Supplemental Information for:

Meta-analysis of genome-wide association study data identifies additional type 1 diabetes loci

Jason D Cooper[1], Deborah J Smyth[1], Adam M Smiles[2], Vincent Plagnol[1], Neil M Walker[1], James Allen[1], Kate Downes[1], Jeffrey C Barrett[1], Barry Healy[1], Josyf C Mychaleckyj[3], James H Warram[2] and John A Todd[1]

[1]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0XY, UK. [2]Research Division, Joslin Diabetes Center, Boston, MA, USA. [3]Center for Public Health Genomics, University of Virginia, VA, USA.

Supplementary Methods

All analyses were carried out using the snpMatrix packaged[1] in R, distributed by the BioConductor project (http://www.bioconductor.org).

Subjects

The T1D cases were recruited as part of the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory's British case collection (Genetic Resource Investigating Diabetes), which is a joint project between the University of Cambridge Department of Paediatrics and the Department of Medical Genetics at the Cambridge Institute for Medical Research. Most cases were less than 16 years of age at the time of collection; all were less than 17 years of age at diagnosis, resided in Great Britain and were of self-reported white ethnicity and of European descent.

The controls were obtained from the British 1958 Birth Cohort, an ongoing follow-up study of all people born in Great Britain during one week in 1958 (National Child Development Study)[2]. All controls were of self-reported white ethnicity and of European descent.

All families were of self-reported white ethnicity and of European descent, with two parents and at least one affected child. The family collection consisted of 468 multiplex families from the Diabetes UK Warren I repository[3], 331 multiplex families from the USA Human Biological Data Interchange[4], 881 multiplex and simplex families from Finland[5], 263 multiplex and simplex families from Northern Ireland[6], 124 simplex families from the Diabetes UK Warren III repository[7], 350 simplex families from Norway[8], 411 simplex families from Romania[9].

The National Institute of Mental Health (NIMH)[10] control data were downloaded on 29th November 2007 and consisted of 1,727 samples; we updated the data on 12th February 2008 to version 6.11.

The Genetics of Kidneys in Diabetes (GoKinD)[11,12] study case data were downloaded on 20th December 2007 and consisted of 1,825 samples: 904 diabetic nephropathy cases; 20 diabetic nephropathy case parents; 881 diabetic controls; and 20 diabetic control parents.

Scoring algorithms

The three GWA studies had been scored using three different algorithms: Wellcome Trust Case Control Consortium (WTCCC)[13] data had been scored using CHIAMO[13]; GoKinD data using Birdseed[14]; and, NIMH data using BRLMM[15]. As we formed the American case-control data using cases from GoKinD and controls from NIMH, to avoid the false-positive associations caused by a differential bias in genotype calling between cases and controls[16], we had to re-score the data. As GoKinD provided both CEL and signal intensity data, we used the signal intensity data for the convenience of not having to normalise the CEL data. However, the NIMH provided only CEL data and normalisation was required to produce the signal intensity data. We used the algorithm JAPL developed by Plagnol *et al.*[17] to re-score the data. One GoKinD plate (86 samples) was dropped as it was being re-genotyped. We note that JAPL scores every sample for every SNP and provides a clustering quality score for each SNP, we rejected SNPs with a score less than 2 (a level determined from the inspected of signal intensity plots).

The WTCCC data was also re-scored using JAPL, but for consistency with the reported WTCCC results, we used the CHIAMO scored data, using a posterior

probability threshold of 0.9. A consequence of using the CHIAMO scored data was that the majority of SNP alleles did not align with the GoKinD and NIMH data as the WTCCC data was aligned with International HapMap Project data[18].

SNP quality control

Although WTCCC SNP exclusion list was provided with the data, we only excluded SNPs that were significantly different ($P = 5.7 \times 10^{-7}$(ref. [13])) between the two control groups (British 1958 Birth Cohort and UK Blood Service). SNP quality control filters consisted of clustering quality metrics, from each study, we excluded: SNPs with a MAF < 0.05 as the performance of scoring algorithms deteriorates with minor allele frequency; SNPs with extreme deviation from Hardy-Weinberg equilibrium (HWE) in controls (defined as $P < 5 \times 10^{-7}$) as this could be indicative of serious genotyping failure; and, WTCCC SNPs with a call-rate < 0.99 (GoKinD and NIMH SNPs had a call-rate = 1, see above).

In addition, we identified poorly clustered SNPs using a Bayes factor for the comparison of the hypothesis that the SNP alleles have been switched with the hypothesis that they have not been switched. A Bayes factor > 20 identified 726 SNPs with poor clustering, inspection of allele signal intensity plots for the first 100 of the 726 SNPs, examined going across the genome, revealed that the main problem (86/100) was when two genotypes were scored in GoKinD and three in NIMH (**Supplementary Figure 8**). We also inspected the signal intensity plots for the most associated 511 SNPs ($P \leq 1 \times 10^{-4}$) outside the major histocompatibility complex (MHC) region and other T1D associated regions, excluding 355 SNPs.

Sample quality control

We applied the WTCCC sample exclusion list provided with the data, except for the sample call-rate filter, which we increased from 0.03 to 0.05 (GoKinD and NIMH samples had a call-rate = 1, see above). The WTCCC sample exclusion list included the filters described below. Duplicates and first or second degree relatives had been removed from the NIMH data before download.

To detect samples with insufficient quality and quantity of DNA, we used genome-wide sample heterozygosity, such samples are clearly visible when heterozygosity is plotted against call-rate[13]. As previously stated, the WTCCC sample exclusion list included this filter (excluding 0.225 < heterozygosity > 0.3) (ref.[13]). We note that the filter may depend on the platform and the calling algorithm, for GoKinD and NIMH, we excluded samples with heterozygosity < 0.3. We excluded two American cases and one American control.

To identify duplicates and first or second degree relatives, we used identity-by-state to estimate the fraction of identical genotypes between each pair of samples (that is, the number of identical-by-state alleles at each locus divided by the number of loci). We excluded four duplicates and two relatives within the American cases.

To exclude subjects with substantial non-European ancestry, we used the International HapMap Project data[18] to identify SNPs, available on the Affymetrix 500K SNP chipset, that differentiated between the three HapMap populations (CEU, YRI and JPT+CHB) to derive two principal component scores for ancestry. We then calculated these scores for the GoKinD and NIMH data whose alleles had been aligned with HapMap, excluding any outlying subjects, defined using a P-value $\leq 1 \times 10^{-4}$ (one false rejection in 10,000 subjects). We started with a 1,000 HapMap population differentiating SNPs and dropped about 50 SNPs whose alleles did not align between the GWA studies and HapMap. We note that far fewer SNPs should be

sufficient to differentiate between the HapMap populations. We excluded 90 American cases (**Supplementary Figure 9**) and 18 American controls (**Supplementary Figure 10**).

Disease models

We performed score tests assuming multiplicative allelic effects (1-degree-of-freedom (df)) and genotypic effects (2-df). The 1-df test is the Cochran-Armitage ("trend") test or when controlling for geographic variation or population structure, the Mantel-extension test.

Population structure

To control for population structure, we stratified the British (WTCCC/follow-up) case-control data by 12 geographical regions of Great Britain[16] and the American (GoKinD/NIMH) case-control data by ten strata based on a "propensity score"[19] derived from principal components[20]. The use of principal components to control for population structure in the American samples is more appropriate than using geographical information, which does not correlate well with the genetic background of the American samples[21]. As principal components often reflect extended linkage disequilibrium, we derived principal components using thinned genome-wide SNP data; working across the genome, we dropped SNPs in $r^2 > 0.2$ with a preceding SNP. A particular problem of using cases and controls from different GWA studies is that the principal components picked out the differences in DNA preparation and scoring. A solution was to standardize the genotype score within study when generating the between subjects covariance matrix. As the resulting principal components had had most of their population information removed, to re-introduce the population information, we re-calculated the principal components for both studies together.

We calculated the propensity score by logistic regression of the case/control indicator on the first six principal components (coefficient *z*-scores ≥ 0.6), the ten strata were defined as the deciles of this score (**Supplementary Figure 11**). The use of the propensity score deciles to control for population structure reduced the inflation of the test statistic from 18% (**Supplementary Figure 12**) to 14% (**Supplementary Figure 13**).

Independent associations

To test whether a SNP was independently associated with disease when there was a known T1D loci in the vicinity, we used logistic regression. For the purposes of this analysis, we assumed no specific mode of inheritance for the T1D loci, for a SNP (A>a), genotype risks of a/a and a/A were modelled relative to the A/A genotype. We then used a 1-df test for adding a SNP to the model by assuming multiplicative allelic effects for the additional SNPs. **Supplementary Table 4** contains a summary of the analysis to test whether the intergenic SNP, rs947474, on 10p15 was associated with T1D independently of the T1D associated *IL2RA* region[22].

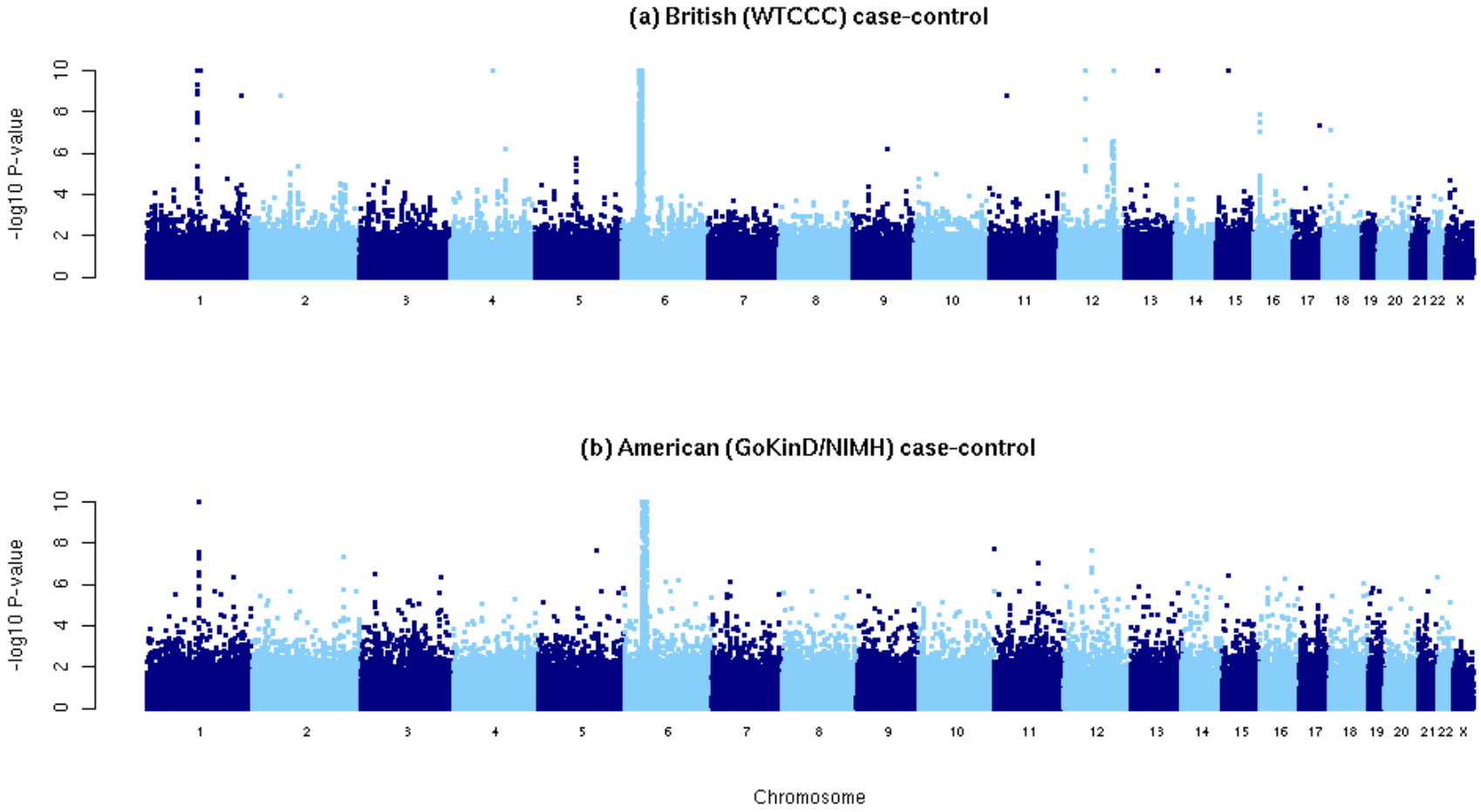
Associated regions

We defined approximate associated regions using 2,942 British controls and the positions of HapMap recombination hotspots that contained the set of Affymetrix SNPs with an $r^2 \geq 0.1$ and $D' \geq 0.1$ with the most associated SNP.

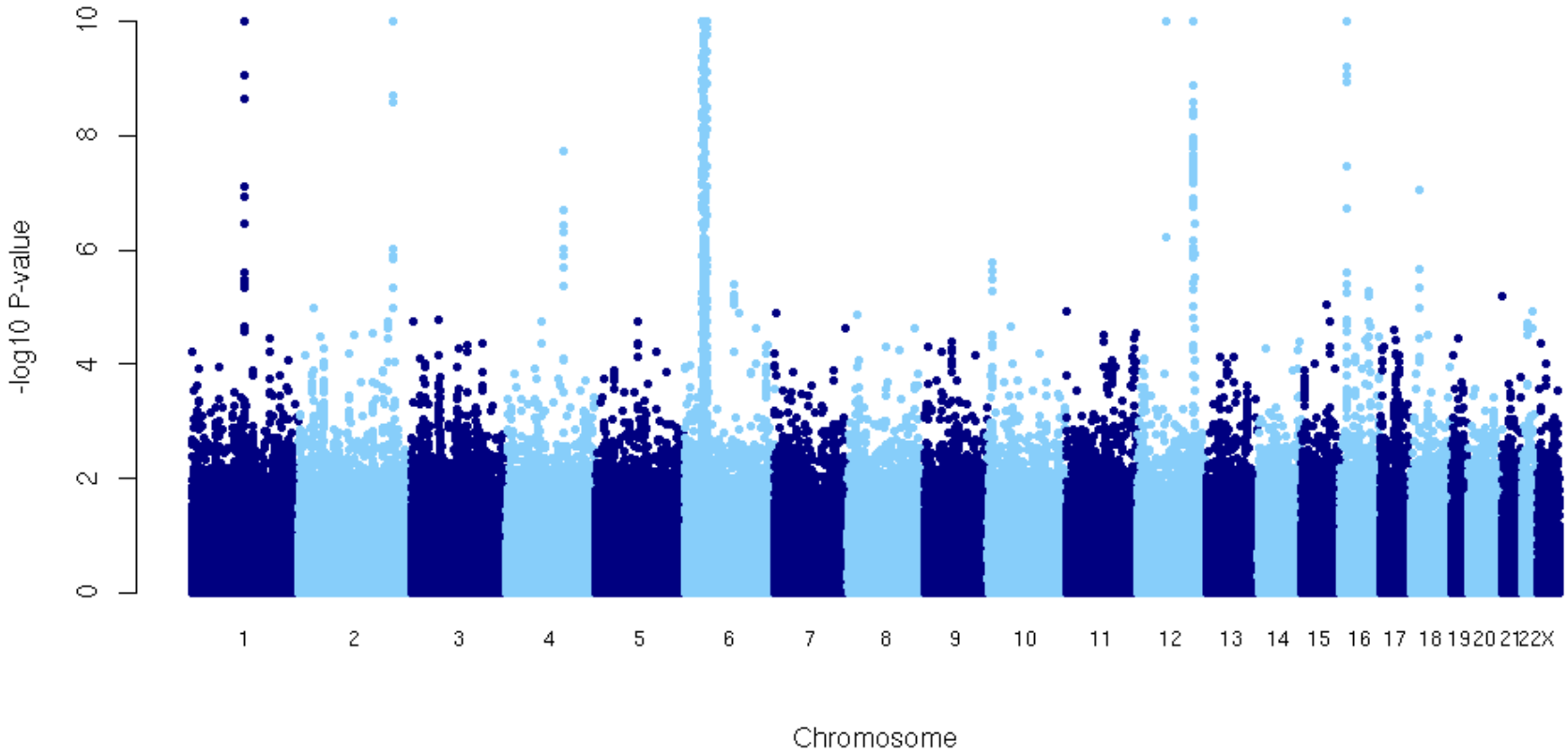
Supplementary Note

The available GoKinD data, such as allele signal intensities and genotype data have had SNPs removed before being made available for download. We downloaded 467,142 unfiltered SNP allele signal intensities for re-scoring.

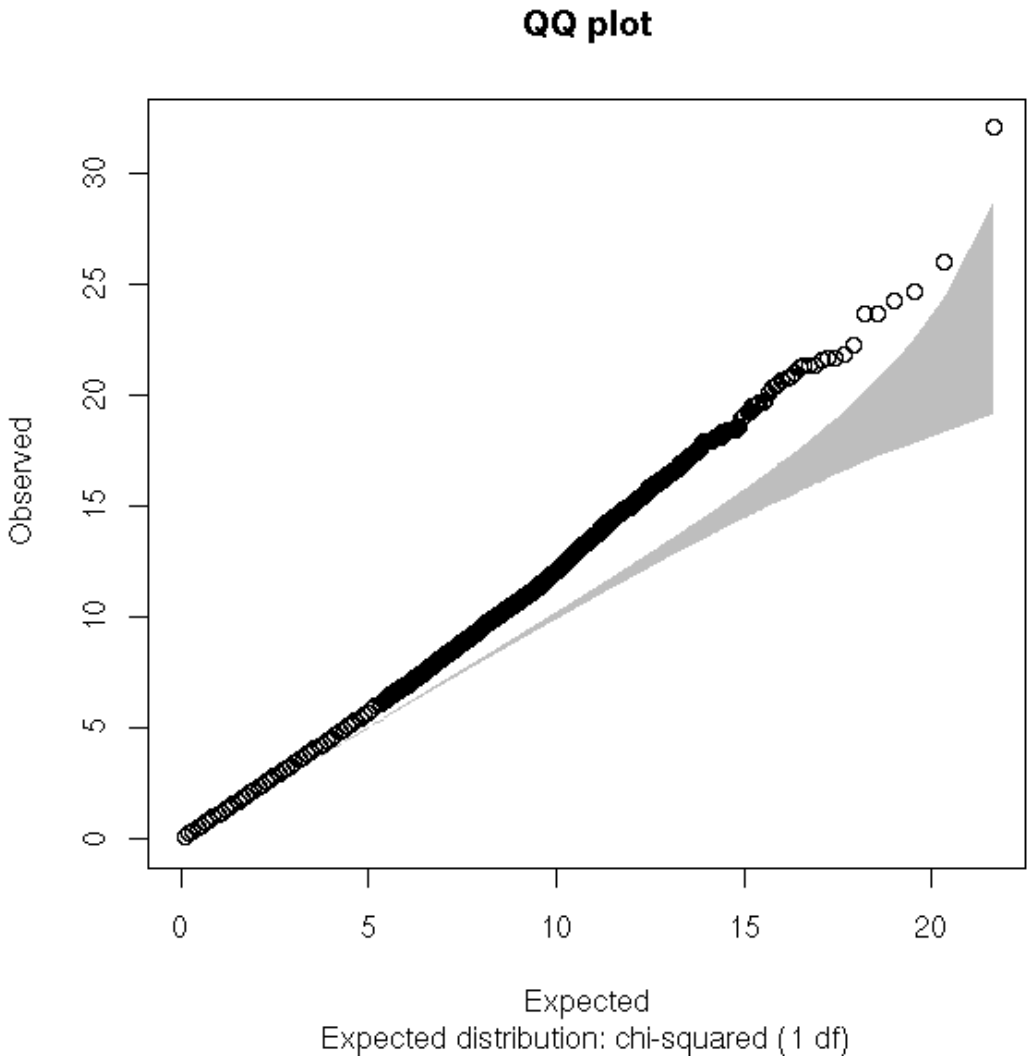
Supplementary Figures



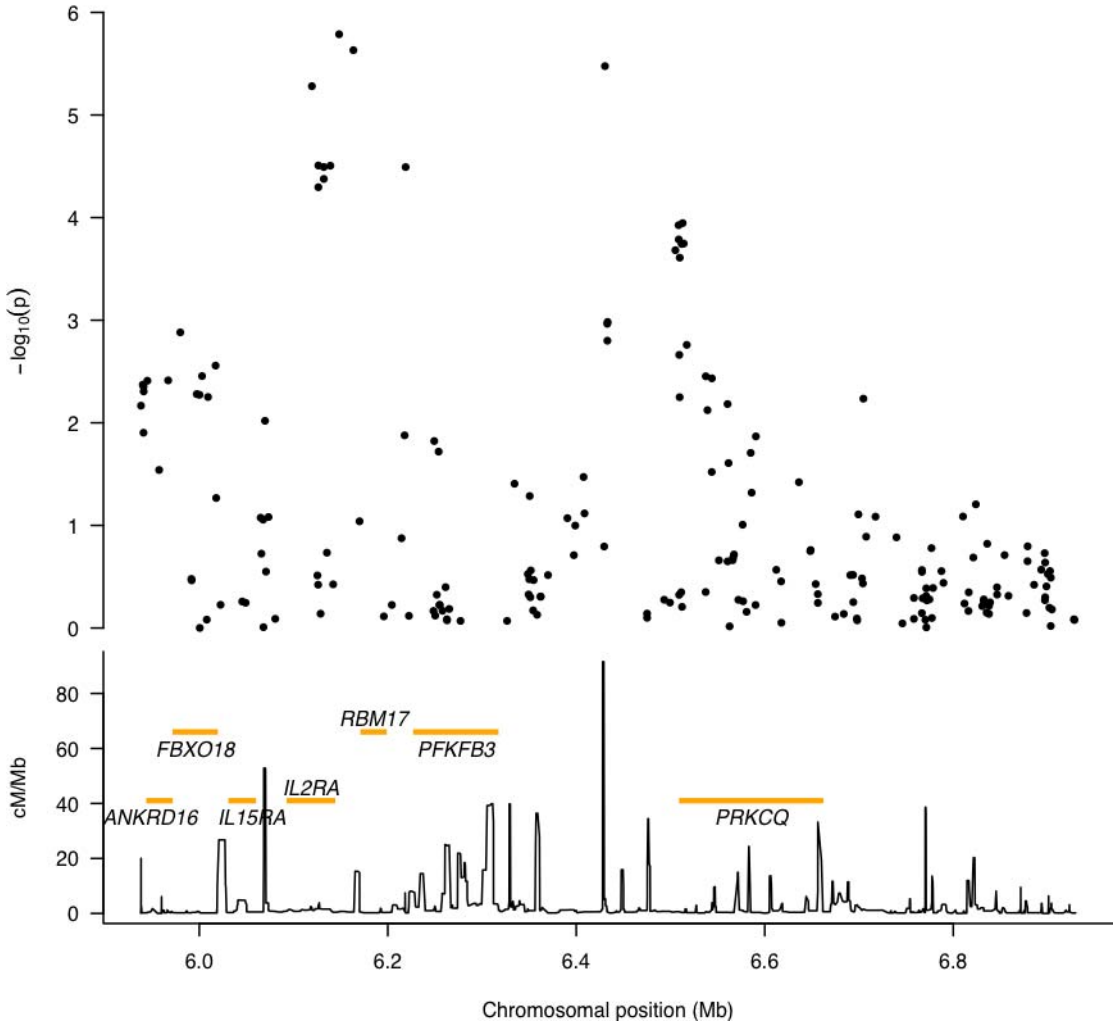
Supplementary Figure 1a. A comparison of the stratified trend test (1-df) *P*-values for the British and American GWA studies. The *P*-values for the T1D loci are shown in **Supplementary Table 1**. The *P*-values were censored at $1x10^{-10}$.



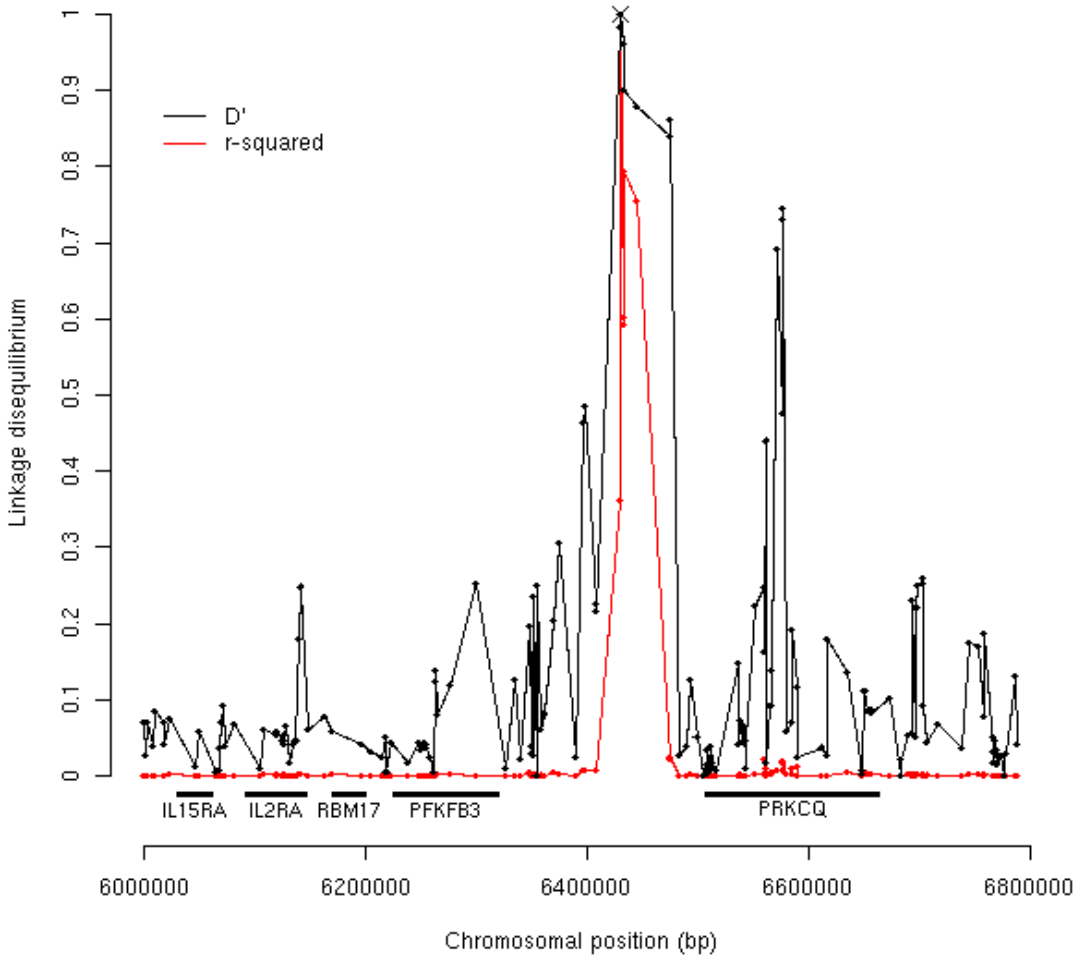
Supplementary Figure 1b. The stratified trend test (1-df) *P*-values for the meta-analysis of British and American GWA studies. The *P*-values for the T1D loci are shown in **Supplementary Table 1**. The *P*-values were censored at $1x10^{-10}$.



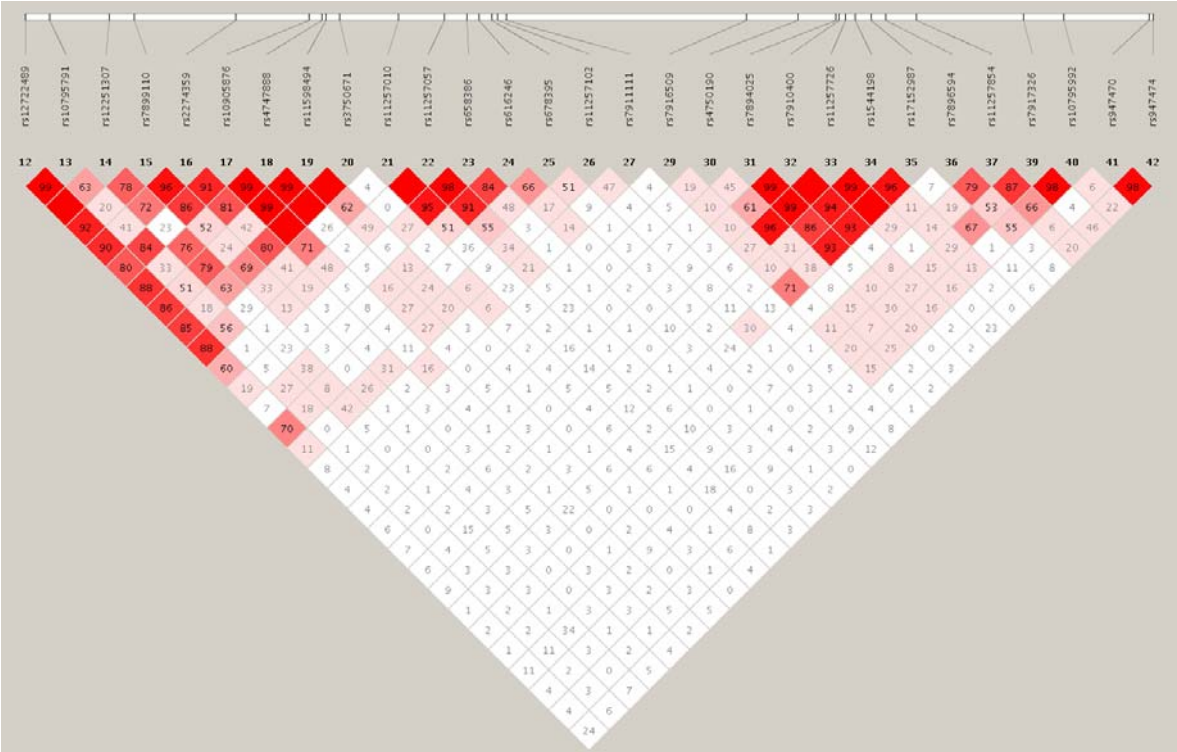
Supplementary Figure 2. A quantile-quantile plot for the meta-analysis stratified 1-df tests. We excluded known T1D loci and SNPs with an $r^2 \geq 0.1$ with them, 303,651 SNPs remained. Overdispersion factor, $\lambda = 1.12$.



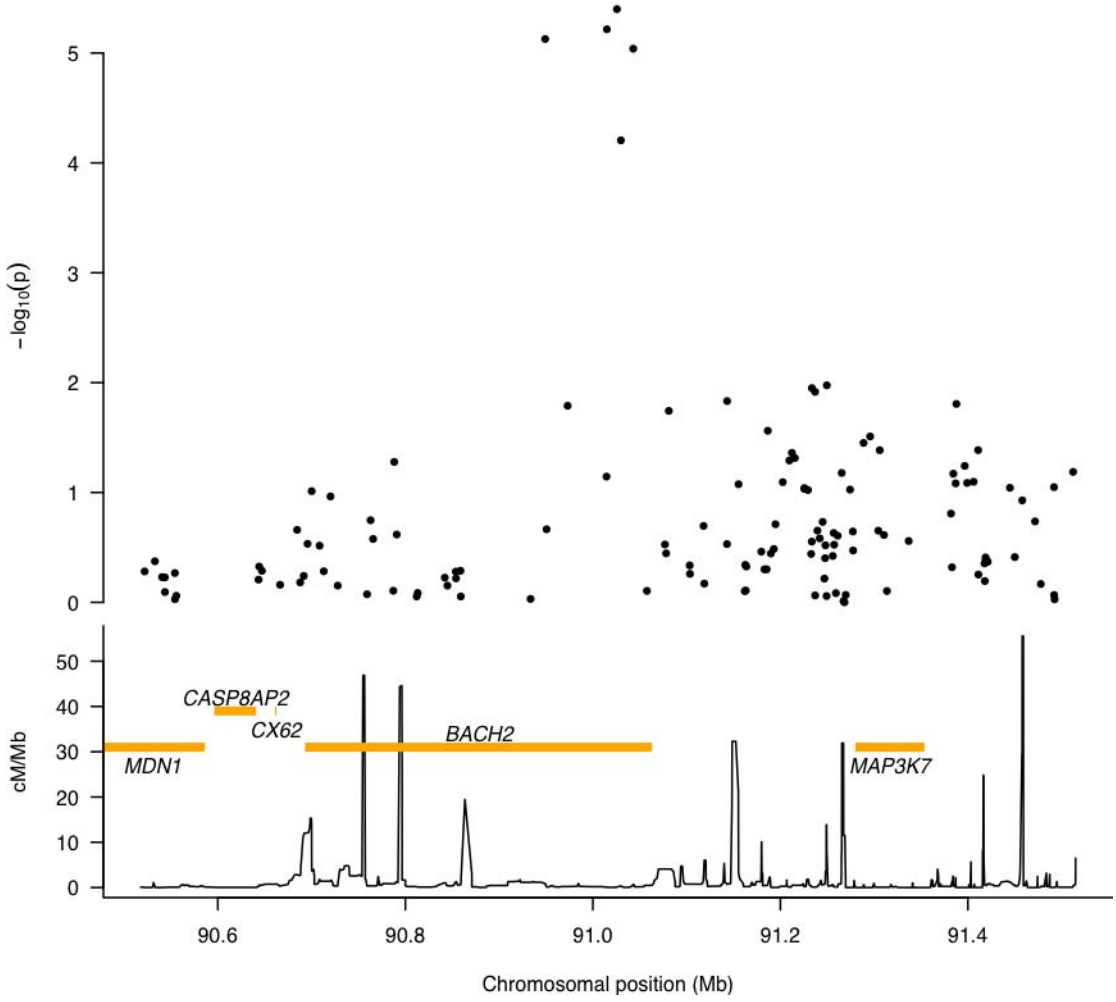
Supplementary Figure 3a. The region about the SNP, rs947474, on chromosome 10p15. The upper panel shows the –log10 *P*-values for the Affymetrix 500K SNPs in the region and the lower panel, shows the recombination map. Genes are shown in yellow. The 234 kb associated region was defined as 6.43 to 6.77 Mb.



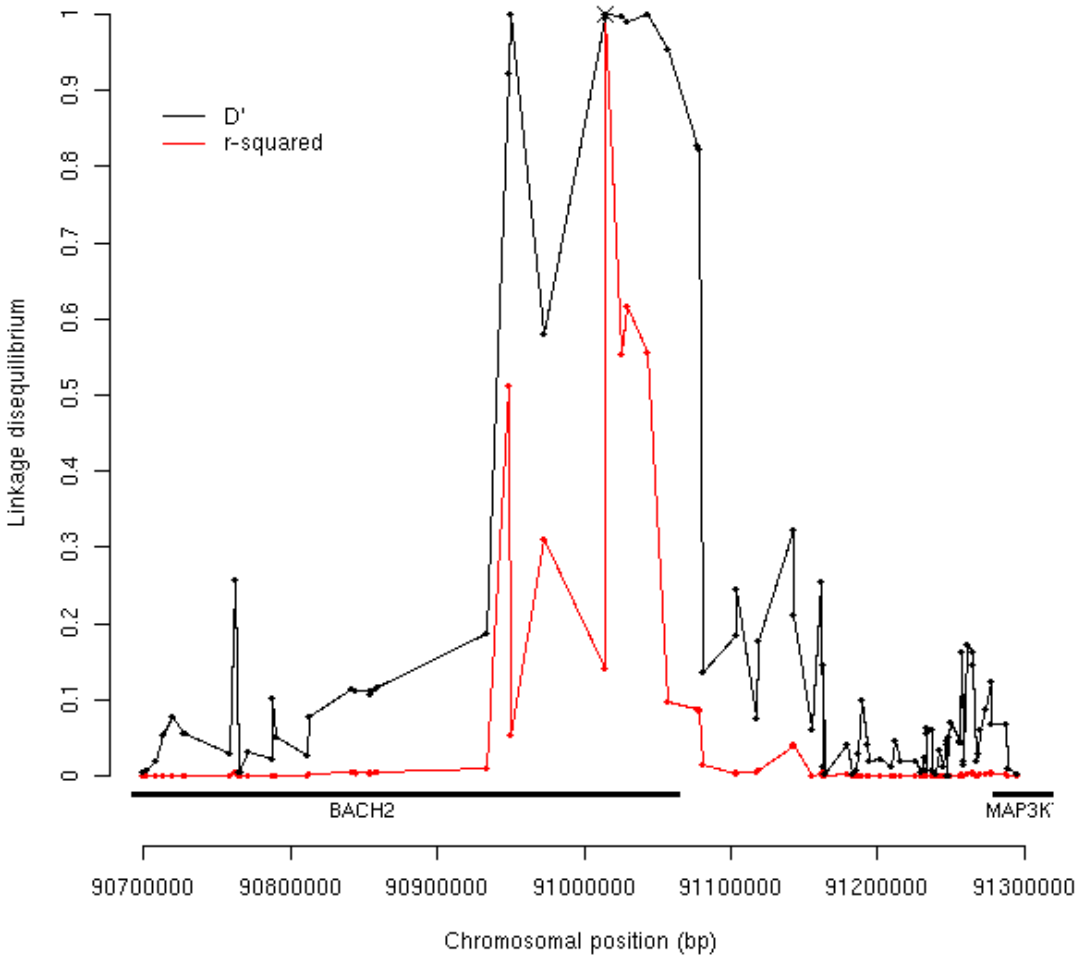
Supplementary Figure 3b. The linkage disequilibrium in 2,942 British controls between rs947474 (X) and neighbouring SNPs (MAF ≥ 0.05) on chromosome 10p15. rs12722489 is the SNP with maximum D' (0.25) in the *IL2RA* region with rs947474. We note that rs12722489 is not associated with T1D (British $P = 0.544$, American $P = 0.507$ and meta-analysis $P = 0.375$; MAF = 0.171). The 234 kb associated region was defined as 6.43 to 6.77 Mb.



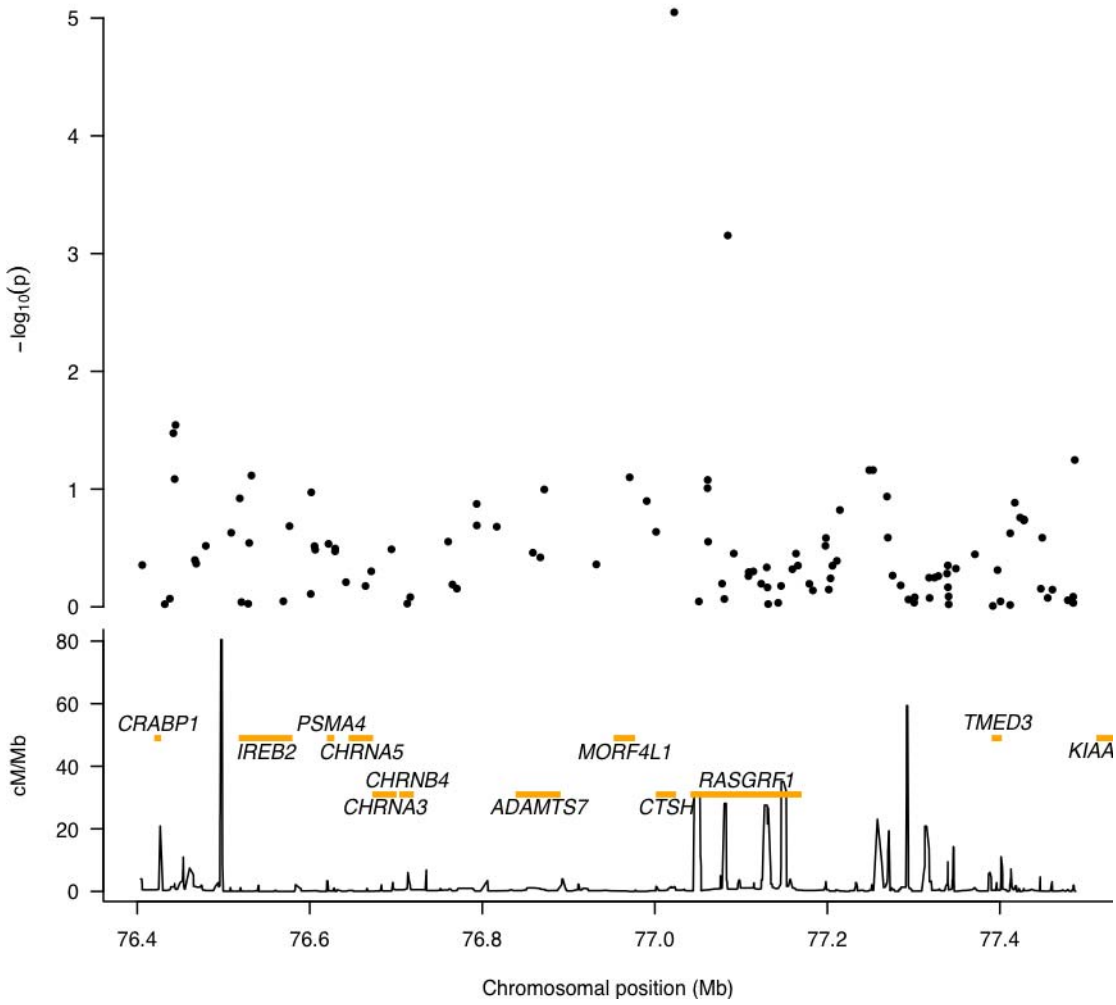
Supplementary Figure 3c. The linkage disequilibrium in 2,942 British controls on chromosome 10p15 between rs12722489 (SNP from the *IL2RA* region with maximum D' with rs947474; **Supplementary Figure 3b**) to rs947474, a 288 kb region. The plot was produced in Haploview (http://www.broad.mit.edu/mpg/haploview/), D' values (%) are shown in the boxes and empty red filled boxes represent D' = 100%.



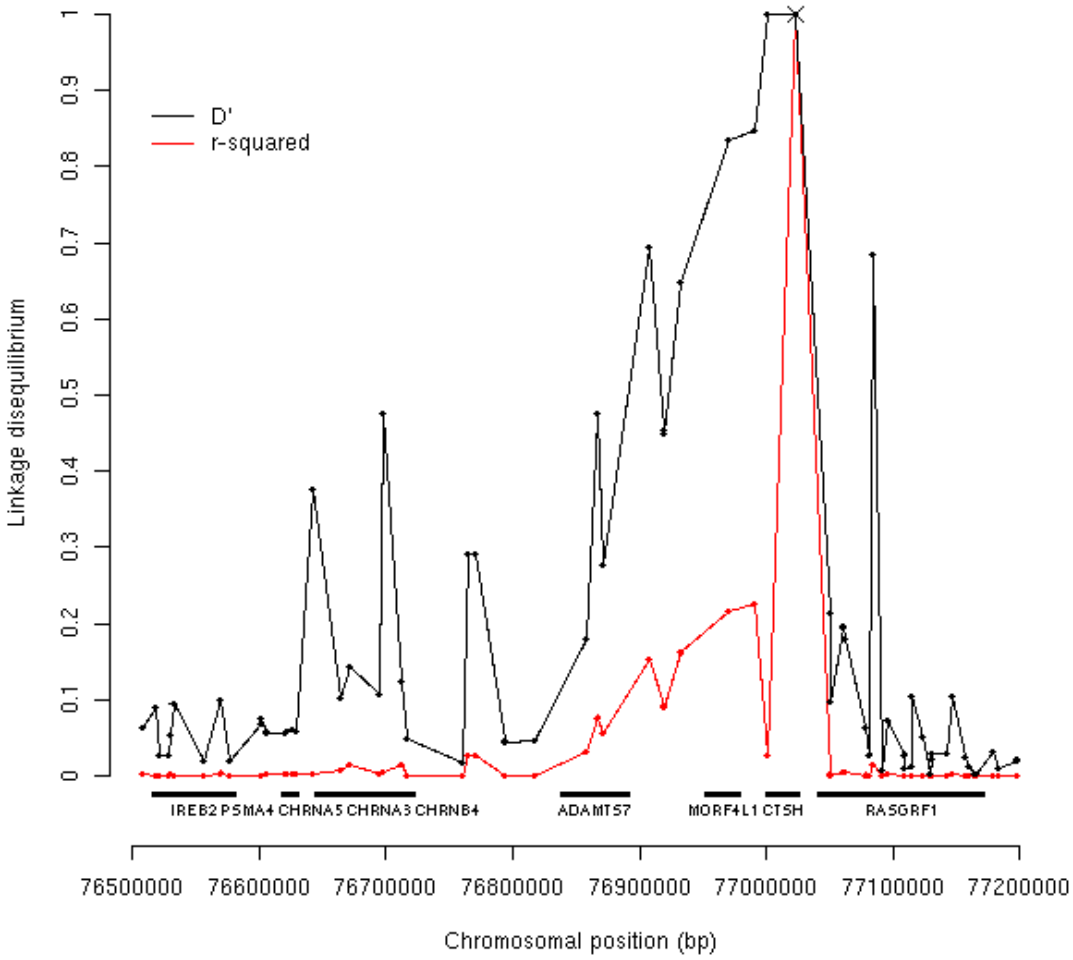
Supplementary Figure 4a. The region about rs11755527 in *BACH2* on chromosome 6q15. The upper panel shows the –log10 *P*-values for the Affymetrix 500K SNPs in the region and the lower panel, shows the recombination map. Genes are shown in yellow. The 365 kb associated region was defined as 90.79 to 91.16 Mb.



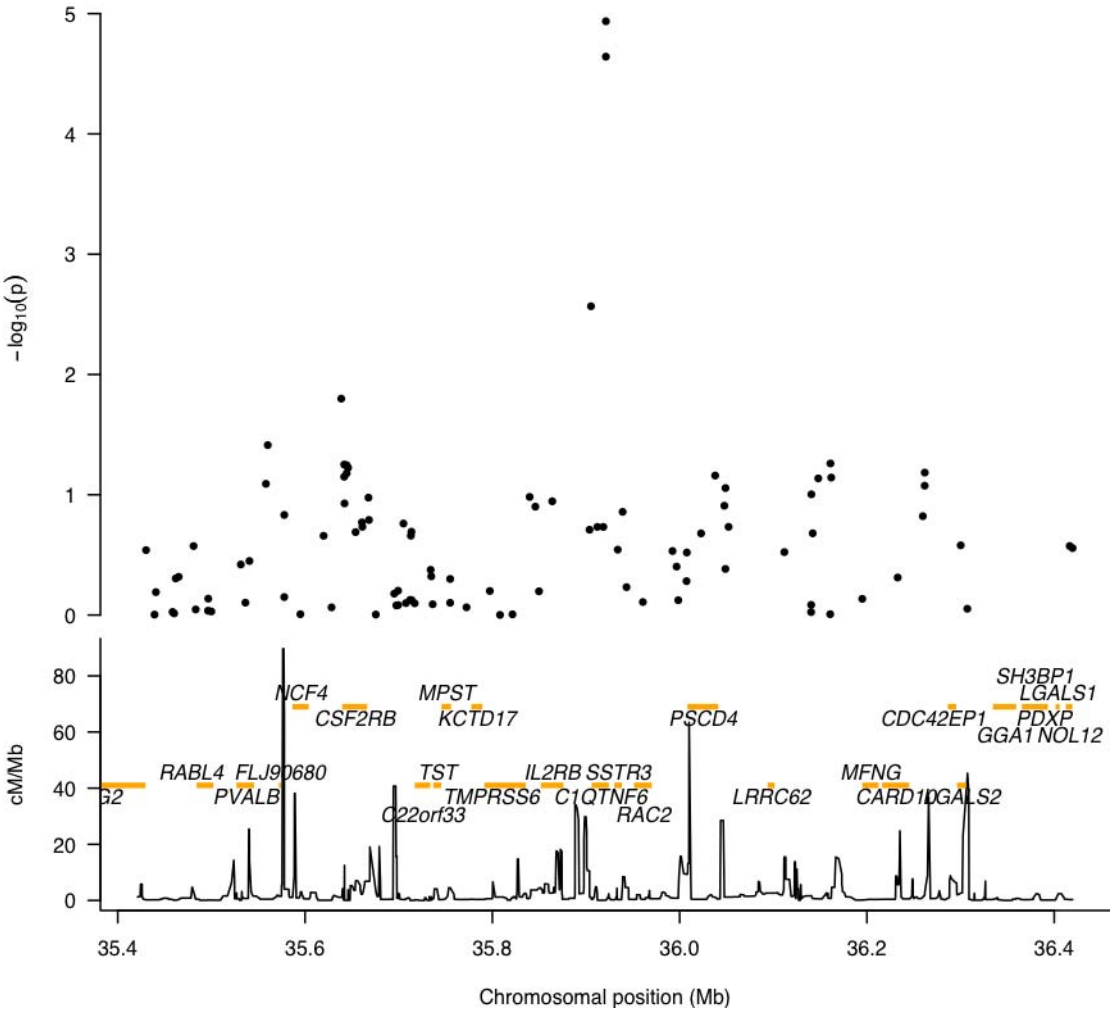
Supplementary Figure 4b. The linkage disequilibrium in 2,942 British controls between rs11755527 (X) in *BACH2* and neighbouring SNPs (MAF ≥ 0.05) on chromosome 6q15. The 365 kb associated region was defined as 90.79 to 91.16 Mb.



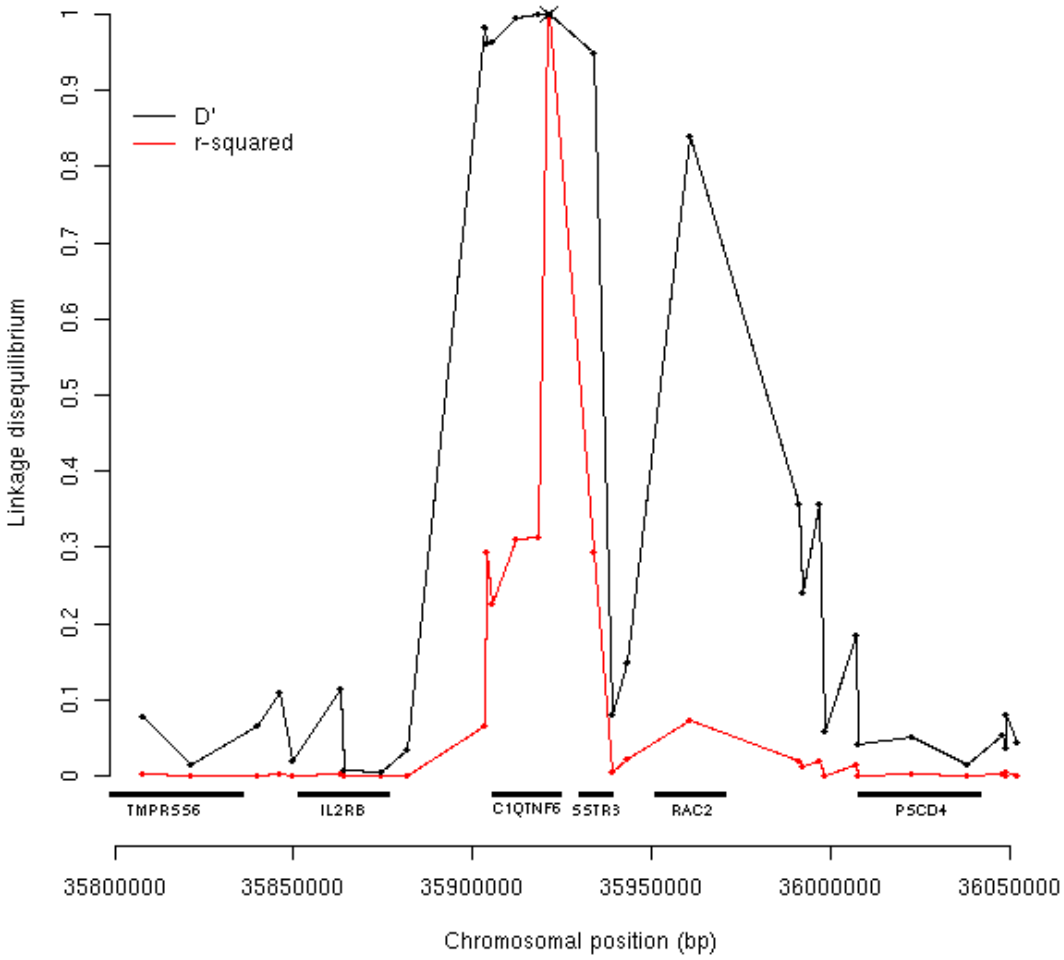
Supplementary Figure 5a. The region about rs3825932 in *CTSH* on chromosome 15q24. The upper panel shows the –log10 *P*-values for the Affymetrix 500K SNPs in the region and the lower panel, shows the recombination map. Genes are shown in yellow. The 660 kb associated region was defined as 76.50 to 77.15 Mb.



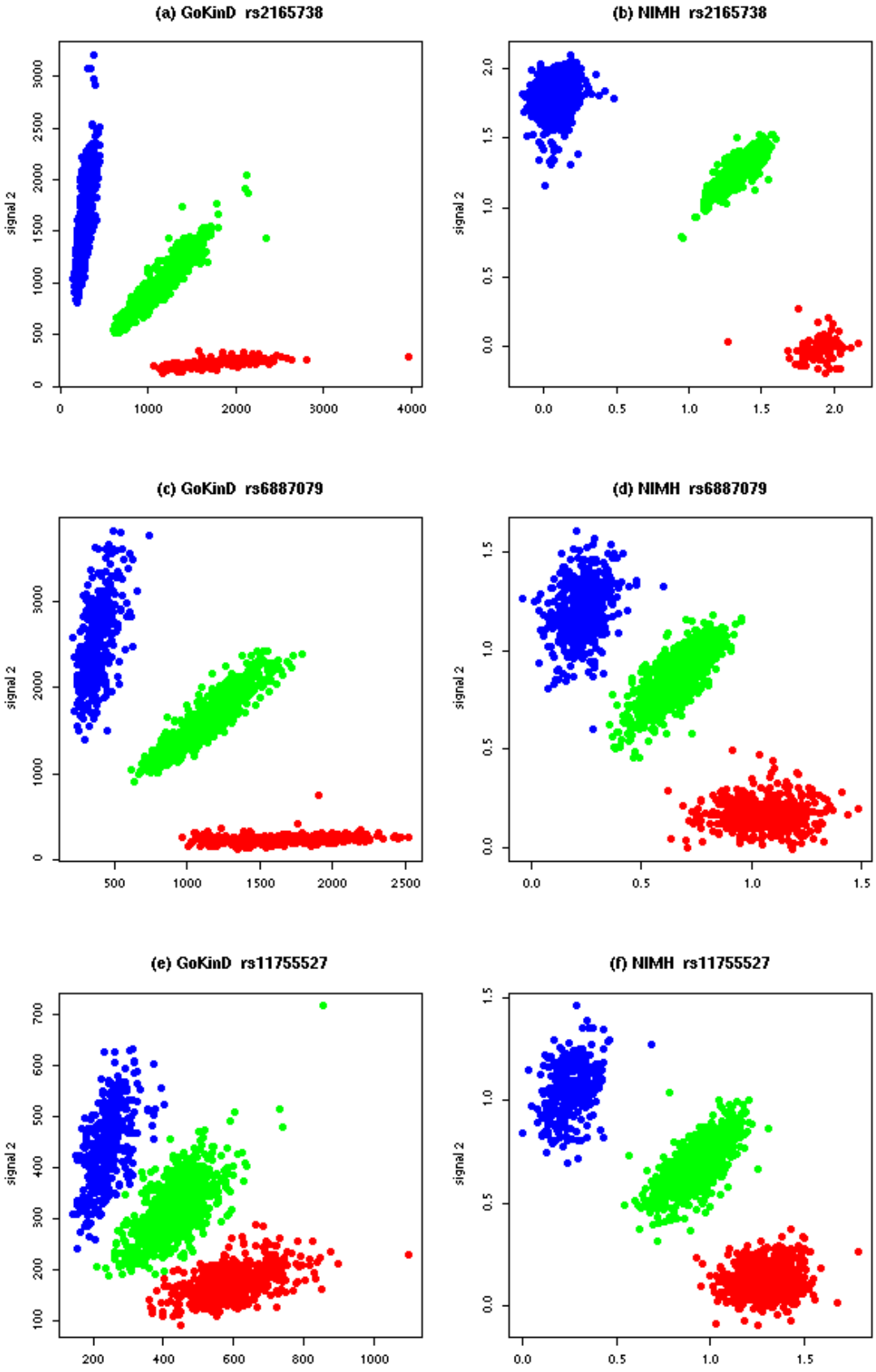
Supplementary Figure 5b. The linkage disequilibrium in 2,942 British controls between rs3825932 (X) in *CTSH* and neighbouring SNPs (MAF ≥ 0.05) on chromosome 15q24. The 660 kb associated region was defined as 76.50 to 77.15 Mb.

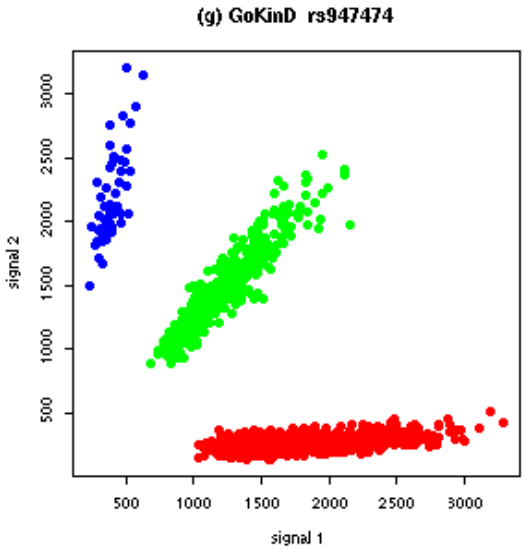


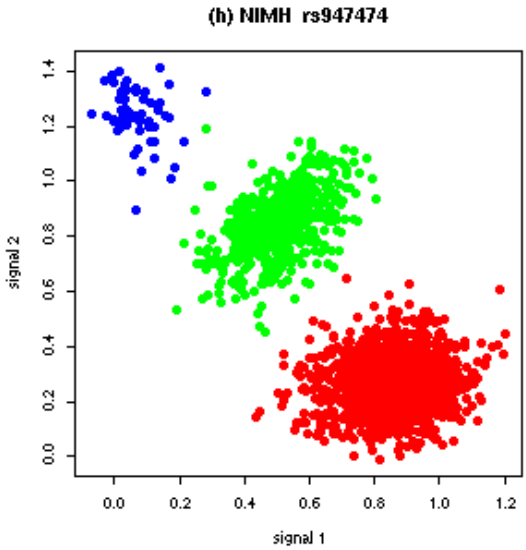
Supplementary Figure 6a. The region about rs229541 in *C1QTNF6* on chromosome 22q13. The upper panel shows the –log10 *P*-values for the Affymetrix 500K SNPs in the region and the lower panel, shows the recombination map. Genes are shown in yellow. The 125 kb associated region was defined as 35.89 to 36.01 Mb.

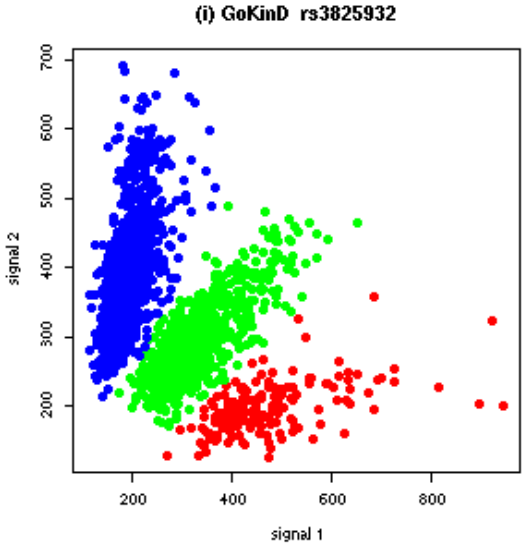


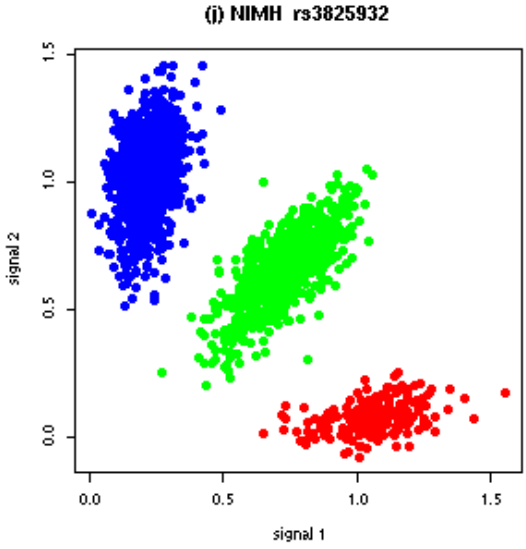
Supplementary Figure 6b. The linkage disequilibrium in 2,942 British controls between rs229541 (X) in *C1QTNF6* and neighbouring SNPs (MAF ≥ 0.05) on chromosome 22q13. The 125 kb associated region was defined as 35.89 to 36.01 Mb.

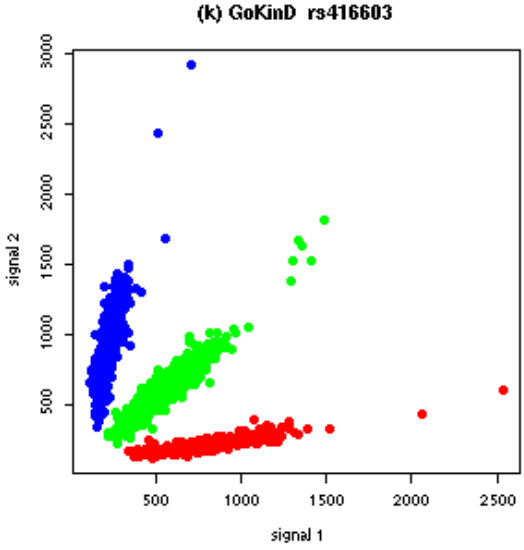

(a) GoKinD rs2165738
signal 2
(b) NIMH rs2165738
signal 2
(c) GoKinD rs6887079
signal 2
(d) NIMH rs6887079
signal 2
(e) GoKinD rs11755527
signal 2
(f) NIMH rs11755527
signal 2

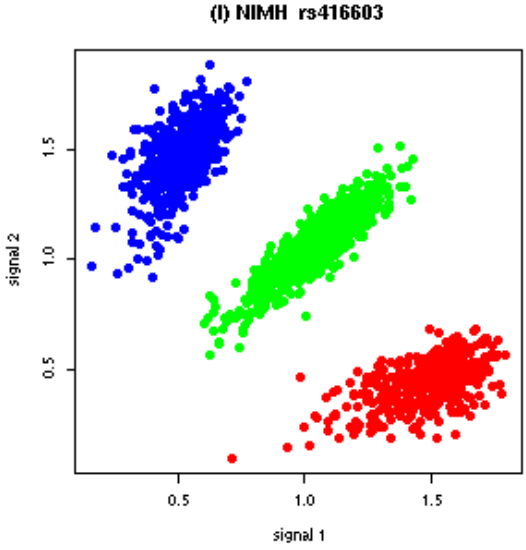

(g) GoKinD rs947474
signal 2
signal 1
500
1000
1500
2000
2500
3000

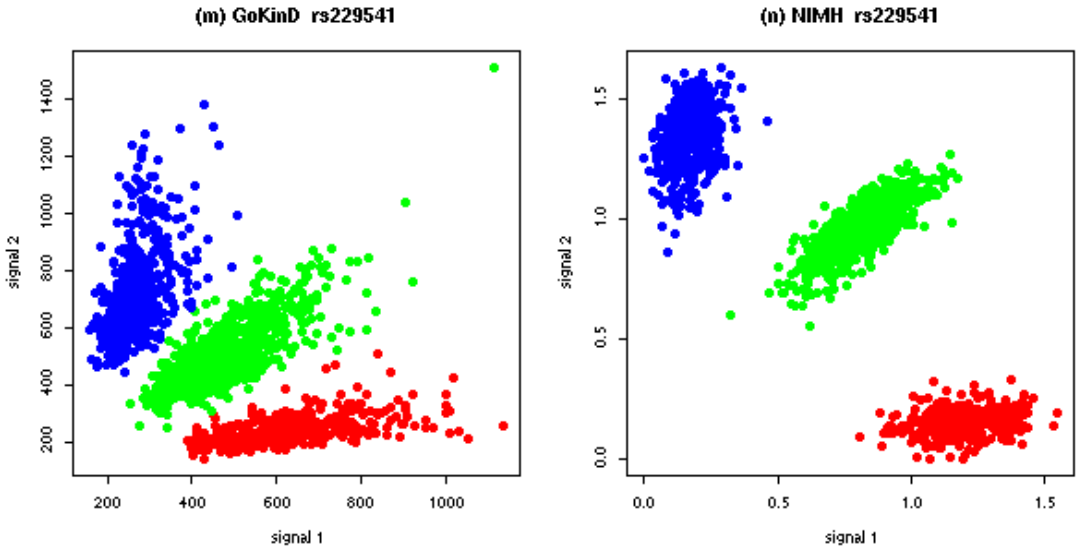

(h) NIMH rs947474
signal 2
signal 1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4


(i) GoKinD rs3825932
signal 2
signal 1
200
300
400
500
600
700
800

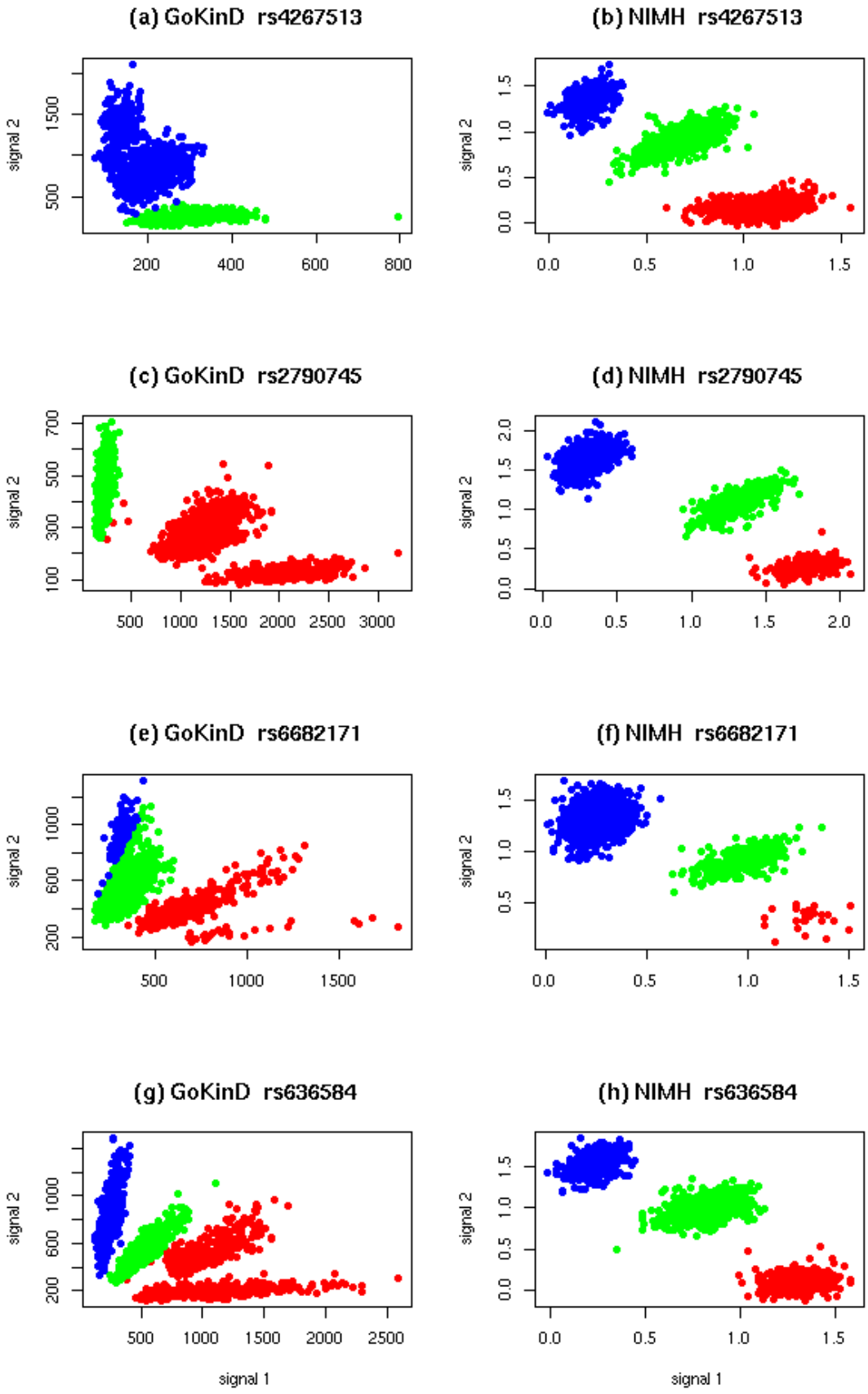

(j) NIMH rs3825932
signal 2
signal 1
0.0
0.5
1.0
1.5


(k) GoKinD rs416603
signal 2
signal 1
500
1000
1500
2000
2500
3000


(l) NIMH rs416603
signal 2
signal 1
0.5
1.0
1.5

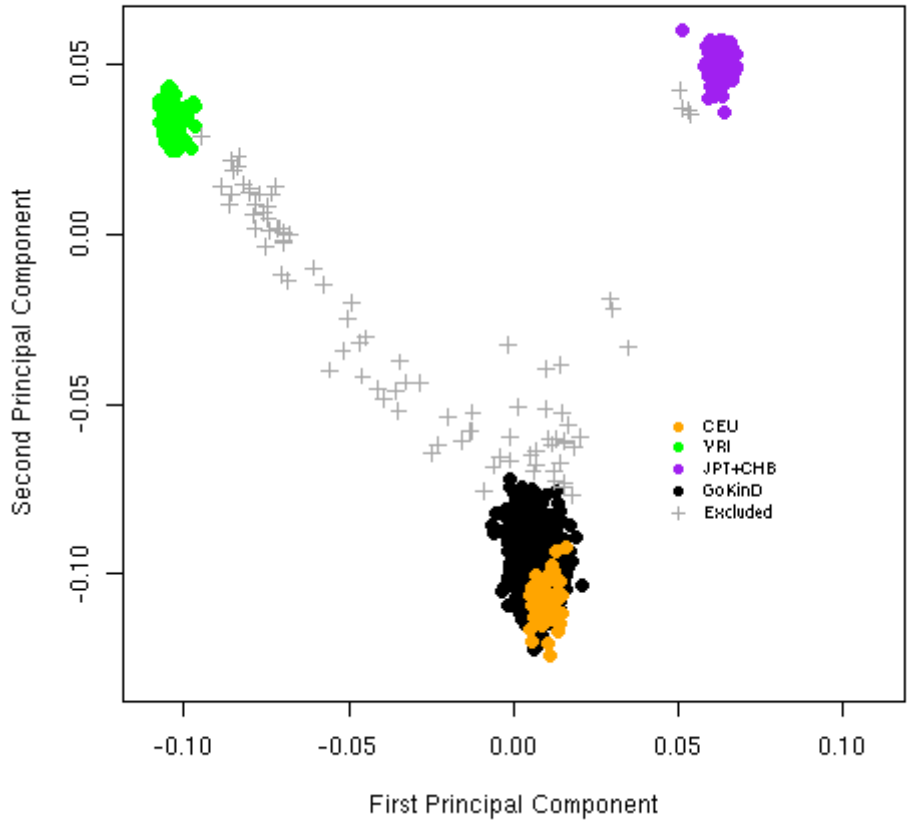

(m) GoKinD rs229541
signal 2
200
400
600
800
1000
1200
1400
200
400
600
800
1000
signal 1
(n) NIMH rs229541
signal 2
0.0
0.5
1.0
1.5
0.0
0.5
1.0
1.5
signal 1

Supplementary Figure 7. Signal intensity plots for the seven SNPs that we attempted to replicate based on the initial study, test results shown in **Table 1**.

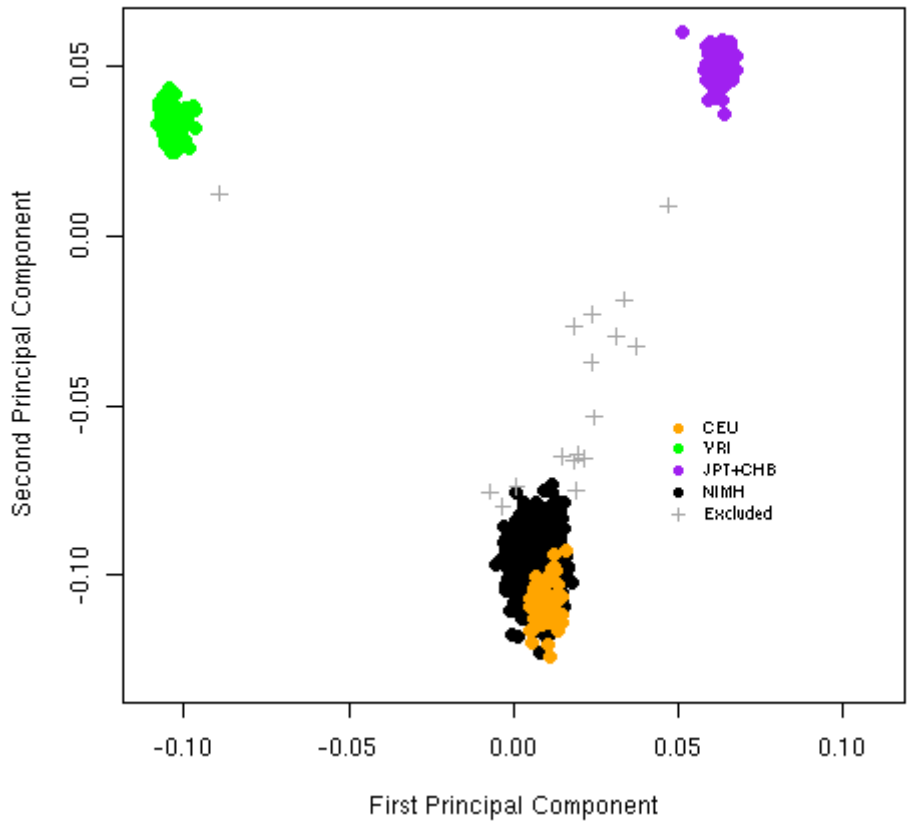

(a) GoKinD rs4267513
signal 2
(b) NIMH rs4267513
signal 2
(c) GoKinD rs2790745
signal 2
(d) NIMH rs2790745
signal 2
(e) GoKinD rs6682171
signal 2
(f) NIMH rs6682171
signal 2
(g) GoKinD rs636584
signal 2
signal 1
(h) NIMH rs636584
signal 2
signal 1

Supplementary Figure 8. Examples of signal intensity plots for poorly clustered SNPs that have passed quality control checks, identified using a Bayes factor for the comparison of the hypothesis that the SNP alleles have been switched with the hypothesis that they have not been switched. The plots show the signal intensity data for each sample coloured by their JAPL genotype call. Homozygotes for the two different alleles are shown in blue and red, and heterozygotes in green.

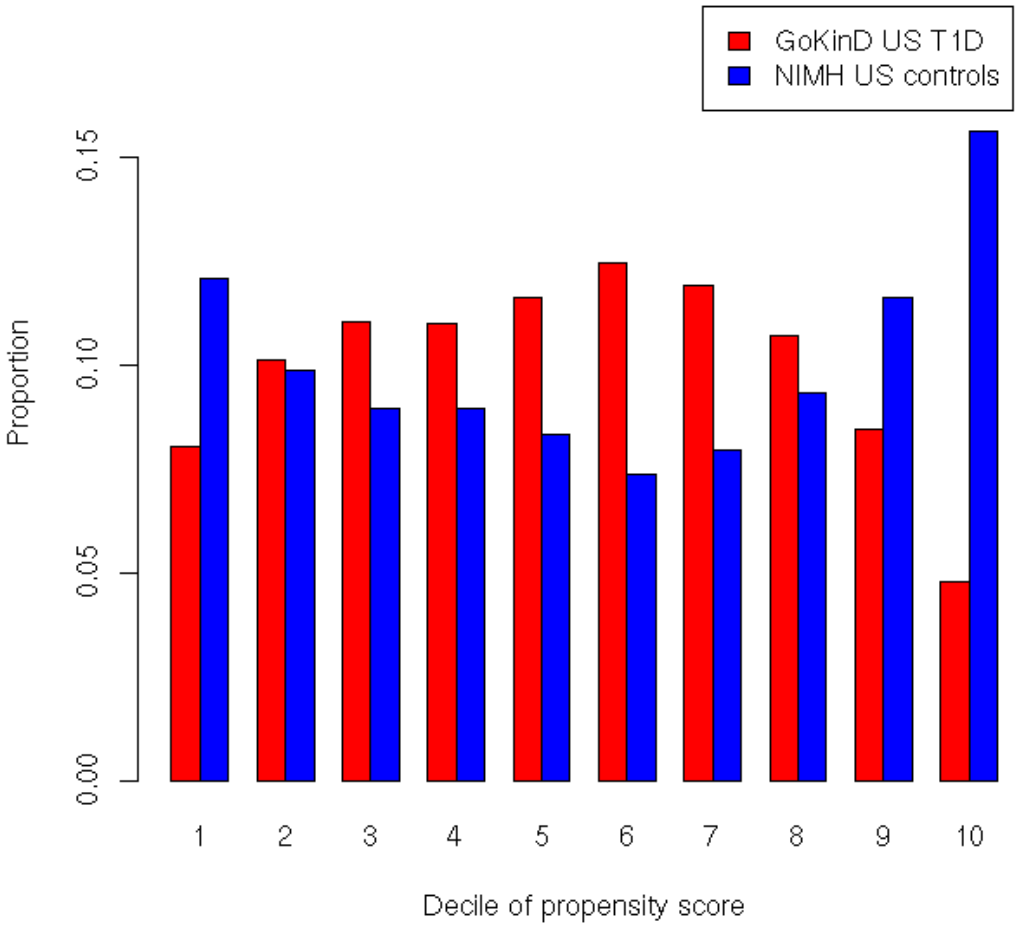
These SNPs have been drawn from the first 100 of 726 SNPs with a Bayes factor > 20, examined going across the genome. The most commonly occurring situation (86/100) is shown in (a) and (b) for rs4267513 on chromosome 2, where there are two genotype clouds in GoKinD and three in NIMH. In (c) and (d), for rs2790745 on chromosome 1, there is another GoKinD two genotype cloud example, however, in this instance (18/86) the clouds have greater separation and modification of the scoring algorithm should result in three GoKinD genotypes. In (e) and (f), for rs6682171 on chromosome 1, there is a GoKinD three genotype cloud example, less common (14/100), which modification of the scoring algorithm might correct. Finally, in (g) and (h), for rs636584 on chromosome 1, there is an additional cloud in GoKinD, but not in NIMH (3/100).



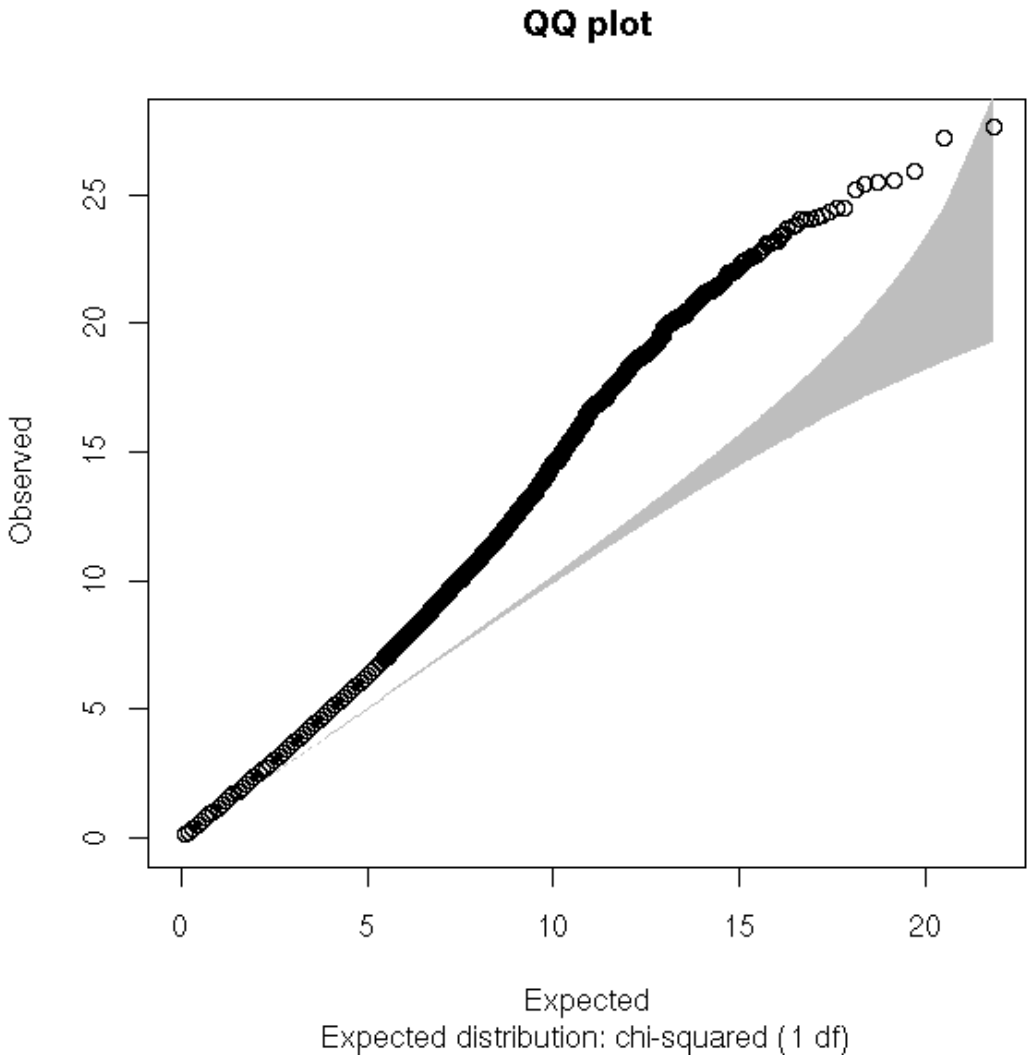
Supplementary Figure 9. The first and second principal component scores for ancestry, based on 947 SNPs, showing HapMap and American GoKinD T1D case samples.



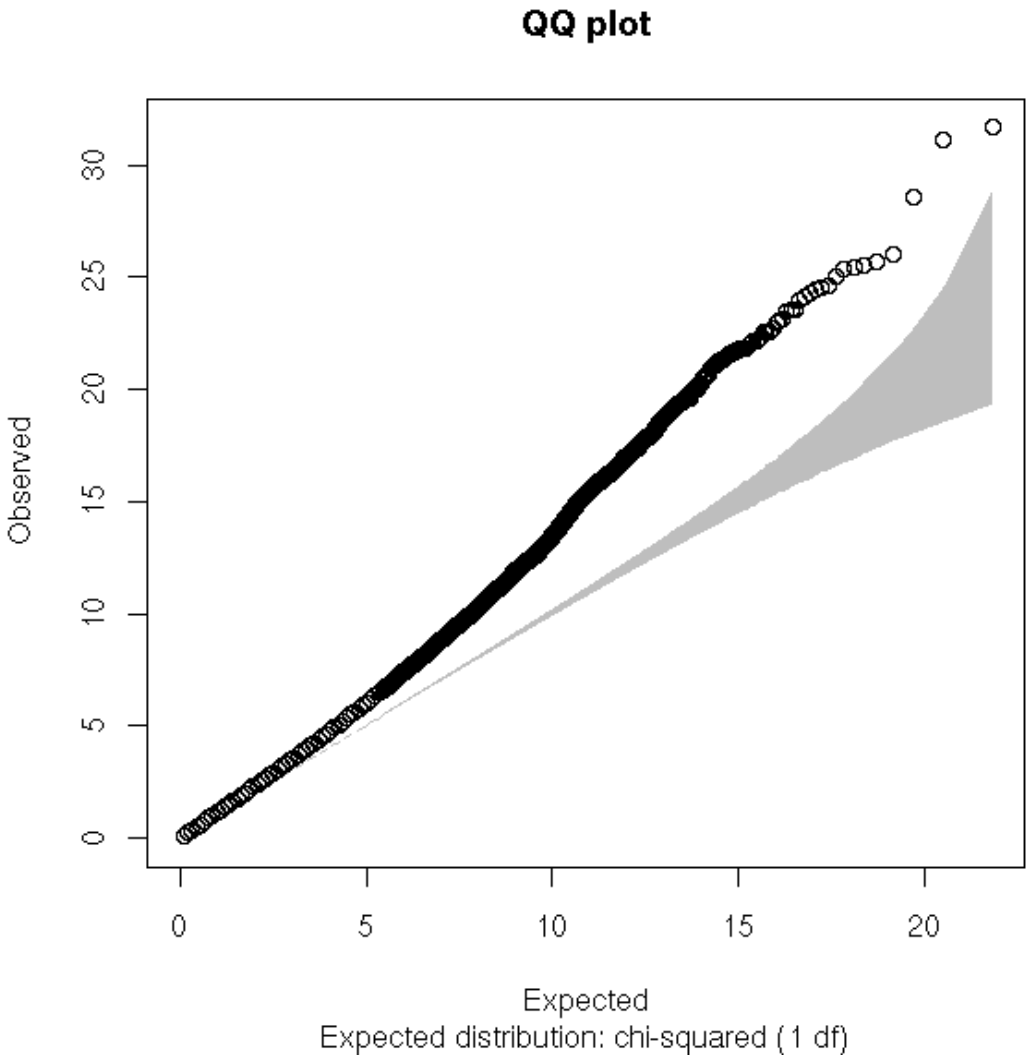
Supplementary Figure 10. The first and second principal component scores for ancestry, based on 951 SNPs, showing HapMap and American NIMH control samples.



Supplementary Figure 11. The deciles of the propensity score use to control for population structure in the American samples.



Supplementary Figure 12. A quantile-quantile plot for the American case-control 1-df tests. We excluded known T1D loci and SNPs with an $r^2 \geq 0.1$ with them, 333,997 SNPs remained. Overdispersion factor, $\lambda = 1.18$.



Supplementary Figure 13. A quantile-quantile plot for the American case-control stratified 1-df tests. We excluded known T1D loci and SNPs with an $r^2 \geq 0.1$ with them, 333,997 SNPs remained. Overdispersion factor, $\lambda = 1.14$.

Supplementary Tables

Chromosome	Gene region	†SNP	British (WTCCC) 1,960 cases and 2,942 controls	American (GoKinD/NIMH) 1,601 cases and 1,704 controls	Meta-analysis of British and American case-control data
1p13	*PTPN22*	rs6679677 C>A	4.76×10^{-24}	3.16×10^{-18}	1.27×10^{-40}
2q24	*IFIH1*	rs3747517 G>A	4.21×10^{-3}	0.0200	2.26×10^{-4}
2q33	*CTLA4*	rs3087243 G>A	8.65×10^{-5}	5.05×10^{-8}	7.67×10^{-11}
6p21	*HLA*	rs9272346 A>G	1.04×10^{-126}	1.21×10^{-94}	5.52×10^{-219}
10p15	*IL2RA*	rs12251307 C>T	3.96×10^{-5}	0.0146	2.33×10^{-6}
		‡(rs11256448 A>G	1.57×10^{-3}	9.22×10^{-4}	5.24×10^{-6})
11p15	*IGF2* (11 kb from *INS*)	rs3741208 C>T	2.28×10^{-7}	N/A	
*12q24	*C12orf30*	rs17696736 A>G	1.02×10^{-13}	5.11×10^{-6}	6.35×10^{-18}
*12q13	*ERBB3*	rs2292239 C>A	2.26×10^{-9}	2.20×10^{-8}	2.91×10^{-16}
*16p13	*CLEC16A*	rs12708716 A>G	9.69×10^{-8}	1.54×10^{-6}	7.28×10^{-13}
*18p11	*PTPN2*	rs2542151 A>C	7.39×10^{-8}	0.0596	8.73×10^{-8}
		‡(rs8087237 C>A	1.34×10^{-3}	3.87×10^{-4}	2.14×10^{-6})

Supplementary Table 1a. The stratified trend test (1-df) *P*-values for SNPs from the ten T1D associated regions, previously reported with convincing evidence for association. *Loci detected by the WTCCC GWA study[13] and follow-up studies[23]. †SNPs previously reported by Todd *et al.*[23], except for rs3747517/*IFIH1* and the HLA class II SNP, rs9272346, which was one of many SNPs available on the Affymetrix 500K SNP chipset in the MHC region (positions 25-35 Mb on chromosome 6) with such high levels of significance. The *IFIH1* SNP rs3747517 is the most associated *IFIH1* SNP on the Affymetrix 500K SNP chipset, it is a nonsynonymous SNP, but not the one we reported [24], rs1990760, when

we first found the association. ‡The most associated SNP in the region in the American study. Although the British (WTCCC) case-control *P*-value for *IGF2* rs3741208 is included in the table, this SNP was excluded as its call-rate < 0.99 (**Supplementary Methods**). N/A = not available in the GoKinD data (**Supplementary Note**).

Chromosome	Gene region	SNP[†]	British (WTCCC) 1,966 cases and 2,955 controls		American (GoKinD/NIMH) 1,601 cases and 1,704 controls	
			MAF in controls	OR (95% c.i.)	MAF in controls	OR (95% c.i.)
1p13	*PTPN22*	rs6679677 C>A	0.096	1.88 (1.66-2.12)	0.095	1.93 (1.66-2.24)
2q24	*IFIH1*	rs3747517 G>A	0.269	0.87 (0.79-0.96)	0.273	0.87 (0.78-0.98)
2q33	*CTLA4*	rs3087243 G>A	0.446	0.84 (0.77-0.92)	0.458	0.76 (0.68-0.84)
6p21	*HLA*	rs9272346 A>G	0.388	0.28 (0.25-0.31)	0.428	0.29 (0.25-0.33)
10p15	*IL2RA*	rs12251307 C>T	0.122	0.75 (0.66-0.86)	0.121	0.82 (0.70-0.96)
		[‡](rs11256448 A>G	0.245	1.16 (1.06-1.28)	0.244	1.21 (1.08-1.35))
11p15	*IGF2* (11 kb from *INS*)	rs3741208 C>T	0.378	1.25 (1.15-1.35)	N/A	
12q24	*C12orf30*	rs17696736 A>G	0.424	1.38 (1.27-1.50)	0.440	1.26 (1.14-1.39)
12q13	*ERBB3*	rs2292239 C>A	0.340	1.30 (1.20-1.42)	0.334	1.34 (1.21-1.48)
16p13	*CLEC16A*	rs12708716 A>G	0.349	0.79 (0.72-0.86)	0.364	0.77 (0.70-0.86)
18p11	*PTPN2*	rs2542151 A>C	0.163	1.35 (1.21-1.50)	0.158	1.14 (0.99-1.29)
		[‡](rs8087237 C>A	0.448	0.87 (0.80-0.95)	0.452	0.84 (0.76-0.92))

Supplementary Table 1b. The minor allele frequency (MAF) and odds ratio (OR) for the minor alleles of SNPs from the ten T1D associated regions, previously reported with convincing evidence for association. [†]SNPs previously reported by Todd *et al.*[23], except for rs3747517/*IFIH1* and rs9272346/HLA class II SNP. [‡]The most associated SNP in the region in the American study. c.i. = confidence interval and N/A = not available in the GoKinD data (**Supplementary Note**).

	Cases (%)	Controls (%)	OR (95% c.i.)	Likelihood ratio test *P* (df)
British (WTCCC): 1,959 case and 2,937 control genotypes[†]				
Alleles				
G	2,713 (69.2)	4,343 (73.9)	1.00 (reference)	-
A	1,205 (30.8)	1,531 (26.1)	1.26 (1.15-1.38)	6.72×10^{-7} (1)
Genotypes				
G/G	943 (48.1)	1,611 (54.9)	1.00 (reference)	-
A/G	827 (42.2)	1,121 (38.2)	1.27 (1.13-1.44)	4.19×10^{-6} (2)
A/A	189 (9.7)	205 (7.0)	1.56 (1.26-1.94)	
American (GoKinD/NIMH): 1,601 case and 1,704 control genotypes				
Alleles				
G	2,238 (69.9)	2,477 (72.7)	1.00 (reference)	-
A	964 (30.1)	931 (27.3)	1.17 (1.05-1.31)	4.80×10^{-3} (1)
Genotypes				
G/G	766 (47.9)	915 (53.7)	1.00 (reference)	-
A/G	706 (44.1)	647 (38.0)	1.35 (1.16-1.56)	3.49×10^{-4} (2)
A/A	129 (8.1)	142 (8.3)	1.11 (0.85-1.45)	
Meta-analysis of stratified trend test (1-df) *P*-values = 1.87×10^{-8}				

Supplementary Table 2. Association analysis of rs17388568 G>A from 4q27/*IL2-IL21*. [†]Previously reported by Todd *et al.*[23]. OR = odds ratio for the minor allele (A), c.i. = confidence interval, and df = degrees-of-freedom.

SNP	Chromosome	Position	Gene region	χ_1^2	1-df *P*
rs10915671	1	4883077		16.03	6.25x10-5
rs12049447	**1**	**113724368**		**21.21**	**4.11x10-6**
rs10458459	**1**	**113768928**	***MAGI3***	**21.65**	**3.28x10-6**
rs12030900	**1**	**113775786**	***MAGI3***	**20.99**	**4.62x10-6**
rs1573996	**1**	**113786979**	***MAGI3***	**22.21**	**2.44x10-6**
rs12022849	**1**	**113865046**	***MAGI3***	**21.6**	**3.36x10-6**
rs1217225	1	113941514	*MAGI3*	18.07	2.13x10-5
rs1217401	1	114240474	*AP4B1*	17.64	2.67x10-5
rs10801121	1	190745497		16.09	6.05x10-5
rs10754012	1	190759518		17.17	3.42x10-5
rs7415895	1	226807770	*LOC644953*	15.49	8.30x10-5
rs736725	2	10607340		15.79	7.07x10-5
rs2165738	**2**	**24546313**		**19.46**	**1.03x10-5**
rs6734118	2	37412859		17.21	3.35x10-5
rs7575024	2	43520495	*THADA*	15.34	8.96x10-5
rs17031005	2	43614803	*THADA*	16.28	5.46x10-5
rs9653442	2	100191799	*LOC150577*	15.99	6.38x10-5
rs6706577	2	111513304	*ACOXL*	17.42	3.00x10-5
rs10196515	2	153948555		17.47	2.93x10-5
rs9967792	2	191682680	*STAT4*	18.35	1.84x10-5
rs4853543	2	191684879	*STAT4*	17.98	2.24x10-5
rs10931483	2	191697495	*STAT4*	17.06	3.62x10-5
rs1551443	2	191704763	*STAT4*	17.16	3.44x10-5
rs1584945	2	191706797	*STAT4*	17.89	2.34x10-5
rs2356350	2	191710783	*STAT4*	15.36	8.87x10-5
rs4147713	2	206724017	*NDUFS1*	15.29	9.24x10-5
rs155390	3	1396673	*CNTN6*	18.43	1.76x10-5
rs7619252	3	11712145	*VGLL4*	15.52	8.16x10-5
rs4857979	3	21154336		15.2	9.67x10-5
rs6441961	3	46327388		18.54	1.66x10-5
rs2157057	3	46341273		15.8	7.05x10-5
rs17749495	3	95665802		16.34	5.30x10-5
rs6786991	3	116409646		16.61	4.59x10-5
rs2085700	3	116417029		16.11	5.97x10-5
rs1449440	3	147302413	*PLOD2*	16.68	4.42x10-5

rs28364983.1	4	75533588	*AREG*	18.39	1.80x10-5
rs11725706	4	75546972	*AREG*	16.68	4.42x10-5
rs28364983.2	4	75703507	*AREG, LOC727738*	18.39	1.80x10-5
rs1465041	4	123147736		15.58	7.92x10-5
rs7722135	5	86330425		16.61	4.59x10-5
rs2640457	5	86355692		16.74	4.30x10-5
rs2544677	5	86435018		18.39	1.80x10-5
rs2112168	5	86440646		15.67	7.55x10-5
rs6873372	5	128260800		16.05	6.18x10-5
rs6887079	**5**	**166799014**		**21.29**	**3.95x10-6**
‡rs2237093	**6**	**36143332**	***MAPK13, MAPK14***	**19.5**	**1.01x10-5**
rs2501716	**6**	**90949120**	***BACH2***	**20.07**	**7.45x10-6**
rs11755527	**6**	**91014952**	***BACH2***	**20.47**	**6.07x10-6**
rs619192	**6**	**91025670**	***BACH2***	**21.27**	**3.98x10-6**
rs1847472	6	91029880	*BACH2*	16.03	6.24x10-5
rs604912	**6**	**91043041**	***BACH2***	**19.69**	**9.13x10-6**
rs9390779	6	102315774	*GRIK2*	19.06	1.27x10-5
rs2327832	6	138014761		17.88	2.36x10-5
rs6920220	6	138048197		15.21	9.60x10-5
rs7755420	6	158677335	*TULP4*	16.46	4.96x10-5
rs6926291	6	158678489	*TULP4*	16.01	6.31x10-5
rs1754405	6	158688454	*TULP4*	16.01	6.30x10-5
rs1041566	6	158836960	*TULP4*	16.56	4.71x10-5
rs17790410	7	2186618	*MAD1L1*	16.02	6.28x10-5
rs2041342	7	2696291	*AMZ1*	19.01	1.30x10-5
rs588958	7	152632167		17.85	2.40x10-5
rs1296023	8	11764226	*CTSB*	18.93	1.36x10-5
rs2250903	8	11772593		15.38	8.79x10-5
rs6985839	8	72911162		16.52	4.81x10-5
rs6991144	8	103127777		16.17	5.78x10-5
rs7815993	8	130020951		17.91	2.31x10-5
rs7020673	9	4281747		16.5	4.88x10-5
rs4339696	9	4285880	*LOC729691*	16.53	4.79x10-5
rs496892	9	22014351	*MTAP*	16.07	6.11x10-5
rs996375	9	71752569		16.36	5.25x10-5
rs7025044	9	71758007		16.89	3.96x10-5
rs1873825	9	71780875		16.15	5.86x10-5
rs4430153	9	71839947		15.48	8.32x10-5

rs2146079	9	113446470	*bA16L21.2.1, GNG10, LOC552891*	15.82	6.95x10-5
rs947474	**10**	**6430456**		**21.61**	**3.34x10-6**
rs2666236	10	33458878		18.08	2.12x10-5
rs286493	10	90011231		15.93	6.56x10-5
rs7928968	11	2006875		19.19	1.18x10-5
rs536962	11	78746092		16.88	3.98x10-5
rs556857	11	78746196		17.31	3.17x10-5
rs10830989	11	92526296	*SLC36A4*	15.39	8.75x10-5
rs10893666	11	126701966		15.64	7.68x10-5
rs609017	11	128088154	*FLI1*	15.36	8.89x10-5
rs587862	11	128100383	*FLI1*	17.06	3.63x10-5
rs1939236	11	128108037	*FLI1*	16.35	5.27x10-5
rs4937700	11	131792538	*OPCML*	15.36	8.88x10-5
rs12417610	11	131796246	*OPCML*	17.53	2.82x10-5
rs4936168	11	131819986	*OPCML*	17.44	2.96x10-5
rs12819116	12	12537788	*DUSP16*	15.58	7.90x10-5
rs2417191	12	12542523	*DUSP16*	15.47	8.37x10-5
rs886125	**12**	**109849707**		**32.06**	**1.50x10-8**
rs4766553	12	110118664	*CUTL2*	16.17	5.78x10-5
rs7962233	12	110139896	*CUTL2*	16.64	4.51x10-5
rs2078851	**12**	**110174962**	***CUTL2***	**20.83**	**5.01x10-6**
rs7300860	**12**	**110238980**	***CUTL2***	**19.59**	**9.61x10-6**
rs1980364	**12**	**111002025**	***C12orf30***	**23.63**	**1.16x10-6**
rs3519	**12**	**111082704**	***FLJ30092***	**19.64**	**9.36x10-6**
rs2285809	**12**	**111192809**	***C12orf51***	**24.67**	**6.80x10-7**
rs1468251	**12**	**111261816**		**21.34**	**3.84x10-6**
rs10774654	**12**	**111447943**		**24.19**	**8.74x10-7**
rs4767879	**12**	**111448453**		**21.82**	**2.99x10-6**
rs4767880	12	111448551		17.86	2.37x10-5
rs2384000	**12**	**111449525**		**23.63**	**1.17x10-6**
rs10850053	**12**	**111462160**		**26**	**3.42x10-7**
rs10850061	12	111500936		15.42	8.59x10-5
rs238271	13	41831354		15.71	7.37x10-5
rs6561906	13	56463195		15.16	9.87x10-5
rs9318020	13	70881756		15.76	7.20x10-5
rs11625107	14	33797878		16.39	5.17x10-5
rs1462275	14	97393096		16.17	5.80x10-5
rs11623343	14	100659083		16.89	3.96x10-5

rs529413	15	51330037		15.23	9.50x10-5
rs3825932	**15**	**77022501**	***CTSH***	**19.73**	**8.93x10-6**
rs2346733	15	84946846		16.42	5.09x10-5
rs2346734	15	84947085		15.94	6.54x10-5
rs2346735	15	84947122		18.39	1.80x10-5
rs1646067	16	11226546	*LOC729954*	15.26	9.35x10-5
rs149310	16	11253200	*C16orf75, SOCS1*	18.17	2.02x10-5
rs416603	**16**	**11271580**	***C16orf75, PRM3, TNP2***	**20.62**	**5.61x10-6**
rs7187741	16	11340604	*C16orf75*	18.53	1.68x10-5
rs151320	16	28476179	*CCDC101*	15.66	7.59x10-5
rs7193402	16	28493628	*CCDC101*	15.33	9.04x10-5
rs17675333	16	64308253		16.93	3.87x10-5
‡**rs6564245**	**16**	**73866400**		**20.74**	**5.26x10-6**
rs4536500	16	73900698	*CFDP1*	18.35	1.84x10-5
rs9934007	16	73948033	*CFDP1*	17.23	3.31x10-5
rs4993971	16	73969135	*CFDP1*	18.4	1.79x10-5
‡**rs247438**	**16**	**73990150**	***CFDP1***	**20.32**	**6.54x10-6**
rs8046416	16	74045259	*TMEM170*	16.45	4.98x10-5
rs11150012	16	76271706		18.03	2.18x10-5
rs4843366	16	84967302		17.29	3.21x10-5
rs733622	17	4127871	*UBE2G1*	15.3	9.19x10-5
rs2325988	17	4146033	*UBE2G1*	16.39	5.16x10-5
rs7216130	17	4153992	*UBE2G1*	15.37	8.86x10-5
rs2053255	17	4167985	*UBE2G1*	15.47	8.39x10-5
rs8074770	17	4188927	*UBE2G1*	16.17	5.80x10-5
rs16956936	17	7574417	*DNAH2*	16.42	5.08x10-5
rs7221109	17	36023812		17.8	2.46x10-5
rs17660595	17	41536766	*KIAA1267*	15.82	6.95x10-5
rs17661428	17	41563921	*KIAA1267*	15.96	6.46x10-5
rs12150320	17	41568981	*KIAA1267*	16.93	3.88x10-5
rs12601816	17	47927239		15.86	6.80x10-5
rs16951546	17	47940821		15.33	9.04x10-5
rs9948279	18	34098218		17.37	3.08x10-5
rs2053074	19	9063085	*OR1M1*	15.85	6.87x10-5
rs12462707	19	35568991	*ZNF536*	17.06	3.62x10-5
‡**rs2823025**	**21**	**15346292**		**20.37**	**6.38x10-6**
rs41176	22	28762403		17.95	2.26x10-5
rs1989870	22	28802254		17.41	3.01x10-5

rs5763779	22	28834652	*HORMAD2*	18.33	1.86x10-5
rs4820830	22	28861091	*HORMAD2*	17.89	2.34x10-5
rs2412976	22	28897907	*HORMAD2*	18.11	2.09x10-5
rs229540	22	35921236	*C1QTNF6*	17.94	2.28x10-5
rs229541	**22**	**35921264**	***C1QTNF6***	**19.23**	**1.16x10-5**
rs5979785	23	12881445		16.72	4.33x10-5
rs6609239	23	43262514		15.14	1.00x10-4

Supplementary Table 3. SNPs ordered by genome position with a *P*-value $\leq 1 \times 10^{-4}$, after excluding T1D loci (**Supplementary Table 1a**) and SNPs with an $r^2 \geq 0.1$ with T1D loci. The top 30 ranked SNPs are shown in bold. We added postfixes ".1" and ".2" to rs28364983 to indicate that it mapped twice to chromosome 4. We note that the associations of rs12049447 – rs1217225 on chromosome 1p13 are explained by PTPN22 Arg620Trp (rs2476601)[25] and that the chromosome 12q24 SNPs are explained by rs3184504 in *SH2B3*. ‡Denotes top 30 ranked SNPs with poor clustering in one or more GWA studies.

SNP	Gene region	Single-locus tests in 6,056 cases and 6,651 controls		Adding SNP to rs41295061 *P* (1-df)	Adding rs41295061 to SNP *P* (1-df)	Adding SNP to rs41295061 and rs11594656 *P* (1-df)
		P (1-df)	OR (95% c.i.)			
rs41295061 C>A	*IL2RA*, *RBM17*	3.90×10^{-24}	0.62 (0.56-0.68)	--	--	--
rs11594656 T>A	*IL2RA*, *RBM17*	8.60×10^{-6}	0.87 (0.82-0.93)	8.62×10^{-10}	4.90×10^{-28}	--
rs947474 A>G	*PRKCQ*	2.10×10^{-6}	0.85 (0.80-0.91)	9.93×10^{-7}	1.99×10^{-24}	7.20×10^{-7}

Supplementary Table 4. A summary of the analysis after rs947474 was genotyped in additional samples, to test whether the *PRKCQ* SNP, rs947474, was associated with T1D independently of the T1D associated *IL2RA* region[22] (**Supplementary Methods**). The decision to genotype rs947474 was based on a similar analysis of the British (WTCCC) data.

	British (WTCCC)		American (GoKinD/NIMH)		Follow-up			
					British cases and controls		Families	
	Cases (%)	Controls (%)	Cases (%)	Controls (%)	Cases (%)	Controls (%)	Cases (%)	†Pseudo-controls (%)
rs2165738 intergenic/2p23								
G/G	971 (49.6)	1,575 (53.6)	784 (49.0)	933 (54.8)	3,242 (52.1)	3,710 (53.4)	N/A	
G/C	828 (42.3)	1,132 (38.6)	679 (42.4)	661 (38.8)	2,438 (39.2)	2,725 (39.2)	N/A	
C/C	157 (8.0)	229 (7.8)	138 (8.6)	110 (6.5)	545 (8.8)	511 (7.4)	N/A	
rs6887079 gene desert/5q34								
T/T	442 (22.6)	800 (27.2)	376 (23.5)	477 (28.0)	1,586 (25.6)	1,791 (26.0)	N/A	
T/C	986 (50.3)	1,400 (47.6)	795 (49.7)	837 (49.1)	3,066 (49.5)	3,435 (49.8)	N/A	
C/C	532 (27.1)	739 (25.1)	430 (26.9)	390 (22.9)	1,538 (24.9)	1,677 (24.3)	N/A	
rs11755527 *BACH2*/6q15								
C/C	510 (26.1)	817 (27.8)	413 (25.8)	515 (30.2)	1,626 (26.3)	1,925 (27.9)	890 (29.1)	2,833 (30.8)
C/G	944 (48.2)	1,486 (50.5)	819 (51.2)	862 (50.6)	2,999 (48.4)	3,553 (51.4)	1,511 (49.3)	4,437 (48.3)
G/G	504 (25.7)	639 (21.7)	369 (23.1)	327 (19.2)	1,568 (25.3)	1,435 (20.8)	663 (21.6)	1,922 (20.9)
rs947474 *PRKCQ*/10p15								
A/A	1,364 (69.6)	1,910 (65.0)	1,129 (70.5)	1,118 (65.6)	4,083 (68.4)	4,615 (66.5)	2,066 (70.2)	6,036 (68.4)
A/G	544 (27.8)	921 (31.3)	421 (26.3)	528 (31.0)	1,710 (28.6)	2,083 (30.0)	801 (27.2)	2,481 (28.1)
G/G	51 (2.6)	109 (3.7)	51 (3.2)	58 (3.4)	180 (3.0)	238 (3.4)	75 (2.6)	309 (3.5)
rs3825932 *CTSH*/15q24								
T/T	987 (50.4)	1,343 (45.7)	830 (51.8)	792 (46.5)	3,120 (50.5)	3,243 (47.0)	1,417 (48.1)	3,994 (45.2)
T/C	809 (41.3)	1,310 (44.5)	647 (40.4)	729 (42.8)	2,586 (41.8)	2,963 (42.9)	1,251 (42.4)	3,895 (44.0)
C/C	162 (8.3)	289 (9.8)	124 (7.8)	183 (10.7)	475 (7.7)	701 (10.2)	280 (9.5)	955 (10.8)
rs416603 *C16orf75, PRM3, TNP2*/16p13								
A/A	690 (35.3)	907 (30.8)	511 (31.9)	489 (28.7)	2,043 (32.9)	2,172 (31.4)	N/A	

A/T	933 (47.7)	1,463 (49.7)	798 (49.8)	860 (50.5)	3,085 (49.7)	3,427 (49.6)	N/A	
T/T	333 (17.0)	572 (19.4)	292 (18.3)	355 (20.8)	1,085 (17.5)	1,317 (19.0)	N/A	
rs229541 *C1QTNF6*/22q13								
C/C	552 (28.2)	969 (33.0)	447 (27.9)	546 (32.0)	1,851 (30.1)	2,249 (32.7)	960 (32.8)	2,935 (33.4)
C/T	987 (50.4)	1,400 (47.7)	810 (50.6)	850 (49.9)	3,048 (49.5)	3,403 (49.5)	1,408 (48.1)	4,228 (48.1)
T/T	421 (21.5)	569 (19.4)	344 (21.5)	308 (18.1)	1,255 (20.4)	1,230 (17.9)	562 (19.2)	1,627 (18.5)

Supplementary Table 5. Follow-up SNP genotype summary. †Pseudo-control (untransmitted) genotypes are estimated[26]. N/A = not attempted.

References

1. Clayton, D. & Leung, H.T. An R package for analysis of whole-genome association studies. *Hum Hered* **64**, 45-51 (2007).

2. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* **35**, 34-41 (2006).

3. Bain, S.C. *et al.* The British Diabetic Association--Warren repository. *Autoimmunity* **7**, 83-85 (1990).

4. Lernmark, A. Human cell lines from families available for diabetes research. *Diabetologia* **34**, 61 (1991).

5. Tuomilehto, J. *et al.* Epidemiology of childhood diabetes mellitus in Finland--background of a nationwide study of type 1 (insulin-dependent) diabetes mellitus. The Childhood Diabetes in Finland (DiMe) Study Group. *Diabetologia* **35**, 70-76 (1992).

6. Patterson, C.C. *et al.* A case-control investigation of perinatal risk factors for childhood IDDM in Northern Ireland and Scotland. *Diabetes Care* **17**, 376-381 (1994).

7. Patel, A. *et al.* The TGF-beta 1 gene codon 10 polymorphism contributes to the genetic predisposition to nephropathy in Type 1 diabetes. *Diabet Med* **22**, 69-73 (2005).

8. Undlien, D.E. *et al*. No difference in the parental origin of susceptibility HLA class II haplotypes among Norwegian patients with insulin-dependent diabetes mellitus. *Am J Hum Genet* **57**, 1511-1514 (1995).

9. Ionescu-Tirgoviste, C. *et al.* Low frequency of HLA DRB1*03 - DQB1*02 and DQB1*0302 haplotypes in Romania is consistent with the country's low incidence of Type I diabetes. *Diabetologia* **44 Suppl 3**, B60-B66 (2001).

10. http://www.nimhgenetics.org/.

11. Mueller, P.W. *et al.* Genetics of Kidneys in Diabetes (GoKinD) study: a genetics collection available for identifying genetic susceptibility factors for diabetic nephropathy in type 1 diabetes. *J Am Soc Nephrol* **17**, 1782-1790 (2006).

12. Manolio, T.A. *et al.* New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet* **39**, 1045-1051 (2007).

13. Wellcome Trust Xase Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).

14. http://www.broad.mit.edu/mpg/birdsuite/birdseed.html.

15. Affymetrix, I. BRLMM: An Improved Genotype Calling Method for the Mapping 500K Array Set.

16. Clayton, D.G. *et al*. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* **37**, 1243-1246 (2005).

17. Plagnol, V. *et al.* A method to address differential bias in genotyping in large-scale association studies. *PLoS Genet* **3**, e74 (2007).

18. Frazer, K.A. *et al*. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).

19. Joffe, M.M. & Rosenbaum, P.R. Invited commentary: propensity scores. *Am J Epidemiol* **150**, 327-333 (1999).

20. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).

21. Novembre, J. & Stephens, M. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* **40**, 646-649 (2008).

22. Lowe, C.E. *et al*. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the *IL2RA* region in type 1 diabetes. *Nat Genet* **39**, 1074-1082 (2007).

23. Todd, J.A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* **39**, 857-864 (2007).

24. Smyth, D.J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. *Nat Genet* **38**, 617-619 (2006).

25. Smyth, D.J. *et al*. *PTPN22* Trp620 explains the association of chromosome 1p13 with type 1 diabetes and shows a statistical interaction with HLA class II genotypes. *Diabetes* **57**, 1730-1737 (2008).

26. Cordell, H.J. & Clayton, D.G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* **70**, 124-141 (2002).