Genetic discovery and translational decision support from exome sequencing of 20,791 type 2 diabetes cases and 24,440 controls from five ancestries

Supplementary Information

Supplementary Tables

Supplementary Table 1: **Samples included in analysis.** Shown are characteristics of the cohorts from which we selected samples for exome sequencing. Subgroup: the label used for the collection of samples throughout the manuscript and figures. Ancestry: the ancestry of the samples. Consortium: the consortium in which samples where first collected and/or analyzed. Study: the study (i.e. cohort) from which samples were drawn. Citation(s): references describing the samples in more detail. T2D Case (Control) Ascertainment: criteria used to define and/or select T2D cases (controls). T1D and MODY exclusion criteria: criteria used (if applicable) to exclude type 1 diabetes or MODY cases from the study. Whole exome sequencing technology: the sequence capture technology used for exome sequencing of the samples. dbGAP (EGA): accession number for download of subgroup data from dbGAP (EGA).

[See separate Excel file]

Supplementary Table 2: **Samples excluded from analysis based on quality control metrics.** To identify samples with evidence of poor sequencing quality, we computed a range of metrics. We then excluded samples who appeared as visual outliers on plots (stratified by sample ancestry and sequencing technology) of these metrics; example plots are shown in **Supplementary Figure 2**. Shown are the number of samples excluded according to each metric, as well as the total number of samples excluded across all metrics.

Metric	Samples Removed
Average (allele balance - 50%)	6
Average allele balance	14
Call rate	260
GWAS concordance (GoT2D samples)	202
GWAS concordance (SIGMA samples)	17
GWAS concordance (T2D-GENES samples)	10
GWAS concordance (newly sequenced samples)	26
Heterozygosity	27
Heterozygosity at common variants	8
Heterozygosity at low frequency variants	41
Number of heterozygous genotypes	15
Number of homozygous non-reference genotypes	12
Number of minor alleles	227
Number of non-reference SNP alleles	2
Number of singleton variants	45
Number of variants	241
Transition:Transversion	30
Total	**481**

Supplementary Table 3: **Variants identified from exome sequencing.** Shown are the number of variants identified by exome sequencing and then advanced for association analysis after quality control. Variant counts are stratified by sequence ontology [1] annotation, produced by the Variant Effect Predictor [2], and further by minor allele frequency (MAF), calculated as the maximum across all ancestries. Rows in the table are shown in decreasing order or predicted deleteriousness.

Annotation	**MAF>0.05**	**0.005<MAF<0.05**	**MAF<0.005**	**Total**
splice_acceptor_variant	82	252	13431	13765
splice_donor_variant	113	297	16898	17308
stop_gained	429	1061	57132	58624
frameshift_variant	756	1501	41963	44221
stop_lost	29	48	1763	1840
start_lost	34	89	3809	3932
inframe_insertion	218	260	3326	3804
inframe_deletion	492	1175	14830	16497
missense_variant	25660	60344	2011159	2097179
protein_altering_variant	10	11	145	166
splice_region_variant	5894	10074	221362	237335
incomplete_terminal_codon_variant	2	1	27	30
stop_retained_variant	18	42	988	1048
synonymous_variant	27748	46978	994052	1068784
coding_sequence_variant	21	21	188	230
mature_miRNA_variant	4	11	421	436
5_prime_UTR_variant	2779	5020	88785	96586
3_prime_UTR_variant	4748	8278	148107	161135
non_coding_transcript_exon_variant	3438	5461	95771	104671
intron_variant	62536	101810	1798630	1963024
upstream_gene_variant	4790	8956	175482	189234
downstream_gene_variant	5730	10705	220197	236636
TF_binding_site_variant	2	2	5	9
regulatory_region_variant	31	58	487	576
intergenic_variant	137	138	1700	1975
Other	894	589	5270	6753
Total	**146595**	**263182**	**5915928**	**6325798**

Supplementary Table 4: **Most significant single-variant associations from exome sequence analysis.** Shown are the most significant results from exome sequence single-variant association analysis. Gene: the closest gene to the variant. Variant: a unique identifier for the variant within our exome sequence analysis. Consequence: the predicted consequence of the variant, defined by sequence ontology annotation and produced by the Variant Effect Predictor. Impact: the impact of the variant, as predicted by the Variant Effect Predictor (High: High; Med: Medium; Low: Low; Mod: Modifier). Change: the predicted protein change, defined according to the "best guess" transcript as described in **Methods**. MAF: the minor allele frequency of the variant, calculated as the maximum across all ancestries. Case: the number of samples with T2D carrying the variant. Ctrl: the number of samples without T2D carrying the variant. OR: the odds-ratio, calculated from the Firth analysis. P: the *p*-value, calculated from the EMMAX analysis.

Gene	Variant	Consequence	Impact	Change	MAF	Case	Ctrl	OR	P
PAX4	rs2233580	missense_variant	Med.	p.Arg192His	0.12	890	563	1.7	7.6e-22
SLC30A8	rs13266634	missense_variant	Med.	p.Arg325Trp	0.43	12258	13756	0.897	3.4e-11
WFS1	rs1801212	missense_variant	Med.	p.Val333Ile	0.27	7101	8456	1.13	1.2e-10
KCNJ11	rs5219	missense_variant	Med.	p.Lys23Glu	0.39	16471	15959	0.898	1.2e-10
KCNJ11	rs5215	missense_variant	Med.	p.Val250Ile	0.39	16687	16132	0.901	3.4e-10
SLC16A11	rs2292351	5_prime_UTR_variant	Low	-	0.34	5244	4249	1.25	1.3e-09
SLC16A11	rs13342692	missense_variant	Med.	p.Asp127Gly	0.34	9468	7492	1.12	1.7e-09
SLC16A11	rs75493593	missense_variant	Med.	p.Pro443Thr	0.3	6262	4929	1.24	3.2e-09
WFS1	rs1801213	synonymous_variant	Low	p.Arg228Arg	0.33	10641	11689	1.1	3.7e-09
SLC16A13	rs76070643	synonymous_variant	Low	p.Tyr166Tyr	0.3	6357	5028	1.2	1.8e-08
WFS1	rs1046317	3_prime_UTR_variant	Mod.	-	0.32	9957	11005	1.09	2.0e-08
SFI1	rs145181683	missense_variant	Med.	p.Arg724Trp	0.16	2861	2144	1.19	3.2e-08
WFS1	rs9998519	intron_variant	Mod.	-	0.39	13395	14741	1.08	4.3e-08
WFS1	rs10010131	intron_variant	Mod.	-	0.39	13046	14406	1.08	5.6e-08
ABCC8	rs757110	missense_variant	Med.	p.Ala1369Ser	0.39	16626	16237	0.913	7.1e-08
WFS1	rs1801214	synonymous_variant	Low	p.Asn500Asn	0.38	12841	14187	1.08	1.6e-07
MC4R	rs79783591	missense_variant	Med.	p.Ile269Asn	0.0089	195	83	2.17	3.4e-07
WFS1	rs1801206	synonymous_variant	Low	p.Val395Val	0.53	15408	16499	1.08	3.7e-07
WFS1	rs1046316	synonymous_variant	Low	p.Ser855Ser	0.32	10412	11572	1.08	5.3e-07
COBLL1	rs7607980	missense_variant	Med.	p.Asn939Asp	0.15	4010	4651	0.857	6.3e-07
PISD	rs12171042	downstream_gene_variant	Mod.	-	0.53	15797	15264	1.09	7.0e-07
PAM	rs35658696	missense_variant	Med.	p.Asp563Gly	0.05	1038	944	1.29	1.3e-06
PPIP5K2	rs36046591	missense_variant	Med.	p.Ser1207Gly	0.049	986	905	1.3	1.4e-06
RAI1	rs3818717	synonymous_variant	Low	p.Ile1867Ile	0.55	14691	16514	0.927	1.8e-06
PPIP5K2	rs116234738	3_prime_UTR_variant	Mod.	-	0.055	956	894	1.29	2.3e-06
MAEA	rs2272481	intron_variant	Mod.	-	0.42	10249	10165	0.898	2.7e-06
COBLL1	rs34305002	intron_variant	Mod.	-	0.25	3646	4625	0.863	3.3e-06
PIK3C2B	rs1553921	synonymous_variant	Low	p.Leu96Leu	0.64	10215	8702	0.912	3.5e-06
WDR13	var_X_48460357	splice_acceptor_variant	Low	-	0.0006	21	2	3.48	3.6e-06
TMCC2	rs1768586	synonymous_variant	Low	p.Ala315Ala	0.36	9113	10374	0.924	4.0e-06
MDM4	rs4252717	intron_variant	Mod.	-	0.42	19786	19287	0.934	4.4e-06
MDM4	rs2290854	intron_variant	Mod.	-	0.72	20466	19245	0.935	4.5e-06
GCKR	rs1260326	missense_variant	Med.	p.Leu446Pro	0.5	15010	16627	1.07	5.4e-06
CDC123	rs12590	3_prime_UTR_variant	Mod.	-	0.24	9078	8543	1.11	5.6e-06
TMCC2	rs1668870	intron_variant	Mod.	-	0.36	8368	9585	0.919	5.7e-06
ANKRD36C	rs188178234	missense_variant	Med.	p.Arg786Trp	0.012	265	364	0.713	6.2e-06
TMCC2	rs1668867	synonymous_variant	Low	p.Tyr562Tyr	0.36	9505	10664	0.926	6.8e-06
PIK3C2B	rs1124777	synonymous_variant	Low	p.Pro199Pro	0.64	10186	8702	0.915	6.9e-06
SIN3A	rs4886696	intron_variant	Mod.	-	0.47	13368	15391	1.08	7.9e-06
ZRANB2	rs11556475	synonymous_variant	Low	p.Tyr114Tyr	0.16	3365	3068	1.12	8.3e-06
CDC123	rs1051055	3_prime_UTR_variant	Mod.	-	0.38	14476	13738	0.902	9.2e-06
TGFB1	rs11466334	intron_variant	Mod.	-	0.082	877	555	1.31	9.5e-06
WFS1	rs1046314	synonymous_variant	Low	p.Lys811Lys	0.49	15003	16045	1.06	9.5e-06
SLC26A3	rs117703371	splice_acceptor_variant	Low	p.Ser438Ser	0.0084	236	169	1.46	1.0e-05
BICD1	rs183649090	intron_variant	Mod.	-	0.0019	28	73	0.426	1.1e-05
RNGTT	rs6937994	3_prime_UTR_variant	Mod.	-	0.075	598	664	0.725	1.1e-05
SNX9	rs80009789	intron_variant	Mod.	-	0.069	740	816	0.783	1.2e-05
PRR14L	rs131224	upstream_gene_variant	Mod.	-	0.19	3986	2813	1.14	1.2e-05
SCN9A	rs12478318	downstream_gene_variant	Mod.	-	0.22	4367	3522	1.13	1.2e-05
PTGER3	rs5671	synonymous_variant	Low	p.Thr206Thr	0.16	3386	3081	1.15	1.3e-05

Supplementary Table 5: **Most significant gene-level associations from exome sequence analysis.** Shown are the most significant results from gene-level analysis. Genes are ranked according the minimum *p*-value achieved across the four gene-level analyses. Gene: a unique identifier for the gene within our exome sequence analysis. Var (CAF): the number of alleles (combined allele frequency of variants) in the mask achieving the strongest association across the four tests (i.e. the 0/5 1% mask for the weighted test, or the mask with the minimum *p*-value for the minimum *p*-value test). Burden: results from the burden analysis. SKAT: results from the SKAT analysis. Min P: results from the minimum *p*-value analysis. Weighted: results from the weighted analysis. OR: the odds-ratio as estimated from the burden analysis. P: the *p*-value from the analysis.

			Burden				SKAT	
	Best Result		Min P		Weighted		Min P	Weighted
Gene	Var	CAF	OR	P	OR	P	P	P
MC4R	41	0.00795	2.07	2.74e-10	2.2	4.81e-09	7.74e-08	3.48e-08
PAM	79	0.0493	1.31	1.58e-08	1.44	2.2e-09	1.53e-07	7.03e-08
SLC30A8	86	0.0116	0.598	1.85e-07	0.397	1.29e-08	0.00011	0.000221
IGFBPL1	33	0.00522	0.564	0.000108	0.208	4.5e-06	0.0222	0.00114
BICD1	188	0.0163	0.857	0.214	0.85	0.575	1.49e-05	0.632
UBE2NL	5	0.000417	12.7	2.71e-05	1.66	0.115	0.00963	0.29
ING3	55	0.00255	2.29	0.000112	7.03	3.47e-05	0.0268	0.0135
HNF1A	131	0.0184	1.23	0.0219	1.47	0.0125	3.62e-05	0.00106
NUMA1	147	0.0459	1.14	0.0249	1.08	0.202	0.000129	4.27e-05
MAP3K15	256	0.0392	0.85	7.76e-05	0.777	0.00239	0.0035	0.12
PDX1	11	0.000371	4.71	0.0214	3.46	0.000166	0.165	0.0573
DPH7	31	0.00631	1.27	0.186	1.3	0.0874	0.000818	0.000184
MGAT4C	28	0.00134	3.13	0.000187	1.9	0.00303	0.00823	0.000886
TMEM216	3	0.301	1.08	0.000294	1.08	0.000207	0.573	0.584
HYAL2	74	0.00308	0.503	0.000728	0.292	0.000228	0.0554	0.024
MBD3	9	0.000649	3.94	0.00415	4.9	0.000239	0.0137	0.00672
SOCS2	37	0.00334	1.96	0.000271	5.16	0.000352	0.00539	0.00334
SLC16A2	57	0.262	0.935	0.000285	0.945	0.000639	0.411	0.575
PTPRC	274	0.033	1.23	0.000297	1.48	0.0162	0.0363	0.11
ANGPTL6	38	0.00598	0.73	0.0585	0.744	0.0443	0.00102	0.0003
STAT4	6	0.00102	0.677	0.00624	0.364	0.000305	0.00294	0.000527
PCBP1	13	0.000371	0.0983	0.000568	0.00288	0.000333	0.243	0.165
MAGEB5	25	0.00183	0.53	0.000454	0.138	0.000885	1.	1.
ARHGEF7	107	0.0201	1.26	0.608	1.18	0.499	0.00165	0.000458
SLC48A1	12	0.000951	3.22	0.000964	22.6	0.000519	0.936	0.777
DGKK	147	0.0121	0.862	0.0519	1.01	0.975	0.00052	0.179
TDRD5	20	0.00805	0.674	0.00203	0.617	0.000521	0.00331	0.00114
TCEA1	12	0.000441	0.0944	0.00053	0.531	0.0331	0.0612	0.354
RXRG	22	0.000788	3.39	0.00291	3.6	0.000546	0.195	0.126
KDM7A	110	0.0136	1.37	0.000559	2.03	0.0017	0.394	0.22
C19orf66	49	0.00413	1.7	0.00165	3.86	0.000566	0.131	0.0712
PDF	27	0.0029	0.578	0.00399	0.776	0.641	0.000584	0.75
OR2M4	10	0.00415	0.554	0.000607	0.699	0.00607	0.0411	0.0263
RHBDD2	8	0.000557	0.721	0.0642	0.458	0.0181	0.00559	0.000626
HLCS	66	0.00364	1.91	0.000644	1.42	0.0278	0.161	0.167
NT5DC1	92	0.0123	1.21	0.0905	1.46	0.0633	0.00276	0.000653
TMEM161B	84	0.00487	1.56	0.00856	2.97	0.000664	0.166	0.0599
LRRTM3	80	0.0083	1.44	0.00078	2.54	0.000884	0.000678	0.000673
PRSS54	8	0.000719	0.569	0.295	0.923	0.646	0.000717	0.0802
MEST	29	0.00148	2.48	0.00341	5.64	0.000743	0.216	0.0942
MIEF1	102	0.0501	1.17	0.000751	1.28	0.000891	0.00996	0.0305
RUVBL2	84	0.00582	1.58	0.00122	2.49	0.000803	0.246	0.0842
FLCN	124	0.0111	0.702	0.00245	0.485	0.000816	0.706	0.945
DHX9	11	0.000441	0.0993	0.000829	0.569	0.153	0.037	0.00564
TRDN	131	0.0227	1.28	0.000854	1.17	0.598	0.00107	0.0517
ABCB11	226	0.0231	1.21	0.0161	1.55	0.000881	0.261	0.0236
MAGEE2	119	0.278	1.07	0.000907	1.06	0.00241	0.473	0.229
ANKS3	186	0.0255	0.808	0.00287	0.664	0.000908	0.614	0.232
DHRS13	73	0.00853	1.66	0.0899	1.3	0.101	0.000917	0.0054
PPP3CA	41	0.00445	0.745	0.0489	0.517	0.0712	0.000921	0.000992

Supplementary Table 6: **Associations by allele mask for most significant gene-level associations.** For the 11 strongest gene-level associations, as determined by the weighted burden, weighted SKAT, minimum *p*-value burden, and minimum *p*-value SKAT analyses, shown are statistics for each mask and each of the burden test and SKAT. We performed analyses without the use of allele weights and including all alleles in each mask (so that the sets of alleles are nested within masks). Gene: a unique identifier for the gene within our exome sequence analysis. Trans: the transcript set used for the analysis (All: all transcripts. Best: "best-guess" transcript). Mask: the allele mask used for analysis. Var: the number of alleles included in the mask. CAF: the combined allele frequency of all alleles in the mask. OR: the aggregate odds-ratio for alleles in the mask, computed from the burden test. Burden: the *p*-value from burden analysis of alleles in the mask. SKAT: the *p*-value from SKAT analysis of alleles in the mask.

Gene	Trans	Mask	Var	CAF	OR	Burden	SKAT
		LofTee	4	0.000325	1.72	0.317	0.772
		16/16	4	0.000325	1.72	0.317	0.772
		11/11	5	0.000371	1.78	0.263	0.823
MC4R	All	5/5	40	0.00791	2.05	1.57e-10	5.38e-08
		5/5 + LofTee LC	41	0.00795	2.07	8.46e-11	4.88e-08
		1/5 1%	94	0.0156	1.45	2.73e-06	2.39e-08
		0/5 1%	105	0.0173	1.41	4.17e-06	6.21e-08
		LofTee	6	0.000255	1.21	0.754	0.684
		16/16	6	0.000255	1.21	0.754	0.684
		11/11	14	0.00058	1.41	0.394	0.304
PAM	All	5/5	79	0.0493	1.31	4.28e-09	1.e-07
		5/5 + LofTee LC	79	0.0493	1.31	4.28e-09	1.e-07
		1/5 1%	196	0.0603	1.27	1.01e-08	4.13e-08
		0/5 1%	213	0.0626	1.26	7.64e-09	1.34e-07
		LofTee	13	0.00118	0.457	0.0072	0.0323
		16/16	13	0.00118	0.457	0.0072	0.0323
		11/11	15	0.00132	0.504	0.0127	0.041
SLC30A8	Best	5/5	37	0.00448	0.473	1.43e-06	0.000724
		5/5 + LofTee LC	37	0.00448	0.473	1.43e-06	0.000724
		1/5 1%	86	0.0116	0.598	4.7e-08	3.57e-05
		0/5 1%	103	0.0135	0.622	5.46e-08	2.8e-05
		LofTee	2	0.000209	0.0577	0.00267	0.00744
		16/16	2	0.000209	0.0577	0.00267	0.00744
		11/11	2	0.000209	0.0577	0.00267	0.00744
IGFBPL1	All	5/5	2	0.000209	0.0577	0.00267	0.00744
		5/5 + LofTee LC	4	0.000441	0.319	0.0214	0.0225
		1/5 1%	33	0.00522	0.564	3.59e-05	0.00745
		0/5 1%	45	0.0101	0.713	0.000692	0.108
		LofTee	2	4.64e-05	0.137	0.142	0.245
		16/16	2	4.64e-05	0.137	0.142	0.245
		11/11	2	4.64e-05	0.137	0.142	0.245
BICD1	All	5/5	34	0.00262	0.865	0.457	0.585
		5/5 + LofTee LC	36	0.00271	0.893	0.552	0.585
		1/5 1%	180	0.0135	0.979	0.8	0.393
		0/5 1%	198	0.0166	0.863	0.0529	2.83e-06
		5/5 + LofTee LC	5	0.000417	12.7	1.03e-05	0.00368
UBE2NL	Best	1/5 1%	19	0.00385	1.27	0.0466	0.185
		0/5 1%	31	0.00466	1.23	0.0225	0.157
		11/11	2	9.28e-05	7.49	0.0855	0.144
		5/5	7	0.000209	3.11	0.115	0.213
ING3	All	5/5 + LofTee LC	8	0.000487	2.37	0.058	0.224
		1/5 1%	55	0.00255	2.29	3.1e-05	0.0106
		0/5 1%	61	0.00385	1.84	0.000132	0.0582
		LofTee	3	6.96e-05	2.77	0.364	0.552
		16/16	3	6.96e-05	2.77	0.364	0.552
		11/11	24	0.00262	1.57	0.0204	0.049
		5/5	32	0.00304	1.34	0.101	0.0458
		5/5 + LofTee LC	33	0.00306	1.32	0.118	0.0458
		1/5 1%	119	0.0166	1.23	0.00504	8.38e-06

HNF1A	Best						
		0/5 1%	131	0.0184	1.14	0.0526	8.26e-06
NUMA1	All	LofTee	9	0.000209	0.813	0.766	0.531
		16/16	9	0.000209	0.813	0.766	0.531
		11/11	9	0.000209	0.813	0.766	0.531
		5/5	147	0.0459	1.13	0.00828	2.47e-05
		5/5 + LofTee LC	149	0.0465	1.13	0.00909	2.54e-05
		1/5 1%	460	0.0893	1.02	0.5	4.72e-05
		0/5 1%	563	0.107	1.03	0.333	2.99e-05
MAP3K15	All	LofTee	26	0.0039	0.732	0.011	0.157
		16/16	26	0.0039	0.732	0.011	0.157
		11/11	28	0.00399	0.732	0.0104	0.152
		5/5	72	0.00916	0.834	0.0198	0.128
		5/5 + LofTee LC	80	0.00974	0.83	0.0137	0.14
		1/5 1%	219	0.031	0.85	9.95e-05	0.00468
		0/5 1%	256	0.0392	0.85	1.38e-05	0.000622
PDX1	Best	LofTee	2	4.64e-05	5.23	0.214	1.
		16/16	2	4.64e-05	5.23	0.214	1.
		11/11	11	0.000371	4.71	0.00546	0.274
		5/5	15	0.00294	1.	0.989	0.274
		5/5 + LofTee LC	15	0.00294	1.	0.989	0.274
		1/5 1%	57	0.00779	1.25	0.043	0.0448
		0/5 1%	65	0.00893	1.17	0.131	0.045

Supplementary Table 7: **Evaluation of association signals in CHARGE.** Shown are results from gene-level analysis within the CHARGE dataset, in which the 50 genes with lowest *p*-value were advanced for analysis. Results are shown for each mask. Var: the number of alleles in the mask; CAC: the combined count of all alleles in the mask; Score: the score statistic from a burden analysis of the mask (positive values denote increased risk, negative values denote decreased risk); Burden: the *p*-value from a burden analysis of the mask; SKAT: the *p*-value from a SKAT analysis of the mask. Best Burden (SKAT) indicate *p*-values from an minimum *p*-value test across all masks for the Burden (SKAT) analyses.

Gene	Mask	Var	CAC	Score	Burden	SKAT
LRRTM3	5/5	4	5	7.1	0.0076	0.12
	5/5 + LofTee LC	4	5	7.1	0.0076	0.12
	1/5 1%	32	139	5.1	0.024	0.48
	0/5 1%	32	139	5.1	0.024	0.48
	Best Burden				0.015	
	Best SKAT					0.23
DHX9	11/11	1	1	3.1	0.076	0.076
	5/5	5	6	0.51	0.47	0.18
	5/5 + LofTee LC	5	6	0.51	0.47	0.18
	1/5 1%	26	36	1.3	0.26	0.43
	0/5 1%	39	95	8.3	0.004	0.0087
	Best Burden				0.016	
	Best SKAT					0.034
MC4R	LofTee	2	2	-0.46	0.5	0.79
	16/16	2	2	-0.46	0.5	0.79
	11/11	3	4	2.7	0.1	0.0054
	5/5	12	18	4.8	0.029	0.018
	5/5 + LofTee LC	12	18	4.8	0.029	0.018
	1/5 1%	33	306	0.27	0.61	0.66
	0/5 1%	40	339	0.023	0.88	0.63
	Best Burden				0.14	
	Best SKAT					0.026
	LofTee	1	1	2.9	0.088	0.088
	16/16	1	1	2.9	0.088	0.088
	11/11	2	2	6.9	0.0086	0.032
	5/5	11	15	4.8	0.029	0.13
	5/5 + LofTee LC	11	15	4.8	0.029	0.13
	1/5 1%	35	107	1.9	0.19	0.29

MGAT4C						
	0/5 1%	37	112	1.1	0.31	0.29
	Best Burden				0.04	
	Best SKAT					0.14
MAGEE2	LofTee	5	6	-0.89	0.35	0.34
	16/16	5	6	-0.89	0.35	0.34
	11/11	5	6	-0.89	0.35	0.34
	5/5	5	6	-0.89	0.54	0.38
	5/5 + LofTee LC	6	7	-0.1	0.75	0.23
	1/5 1%	32	122	2.2	0.14	0.19
	0/5 1%	57	230	-0.3	0.58	0.01
	Best Burden				0.45	
	Best SKAT					0.041
UBE2NL	5/5	3	4	5.7	0.017	0.12
	5/5 + LofTee LC	3	4	5.7	0.017	0.12
	1/5 1%	13	44	1.1	0.29	0.56
	0/5 1%	22	62	0.8	0.37	0.77
	Best Burden				0.046	
	Best SKAT					0.3
PCBP1	5/5	2	4	5.8	0.032	0.075
	5/5 + LofTee LC	2	4	5.8	0.031	0.072
	1/5 1%	13	24	1.9	0.17	0.33
	0/5 1%	17	63	1.7	0.2	0.18
	Best Burden				0.062	
	Best SKAT					0.14
DGKK	LofTee	2	2	5.2	0.023	0.072
	16/16	2	2	5.2	0.023	0.072
	11/11	2	2	5.2	0.023	0.072
	5/5	2	2	5.2	0.023	0.072
	5/5 + LofTee LC	2	2	5.2	0.023	0.072
	1/5 1%	36	119	0.11	0.74	0.9
	0/5 1%	71	304	-0.013	0.91	0.96
	Best Burden				0.066	
	Best SKAT					0.2
PRSS54	LofTee	1	1	-0.54	0.7	0.68
	16/16	1	1	-0.54	0.7	0.68
	11/11	1	1	-0.54	0.7	0.68
	5/5	4	5	-1.7	0.58	0.38
	5/5 + LofTee LC	4	5	-1.7	0.35	0.24
	1/5 1%	33	275	-3.	0.084	0.3
	0/5 1%	13	34	-6.6	0.021	0.039
	Best Burden				0.078	
	Best SKAT					0.12
MAP3K15	LofTee	5	73	-3.8	0.082	0.018
	16/16	5	73	-3.8	0.082	0.018
	11/11	6	78	-4.2	0.063	0.028
	5/5	28	182	-3.6	0.1	0.14
	5/5 + LofTee LC	30	186	-3.4	0.085	0.11
	1/5 1%	77	460	-3.3	0.099	0.18
	0/5 1%	92	493	-2.7	0.15	0.17
	Best Burden				0.23	
	Best SKAT					0.1
IGFBPL1	LofTee	1	2	0.49	0.48	0.48
	16/16	1	2	0.49	0.48	0.48
	11/11	1	2	0.49	0.48	0.48
	5/5	5	82	4.	0.089	0.13
	5/5 + LofTee LC	5	82	4.	0.045	0.068
	1/5 1%	18	236	1.5	0.22	0.046
	0/5 1%	27	382	0.26	0.61	0.066
	Best Burden				0.17	
	Best SKAT					0.17
	1/5 1%	3	26	2.4	0.12	0.16
	0/5 1%	4	27	1.9	0.17	0.17

RAPGEF3	Best Burden				0.17	
	Best SKAT					0.23
ARHGEF7	5/5	12	81	0.91	0.34	0.7
	5/5 + LofTee LC	12	81	0.91	0.34	0.7
	1/5 1%	56	316	1.9	0.18	0.89
	0/5 1%	59	328	3.1	0.083	0.85
	Best Burden				0.18	
	Best SKAT					0.95
TMEM161B	LofTee	3	3	-0.74	0.62	0.85
	16/16	3	3	-0.74	0.62	0.85
	11/11	3	3	-0.74	0.62	0.85
	5/5	10	16	-4.6	0.095	0.82
	5/5 + LofTee LC	10	16	-4.6	0.061	0.72
	1/5 1%	27	204	-2.4	0.15	0.98
	0/5 1%	29	206	-2.1	0.18	0.98
	Best Burden				0.21	
	Best SKAT					0.98
KDM7A	5/5	7	17	1.1	0.49	0.12
	5/5 + LofTee LC	8	18	0.9	0.34	0.064
	1/5 1%	39	207	0.063	0.8	0.28
	0/5 1%	53	324	-0.017	0.9	0.41
	Best Burden				0.74	
	Best SKAT					0.22
HNF1A	LofTee	1	1	-0.18	0.67	0.67
	16/16	1	1	-0.18	0.67	0.67
	11/11	15	35	2.5	0.11	0.046
	5/5	22	44	1.4	0.23	0.05
	5/5 + LofTee LC	22	44	1.4	0.23	0.05
	1/5 1%	71	218	1.6	0.21	0.44
	0/5 1%	88	271	1.9	0.21	0.56
	Best Burden				0.47	
	Best SKAT					0.22
TRDN	LofTee	19	41	0.1	0.75	0.81
	16/16	19	41	0.1	0.75	0.81
	11/11	19	41	0.1	0.75	0.81
	5/5	30	61	2.4	0.15	0.083
	5/5 + LofTee LC	31	62	2.3	0.13	0.066
	1/5 1%	67	348	0.7	0.4	0.66
	0/5 1%	94	806	0.18	0.74	0.96
	Best Burden				0.44	
	Best SKAT					0.27
RHBDD2	5/5	6	14	2.6	0.11	0.097
	5/5 + LofTee LC	6	14	2.6	0.11	0.097
	1/5 1%	27	45	-0.0037	0.97	0.48
	0/5 1%	36	160	0.013	0.91	0.79
	Best Burden				0.3	
	Best SKAT					0.27
NUMA1	LofTee	1	1	4.1	0.083	0.083
	16/16	1	1	4.1	0.083	0.083
	11/11	1	1	4.1	0.083	0.083
	5/5	25	128	2.4	0.23	0.19
	5/5 + LofTee LC	25	128	2.4	0.22	0.19
	1/5 1%	200	1194	-0.57	0.69	0.9
	0/5 1%	246	1534	-0.84	0.57	0.95
	Best Burden				0.28	
	Best SKAT					0.28
	LofTee	3	10990	2.5	0.11	0.13
	16/16	3	10990	2.5	0.11	0.13
	11/11	4	10990	2.5	0.11	0.13
	5/5	4	10990	2.5	0.18	0.2
	5/5 + LofTee LC	5	12005	2.2	0.14	0.16
	1/5 1%	6	12	0.068	0.79	0.24

CPSF7						
	0/5 1%	7	23	0.29	0.59	0.46
	Best Burden				0.3	
	Best SKAT					0.34
CCDC113	LofTee	1	2	1.	0.32	0.32
	5/5	7	24	-0.077	0.78	0.11
	5/5 + LofTee LC	7	24	-0.077	0.78	0.11
	1/5 1%	21	241	-0.72	0.4	0.5
	0/5 1%	37	413	-1.5	0.22	0.67
	Best Burden				0.61	
	Best SKAT					0.36
SLC16A2	5/5	1	28	-1.6	0.2	0.2
	5/5 + LofTee LC	1	28	-1.6	0.2	0.2
	1/5 1%	12	54	-0.058	0.81	0.15
	0/5 1%	15	63	0.031	0.86	0.14
	Best Burden				0.49	
	Best SKAT					0.37
SLC48A1	1/5 1%	15	52	0.86	0.35	0.22
	0/5 1%	19	58	0.17	0.68	0.23
	Best Burden				0.57	
	Best SKAT					0.38
PHF12	LofTee	2	2	0.2	0.65	0.44
	16/16	3	3	0.037	0.85	0.57
	11/11	14	24	-0.82	0.36	0.12
	5/5	14	24	-0.82	0.36	0.12
	5/5 + LofTee LC	14	24	-0.82	0.36	0.12
	1/5 1%	28	71	-0.91	0.34	0.26
	0/5 1%	32	82	-1.6	0.21	0.26
	Best Burden				0.62	
	Best SKAT					0.39
HLCS	LofTee	6	7	0.57	0.45	0.077
	16/16	6	7	0.57	0.45	0.077
	11/11	21	34	0.38	0.6	0.29
	5/5	25	44	1.5	0.32	0.26
	5/5 + LofTee LC	25	44	1.5	0.27	0.22
	1/5 1%	64	283	-0.16	0.75	0.69
	0/5 1%	93	714	0.63	0.5	0.2
	Best Burden				0.79	
	Best SKAT					0.4
ING3	11/11	2	2	0.61	0.44	0.21
	5/5	5	5	-0.049	0.83	0.53
	5/5 + LofTee LC	5	5	-0.049	0.83	0.53
	1/5 1%	14	19	-0.38	0.54	0.35
	0/5 1%	16	42	0.59	0.44	0.12
	Best Burden				0.9	
	Best SKAT					0.41
SLC30A8	LofTee	8	15	7.3e-05	0.99	0.89
	16/16	8	15	7.3e-05	0.99	0.89
	11/11	8	15	7.3e-05	0.99	0.89
	5/5	23	69	-2.4	0.12	0.83
	5/5 + LofTee LC	23	69	-2.4	0.12	0.83
	1/5 1%	36	137	-0.91	0.34	0.67
	0/5 1%	46	152	-1.2	0.27	0.69
	Best Burden				0.41	
	Best SKAT					0.99
	LofTee	6	7	3.	0.081	0.21
	16/16	9	13	2.1	0.15	0.55
	11/11	32	44	-0.067	0.8	0.9
	5/5	37	49	0.073	0.85	0.84
	5/5 + LofTee LC	38	50	0.038	0.84	0.79
	1/5 1%	112	523	-0.16	0.69	0.66
	0/5 1%	133	654	-0.47	0.49	0.64
	Best Burden				0.41	

ABCB11						
	Best SKAT					0.76
IL18BP	LofTee	1	1	-0.38	0.54	0.54
	16/16	1	1	-0.38	0.54	0.54
	11/11	1	1	-0.38	0.54	0.54
	5/5	42	200	-0.33	0.57	0.74
	5/5 + LofTee LC	42	200	-0.33	0.57	0.74
	1/5 1%	82	633	-2.1	0.15	0.53

Supplementary Table 8: **Evaluation of association signals in GHS.** Shown are results from the precomputed gene-level analysis of the GHS dataset. As custom analytical results were unavailable, the precise masks and testing methodologies are only broadly similar to those used in our exome-wide gene-level analysis. Genes are sorted in order of increasing *p*-value in the GHS dataset, reading from left to right and then top to bottom. The top 50 genes were advanced for analysis in the GHS dataset, but only results for the top 44 genes were available. Mask: the grouping of alleles used in the GHS analysis; Var: the number of alleles in the mask; CAF: the combined allele frequency of all alleles in the mask; OR: the aggregate odds-ratio calculated from a burden analysis of the mask; Burden: the *p*-value from a burden analysis of the mask. M1: predicted loss-of-function variants, according to the Variant Effect Predictor, with MAF$<$1% (similar to the LofTee mask but without an additional filter on LofTee and with an additional filter on MAF); M2: nonsynonymous variants predicted deleterious by 5/5 prediction algorithms with MAF$<$1% (similar to the 5/5 mask but with an additional filter on MAF); M3: all nonsynonymous variants predicted deleterious by $\geq$1/5 bioinformatic algorithms with MAF$<$1% (similar to the 1/5 1% mask); M4: all nonsynonymous variants with MAF$<$1% (similar to the 0/5 1% mask); Best: the minimum *p*-value calculated across all four masks, as described in **Methods**.

Gene	Mask	Var	CAF	OR	Burden	Gene	Mask	Var	CAF	OR	Burden
MC4R	M1	12	0.00035	2.3	0.022	*SLC30A8*	M1	13	0.00062	0.77	0.44
	M2	58	0.0021	1.6	0.0018		M2	46	0.0018	0.61	0.013
	M3	86	0.0039	1.2	0.057		M3	79	0.0037	0.8	0.088
	M4	94	0.0041	1.2	0.059		M4	94	0.004	0.81	0.085
	Best				0.0058		Best				0.043
ANGPTL6	M1	18	0.0015	1.2	0.23	*PAM*	M1	18	0.00025	1.4	0.41
	M2	59	0.0031	0.83	0.17		M2	98	0.011	1.2	0.032
	M3	111	0.0062	0.79	0.019		M3	209	0.016	1.1	0.11
	M4	127	0.007	0.82	0.034		M4	220	0.016	1.1	0.17
	Best				0.058		Best				0.096
MGAT4C	M1	8	0.00012	4.	0.032	*PDF*	M1	9	0.00015	0.41	0.25
	M2	34	0.0009	0.75	0.29		M2	9	0.00015	0.41	0.25
	M3	88	0.0051	0.9	0.35		M3	46	0.0012	0.67	0.085
	M4	96	0.0054	0.92	0.42		M4	50	0.0014	0.8	0.28
	Best				0.1		Best				0.18
DPH7	M1	18	0.00019	2.4	0.066	*STAT4*	M1	6	9.2e-05	1.7	0.43
	M2	51	0.001	0.9	0.66		M2	43	0.0015	0.94	0.75
	M3	94	0.005	1.1	0.64		M3	83	0.0089	0.86	0.076
	M4	127	0.0064	1.1	0.5		M4	95	0.0091	0.88	0.12
	Best				0.22		Best				0.24
TRDN	M1	53	0.0033	0.94	0.67	*HNF1A*	M1	11	0.0024	1.2	0.2
	M2	73	0.0042	0.92	0.49		M2	59	0.0041	1.2	0.09
	M3	185	0.01	0.89	0.13		M3	141	0.0091	1.1	0.18
	M4	230	0.012	0.89	0.09		M4	163	0.01	1.1	0.42
	Best				0.25		Best				0.26
TMEM161B	M1	8	0.00017	1.1	0.9	*PRSS54*	M1	10	0.00013	0.56	0.46
	M2	35	0.00086	1.	0.97		M2	22	0.00085	0.63	0.11
	M3	82	0.013	0.89	0.099		M3	68	0.011	0.88	0.084
	M4	83	0.013	0.89	0.094		M4	92	0.011	0.89	0.1
	Best				0.26		Best				0.27
MIEF1	M1	10	0.00026	2.	0.11	*KDM7A*	M1	6	0.0001	1.8	0.42
	M2	55	0.012	1.1	0.11		M2	39	0.0011	1.2	0.4
	M3	127	0.022	1.	0.36		M3	132	0.0094	1.1	0.2
	M4	133	0.022	1.1	0.31		M4	167	0.01	1.1	0.11
	Best				0.29		Best				0.34

	M1	7	8.1e-05	2.9	0.14		M1	5	4.1e-05	0.65	0.7
	M2	10	0.00017	1.2	0.75		M2	23	0.0071	1.1	0.46
OR2M4	M3	64	0.0082	1.	0.56	*TDRD5*	M3	124	0.0096	1.	0.72
	M4	89	0.011	0.99	0.94		M4	187	0.017	1.1	0.16
	Best				0.41		Best				0.46
	M1	12	0.00053	1.	0.91		M1	25	0.00039	1.3	0.49
	M2	37	0.0015	1.2	0.3		M2	194	0.013	0.96	0.57
NT5DC1	M3	80	0.007	1.1	0.35	*NUMA1*	M3	504	0.041	0.97	0.47
	M4	104	0.0079	1.1	0.17		M4	582	0.058	0.96	0.17
	Best				0.47		Best				0.48

Supplementary Table 9: **Evaluation of association signal concordance in CHARGE and GHS.** For each of the 50 genes from our exome sequence analysis with lowest gene-level *p*-value, we compared the direction of effect from our burden analysis to those in the CHARGE and GHS datasets. In this analysis, we only included genes which achieved $p<0.05$ for the burden test (i.e. we excluded genes significant under SKAT but not the burden test). For each gene, we took the direction of effect from the mask achieving the lowest *p*-value and compared it to the direction of effect in the analogous mask in CHARGE or GHS (for GHS, as discussed in **Methods**, we matched the LofTee mask to M1; the 15/15, 10/10, 5/5, and 5/5+LofTee LC mask to M2; the 1/5 1% mask to M3; and the 0/5 1% mask to M4). Best Test, $\log(OR)$, and P: the test with the lowest *p*-value within the chosen mask as well as the logarithm of the estimated odds-ratio and *p*-value; for genes in which the lowest *p*-value is achieved by the SKAT test, no direction of effect is shown and no comparison with CHARGE and GHS is performed (genes achieving $p<0.05$ for both SKAT and burden analyses are shown as two separate rows of the table). CHARGE Var (CAC, Score, P): the number of alleles (combined allele count, score statistic, and *p*-value) in the analogous mask in the CHARGE analysis. GHS Var (CAF, $\log(OR)$, P): the number of alleles (combined allele frequency, logarithm of odds-ratio, *p*-value) in the matched mask in the GHS analysis.

Gene	**Best**				**CHARGE**				**GHS**			
	Mask	**Test**	**log(OR)**	**P**	**Var**	**CAC**	**Score**	**P**	**Var**	**CAF**	**log(OR)**	**P**
MC4R	5/5 + LofTee LC	Burden	0.726	8.46e-11	12	18	**4.75**	0.0294	58	0.00212	**0.477**	0.00184
PAM	5/5 + LofTee LC	Burden	0.269	4.28e-09	40	1201	**1.86**	0.173	98	0.0106	**0.153**	0.0324
SLC30A8	1/5 1%	Burden	-0.514	4.7e-08	36	137	**-0.912**	0.34	79	0.0037	**-0.223**	0.0881
BICD1	0/5 1%	SKAT	-	2.83e-06	75	174	-	0.886	-	-	-	-
HNF1A	0/5 1%	SKAT	-	8.26e-06	88	271	-	0.471	-	-	-	-
UBE2NL	5/5 + LofTee LC	Burden	2.54	1.03e-05	3	4	**5.7**	0.017	-	-	-	-
MAP3K15	0/5 1%	Burden	-0.163	1.38e-05	92	493	**-2.66**	0.103	-	-	-	-
NUMA1	5/5	SKAT	-	2.47e-05	67	326	-	0.36	-	-	-	-
ING3	1/5 1%	Burden	0.829	3.1e-05	14	19	-0.382	0.536	44	0.00125	-0.187	0.387
IGFBPL1	1/5 1%	Burden	-0.573	3.59e-05	18	236	1.48	0.224	41	0.00359	**-0.0434**	0.733
MGAT4C	5/5 + LofTee LC	Burden	1.14	4.35e-05	11	15	**4.76**	0.0291	34	0.000904	-0.285	0.285
PTPRC	0/5 1%	Burden	0.21	7.35e-05	124	848	**0.284**	0.594	255	0.0232	**0.0266**	0.595
TMEM216	LofTee	Burden	0.0803	8.36e-05	4	10990	**2.61**	0.106	7	0.000183	**0.308**	0.587
HLCS	5/5 + LofTee LC	Burden	0.648	9.84e-05	27	48	**1.52**	0.218	49	0.00149	-0.121	0.54
SOCS2	0/5 1%	Burden	0.671	0.000102	7	13	-0.373	0.542	42	0.00109	**0.291**	0.193
TRDN	0/5 1%	Burden	0.247	0.000114	94	806	**0.184**	0.668	230	0.0119	-0.12	0.0902
TCEA1	5/5 + LofTee LC	Burden	-2.36	0.000157	4	4	**-1.09**	0.296	19	0.000274	**-0.375**	0.422
KDM7A	0/5 1%	Burden	0.315	0.000157	53	324	-0.0166	0.898	167	0.0104	**0.115**	0.113
PRSS54	11/11	SKAT	-	0.000183	1	1	-	0.461	-	-	-	-
DPH7	5/5 + LofTee LC	SKAT	-	0.000204	16	51	-	0.789	-	-	-	-
DHRS13	1/5 1%	SKAT	-	0.000211	44	118	-	0.463	-	-	-	-
ANGPTL6	5/5 + LofTee LC	SKAT	-	0.000222	22	45	-	0.997	-	-	-	-
OR2M4	5/5 + LofTee LC	Burden	-0.59	0.000248	7	56	0.526	0.468	10	0.000173	0.174	0.75
DGKK	0/5 1%	SKAT	-	0.000252	71	304	-	0.961	-	-	-	-
SLC16A2	0/5 1%	Burden	-0.0674	0.000266	15	63	0.031	0.86	-	-	-	-
HYAL2	1/5 1%	Burden	-0.687	0.000268	28	48	**-0.29**	0.59	91	0.00391	0.003	0.979
MAGEB5	0/5 1%	Burden	-0.635	0.000276	18	34	**-0.443**	0.506	-	-	-	-
FLCN	1/5 1%	Burden	-0.354	0.000277	19	57	**-0.0389**	0.844	138	0.00907	**-0.025**	0.754
DHX9	5/5 + LofTee LC	Burden	-2.31	0.000314	5	6	0.514	0.473	27	0.00375	0.128	0.293
MIEF1	1/5 1%	Burden	0.161	0.000317	35	551	**0.121**	0.728	127	0.0216	**0.0469**	0.363

MAGEE2	0/5 1%	Burden	0.0655	0.000334	57	230	-0.298	0.585	-	-	-	-
ARHGEF7	0/5 1%	SKAT	-	0.000338	60	330	-	0.835	-	-	-	-
PDF	0/5 1%	SKAT	-	0.000361	7	21	-	0.754	-	-	-	-
RUVBL2	1/5 1%	Burden	0.458	0.000433	25	286	**0.00663**	0.935	83	0.00684	-0.0402	0.66
C19orf66	1/5 1%	Burden	0.529	0.000483	22	52	**0.00353**	0.953	61	0.00224	**0.12**	0.442
SLC48A1	1/5 1%	Burden	1.17	0.000526	15	52	**0.861**	0.353	38	0.00149	**0.0583**	0.76
MEST	1/5 1%	Burden	0.908	0.000573	16	114	**0.0777**	0.78	41	0.00128	-0.0574	0.782
ANKS3	1/5 1%	Burden	-0.213	0.000617	69	400	2.43	0.119	227	0.0131	0.0373	0.566
TDRD5	5/5 + LofTee LC	Burden	-0.395	0.000648	16	156	0.00304	0.956	23	0.00709	0.0658	0.457
LRRTM3	1/5 1%	SKAT	-	0.000678	32	139	-	0.476	-	-	-	-
STAT4	11/11	SKAT	-	0.000706	7	24	-	0.422	-	-	-	-
LRRTM3	1/5 1%	Burden	0.367	0.00078	32	139	**5.08**	0.0242	83	0.00642	-0.0624	0.514
NT5DC1	1/5 1%	SKAT	-	0.000879	42	176	-	0.369	-	-	-	-
PPP3CA	1/5 1%	SKAT	-	0.000908	19	27	-	0.747	-	-	-	-
MBD3	5/5 + LofTee LC	Burden	1.37	0.00111	4	10	-0.843	0.359	7	9.15e-05	**0.347**	0.635
RXRG	5/5 + LofTee LC	Burden	1.22	0.00116	7	9	-1.05	0.307	32	0.000844	**0.276**	0.273
STAT4	1/5 1%	Burden	-0.39	0.0015	40	382	**-0.095**	0.758	83	0.00887	**-0.146**	0.0762
TMEM161B	1/5 1%	Burden	0.445	0.0016	27	204	-2.41	0.121	82	0.0129	-0.111	0.0991
RHBDD2	5/5 + LofTee LC	SKAT	-	0.0019	6	14	-	0.097	-	-	-	-
PDF	0/5 1%	Burden	-0.549	0.00247	7	21	**-0.0179**	0.894	50	0.00144	**-0.22**	0.279
ABCB11	1/5 1%	Burden	0.188	0.00275	112	523	-0.159	0.69	229	0.0113	-0.00955	0.894
NUMA1	5/5	Burden	0.13	0.00481	67	326	**0.221**	0.638	194	0.0129	-0.0378	0.571
HNF1A	1/5 1%	Burden	0.206	0.00504	71	218	**1.56**	0.212	141	0.00906	**0.104**	0.184
PDX1	11/11	Burden	1.55	0.00546	1	4	**1.29**	0.257	30	0.00314	**0.126**	0.339
ANGPTL6	5/5 + LofTee LC	Burden	-0.315	0.013	22	45	**-0.658**	0.417	59	0.00314	**-0.188**	0.168
DHRS13	16/16	Burden	0.508	0.0215	7	32	**0.457**	0.499	34	0.0013	-0.0998	0.637
RHBDD2	1/5 1%	Burden	-0.327	0.0223	27	45	**-0.00373**	0.951	83	0.00649	0.0363	0.695
DGKK	0/5 1%	Burden	-0.148	0.0254	71	304	**-0.0132**	0.909	-	-	-	-
NT5DC1	1/5 1%	Burden	0.194	0.0298	42	176	**0.0101**	0.92	80	0.00699	**0.0825**	0.347
BICD1	0/5 1%	Burden	-0.154	0.0448	75	174	**-2.13**	0.144	181	0.00566	**-0.0651**	0.519
PPP3CA	0/5 1%	Burden	-0.294	0.0483	19	27	0.0708	0.79	48	0.000955	**-0.19**	0.449

Supplementary Table 10: **Genes used in gene set analysis.** We selected various sets of genes, as described in **Methods**, to test for stronger-than-expected gene-level associations. Shown are the set of genes used in each gene set.

[See separate Excel file]

Supplementary Table 11: **Genes within T2D GWAS loci with nominally significant gene-level associations.** Shown are all genes within established T2D GWAS loci that achieved a $p<0.05$ for the minimum p-value burden analysis. Columns are analogous to those in **Supplementary Table 5**. Locus: an identifier for the T2D GWAS locus containing the gene.

				Burden				SKAT	
		Best Result		Min P		Weighted		Min P	Weighted
Gene	Locus	Var	CAF	OR	P	OR	P	P	P
MC4R	*MC4R*	41	0.00795	2.07	2.74e-10	2.2	4.81e-09	7.74e-08	3.48e-08
PAM	*PAM*	79	0.0493	1.31	1.58e-08	1.44	2.2e-09	1.53e-07	7.03e-08
SLC30A8	*SLC30A8*	86	0.0116	0.598	1.85e-07	0.397	1.29e-08	0.00011	0.000221
MNT	*SRR*	105	0.00939	0.708	0.00131	0.444	0.0151	0.962	0.922
ZFP1	*BCAR1*	120	0.0134	0.762	0.00664	0.552	0.00755	0.157	0.0561
SLC30A3	*GCKR*	74	0.00823	1.44	0.00699	2.19	0.00808	0.189	0.00168
RASGRP1	*RASGRP1*	22	0.000812	3.1	0.00879	1.63	0.123	0.856	0.693
TH	*IGF2*	35	0.00223	0.509	0.00879	0.641	0.003	0.0601	0.0121
WFS1	*WFS1*	120	0.146	1.09	0.0118	1.12	0.000944	0.0261	0.00167
ACADS	*HNF1A*	53	0.00684	1.44	0.013	1.41	0.0175	0.472	0.719
NUDCD3	*GCK*	60	0.0048	1.5	0.013	1.77	0.0239	0.825	0.758
FSCN3	*GCC1/PAX4*	35	0.0029	1.7	0.0145	1.25	0.115	0.123	0.237
BCL11A	*BCL11A*	115	0.00737	1.35	0.0166	1.73	0.0466	0.413	0.263
INS-IGF2	*IGF2*	43	0.00728	0.743	0.0187	0.462	0.0708	0.239	0.179
TMEM19	*LGR5*	73	0.0164	1.25	0.0213	1.52	0.00955	0.154	0.0752
PDX1	*PDX1*	11	0.000371	4.71	0.0214	3.46	0.000166	0.165	0.0573
SYN2	*PPARG*	36	0.00167	0.5	0.0214	0.689	0.138	0.556	0.274
HNF1A	*HNF1A*	131	0.0184	1.23	0.0219	1.47	0.0125	3.62e-05	0.00106
IGF2	*IGF2*	53	0.00457	0.66	0.0235	0.264	0.0912	0.867	0.649
PHLPP1	*BCL2*	271	0.0501	0.898	0.0236	0.822	0.156	0.261	0.182
PAX4	*GCC1/PAX4*	11	0.00151	2.16	0.0237	1.14	0.176	0.342	0.34
HIBADH	*JAZF1*	65	0.00459	1.38	0.0265	1.98	0.0551	0.255	0.24
POLM	*GCK*	180	0.0597	1.12	0.0266	1.1	0.101	0.704	0.379
MEF2B	*CILP2*	20	0.00162	1.92	0.0294	1.35	0.349	0.698	0.7
VPS33B	*PRC1*	23	0.0013	1.97	0.0306	1.21	0.193	0.025	0.577
PLEKHH2	*THADA*	146	0.0185	1.21	0.031	1.22	0.0167	0.165	0.0441
CCND2	*CCND2*	50	0.00408	1.47	0.0329	1.34	0.741	0.0632	0.0971
MYCBP	*MACF1*	26	0.00339	0.643	0.0344	0.593	0.0373	0.0999	0.0653
FAM135A	*C6orf57*	7	0.000348	4.39	0.0371	1.08	0.46	0.0627	0.113
DNLZ	*GPSM1*	40	0.00821	1.32	0.0373	1.77	0.0613	0.386	0.159
ATG16L2	*ARAP1*	166	0.0293	1.17	0.038	1.26	0.0101	0.982	0.868
UBE2E2	*UBE2E2*	23	0.00371	1.5	0.0391	1.51	0.0355	0.879	0.733
BORCS8-MEF2B	*CILP2*	21	0.00165	1.96	0.0397	1.36	0.275	0.888	0.634
ATXN7	*PSMD6*	199	0.106	1.08	0.0428	1.12	0.0192	0.0651	0.0288
LPP	*LPP*	160	0.0284	0.855	0.0432	0.754	0.00952	0.592	0.132
ZNF14	*CILP2*	106	0.0109	0.794	0.0432	0.728	0.0392	0.911	0.842
FAH	*ZFAND6*	119	0.0447	1.15	0.0437	1.16	0.0399	0.00492	0.00104
PHGDH	*NOTCH2*	50	0.0115	0.791	0.0445	0.792	0.0304	0.576	0.233
GLP1R	*KCNK16*	17	0.00116	1.97	0.0465	1.72	0.0194	0.0738	0.0171
ST20-MTHFS	*ZFAND6*	48	0.00631	0.742	0.0468	0.78	0.31	0.818	0.587

Supplementary Table 12: **Sample and variant counts for imputed GWAS analysis.** Shown are the sample subgroups with SNP array data analyzed as part of the imputed GWAS analysis. Subgroup, Ancestry, Sequence tech: exome sequencing subgroup characteristics. SNP array: technology used for imputed GWAS genotyping. Samples (Cases, Ctrls): Number of samples (T2D cases, controls) included in imputed GWAS analysis. Variants: Number of variants passing quality control and included in imputed GWAS analysis. Prior to analysis, all subgroups had genotypes imputed from the 1000G Phase 3 reference panel.

Subgroup	Ancestry	Sequence tech	SNP array	Samples	Cases	Ctrls	Variants
Wake Forest	African-American	Agilent	Affy6	1053	531	522	27,973,694
JHS	African-American	Agilent	Affy6	989	513	476	27,748,674
BioMe	African-American	Illumina	G4L	2518	1233	1285	26,084,594
Singapore	East-Asian	Agilent	Illumina610/Illumina1Mdov3	1077	591	486	10,897,305
Singapore	East-Asian	Illumina	G4L	1969	985	984	9,209,756
KARE	East-Asian	Agilent	Affy5	1096	567	529	7,770,075
SNUH	East-Asian	Illumina	G4L	917	474	443	8,839,178
Hong Kong	East-Asian	Illumina	G4L	960	481	479	9,231,055
GoT2D	European	Agilent	HumanOmni2.5	2657	1326	1331	17,692,443
METSIM	European	Agilent	HumanOmni2.5	970	494	476	14,135,914
Ashkenazi	European	Agilent	Illumina Cardio-MetaboChip	732	359	373	2,602,793
GoDARTS	European	Illumina	G4L	1886	941	945	12,197,555
FHS	European	Illumina	G4L	973	584	389	10,939,434
SIGMA	Latino	Agilent	HumanOmni2.5	3542	1712	1830	35,256,845
SIGMA	Latino	Illumina	G4L	5851	3020	2831	19,591,358
Starr County	Hispanic	Agilent	Affy6	1383	673	710	20,401,781
Starr County	Hispanic	Illumina	G4L	933	608	325	16,157,686
San Antonio	Hispanic	Agilent	Illumina Cardio-MetaboChip	445	202	243	2,996,739
Singapore	South-Asian	Agilent	Illumina610	1112	576	536	14,471,389
Singapore	South-Asian	Illumina	G4L	1932	882	1050	11,989,365
LOLIPOP	South-Asian	Agilent	Illumina610	1199	599	600	15,256,850
PGR	South-Asian	Illumina	G4L	1718	882	836	12,580,193

Supplementary Table 13: **Loci with most significant associations from imputed GWAS analysis.** Shown are the most significant associations from the imputed GWAS analysis, with only one association shown per 250kb of genomic sequence. Closest Gene: the closest gene to the variant. rsID: the dbSNP ID of the variant (as predicted by the Variant Effect Predictor), if applicable. Chrom/Position: the chromosome and position of the variant. E.A./O.A.: the effect and non-effect alleles of the variant. Samples: the number of samples analyzed for the variant (i.e. the number of samples within subgroups in which the variant was polymorphic and passed quality control). MAF: the minor allele frequency of the variant, calculated across all samples. OR: the estimated odds-ratio of the variant. P: the *p*-value of the variant.

Closest Gene	rsID	Chrom	Position	E.A.	O.A.	Samples	MAF	OR	P
TCF7L2	rs7903146	10	114758349	T	C	30683	0.25	1.34	1.48e-41
KCNQ1	rs2237896	11	2858440	A	G	27249	0.26	0.734	1.23e-32
CDC123	rs11257600	10	12255657	T	G	34529	0.31	1.16	6.14e-17
CDKAL1	rs9460550	6	20719561	A	G	34527	0.32	1.15	6.63e-15
SLC30A8	rs3802177	8	118185025	A	G	27745	0.31	0.862	7.7e-14
IGF2BP2	rs4414887	3	185506892	T	C	33555	0.32	1.13	3.72e-13
CTBP1	rs72501962	4	1246038	A	T	33797	0.28	1.15	7.56e-12
ASCL2	rs17737404	11	2270342	A	G	30237	0.23	1.22	3.7e-10
KCNJ11	rs10734252	11	17404839	A	G	34524	0.38	0.9	3.83e-10
HNF4A	rs6103716	20	42999630	A	C	34529	0.39	0.892	4.54e-10
KIF11	rs2153827	10	94424073	T	G	34442	0.34	1.11	9.97e-10
ZMIZ1	rs703978	10	80944147	C	G	34528	0.31	1.11	1.91e-09
IRS1	rs2943657	2	227123439	T	C	34529	0.26	1.12	2.82e-09
JAZF1	rs1635852	7	28189411	T	C	27745	0.34	1.12	3.67e-09
SFI1	rs2236033	22	32001050	A	G	34511	0.3	1.11	4.37e-09
GPSM1	rs376993806	9	139246588	A	G	29705	0.26	0.885	1.82e-08
SPRY2	rs1359790	13	80717156	A	G	34529	0.31	0.896	2.27e-08
EML4	-	2	42506923	T	TA	2657	0.44	1.43	3.02e-08
PPARG	rs4684848	3	12395645	A	G	34529	0.2	0.883	4.11e-08
WFS1	-	4	6301628	T	TTG	28470	0.21	1.14	4.45e-08
SOX11	rs896911	2	5376965	T	C	32372	0.38	0.908	5.84e-08
CCND2	rs76895963	12	4384844	T	G	4826	0.023	2.84	1.14e-07
AUTS2	-	7	69534694	G	GT	29821	0.23	0.878	1.17e-07
NTRK2	rs1573219	9	87387622	A	G	26568	0.3	0.903	2.1e-07
COBLL1	rs12692738	2	165558252	T	C	34529	0.23	1.12	2.25e-07

Supplementary Table 14: **Most significant nonsynonymous variants within T2D GWAS loci.** Shown are the 50 nonsynonymous variants within established T2D GWAS loci that achieved the lowest *p*-values in the exome sequence single-variant analysis. Columns are analogous to those in **Supplementary Table 4**. Locus: an identifier for the T2D GWAS locus containing the variant.

Gene	Locus	Variant	Consequence	Change	MAF	Case	Ctrl	OR	P
PAX4	*GCC1/PAX4*	rs2233580	missense_variant	p.Arg192His	0.12	890	563	1.7	7.6e-22
SLC30A8	*SLC30A8*	rs13266634	missense_variant	p.Arg325Trp	0.43	12258	13756	0.897	3.4e-11
WFS1	*WFS1*	rs1801212	missense_variant	p.Val333Ile	0.27	7101	8456	1.13	1.2e-10
KCNJ11	*KCNJ11*	rs5215	missense_variant	p.Val250Ile	0.39	16687	16132	0.901	3.4e-10
ABCC8	*KCNJ11*	rs757110	missense_variant	p.Ala1369Ser	0.39	16626	16237	0.913	7.1e-08
MC4R	*MC4R*	rs79783591	missense_variant	p.Ile269Asn	0.0089	195	83	2.17	3.4e-07
COBLL1	*COBLL1/GRB14*	rs7607980	missense_variant	p.Asn939Asp	0.15	4010	4651	0.857	6.3e-07
PAM	*PAM*	rs35658696	missense_variant	p.Asp563Gly	0.05	1038	944	1.29	1.3e-06
PPIP5K2	*PAM*	rs36046591	missense_variant	p.Ser1207Gly	0.049	986	905	1.3	1.4e-06
GCKR	*GCKR*	rs1260326	missense_variant	p.Leu446Pro	0.5	15010	16627	1.07	5.4e-06
TM6SF2	*CILP2*	rs58542926	missense_variant	p.Glu167Lys	0.1	2899	2694	1.14	2.6e-05
FES	*PRC1*	var_15_91434859	missense_variant	p.Pro536Ser	0.0071	63	24	1.82	3.1e-05
SENP2	*IGF2BP2*	rs6762208	missense_variant	p.Thr301Lys	0.49	18375	17667	1.07	3.8e-05
PPARG	*PPARG*	rs1801282	missense_variant	p.Pro12Ala	0.12	4241	4935	0.894	6.5e-05
HNF1A	*HNF1A*	var_12_121437091	missense_variant	p.Glu508Lys	0.006	93	33	2.25	9.1e-05
TCF19	*HLA-B*	rs2073721	missense_variant	p.Met211Val	0.3	11485	11721	1.07	0.00013
NUCB2	*KCNJ11*	rs757081	missense_variant	p.Gln338Glu	0.36	13223	13362	1.08	0.00017
RREB1	*SSR1/RREB1*	rs9379084	missense_variant	p.Asp1171Asn	0.13	3641	4117	0.916	0.00026
NUCB2	*KCNJ11*	rs3842269	inframe_deletion	p.401-402LeuGln/Leu	0.5	10255	11247	0.934	0.0003
GPSM1	*GPSM1*	var_9_139235415	missense_variant	p.391-392SerGlu/LeuGlu	0.28	8404	8458	0.931	0.00033
C6orf136	*POU5F1/TCF19*	rs150233869	missense_variant	p.Arg220Cys	0.0086	69	92	0.527	0.00038
GTF2H4	*POU5F1/TCF19*	rs140816086	missense_variant	p.Arg453His	0.001	16	34	0.476	0.00043
SDCCAG3	*GPSM1*	rs1131992	missense_variant	p.Val356Met	0.26	7300	7616	0.928	0.00043
THADA	*THADA*	rs35720761	missense_variant	p.Cys1605Tyr	0.15	4491	4736	0.927	0.00048
WFS1	*WFS1*	rs734312	missense_variant	p.Arg611His	0.89	21467	21951	1.05	0.0005
RBL2	*FTO*	rs199555150	missense_variant	p.Ser995Gly	0.01	51	97	0.619	0.0005
HNF1A	*HNF1A*	rs1800574	missense_variant	p.Ala98Val	0.061	1059	915	1.17	0.00054
VPS33B	*PRC1*	rs11073964	missense_variant	p.Gly514Ser	0.58	13333	15422	1.07	0.00071
NOTCH1	*GPSM1*	rs61751489	missense_variant	p.Val2285Ile	0.16	2748	2756	0.882	0.00079
SDCCAG3	*GPSM1*	rs3812577	missense_variant	p.Arg281Gln	0.26	7124	7422	0.929	0.00086
C6orf15	*HLA-B*	rs2233977	missense_variant	p.Val81Ala	0.37	9445	9951	0.947	0.00086
ACMSD	*TMEM163*	var_2_135621062	missense_variant	p.Thr116Met	0.0032	9	29	0.462	0.00089
IFT172	*GCKR*	rs139229844	missense_variant	p.Gln866Arg	0.0017	27	5	3.	0.00089
LST1	*HLA-B*	rs184203129	missense_variant	p.Ala46Thr	0.00067	16	4	2.96	0.00093
TM6SF2	*CILP2*	rs187429064	missense_variant	p.Leu156Pro	0.019	415	363	1.3	0.00093
SLC30A8	*SLC30A8*	rs73317647	missense_variant	p.Arg165Cys	0.0061	34	70	0.516	0.001
MUC22	*POU5F1*	rs117024916	missense_variant	p.Thr71Ala	0.013	80	112	0.638	0.0012
FAT3	*MTNR1B*	var_11_92620216	missense_variant	p.Lys665Glu	0.0013	9	31	0.51	0.0015
GATAD2A	*CILP2*	rs370240766	inframe_insertion	p.67Thr/ThrAlaMet	0.00066	4	15	0.464	0.0016
RCCD1	*PRC1*	rs75390535	missense_variant	p.Leu249Val	0.099	583	701	0.828	0.0016
MC4R	*MC4R*	var_18_58038669	missense_variant	p.Arg305Gln	0.00033	8	0	4.89	0.0017
FSCN3	*GCC1/PAX4*	rs144391719	missense_variant	p.Arg356His	0.00023	11	1	3.85	0.0018
FLT3	*PDX1*	rs62636526	missense_variant	p.Val16Leu	0.0089	94	52	1.7	0.0018
MDGA1	*ZFAND3*	rs143644874	missense_variant	p.Glu756Gln	0.0069	54	72	0.492	0.0018
WFS1	*WFS1*	rs1801208	missense_variant	p.Arg456His	0.091	3063	2828	1.1	0.0019
C6orf15	*HLA-B*	rs2233978	missense_variant	p.Ala145Pro	0.37	7733	7907	0.938	0.0019
SLC30A8	*SLC30A8*	rs145677283	missense_variant	p.Arg165His	0.0014	15	34	0.447	0.0021
LAMA1	*LAMA1*	rs115759032	missense_variant	p.Leu1932Val	0.006	33	60	0.564	0.0021
CARD9	*GPSM1*	rs4077515	missense_variant	p.Ser12Asn	0.54	19891	20405	0.956	0.0022
F2RL1	*ZBED3*	rs148584357	missense_variant	p.His135Arg	0.00026	8	0	5.74	0.0022

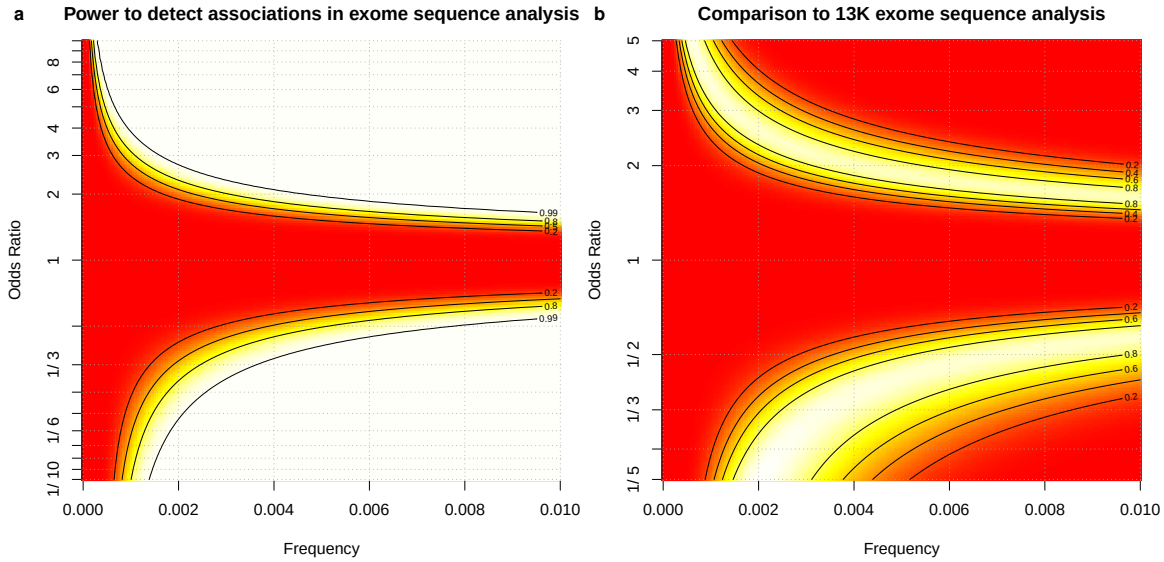
Supplementary Table 15: **Most significant protein-truncating variants within T2D GWAS loci.** Shown are all protein-truncating variants (as annotated by the Variant Effect Predictor) within established T2D GWAS loci that achieved $p<0.05$ in the exome sequence single-variant analysis. Columns are analogous to those in **Supplementary Table 14**.

Gene	Locus	Variant	Consequence	Change	MAF	Case	Ctrl	OR	P
STOX1	*VPS26A*	var_10_70644927	stop_gained	p.Gln459Ter	0.00066	8	0	4.85	0.0034
CRIPAK	*MAEA*	var_4_1388335	stop_gained	p.Cys12Ter	0.0045	65	101	0.657	0.004
PPM1N	*GIPR*	var_19_46005322	frameshift_variant	p.-106-107Ter	0.001	7	23	0.323	0.0046
CDSN	*HLA-B*	var_6_31084723	frameshift_variant	p.218-223CysSerSerAspIlePro/Ter	0.016	91	142	0.707	0.0064
THADA	*THADA*	var_2_43817961	splice_acceptor_variant	-	0.0004	14	9	0.238	0.0073
THADA	*THADA*	var_2_43817962	splice_acceptor_variant	-	0.0004	12	9	0.239	0.0075
IFT172	*GCKR*	rs150246251	stop_gained	p.Arg1507Ter	0.00024	4	0	5.34	0.011
ABCB9	*MPHOSPH9*	var_12_123419979	splice_acceptor_variant	-	0.00019	4	0	8.12	0.013
DPY19L4	*TP53INP1*	rs200830188	splice_acceptor_variant	-	0.0011	10	2	3.71	0.014
SNAPC4	*GPSM1*	var_9_139278009	frameshift_variant	p.Ser537Ter	0.00042	5	4	5.19	0.017
C11orf21	*IGF2*	rs3214127	frameshift_variant	p.-119-120Ter	0.034	166	130	1.29	0.017
SPATA31D5P	*TLE1*	var_9_84528573	splice_acceptor_variant	-	0.00034	0	8	0.0938	0.017
RBM28	*LEP*	var_7_127961426	stop_gained	p.Arg486Ter	0.00025	0	3	0.118	0.018
KRTCAP3	*GCKR*	rs140428163	frameshift_variant	p.119Leu/LeuTer	0.001	10	20	0.437	0.019
C15orf53	*RASGRP1*	var_15_38988925	frameshift_variant	p.39-40AlaSer/AlaTer	0.006	63	54	1.32	0.02
SNX17	*GCKR*	var_2_27599220	frameshift_variant	p.408-410AspSerGln/Ter	0.002	46	24	1.64	0.021
VPS13C	*C2CD4A/C2CD4B*	var_15_62169177	stop_gained	p.Glu3364Ter	0.00016	6	0	5.3	0.023
KCNQ1	*KCNQ1*	rs11601907	stop_gained	p.Tyr662Ter	0.27	4247	4392	1.08	0.025
VPS33B	*PRC1*	var_15_91553029	splice_acceptor_variant	-	0.00041	5	0	5.58	0.026
TIMP4	*PPARG*	var_3_12195162	stop_gained	p.Cys176Ter	0.00065	7	1	3.55	0.026
SLC16A13	*SLC16A11*	rs202121781	stop_gained	p.Arg282Ter	0.00017	5	0	2.65	0.028
PEX11A	*AP3S2*	var_15_90229661	splice_acceptor_variant	-	0.00012	4	0	4.3	0.028
BLM	*PRC1*	rs367543013	frameshift_variant	p.256-257AspSer/AspTer	0.00018	0	5	0.261	0.029
TIGD4	*TMEM154*	var_4_153691112	frameshift_variant	p.Phe348Ter	0.00072	5	10	0.432	0.035
DHTKD1	*CDC123/CAMK1D*	var_10_12159670	splice_acceptor_variant	-	0.0002	1	5	0.209	0.035
RHBDL2	*MACF1*	var_1_39381293	stop_gained	p.Tyr112Ter	0.0014	29	13	1.85	0.036
ATG16L2	*ARAP1*	var_11_72535166	splice_acceptor_variant	-	0.00013	4	0	8.64	0.036
COBLL1	*COBLL1/GRB14*	var_2_165551295	frameshift_variant	p.907Leu/PheTer	0.019	81	70	0.331	0.037
KIF6	*KCNK16*	rs202222855	stop_gained	p.Ser244Ter	0.00025	3	0	3.62	0.038
IGF2BP2	*IGF2BP2*	var_3_185375092	frameshift_variant	p.Phe456Ter	0.00033	7	1	4.09	0.039
SNAPC4	*GPSM1*	rs3812565	frameshift_variant	p.1259L/LPQPGPEKGALDLEX	0.46	15090	13994	0.958	0.039
ZNF14	*CILP2*	var_19_19822199	stop_gained	p.630-631PheArg/PheTer	0.00025	0	2	0.229	0.04
KRTAP5-5	*DUSP8*	var_11_1651596	frameshift_variant	p.176-194SSCCKPYCCQSSCCKPYCC/X	0.16	4304	4202	0.934	0.041
HMGCS2	*NOTCH2*	rs1048438	stop_gained	p.297Tyr/TerTyr	0.00025	0	3	0.215	0.043
FAM135A	*C6orf57*	var_6_71195923	stop_gained	p.Arg250Ter	0.00035	9	2	3.63	0.043
ZNF14	*CILP2*	var_19_19822283	stop_gained	p.Arg603Ter	0.00012	0	3	0.251	0.043
OTOG	*KCNJ11*	var_11_17631829	frameshift_variant	p.Ser679Ter	0.00042	0	7	0.155	0.044
NANOS2	*GIPR*	var_19_46417550	frameshift_variant	p.134Arg/ArgTer	0.0003	1	5	0.428	0.045
SPDYE1	*GCK*	var_7_44042207	frameshift_variant	p.93Ser/SerTer	9.8e-05	0	5	0.304	0.046
PMPCA	*GPSM1*	var_9_139311505	stop_gained	p.Tyr246Ter	0.00013	3	0	0.986	0.046
P2RX7	*HNF1A*	var_12_121603952	stop_gained	p.Arg236Ter	0.00016	1	6	0.353	0.047
PLIN1	*AP3S2*	var_15_90213359	stop_gained	p.Cys150Ter	0.00013	0	4	0.183	0.047
FGF6	*CCND2*	rs375467953	initiator_codon_variant	p.Met1Leu	0.00017	4	2	2.33	0.047
DDX52	*HNF1B*	var_17_35985993	stop_gained	p.Gln362Ter	0.00013	1	4	0.181	0.047
PRR3	*POU5F1/TCF19*	rs371871050	stop_gained	p.Arg171Ter	0.00082	8	2	2.51	0.048
CRIPAK	*MAEA*	rs373049641	stop_gained	p.Arg39Ter	0.006	111	98	0.805	0.048
VPS4B	*BCL2*	var_18_61067295	frameshift_variant	p.259Leu/ProTer	0.0002	6	0	6.53	0.048
KIF6	*KCNK16*	var_6_39330288	frameshift_variant	p.74Ser/CysTer	0.00033	2	4	0.261	0.049
RFC4	*ST6GAL1*	rs370046824	splice_acceptor_variant	-	0.00066	2	6	0.21	0.049
DHX16	*POU5F1/TCF19*	rs368358552	stop_gained	p.Gln370Ter	0.00012	2	0	4.61	0.049

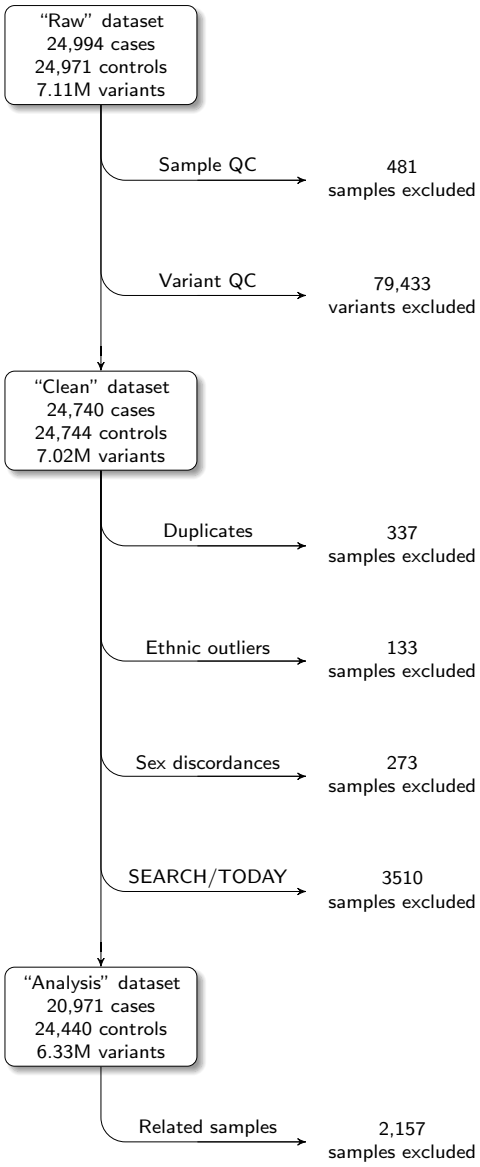
Supplementary Table 16: **Posterior probability conversion table.** Based on *p*-values from the exome sequence analysis for nonsynonymous variants within established T2D GWAS loci, together with an independent analysis of a subset of these variants on the Illumina Exome Array, we estimated the posterior probability of association for arbitrary nonsynonymous variants within the exome sequence analysis. The posterior probability estimates are a function of the observed *p*-value in the exome sequence analysis (rows in the table, with $-\log_{10}(p)$ shown in the first column) and the prior likelihood that the variant is associated with T2D. The prior likelihood, which quantifies belief in causal variant association before observing any results from our sequence analysis, can be specified in two ways. First (top two rows), via a "gene prior", or prior probability that loss of function of the gene is associated with T2D risk, which could be based on (for example) literature or experimental data implicating the gene in T2D pathogenesis. Second (third and fourth row), via a "variant prior", or the prior probability that the variant itself is associated with T2D risk. Calculations based on the gene prior (top two rows) use estimates from our allelic mask weights (**Methods**) that 33% of missense variants result in gene loss of function.

	Prior probability of gene relevance																	
	0.01	**0.07**	**0.13**	**0.19**	**0.25**	**0.31**	**0.37**	**0.43**	**0.49**	**0.55**	**0.61**	**0.67**	**0.73**	**0.79**	**0.85**	**0.91**	**0.97**
	Prior probability of variant association																
-log$_{10}$(p)	**0.003**	**0.023**	**0.04**	**0.06**	**0.08**	**0.1**	**0.12**	**0.14**	**0.16**	**0.18**	**0.2**	**0.22**	**0.24**	**0.26**	**0.28**	**0.3**	**0.32**
1.00	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30	0.32
1.01	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.23	0.25	0.27	0.29	0.31
1.02	0.00	0.02	0.04	0.06	0.08	0.09	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.24	0.26	0.28	0.30
1.03	0.00	0.02	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.29
1.04	0.00	0.02	0.04	0.05	0.07	0.09	0.11	0.12	0.14	0.16	0.18	0.20	0.21	0.23	0.25	0.27	0.29
1.05	0.00	0.02	0.04	0.05	0.07	0.09	0.11	0.12	0.14	0.16	0.18	0.19	0.21	0.23	0.25	0.27	0.28
1.06	0.00	0.02	0.04	0.05	0.07	0.09	0.10	0.12	0.14	0.16	0.17	0.19	0.21	0.23	0.25	0.26	0.28
1.07	0.00	0.02	0.04	0.05	0.07	0.09	0.10	0.12	0.14	0.16	0.17	0.19	0.21	0.23	0.25	0.26	0.28
1.08	0.00	0.02	0.04	0.05	0.07	0.09	0.10	0.12	0.14	0.16	0.18	0.19	0.21	0.23	0.25	0.27	0.28
1.09	0.00	0.02	0.04	0.05	0.07	0.09	0.11	0.12	0.14	0.16	0.18	0.20	0.21	0.23	0.25	0.27	0.29
1.10	0.00	0.02	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.29
1.11	0.00	0.02	0.04	0.06	0.08	0.10	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.25	0.27	0.29	0.30
1.12	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30	0.32
1.14	0.00	0.02	0.04	0.06	0.09	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.25	0.27	0.29	0.31	0.33
1.15	0.00	0.03	0.05	0.07	0.09	0.11	0.13	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30	0.32	0.34
1.16	0.00	0.03	0.05	0.07	0.10	0.12	0.14	0.16	0.19	0.21	0.23	0.25	0.27	0.30	0.32	0.34	0.36
1.17	0.00	0.03	0.05	0.08	0.10	0.12	0.15	0.17	0.19	0.22	0.24	0.26	0.28	0.31	0.33	0.35	0.37
1.19	0.00	0.03	0.06	0.08	0.11	0.13	0.15	0.18	0.20	0.23	0.25	0.27	0.29	0.32	0.34	0.36	0.38
1.20	0.00	0.03	0.06	0.08	0.11	0.13	0.16	0.18	0.21	0.23	0.26	0.28	0.30	0.32	0.35	0.37	0.39
1.22	0.00	0.03	0.06	0.09	0.11	0.14	0.16	0.19	0.21	0.24	0.26	0.28	0.31	0.33	0.35	0.37	0.40
1.23	0.00	0.03	0.06	0.09	0.11	0.14	0.17	0.19	0.22	0.24	0.26	0.29	0.31	0.33	0.36	0.38	0.40
1.25	0.00	0.03	0.06	0.09	0.12	0.14	0.17	0.19	0.22	0.24	0.27	0.29	0.32	0.34	0.36	0.38	0.41
1.26	0.00	0.03	0.06	0.09	0.12	0.14	0.17	0.20	0.22	0.25	0.27	0.29	0.32	0.34	0.36	0.39	0.41
1.28	0.00	0.03	0.06	0.09	0.12	0.14	0.17	0.20	0.22	0.25	0.27	0.29	0.32	0.34	0.36	0.39	0.41
1.30	0.00	0.03	0.06	0.09	0.12	0.14	0.17	0.20	0.22	0.25	0.27	0.30	0.32	0.34	0.37	0.39	0.41
1.32	0.00	0.03	0.06	0.09	0.12	0.15	0.17	0.20	0.22	0.25	0.27	0.30	0.32	0.35	0.37	0.39	0.41
1.34	0.00	0.03	0.06	0.09	0.12	0.15	0.17	0.20	0.22	0.25	0.27	0.30	0.32	0.35	0.37	0.39	0.41
1.36	0.01	0.03	0.06	0.09	0.12	0.15	0.18	0.20	0.23	0.25	0.28	0.30	0.33	0.35	0.37	0.40	0.42
1.38	0.01	0.04	0.07	0.10	0.12	0.15	0.18	0.21	0.23	0.26	0.28	0.31	0.33	0.36	0.38	0.40	0.42
1.40	0.01	0.04	0.07	0.10	0.13	0.16	0.18	0.21	0.24	0.26	0.29	0.31	0.34	0.36	0.39	0.41	0.43
1.42	0.01	0.04	0.07	0.10	0.13	0.16	0.19	0.21	0.24	0.27	0.29	0.32	0.34	0.37	0.39	0.41	0.44
1.45	0.01	0.04	0.07	0.10	0.14	0.17	0.20	0.22	0.25	0.28	0.31	0.33	0.36	0.38	0.40	0.43	0.45
1.48	0.01	0.04	0.08	0.11	0.14	0.17	0.20	0.23	0.26	0.29	0.32	0.34	0.37	0.39	0.42	0.44	0.46
1.50	0.01	0.04	0.08	0.11	0.15	0.18	0.21	0.24	0.27	0.30	0.33	0.35	0.38	0.40	0.43	0.45	0.47
1.53	0.01	0.05	0.08	0.12	0.15	0.19	0.22	0.25	0.28	0.31	0.34	0.36	0.39	0.42	0.44	0.46	0.49
1.57	0.01	0.05	0.09	0.12	0.16	0.19	0.23	0.26	0.29	0.32	0.34	0.37	0.40	0.42	0.45	0.47	0.50
1.60	0.01	0.05	0.09	0.13	0.16	0.20	0.23	0.26	0.29	0.32	0.35	0.38	0.40	0.43	0.45	0.48	0.50
1.64	0.01	0.05	0.09	0.13	0.17	0.20	0.24	0.27	0.30	0.33	0.36	0.39	0.41	0.44	0.46	0.49	0.51
1.69	0.01	0.05	0.09	0.13	0.17	0.20	0.24	0.27	0.30	0.33	0.36	0.39	0.41	0.44	0.47	0.49	0.51
1.74	0.01	0.05	0.09	0.13	0.17	0.21	0.24	0.27	0.31	0.34	0.37	0.39	0.42	0.45	0.47	0.50	0.52
1.79	0.01	0.05	0.10	0.14	0.18	0.22	0.25	0.29	0.32	0.35	0.38	0.41	0.44	0.46	0.49	0.51	0.54
1.85	0.01	0.06	0.11	0.15	0.19	0.23	0.27	0.31	0.34	0.37	0.40	0.43	0.46	0.48	0.51	0.53	0.56
1.93	0.01	0.07	0.12	0.17	0.21	0.25	0.29	0.33	0.36	0.40	0.43	0.46	0.49	0.51	0.54	0.56	0.58
2.01	0.01	0.07	0.13	0.18	0.23	0.28	0.32	0.36	0.39	0.43	0.46	0.49	0.52	0.54	0.57	0.59	0.61
2.08	0.01	0.08	0.14	0.20	0.25	0.30	0.34	0.38	0.41	0.45	0.48	0.51	0.54	0.56	0.59	0.61	0.63
2.15	0.01	0.08	0.15	0.21	0.26	0.31	0.35	0.39	0.43	0.47	0.50	0.53	0.55	0.58	0.60	0.63	0.65
2.25	0.01	0.09	0.15	0.21	0.27	0.32	0.36	0.40	0.44	0.47	0.51	0.54	0.56	0.59	0.61	0.64	0.66
2.37	0.01	0.10	0.17	0.23	0.29	0.34	0.39	0.43	0.47	0.50	0.53	0.56	0.59	0.61	0.64	0.66	0.68
2.54	0.02	0.11	0.18	0.25	0.31	0.36	0.41	0.45	0.49	0.53	0.56	0.59	0.61	0.64	0.66	0.68	0.70
2.78	0.02	0.12	0.20	0.27	0.33	0.39	0.43	0.48	0.52	0.55	0.58	0.61	0.64	0.66	0.68	0.70	0.72
3.64	0.02	0.13	0.21	0.29	0.35	0.41	0.46	0.50	0.54	0.57	0.60	0.63	0.66	0.68	0.70	0.72	0.74

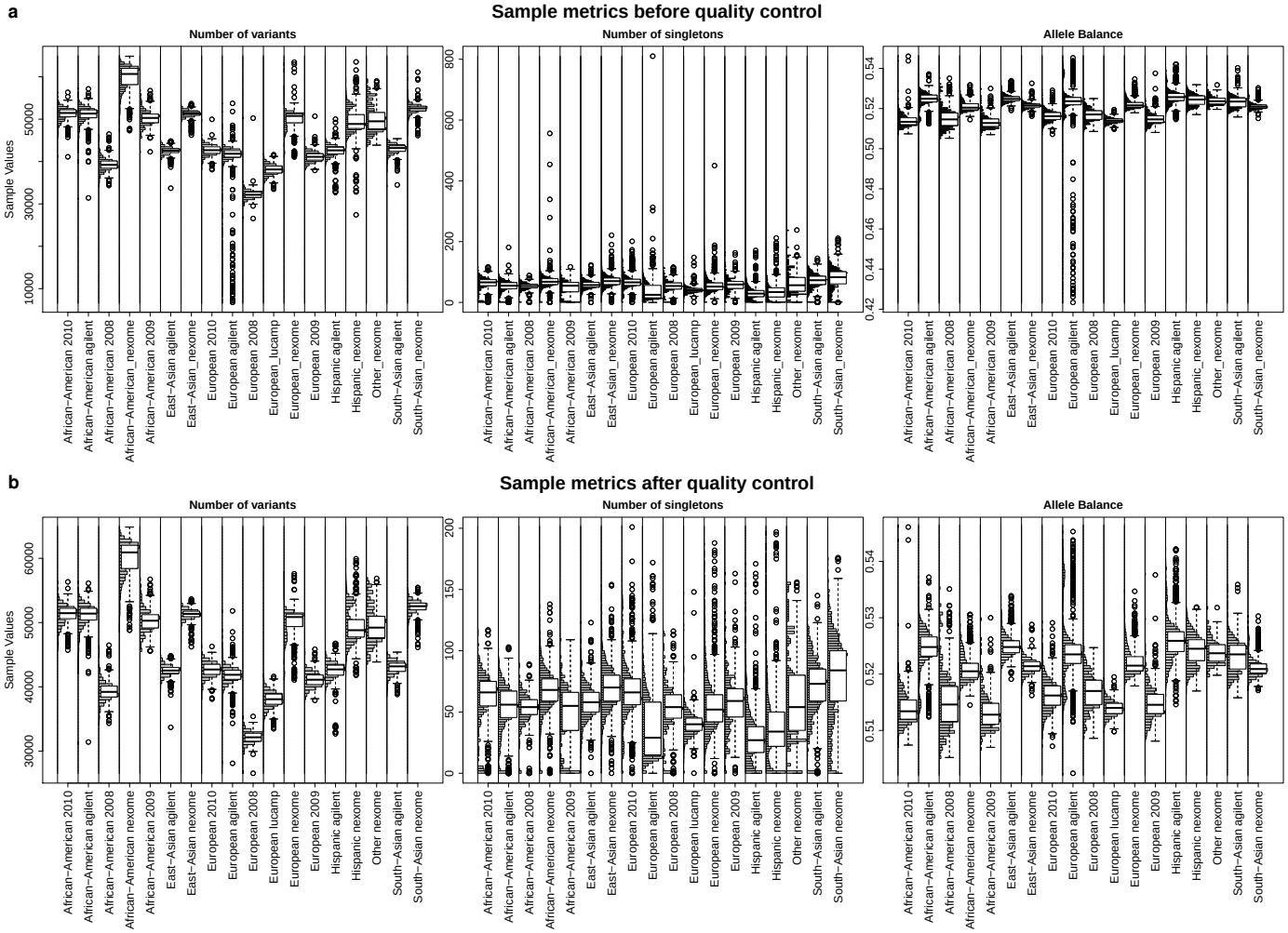
Supplementary Figures



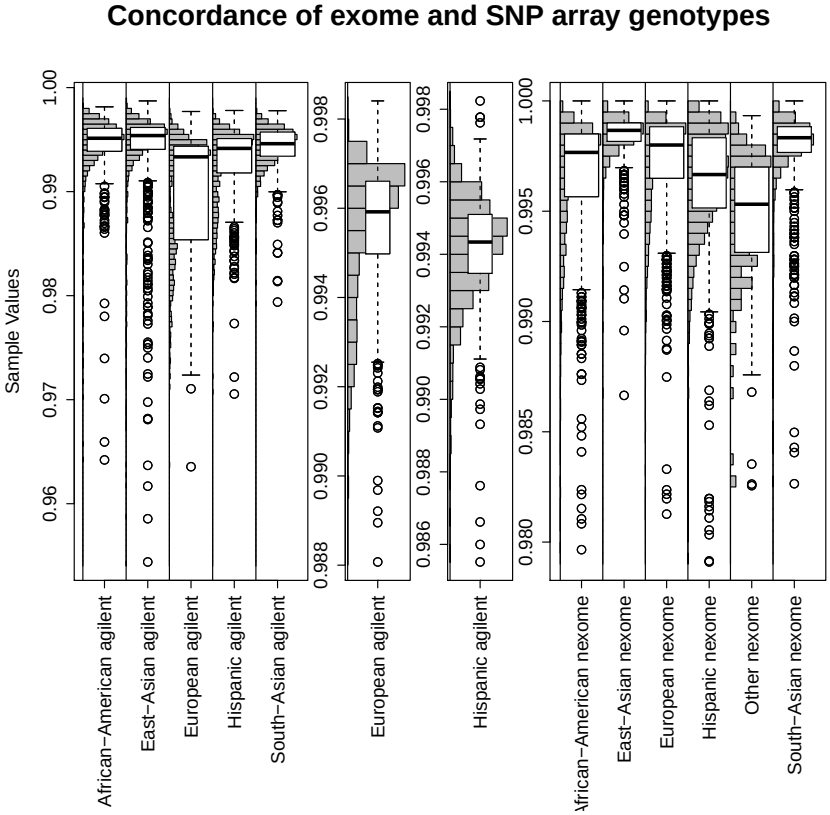
Supplementary Figure 1: **Power analysis.** Shown is the power to detect association below $p < 5 \times 10^{-8}$ for variants (or collections of variants) with a given minor allele frequency (x-axis) and odds ratio (y-axis) measured as the average across all ancestries. **(a)** Cells are shaded according to the power of the current study of 20,791 T2D cases and 24,440 controls, with white indicating high power and red indicating low power. **(b)** Cells are shaded according to the difference in power between the current study and a previously published study of 12,940 individuals [4], with yellow/white indicating a large increase in power and red indicating a small increase in power. For each plot, 20%, 50%, 80%, and 99% contour lines are labeled.



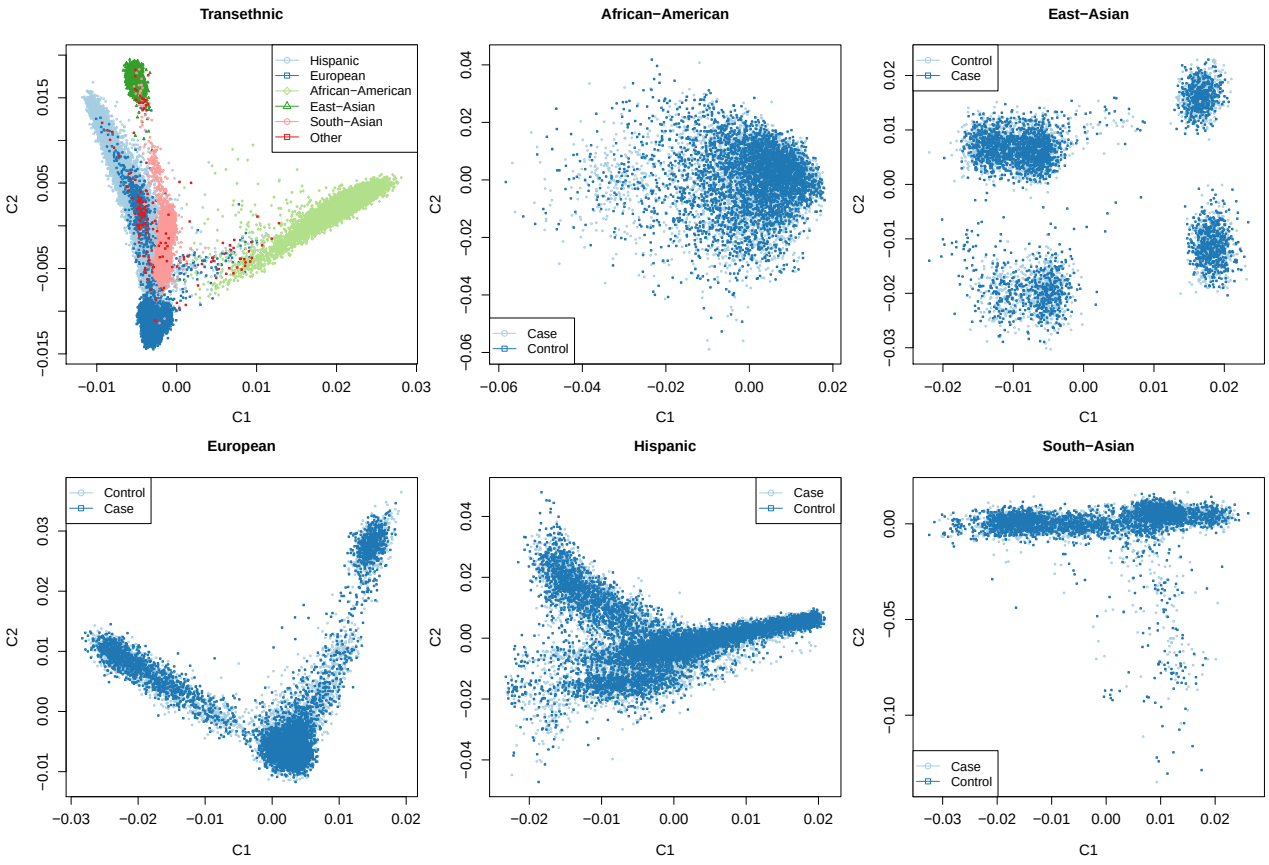
Supplementary Figure 2: **Data quality control workflow.** Shown is a schematic of the steps involved in sample and variant quality control, conducted as described in **Methods** to construct a final set of samples and variants included in association analysis. Each step is depicted as an arrow, with the number of samples or variants excluded by the step shown at the end of the arrow. The final set of samples and variants analyzed are represented by the "Analysis" dataset; we further excluded samples of high relatedness to other samples in the dataset from some but not all analyses. After each step that removed samples, we also removed newly monomorphic variants (hence the decrease in variants between the "Clean" and "Analysis" datasets).



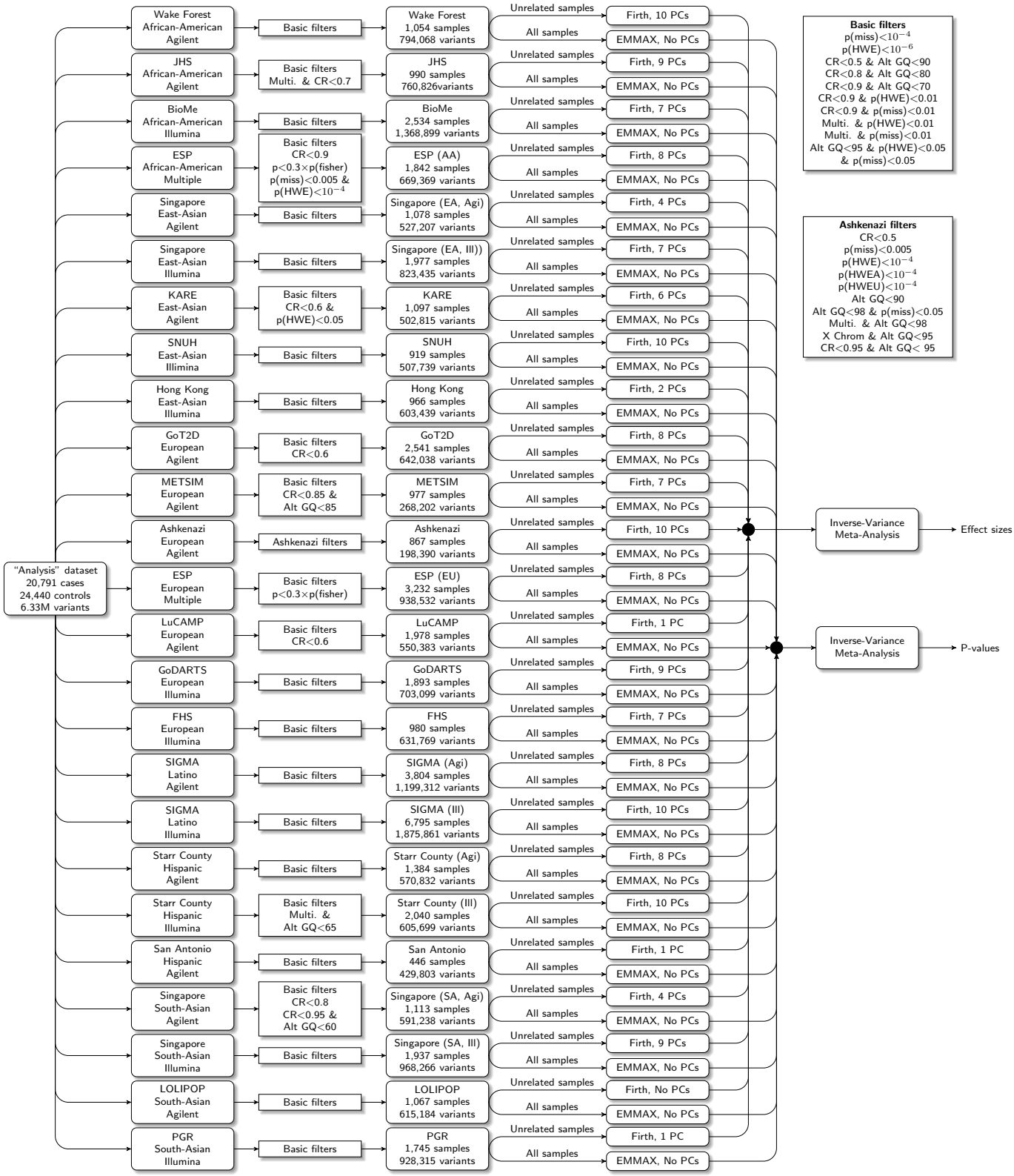
Supplementary Figure 3: **Sample quality control metrics.** To perform sample quality control, we computed a series of metrics that informed on the sequencing quality of a sample. We then stratified samples by ancestry and sequencing technology (i.e. capture technology and year of sequencing), plotted the distribution of metrics for each stratum of samples, and used these plots to visually identify outlier samples for removal by quality control. Shown are (left to right) distributions of the number of variant alleles carried by each sample, the number of variant alleles unique to a sample carried by each sample, and the average fraction of sequence reads supporting a non-reference allele at heterozygous sites within each sample. Distributions are shown for **(a)** all samples from the "Raw" dataset and **(b)** all samples from the "Clean" dataset. Sample strata are labeled by a combination of ancestry and (internal names for) sequencing technology.



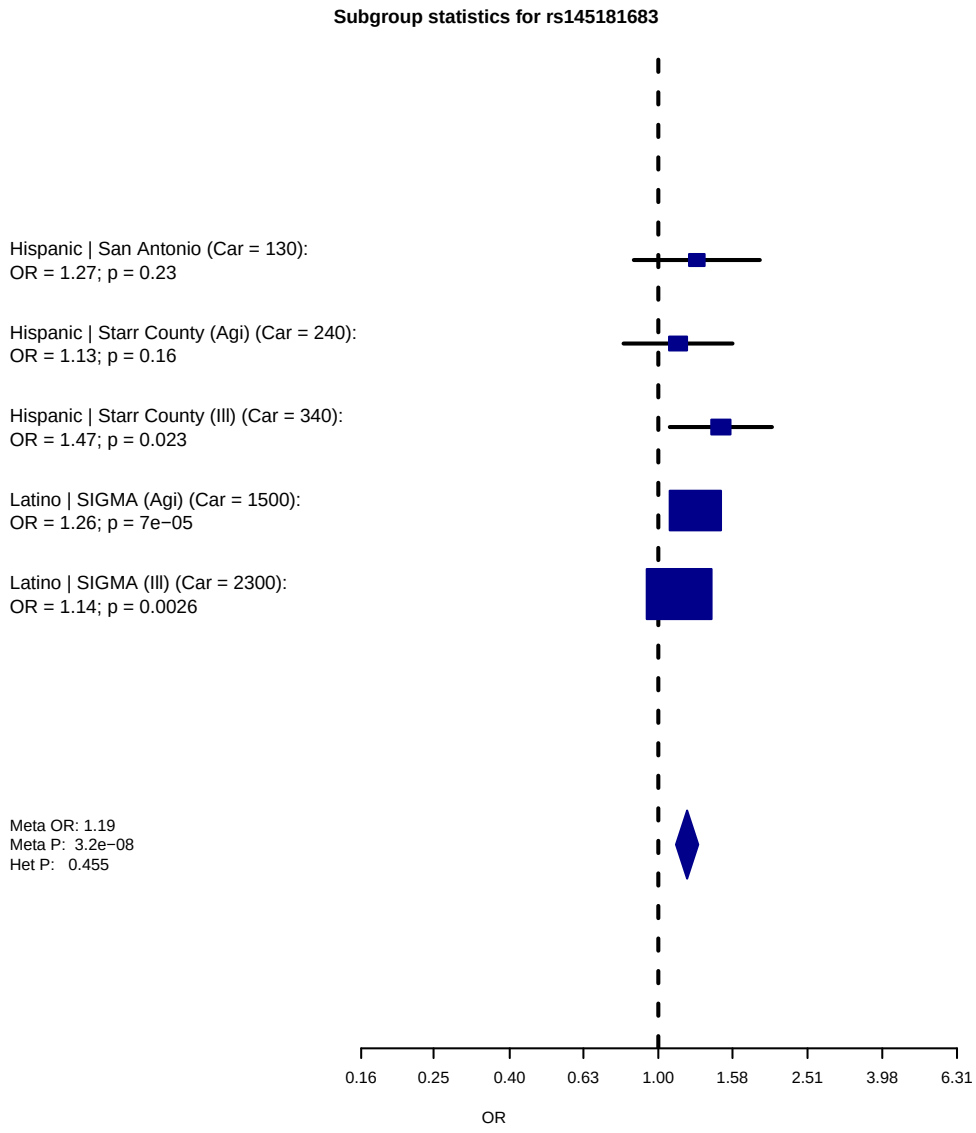
Supplementary Figure 4: **Concordance of exome sequence and SNP array genotypes.** We measured concordance between genotypes called non-reference from sequence data and genotypes called at the same sites in the same samples from SNP array data. Samples are stratified via the same manner as in **Supplementary Figure 3**; the y-axis plots the fraction of non-reference genotypes with an identical genotype call in the corresponding SNP array data. We used four different groups of SNP array data in the analysis (**Methods**), resulting in different y-axis scales for different SNP arrays. Hispanic refers to individuals of either Hispanic or Latino ancestry.



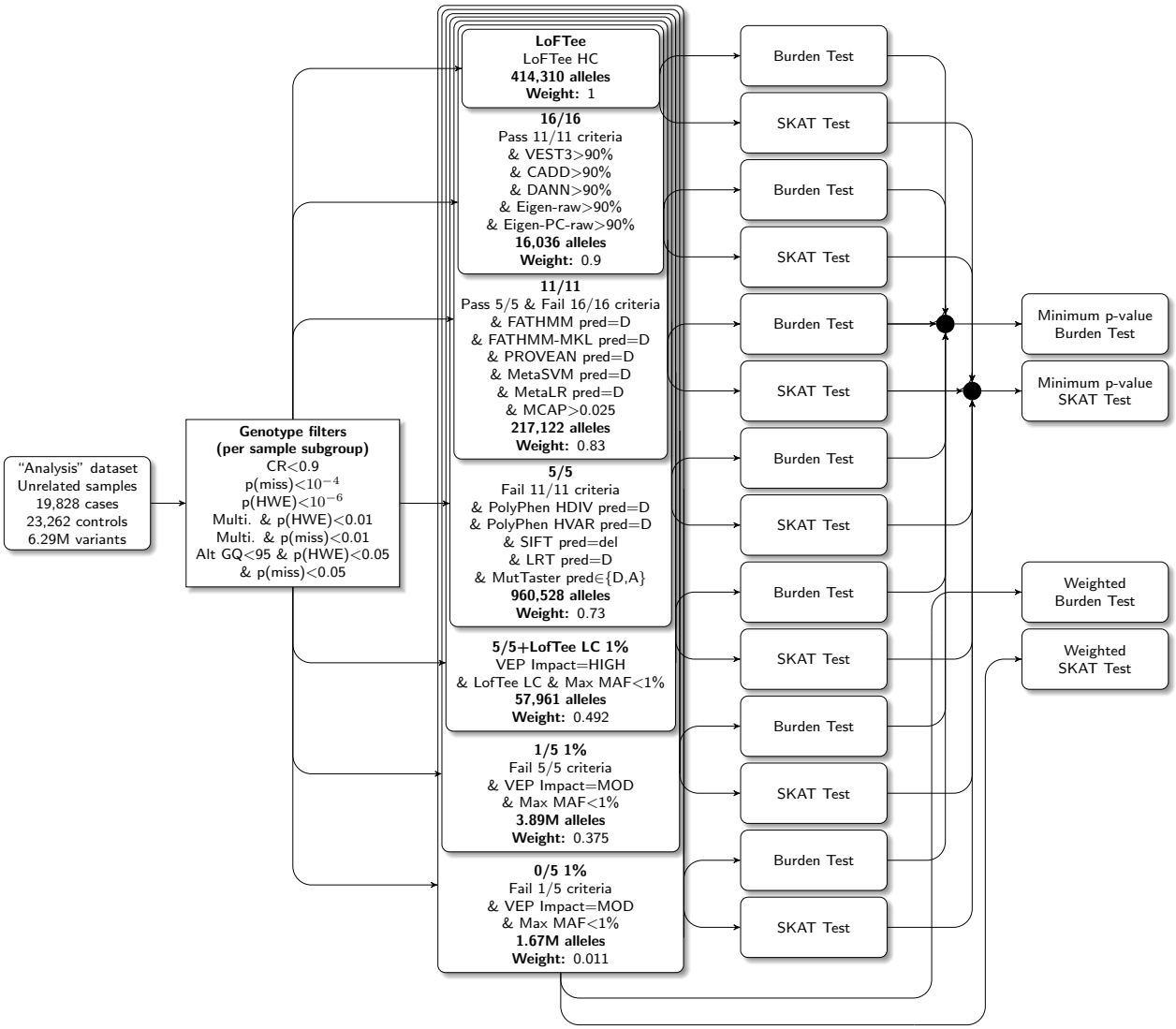
Supplementary Figure 5: **Principal component analysis.** We computed principal component analysis (PCA) based on an LD-pruned collection of variants from exome sequence data. We computed a PCA across all samples (Transethnic; samples colored by reported ancestry) using SNPs common (MAF>1%) in each ancestry, as well as additional PCAs specific to samples from each ancestry (Ancestry labeled plots; samples colored by case/control status for T2D) using a broader set of SNPs common (MAF>1%) in the relevant ancestry.


"Analysis" dataset
20,791 cases
24,440 controls
6.33M variants
Wake Forest
African-American
Agilent
Basic filters
Wake Forest
1,054 samples
794,068 variants
JHS
African-American
Agilent
Basic filters
Multi. & CR<0.7
JHS
990 samples
760,826variants
BioMe
African-American
Illumina
Basic filters
BioMe
2,534 samples
1,368,899 variants
ESP
African-American
Multiple
Basic filters
CR<0.9
p<0.3×p(fisher)
p(miss)<0.005 &
p(HWE)<10^-4
ESP (AA)
1,842 samples
669,369 variants
Singapore
East-Asian
Agilent
Basic filters
Singapore (EA, Agi)
1,078 samples
527,207 variants
Singapore
East-Asian
Illumina
Basic filters
Singapore (EA, Ill))
1,977 samples
823,435 variants
KARE
East-Asian
Agilent
Basic filters
CR<0.6 &
p(HWE)<0.05
KARE
1,097 samples
502,815 variants
SNUH
East-Asian
Illimina
Basic filters
SNUH
919 samples
507,739 variants
Hong Kong
East-Asian
Illumina
Basic filters
Hong Kong
966 samples
603,439 variants
GoT2D
European
Agilent
Basic filters
CR<0.6
GoT2D
2,541 samples
642,038 variants
METSIM
European
Agilent
Basic filters
CR<0.85 &
Alt GQ<85
METSIM
977 samples
268,202 variants
Ashkenazi
European
Agilent
Ashkenazi filters
Ashkenazi
867 samples
198,390 variants
ESP
European
Multiple
Basic filters
p<0.3×p(fisher)
ESP (EU)
3,232 samples
938,532 variants
LuCAMP
European
Agilent
Basic filters
CR<0.6
LuCAMP
1,978 samples
550,383 variants
GoDARTS
European
Illumina
Basic filters
GoDARTS
1,893 samples
703,099 variants
FHS
European
Illumina
Basic filters
FHS
980 samples
631,769 variants
SIGMA
Latino
Agilent
Basic filters
SIGMA (Agi)
3,804 samples
1,199,312 variants
SIGMA
Latino
Illumina
Basic filters
SIGMA (Ill)
6,795 samples
1,875,861 variants
Starr County
Hispanic
Agilent
Basic filters
Starr County (Agi)
1,384 samples
570,832 variants
Starr County
Hispanic
Illumina
Basic filters
Multi. &
Alt GQ<65
Starr County (Ill)
2,040 samples
605,699 variants
San Antonio
Hispanic
Agilent
Basic filters
San Antonio
446 samples
429,803 variants
Singapore
South-Asian
Agilent
Basic filters
CR<0.8
CR<0.95 &
Alt GQ<60
Singapore (SA, Agi)
1,113 samples
591,238 variants
Singapore
South-Asian
Illumina
Basic filters
Singapore (SA, Ill)
1,937 samples
968,266 variants
LOLIPOP
South-Asian
Agilent
Basic filters
LOLIPOP
1,067 samples
615,184 variants
PGR
South-Asian
Illumina
Basic filters
PGR
1,745 samples
928,315 variants
Unrelated samples
All samples
Firth, 10 PCs
EMMAX, No PCs
Firth, 9 PCs
Firth, 7 PCs
Firth, 8 PCs
Firth, 4 PCs
Firth, 6 PCs
Firth, 2 PCs
Firth, 1 PC
Firth, No PCs
Basic filters
p(miss)<10^-4
p(HWE)<10^-6
CR<0.5 & Alt GQ<90
CR<0.8 & Alt GQ<80
CR<0.9 & Alt GQ<70
CR<0.9 & p(HWE)<0.01
CR<0.9 & p(miss)<0.01
Multi. & p(HWE)<0.01
Multi. & p(miss)<0.01
Alt GQ<95 & p(HWE)<0.05
& p(miss)<0.05
Ashkenazi filters
CR<0.5
p(miss)<0.005
p(HWE)<10^-4
p(HWEA)<10^-4
p(HWEU)<10^-4
Alt GQ<90
Alt GQ<98 & p(miss)<0.05
Multi. & Alt GQ<98
X Chrom & Alt GQ<95
CR<0.95 & Alt GQ< 95
Inverse-Variance
Meta-Analysis
Effect sizes
Inverse-Variance
Meta-Analysis
P-values

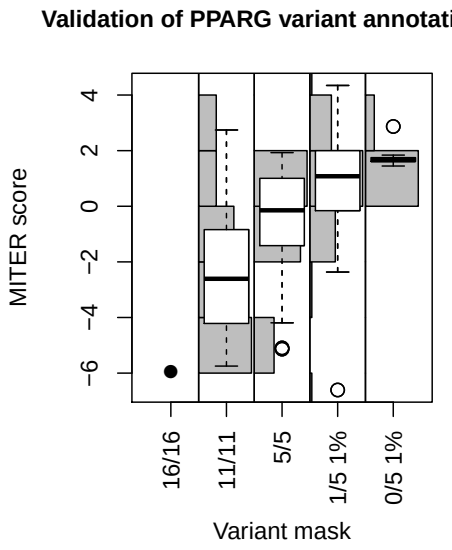
Supplementary Figure 6 ***(preceding page)***: **Single-variant association analysis workflow.** Shown is a schematic of the steps involved in single-variant exome sequence association analysis, as described in **Methods**. We began analysis with a division of samples in the "Analysis" dataset (leftmost column) into 25 different subgroups (second column from left) based on cohort, ancestry and sequencing technology (labeled in each box in the second column). We then filtered variants according to metrics computed separately for each subgroup; we applied the filters listed in the "Basic filters" box to all subgroups, and for some subgroups we applied additional (more stringent) filters as indicated by boxes in the third column from left. The resulting number of variants and samples advanced for analysis in each subgroup are indicated in the fourth column from left. We analyzed each subgroup with both the EMMAX test (to measure association strength) and the Firth test (to measure allelic odds ratios); the number of principal components included as covariates in the Firth test is shown in the fifth column from the left. Finally, we combined each of the EMMAX and Firth results via a 25-group meta-analysis to produce the final *p*-values and odds ratios reported for each variant. Multi: variant is multiallelic; CR: call rate; p: variant subgroup-level *p*-value; p(fisher): variant subgroup-level *p*-value from fisher exact test; p(miss): *p*-value for subgroup-level variant differential missingness between T2D cases and controls; p(HWE): *p*-value for deviation from subgroup-level hardy-weinberg equilibrium; Alt GQ: mean genotype quality of non-reference genotypes (across all samples); X Chrom: variant is on X chromosome.



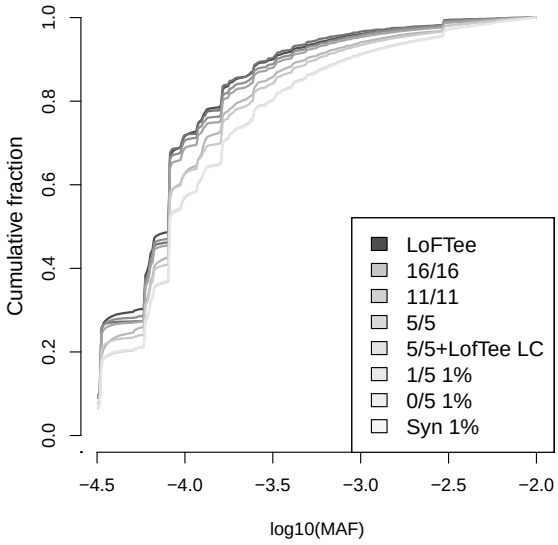
Supplementary Figure 7: ***SFI1* subgroup-level associations.** Shown are the *p*-values and odds-ratio estimates for each sample subgroup with at least 10 carriers of the rs145181683 variant in *SFI1*. Blue boxes indicate odds ratios (sized proportionately to the number of carriers in the subgroup) and black bars indicate standard errors. Car: number of variant carriers. Meta: results from the full analysis across all 25 sample subgroups (including those not shown in this figure). Het P: *p*-value of test for heterogeneity in odds ratios across sample subgroups.



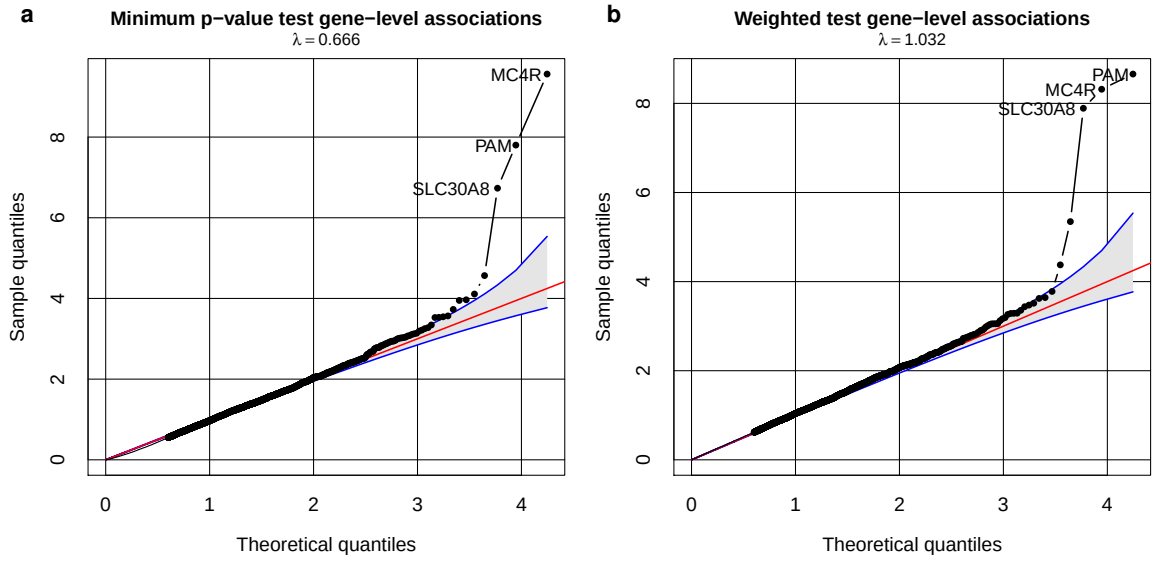
Supplementary Figure 8: **Gene-level association analysis workflow.** Shown is a schematic of the steps involved in gene-level exome sequence association analysis, as described in **Methods**. We began analysis with subgroup-level genotype filtering (second column from left) of unrelated samples in the "Analysis" dataset (leftmost column); we then applied genotype filters for each subgroup (filtering genotypes for either all or no samples in each subgroup) that were similar to those used in subgroup-level single-variant analysis. We then annotated each non-reference variant allele with 16 different bioinformatic algorithms to assess allele deleteriousness, and we grouped alleles into one of seven nested masks (third column from left; the number of variants and weights shown correspond to alleles absent from "higher", or more stringent, nested masks). We computed burden and SKAT analyses via one of two approaches to combine alleles across masks (**Methods**): first, by analyzing all alleles at once with weights assigned according to the most stringent mask containing the allele (weighted test); and second, by analyzing each mask independently and then calculating the lowest *p*-value corrected for the effective number of tests (minimum *p*-value test).



Supplementary Figure 9: **Validation of allele deleteriousness within variant masks.** To assess whether the severity ordering of masks in **Supplementary Figure 8** corresponded to an increasing likelihood that an allele in the mask was deleterious, we used previously published data assessing the extent to which missense variants in the gene *PPARG* impede adipocyte differentiation. For the five masks containing at least one *PPARG* allele, shown are box plots or strip charts of allelic MITER scores (a measure of predicted *PPARG* loss of function, with lower scores suggesting lower function).



Supplementary Figure 10: **Annotation mask weight estimation.** For each variant mask, we estimated allelic weights corresponding to the fraction of loss-of-function alleles in the mask, under a previously presented [3] model whereby a set of missense alleles is a mixture of fully loss-of-function or fully benign alleles. Weights corresponded to the fraction of loss-of-function alleles in each mask, estimated to maximize the likelihood of the allele frequency distribution, with the LofTee mask used as a reference for loss-of-function alleles and the set of synonymous alleles with frequency below 1% used as a reference for benign alleles. Shown are the cumulative frequency distributions for alleles “unique” to each mask (i.e. absent from all more stringent masks).

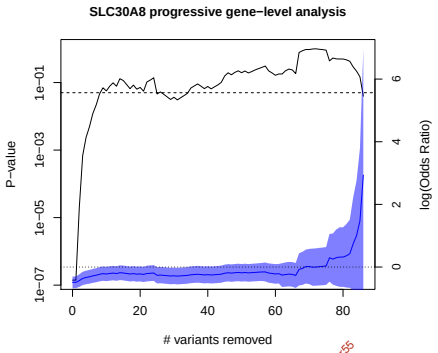


Supplementary Figure 11: **Calibration of gene-level association analyses.** For both the burden and SKAT tests, we tested for gene-level association within seven different allelic masks. As this produced seven *p*-values for each test, we developed two means to consolidate these results (**Methods**). Shown are quantile-quantile (QQ) plots of associations for the **(a)** minimum *p*-value burden test and **(b)** weighted burden test. Only genes with combined minor allele count of 20 or greater are shown in the QQ plots, in order to avoid deflation from genes with too few variants to produce *p*-values asymptotically uniform under the null. Lambda values indicate genomic control, as measured by the ratio in observed median χ^2 statistic to that expected under the null. The three genes with exome-wide significant associations are labeled.

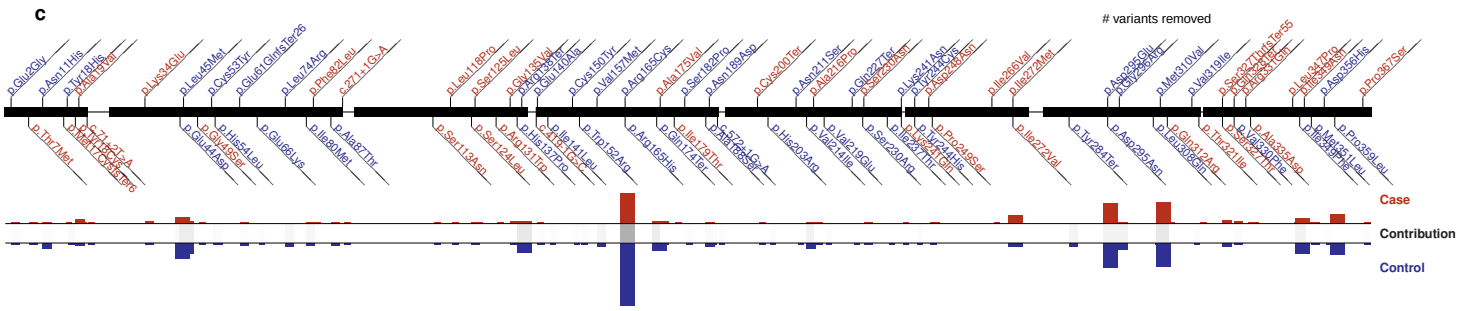
a

Mask	Total					Weighted		Unique				
	# Var	CAF	OR	Burden	SKAT	OR	Burden	# Var	CAF	OR	Burden	SKAT
LoFTee	13	0.0012	0.46	0.0072	0.032	0.46	0.0072	13	0.0012	0.46	0.0072	0.032
16/16	13	0.0012	0.46	0.0072	0.032	0.46	0.0072	0	-	-	-	-
11/11	15	0.0013	0.5	0.013	0.041	0.49	0.011	2	0.00014	1.1	0.87	0.93
5/5	37	0.0045	0.47	1.4e-06	0.00072	0.4	1.8e-06	22	0.0032	0.46	3.5e-05	0.0011
5/5+LofTee LC 1%	37	0.0045	0.47	1.4e-06	0.00072	0.4	1.8e-06	0	-	-	-	-
1/5 1%	86	0.012	0.6	4.7e-08	3.6e-05	0.39	1.1e-08	49	0.0072	0.69	0.0015	0.0015
0/5 1%	103	0.014	0.62	5.5e-08	2.8e-05	0.4	1.3e-08	17	0.0019	0.8	0.34	0.12

b



c

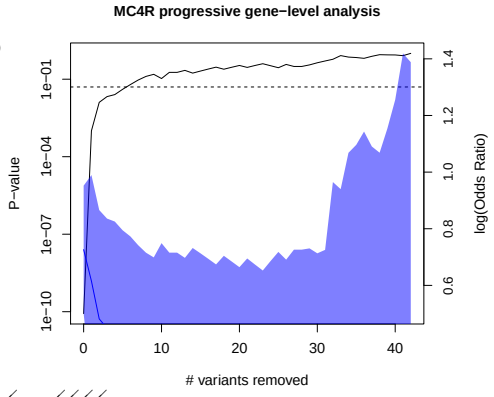


Supplementary Figure 12: ***SLC30A8* gene-level analysis.** Shown is a dissection of the gene-level associations for *SLC30A8*. **(a)** Mask-level statistics for the burden and SKAT tests, as well as the weighted burden test. Each row in the table corresponds to one of the allele masks defined in **Supplementary Figure 8**. The first five columns ("Total") show association results for an analysis of all alleles in the mask; the final five columns ("Unique") show association results for analysis of alleles unique to the mask (i.e. are not present in more deleterious masks). The "Weighted" columns show association results for a weighted burden test of all alleles in each mask; the weighted burden result used in the main analysis is that in the final row. #Var: the number of variants in the association analysis. CAF: the total combined frequency of all alleles in the analysis. OR: the odds-ratio estimated from the burden (or weighted burden) analysis. Burden: the *p*-value from the burden test. SKAT: the *p*-value from the SKAT analysis. The #Var and CAF columns for the "Total" analysis also apply to the "Weighted" analysis. **(b)** Gene-level association *p*-values for *SLC30A8*, using the burden test on alleles in the 1/5 1% mask (that achieving greatest statistical significance) after progressive removal of variants ordered by increasing single-variant *p*-value. The left-axis (black line) shows the observed $-\log_{10}(p)$ value, with dashed line indicating nominal significance of $p<0.05$. The right-axis (blue line) shows the estimated effect size ($log(OR)$), with shaded blue indicating the 95% confidence interval and dotted line indicating effect size=0. This figure is repeated from **Figure 1c** in the main text. **(c)** A graphical plot of variants observed in *SLC30A8* within the 1/5 1% mask. Variants are colored blue (if individual OR < 1) or red (OR > 1). Case (red) and control (blue) frequencies are shown below for each variant, with black boxes shaded according to the contribution of each variant to the gene-level signal (computed by the difference in $\log_{10}(p)$ observed after removal of the variant from the test). This figure is repeated from **Figure 1d** in the main text. OR: odds ratio.

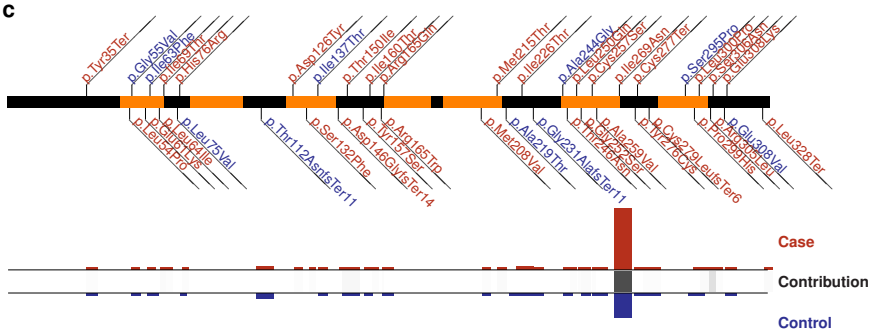
a

Mask	Total					Weighted		Unique				
	# Var	CAF	OR	Burden	SKAT	OR	Burden	# Var	CAF	OR	Burden	SKAT
LoFTee	4	0.00032	1.7	0.32	0.77	1.7	0.32	4	0.00032	1.7	0.32	0.77
16/16	4	0.00032	1.7	0.32	0.77	1.7	0.32	0	-	-	-	-
11/11	5	0.00037	1.8	0.26	0.82	1.8	0.27	1	4.6e-05	2.2	0.55	0.6
5/5	40	0.0079	2	1.6e-10	5.4e-08	2.6	2e-10	35	0.0075	2.1	2.9e-10	5.8e-08
5/5+LofTee LC 1%	41	0.008	2.1	8.5e-11	4.9e-08	2.6	1.3e-10	1	4.6e-05	5.1	0.22	0.15
1/5 1%	94	0.016	1.5	2.7e-06	2.4e-08	2.2	6e-09	53	0.0077	1	0.97	0.46
0/5 1%	105	0.017	1.4	4.2e-06	6.2e-08	2.2	4.8e-09	10	0.00093	1.4	0.28	0.57

b



c

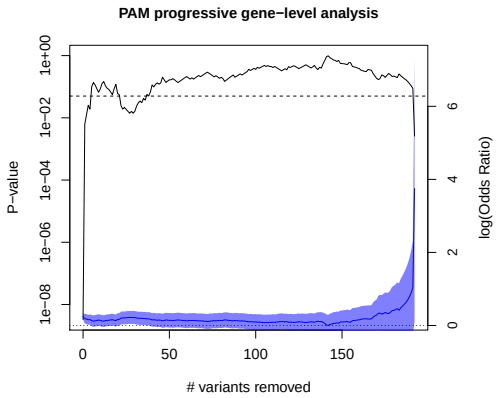


Supplementary Figure 13: ***MC4R* gene-level analysis.** Shown is a dissection of the gene-level associations for *MC4R*. Panels are analogous to those in **Supplementary Figure 12**.

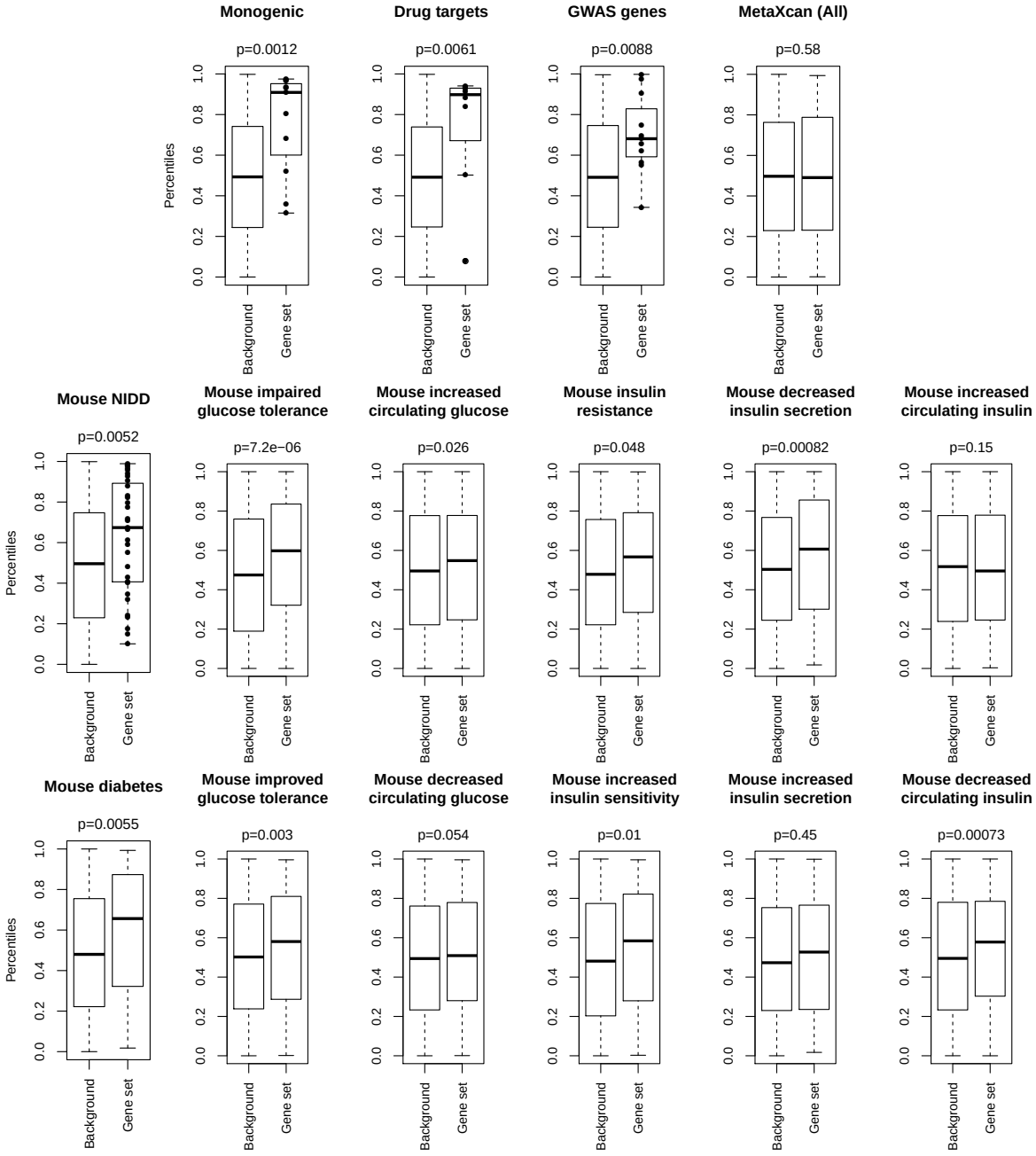
a

Mask	Total					Weighted		Unique				
	# Var	CAF	OR	Burden	SKAT	OR	Burden	# Var	CAF	OR	Burden	SKAT
LoFTee	4	9.3e-05	0.67	0.69	0.38	0.67	0.69	4	9.3e-05	0.67	0.69	0.38
16/16	4	9.3e-05	0.67	0.69	0.38	0.67	0.69	0	-	-	-	-
11/11	12	0.00042	1.3	0.58	0.15	1.3	0.62	8	0.00032	1.6	0.39	0.15
5/5	70	0.049	1.3	9e-09	9.6e-08	1.4	9.5e-09	58	0.048	1.3	1e-08	9.1e-08
5/5+LofTee LC 1%	73	0.049	1.3	8.5e-09	9.9e-08	1.4	7.5e-09	3	0.00019	1.3	0.72	0.7
1/5 1%	193	0.06	1.3	6.7e-09	6.5e-08	1.4	1.8e-09	120	0.011	1.1	0.17	0.14
0/5 1%	213	0.063	1.3	7.6e-09	1.3e-07	1.4	2.2e-09	19	0.0024	0.99	0.95	0.75

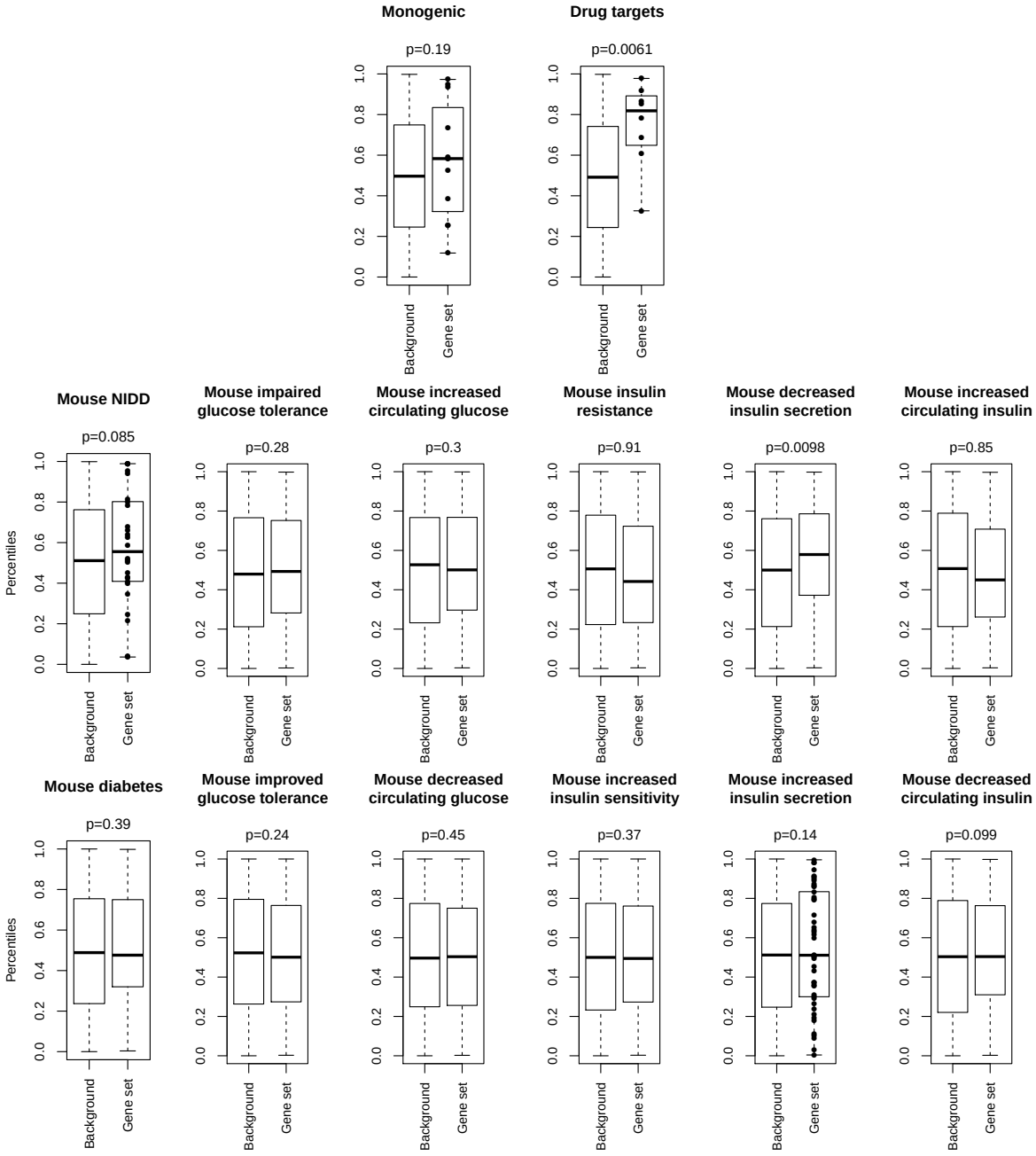
b



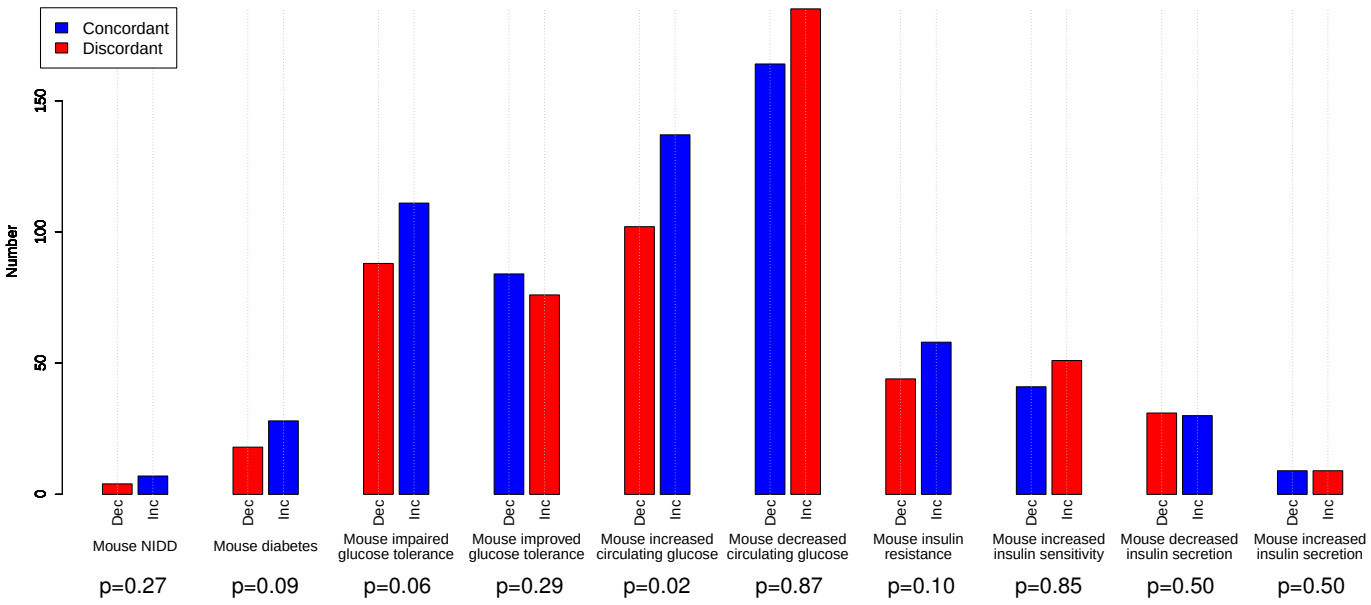
Supplementary Figure 14: ***PAM* gene-level analysis.** Shown is a dissection of the gene-level associations for *PAM*. Panels are analogous to those in **Supplementary Figures 12** and **13**. A graphical plot of variants is not shown due to the large number of variants in *PAM*.



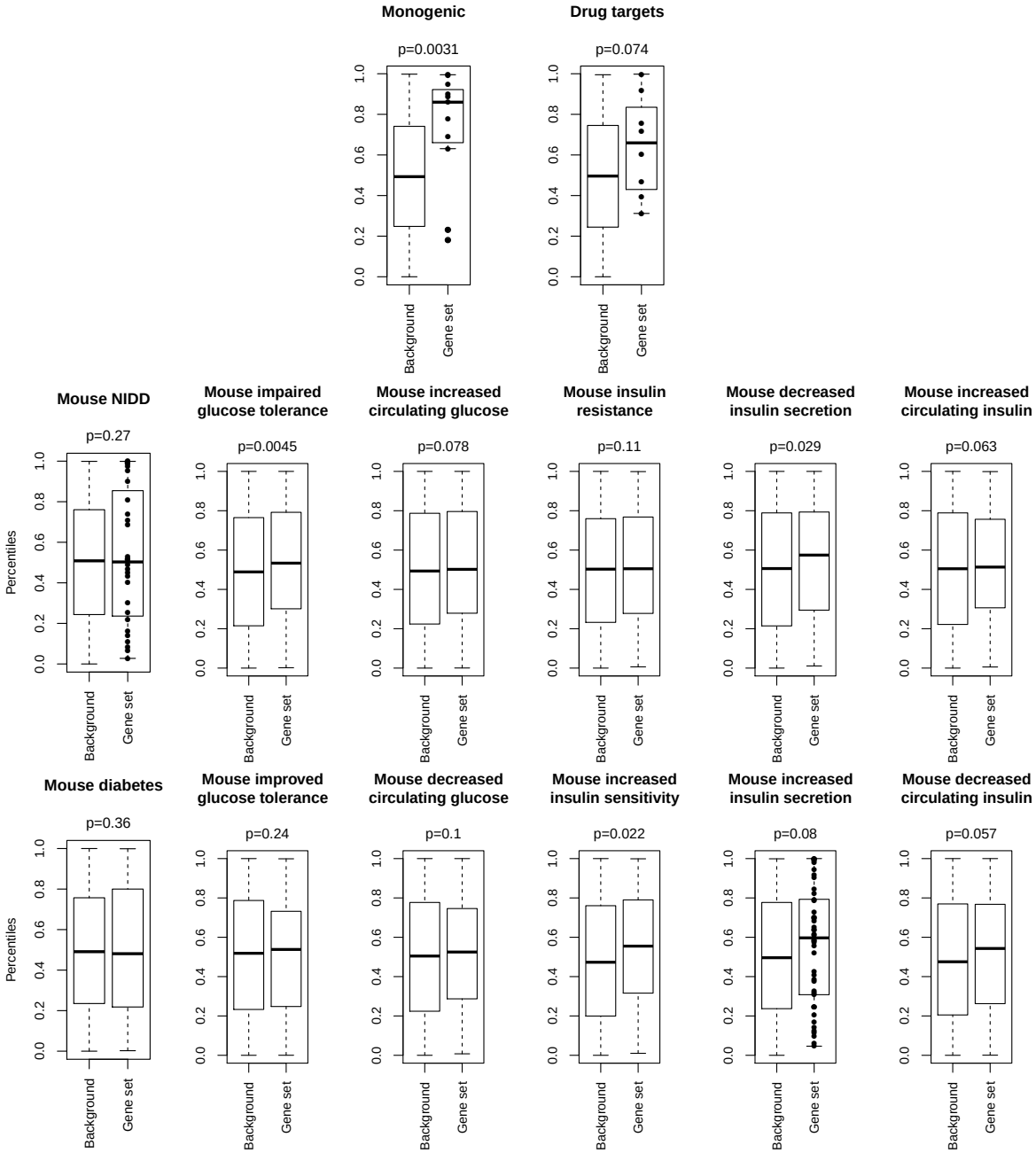
Supplementary Figure 15: **Full results from exome sequence gene set analysis.** For various sets of genes implicated as relevant to T2D based on knockout mouse phenotypes, we used a one-side Wilcoxon rank-sum test to compare gene-level association statistics to those of matched comparison genes (**Methods**). Shown are box plots of the distributions of rank percentiles (1 being the highest) for each gene set analyzed. Plots are analogous to those in **Figure 2**.



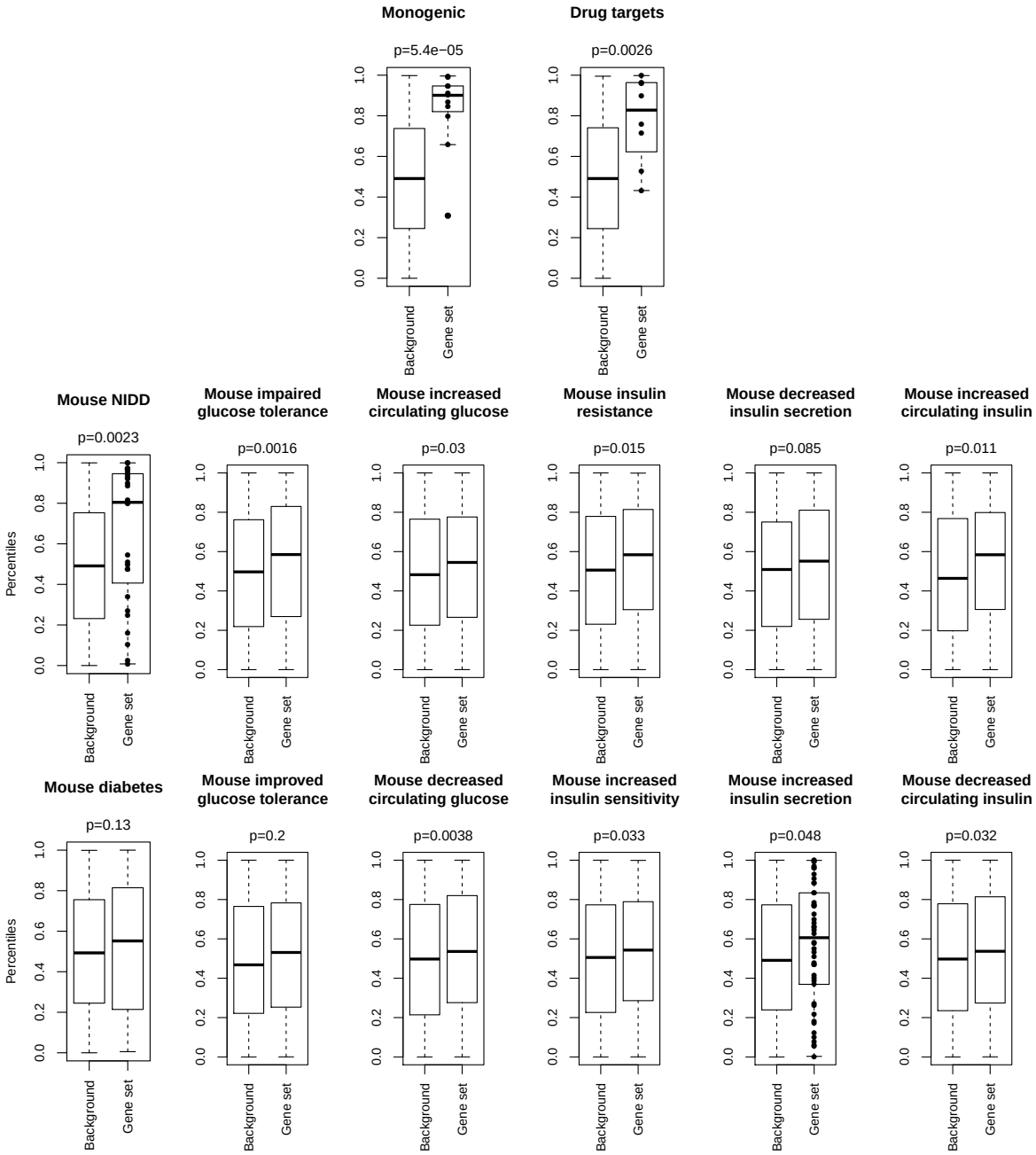
Supplementary Figure 16: **Gene set analysis from protein-truncating variants.** To assess the value of nonsynonymous variants in exome sequence gene set analysis, we conducted a similar rank-sum comparison of gene sets as that described in the main text – only using the burden test of protein truncating variants (PTVs, those included in the LofTee mask), rather than the minimum *p*-value burden test, to calculate gene-level associations. Shown are plots, analogous to those in **Supplementary Figure 15**, summarizing these PTV-based comparisons.



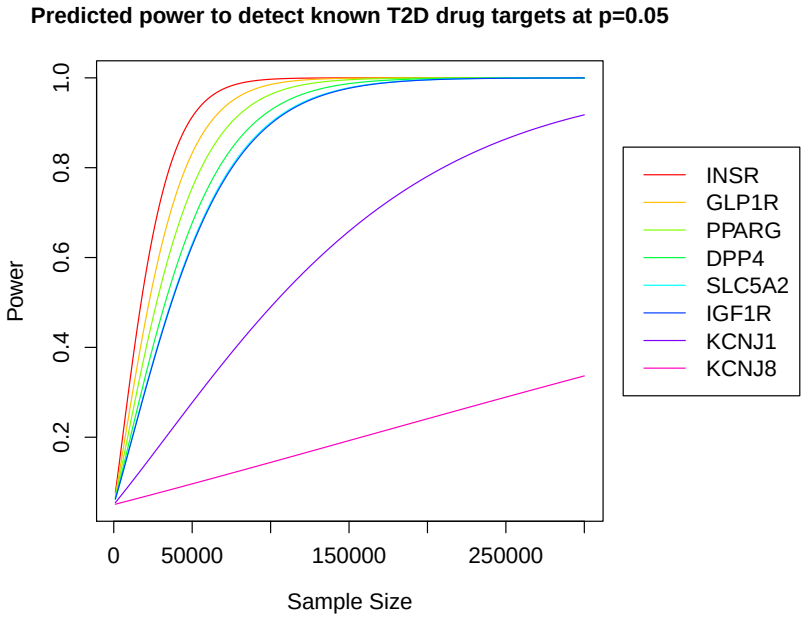
Supplementary Figure 17: **Directional consistency of genetic odds ratio estimates and knockout mouse phenotypes.** For each gene set associated with a knockout mouse phenotype for which there was a analogous human phenotype of increased or decreased T2D risk (**Methods**), we calculated the fraction of genes for which the odds-ratio (OR) estimated from the weighted burden test had a direction consistent with what would be predicted from the knockout mouse phenotype. Blue bars correspond to the number of genes with OR estimates concordant with that predicted from the mouse phenotype, while red bars correspond to the number with discordant OR estimates. *p*-values shown below the bars are calculated from a one-sided binomial test of the null hypothesis that $< 50\%$ of estimates are concordant. Dec: OR estimate is <1. Inc: OR estimate is >1.



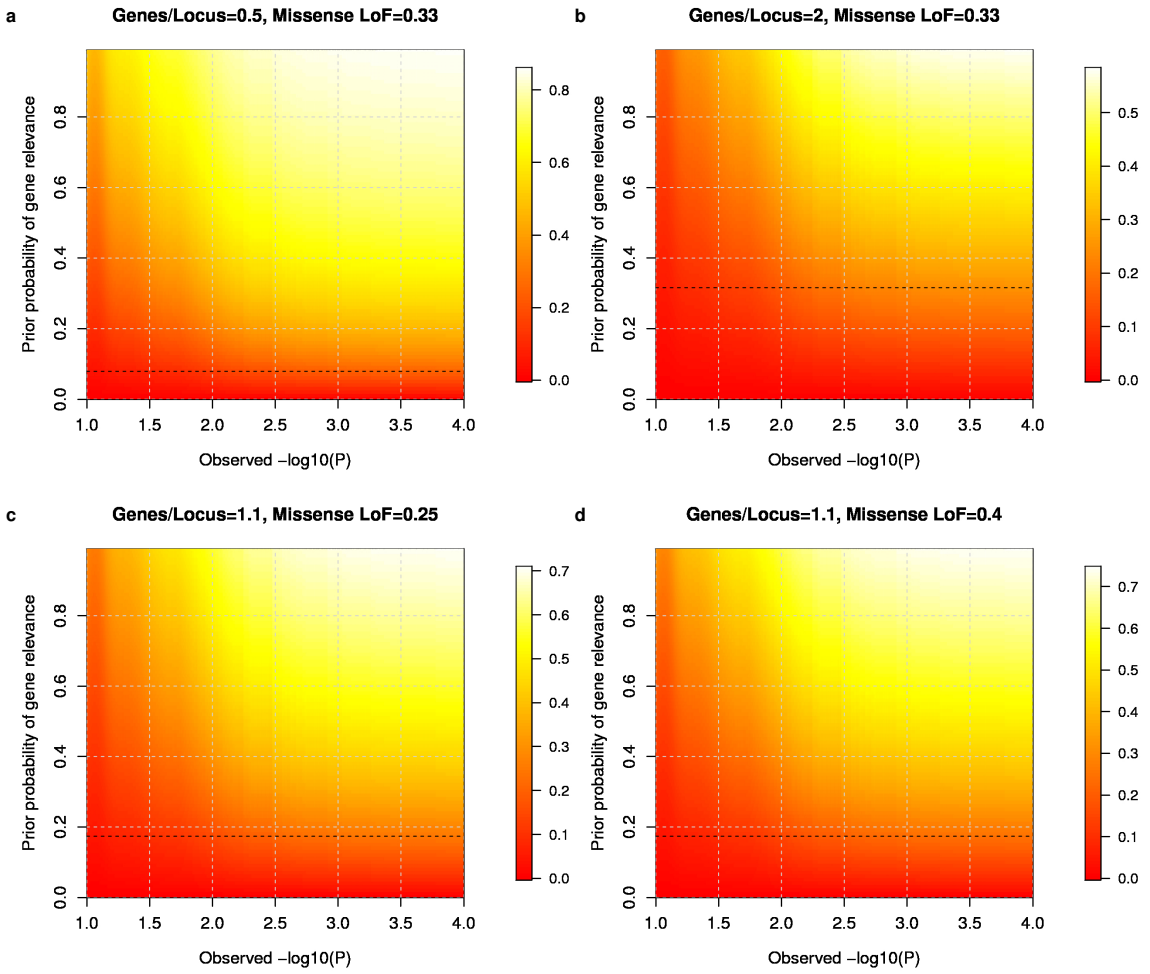
Supplementary Figure 18: **Gene set analysis from imputed GWAS statistics.** To assess how similarly array-based GWAS association statistics could identify gene set associations, as compared to exome sequence gene-level association statistics, we conducted a similar rank-sum comparison of gene sets as that described in the main text – only using gene MAGENTA [7] scores from the imputed GWAS rather than the minimum *p*-value burden test to calculate ranks. Shown are plots, analogous to those in **Supplementary Figure 15**, summarizing these GWAS-based comparisons.



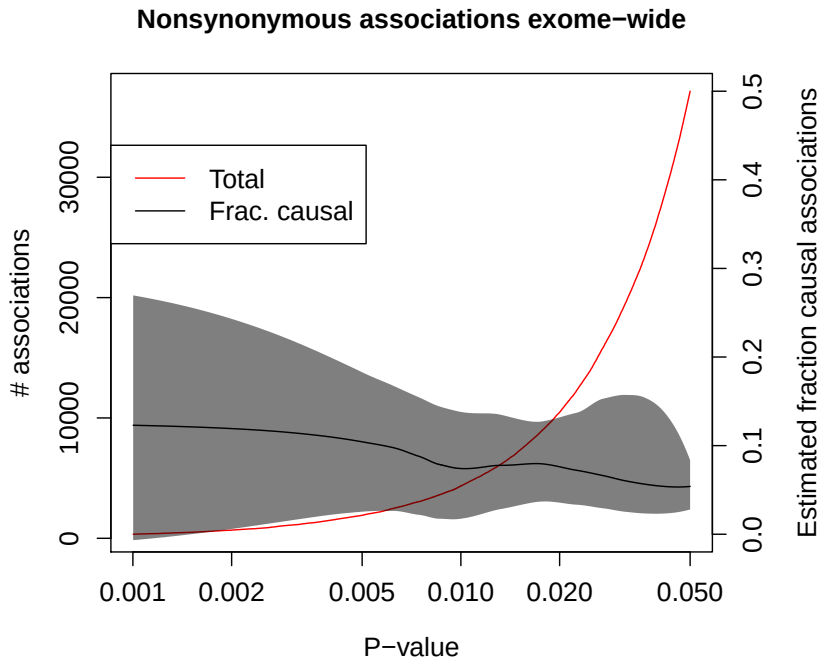
Supplementary Figure 19: **Gene set analysis from a larger array-based GWAS.** To assess whether whether the ability of GWAS statistics to prioritize genes was driven by sample size, rather than fundamental limitations of SNP arrays and imputation, we repeated our rank-sum analysis using gene MAGENTA [7] scores but from a large transethnic T2D GWAS [8] rather than the imputed GWAS in our study. Shown are plots, analogous to those in **Supplementary Figure 18**, summarizing these comparisons.



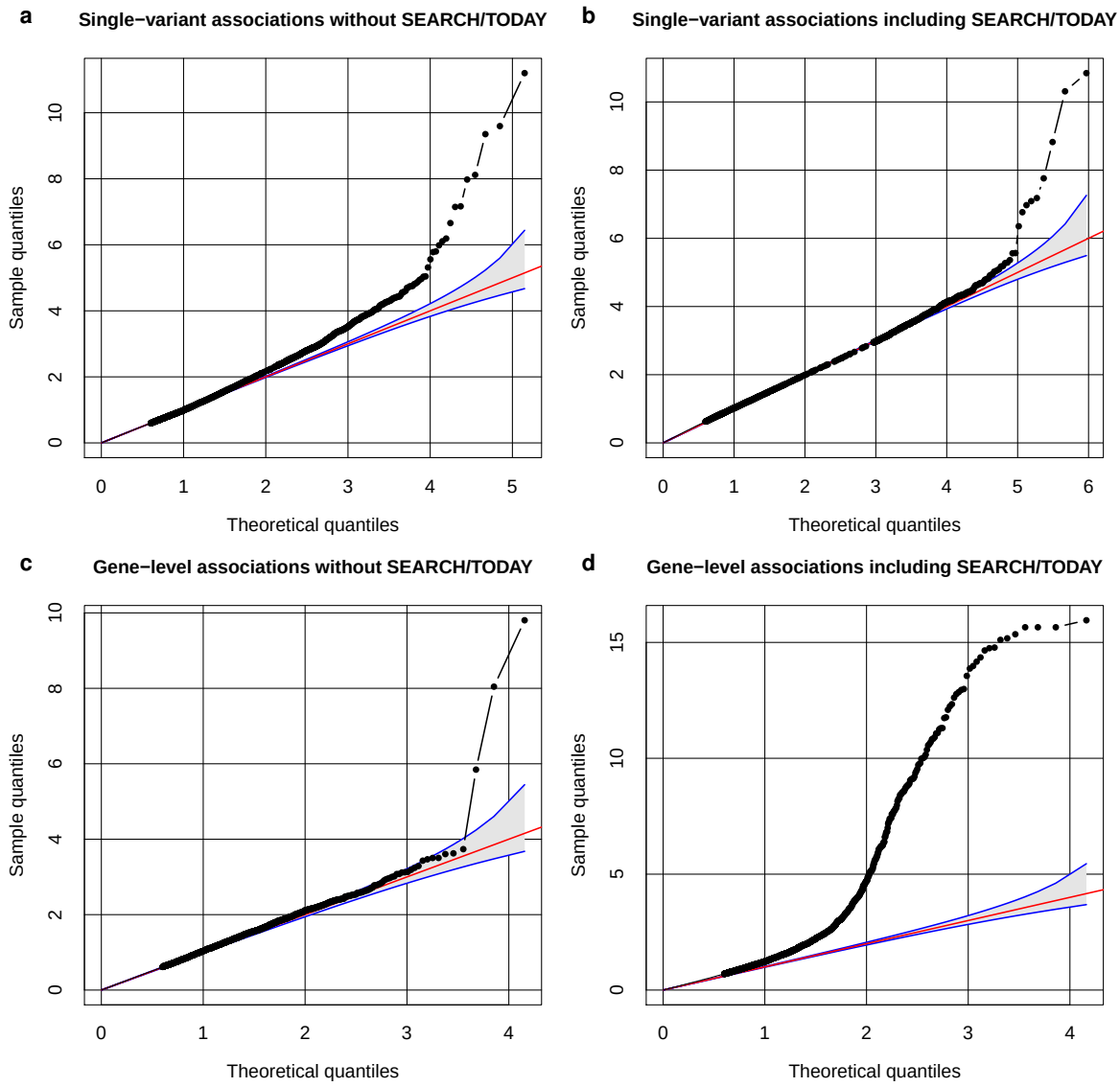
Supplementary Figure 20: **Power to exceed nominal significance for T2D drug targets.** Estimated power, as a function of sample size, to detect T2D gene-level associations (at significance $p<0.05$) for genes with genetic effects (aggregate frequency and odds ratios) equal to those estimated for eight established T2D drug targets. Power curves are shown and colored separately for each target. This figure is identical to that in **Figure 4a** except with a lower significance threshold used in power calculations.



Supplementary Figure 21: **Estimated posterior probability of association for different prior hypotheses.** We estimated the posterior probability of association for nonsynonymous variants meeting various *p*-value thresholds in our analysis, as described in **Methods** and shown in **Figure 4c-f**. In order to perform the needed calculations, we assumed that, on average, 1.1 genes were within each T2D GWAS locus are relevant to T2D and 33% of missense mutations within these genes cause gene loss-of-function. To assess the sensitivity of our analysis to this assumption, we repeated the calculations with different assumptions of **(a)** 0.5 and **(b)** two T2D-relevant genes within each GWAS locus, as well as **(c)** 25% and **(d)** 40% of missense variants leading to loss-of-function. All figures shown assume the default modeling parameters that 30% of true nonsynonymous associations are causal associations; different values for this parameter would scale posterior probability estimates linearly.



Supplementary Figure 22: **Exome-wide posterior estimates.** In addition to estimation of the posterior probability of association (PPA) for nonsynonymous variants within T2D GWAS loci, we also calculated PPA estimates for arbitrary variants exome-wide. Shown are these estimates (black line, gray 95% confidence interval; right axis), as well as the number of total variants (red line; left axis), as a function of single-variant *p*-value observed in our analysis. This plot is analogous to that in **Figure 4d**.



Supplementary Figure 23: **Analysis with SEARCH and TODAY samples included.** Among the cohorts initially sequenced for our exome sequence analysis were childhood diabetes cases from the SEARCH and TODAY studies (**Supplementary Table 1**). We initially hoped to include these cases in our analysis, but the lack of matched controls within these studies raised concerns about potential artifacts that could be introduced during association analysis. To evaluate the possibility of including these cohorts, we compared **(ab)** single-variant and **(cd)** gene-level associations with and without SEARCH and TODAY samples included. **(a)** A quantile-quantile (QQ) plot of single-variant associations computed without SEARCH and TODAY samples. By contrast with the results reported in the manuscript, association statistics here are computed via a meta-analysis of ancestry-level (rather than subgroup-level) association statistics, in order to match an analysis with SEARCH and TODAY samples as closely as possible (a subgroup-level meta-analysis is not possible with SEARCH and TODAY due to the absence of controls in those studies). Only variants with minor allele count above 15 are shown in the QQ plot. **(b)** A QQ plot of single-variant associations, identical to that in **(a)** but with SEARCH and TODAY samples included. **(c)** A QQ plot of gene-level burden associations from the 5/5 mask. Only genes with a total of 15 aggregate alternate alleles are shown in the QQ plot. **(d)** A QQ plot of gene-level burden associations from the 5/5 mask, identical to that in **(c)** but with SEARCH and TODAY samples included.

References

[1] Eilbeck, K. *et al.* The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* **6**, R44 (2005).

[2] McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

[3] Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E455–464 (2014).

[4] Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

[5] Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).

[6] Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).

[7] Segre, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6** (2010).

[8] Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).