
STK33	Serine/threonine protein kinase which phosphorylates VIME	BMI, small cell lung carcinoma
ALDH1L1	catalyzes the conversion of 10-formyltetrahydrofolate, NADP, and water to tetrahydrofolate, NADPH, and carbon dioxide	homocytosine, insulin sensitivity
BCAT2	catabolism of the branched chain amino acids leucine, isoleucine and valine	urinary metabolite, eye measurement, reticulocyte