The American Journal of Human Genetics, Volume *100*

Supplemental Data

High-Resolution Genetic Maps Identify Multiple Type 2 Diabetes Loci at Regulatory Hotspots in African Americans and Europeans

Winston Lau, Toby Andrew, and Nikolas Maniatis

SUPPLEMENTARY METHODS

WELLCOME TRUST CASE CONTROL CONSORTIUM AND AFRICAN AMERICAN SAMPLE SELECTION

The Wellcome Trust Case Control Consortium (WTCCC phase I) described the diagnosis and selection of T2D cases for the original study as "based on either current prescribed treatment with sulphonylureas, biguanides, other oral agents and/or insulin or, in the case of individuals treated with diet alone, historical or contemporary laboratory evidence of hyperglycaemia (as defined by the World Health Organization)"[1]. The two pooled control groups from the 1958 Birth Cohort (aged 44-45) and the UK Blood Service (aged 18-69) are used as shared controls for all seven disease cases (including T2D) in the original study[1], which implies that the controls cannot be assumed to be group matched by BMI, age or sex. The WTCCC2 (phase II) indirectly describes case selection in relation to the Metabochip array design[2], which targets T2D genomic disease loci identified by the DIAGRAM consortium[3]. Although the T2D diagnostic criteria used by the >20 participating research groups varied, most used ADA[4; 5] and WHO[6] guidelines and/or based on treatment with oral anti-diabetic medication or insulin[3]. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)[7] recruited individuals with T2DM and End Stage Renal Disease (ESRD) from dialysis facilities, with the stipulation that cases had to meet at least one of the following three criteria to be included: i) T2DM diagnosed at least 5 years before initiating renal replacement therapy; ii) diabetic retinopathy and/or c) diabetic nephropathy (T2D-ESRD cases).

Given the WTCCC groups did not specify if case and controls were matched by BMI, it cannot be ruled out that gene-mapping studies based upon these European samples might identify genetic loci that are confounded by BMI / adiposity rather than being associated with T2D alone. By contrast, this possibility is to some extent mitigated for this study, where the National Institute of

Diabetes and Digestive and Kidney Diseases (NIDDK)[7] samples are BMI matched. WTCCC1 used population controls, which (as noted in the publication[1]) reduces power due the control group including an expected proportion of T2D cases equal to the population prevalence. No published documentation for the selection of WTCCC2 controls is recorded[8]. For the NIDDK, unrelated African-American controls screened for no diagnosis of diabetes or renal disease were recruited from the community and internal medicine clinics (controls)[7]. This suggests that where cosmopolitan T2D disease loci (i.e. the co-location of disease loci for European and African American samples) are identified in this study, we can be more confident that these are not confounded by BMI/ adiposity.

TARGETED RE-SEQUENCING - EUROPEAN CASE/CONTROL SAMPLES

For the purposes of targeted re-sequencing at the loci *ACTL7B*, *KCNK3* and *TCF7L2*, we used French samples with cases selected from multiplex families from linkage studies with a history of T2D[9; 10] and unrelated controls selected from families with obese individuals, but no history of T2D[11; 12] and 1:1 matched for age, sex and body mass index (BMI, see Table S2).

CONSTRUCTION OF THE GENETIC LDU MAPS

HapMap (Release 28) was used to construct genetic maps for the European samples using 56 unrelated European (EUR) individuals genotyped for 2,270,218 SNPs (screened for quality control) and a second genetic map constructed for the African-American samples using 57 unrelated individuals of African ancestry from South West USA (ASW) genotyped for 1,333,297 quality-control SNPs. The Linkage Disequilibrium (LD) maps are based upon HapMap data with genetic distance provided in additive LD units (LDU). The power of the multi-marker approach

compared with conventional GWA analysis (single-SNP tests) is primarily provided by the additional information contained in the high-resolution LDU genetic maps[13-15] and the reduced number of genomic tests that reduces the multiple-testing burden. The construction of the LDU maps is based on the Malécot-Morton model, which describes the observed decline of pair-wise LD between SNPs as measured by rho, ρ, as an exponential function of physical distance in kilobases (d). The expected decline in pairwise LD is modelled as: $\hat{\rho} = (1-L)Me^{-\varepsilon d} + L$, with M being the intercept, reflecting the maximum value of LD prior to LD breakdown (~1 for monophyletic origin, i.e. one ancestral haplotype) and L being the asymptote, reflecting spurious LD at large distance, not due to linkage. The parameter ε is the exponential decline of LD and, together with distance d, in kb, an estimate of LDU = $\Sigma\varepsilon_i d_i$ is provided for every ith interval. In this way, all SNPs in the T2D datasets have genetic locations measured in LDU[16]. The parameters M, L and ε are the iterative maximum likelihood estimations. The autosomal genome was divided into 4,800 non-overlapping analytical windows of approximately equal size on the genetic map.

ASSOCIATION MAPPING USING LDU MAPS

We carried out association analyses for all autosomal chromosomes using three T2D datasets with a total of 5,800 cases and 9,691 controls. The first genome-wide association (GWA) dataset was obtained from the WTCCC (phase I) and included T2D cases (n=1,925) and controls (n=2,938) of North European ancestry with available genotypes (Affymetrix, ~500K SNPs)[1; 2; 7]. The second independent dataset was also from the WTCCC (phase II), and included T2D cases (n=2,910) and controls (n=5,724) of North European ancestry (UK)[2], who were genotyped using the Metabochip array (Illumina, ~200K SNPs). The third dataset was obtained from a GWA study for a population of predominantly African ancestry conducted by the NIDDK[7] and included African American

(AA) T2D cases (965) and AA controls (1029). The AA NIDDK T2D cases and controls were genotyped at a much higher SNP resolution array (Affymetrix, ~1M SNPs)[7]. All three datasets were screened using standard quality control filters described in previous publications[1; 2; 7] and online data sources. For the eQTLs analysis we used data generated by the MuTHER consortium[17]. Subcutaneous adipose mRNA levels were measured in 825 European twins (TwinsUK) by the MuTHER consortium with data generation and normalization methods described in their initial report and online data sources[17].

The multi-marker association test[18] is based on composite likelihood (Λ)[16; 19], in which all observed genotyped SNPs within each window are simultaneously tested. We therefore do not use imputation and conditional analysis, because the aim of LDU analysis is to estimate the location of functional variants in any given genomic region that provides the strongest evidence of association with disease. For this approach observed (not imputed missing) genotype data are required for reasons explained in the main manuscript. Application of this method to each analytical window returns one estimated location ($\hat{S}$) for the causal variant (± standard error) at the strongest signal, along with the association test *P*-value. The association test is based upon the same Malécot model used to construct the LDU maps described above, although in this case the T2D-by-SNP association (z)[14] is included in the model instead of SNP-by-SNP association (ρ), along with an additional parameter of causal variant location ($\hat{S}$), with all distances measured in LDU. Therefore the Malécot model prediction of association between disease and markers is estimated by the equation $\hat{z}_i = (1-L)Me^{-\varepsilon|(S_i - \hat{S})|} + L$, where S_i the ith SNP LDU location and $\hat{S}$ the estimated location of the putative functional variant on the genetic LDU map. The genetic distance standard errors of $\hat{S}$ were used to obtain the 95% confidence intervals (CI) of the putative causal

variants[14], but in this study we only present the co-location intervals (distance between the $\hat{S}$ location estimates), which are the genomic regions within the CIs that most plausibly include the functional variants that confer risk of T2D. For the three gene regions used as examples (Figures 2, 3 and 4 in the manuscript), we constructed LDU maps from the 1000 Genomes Project data, but no differences on the T2D locations estimates were observed based upon the 1000G and HapMap LDU maps.

The regression coefficient b was used instead of z for the adipose expression quantitative phenotype (eQTL analysis). All the regression coefficients, standard errors and P-values for expression probes regressed upon SNPs and probe corresponding gene names were obtained from the MuTHER website (http://www.muther.ac.uk). Our eQTL analysis targeted 173 replicated T2D signals (111 additional loci in Table 1, 62 from the previously found list[20]). The Malécot model was then applied after assigning the EUR LDU locations to the SNPs used from the MuTHER data for these 173 signals. The MuTHER probe gene names were updated based on common nomenclature as provide by the UCSC website and to be consistent with the publications that we have referenced in the manuscript.

For convenience, all the functional location estimates ($\hat{S}$) for T2D and eQTLs were converted back to the physical map Build 36 (B36, NCBI36/hg18) in kb by linear interpolation of the two flanking SNPs on the HapMap LDU map. When the $\hat{S}$ was located in an LDU block (horizontal LDU line) then all markers within that block have the same LDU location. In such cases, we took the midpoint of that block as an estimate of $\hat{S}$ in kb. All eQTL locations ($\hat{S}$) had to co-locate within 50kb from the T2D $\hat{S}$ estimates. A detailed description of the Malécot multi-marker test of association is

provided by Maniatis et al.[18] and the construction of the LDU maps for this study using the HapMap phase II data are described in more detail elsewhere[14; 16].

Analytical window *P*-values were meta-analysed using Fisher's method to provide overall evidence of association. We did not use other types of meta-analysis (e.g. fixed or random effects), because the multi-marker test of association estimates the causal variant location, but not the association effect size. In order to account for multiple testing, analytical windows were filtered for having a meta *P*-value less than the Bonferroni corrected, genomic *P*-value threshold of $1x10^{-5}$, based on the total number of tests performed (n=4,800; $\alpha = 0.05/4,800$). Loci were only considered biologically plausible if the significant $\hat{S}$ location estimates from different datasets were within a <100 kb interval.

For the eQTL analysis, adipose tissue expression probes were tested for *cis*-association and co-localization, with *cis* defined in this study to be within ±1.5 Mb distance either side of the replicated T2D causal location estimate. This approach provided eQTL location estimates on the LDU maps after Bonferroni correction for the total number of probes tested per 3Mb window. If the eQTL location estimate was within 50kb of the disease susceptibility location, this locus was considered to be a disease eQTL (i.e. associated with both T2D and *cis*-gene expression) and only these eQTL are presented in the result tables. The results table includes a column for the list of *cis*-genes regulated by identified T2D disease loci.

Here we make an important distinction between an eQTL and an eSNP, which relates to ability to make functional inferences about disease loci. In this study an eQTL is defined by a location

estimate for a putative functional variant(s) that regulates gene expression levels for one or more neighbouring genes in a relevant tissue *and* is associated with T2D. In other words, a potential molecular mechanism is immediately suggested for how risk may be conferred by a disease locus, which previously was unknown. By contrast, an eSNP study is defined only by the location of a SNP that is most strongly associated with neighbouring gene expression levels (and may or may not be associated with disease). For eSNP studies, the problems of inconsistency between different lead SNPs associated with disease and expression, between different arrays and across different populations can only be indirectly addressed using genotype imputation methods[21; 22]. For this study it has been established that the majority of the 111 additional susceptibility loci are also eQTLs. This implies that these disease loci may confer risk of T2D, at least in part, via the *cis*-regulation of expression levels for a large number of neighbouring genes (conservatively, a total of 173 genomic disease loci, both new and previously known, that regulate the expression levels of a further 266 cis-genes).

Our final set of *cis*-genes (from Tables 1, 2 and S1) were then further investigated in order to identify which adipose and liver gene expression profiles have previously also shown evidence of association with body mass index (BMI), a well-established co-morbidity of T2D. We used the results generated by an independent gene expression study[23] which was based upon 701 subcutaneous adipose and liver samples collected at Massachusetts General Hospital (MGH study) from morbidly obese individuals (BMI >30) who underwent Roux-en-Y gastric bypass surgery.

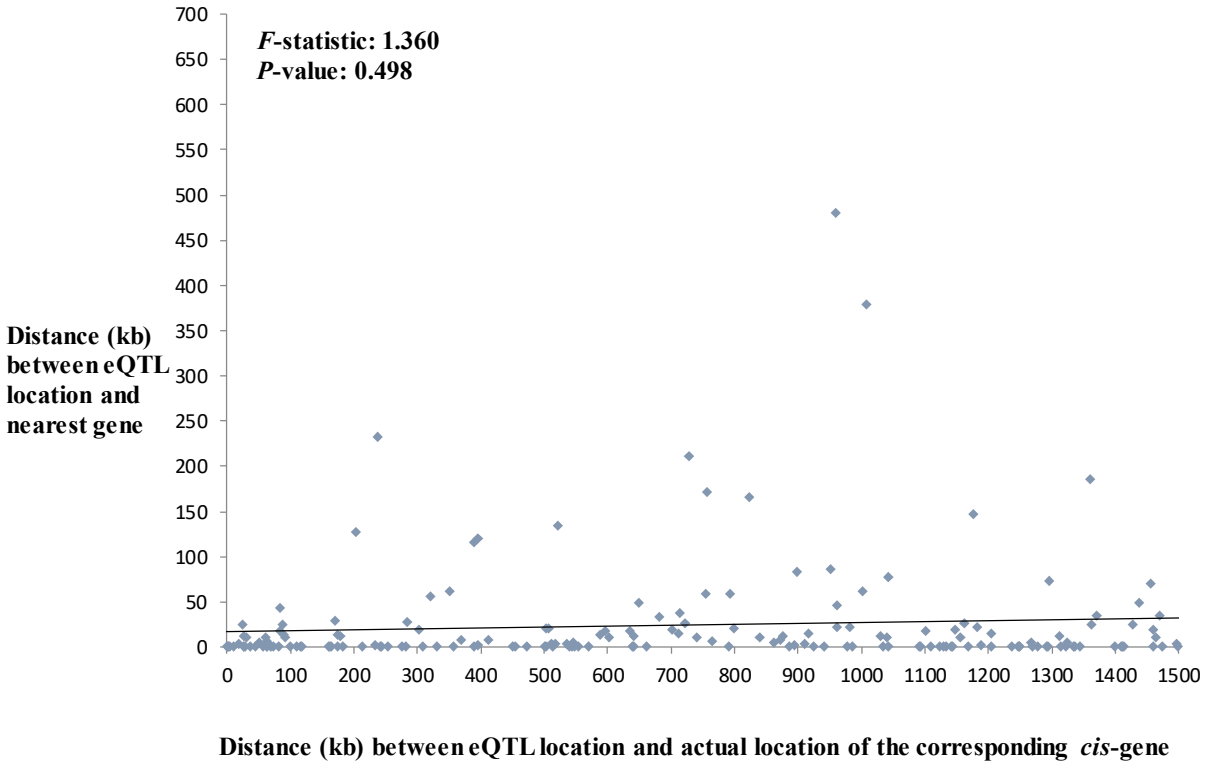
PREVIOUSLY KNOWN T2D LOCI

We also analysed 76 previously known T2D loci[20] to obtain refined location estimates on the same genetic maps. For these loci, we undertook commensurable association analyses by centralising the analytical window on the reported lead SNP. These 76 windows were then examined using the same procedures described above to identify T2D locations and assess whether these are eQTL or not. We confirmed 62 out of 76 loci (signals 112-173) and provide T2D location estimates along with associated *cis*-regulated genes in the supplementary Table S1 (including one, signal 174, from our previous work[24]). Results from the further investigation of the *TCF7L2* locus (signal 117) are provided in the main manuscript. Other notable examples from the supplementary Table S1 are the *HHEX* (signal 149, [MIM: 604420]) and *FTO* (signal 152, [MIM: 610966]). *HHEX* is observed to regulate *MARCH5* [MIM: 610637] expression levels, which codes for a mitochondrial E3 ubiquitin-protein ligase that plays a crucial role in the control of mitochondrial morphology by acting as a positive regulator of mitochondrial fission[25]. This is the first time a mitochondrial fission gene has been implicated as a risk factor for metabolic disease. Despite testing T2D and not obesity, we observed *FTO* to also be a European T2D disease susceptibility locus with a co-located eQTL that regulates *IRX3*[26] [MIM: 612985]. It is possible this observed association may reflect that the Wellcome Trust T2D cases for this study are overweight and/or poorly matched for BMI with the controls[1; 2]. However, we do not present this result in the table, because while nominally significant (P=0.03) and similar to previous studies[26], the eQTL location for *IRX3* did not pass Bonferroni correction for the total number of probes tested for this window. We also observed an eQTL within the promoter of *IRX3* that regulates *IRX5* [MIM: 606195], but we did not further investigate the regulatory landscape of *IRX3*, since the focus of this study was to

identify T2D loci that are also eQTL. The *IRX3* was not observed to be associated with T2D either for this or in other studies.

Supplementary Figure S1: No relationship between the distance of the eQTL to the nearest gene (Y-axis) and the distance of the eQTL from the corresponding *cis*-regulated gene (X-axis).

This regression analysis plot demonstrates that the practice of giving the nearest genes in GWA studies is misleading, since the implicated functional genes the eQTLs regulate are just as likely to be distant or near to the eQTL. The same analysis of the Y and X variables, but only including signals where the distance between T2D sample location estimates (Tables 1 and 2) were < 5kb yielded the same result (P>0.05).



Supplementary Figure S2: No relationship between the distance of the eQTL to the T2D location (Y-axis) and the distance between T2D sample location estimates (X-axis; i.e. between EUR and AA in Table 1 and between the two EUR samples in Table 2).

This regression analysis plot shows no relationship between eQTL co-location and disease co-location and demonstrates that the threshold of <100kb used as the criterion for considering estimated disease loci to be co-located and replicated, does not introduce any bias compared to the more conservative threshold of <50kb for the co-location of disease and eQTL.

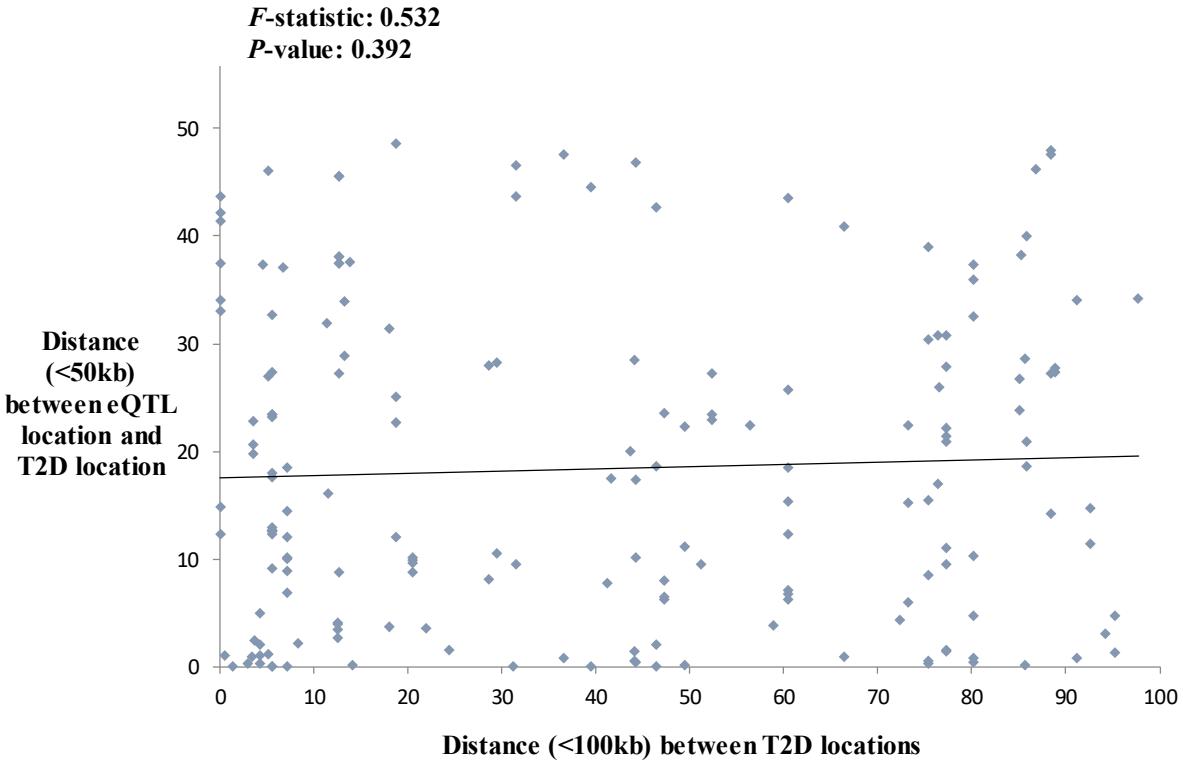


Table S1. Refined information on the previously known T2D loci and their regulatory role of neighbouring gene expression

All locations and distances are given in build 36; ¶Replication with the WTCCC (W), NIDDK AA (A), Metaboship (M) datasets; [a]T2D associated intervals in kb (<100) that harbour T2D locations between datasets; [b]Location estimates for the European (E) GWAS; [c]Location estimates for the African-American (A) GWAS; [d]Location estimates for the Metaboship European (E) samples, signals with low SNP coverage '-' were not meta-analysed; [e]Genes in bold denote the intragenic localization and genes with '+' for self-regulatory; [f]Number of *cis*-genes regulated by the eQTL; [g]List of *cis*-genes associated with eQTLs that co-located within <50kb of the T2D locations on the genetic maps; cis-genes with '*' have previously shown evidence of association between Body Mass Index for morbidly obese and adipose/liver expression profiles[23]; [h]Distance in kb (<50) between eQTL and T2D locations, the minimum is given when more than one *cis*-gene is implicated. Previously observed loci for signals 112-173 are derived from[20] and signal **174** from[24].

Signal	Known locus	Lead SNP	Lead SNP b36	chr	Data¶	Meta *P*-value	Distance between locations[a]	T2D location GWAS-E[b]	T2D location GWAS-A[c]	T2D location metabo-E[d]	Nearest gene to T2D locations[e]	no. of *cis*-genes[f]	eQTL associated cis-genes[g]	eQTL distance from T2D[h]
112	*BCL11A*	rs243088	60422	2p	WAM	4.22E-34	0	60441	60427	60441	*MIR4432*	1	*PEX13*	31
113	*TMEM154*	rs6813195	153740	4q	WAM	1.38E-08	1	153747	153739	153740	*TMEM154*	0	-	-
114	*ANKRD55*	rs459193	55843	5q	WAM	2.15E-14	2	55834	55926	55832	***LOC101928448***	0	-	-
115	*CDKAL1*	rs7756992	20788	6p	WAM	5.99E-180	0	20787	20750	20787	***CDKAL1***	0	-	-
116	*CDKN2A/B*	rs944801	22042	9p	WAM	8.16E-41	1	21987	21986	22022	***CDKN2A/B***	2	*KIAA1797, MTAP*	26
117	*TCF7L2*	rs7903146	114748	10q	WAM	3.55E-86	9	114736	114745	114737	***TCF7L2***	1	*GPAM*	28
118	*RBMS1*	rs7569522	161055	2q	WA	4.42E-21	95	160935	160840	>100kb	*RBMS1*	1	*RBMS1**	4
119	*KCNK16*	rs1535500	39392	6p	WA	1.04E-07	86	39505	39419	-	***KIF6***	0	-	-
120	*ZFAND6*	rs11634397	78219	15q	WA	1.95E-03	67	78193	78126	-	***ZFAND6***	0	-	-
121	*TMEM163*	rs6723108	135196	2q	AM	3.54E-12	1	>100kb	135313	135312	***ACMSD***	0	-	-
122	*KCNQ1*	rs231361	2648	11p	AM	1.96E-14	15	ns	2648	2663	***KCNQ1***	0	-	-
123	*KCNJ11*	rs5215	17365	11p	AM	5.32E-26	2	ns	17384	17382	***ABCC8***	2	*MYOD1, UEVLD*	0
124	*HNF1B*	rs4430796	33172	17q	AM	1.21E-11	30	ns	33135	33165	***HNF1B***	0	-	-
125	*PSMD6*	rs12497268	64065	3p	A	1.87E-05	-	>100kb	63759	-	*C3orf49*	0	-	-
126	*HMGA2*	rs2261181	64499	12q	A	5.53E-03	-	>100kb	64404	ns	*RPSAP52*	0		-
127	*MAEA*	rs6815464	1300	4p	A	1.84E-03	-	ns	1267	ns	*MAEA*	3	*CTBP1, KIAA1530, CRIPAK**	8
128	*ANK1*	rs516946	41638	8p	A	7.57E-07	-	ns	41608	-	*AGPAT6*	1	*ANK1*	14
129	*TLE4*	rs13292136	81142	9q	A	2.97E-02	-	ns	81146	-	*CHCHD9*	0	-	-
130	*FAF1*	rs17106184	50683	1p	A	5.38E-02	-	ns	50894	-	***FAF1***	2	*EPS15, TXNDC12**	4
131	*BCAR1*	rs7202877	73805	16q	A	2.92E-03	-	ns	73490	-	***WDR59***	1	*FA2H*	16
132	*SRR*	rs2447090	2246	17p	A	2.28E-07	-	ns	2039	-	***SMG6***	7	*SRR, RPA1,CAMKK1, ZZEF1, TSR1, SMG6*, TMEM93*	0
133	*PEPD*	rs8182584	38602	19q	A	7.27E-03	-	ns	38543	-	*CEBPG*	0	-	-
134	*ADCY5*	rs11717195	124565	3q	WM	3.23E-23	7	124531	>100kb	124538	*ADCY5*	2	*SEC22A, CCDC14**	15
135	*POU5F1*	rs3130501	31244	6p	WM	3.21E-35	4	31773	>100kb	31777	*LINC00243*	13	*LST1, LY6G6C, C6ORF25, MSH5, SLC44A4*, VARS2, DDR1, FLOT1, ABCF1, HLA-DQB2, TAP2*, TRIM15, TRIM40*	0
136	*DGKB*	rs6960043	15019	7p	WM	3.90E-47	2	15034	>100kb	15032	***DGKB***	0	-	-
137	*TSPAN8*	rs7955901	69720	12q	WM	4.01E-37	11	69867	>100kb	69878	*TSPAN8*	2	*LRRC10, FRS2*	0
138	*MPHOSPH9*	rs4275659	122014	12q	WM	4.46E-05	61	121953	>100kb	122014	*VPS37B, **ABCB9**+*	9	*DNAH10, PITPNM2, ABCB9, VPS37B, TMED2, RSRC2, ZCCHC8, NCOR2, DIABLO*	0
139	*HMG20A*	rs7177055	75620	15q	WM	2.25E-04	40	75058	>100kb	75098	*PSTPIP1*	2	*HMG20A, TSPAN3*	0
140	*IRS1*	rs7578326	226729	2q	WM	3.08E-41	59	226788	ns	226729	***LOC646736***	0	-	-
141	*PPARG*	rs13081389	12265	3p	WM	4.33E-17	18	12311	ns	12292	***PPARG***	1	*WNT7A*	6
142	*ADAMTS9*	rs6795735	64680	3p	WM	1.01E-13	1	64707	ns	64706	***ADAMTS9***	0	-	-
143	*IGF2BP2*	rs4402960	186994	3q	WM	1.89E-23	1	187031	ns	187032	***IGF2BP2***	0	-	-
144	*ARL15*	rs702634	53307	5q	WM	2.85E-04	99	53347	ns	53248	***ARL15***	1	*FST**	19
145	*ZBED3*	rs6878122	76463	5q	WM	1.81E-11	0	76457	ns	76457	***ZBED3***	1	*PDE8B**	16
146	*JAZF1*	rs849135	28163	7p	WM	2.51E-67	90	28226	ns	28136	***JAZF1***	0	-	-
147	*KLF14*	rs13233731	130088	7q	WM	2.59E-06	46	130074	ns	130120	*KLF14*	0	-	-

148	*TP53INP1*	rs7845219	96007	8q	WM	1.28E-11	97	96132	ns	96035	*NDUFAF6, TP53INP1*	6	*GDF6, GEM, MTERFD1, FAM92A1, C8orf37, KIAA1429*	0
149	*HHEX/IDE*	rs1111875	94453	10q	WM	2.27E-61	21	94490	ns	94469	*HHEX*	1	*MARCH5*	1
150	*HNF1A*	rs12427353	119911	12q	WM	3.79E-31	74	119794	ns	119720	***SPPL3***	1	*MSI1*	48
151	*PRC1*	rs8042680	89322	15q	WM	3.68E-56	59	89245	ns	89304	*MAN2A2,* ***RCCD1***$^{+}$	4	*RCCD1, UNC45A, IQGAP1*, FAM174B**	1
152	*FTO*	rs9936385	52377	16q	WM	2.52E-176	11	52368	ns	52357	***FTO***	0	-	-
153	*MC4R*	rs12970134	56036	18q	WM	5.19E-21	1	55879	ns	55880	*RPS3A*	0	-	-
154	*GCKR*	rs780094	27595	2p	M	2.16E-17	-	>100kb	>100kb	27228	***TCF23***	2	*PLB1, KHK*	6
155	*GCC1*	rs17867832	126784	7q	M	1.01E-02	-	>100kb	ns	126964	*GCC1*	1	*IMPDH1*	23
156	*NOTCH2*	rs10923931	120319	1p	M	1.42E-24	-	ns	ns	120238	***ADAM30***	0	-	-
157	*PROX1*	rs2075423	212221	1q	M	1.71E-04	-	ns	ns	212226	***PROX1***	0	-	-
158	*THADA*	rs10203174	43544	2p	M	4.35E-15	-	ns	ns	43555	***THADA***	0	-	-
159	*GRB14*	rs13389219	165237	2q	M	7.31E-03	-	ns	ns	165209	*GRB14*	1	*SCN2A*	32
160	*WFS1*	rs4458523	6341	4p	M	2.61E-46	-	ns	ns	6359	*WFS1*	3	*GRPEL1, STK32B, KIAA0232*	2
161	*SLC30A8*	rs3802177	118254	8q	M	6.10E-05	-	ns	ns	118251	***SLC30A8***	1	*SAMD12*	0
162	*GLIS3*	rs10758593	4282	9p	M	2.45E-10	-	ns	ns	4273	***GLIS3***	0	-	-
163	*CDC123*	rs11257655	12348	10p	M	4.54E-06	-	ns	ns	12189	***DHTKD1***	0	-	-
164	*ARAP1*	rs1552224	72111	11q	M	1.67E-03	-	ns	ns	72534	*FCHSD2*	1	*POLD3*	8
165	*CILP2*	rs10401969	19269	19p	M	1.35E-39	-	ns	ns	19188	***NCAN***	3	*ATP13A1, KIAA0892, TM6SF2**	2
166	*GIPR*	rs8108269	50850	19q	M	1.49E-04	-	ns	ns	51124	*NOVA2*	0	-	-
167	*RND3*	rs7560163	151346	2q	W	4.53E-02	-	151248	ns	-	*LOC101929282*	0	-	-
168	*SSR1*	rs9505118	7235	6p	W	4.31E-02	-	7229	>100kb	-	***SSR1***	1	*BMP6*	46
169	*ZMIZ1*	rs12571751	80613	10q	W	4.06E-04	-	80700	ns	-	***ZMIZ1***	1	*DYDC2*	9
170	*GRK5*	rs10886471	121139	10q	W	3.28E-02	-	121233	ns	-	*RGS10*	0	-	-
171	*CCND2*	rs11063069	4245	12p	W	1.04E-02	-	4170	ns	-	*CCND2*	0	-	-
172	*VPS26A*	rs1802295	70601	10q	W	1.03E-02	-	70421	ns	ns	***KIAA1279***	1	*HERC4*	49
173	*C2CD4A*	rs4502156	60170	15q	W	2.02E-02	-	59905	ns	ns	*VPS13C*	4	*APH1B, RORA, VPS13C, TPM1*	14
174	*ABCC5*	-	-	3q	WA	1.00E-07	0	185136	185136	-	***ABCC5***$^{+}$	1	*ABCC5*	0

Table S2. Demographic characteristics for targeted re-sequence T2D case/ control European samples.

Cases	Variable	Obs	Mean	Std Dev.	Min	Max
Female	Age	57	47.4	7.0	26.0	72.0
	BMI	57	27.1	4.6	17.6	34.7
Male	Age	49	43.5	7.5	20.0	53.0
	BMI	49	25.9	3.5	17.6	34.5
Controls						
Female	Age	57	47.8	7.3	26.0	72.0
	BMI	57	27.7	4.0	21.1	34.8
Male	Age	49	40.7	7.1	20.0	53.0
	BMI	49	27.0	3.6	18.7	34.5

References

1. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678.
2. Voight, B.F., Kang, H.M., Ding, J., Palmer, C.D., Sidore, C., Chines, P.S., Burtt, N.P., Fuchsberger, C., Li, Y., Erdmann, J., et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 8, e1002793.
3. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., et al. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42, 579-589.
4. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 20 (1997), 1183-1197.
5. Expert Committee on the, D., and Classification of Diabetes, M. (2003). Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 26 Suppl 1, S5-20.
6. Definition, diagnosis and classification of diabetes mellitus, report of a WHO consultation, part 1: diagnosis and classification of diabetes mellitus (1999). In, W.H. Organization, ed. Geneva.
7. Palmer, N.D., McDonough, C.W., Hicks, P.J., Roh, B.H., Wing, M.R., An, S.S., Hester, J.M., Cooke, J.N., Bostrom, M.A., Rudock, M.E., et al. (2012). A genome-wide association search for type 2 diabetes genes in African Americans. PLoS One 7, e29202.
8. Wellcome Trust Case Control Consortium 2 (2008). In, W.T.S. Institute, ed. (Wellcome Trust Sanger Institute.
9. Martin, L.J., Comuzzie, A.G., Dupont, S., Vionnet, N., Dina, C., Gallina, S., Houari, M., Blangero, J., and Froguel, P. (2002). A quantitative trait locus influencing type 2 diabetes susceptibility maps to a region on 5q in an extended French family. Diabetes 51, 3568-3572.
10. Vionnet, N., Hani, E.H., Dupont, S., Gallina, S., Francke, S., Dotte, S., De Matos, F., Durand, E., Lepretre, F., Lecoeur, C., et al. (2000). Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. Am J Hum Genet 67, 1470-1480.
11. Hager, J., Dina, C., Francke, S., Dubois, S., Houari, M., Vatin, V., Vaillant, E., Lorentz, N., Basdevant, A., Clement, K., et al. (1998). A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. Nat Genet 20, 304-308.
12. Meyre, D., Lecoeur, C., Delplanque, J., Francke, S., Vatin, V., Durand, E., Weill, J., Dina, C., and Froguel, P. (2004). A genome-wide scan for childhood obesity-associated traits in French families shows significant linkage on chromosome 6q22.31-q23.2. Diabetes 53, 803-811.
13. Maniatis, N., Collins, A., Xu, C.F., McCarthy, L.C., Hewett, D.R., Tapper, W., Ennis, S., Ke, X., and Morton, N.E. (2002). The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. Proc Natl Acad Sci U S A 99, 2228-2233.
14. Maniatis, N. (2007). Linkage disequilibrium maps and disease-association mapping. Methods Mol Biol 376, 109-121.
15. Lau, W., Kuo, T.Y., Tapper, W., Cox, S., and Collins, A. (2007). Exploiting large scale computing to construct high resolution linkage disequilibrium maps of the human genome. Bioinformatics 23, 517-519.
16. Maniatis, N., Collins, A., Gibson, J., Zhang, W., Tapper, W., and Morton, N.E. (2004). Positional cloning by linkage disequilibrium. Am J Hum Genet 74, 846-855.
17. Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A., et al. (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 44, 1084-1089.
18. Maniatis, N., Collins, A., and Morton, N.E. (2007). Effects of single SNPs, haplotypes, and whole-genome LD maps on accuracy of association mapping. Genet Epidemiol 31, 179-188.

19. Morton, N., Maniatis, N., Zhang, W., Ennis, S., and Collins, A. (2007). Genome scanning by composite likelihood. Am J Hum Genet 80, 19-28.
20. Mahajan, A., Go, M.J., Zhang, W., Below, J.E., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 46, 234-244.
21. Morris, A.P. (2014). Fine mapping of type 2 diabetes susceptibility loci. Curr Diab Rep 14, 549.
22. Price, A.L., Spencer, C.C., and Donnelly, P. (2015). Progress and promise in understanding the genetic basis of common diseases. Proceedings Biological sciences / The Royal Society 282.
23. Greenawalt, D.M., Dobrin, R., Chudin, E., Hatoum, I.J., Suver, C., Beaulaurier, J., Zhang, B., Castro, V., Zhu, J., Sieberts, S.K., et al. (2011). A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome research 21, 1008-1016.
24. Direk, K., Lau, W., Small, K.S., Maniatis, N., and Andrew, T. (2014). ABCC5 transporter is a novel type 2 diabetes susceptibility gene in European and African American populations. Ann Hum Genet 78, 333-344.
25. Chan, D.C. (2006). Mitochondria: dynamic organelles in disease, aging, and development. Cell 125, 1241-1252.
26. Claussnitzer, M., Hui, C.C., and Kellis, M. (2016). FTO Obesity Variant and Adipocyte Browning in Humans. The New England journal of medicine 374, 192-193.