

The American Journal of Human Genetics, Volume 107

Supplemental Data

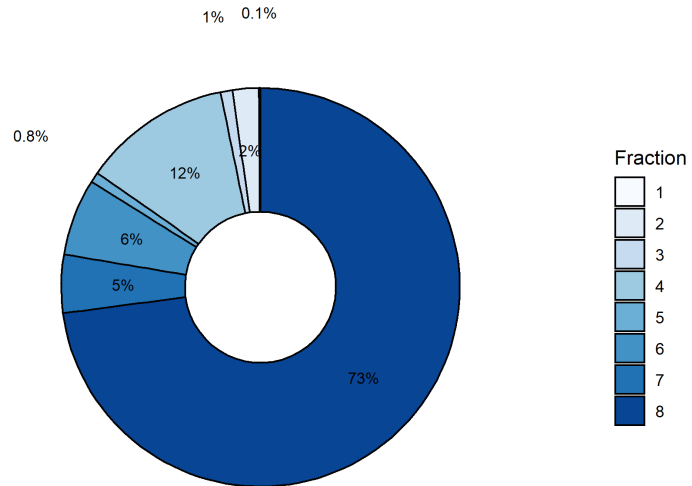
**Characterization of Exome Variants and Their Metabolic
Impact in 6,716 American Indians from the Southwest US**

Hye In Kim, Bin Ye, Nehal Gosalia, Regeneron Genetics Center, Çiğdem Köroğlu, Robert L. Hanson, Wen-Chi Hsueh, William C. Knowler, Leslie J. Baier, Clifton Bogardus, Alan R. Shuldiner, and Cristopher V. Van Hout

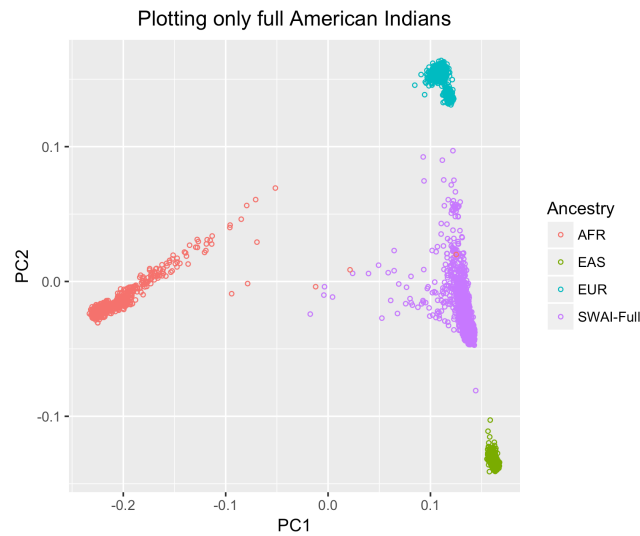
Figure S1. Population structure in the SWAI study

A

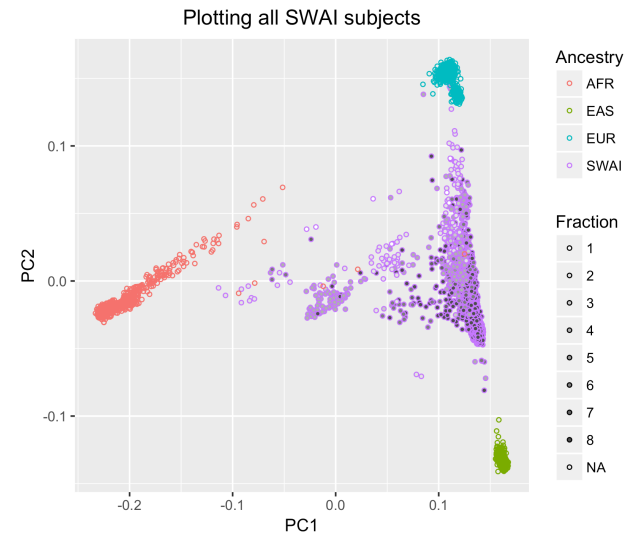
| Fraction Number | | % |
|-----------------|------|-------------------|
| 1 | 8 | 0.1 |
| 2 | 144 | 2.1 |
| 3 | 65 | 1.0 |
| 4 | 809 | 12.0 |
| 5 | 57 | 0.8 |
| 6 | 419 | 6.2 |
| 7 | 317 | 4.7 |
| 8 | 4897 | 72.9 |
| Total | | 6716 100.0 |



B

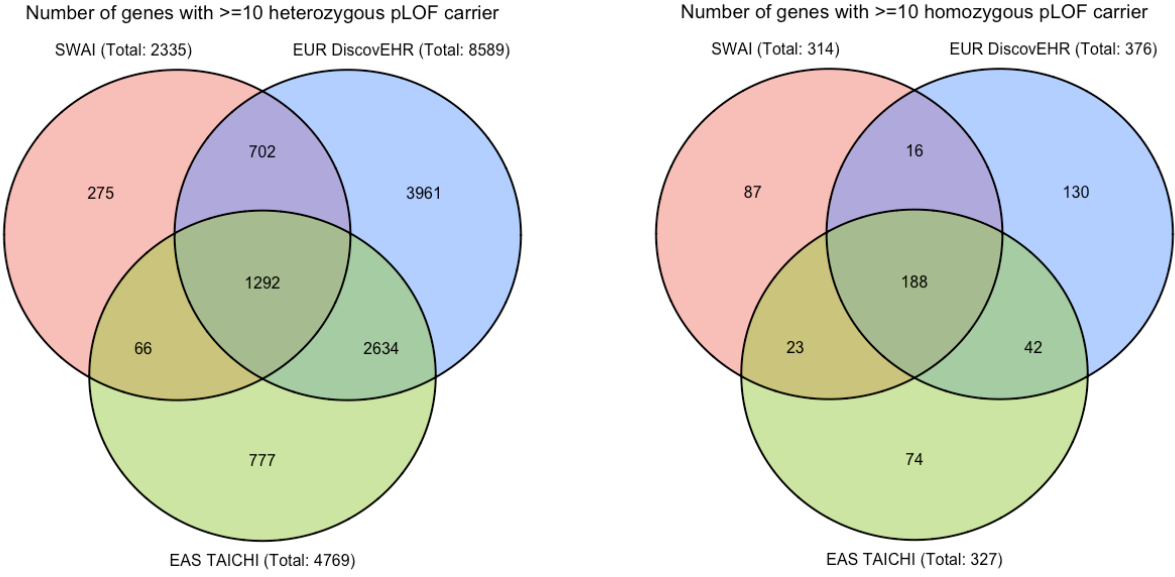


C



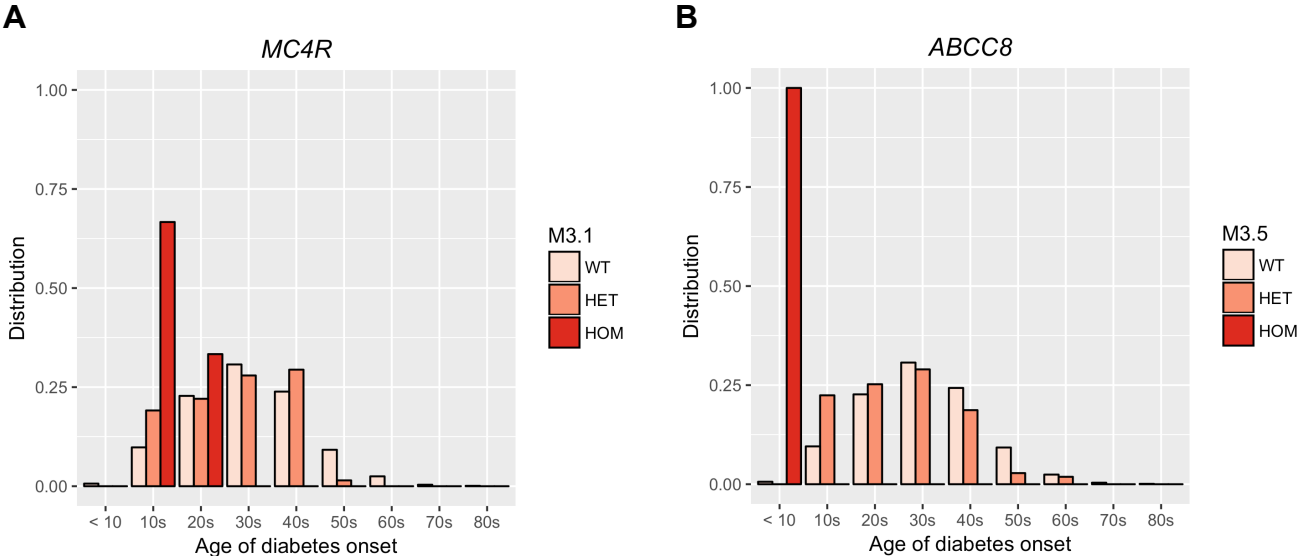
(A) The number and percentage of individuals according to the self-reported admixture (fraction of eight great grandparents that were American Indian). (B) Individuals with full self-reported American Indian ancestry in the SWAI study were projected onto the PC space calculated from African, East Asian, and European ancestries from the 1000 genomes project as reference. (C) All individuals from the SWAI study were projected onto the PC space with the fill intensity indicating the self-reported admixture.

Figure S2. Overlap of genes with at least 10 heterozygous or homozygous pLOF carriers among SWAI, European, and East Asian exomes



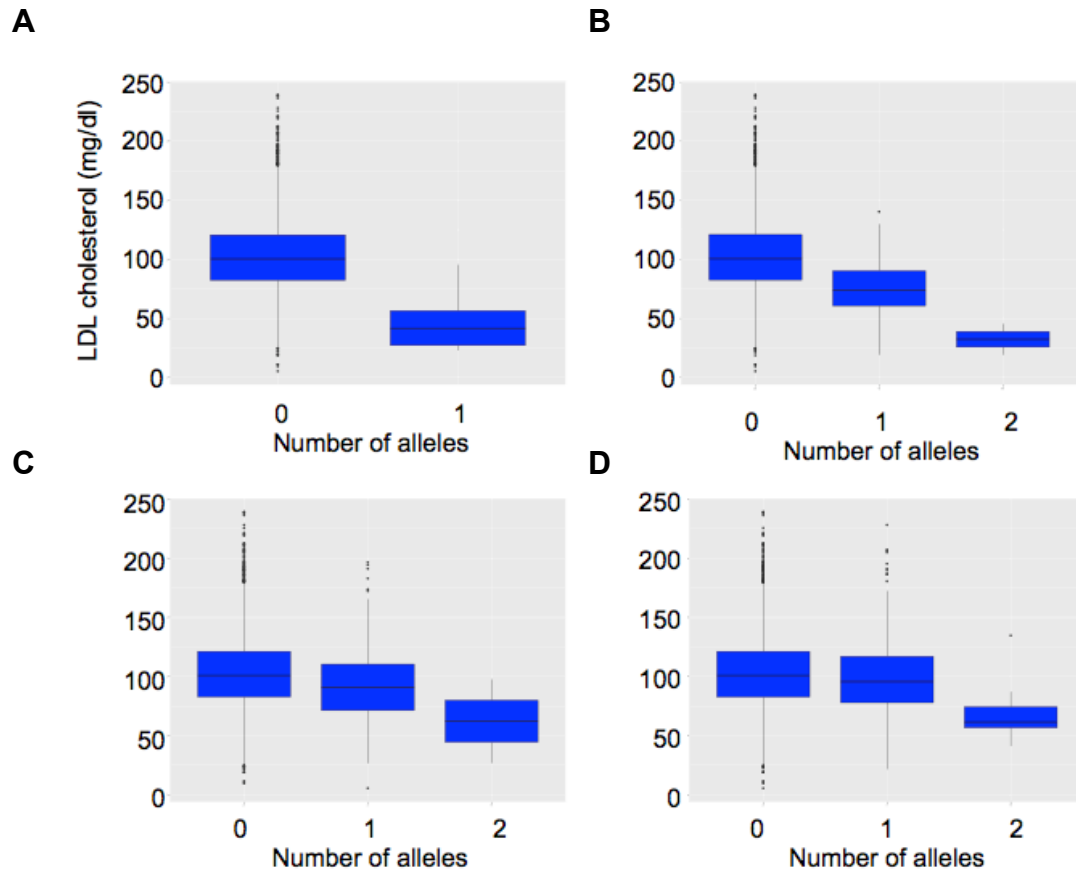
The Venn diagrams represent the overlap of the genes with at least 10 heterozygous (left) and homozygous (right) pLOF carriers in 6,716 SWAI, 29,575 European, and 13,947 East Asian exomes.

Figure S3. The distribution of T2D age of onset per genotype



The distribution of age of T2D onset per the genotype of (A) *MC4R* M3.1 mask and (B) *ABCC8* M3.5 mask.

Figure S4. Plasma LDL cholesterol levels per the genotype of SWAI-enriched variants in *APOB*, *APOE*, *PCSK9*, and *TM6SF2*



| Variant | Genotypic counts | Genotypic means (stdev) | Genotypic medians |
|---------------------------|------------------|--|---------------------|
| <i>APOB</i> p.Ala3175fs | 5043 / 13 / 0 | 102.6 (30.34) / 44.9 (21.4) / - | 100.4 / 41.6 / - |
| <i>APOE</i> p.Ala184Asp | 4984 / 70 / 2 | 102.9 (30.4) / 76.4 (25.2) / 32.4 (18.7) | 100.6 / 73.9 / 32.4 |
| <i>PCSK9</i> p.Gly244Asp | 4823 / 231 / 2 | 103.0 (30.4) / 92.5 (30.5) / 62.1 (50.2) | 100.8 / 90.6 / 62.1 |
| <i>TM6SF2</i> p.Arg138Trp | 4591 / 457 / 8 | 103.0 (30.4) / 98.1 (31.0) / 70.5 (29.2) | 100.8 / 95.6 / 61.4 |

Plasma LDL cholesterol levels per the genotype of (A) *APOB* p.Ala3175fs, (B) *APOE* p.Ala184Asp, (C) *PCSK9* p.Gly244Asp, and (D) *TM6SF2* p.Arg138Trp variants. The genotypic counts, means (with standard deviations), and medians are indicated in the table.

Table S2. Association of the individual variants in MC4R M3.1 mask with maximum body mass index

| Gene Mask | Variant ^a | Variant effect | Amino acid change | Beta ^b | P-value | AAF ^c | Allele frequency ratio (AAF in full American Indians / AAF in respective populations) | | | | | | | |
|-----------|----------------------|----------------|-------------------|-------------------|---------|------------------|---|------------|-----------|------------|------------|------------|------------|------------|
| | | | | | | | GHS EUR | TAICHI EAS | TAICHI NC | gnomAD ALL | gnomAD AMR | gnomAD EUR | gnomAD EAS | gnomAD AFR |
| MC4R M3.1 | 18:60371857:G:C | missense | p.Arg165Gly | 0.666 | 9.8E-04 | 0.0025 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | Inf |
| | 18:60372250:C:CT | frameshift | p.Gly34fs | 0.622 | 1.0E-03 | 0.0023 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | Inf |
| | 18:60371443:C:G | missense | p.Ala303Pro | 0.881 | 1.0E-02 | 0.0008 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | Inf |
| | 18:60371856:C:T | missense | p.Arg165Gln | 0.353 | 2.0E-02 | 0.0044 | 132.9 | 125.3 | 102.1 | 188.2 | 82.6 | Inf | Inf | Inf |

^a Variant is designated as chromosome:position:reference allele:alternate allele on GRCh38 reference.

^b The unit is standard deviation of normalized traits.

^c AAF: alternate allele frequency

The individual variants included in MC4R M3.1 mask with ≥ 10 alternate allele counts were tested for association with maximum body mass index in SWAI. The allele frequency ratios were derived by dividing the allele frequency in the full American Indians from the SWAI study by the allele frequencies in the European and East Asian individuals from DiscovEHR and TAICHI studies, and in the ancestral populations from the gnomAD exome database. When variants were not detected and fall outside of the consistently covered regions in a study, they are indicated as “not captured (NC)”.

Table S3. Association of the individual variants in MC4R M3.1 and ABCC8 M3.5 masks with T2D

| Gene Mask | Variant ^a | Variant effect | Amino acid change | OR ^b | P-value | AAF ^c | Allele frequency ratio (AAF in full American Indians / AAF in respective populations) | | | | | | | | | |
|------------|----------------------|----------------|-------------------|-----------------|----------|------------------|---|------------|------------|------------|------------|------------|------------|-----|--|--|
| | | | | | | | GHS EUR | TAICHI EAS | gnomAD ALL | gnomAD AMR | gnomAD EUR | gnomAD EAS | gnomAD AFR | | | |
| MC4R M3.1 | 18:60372250:C:CT | frameshift | p.Gly34fs | 3.04 | 7.90E-03 | 0.0021 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | Inf | | |
| | 18:60371857:G:C | missense | p.Arg165Gly | 2.92 | 1.61E-02 | 0.0024 | Inf | NC | Inf | Inf | Inf | Inf | Inf | Inf | | |
| | 18:60371856:C:T | missense | p.Arg165Gln | 2.08 | 3.91E-02 | 0.0043 | 132.9 | 125.3 | 188.2 | 102.1 | 82.6 | Inf | Inf | Inf | | |
| ABCC8 M3.5 | 18:60371443:C:G | missense | p.Ala303Pro | 2.55 | 2.62E-01 | 0.0008 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | Inf | | |
| | 11:17395658:C:T | missense | p.Arg1420His | 2.23 | 1.53E-05 | 0.0162 | 489.2 | 115.3 | 701.8 | 107.0 | Inf | Inf | Inf | Inf | | |

^a Variant is designated as chromosome:position:reference allele:alternate allele on GRCh38 reference.

^b OR: odds ratio

^c AAF: alternate allele frequency

The individual variants included in MC4R M3.1 and ABCC8 M3.5 masks with ≥ 10 alternate allele counts were tested for association with T2D in SWAI. The allele frequency ratios were derived by dividing the allele frequency in the full American Indians from the SWAI study by the allele frequencies in the European and East Asian individuals from DiscovEHR and TAICHI studies, and in the ancestral populations from the gnomAD exome database. When variants were not detected and fall outside of the consistently covered regions in a study, they are indicated as “not captured (NC)”.

| | | | | | | | | | | | | | | | | | | | |
|--|------------------|----------|-------------|-------|---------|-------|---------|-------|---------|-------|---------|--------|--|--------|-------------------|-------|-------|-------|------|
| | 1:55039995:C:T | missense | p.Ala53Val | -0.22 | 5.0E-02 | -0.25 | 4.1E-02 | 0.12 | 3.8E-01 | -0.14 | 2.1E-01 | 0.0083 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 |
| | 1:55043922:G:A | missense | p.Arg96His | -0.16 | 8.6E-02 | -0.22 | 4.4E-02 | 0.05 | 6.7E-01 | -0.11 | 2.6E-01 | 0.0114 | Inf | 1602.8 | Inf | Inf | 234.7 | Inf | Inf |
| | 1:55039974:G:T | missense | p.Arg46Leu | -0.45 | 1.3E-01 | -0.34 | 3.0E-01 | -0.67 | 5.7E-02 | 0.20 | 5.2E-01 | 0.0011 | 0.0 | Inf | 0.0 | 0.0 | Inf | 0.2 | 0.2 |
| | 1:55068129:A:G | missense | p.Asn425Ser | -0.41 | 1.6E-01 | -0.45 | 1.6E-01 | 0.08 | 8.0E-01 | 0.06 | 8.3E-01 | 0.0012 | Variant not present in full American Indians | | | | | | |
| | 1:55052746:G:A | missense | p.Val252Met | 0.38 | 1.7E-01 | 0.66 | 4.2E-02 | -0.42 | 2.0E-01 | 0.17 | 5.8E-01 | 0.0014 | 0.8 | Inf | 68.2 | Inf | 76.4 | Inf | 11.0 |
| | 1:55068549:C:T | missense | p.Arg469Trp | 0.28 | 3.0E-01 | 0.40 | 1.9E-01 | -0.66 | 4.2E-02 | 0.48 | 9.2E-02 | 0.0013 | 24.2 | Inf | 1.4 | 3.5 | 30.9 | Inf | 0.1 |
| | 1:55061388:C:G | missense | p.His565Gln | 0.21 | 3.5E-01 | 0.44 | 8.8E-02 | 0.00 | 9.9E-01 | -0.17 | 4.9E-01 | 0.0020 | 1.1 | Inf | Variant failed QC | | | | |
| | 1:55068650:G:A | missense | p.Arg469Gln | 0.11 | 6.8E-01 | 0.37 | 2.0E-01 | -0.04 | 9.0E-01 | 0.11 | 6.8E-01 | 0.0014 | Inf | 436.3 | Inf | Inf | 31.9 | Inf | Inf |
| | 1:55068182:G:A | missense | p.Ala443Thr | -0.05 | 7.9E-01 | -0.15 | 4.9E-01 | 0.16 | 4.9E-01 | 0.04 | 8.5E-01 | 0.0026 | 0.3 | 0.7 | 0.0 | 0.4 | 0.1 | 0.0 | 0.0 |
| | 19:19269759:G:A | missense | p.Arg138Trp | -0.27 | 3.3E-08 | -0.20 | 1.2E-04 | 0.10 | 7.0E-02 | -0.33 | 2.5E-11 | 0.0461 | 2843.9 | Inf | 31.6 | 4.4 | Inf | 432.3 | Inf |
| | 19:19268740:C:T | missense | p.Glu167Lys | -0.23 | 1.4E-03 | -0.27 | 6.4E-04 | -0.06 | 4.7E-01 | -0.15 | 4.3E-02 | 0.0201 | 0.2 | 0.2 | 0.3 | 0.4 | 0.2 | 0.2 | 0.5 |
| | 19:19269704:A:G | missense | p.Leu156Pro | -0.08 | 7.5E-01 | 0.12 | 6.4E-01 | -0.09 | 7.6E-01 | -0.21 | 4.1E-01 | 0.0017 | 0.2 | 10.7 | 0.1 | 0.3 | 0.1 | Inf | 0.9 |
| | 10:112159980:G:T | missense | p.Ser611Arg | 0.19 | 3.2E-03 | 0.11 | 1.4E-01 | 0.57 | 3.8E-14 | -0.17 | 9.0E-03 | 0.0250 | Inf | 383.1 | Inf | Inf | Inf | Inf | Inf |
| | 15:58548359:G:A | missense | p.Glu280Lys | 0.01 | 9.5E-01 | -0.19 | 1.6E-01 | 0.38 | 9.4E-03 | 0.14 | 2.9E-01 | 0.0068 | 386.5 | Inf | 410.0 | 226.0 | 370.3 | 120.1 | Inf |
| | 15:58541878:A:G | missense | p.Ile123Val | -0.35 | 7.6E-02 | -0.35 | 1.1E-01 | 0.60 | 1.1E-02 | -0.62 | 2.3E-03 | 0.0026 | Inf | NC | Inf | Inf | Inf | Inf | Inf |
| | 15:58548541:C:G | missense | p.Phe340Leu | -0.07 | 7.6E-01 | -0.05 | 8.3E-01 | 0.24 | 3.5E-01 | -0.25 | 2.6E-01 | 0.0021 | 0.8 | Inf | Inf | Inf | Inf | Inf | Inf |
| | 11:68773378:C:T | missense | p.Asp543Asn | 0.16 | 2.4E-01 | 0.23 | 1.2E-01 | -0.54 | 6.6E-04 | 0.07 | 5.9E-01 | 0.0058 | Inf | Inf | Inf | Inf | Inf | Inf | Inf |
| | 11:68794860:C:T | missense | p.Ala275Thr | -0.05 | 7.6E-01 | 0.00 | 1.0E+00 | -0.62 | 1.1E-03 | 0.24 | 1.5E-01 | 0.0039 | 0.0 | 3.4 | 0.0 | 0.1 | 0.0 | 22.5 | 0.1 |
| | 11:68780662:G:A | missense | p.Pro479Leu | -0.05 | 8.6E-01 | -0.25 | 4.4E-01 | -0.54 | 9.7E-02 | -0.23 | 4.5E-01 | 0.0014 | Inf | Inf | 51.4 | 28.3 | Inf | Inf | Inf |
| | 11:68807515:C:G | missense | p.Trp135Cys | 0.21 | 4.8E-01 | 0.62 | 5.5E-02 | -0.44 | 2.1E-01 | -0.01 | 9.9E-01 | 0.0011 | Inf | Inf | Inf | Inf | Inf | Inf | Inf |
| | 2:164525023:G:T | missense | p.Ser220Tyr | 0.05 | 7.1E-01 | 0.17 | 2.2E-01 | -0.76 | 2.2E-07 | 0.21 | 1.1E-01 | 0.0068 | Inf | 1375.3 | 185.0 | Inf | Inf | Inf | Inf |
| | 2:164525032:C:A | missense | p.Gly217Val | -0.23 | 3.1E-01 | -0.18 | 4.7E-01 | -0.03 | 9.2E-01 | -0.21 | 3.7E-01 | 0.0020 | 0.0 | Inf | 0.0 | 0.1 | 0.0 | Inf | 0.2 |
| | 11:116830844:G:A | missense | p.Ala43Thr | -0.35 | 2.7E-08 | -0.24 | 6.3E-04 | 0.74 | 8.6E-23 | -1.17 | 1.1E-71 | 0.0259 | 122.1 | 403.1 | 33.3 | 155.2 | 265.2 | 11.1 | 11.1 |

^a Variant is designated as chromosome:position:reference allele:alternate allele on GRCh38 reference.

^b The unit is standard deviation of normalized traits.

^c AAF: alternate allele frequency

The individual variants included in the masks with ≥ 10 alternate allele counts were tested for association with plasma lipid levels in SWAI. The allele frequency ratios were derived by dividing the allele frequency in the full American Indians from the SWAI study by the allele frequencies in the European and East Asian individuals from DiscovEHR and TAICHI studies, and in the ancestral populations from the gnomAD exome database. When variants were not detected and fall outside of the consistently covered regions in a study, they are indicated as “not captured (NC)”.

Table S5. Gene-burden associations of genes within ± 250 kb window of GWAS sentinel variants when the genes with significant gene-burden associations in SWAI were not the closest genes to the GWAS sentinel variants

| Associations from European GWAS | | | | Gene-burden associations in SWAI | | | | | | | |
|---------------------------------|----------------|----------------|------------------|----------------------------------|-------------------|----------|--------------|-------------|--------------|-------------------|----------------|
| Sentinel variant | Closest gene | Variant effect | AAF ^a | Trait | Beta ^b | P-value | Gene | Top mask | Freq | Beta ^b | P-value |
| rs622082 | <i>IGHMBP2</i> | missense | 0.31 | HDLC | -0.02 | 5.9E-10 | CPT7A | M2.1 | 0.014 | -0.50 | 1.3E-06 |
| | | | | | | | GAL | M2.1 | 0.004 | -0.41 | 2.3E-02 |
| | | | | | | | MRPL21 | M4.1 | 0.006 | -0.27 | 8.7E-02 |
| | | | | | | | MRGPRD | M4.5 | 0.045 | 0.08 | 1.5E-01 |
| | | | | | | | TPCN2 | M2.5 | 0.021 | 0.11 | 1.8E-01 |
| | | | | | | | MRGPRF | M2.1 | 0.001 | 0.40 | 3.0E-01 |
| | | | | | | | IGHMBP2 | M4.1 | 0.004 | -0.13 | 4.7E-01 |
| | | | | | | | TESMIN | M2.1 | 0.001 | 0.17 | 6.4E-01 |
| rs12328675 | COBLL1 | 3' UTR | 0.12 | HDLC | 0.05 | 3.1E-37 | GRB14 | M2.1 | 0.013 | -0.44 | 1.8E-05 |
| | | | | | | | COBLL1 | M4.1 | 0.006 | 0.22 | 1.5E-01 |
| | | | | | | | SLC38A11 | M2.1 | 0.010 | 0.07 | 5.6E-01 |
| rs964184 | ZPR1 | 3' UTR | 0.85 | HDLC | 0.11 | 2.6E-217 | APOC3 | M4.5 | 0.026 | 0.74 | 6.0E-23 |
| | | | | | | | BUD13 | M2.1 | 0.005 | -0.38 | 2.1E-02 |
| | | | | | | | APOA4 | M2.5 | 0.052 | -0.11 | 5.1E-02 |
| | | | | | | | APOA5 | M2.1 | 0.002 | -0.40 | 1.6E-01 |
| | | | | | | | S/K3 | M2.1 | 0.006 | 0.13 | 4.1E-01 |
| | | | | | | | ZPR1 | M4.1 | 0.003 | 0.08 | 7.2E-01 |
| | | | | | | | APOA1 | M2.1 | 0.003 | 0.05 | 8.2E-01 |

^a AAF: alternate allele frequency

^b The unit is standard deviation of normalized traits.

Significant associations (P-value < 2.4×10^{-5}) are bolded.

Table S6. The frequency and trait association of the GWAS sentinel variants in SWAI and gene-burden associations before and after conditioning on the GWAS sentinel variants

| Sentinel variant | Closest gene | Variant effect | Trait | Association of sentinel variants in original GWAS | | | Association of sentinel variants in SWAI | | | Association of gene-burden in SWAI | | | | | | |
|------------------|---------------|----------------|---------------|---|---------------------|----------|--|----------------------|---------|------------------------------------|----------|-------|----------------------|---------|---------------------------------------|---------|
| | | | | AAF ^a | Effect ^b | P-value | AAF ^a | Effect ^b | P-value | Gene | Top mask | Freq | Without conditioning | | Upon conditioning on sentinel variant | |
| | | | | | | | | | | | | | Effect ^b | P-value | Effect ^b | P-value |
| rs6567160 | <i>MC4R</i> | intergenic | BMI MaxBMI | 0.230 | 0.06 | 1.8E-178 | 0.016 | -0.04 | 6.3E-01 | <i>MC4R</i> | M3.1 | 0.011 | 0.56 | 5.2E-09 | | |
| rs523288 | <i>MC4R</i> | intergenic | T2D | 0.238 | 1.05 | 7.6E-13 | 0.024 | 0.27 | 4.2E-02 | <i>MC4R</i> | M3.1 | 0.010 | 2.62 | 1.2E-05 | | |
| rs67254669 | <i>ABCC8</i> | missense | T2D | 0.001 | 1.89 | 1.1E-08 | | Variant not detected | | <i>ABCC8</i> | M3.5 | 0.018 | 2.21 | 9.3E-06 | | |
| rs541041 | <i>APOB</i> | intergenic | TC | 0.810 | 0.11 | 5.3E-237 | 0.950 | 0.22 | 2.5E-06 | <i>APOB</i> | M4.5 | 0.062 | -0.21 | 1.4E-06 | -0.24 | 2.1E-07 |
| | | | LDLC | 0.12 | 1.3E-287 | 0.23 | 8.4E-07 | -0.29 | 1.6E-09 | | | | -0.28 | 8.1E-10 | | |
| rs445925 | <i>APOE</i> | downstream | TC | 0.110 | -0.21 | 0 | 0.002 | -0.57 | 8.2E-03 | <i>APOE</i> | M4.1 | 0.014 | -0.63 | 9.4E-14 | | |
| | | | LDLC | -0.32 | 0 | 0.002 | -0.62 | 5.9E-03 | -0.86 | | | | 4.7E-20 | | | |
| rs11591147 | <i>PCSK9</i> | missense | TC | 0.015 | -0.41 | 0 | 0.001 | -0.45 | 1.3E-01 | <i>PCSK9</i> | M2.5 | 0.057 | -0.20 | 6.2E-06 | | |
| | | | LDLC | -0.48 | 0 | 0.001 | -0.34 | 3.0E-01 | -0.44 | | | | 9.1E-11 | | | |
| rs58542926 | <i>TM6SF2</i> | missense | TC | | -0.13 | 7.0E-155 | | -0.23 | 1.4E-03 | <i>TM6SF2</i> | M4.5 | 0.067 | -0.26 | 4.6E-10 | | |
| | | | LDLC | 0.074 | -0.10 | 6.5E-93 | 0.020 | -0.27 | 6.4E-04 | | | | -0.22 | 7.9E-07 | -0.14 | 5.4E-03 |
| | | | TG | | -0.12 | 3.7E-125 | | -0.15 | 4.3E-02 | | | | -0.28 | 2.5E-11 | | |
| rs2792751 | <i>GPAM</i> | missense | LDLC | 0.730 | -0.03 | 3.8E-21 | 0.660 | -0.13 | 1.2E-07 | <i>GPAM</i> | M3.5 | 0.026 | 0.58 | 5.1E-15 | 0.33 | 1.1E-06 |
| rs1800588 | <i>LIPC</i> | upstream | LDLC | 0.240 | 0.12 | 0 | 0.810 | 0.08 | 4.1E-03 | <i>LIPC</i> | M4.1 | 0.013 | 0.47 | 9.8E-06 | 0.33 | 4.8E-04 |
| rs622082 | <i>GHMBP2</i> | missense | LDLC | 0.31 | -0.02 | 5.90E-10 | 0.022 | -0.01 | 9.2E-01 | <i>CPT1A</i> | M2.1 | 0.014 | -0.50 | 1.3E-06 | | |
| rs12328675 | <i>COBLL1</i> | 3' UTR | LDLC | 0.12 | 0.05 | 3.1E-37 | | Variant not detected | | <i>GRB14</i> | M2.1 | 0.013 | -0.46 | 1.8E-05 | | |
| rs964184 | <i>ZPR1</i> | 3' UTR | TC | | -0.09 | 4.7E-135 | | Variant not detected | | <i>APOC3</i> | M4.5 | 0.026 | -0.35 | 3.1E-08 | | |
| | | | LDLC | 0.85 | 0.11 | 2.6E-217 | | Variant not detected | 0.74 | | | | 6.0E-23 | | | |
| | | | TG | | -0.25 | 0 | | | -1.16 | | | | 7.5E-72 | | | |

^a AAF: alternate allele frequency

^b For quantitative traits: beta coefficients in standard deviation unit of normalized traits. For binary traits: odds ratio (OR).