Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls

Supplementary Information

The Wellcome Trust Case Control Consortium*

June 5, 2007

Data Quality

One important step in the analysis of genome-wide SNP data is determining a set of quality-assurance filters identifying problem SNPs while discarding as few high quality SNPs as possible. To this end the WTCCC Design and Analysis Group (DAG) investigated a number of different statistics related to genotype clustering, including ratios of cluster mean and variance within and across sample collections (2 control groups plus 7 diseases). As noted in the main text, the best "metric" we found for CHIAMO for identifying difficult-to-call SNPs is the proportion of missing data.

One interesting feature of our QC analyses was the identification of several hundred SNPs which resemble Supplementary Figure 24. For these SNPs, some collections show three tight clusters, whereas others show six distinct clusters: two for each genotype. In fact, all such SNPs show the three cluster pattern in UKBS, RA and CAD and the six cluster pattern in 58C, BD, CD, HT, T1D and T2D. Furthermore, investigation of which specific samples appear in the clusters closer to the origin of the graph revealed that these samples were always the same, and came from early plates in the experiment run at the Affymetrix R&D facility, whereas the samples in the more distantly spaced clusters (and all samples from UKBS, RA and CAD) were processed at the Affymetrix Services Lab. (Based on our QC metrics both laboratories generated data of similar quality.)

We note this observation for two reasons: first that specific laboratory conditions can significantly affect the raw data generated on the Affymetrix 500K chip, and second that such experimental aberrations might be incorrectly followed up as possible copy number variants or other biological processes of interest. While we have ruled out any such explanations, we are left with questions about the experimental process such as why only a few SNPs are affected in this fashion. Note that CHIAMO had no difficulty in correctly calling genotypes when there are these six clusters present, so this artefact had no effect on our analyses.

The plots of missing data rates per sample (Supplementary Figure 25) show a change-point roughly midway through our study, where there is a small but noticeable increase in overall rates. Although we investigated several possibilities, the reasons for this are not clear. There are also plate effects in missing data rates. Again, these phenomena do not affect our analyses.

Interpreting Cluster Plots

Genotype calling errors can easily lead to spurious associations. These can be attributed either to errors made by the calling algorithm or to poor data quality. The latter can make genotype calling very difficult, or in some cases even impossible.

The calling algorithm we use, CHIAMO, correctly calls the vast majority of SNPs. However, the large number of SNPs in the study means that even a small error rate or a small amount of poor quality data can lead to non-trivial numbers of false-positive associated SNPs. Thus, all SNPs that showed strong association were subject to cluster plot inspection to verify their genotype calls.

A cluster plot is a graphical representation of the results of both the genotyping and genotype calling of a SNP. It is a scatter plot of normalised summary probe intensities from the genotyping, with each point representing one individual. Each point is then coloured to indicate how the genotype calling algorithm decided to

*List of participants and affiliations appear at the end of the Main paper.

classify that individual (either as a homozygote for one of the two alleles, a heterozygote, or a "null" (missing) call).

The aim of examining a cluster plot is twofold: to determine whether the given SNP has been genotyped well — in particular, whether we see clear, distinct, clusters on the plot that would correspond to the three genotypes — and also to determine whether the calling algorithm has called the clusters correctly. If both of these are true, we can usually be confident that the genotype counts are accurate. If these are not true, any associations that we observe at such SNPs may well be caused by the resulting incorrect genotype counts.

SNPs observed to have a cluster plot error are generally excluded from the analysis. However, depending on what kind of error it is and any other evidence of association, some method of error recovery could be attempted — e.g. re-running CHIAMO, re-genotyping using a different platform, or imputation of genotypes.

One particularly challenging but relatively rare scenario relates to differential missingness. Some samples may be hard to call, especially if they lie on cluster boundaries, so instead of calling them a certain genotype the algorithm may call them as "null" (i.e. they will be missing data). Ordinarily, this isn't a problem and very few individuals will be called null. However, if clusters are close together, or even overlapping, many of the individuals on the adjacent boundaries will be called null (and these represent sensible calls given the uncertainty). Often, only two of the clusters are close together (the heterozygote and one of the homozygotes), which leads to most of the missing individuals to be those that preferentially have one of the two alleles, resulting in biased allele frequency estimates. If this happens differentially between cases and controls it can lead to spurious associations. We used a posterior probability threshold in CHIAMO of 0.9, treating the genotype as missing when the most probable call fell below this threshold. With this choice of threshold the call rate was high. Counterintuitively, we found that increasing the threshold in an attempt to improve data quality was counter-productive, leading to increased false positives because of differential missingness.

Despite the multiple potential sources of error, CHIAMO and BRLMM get nearly all the calls correct (Supplementary Table 3). Using CHIAMO as described, typically only about 100 cluster plots required inspection per disease for a genomewide scan.

Interpreting Signal Plots

Signal plots can help us learn about the believability and characteristics of putative disease associations. Most real susceptibility loci should show elevated signals at multiple nearby SNPs unless the recombination rate in the region is very high or the SNP density very low. This is because sets of SNPs that are near a disease locus will often be correlated and consequently share an allele frequency disparity between cases and controls. This effect should decline with genetic distance (and decreasing LD) but also depends on factors like minor allele frequency. A single elevated signal is often the spurious result of data or analysis artifacts, such as miscalled genotypes at a SNP.

Once a plausible hit region has been identified, we can learn more about it by delimiting its boundaries. This provides some basic information about the part of the genome where the association signal was found and guides follow-up studies of interesting regions. We specify hit region boundaries by looking for positions flanking a hit SNP where the signal returns to background levels; if this coincides with a recombination hotspot, as it often does, we choose that as the boundary. Historical recombinations — particularly those concentrated in hotspots —- will tend to decrease correlations between marker and disease loci. The gene and sequence conservation tracks on our signal plots also provide useful information about genomic context in the neighborhood of a putative susceptibility locus: highly conserved tracts and known genes in the hit region suggest obvious candidates for further scrutiny. Inclusion of imputed SNPs gives an additional level of resolution and, as in some of our examples, can potentially identify a stronger signal of association at a SNP not assayed directly.

Bayes Factors

Consider a SNP with two alleles coded 0 and 1. Suppose we have genotypes at this SNP in a set of N individuals (N_1 cases and N_2 controls). We use Y_i to denote the binary disease phenotype of individual i (cases have Y_i = 1, controls have Y_i = 0). Let Z_i denote the count of alleles coded as 1 for individual i. The data at the SNP can be summarised in the following table:

Z	0	1	2
Cases	s_0	s_1	s_2
Controls	r_0	r_1	r_2

We use M_0 to denote a model of no association, M_1 for a model with an additive effect on the log-odds scale and M_2 for a general 3 parameter model of association. The Bayes Factor between models M_1 and M_0 is defined as

$$BF_1 = \frac{P(D|M_1)}{P(D|M_0)} = \frac{\int P(D|\theta_1, M_1)P(\theta_1|M_1)d\theta_1}{\int P(D|\theta_0, M_0)P(\theta_0|M_0)d\theta_0}$$

where D is used to denote the data, θ_1 and θ_0 are the parameters of the models M_1 and M_0. It can be clearly seen that instead of maximising the likelihood (as is the case for frequentist tests) under the two models the parameters are integrated out of the likelihood with a weighting given by the prior distribution on the parameters.

For both models we use a logistic regression model for the likelihood

$$P(D|\theta) = \prod_{i=1}^{N} p_i^{Y_i}(1-p_i)^{1-Y_i}$$

where for model M_1 we have

$$\theta_1 = (\mu, \gamma) \qquad \log\frac{p_i}{1-p_i} = \mu + \gamma Z_i,$$

and for model M_0 we have

$$\theta_0 = (\mu) \qquad \log\frac{p_i}{1-p_i} = \mu.$$

We now need to specify the prior distribution $P(\theta_1|M_1) = P(\mu, \gamma|M_1)$. The parameter μ represents the baseline odds of disease. This parameter will be influenced by the numbers of cases and controls in the dataset. In a case-control design the numbers of cases in the sample have been elevated artificially which will have a large effect on likely values of μ. For this reason we wish to use a prior distribution that allows flexibility in the specification of our beliefs on μ so we use a $\mathrm{N}(\alpha_1, \beta_1)$ distribution. In practice we have used $\mu \sim \mathrm{N}(0, 1)$.

The parameter γ is the increase in log-odds of disease for every copy of the risk allele and e^γ is the additive model odds ratio. We have some good prior information about likely values of this parameter. For example, it is widely believed that the genetic variants underlying common disease will have risk allele odds-ratios in the range 1–2 with substantially more weight on the values between 1–1.5. Note that this implies a protective allele odds-ratio in the range 0.5–1 with substantially more weight on values between 0.67–1. After some experimentation we settled on a flexible prior distribution for γ of a $\mathrm{N}(\alpha_2, \beta_2)$ distribution. For example, Figure 1 shows a density plot for the additive model odds-ratio (e^γ) from a sample of 1,000,000 draws from the prior $\gamma \sim \mathrm{N}(0, 0.2)$.

Overall, the prior distribution on the parameters has the form

$$P(\theta_1|M_1) \propto \frac{1}{\beta_1} e^{-\frac{(\mu-\alpha_1)^2}{2\beta_1^2}} \frac{1}{\beta_2} e^{-\frac{(\gamma-\alpha_2)^2}{2\beta_2^2}}.$$

For $P(\theta_0|M_0) = P(\mu|M_0)$ we used the same prior on μ as in the model M_1. That is,

$$P(\theta_0|M_0) \propto \frac{1}{\beta_1} e^{-\frac{(\mu-\alpha_1)^2}{2\beta_1^2}}.$$

It is well understood that the priors on the parameters of the model can have a non-negligible impact on the value of the Bayes Factor[1] even as the amount of data gets large. In line with this we have found that using different priors on μ for the two models can substantially change the Bayes Factor. We have little strong prior information about μ and as noted above the case-control ratio will have a large effect on the values that best fit the data. For these reasons we use a reasonably diffuse prior distribution on this parameter that is the same for both models. This acts to focus the comparison between the models on the parameter γ which is the main parameter of interest.

To evaluate the marginal likelihood for the model $P(D|M_1)$ we need to evaluate the integral

$$\int P(D|\theta_1, M_1)P(\theta_1|M_1)d\theta_1.$$

We do this using a Laplace Approximation[2] in which the posterior distribution is approximated using a Gaussian distribution centred on its mode. More specifically, we use

$$\log P(D|M_1) \approx \log P(D|\hat{\theta_1}, M_1) + \log P(\hat{\theta_1}|M_1) + \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|A|$$

where $\hat{\theta_1}$ is the value of θ_1 that maximises $P(D|\theta_1, M_1)P(\theta_1|M_1)$, and is known as the *maximum a posteriori* (MAP) estimate of θ_1. Also, A is the negative Hessian of $P(D|\theta_1, M_1)P(\theta_1|M_1)$ evaluated at $\hat{\theta_1}$ and d is the dimension of θ_1. We use Newton-Raphson optimisation to find $\hat{\theta_1}$ but if this fails to converge we use a line-search method. Both approaches are numerically efficient for this low dimensional integral.

In addition, we note that the evaluation of this marginal likelihood will depend upon the way the alleles at the SNP have been coded 0 and 1. Thus, to calculate the

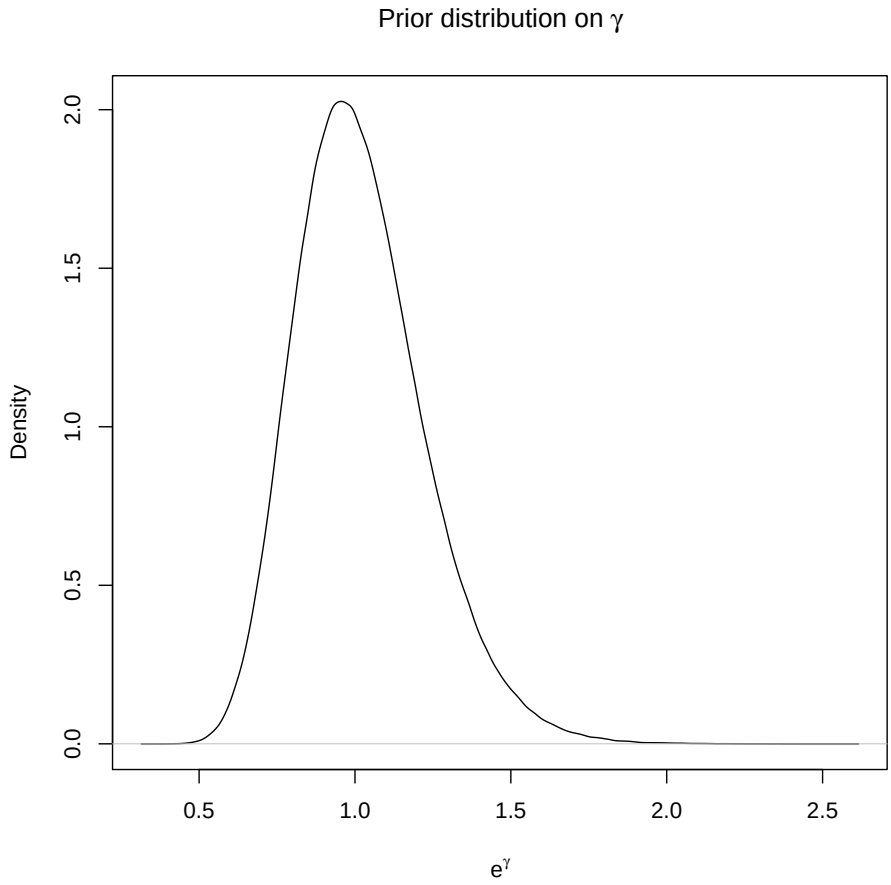


Figure 1: Density plot of the empirical distribution of e^{γ} from a sample of size 10^6 from the distribution $\gamma \sim \mathrm{N}(0, 0.2)$

marginal likelihood for the additive model of the SNP we average over the two possible codings with equal weight. We have used a similar formulation for dominant and recessive models[?] .

The general 3 parameter model, M_2, is slightly more complicated in that we require a prior distribution on an additional parameter. We use the following model for the log-odds

$$\log \frac{p_i}{1-p_i} = \mu + \gamma \mathrm{I}(Z_i = 1) + 2\phi\gamma \mathrm{I}(Z_i = 2)$$

which has an additive genetic effect parametrised by γ and then an additional recessive effect parametrised by ϕ. In this model the additive model occurs when $\phi = 1$. We use a Gaussian prior, $\mathrm{N}(\alpha_3, \beta_3)$ for ϕ. In practice we use a $\mathrm{N}(1, 1)$ for ϕ which results in a symmetric departure from the additive model and we use the same prior for γ i.e. $\mathrm{N}(0, 0.2)$ as we did above when we considered the additive model. As with the additive model we use a Laplace approximation to evaluate the required integral we average the marginal likelihood over the two possible codings for the SNP.

Other priors that are more computationally efficient are possible[3]. For example, for the general 3 parameter model if we use the formulation

$$\log \frac{p_i}{1-p_i} = \mu \mathrm{I}(Z_i = 0) + \gamma \mathrm{I}(Z_i = 1) + \phi \mathrm{I}(Z_i = 2)$$

in which each genotype is given its own log-odds parameter then the likelihood can be re-written as

$$P(D|\theta_2, M_2) = p_0^{s_0}(1-p_0)^{r_0} p_1^{s_1}(1-p_1)^{r_1} p_2^{s_2}(1-p_2)^{r_2}$$

where $p_0 = \frac{e^{\mu}}{1+e^{\mu}}$, $p_1 = \frac{e^{\gamma}}{1+e^{\gamma}}$ and $p_2 = \frac{e^{\phi}}{1+e^{\phi}}$. This has the form of an independent Binomial Likelihood for each of the three penetrance parameters p_0, p_1 and p_2. Thus we could use a conjugate Beta prior for these parameters which facilitates the exact calculation of the integrals. That is, if we let

$$P(\theta_2|M_2) = \prod_{g=0}^{2} \frac{1}{\beta(\psi_g, \eta_g)} p_g^{\psi_g - 1}(1-p_g)^{\eta_g - 1}$$

where $\beta(\psi_g, \eta_g) = \frac{\Gamma(\psi_g)\Gamma(\eta_g)}{\Gamma(\psi_g+\eta_g)}$ then

$$P(D|M_2) = \prod_{g=0}^{2} \frac{\beta(s_g + \psi_g, r_g + \eta_g)}{\beta(\psi_g, \eta_g)}.$$

Similar expressions can be derived for the marginal likelihoods for dominant and recessive models using this class of conjugate priors. For the null model M_0 of no association we obtain a marginal likelihood of

$$P(D|M_0) = \frac{\beta(s_0 + s_1 + s_2 + \psi_0, r_0 + r_1 + r_2 + \eta_0)}{\beta(\psi_0, \eta_0)}$$

where a Beta(ψ_0, η_0) is used for the baseline penetrance.

It is interesting to consider what the conjugate Beta priors on penetrance actually mean in terms of odds-ratios. It can be shown that a $\mathrm{Beta}(a, b)$ prior on a probability p is equivalent to a Generalised Logistic distribution on the log-odds $(\log \frac{p}{1-p})$ with mean $\Psi^{(0)}(a) - \Psi^{(0)}(b)$ and variance $\Psi^{(1)}(a) + \Psi^{(1)}(b)$ where $\Psi^{(r)}$ is the polygamma function[4]. For example, a uniform distribution, $p \sim \mathrm{Beta}(1, 1)$, results in a distribution for log-odds centred on 0 with a variance of $\pi^2/3$. This implies that the prior on the difference in log-odds between the heterozygote genotype and the baseline homozygote genotype has mean 0 and variance $2\pi^2/3$. This is considerably more diffuse than the $\mathrm{N}(0, 0.2)$ prior we use in the additive model above. Using simulation from this prior we found that it corresponds to a prior distribution on the risk-allele odds ratio with a mean of approximately 80, which is rather larger than might be expected for common human diseases. This suggests that for the General, Dominant and Recessive models in which Beta priors are applicable it might be more reasonable to set the hyper-parameters a and b to be greater than 1. This would bring the priors closer to those we have suggested above.

CHIAMO

Normalisation

The raw data for the sth SNP of the ith individual/array can be denoted as $I_{isk} = (I^{PA}_{isk}, I^{PB}_{isk}, I^{MA}_{isk}, I^{MB}_{isk})$. These denote the kth perfect match probe intensities for the A and B alleles and the mismatch probes probe intensities for the A and B alleles respectively, where $k = 1, \ldots, K$ and $K \in \{6, 10\}$. We create a set of normalised intensities, I', using the standard pre-processing step of quantile normalisation to reduce variability across chips[5,6]. To carry out this step on the 17,000 arrays in the WTCCC study we wrote our own software to improve efficiency. Second, we log transform the quantile normalised intensities $Y = \log(I')$ to reduce the skewness of the intensities[6].

We use $Y_{iks} = (Y^{PA}_{iks}, Y^{PB}_{iks}, Y^{PA}_{iks}, Y^{PB}_{iks})$ to denote the vector of log-normalised intensities for the kth probe quartet of individual i at SNP s. We use the following transformation to correct for the average background hybridisation across the A and B alleles

$$Y^A_{iks} = \begin{cases} Y^{PA}_{iks} - \frac{1}{2}\left(Y^{MA}_{iks} + Y^{MB}_{iks}\right) & \text{if } Y^{PA}_{iks} \geq Y^{MB}_{iks} \\ 0 & \text{if } Y^{PA}_{iks} < Y^{MB}_{iks} \end{cases}$$

$$Y^B_{iks} = \begin{cases} Y^{PB}_{iks} - \frac{1}{2}\left(Y^{MA}_{iks} + Y^{MB}_{iks}\right) & \text{if } Y^{PB}_{iks} \geq Y^{MB}_{iks} \\ 0 & \text{if } Y^{PB}_{iks} < Y^{MB}_{iks}. \end{cases}$$

We then pool signals across probes using a simple arithmetic mean to create a pair of intensities $X_{is} = (X^A_{is}, X^B_{is})$ for individual i at SNP s,

$$X^A_{is} = \frac{1}{n_s}\sum_{k=1}^{n_s} Y^A_{iks}, \qquad X^B_{is} = \frac{1}{n_s}\sum_{k=1}^{n_s} Y^B_{iks},$$

where n_s is the number of probe quartets at SNP s. Through visual inspection of thousands of SNP intensity plots we found that our normalisation scheme produced tighter clusters with fewer outlying observations than a median polish approach[6].

A Bayesian Hierarchical Mixture Model

We use X_{ij} to denote the bi-variate intensity vector for jth individual in ith collection. We use N_i to denote the number of individuals in the ith collection and C to denote the number of collections. The basic model for the set of intensities within each collection is a 4-class Gaussian mixture model. We have a class for each of the genotypes $\{AA, AB, BB\}$ as well as a $NULL$ outlier class to capture the clear outlying observations in each collection and to add robustness to the model fit of the other 3 genotype classes. We use Z_{ij} to denote the genotype call for the jth individual in the ith collection where $Z_{ij} \in \{0, 1, 2, 3\} \equiv \{AA, AB, BB, NULL\}$. The mean and covariance matrix of the kth cluster in ith collection are denoted by μ_{ik} and Σ_{ik} where

$$\mu_{ik} = (\mu_{ikA}, \mu_{ikB}) \quad \text{and} \quad \Sigma_{ik} = \begin{pmatrix} \sigma^2_{iAA} & \sigma^2_{iAB} \\ \sigma^2_{iAB} & \sigma^2_{iBB} \end{pmatrix}$$

so that

$$X_{ij}|Z_{ij}, \mu_{iZ_{ij}}, \Sigma_{iZ_{ij}} \sim \mathrm{MVN}(\mu_{iZ_{ij}}, \Sigma_{iZ_{ij}}).$$

All individuals are conditionally independent given the class labels, means and covariances which gives

$$p(X|Z,\mu,\Sigma) = \prod_{i=1}^{C}\prod_{j=1}^{N_i} p(X_{ij}|Z_{ij},\mu_{iZ_{ij}},\Sigma_{iZ_{ij}}). \quad (1)$$

The class labels Z_{ij} depend upon a vector of class proportions for the ith collection which we denote as $\lambda_i = \{\lambda_{i0},\lambda_{i1},\lambda_{i2},\lambda_{i3}\}$. We use a multinomial distrbution for each of the class labels such that

$$Z_{ij}|\lambda_i \sim \text{Multinomial}(1,\lambda_i),$$

and assume independence of labels within and across collections conditional upon collection genotype proportions

$$p(Z|\lambda) = \prod_{i=1}^{C}\prod_{j=1}^{N_i} p(Z_{ij}|\lambda_i). \quad (2)$$

The proportions of the classes 1–3 correspond to the genotype proportions in the collection. These proportions depend upon the allele frequency and the proportion of $NULL$ data in the ith collection which we denote α_i and η_i respectively. Since we expect most SNPs to conform closely but not exactly to the Hardy-Weinberg Law we use a Dirichlet distribution with a variance parameter θ to model the dependence of the class proportions on α_i and η_i as follows

$$\begin{aligned}\lambda_i|\alpha_i,\theta,\eta_i \sim \text{Dirichlet}((&(1-\eta_i)\alpha_i^2\theta,\\ &(1-\eta_i)2\alpha_i(1-\alpha_i)\theta,\\ &(1-\eta_i)(1-\alpha_i)^2\theta,\\ &\eta_i\theta\ ).\end{aligned}$$

We assume independence of the genotype proportions across collections conditional on the collection allele frequencies and proportions of $NULL$ data

$$p(\lambda|\alpha,\eta) = \prod_{i=1}^{C} p(\lambda_i|\alpha_i,\eta_i,\theta). \quad (3)$$

The parameter θ is a fixed constant that controls the degree to which the data should conform to the Hardy-Weinberg Law. We have used $\theta = 10$.

We use a model for the means μ_i of each collection that is based on the following parametrisation

$$\begin{aligned}\mu &= \{\mu_i; i = 1,\ldots,C\}\\ \Rightarrow \mu' &= \{c_i, l_i, g_i, f_i, d_i, \mu_{i3}; i = 1,\ldots,C\}.\end{aligned}$$

We use c_i, l_i and g_i to denote the midpoint, length and negative gradient of the line joining the centres of the homozygote clusters μ_{i0} and μ_{i2}. We let e_i be the point of intersection of the line between μ_{i0} and μ_{i2} and a perpendicular line through μ_{i1}. We use f_i to denote the distance from e_i to μ_{i0} expressed as a fraction of the length l_i. More specifically,

$$f_i = \frac{(\mu_{i0}-\mu_{i2}).(\mu_{i0}-\mu_{i1})}{(\mu_{i0}-\mu_{i2}).(\mu_{i2}-\mu_{i0})}.$$

Finally, we use d_i to denote the distance from μ_{i1} to e_i expressed as a fraction of its maximum possible distance. This maximum distance is the length of $h_i - e_i$ where h_i is defined such that μ_{i1} cannot have an x-coordinate larger than the x-coordinate of μ_{i0} or a y-coordinate larger than the y-coordinate of μ_{i2}. The centre of the $NULL$ class is fixed to a constant and is described in more detail below.

Given this parametrisation we model the variable components of μ' using a hierarchical structure that links the parameters of each collection together conditional upon a set of mean and variance parameters

$$\begin{aligned}p(\mu') \;=\; &\prod_{i=1}^{C} p(c_i|c_\mu,c_{\sigma^2})p(c_\mu)p(c_{\sigma^2})\\ &\times p(l_i|l_\mu,l_{var})p(l_\mu)p(l_{var})\\ &\times p(g_i|g_\mu,g_{var})p(g_\mu)p(g_{var})\\ &\times p(f_i|f_\mu,f_{var})p(f_\mu)p(f_{var})\\ &\times p(d_i|d_\mu,d_{var})p(d_\mu)p(d_{var}). \quad (4)\end{aligned}$$

The prior structure on the parameters c_i, l_i, f_i and d_i is as follows:

$$\begin{aligned}c_i &\sim \text{N}(c_\mu, c_{\sigma^2}I_2),\\ c_\mu &\sim \text{N}((1,1),1),\\ c_{\sigma^2} &\sim \text{Scale}-\text{Inv}{-}\chi^2(0.0001,100),\\ l_i &\sim \text{Gamma}\Big(\frac{l_\mu^2}{l_{var}},\frac{l_\mu}{l_{var}}\Big),\\ g_i &\sim \text{Gamma}\Big(\frac{g_\mu^2}{g_{var}},\frac{g_\mu}{g_{var}}\Big),\\ l_\mu &\sim \text{Gamma}(20,20),\\ l_{var} &\sim \text{Gamma}(1,100),\\ g_\mu &\sim \text{Gamma}(10,10),\\ g_{var} &\sim \text{Gamma}(1,100),\\ f_i &\sim \text{Beta}\Big(f_\mu\frac{1-f_{var}}{f_{var}},(1-f_\mu)\frac{1-f_{var}}{f_{var}}\Big),\\ d_i &\sim \text{Beta}\Big(d_\mu\frac{1-d_{var}}{d_{var}},(1-d_\mu)\frac{1-d_{var}}{d_{var}}\Big),\\ f_\mu &\sim \text{Beta}(30,30),\\ f_{var} &\sim \text{Beta}(1,100),\\ d_\mu &\sim \text{Beta}(2,8),\\ d_{var} &\sim \text{Beta}(1,100).\end{aligned}$$

We use the following reparameterization of the covariance matrices Σ_i

$$\begin{aligned}\Sigma &= \{\Sigma_i; i = 1, \ldots, C\} \\ \Rightarrow \Sigma' &= \{\sigma_i^2, \phi_i, r_{i0}, r_{i1}, r_{i2}, \Sigma_{i3}; i = 1, \ldots, C\},\end{aligned}$$

where

$$\begin{aligned}\Sigma_{i0} &= \begin{pmatrix} \sigma_i^2 & r_{i0}\phi_i\sigma_i^2 \\ r_{i0}\phi_i\sigma_i^2 & \phi_i^2\sigma_i^2 \end{pmatrix}, \\ \Sigma_{i1} &= \begin{pmatrix} \sigma_i^2 & r_{i1}\sigma_i^2 \\ r_{i1}\sigma_i^2 & \sigma_i^2 \end{pmatrix}, \\ \Sigma_{i2} &= \begin{pmatrix} \phi_i^2\sigma_i^2 & r_{i2}\phi_i\sigma_i^2 \\ r_{i2}\phi_i\sigma_i^2 & \sigma_i^2 \end{pmatrix}.\end{aligned}$$

The prior structure on Σ' is as follows

$$\begin{aligned}p(\Sigma') &= \prod_{i=1}^{C} p(\sigma_i|\sigma_\mu, \sigma_{var})p(\sigma_\mu)p(\sigma_{var}) \\ &\quad \times p(\phi_i|\phi_\mu, \phi_{var})p(\phi_\mu)p(\phi_{var}) \\ &\quad \times \prod_{j=1}^{3} p(r_{ij}|r_{j\mu}, r_{jvar}) \\ &\quad \times p(r_{j\mu})p(r_{jvar}),\end{aligned} \tag{5}$$

$$\begin{aligned}\sigma_i &\sim \text{Gamma}(\frac{\sigma_\mu^2}{\sigma_{var}}, \frac{\sigma_\mu}{\sigma_{var}}), \\ \sigma_\mu &\sim \text{Gamma}(1, 4), \\ \sigma_{var} &\sim \text{Gamma}(1, 100), \\ \phi_i &\sim \text{Beta}(\phi_\mu\frac{1-\phi_{var}}{\phi_{var}}, (1-\phi_\mu)\frac{1-\phi_{var}}{\phi_{var}}), \\ \sigma_\mu &\sim \text{Beta}(1, 4) \\ \sigma_{var} &\sim \text{Beta}(1, 100), \\ r_{ij} &\sim \text{Beta}(r_{j\mu}\frac{1-r_{jvar}}{r_{jvar}}, (1-r_{j\mu})\frac{1-r_{jvar}}{r_{jvar}}), \\ r_{0\mu} &\sim \text{Beta}(1, 50), \\ r_{1\mu} &\sim \text{Beta}(10, 1), \\ r_{2\mu} &\sim \text{Beta}(1, 50), \\ r_{0var} &\sim \text{Beta}(1, 100), \\ r_{1var} &\sim \text{Beta}(1, 100), \\ r_{2var} &\sim \text{Beta}(1, 100).\end{aligned}$$

We fix the parameters of the outlier class to be a very flat density to capture outlier observation in an approximately uniform way across the intensity space.

$$\mu_{i3} = (1, 1), \quad \Sigma_{i3} = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}.$$

We model the dependency of allele frequencies across collections by conditioning on an unknown 'global' or mean allele frequency α_μ and an allele frequency variance parameter α_{var} such that

$$\alpha_i \sim \text{Beta}(\alpha_\mu\frac{1-\alpha_{var}}{\alpha_{var}}, (1-\alpha_\mu)\frac{1-\alpha_{var}}{\alpha_{var}}).$$

For the WTCCC we use $\alpha_{var} = 0.005$ and $\alpha_\mu \sim \text{Beta}(1, 1)$. Further, we allow for additional prior information on the allele frequency of the collections through the specification of an extra fixed allele frequency α' with distribution

$$\alpha' \sim \text{Beta}(\alpha_\mu\frac{1-\alpha'_{var}}{\alpha'_{var}}, (1-\alpha_\mu)\frac{1-\alpha'_{var}}{\alpha'_{var}}).$$

In the current implementation we have set α' equal to the empirical allele frequency of the SNP from the CEU HapMap population if available and set $\alpha'_{var} = 0.01$. Overall, the distribution for α is given by

$$p(\alpha) = \prod_{i=1}^{C} p(\alpha_i|\alpha_\mu, \alpha_{var})p(\alpha_\mu)p(\alpha'|\alpha_\mu). \tag{6}$$

We model dependency of the $NULL$ class proportions across collections by conditioning on an unknown mean proportion η_μ and a variance parameter η_{var} such that

$$\eta_i \sim \text{Beta}(\eta_\mu\frac{1-\eta_{var}}{\eta_{var}}, (1-\eta_\mu)\frac{1-\eta_{var}}{\eta_{var}}),$$

so that

$$p(\eta) = \prod_{i=1}^{C} p(\eta_i|\eta_\mu, \eta_{var})p(\eta_\mu)p(\eta_{var}), \tag{7}$$

$$\eta_\mu \sim \text{Beta}(3, 100), \quad \eta_{var} \sim \text{Beta}(5, 100).$$

Overall, the posterior distribution is constructed using equations (1), (2), (3), (4), (5), (6) and (7) to give

$$\begin{aligned}p(Z,&\mu, \Sigma, \lambda, \alpha, \eta|X) \propto \\ &p(X|Z, \mu, \Sigma)\, p(\mu')\, p(\Sigma') \\ &\times p(Z|\lambda)\, p(\lambda|\alpha, \eta)\, p(\alpha)\, p(\eta).\end{aligned} \tag{8}$$

Software

A C++ implementation of the algorithm called CHIAMO* is available. Separate normalization software is also available. Please email Jonathan Marchini at `marchini@stats.ox.ac.uk` in the first instance to obtain this software.

*chiamo means 'I call' in Italian and is derived from the verb chiamare.

Membership

The following individuals were responsible for the stated activities within the WTCCC

Study Working Group (2003-2005): Lon R. Cardon, David G. Clayton, Panos Deloukas, Peter Donnelly, Marcus Pembrey, David P. Strachan, John A. Todd, David R. Bentley (Chair);

Publications Committee: Nick Craddock, Peter Donnelly, Willem H. Ouwehand, Nilesh J. Samani, Mark I. McCarthy (Chair);

Writing Committee: David G. Clayton, Nick Craddock, Panos Deloukas, Mark I. McCarthy, Peter Donnelly (Chair);

Joint Steering Committee: Simon Cawley, Alan Dance*, Hossein Fakhrai-Rad, Rui Mei*, Raji Pillai (Affymetrix); Suzannah J. Bumpstead, Claire Bryan, David G. Clayton, Lon R. Cardon, Panos Deloukas*, Sarah Nutland, Simon Potter, Helen E. Stevens, John A. Todd*, Neil M. Walker, Pamela Whittaker (*=voting member).

Membership of BRAGGS and BCSC

BRAGGS

- Anne Barton, arc Epidemiology Unit, University of Manchester, Oxford Rd, Manchester, M13 9PT
- John D. Isaacs, Department of Rheumatology, University of Newcastle-Upon-Tyne, Framlington Place, Newcastle-Upon-Tyne NE2 4HH
- Ann W. Morgan, Leeds Institute of Molecular Medicine, Section of Academic Unit of Musculoskeletal Disease Wellcome Trust Brenner Building, St. James's University Hospital, Beckett Street, Leeds LS9 7TF
- Gerry D. Wilson, Genomic Medicine, The University of Sheffield, Western Bank, Sheffield, S10 2TN

Breast Cancer Susceptibility Collaboration (UK)

The Breast Cancer Susceptibility Collaboration (UK) consists of the following contributors:

A Ardern-Jones, J Berg, A Brady, N Bradshaw, C Brewer, G Brice, B Bullman, J Campbell, B Castle, R Cetnarsryj, C Chapman, C Chu, N Coates, T Cole, R Davidson, A Donaldson, H Dorkins, F Douglas, D Eccles, R Eeles, F Elmslie, DG Evans, S Goff, S Goodman, D Goudie, J Gray, L Greenhalgh, H Gregory, SV Hodgson, T Homfray, RS Houlston, L Izatt, L Jackson, L Jeffers, V Johnson-Roffey, F Kavalier, C Kirk, F Lalloo, C Langman, I Locke, M Longmuir, J Mackay, A Magee, S Mansour, Z Miedzybrodzka, J Miller, P Morrison, V Murday, J Paterson, M Porteous, N Rahman, M Rogers, S Rowe, S Shanley, A Saggar, G Scott, L Side, L Snadden, M Steel, M Thomas, S Thomas

References

1. O'Hagan, A. and Forster, J. *Bayesian Inference*. Arnold, (2004).

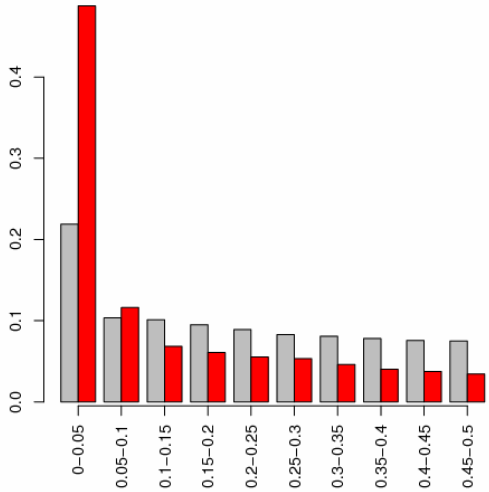
2. Jordan, M. I., editor. *Learning in Graphical Models*. Kluwer Academic Publishers, (1998).

3. Balding, D. J. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* **7**, 781–791 (2006).

4. Jong-Wuu, W., Wen-Liang, H., and Lee, H. Some moments and limit behaviours of the generalized logistic distribution with applications. *Proc. Natl. Sci, Counc. ROC(A)* **24**(1), 7–14.

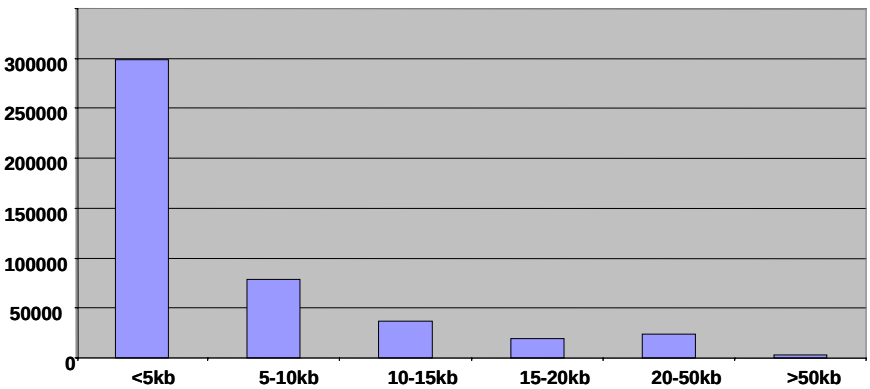
5. Bolstad, B. M., A., I. R., strand, M., and Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).

6. Rabbee, N. and Speed, T. A genotype calling algorithm for affymetrix snp arrays. *Bioinformatics* **22**(1), 7–12 (2006).

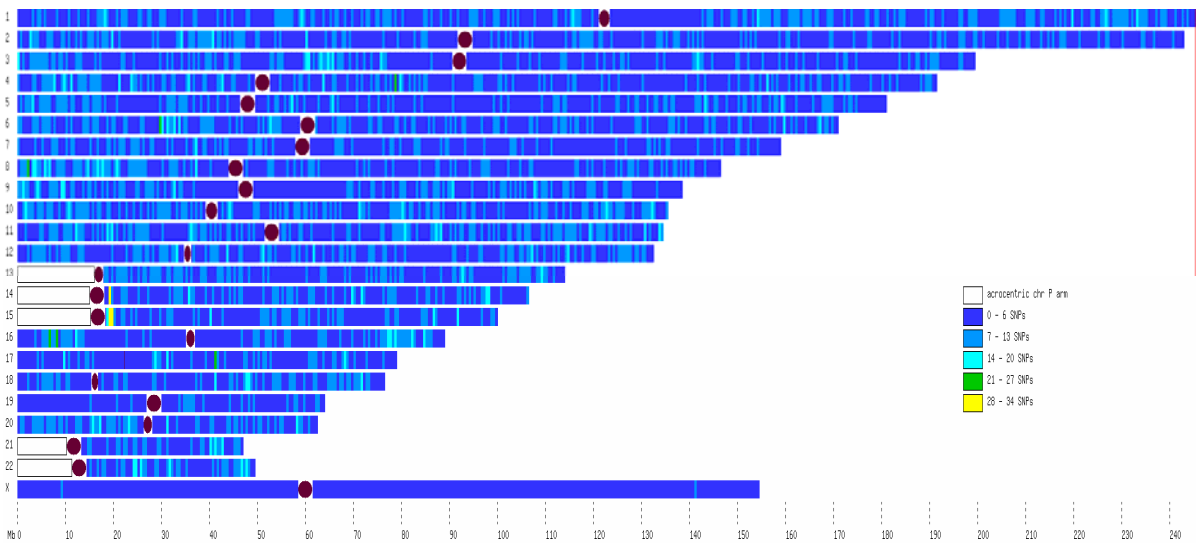


Supplementary Figure 1 | Minor allele frequency (MAF) spectrum of SNPs included and excluded from the study. Red bars show the proportion of SNPs excluded from the study (see text and Methods) in 10 MAF bins. Grey bars show the frequency spectrum of included SNPs.

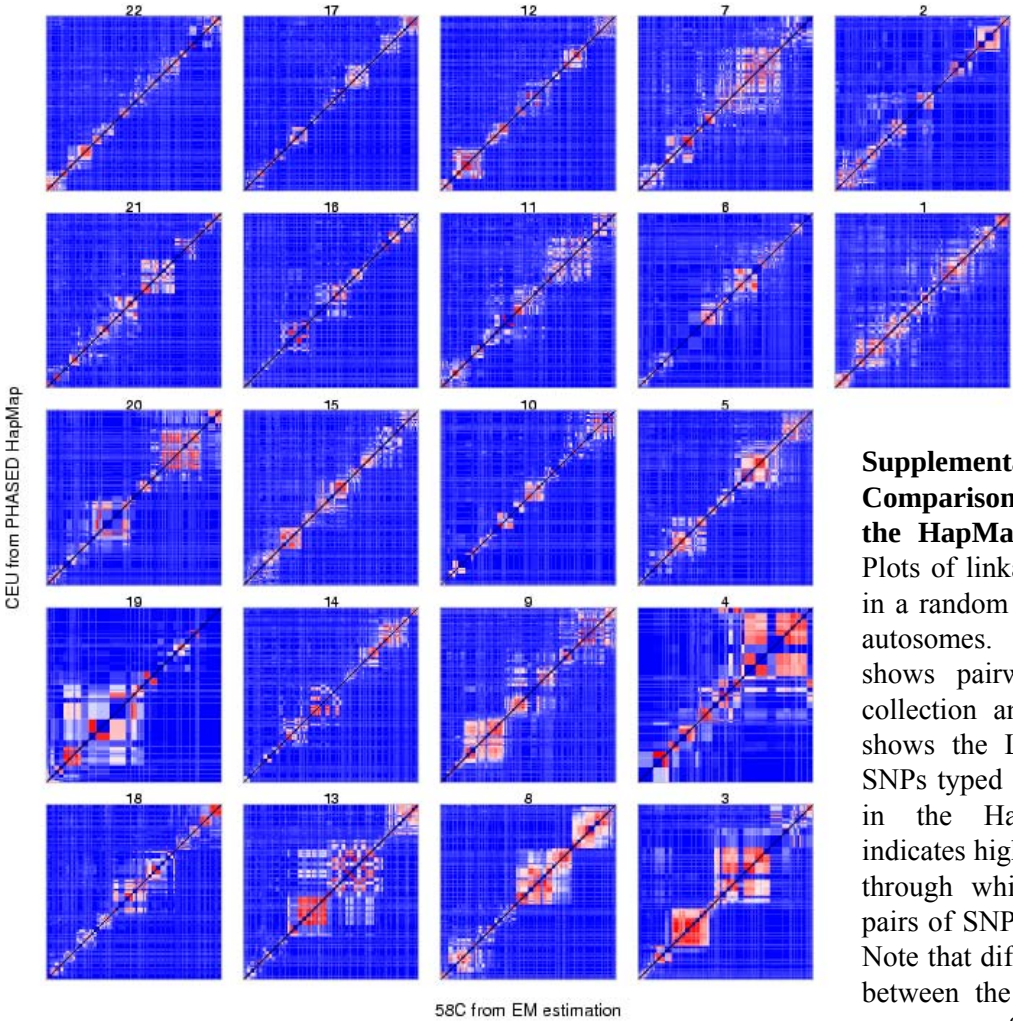
a)



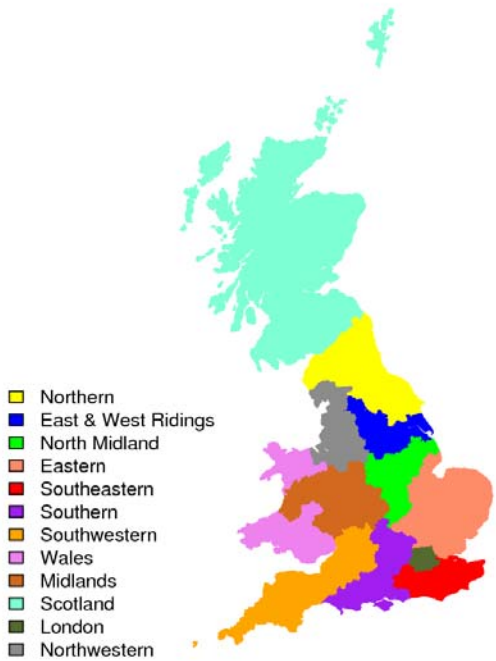
b)



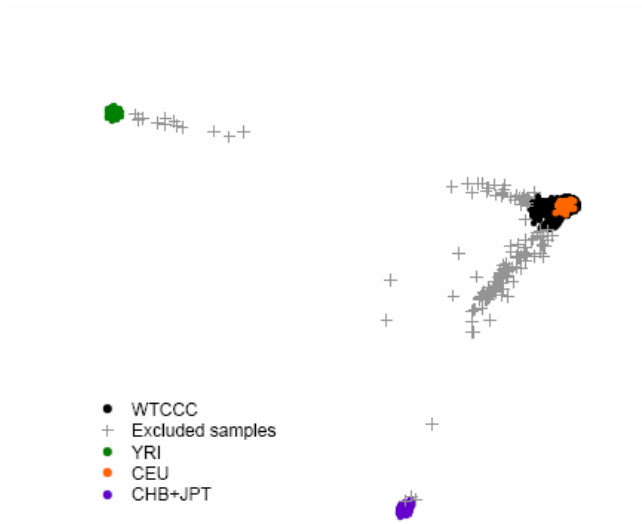
Supplementary Figure 2 | Marker density on the Affymetrix 500k chip. Distribution of SNP markers on the 500K Affymetrix array based on QC. a) Distribution of inter-marker distances for the ~459,000 autosomal SNPs Passing QC filters. b) Chromosomal distribution of SNPs failing QC criteria in 500 kb windows. Each window is coloured according to SNP density as follows 0-6, blue; 7-13, light blue; 14-20 turquoise; 21-27, green; and 28-34 yellow. Centromeres are depicted as black ovals and the p-arm of the acrocentric chromosomes as white rectangles.



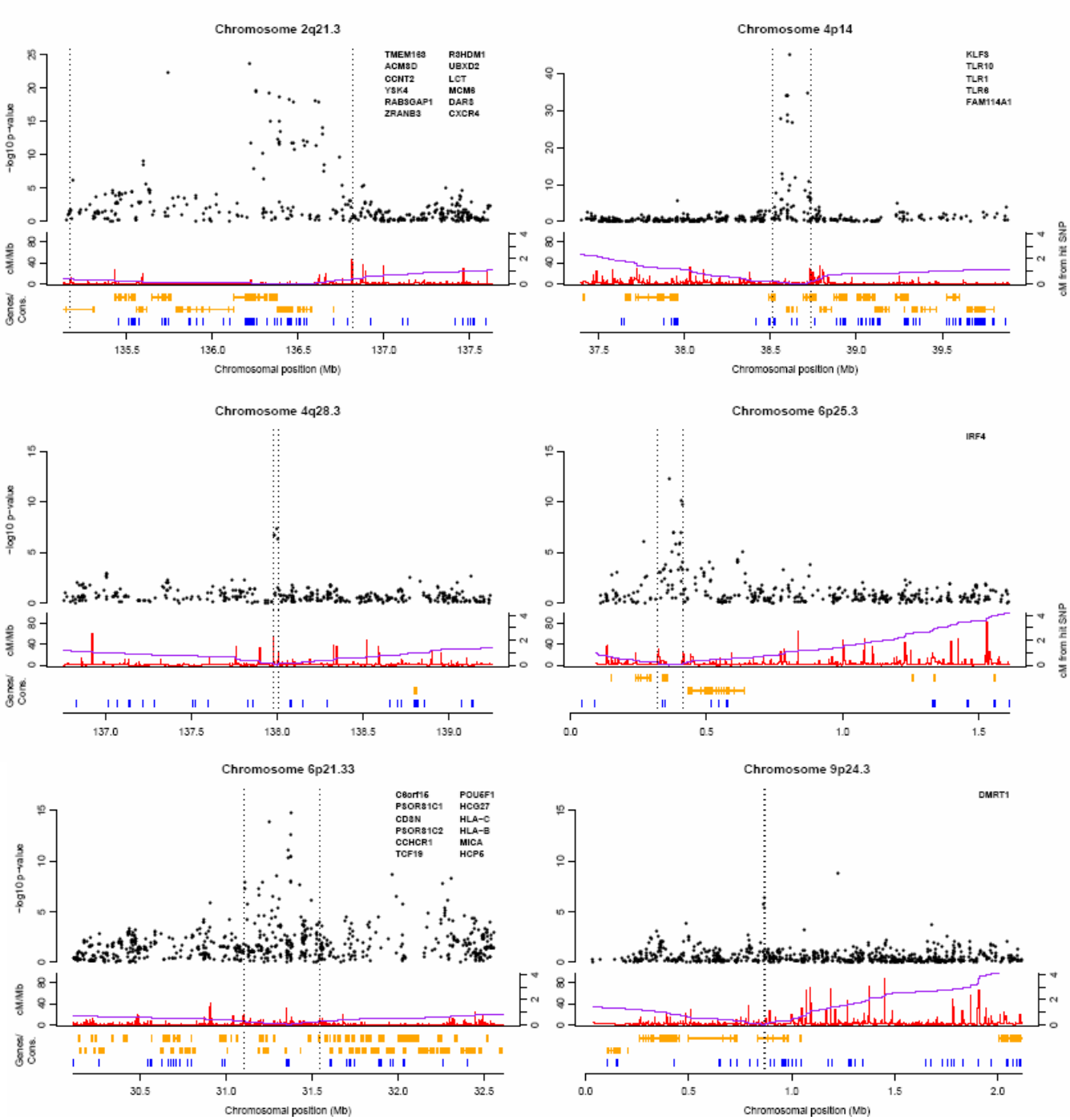
Supplementary Figure 3 | Comparison of patterns of LD in the HapMap and the WTCCC. Plots of linkage disequilibrium (r2) in a random 1Mb from each of the autosomes. The lower diagonal shows pairwise LD in the 58C collection and the upper diagonal shows the LD between the same SNPs typed in the CEU population in the HapMap sample. Red indicates high r2, strong correlation, through white and then blue for pairs of SNPs with low correlation. Note that difference in SNP density between the regions can alter the appearance. See Methods.

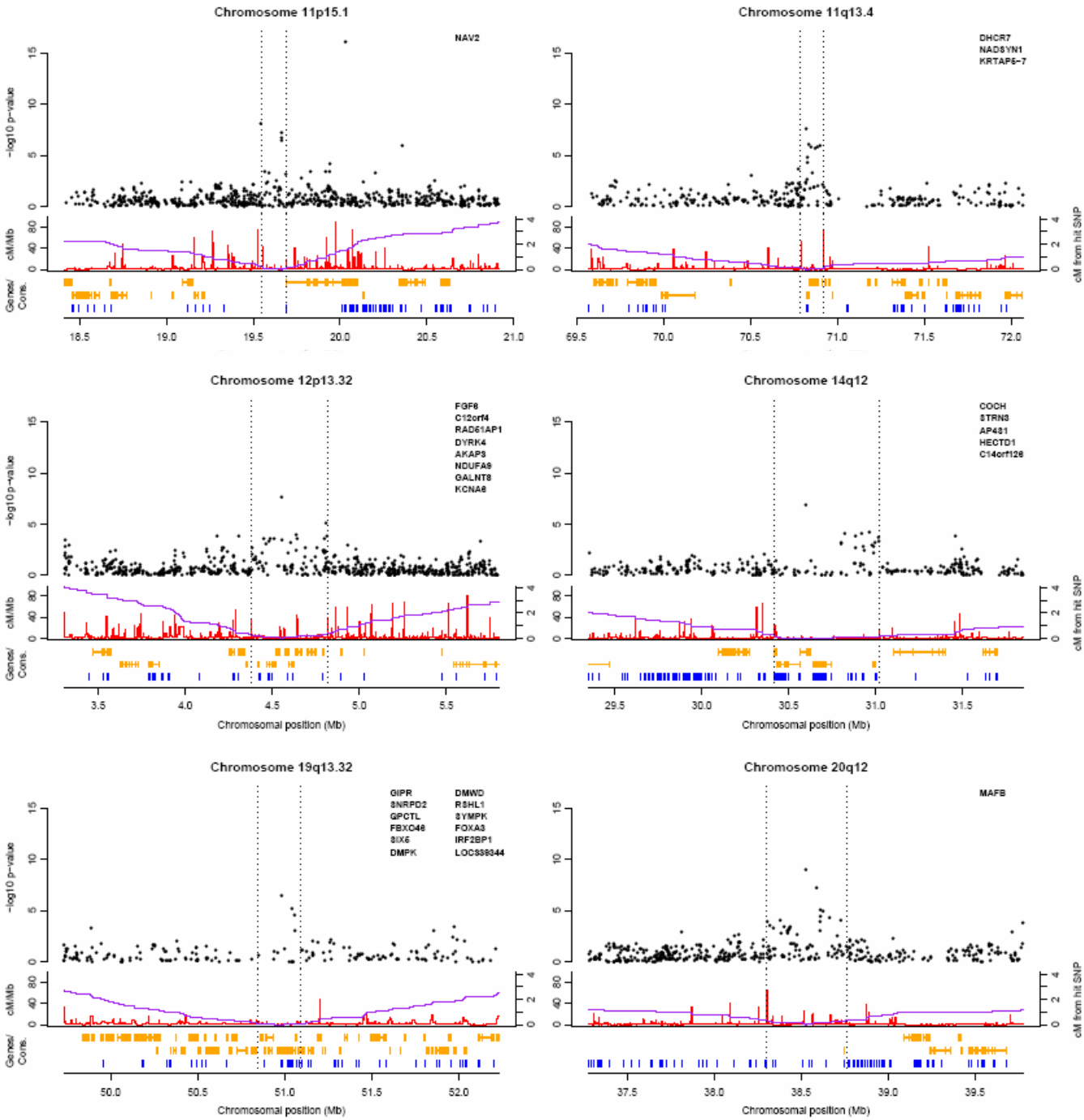


Supplementary Figure 4 | Geographical region definitions. Geographical regions are defined by postcode and their boundaries do not necessarily coincide with current administrative boundaries. For example, part of Wales falls within the SY postcode which is assigned to the "Midlands" region (however this is an area of low population density). The correspondence between postcode and region is shown in the relevant section of the Supplementary Information.

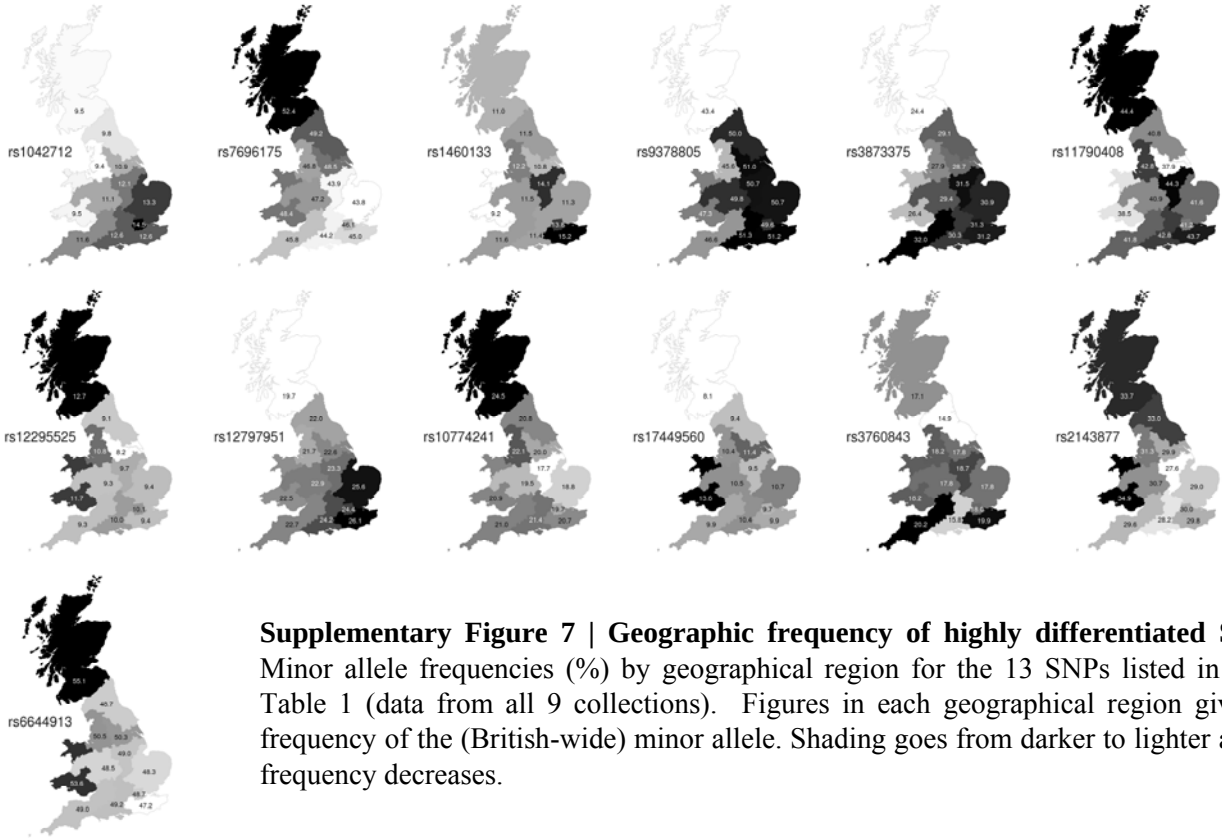


Supplementary Figure 5 | Multidimensional Scaling (MDS). WTCCC and HapMap samples plotted for the first two principal components obtained by multidimensional scaling of a matrix of pairwise IBS values between samples. Samples near the YRI cluster were subsequently identified in sample records as Afro-Caribbean; the large cluster one-third of the way between CEU and CHB+JPT were subsequently identified as South Asian (India/Pakistan). Samples showing evidence of non-European ancestry were excluded from further analyses (grey crosses).

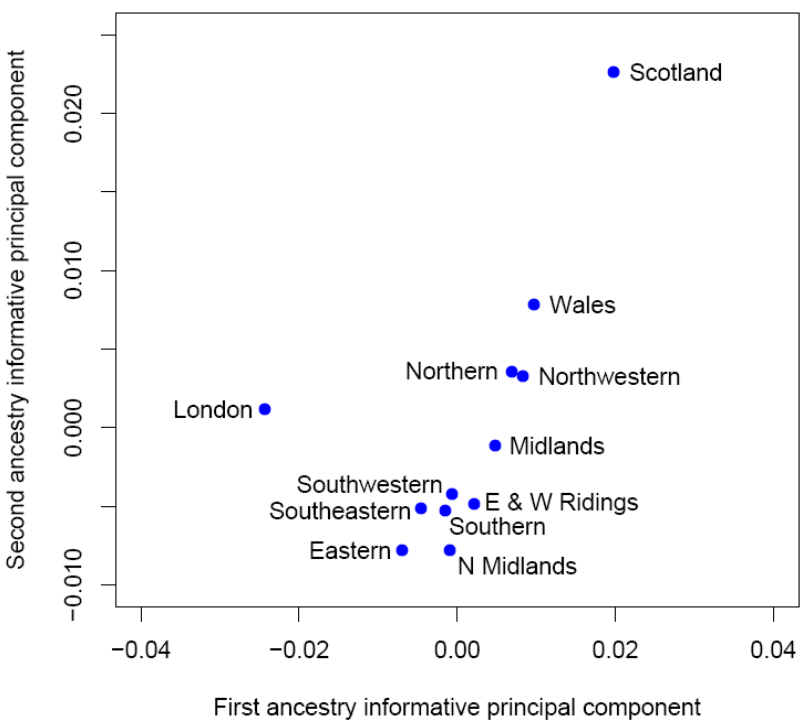




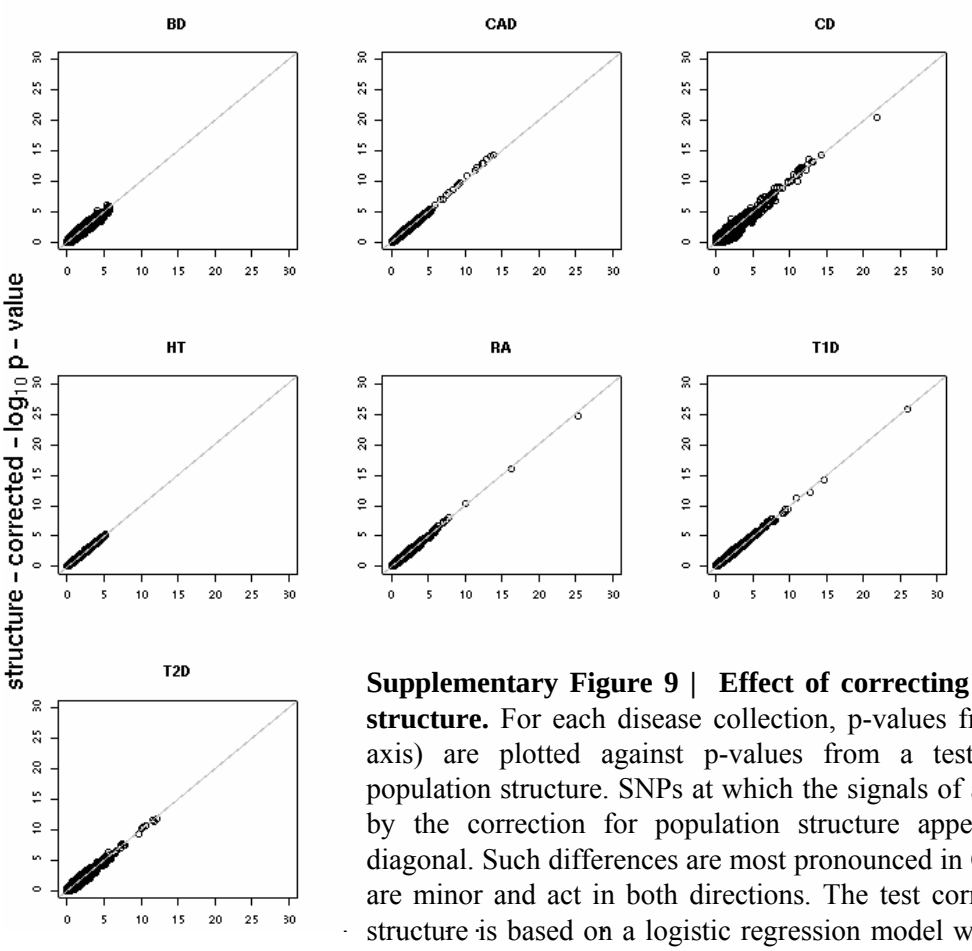
Supplementary Figure 6 | Signal plots for regions of strong geographic differentiation. Characteristics of genomic regions with strong evidence for geographic differentiation between WTCCC samples. Region boundaries (vertical dotted lines) were chosen to coincide with locations where test statistics returned to background levels and, where possible, recombination hotspots. Top panel: -log10 p-values for an 11 d.f. test for allele frequency differences between geographic regions. See Methods. Note that the y-axis scales used for regions 2q21 and 4p14 differ from each other and from the scale used in all other plots in this figure. Middle panel: Fine-scale recombination rate (cM/Mb). The purple line shows the cumulative genetic distance (in cM) from the hit SNP. Bottom panel: Known genes and sequence conservation in 17 vertebrates. Genes (orange) in the hit region are listed in the upper righthand part of each plot in chromosomal order, starting at the lefthand edge of the region. The top track shows plus-strand genes and the middle track shows minus-strand genes. Sequence conservation (bottom track) scores are based on the phylogenetic hidden Markov model phastCons. Highly conserved regions (phastCons score ≥600) are shown in blue. Further instruction on interpreting these plots can be found in the Supplementary Information.



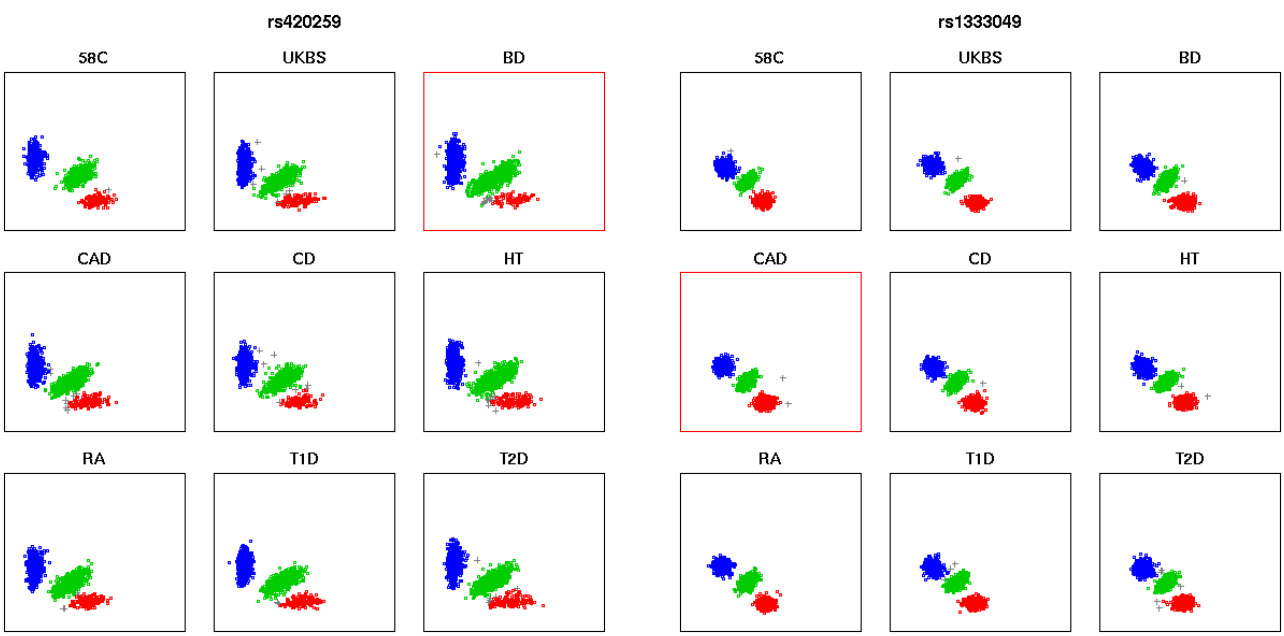
Supplementary Figure 7 | Geographic frequency of highly differentiated SNPs. Minor allele frequencies (%) by geographical region for the 13 SNPs listed in Main Table 1 (data from all 9 collections). Figures in each geographical region give the frequency of the (British-wide) minor allele. Shading goes from darker to lighter as this frequency decreases.

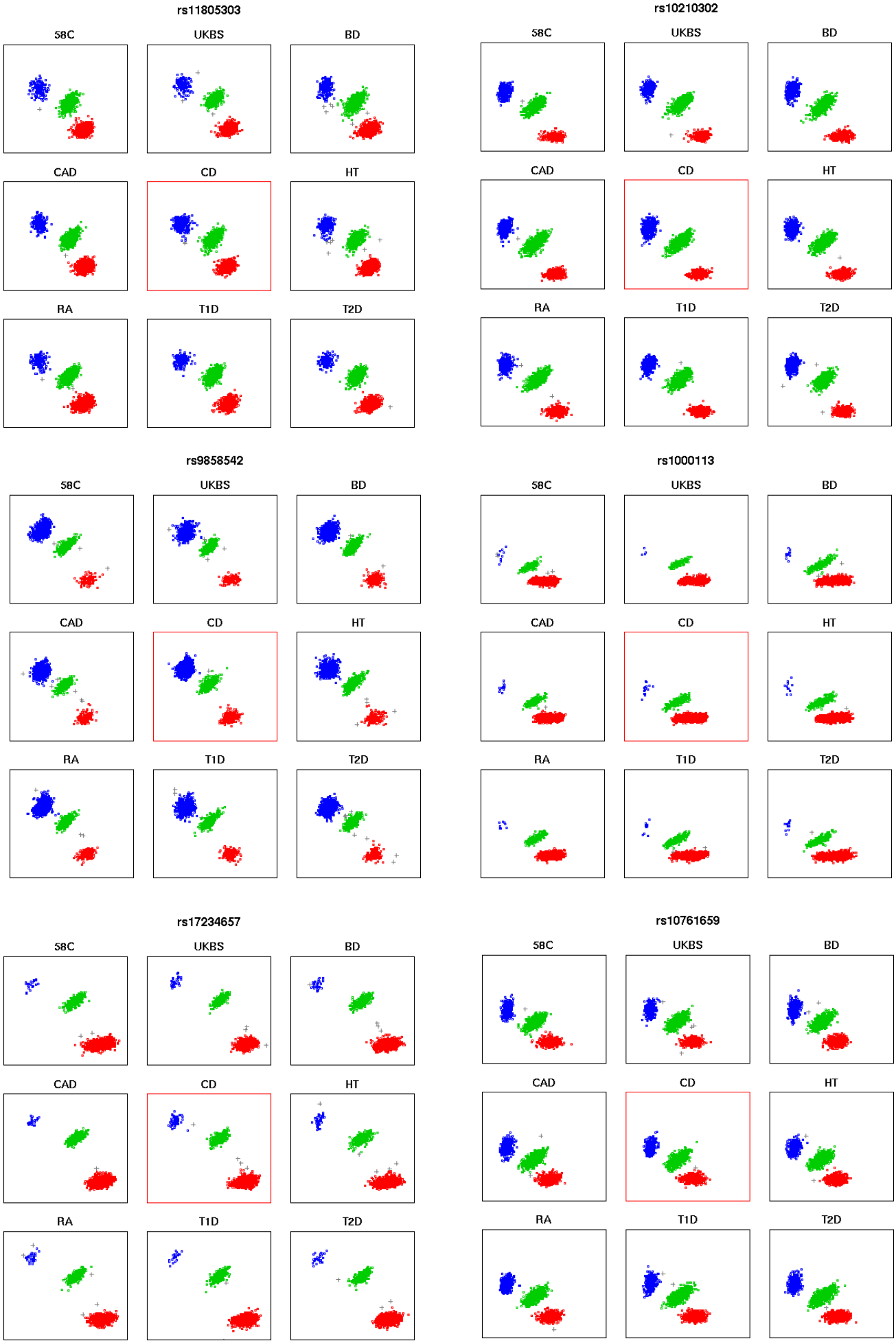


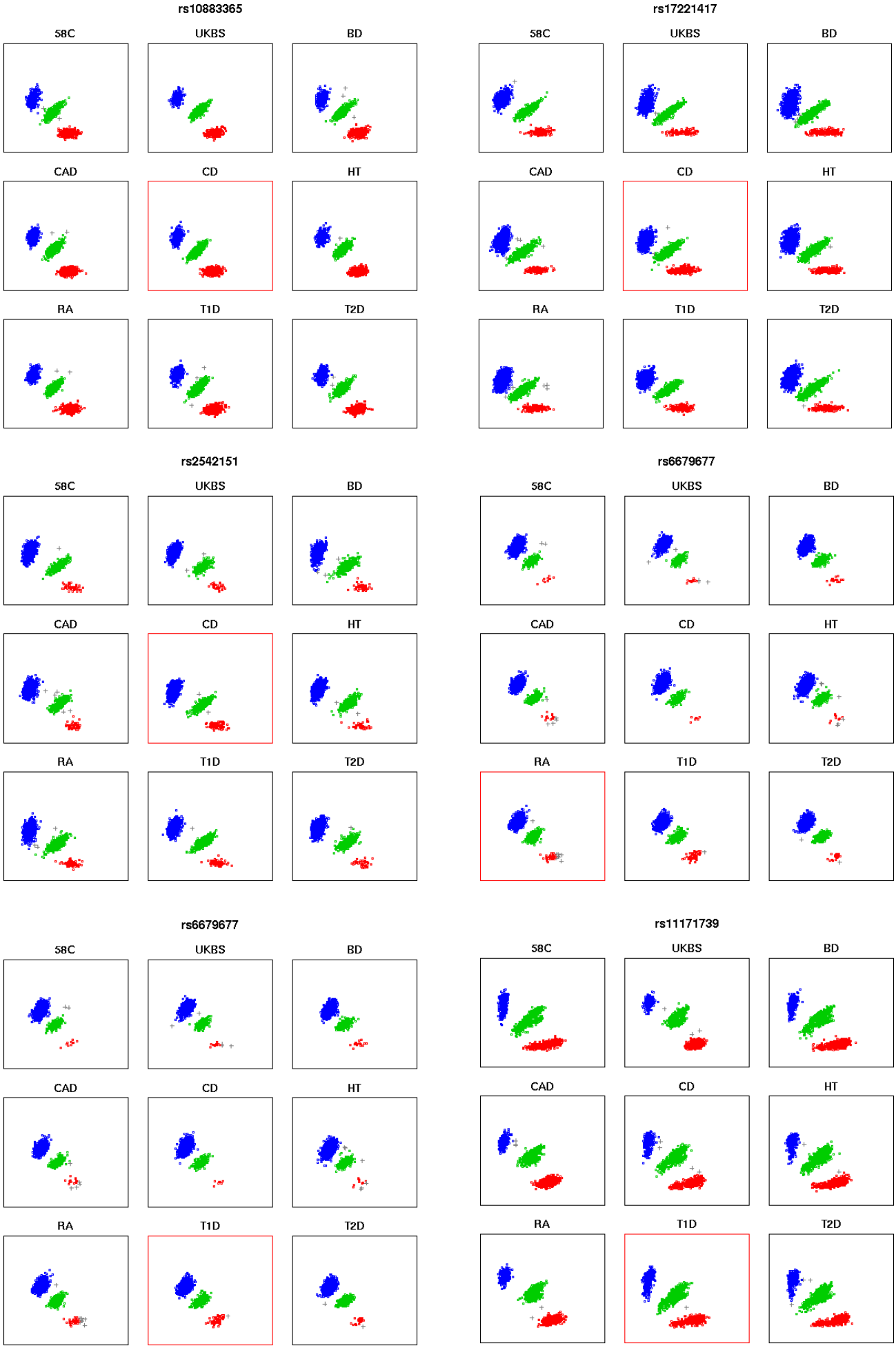
Supplementary Figure 8 | UK geographic population structure. Means, by geographical region, of the two principal components we judged to be informative of ancestral population admixture. Note that the means reflect the geographical configuration of the regions to some extent, confirming that the principal components are informative about geographical population structure. However, the distributions of principal component scores for individuals from each region overlap very extensively (data not shown), indicating that the population structure is very weak.

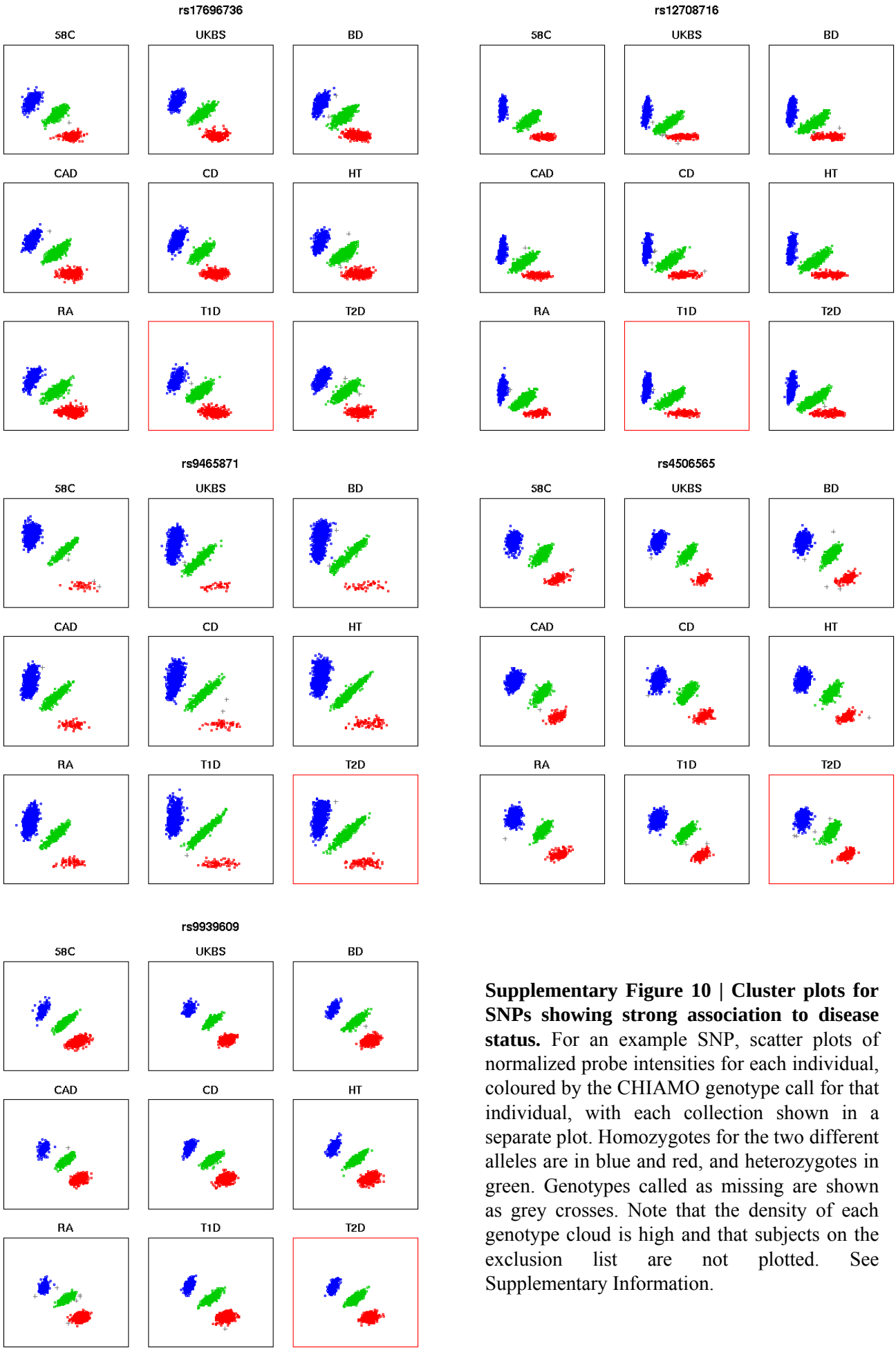


Supplementary Figure 9 | Effect of correcting tests for population structure. For each disease collection, p-values from the trend test (x-axis) are plotted against p-values from a test which corrects for population structure. SNPs at which the signals of association is affected by the correction for population structure appear as points off the diagonal. Such differences are most pronounced in CD, BD and T2D, but are minor and act in both directions. The test correcting for population structure is based on a logistic regression model which includes the two ancestry informative principal components as covariates (while the trend test is equivalent to a score test in a logistic regression model with no covariates).

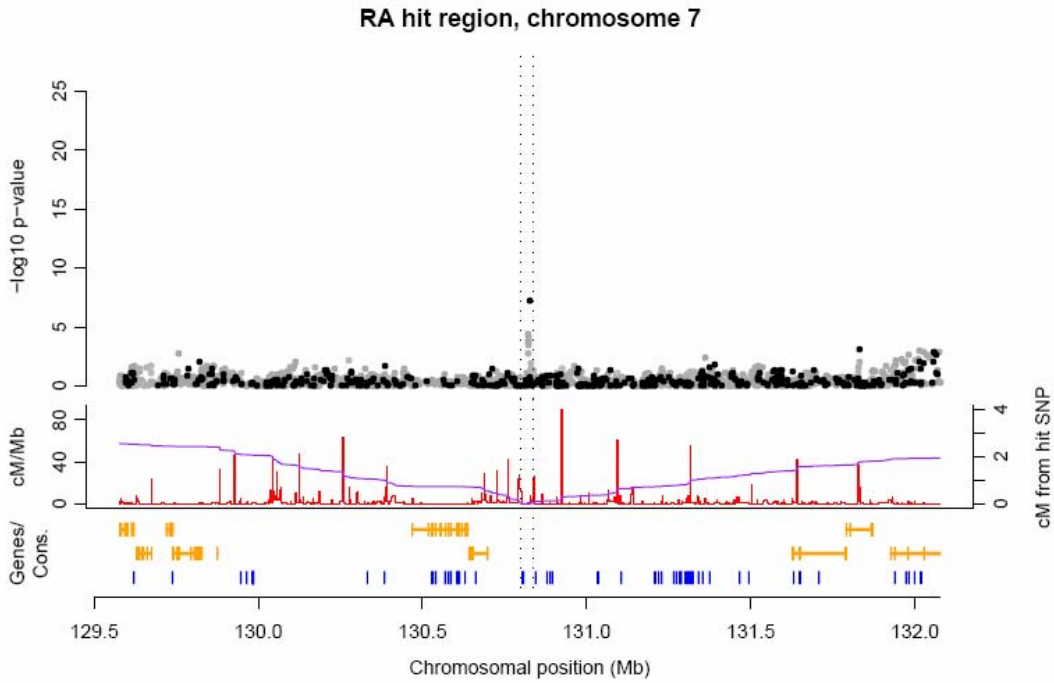



rs11805303
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs10210302
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs9858542
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs1000113
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs17234657
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs10761659
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D

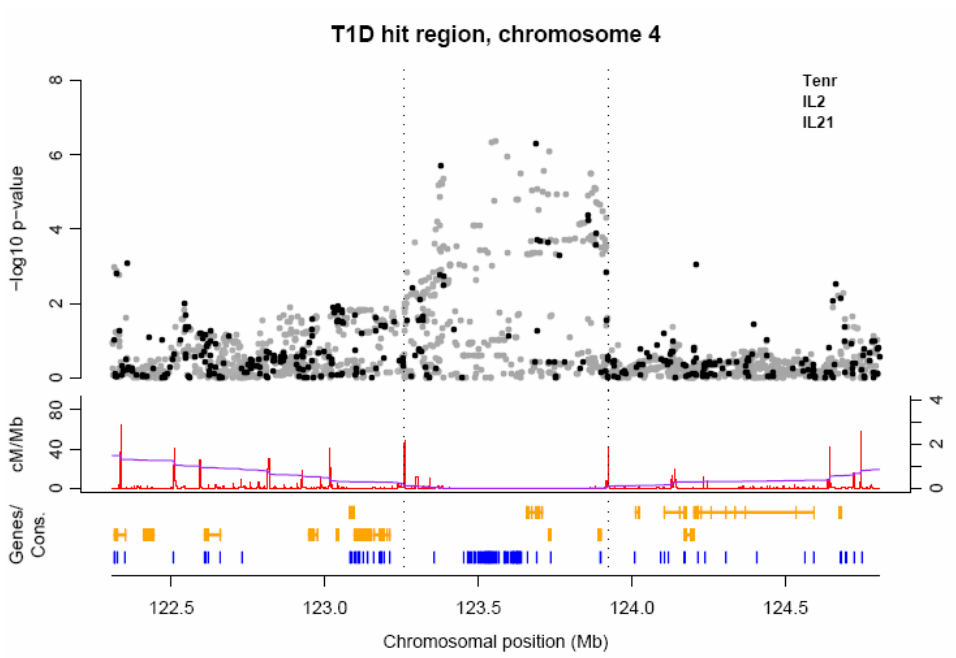

rs10883365
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs17221417
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs2542151
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs6679677
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs6679677
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D
rs11171739
58C
UKBS
BD
CAD
CD
HT
RA
T1D
T2D

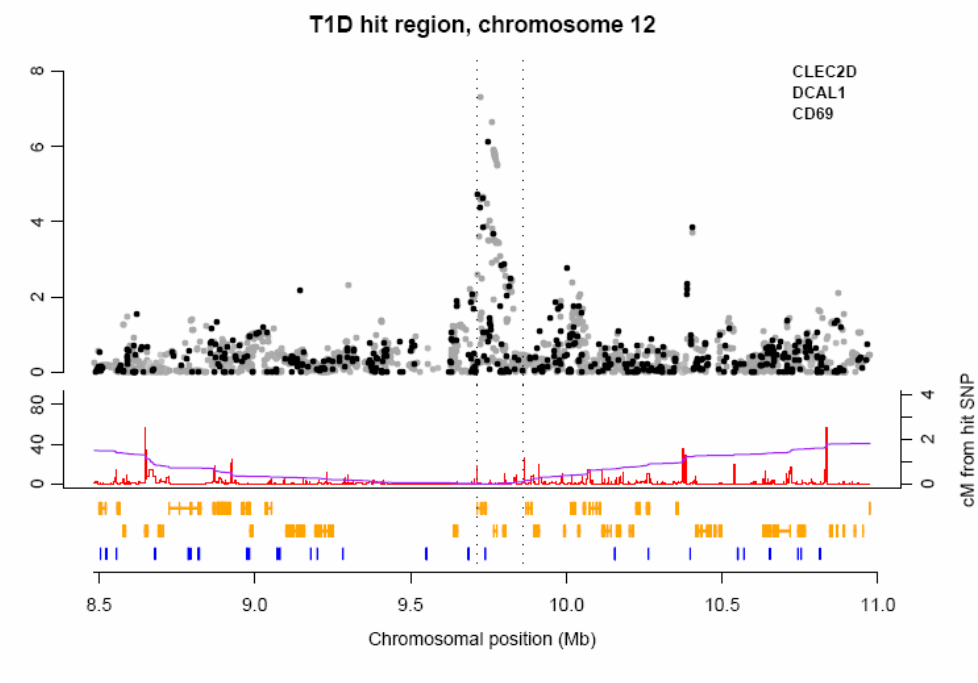


Supplementary Figure 10 | Cluster plots for SNPs showing strong association to disease status. For an example SNP, scatter plots of normalized probe intensities for each individual, coloured by the CHIAMO genotype call for that individual, with each collection shown in a separate plot. Homozygotes for the two different alleles are in blue and red, and heterozygotes in green. Genotypes called as missing are shown as grey crosses. Note that the density of each genotype cloud is high and that subjects on the exclusion list are not plotted. See Supplementary Information.

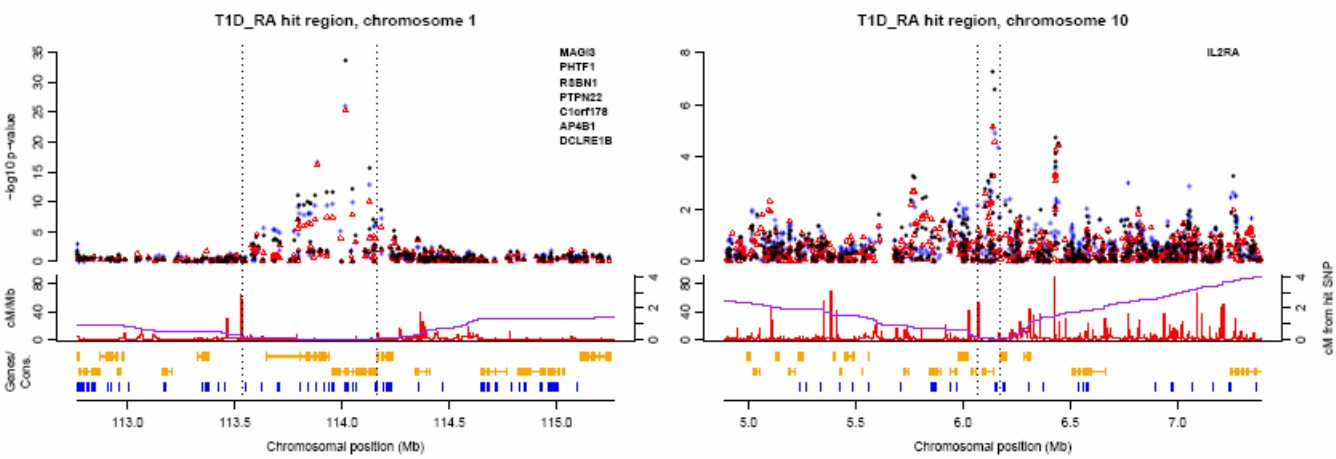


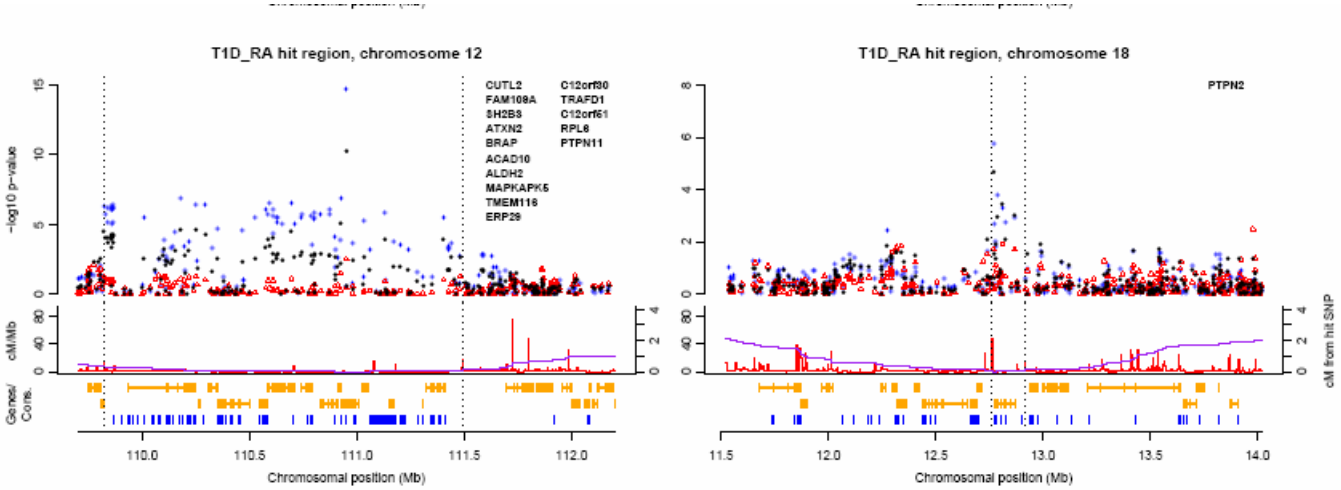
Supplementary Figure 11 | Signal plot for region identified by the sex-differentiated test. Characteristics of genomic region surrounding the hit SNP identified by the sex-differentiated test. Since further analysis suggested that the association was considerably stronger in females, this figure displays the results of an analysis which uses only females. Region boundaries (vertical dotted lines) were chosen to coincide with the recombination hotspots immediately flanking the hit SNP. Top panel: -log10 p-values for the (1 d.f.) trend test calculated using only females. Black points represent SNPs typed in the study, and grey points represent SNPs whose genotypes were imputed. SNPs imputed with higher confidence are shown in darker grey. Middle panel: Fine-scale recombination rate (cM/Mb). The purple line shows the cumulative genetic distance (in cM) from the hit SNP. Bottom panel: Known genes and sequence conservation in 17 vertebrates. No genes (orange) lie in the hit region, so none is listed on the plot. The top track shows plus-strand genes and the middle track shows minus-strand genes. Sequence conservation (bottom track) scores are based on the phylogenetic hidden Markov model phastCons . Highly conserved regions (phastCons score ≥600) are shown in blue. Further instruction on interpreting these plots can be found in the Supplementary Information.



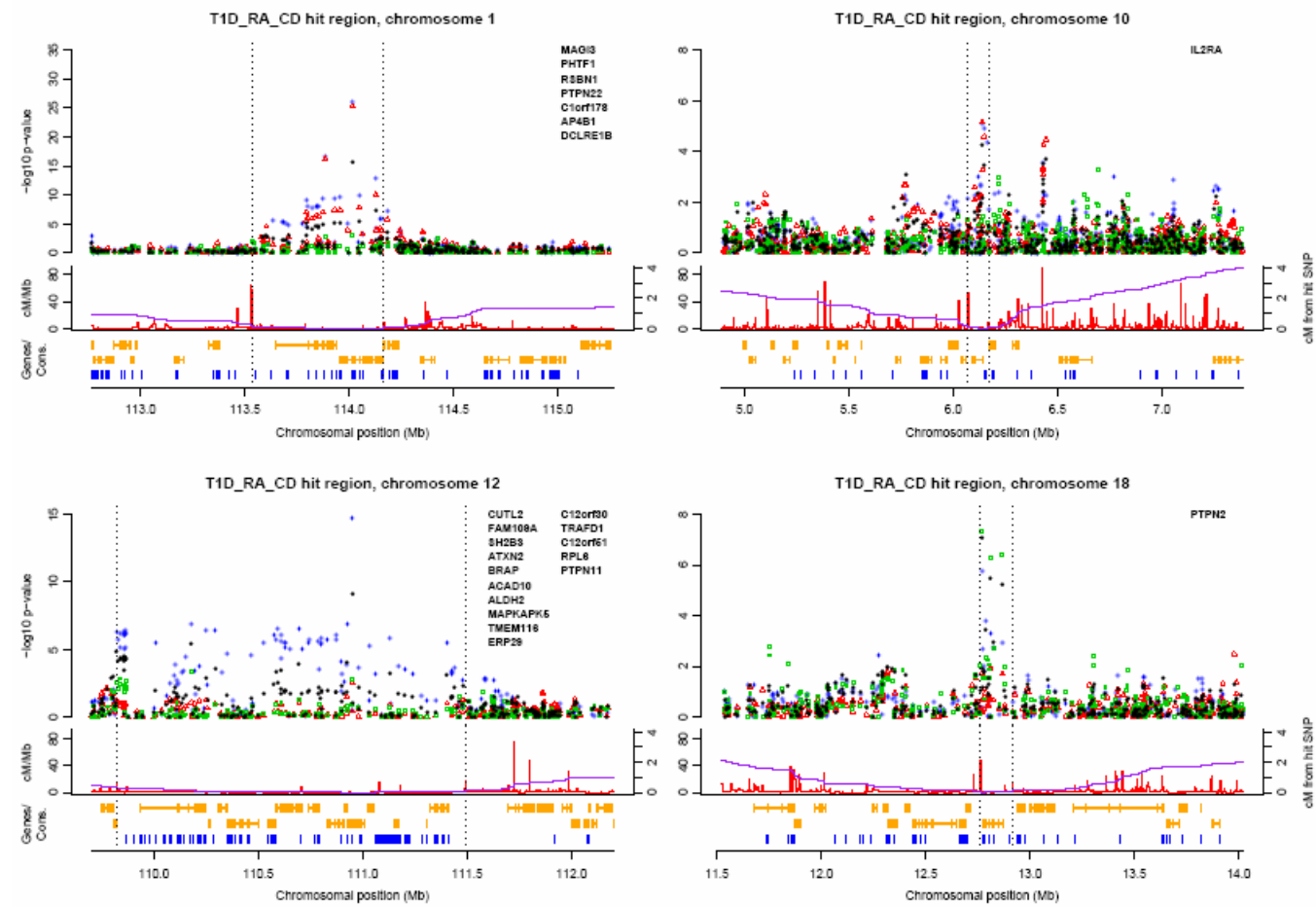


Supplementary Figure 12 | Signal plots for regions identified through imputation. Characteristics of genomic regions that reached a p-value threshold of 5x10-7 only at imputed SNPs. Region boundaries (vertical dotted lines) were chosen to coincide with locations where test statistics returned to background levels and, where possible, recombination hotspots. Top panel: -log10 p-values for the frequentist test with the smallest p-value at the imputed hit SNP. Black points represent SNPs typed in the study, and grey points represent SNPs whose genotypes were imputed. SNPs imputed with higher confidence are shown in darker grey. Middle panel: Fine-scale recombination rate (cM/Mb). The purple line shows the cumulative genetic distance (in cM) from the hit SNP. Bottom panel: Known genes and sequence conservation in 17 vertebrates. Genes (orange) in the hit region are listed in the upper righthand part of each plot in chromosomal order, starting at the lefthand edge of the region. The top track shows plus-strand genes and the middle track shows minus-strand genes. Sequence conservation (bottom track) scores are based on the phylogenetic hidden Markov model phastCons . Highly conserved regions (phastCons score ≥600) are shown in blue. Further instruction on interpreting these plots can be found in the Supplementary Information.

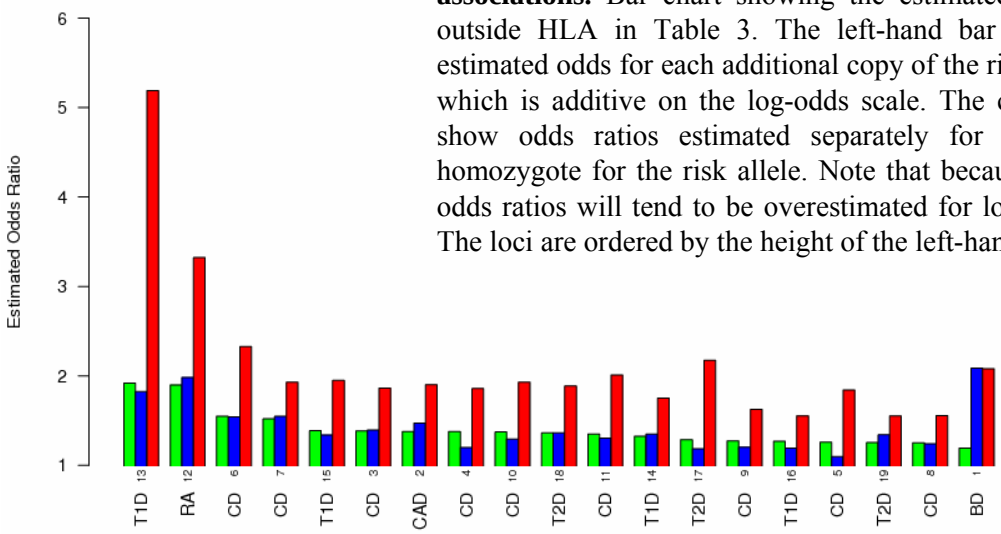




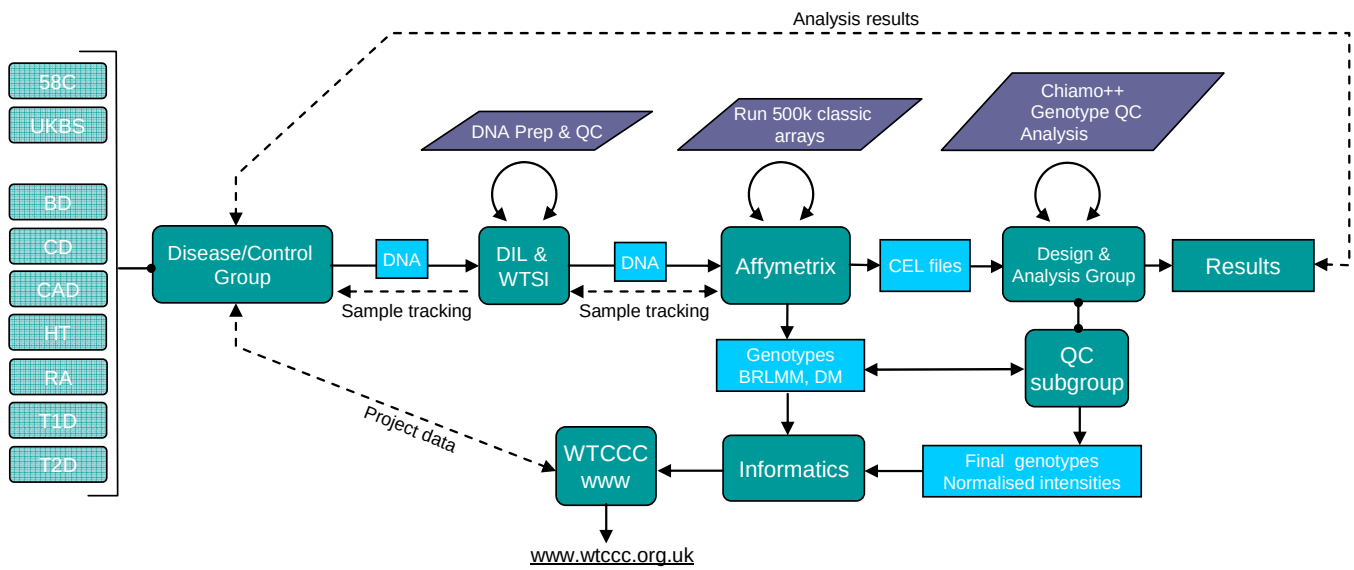
Supplementary Figure 13 | Signal plots for regions identified by pooling T1D and RA cases. Characteristics of genomic regions surrounding hit SNPs found by combining case groups. Region boundaries (vertical dotted lines) were chosen to coincide with locations where test statistics returned to background levels and, where possible, recombination hotspots. Top panel: -log10 p-values for the trend test in pooled cases (black dots), T1D cases (blue crosses), and RA cases (red triangles). Middle panel: Fine-scale recombination rate (cM/Mb). The purple line shows the cumulative genetic distance (in cM) from the hit SNP. Bottom panel: Known genes and sequence conservation in 17 vertebrates. Genes (orange) in the hit region are listed in the upper righthand part of each plot in chromosomal order, starting at the lefthand edge of the region. The top track shows plus-strand genes and the middle track shows minus-strand genes. Sequence conservation (bottom track) scores are based on the phylogenetic hidden Markov model phastCons . Highly conserved regions (phastCons score ≥600) are shown in blue. Further instruction on interpreting these plots can be found in the Supplementary Information.



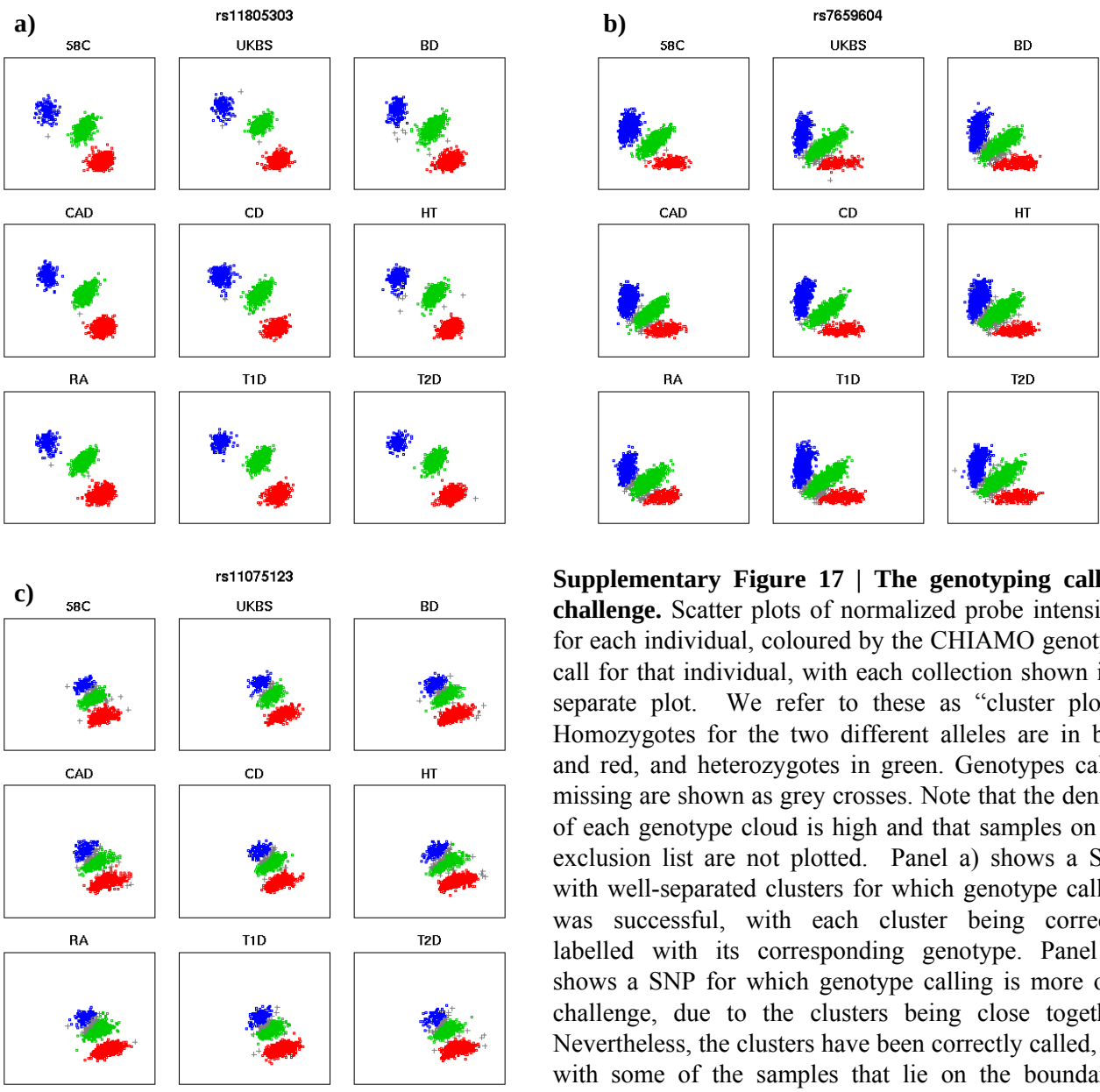
Supplementary Figure 14 | Signal plots for regions identified by pooling T1D, RA, and CD cases. Characteristics of genomic regions surrounding hit SNPs identified by combining case groups. Region boundaries (vertical dotted lines) were chosen to coincide with locations where test statistics returned to background levels and, where possible, recombination hotspots. Top panel: -log10 p-values for the trend test in pooled cases (black dots), T1D cases (blue crosses), RA cases (red triangles), and CD cases (green squares). Middle panel: Fine-scale recombination rate (cM/Mb). The purple line shows the cumulative genetic distance (in cM) from the hit SNP. Bottom panel: Known genes and sequence conservation in 17 vertebrates. Genes (orange) in the hit region are listed in the upper righthand part of each plot in chromosomal order, starting at the lefthand edge of the region. The top track shows plus-strand genes and the middle track shows minus-strand genes. Sequence conservation (bottom track) scores are based on the phylogenetic hidden Markov model phastCons . Highly conserved regions (phastCons score ≥600) are shown in blue. Further instruction on interpreting these plots can be found in the Supplementary Information.



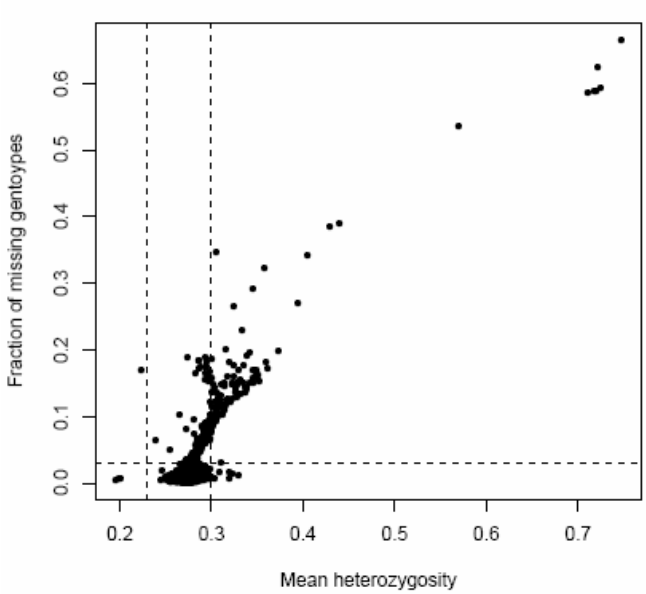
Supplementary Figure 15 | Estimated odds ratios for strong associations. Bar chart showing the estimated odds ratio for the SNPs outside HLA in Table 3. The left-hand bar for each SNP shows the estimated odds for each additional copy of the risk allele in a disease model which is additive on the log-odds scale. The central and right-hand bars show odds ratios estimated separately for the heterozygote and the homozygote for the risk allele. Note that because of the "winner's curse" odds ratios will tend to be overestimated for loci discovered in our study. The loci are ordered by the height of the left-hand bar.



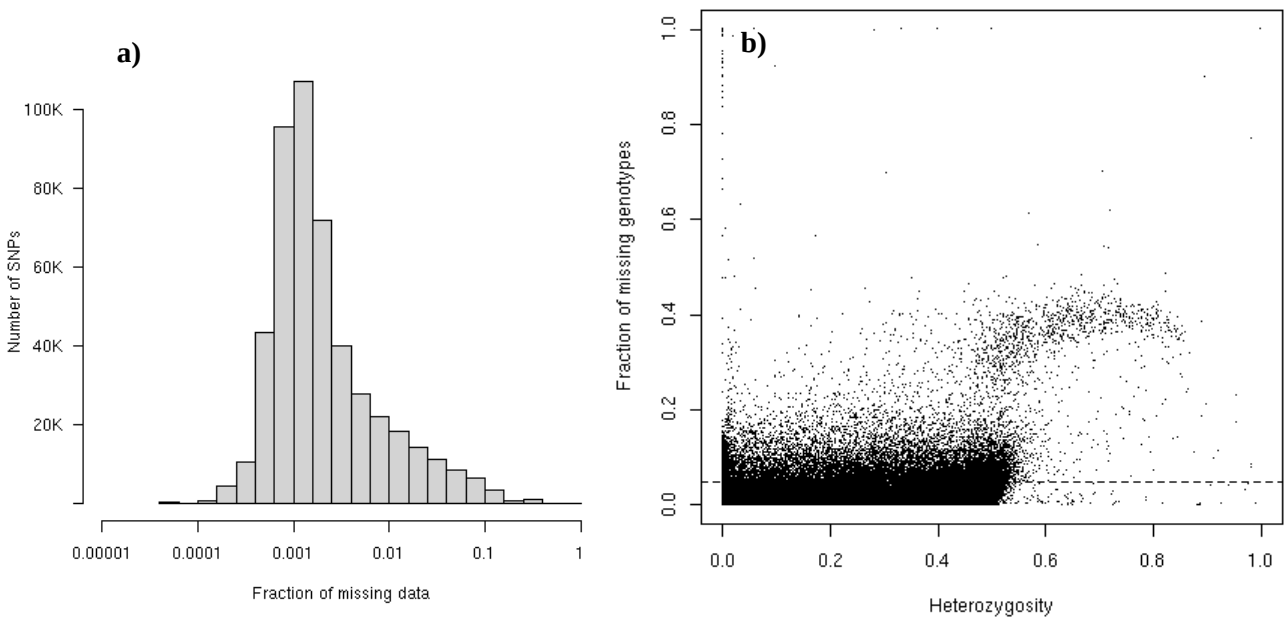
Supplementary Figure 16 | Workflow for the WTCCC. Initial DNA processing, QC and tracking steps were undertaken at the Wellcome Trust/JDRF Diabetes and Inflammation Laboratory (DIL) and the Wellcome Trust Sanger Institute (WTSI) in the United Kingdom. Re-arrayed plates of DNA were shipped to Affymetrix in the United States for outsourced genotyping. Raw intensity information from the Affymetrix platform was shipped back to the UK for genotype calling (CHIAMO) by the WTCCC Design and Analysis Group (DAG). The size and volume of this information exceeded the capacity for transfer over the internet and necessitated shipping of physical hard disks. The BRLMM and DM genotype calls were not used by the WTCCC, but are shown to emphasize their availability for external access via the Data Access Committee.



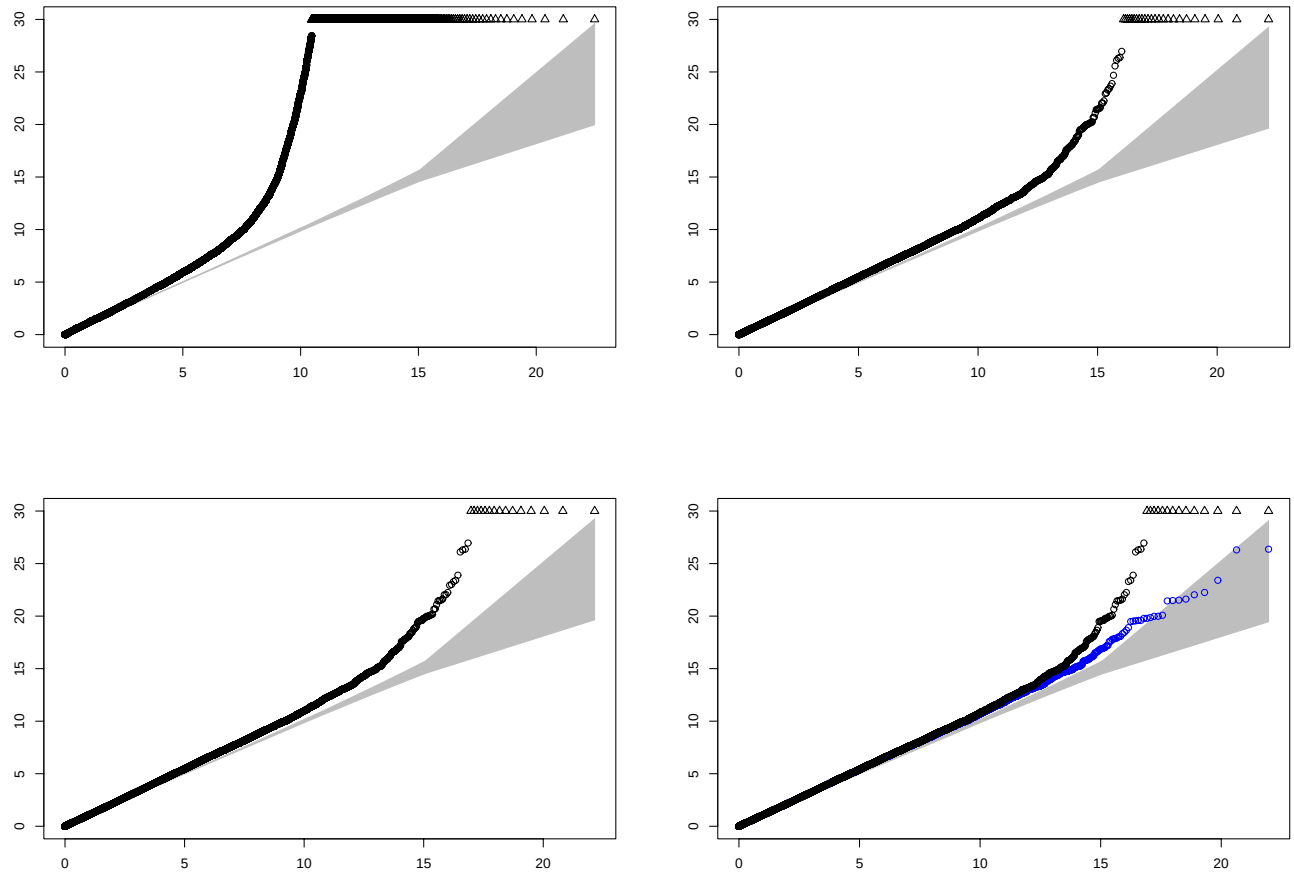
Supplementary Figure 17 | The genotyping calling challenge. Scatter plots of normalized probe intensities for each individual, coloured by the CHIAMO genotype call for that individual, with each collection shown in a separate plot. We refer to these as "cluster plots". Homozygotes for the two different alleles are in blue and red, and heterozygotes in green. Genotypes called missing are shown as grey crosses. Note that the density of each genotype cloud is high and that samples on the exclusion list are not plotted. Panel a) shows a SNP with well-separated clusters for which genotype calling was successful, with each cluster being correctly labelled with its corresponding genotype. Panel b) shows a SNP for which genotype calling is more of a challenge, due to the clusters being close together. Nevertheless, the clusters have been correctly called, but with some of the samples that lie on the boundaries between clusters being called as missing (they are shown as grey crosses). Panel c) shows a SNP for which genotype calling is problematic, due to two clusters being very close together but separated from the third cluster. While the clusters have been correctly called, many samples with signals lying between the two adjacent clusters are called as missing, causing bias in allele frequency estimates.



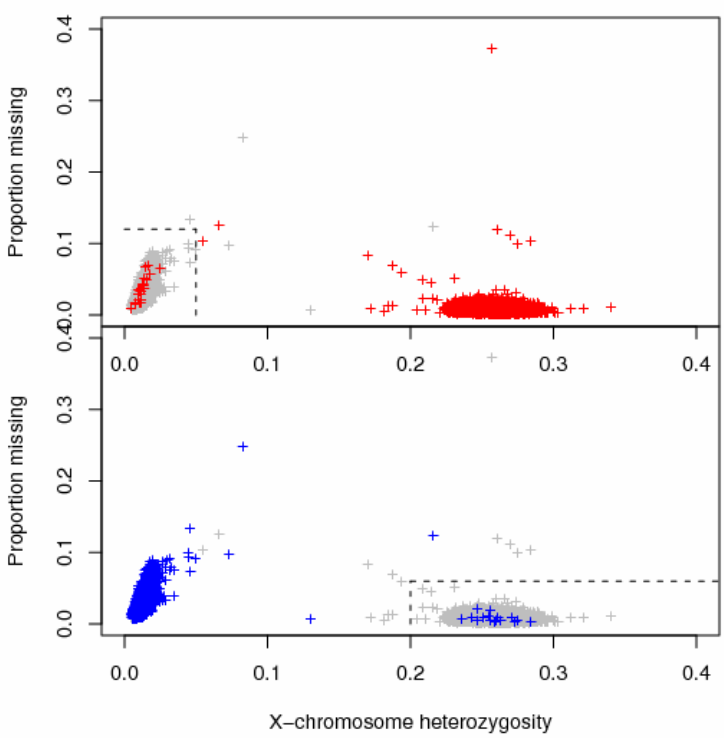
Supplementary Figure 18 | Individual missing data and heterozygosity. Scatter plot of the proportion of SNPs called heterozygote (x-axis) against the proportion called missing at a posterior probability threshold of 0.9 (y-axis) for each individual in the study. Dotted lines delimit the threshold used for exclusion of individuals from further analysis.



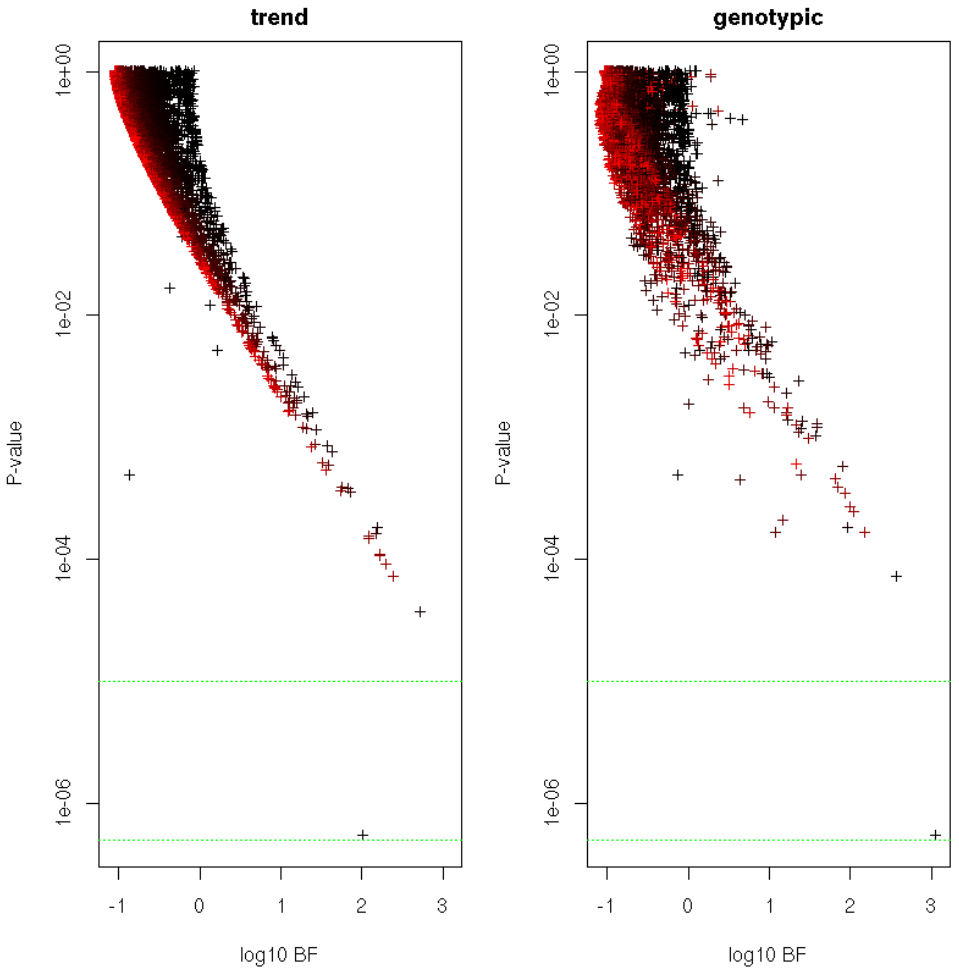
Supplementary Figure 19 | Missing data and heterozygosity per SNP. a) Histogram of proportion of individuals called missing for each SNP (i.e with posterior probability < 0.9) b) Scatter plot of the proportion of individuals called heterozygote (x-axis) against the proportion called missing at a posterior probability threshold of 0.9 (y-axis) for each SNP assayed. The dotted line shows the threshold over which a SNP was excluded from further analyses.



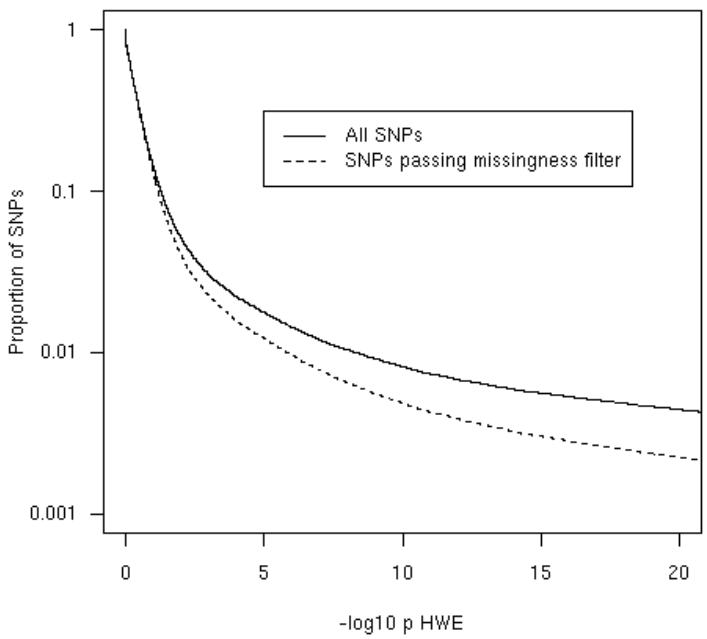
Supplementary Figure 20 | Quantile-quantile plots at four different stages of filtering. Each panel shows a QQ plot for the trend test results in T2D for the following subsets of SNPs (λ estimates for each subset in parentheses – see Methods for details), observed test statistics (y-axis) >30 are shown as triangles: top left, all SNPs (λ = 1.17); top right, SNPs passing standard project filter described in text and having minor allele frequency > 1% (λ = 1.09); bottom left, those passing the previous filter but excluding SNPs for which visual inspection of cluster plots revealed poor genotype calls (λ = 1.09); bottom right is as bottom left, but the plot which excludes regions with strong evidence for association is superimposed in blue, as in Main Figure 3.



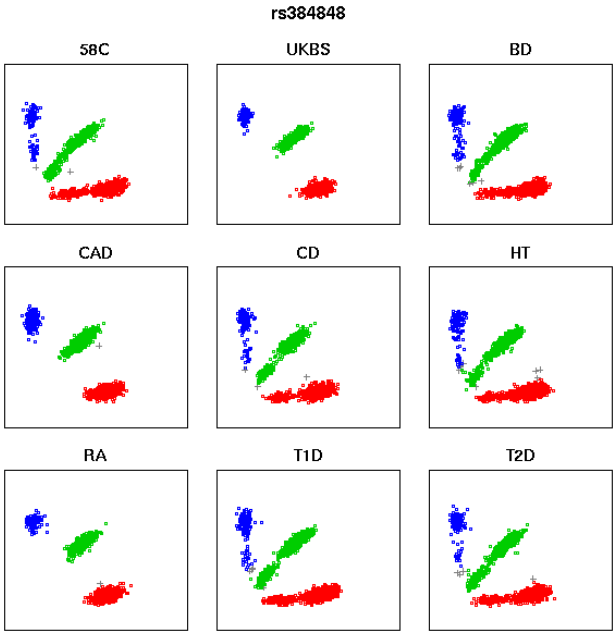
Supplementary Figure 21 | Individual missing data and heterozygosity on the X chromosome. Scatter plot of the proportion of SNPs called heterozygote (x-axis) against the proportion called missing (y-axis) for each individual in the study. For each collection the individuals are plotted twice; samples whose gender were initially reported as male are coloured blue and those reported as female are coloured red.



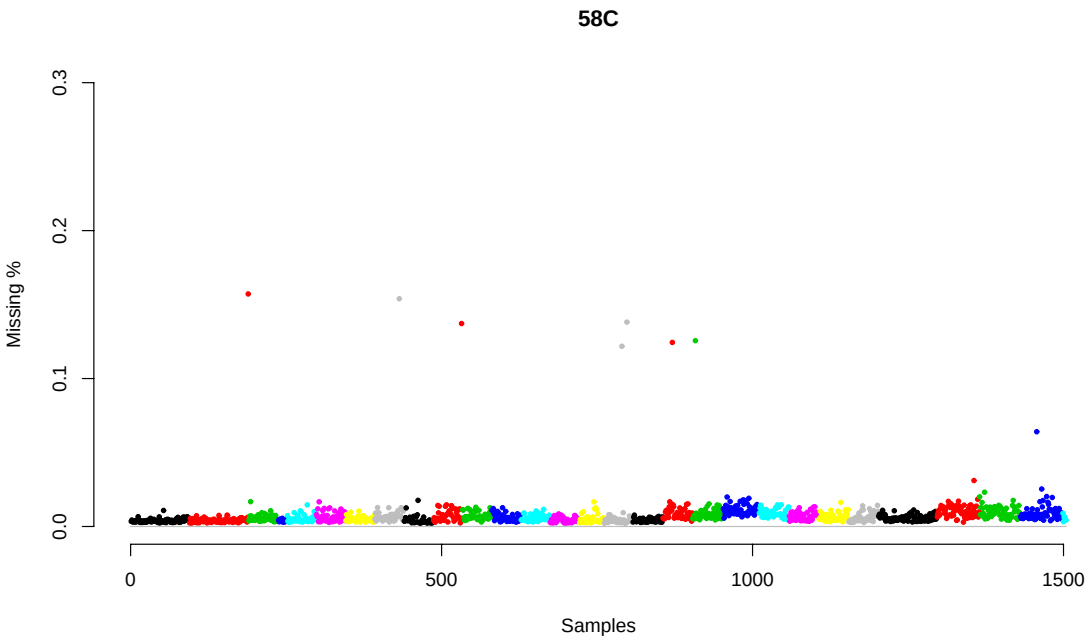
Supplementary Figure 22 | Comparison of p-values and Bayes factors. Scatter plot showing p-values and Bayes factors for all SNPs on chromosome 22 for BD. The points are coloured according to the minor allele frequency (MAF), ranging from fully black for a MAF of 0 to fully red for a MAF of 0.5. Two panels are shown, one for the trend test and the other for the genotypic test. The horizontal lines denote threshold values for inclusion on various tables.

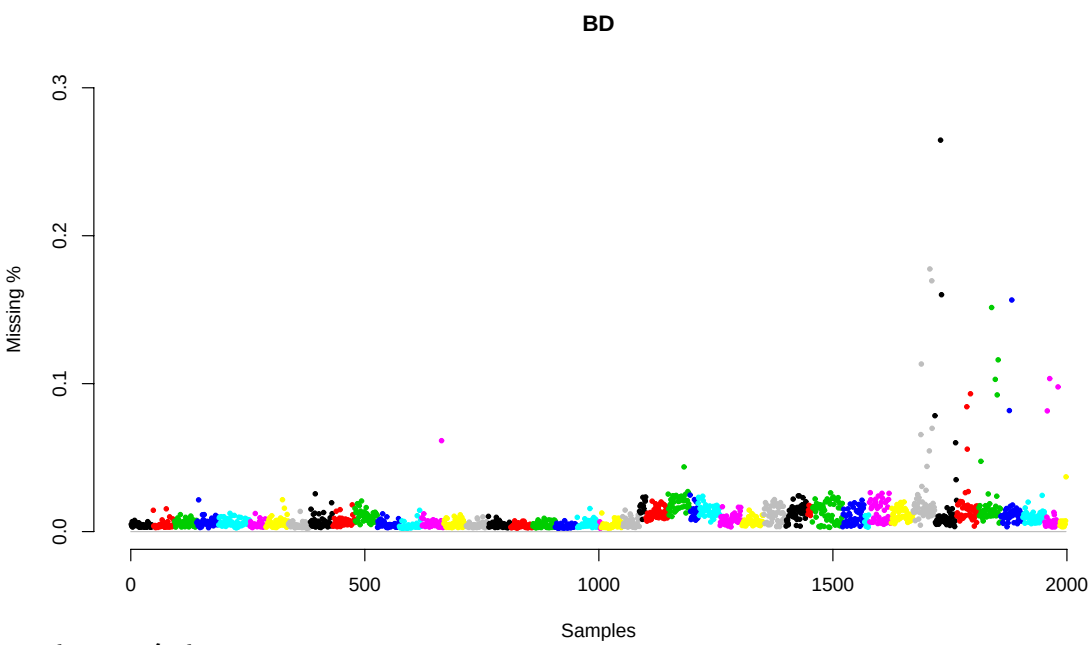


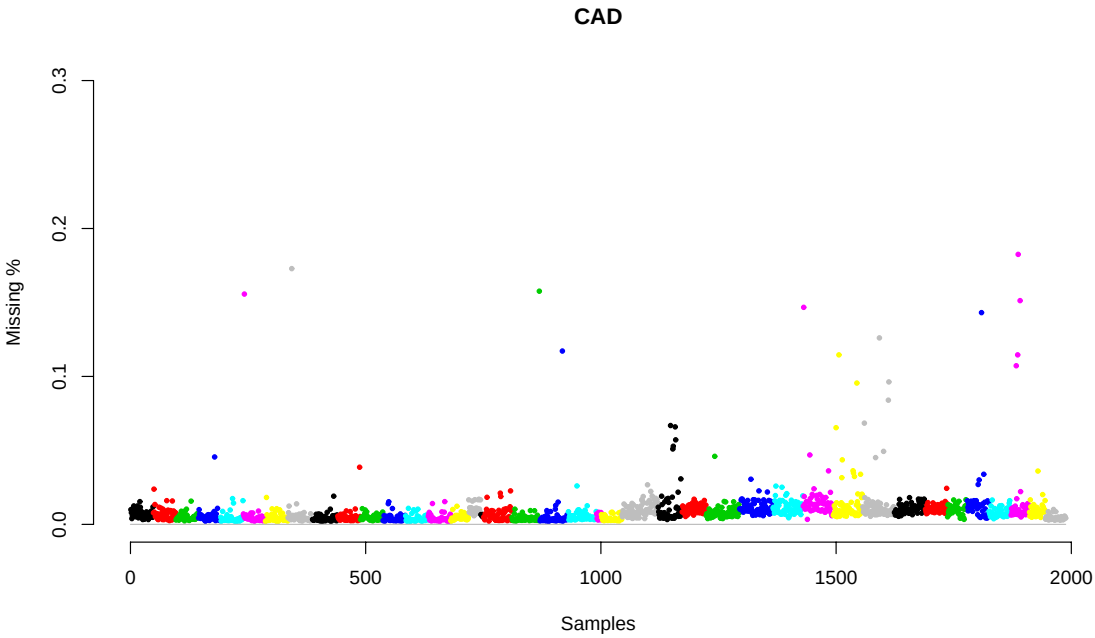
Supplementary Figure 23 | Proportion of SNPs failing test for HWE in controls. Proportion of SNPs with HWE p-values (in controls) below a given threshold. The solid line corresponds to all SNPs, while the dashed line corresponds to the SNPs which pass the call rate filter. SNPs which fail the call rate filter account for a substantial fraction of the extreme HWE p-values.

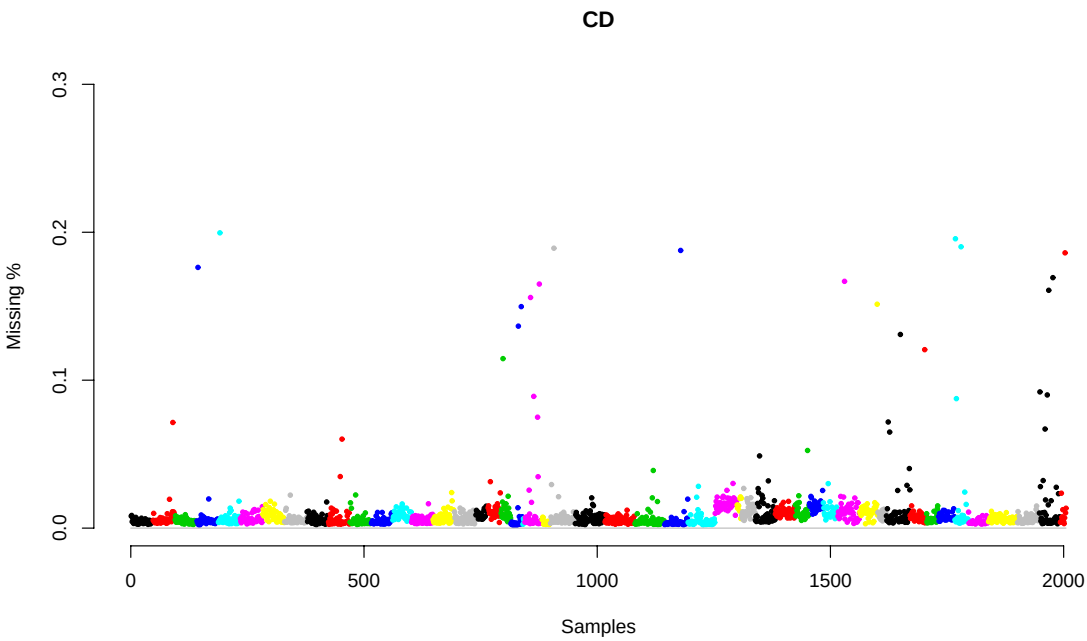


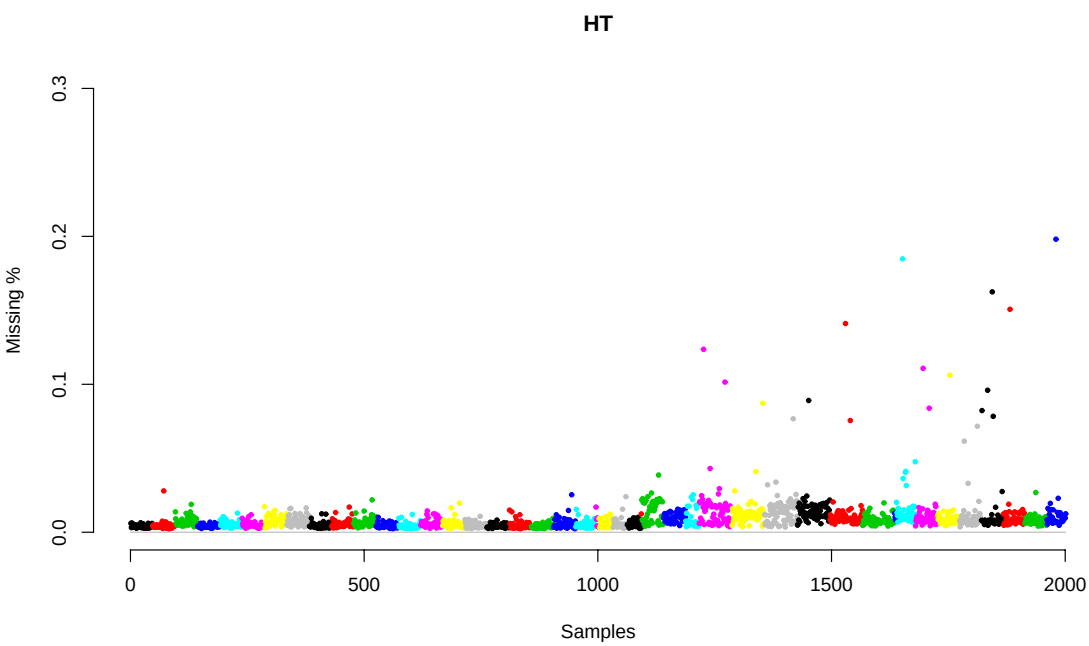
Supplementary Figure 24 | Radial shift. This cluster plot (see Supplementary Figure 19) shows the shift towards the origin that occurs in a subset of samples in a small number of SNPs. The shift did not occur in samples from the UKBS, CAD and RA collections. These collections differ from the others in that no samples were typed in the early stages of the study.

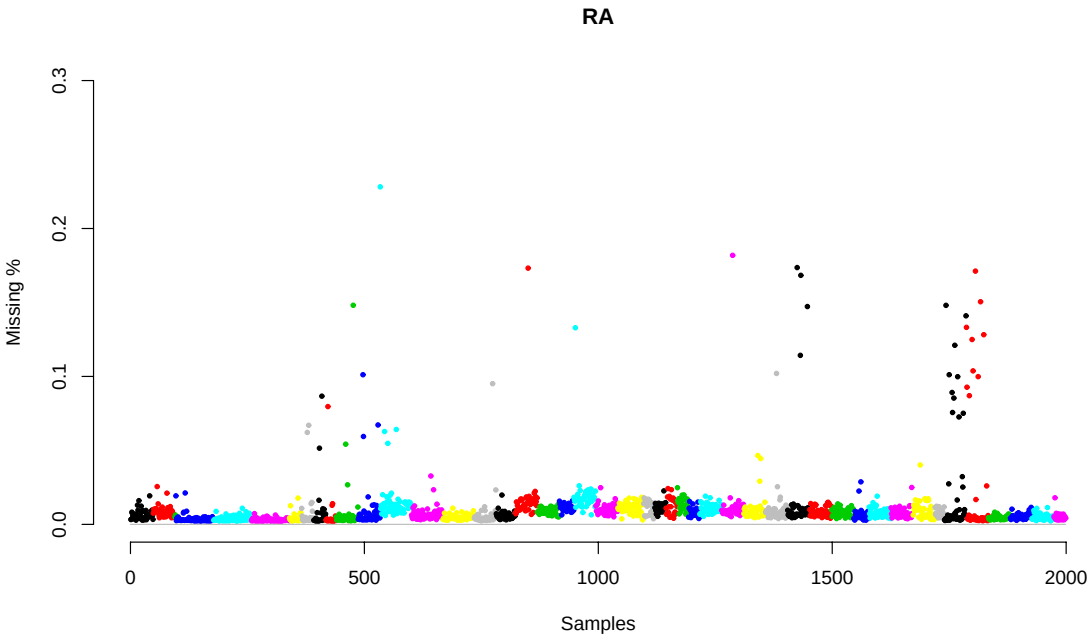


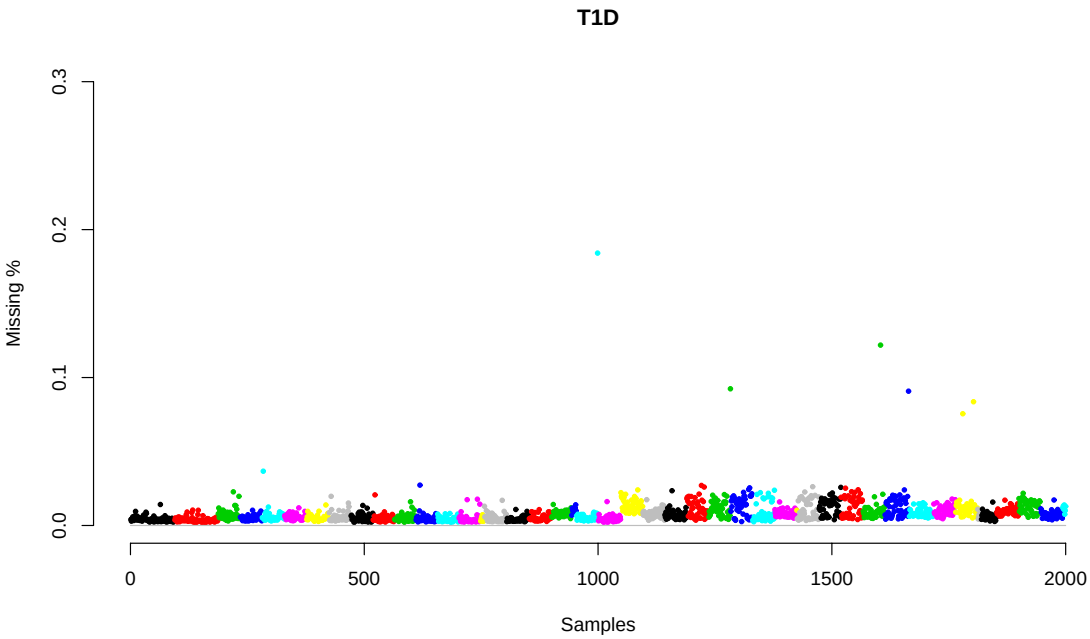


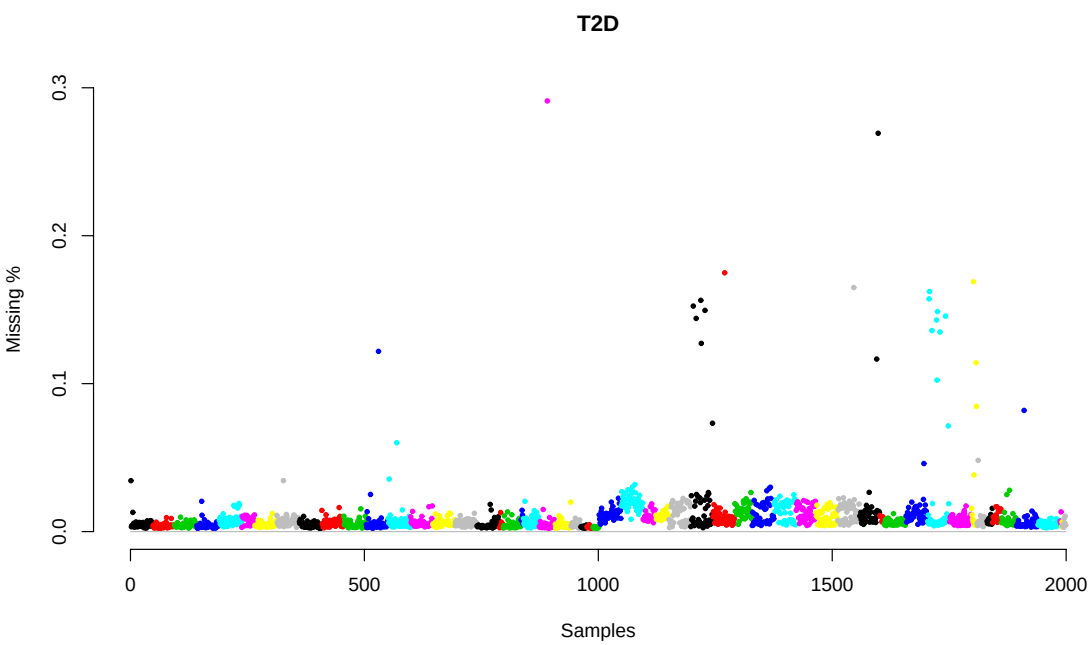

CAD
Missing %
0.0
0.1
0.2
0.3
0
500
1000
1500
2000
Samples

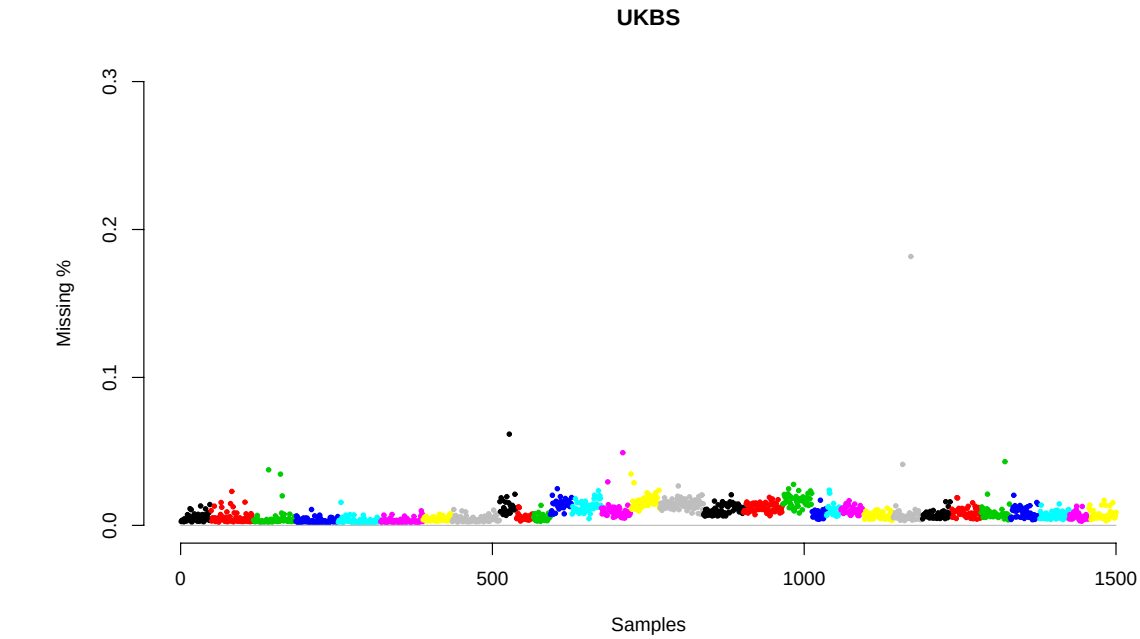

CD
Missing %
0.0
0.1
0.2
0.3
0
500
1000
1500
2000
Samples

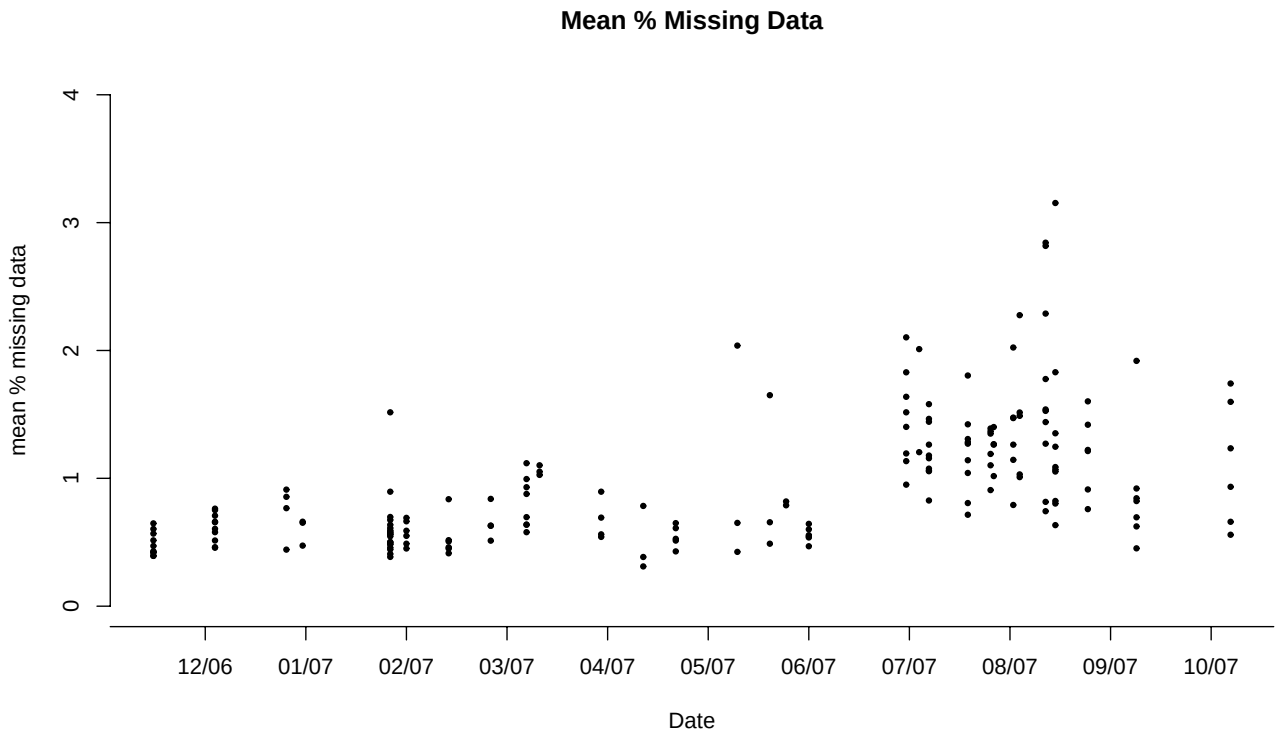

HT
Missing %
0.0
0.1
0.2
0.3
0
500
1000
1500
2000
Samples


RA
Missing %
0.0
0.1
0.2
0.3
0
500
1000
1500
2000
Samples


T1D
Missing %
0.0
0.1
0.2
0.3
0
500
1000
1500
2000
Samples


T2D
Missing %
0.0
0.1
0.2
0.3
0
500
1000
1500
2000
Samples





Supplementary Figure 25 | Missing genotypes per collection over time. The plots show the percentage of missing data for each individual in the unfiltered dataset for each collection. The individuals are coloured by plate, and plates of individuals are arranged left to right in ascending order of the time the plate was shipped to Affymetrix. The plot titled "Mean % Missing Data" shows the mean percentage of missing data for each plate of individuals in the unfiltered dataset. The plates of individuals are arranged by the date the plate was shipped to Affymetrix.

	Collection	58C	UKBS	BD	CAD	CD	HT	RA	T1D	T2D
	No Samples	1480	1458	1868	1926	1748	1952	1860	1963	1924
	% Male / % Female	50/50	48/52	37/63	79/21	39/61	40/60	25/75	51/49	58/42
Geographic region - % samples	Eastern	11	12	3	10	25	16	20	17	26
	E&WRidings	9	6	1	26	0	1	10	10	0
	London	8	5	7	2	22	18	2	4	10
	Midlands	9	12	24	6	1	4	16	6	1
	Northern	8	10	9	10	20	3	3	8	15
	North Midlands	7	3	6	15	1	13	8	7	5
	Northwestern	11	11	3	8	1	3	19	11	3
	Southeastern	7	11	5	4	1	8	2	3	2
	Southern	8	8	6	4	14	5	12	8	18
	Southwestern	8	9	6	6	1	3	1	9	19
	Scotland	10	9	10	5	14	24	3	10	0
	Wales	5	5	21	5	0	1	1	7	1
	Unknown	0	0	0	0	0	0	3	0	0
Age band % samples	<40	0	37	30	1	40	8	28*	65	5
	40-49	100	27	29	11	20	20	20*	1	16
	50-59	0	28	24	37	18	31	20*	0	33
	60-69	0	8	14	42	10	30	12*	0	34
	>70	0	0	3	9	7	11	5*	0	11
	Unknown	0	0	0	0	4	0	14*	33	1

Supplementary Table 1 | Descriptive statistics. For each collection: number of samples after QC filtering, percentage Male/Female, percentage of samples in each of 12 geographic regions (see Supplementary Information for region definitions) and percentage of samples in each age band for age at entry of individual into the study. *RA patients age-at-onset.

Relative Risk	1.3	1.5	1.7
Power (p-value threshold $1x10^{-6}$)	0.461	0.813	0.91
Power (p-value threshold $5x10^{-7}$)	0.429	0.798	0.902

Supplementary Table 2 | Power of study design. Estimate of the power of the study, with 3,000 controls and 2,000 cases using for SNPs above 5% MAF in HapMap. See Methods section for details.

a)

	Percent discordance	
	Illumina filtering	Illumina and CHIAMO filtering
BRLMM vs. Illumina	0.60	0.35
CHIAMO vs. Illumina	0.37	0.22
BRLMM vs. CHIAMO	0.39	0.18

b)

	Percent missing data		
	Illumina filtering	Illumina and CHIAMO filtering	Study-wide
BRLMM	0.79	0.63	0.7
CHIAMO	0.60	0.33	0.3
Illumina	0.57	0.54	---

Supplementary Table 3 | Agreement of genotype calls. Percent discordance and percent missing data (the percentage of genotypes for which an algorithm could not make a call with high certainty) for BRLMM, CHIAMO, and Illumina calling algorithms. The CHIAMO and BRLMM calls were made on Affymetrix 500K intensity data, and the Illumina calls were made on the Infinium platform for the same individuals from the 58C collection (Typed as part of WTCCC non-synonymous study, see www.wtccc.org.uk). Results are shown for 1,489 SNPs and 1,456 individuals that passed Illumina filters, and for 1,396 SNPs and 1,444 individuals that passed additional CHIAMO filters. Study-wide missing data figures are for 453,509 CHIAMO-filtered SNPs, of which the 1,396 used in the Illumina comparison are a subset.

Collection	Missingness	Heterozygosity	External discordance	Non-European ancestry	Duplicate	Relative	Total
58C	9	0	4	6	4	1	**24**
UKBS	8	0	5	14	0	15	**42**
BD	30	0	0	9	77	13	**129**
CAD	41	1	0	13	2	5	**62**
CD	43	4	6	54	131	18	**256**
HT	29	0	0	2	6	11	**48**
RA	47	1	0	26	53	9	**136**
T1D	7	2	1	18	6	3	**37**
T2D	36	1	0	11	16	11	**75**
Total	**250**	**9**	**16**	**153**	**295**	**86**	**809**

Supplementary Table 4 | Exclusion summary by collection. Six filters were applied for sample exclusion: 1. SNP call rate < 97% (missingness). 2. Heterozygosity > 30% or < 23% across all SNPs. 3. External discordance with genotype or phenotype data. 4. Individuals identified as having recent non-European ancestry by the Multidimensional Scaling analysis (see Methods). 5. Duplicates (the copy with more missing data was removed) 6. Individuals with too much IBS sharing (>86%); likely relatives. Where individuals could be excluded for more than one reason, they appear in the leftmost such column.

Region	Postcodes
Northern	CA DH DL HG NE SR TS YO
East & West Ridings	BD DN HD HU HX LS S WF
North Midland	DE LE LN NG NN
Eastern	AL CB CM CO EN IG IP LU MK NR PE RM SG SS WD
Southeast	BN CT DA GU ME RH TN
Southern	BH DT HP OX PO RG SL SO SP
Southwest	BA BS EX GL GY PL SN TA TQ TR
Wales	CF LD LL NP SA
Midlands	B CV DY HR ST SY TF WR WS WV
Scotland	AB DD DG EH FK G HS IV KA KW KY ML PA PH TD ZE
London	BR CR E EC HA KT N NW SE SM SW TW UB W WC
Northwest	BB BL CH CW FY IM L LA M OL PR SK WA WN

Supplementary Table 5 | Postcode definition of geographic regions. The prefix to the postcodes which defines the regions of the U.K. referred to in the Consortium.

	BD	CAD	CD	HT	RA	T1D	T2D
No covariates	1.11	1.07	1.11	1.06	1.03	1.05	1.08
PCA covariates	1.09	1.06	1.07	1.07	1.03	1.05	1.06

Supplementary Table 6 | Estimated over-dispersion of tests for association. Values greater than one indicate that the distribution of the test statistic is shifted towards larger values relative to the expected Chi-squared distribution. Values are given for the trend test (equivalent to a score test in an additive logistic regression model with no covariates), and for a test based on an additive logistic regression model in which the two ancestry informative principal components were included as covariates. Note that inclusion of the principal components does reduce overdispersion, but only by a small amount. The SNPs used in making these estimates were those that passed the filter described in the legend for Figure 3.

a) Bipolar Disorder

Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	Sex-differentiated Trend p-value	Sex-differentiated Genotypic p-value
Strong or moderate association (autosomes)									
1	60.77	rs2989476	imputed	1.61E-05	7.47E-06	3.09	3.42	6.74E-05	5.71E-05
2	11.94 - 12.00	rs4027132	chip	1.31E-05	9.69E-06	3.07	3.27	5.15E-05	9.50E-05
2	104.41 - 104.58	rs7570682	chip	3.11E-06	1.64E-05	3.68	3.23	3.35E-05	3.93E-05
2	115.63 - 116.11	rs1375144	chip	2.43E-06	1.31E-05	3.80	3.34	1.17E-05	1.25E-04
2	181.18 - 181.34	rs11888446	imputed	7.01E-07	2.96E-06	4.53	4.16	2.14E-06	1.56E-05
2	200.99	rs4673905	imputed	9.72E-06	5.44E-05	2.86	2.37	2.98E-05	3.07E-04
2	241.23 - 241.28	rs2953145	chip	1.11E-05	6.57E-06	3.22	3.50	7.32E-05	1.39E-05
3	32.26 – 32.33	rs4276227	chip	4.57E-06	2.62E-05	3.52	3.04	3.39E-05	3.38E-04
3	36.83	rs9834970	imputed	1.21E-06	7.00E-06	4.18	3.72	6.39E-06	6.02E-05
3	184.29 - 184.40	rs683395	chip	2.30E-06	5.11E-06	3.87	3.73	4.26E-06	2.10E-05
6	42.82 - 42.86	rs6458307	chip	3.43E-01	4.35E-06	-0.80	2.84	5.14E-01	4.42E-05
6	123.82	rs6901299	imputed	3.13E-06	1.08E-05	4.08	3.81	1.72E-05	1.08E-04
7	11.48	rs1405318	imputed	4.54E-06	2.72E-05	3.37	3.06	1.11E-05	5.24E-05
8	34.22 - 34.61	rs2609653	chip	6.86E-06	2.31E-05	3.44	3.21	2.97E-05	9.51E-05
9	114.31 - 114.39	rs10982256	chip	8.80E-06	4.41E-05	3.23	2.78	6.76E-06	7.92E-05
14	57.17 - 57.24	rs10134944	chip	3.21E-06	6.89E-06	3.73	3.59	2.63E-05	9.90E-05
14	103.43 - 103.62	rs11622475	chip	2.10E-06	8.14E-06	3.87	3.24	9.15E-06	8.01E-05
16	23.3 – 23.62	rs420259	chip	2.19E-04	**6.29E-08**	1.96	**4.78**	1.16E-03	6.56E-07
16	51.36 - 51.50	rs1344484	chip	1.65E-06	1.03E-05	3.94	3.41	8.30E-06	3.69E-05
20	3.70 - 3.73	rs3761218	chip	4.43E-05	6.71E-06	2.58	3.18	1.51E-04	3.44E-05
Strong or moderate association (X chromsome)									
X	110.32	Rs975687	chip	2.09E-06	9.99E-06				
$1x10^{-5}$ < p-value < $1x10^{-4}$									
1	54.96	rs10888879	chip	1.35E-05	1.00E-00				
1	60.74	rs10889189	chip	5.51E-05	2.09E-04				
1	65.39	rs4916031	chip	3.40E-02	9.80E-05				
1	70.01	rs6691577	chip	9.23E-05	2.37E-04				
1	101.66	rs1776905	chip	5.70E-05	3.03E-04				
1	213.31	rs10779279	chip	3.98E-05	2.15E-04				
1	224.13	rs12070036	chip	9.89E-05	2.85E-04				
2	62.70	rs2049674	chip	7.48E-05	1.00E-00				
2	104.44	rs17029753	chip	8.10E-05	4.17E-04				
2	115.74	rs13386690	chip	5.14E-05	2.43E-04				
2	181.18	rs4407218	chip	8.35E-05	3.73E-04				
2	200.99	rs4673905	chip	2.35E-05	1.30E-04				
3	7.63	rs1485171	chip	1.31E-01	9.73E-05				

3	21.67	rs6762678	chip	7.59E-05	2.31E-04
3	22.99	rs711715	chip	1.99E-02	5.25E-05
3	24.25	rs4858594	chip	4.02E-05	2.16E-04
3	42.38	rs33460	chip	9.56E-05	1.00E-00
3	61.56	rs13074575	chip	3.49E-05	5.43E-05
4	46.99	rs7680321	chip	6.23E-05	1.55E-04
4	54.65	rs1996755	chip	9.28E-05	3.01E-04
5	23.40	rs5009031	chip	5.17E-05	2.76E-04
5	116.21	rs1428006	chip	4.02E-05	2.17E-04
5	135.30	rs17701996	chip	7.15E-05	9.91E-05
5	162.73	rs999580	chip	9.70E-05	4.78E-04
6	18.29	rs365237	chip	1.73E-04	8.59E-05
6	33.96	rs6926599	chip	7.45E-05	1.00E-00
6	123.86	rs17739564	chip	4.28E-05	2.22E-04
6	132.77	rs6906574	chip	1.75E-04	8.76E-05
6	152.62	rs2763025	chip	5.10E-05	1.92E-05
7	22.76	rs2286492	chip	5.06E-01	2.04E-05
8	58.48	rs2875734	chip	1.20E-03	3.13E-05
8	83.20	rs16919670	chip	6.02E-05	1.58E-04
8	83.83	rs9643449	chip	9.75E-05	1.76E-04
8	102.35	rs10097578	chip	3.72E-05	1.46E-04
8	118.68	rs1993980	chip	9.12E-05	2.13E-04
9	11.21	rs7030123	chip	6.04E-05	3.11E-04
9	36.89	rs1573257	chip	3.62E-04	7.45E-05
9	90.66	rs10993698	chip	7.71E-05	3.56E-04
9	110.28	rs4978927	chip	9.92E-05	5.12E-04
9	114.33	rs10982246	chip	2.58E-05	1.27E-04
10	42.76	rs788261	chip	4.93E-05	9.13E-05
10	60.39	rs10826258	chip	7.26E-05	2.70E-04
10	79.20	rs1866437	chip	4.72E-02	4.94E-05
10	94.54	rs7896131	chip	1.56E-04	4.65E-05
10	129.77	rs2096285	chip	4.64E-05	2.47E-04
11	81.99	*(no rsID)*	chip	6.50E-05	1.00E-00
11	129.67	rs858719	chip	1.11E-03	2.87E-05
12	23.95	rs7136898	chip	1.32E-05	2.97E-05
12	93.57	rs17309820	chip	8.71E-05	1.00E-00
13	22.59	rs4770394	chip	1.20E-05	1.00E-00
13	45.42	rs2806922	chip	7.90E-03	9.45E-05
13	67.96	rs12584910	chip	5.94E-05	3.14E-04
14	23.20	rs221703	chip	4.49E-05	1.00E-00
14	38.38	rs17108400	chip	3.02E-05	2.39E-05
14	42.67	rs17113911	chip	5.89E-05	1.00E-00
14	49.23	rs10146912	chip	6.80E-05	3.29E-04
14	75.15	rs3784005	chip	3.57E-05	1.00E-00
14	103.42	rs10438244	chip	7.94E-05	1.71E-04
15	71.95	rs7163502	chip	8.72E-05	3.40E-04
16	51.43	rs1420239	chip	4.47E-05	2.28E-04
16	53.86	rs4567706	chip	1.55E-05	6.62E-05

16	72.66	rs12149894	chip	8.99E-05	2.60E-04
16	81.17	rs7184080	chip	6.73E-01	3.67E-05
16	85.85	rs10220973	chip	9.18E-05	1.00E-00
17	19.75	rs203466	chip	4.02E-05	9.23E-05
18	8.45	rs7243929	chip	2.86E-05	1.00E-00
18	8.98	rs1893146	chip	8.13E-05	4.14E-04
19	12.58	rs12979795	chip	4.12E-05	2.03E-04
19	48.49	rs7408169	chip	6.09E-05	3.13E-04
19	49.31	rs2061332	chip	8.72E-05	4.44E-04
19	63.40	rs7248493	chip	3.07E-05	1.48E-04
20	3.72	rs4815603	chip	7.50E-05	1.77E-05
20	43.16	rs6031991	chip	6.18E-05	2.66E-04
21	31.31	rs2833193	chip	5.74E-05	1.00E-00
22	31.69	rs11089599	chip	7.16E-05	1.67E-04
22	35.66	rs16997510	chip	3.70E-05	1.00E-00

b) Coronary Artery Disease

Strong or moderate association

Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	Sex-differentiated: Trend p-value	Sex-differentiated: Genotypic p-value
1	236.77 - 236.85	rs17672135	chip	1.04E-04	2.35E-06	2.36	3.88	2.50E-04	8.37E-05
5	99.98 - 100.11	rs383830	chip	5.71E-06	1.34E-05	3.49	3.26	1.56E-04	1.38E-04
6	151.34 - 151.42	rs6922269	chip	6.33E-06	1.50E-05	3.38	3.14	2.32E-05	2.04E-04
9	21.93 - 22.12	rs1333049	chip	**1.79E-14**	**1.16E-13**	**11.66**	**11.03**	**8.86E-12**	**3.67E-11**
15	76.64	rs7173512	imputed	4.58E-06	1.57E-05	3.02	2.69	2.53E-04	1.10E-03
16	81.72 - 81.79	rs8055236	chip	9.73E-06	5.60E-06	3.28	3.59	6.91E-05	5.20E-05
19	34.74 - 34.78	rs7250581	chip	9.12E-06	2.50E-05	3.30	2.87	4.56E-04	1.65E-03
22	25.01 - 25.06	rs688034	chip	6.90E-06	3.75E-06	3.33	3.15	1.05E-04	9.34E-05
				$1x10^{-5}$ < p-value < $1x10^{-4}$					
1	45.10	rs11211059	chip	2.28E-02	7.16E-05				
1	55.64	rs12068336	chip	6.51E-05	2.39E-04				
1	176.88	rs16855395	chip	9.63E-02	9.46E-05				
1	227.56	rs2883720	chip	5.07E-05	2.71E-04				
2	3.78	rs17018897	chip	3.11E-05	1.71E-04				
2	145.76	rs2044369	chip	4.86E-01	6.70E-05				
2	159.07	rs13391688	chip	2.93E-05	1.00E-00				
2	226.89	rs2943634	chip	1.19E-05	6.86E-05				
3	39.42	rs4464383	chip	7.86E-05	3.09E-04				
3	82.23	rs17728526	chip	1.61E-04	3.16E-05				
3	82.74	rs834858	chip	6.52E-05	8.19E-05				

3	108.65	rs7627215	chip	4.29E-01	1.73E-05
3	140.70	rs295470	chip	1.87E-01	5.47E-05
3	149.31	rs16860117	chip	2.18E-05	1.00E-00
3	152.29	rs906766	chip	1.44E-05	1.00E-00
3	195.18	rs7649230	chip	3.82E-03	9.55E-05
4	44.87	rs4456994	chip	7.55E-05	2.86E-04
4	132.32	rs17051141	chip	2.72E-05	1.34E-04
4	138.29	rs6841127	chip	1.50E-02	2.95E-05
4	161.70	rs6536520	chip	1.81E-02	9.97E-05
5	36.56	rs2562544	chip	1.18E-05	1.00E-00
5	66.00	rs4374700	chip	2.46E-01	7.54E-05
5	99.97	rs247951	chip	6.73E-05	3.12E-04
5	103.15	rs635331	chip	1.81E-01	6.22E-05
5	109.07	rs6897334	chip	4.94E-05	2.63E-04
7	80.13	rs17154557	chip	8.65E-05	4.00E-04
7	129.26	rs11556924	chip	2.35E-05	1.26E-04
8	106.33	rs12543827	chip	9.26E-05	4.75E-04
8	127.41	rs2471935	chip	5.82E-04	3.17E-05
8	143.23	rs10099944	chip	2.21E-05	8.03E-05
9	16.89	rs10810661	chip	5.03E-05	2.55E-04
10	5.09	rs7099555	chip	6.69E-01	3.06E-05
10	11.44	rs7917066	chip	9.84E-05	6.28E-05
10	83.75	rs10884019	chip	7.33E-03	8.01E-05
10	85.10	rs11198250	chip	1.70E-05	1.00E-00
11	133.55	rs11606866	chip	6.98E-05	1.76E-04
12	54.93	rs808919	chip	7.03E-05	2.67E-04
12	106.28	rs1426466	chip	1.46E-02	9.49E-05
12	128.80	rs2398486	chip	5.48E-05	2.92E-04
13	51.32	rs9535772	chip	7.23E-05	3.47E-04
13	75.72	rs9544230	chip	7.83E-05	3.55E-04
13	109.13	rs9521469	chip	7.00E-05	3.46E-04
14	57.21	rs10431700	chip	2.21E-05	1.09E-04
14	61.88	rs4644784	chip	2.36E-03	2.74E-05
15	38.23	rs4924428	chip	2.71E-01	6.97E-05
15	76.67	rs514743	chip	3.18E-05	1.15E-04
16	7.05	rs11647415	chip	4.37E-05	2.17E-04
16	79.38	rs17761499	chip	7.01E-01	8.50E-05
16	85.07	rs7199903	chip	1.52E-05	7.14E-05
17	5.51	rs6502872	chip	5.02E-05	2.69E-04
17	12.64	rs16946601	chip	6.56E-05	3.37E-04
17	71.45	rs2608881	chip	8.03E-05	4.16E-04
18	61.00	rs9954012	chip	7.26E-05	3.34E-04
20	59.07	rs6071467	chip	5.10E-05	2.72E-04
21	45.28	rs2838756	chip	2.47E-05	1.29E-04
22	27.46	rs5762763	chip	8.28E-01	8.33E-05

c) Crohn's Disease

Strong or moderate association									
Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	log_{10} BF additive	log_{10} BF general	Sex-differentiated Trend p-value	Sex-differentiated Genotypic p-value
1	67.3 – 67.48	rs11805303	chip	**6.45E-13**	**5.85E-12**	**10.07**	**9.19**	**6.79E-12**	**1.36E-10**
1	169.53 - 169.67	rs12037606	chip	1.79E-06	1.08E-05	3.89	3.35	8.71E-06	1.00E-04
2	27.64 - 28.55	rs7606480	imputed	6.41E-06	2.62E-05	3.02	2.69	7.68E-05	5.05E-04
2	105.36	rs3792048	chip	1.58E-02	2.77E-02	0.47	0.34	3.80E-02	**2.13E-07**
2	233.92 – 234	rs10210302	chip	**7.10E-14**	**5.26E-14**	**11.11**	**11.07**	**2.36E-14**	**5.27E-14**
3	49.3 – 49.87	rs9858542	chip	**7.71E-07**	**3.58E-08**	**4.24**	**5.22**	1.76E-06	**3.25E-07**
5	40.32 – 40.66	rs17234657	chip	**2.13E-13**	**1.99E-12**	**10.41**	**9.89**	**4.16E-12**	**5.66E-11**
5	131.40 - 131.90	rs6596075	chip	5.40E-07	3.19E-06	**4.54**	**4.01**	1.65E-06	2.15E-05
5	150.15 – 150.31	rs1000113	chip	**5.10E-08**	**3.15E-07**	5.36	5.01	2.35E-07	1.87E-06
		rs11747270	chip	4.26E-08	2.60E-07	**5.43**	**5.07**	**2.26E-07**	3.12E-06
6	20.83 - 20.85	rs6908425	chip	5.13E-06	1.10E-05	3.55	3.38	2.54E-05	6.48E-06
6	32.79 - 32.91	rs9469220	chip	8.65E-07	2.28E-06	**4.19**	3.92	3.28E-06	2.26E-05
6	138.06 - 138.17	rs7753394	chip	4.42E-06	2.59E-05	3.52	2.99	5.43E-05	1.98E-04
7	147.62 - 147.70	rs7807268	chip	6.89E-06	4.43E-06	3.33	3.58	2.96E-05	4.17E-05
9	114.66	rs7869487	imputed	3.25E-05	4.72E-06	3.12	3.78	1.63E-04	2.05E-05
10	38.52 - 38.57	rs6601764	chip	2.56E-06	8.95E-06	3.74	3.01	1.53E-05	1.29E-04
10	64.06 – 64.31	rs10761659	chip	**2.68E-07**	1.75E-06	**4.69**	3.80	8.70E-07	1.13E-05
10	101.26 – 101.32	rs10883365	chip	**1.41E-08**	**5.82E-08**	**5.91**	5.13	**7.29E-08**	**4.09E-07**
12	30.08	rs11610584	imputed	8.06E-06	3.70E-05	3.55	3.14	4.37E-05	1.94E-04
16	49.02 – 49.4	rs17221417	chip	**9.36E-12**	**3.98E-11**	8.93	8.47	1.40E-10	1.21E-09
		rs2066843	chip	1.16E-12	1.79E-12	**9.79**	**9.67**	**1.48E-11**	**7.29E-11**
17	37.77	rs744166	imputed	7.19E-06	3.85E-05	2.97	2.52	3.03E-05	2.90E-04
17	73.55	rs4362447	imputed	6.46E-06	3.69E-05	3.18	2.69	4.74E-05	4.69E-04
18	12.76 – 12.91	rs2542151	chip	**4.56E-08**	**2.03E-07**	**5.42**	**5.00**	**2.28E-07**	1.73E-06
19	50.89 - 51.07	rs8111071	chip	6.13E-06	1.75E-05	3.48	3.29	5.20E-05	2.89E-04
21	39.21	rs2836753	imputed	8.28E-06	4.74E-05	3.15	2.65	1.33E-05	5.33E-05
$1x10^{-5}$ < p-value < $1x10^{-4}$									
1	17.84	rs7547331	chip	5.86E-05	2.79E-04				
1	50.89	rs11205760	chip	1.67E-05	8.86E-05				
1	54.16	rs7515322	chip	7.26E-05	1.00E-00				
1	63.81	rs2269252	chip	2.64E-05	1.05E-04				
1	169.58	rs12037853	chip	9.46E-05	4.86E-04				
1	188.29	rs10801047	chip	8.15E-05	3.40E-04				
1	215.98	rs2791559	chip	4.13E-05	2.10E-04				
1	221.01	rs7550648	chip	3.15E-05	6.13E-05				
1	226.24	rs541010	chip	2.36E-04	5.94E-05				

1	227.18	rs16852515	chip	1.50E-05	8.52E-05
2	27.65	rs780094	chip	1.88E-05	1.04E-04
2	28.52	rs906805	chip	9.35E-05	4.13E-04
2	45.58	rs3755076	chip	4.96E-01	8.71E-05
2	51.43	rs723713	chip	5.68E-05	1.24E-04
2	101.75	*(no rsID)*	chip	6.61E-04	2.72E-05
2	188.49	rs6434236	chip	6.25E-05	2.76E-04
2	192.22	rs17351803	chip	6.47E-05	2.53E-04
2	230.92	rs13397985	chip	7.40E-05	3.15E-04
3	29.78	rs9821929	chip	1.91E-04	7.62E-05
3	49.30	rs2304442	chip	5.07E-05	1.05E-04
3	56.45	rs9855289	chip	9.33E-05	4.62E-04
3	57.50	rs9870678	chip	3.67E-05	1.97E-04
3	141.19	rs1426036	chip	2.66E-04	7.83E-05
3	149.23	rs16860030	chip	3.34E-05	1.66E-04
3	152.12	rs936189	chip	8.61E-05	2.89E-04
3	162.94	rs498051	chip	9.58E-01	3.60E-05
4	11.12	rs425002	chip	1.01E-02	9.11E-05
4	25.81	rs4692386	chip	2.31E-05	1.22E-04
4	38.70	rs2130296	chip	3.06E-01	5.10E-05
4	57.87	rs4315858	chip	9.50E-05	4.91E-04
4	89.51	rs2046132	chip	7.20E-05	3.52E-04
4	127.57	rs10011584	chip	2.56E-01	3.00E-05
5	40.32	rs348621	chip	2.10E-05	4.23E-05
5	57.95	rs2279980	chip	5.97E-05	8.50E-05
5	82.19	rs12517180	chip	9.13E-05	1.61E-04
5	119.00	rs17145587	chip	1.42E-02	5.34E-05
5	131.37	rs7714191	chip	7.20E-05	3.74E-04
5	131.68	rs3792884	chip	8.92E-05	4.48E-04
5	158.72	rs1363670	chip	3.19E-05	1.62E-04
6	3.36	rs9405639	chip	5.30E-05	2.72E-04
6	29.54	rs2107192	chip	8.62E-05	4.10E-04
6	29.91	rs2253981	chip	2.31E-05	9.10E-05
6	30.23	rs2517646	chip	3.39E-05	4.73E-05
6	30.44	rs3094055	chip	7.07E-05	6.76E-05
6	33.92	rs9469615	chip	1.40E-05	5.62E-05
7	19.77	rs10486379	chip	5.70E-01	8.63E-05
7	90.36	rs3779585	chip	5.97E-05	1.00E-00
7	130.91	rs1477226	chip	7.01E-05	3.51E-04
7	135.31	rs834771	chip	9.05E-04	9.87E-05
7	147.69	rs887822	chip	9.48E-05	2.29E-04
8	9.51	rs4240626	chip	8.61E-05	4.11E-04
8	77.90	rs10957818	chip	1.87E-05	1.00E-00
8	83.70	rs10958116	chip	5.21E-03	4.14E-05
8	126.60	rs2168131	chip	8.20E-05	3.67E-04
9	71.83	rs2147240	chip	7.75E-05	1.00E-00
9	109.15	rs3763640	chip	7.80E-04	7.43E-05
9	114.64	rs6478108	chip	9.00E-05	2.46E-05

1035.33	rs17582416	chip	3.12E-05	1.44E-04	
10	104.31	rs10786682	chip	3.85E-02	9.82E-05
11	115.93	rs12362410	chip	1.45E-03	2.46E-05
12	30.08	rs11609984	chip	7.00E-05	2.23E-04
12	111.83	rs7134391	chip	6.63E-01	5.92E-05
13	43.36	rs3764147	chip	3.97E-05	4.11E-05
13	70.48	rs17620171	chip	7.72E-05	3.98E-04
14	51.37	rs4898718	chip	6.65E-05	3.47E-04
14	53.06	rs2251589	chip	8.73E-05	4.54E-04
14	75.07	rs7161377	chip	4.13E-05	1.87E-04
14	77.10	rs4903604	chip	2.59E-03	2.65E-05
14	95.70	rs10149792	chip	2.35E-04	5.64E-05
15	30.95	rs1451890	chip	9.69E-05	4.21E-04
15	93.06	rs4984405	chip	4.42E-05	1.84E-04
17	31.42	rs2015070	chip	9.38E-05	4.86E-04
17	37.75	rs3816769	chip	3.10E-05	1.69E-04
17	42.18	rs13341140	chip	6.80E-05	3.30E-04
17	73.52	rs4789523	chip	6.95E-05	3.50E-04
18	48.70	rs16955848	chip	3.40E-02	4.69E-05
18	54.29	rs7235137	chip	6.27E-05	3.26E-04
19	1.07	rs4807569	chip	2.81E-05	1.03E-04
19	49.16	rs413061	chip	8.40E-05	4.09E-04
20	18.75	*(no rsID)*	chip	9.38E-05	4.42E-04
20	49.49	rs880324	chip	1.68E-05	9.48E-05
20	57.35	rs6128541	chip	3.32E-05	1.81E-04
20	61.81	rs6011040	chip	8.92E-05	3.92E-04
21	26.39	rs2830050	chip	8.60E-05	4.47E-04
21	39.21	rs2836753	chip	1.29E-05	7.28E-05

d) Hypertension

Strong or moderate association

Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	Trend p-value	Genotypic p-value
								Sex-differentiated	
1	71.81 – 71.82	rs1577396	imputed	1.24E-01	5.11E-06	-0.51	2.85	2.50E-01	6.12E-05
1	235.67 - 235.79	rs2820037	chip	5.76E-05	7.66E-07	2.54	3.99	1.18E-04	1.62E-06
8	140.17 - 140.35	rs6997709	chip	7.88E-06	4.36E-05	3.32	2.60	3.35E-05	3.42E-04
10	49.93 – 49.96	rs7897289	imputed	3.31E-06	1.82E-05	3.71	3.24	1.51E-05	1.50E-04
12	24.86 - 24.95	rs7961152	chip	7.39E-06	3.03E-05	3.29	2.87	3.63E-05	2.40E-04
12	100.52 - 100.58	rs11110912	chip	9.18E-06	1.94E-05	3.27	3.11	2.64E-05	1.18E-04

13	66.90 - 67.04	rs1937506	chip	9.23E-06	4.53E-05	3.25	2.85	2.17E-05	1.26E-04
15	94.60 - 94.67	rs2398162	chip	7.85E-06	5.67E-06	3.33	3.40	4.84E-05	6.64E-05
19	48.99	rs10426528	imputed	3.80E-06	2.23E-05	3.76	3.24	2.07E-05	1.14E-04
				$1x10^{-5}$ < p-value < $1x10^{-4}$					
1	71.81	rs10889923	chip	2.14E-01	1.91E-05				
1	226.22	rs557208	chip	6.56E-05	2.89E-04				
2	12.68	rs4668771	chip	4.84E-05	1.18E-04				
2	22.96	rs13421717	chip	6.37E-05	2.57E-04				
2	25.90	rs17680828	chip	9.56E-05	4.36E-04				
2	79.68	rs17017233	chip	8.64E-05	1.52E-04				
2	96.44	rs17633463	chip	6.25E-05	1.55E-04				
2	212.20	rs6435632	chip	3.14E-02	2.82E-05				
2	240.14	rs11894982	chip	8.40E-05	4.32E-04				
3	6.46	rs17234606	chip	4.33E-05	1.21E-04				
3	28.68	rs4399848	chip	9.94E-05	4.44E-04				
3	65.45	rs11710619	chip	9.47E-05	1.20E-04				
4	15.57	rs2191003	chip	6.75E-02	1.23E-05				
4	54.56	rs6824846	chip	9.22E-05	4.71E-04				
4	144.63	rs300917	chip	9.29E-05	4.44E-04				
4	145.63	rs17709487	chip	1.52E-03	9.65E-05				
5	10.25	rs9312724	chip	1.53E-05	1.00E-00				
5	100.02	rs4702982	chip	5.75E-05	2.17E-04				
6	0.46	rs2493013	chip	2.53E-03	2.92E-05				
6	76.51	rs276699	chip	9.37E-05	1.00E-00				
6	99.62	rs1884184	chip	3.53E-04	7.51E-05				
6	134.78	rs4896044	chip	8.97E-04	7.21E-05				
6	151.34	rs12201472	chip	3.06E-05	8.30E-05				
7	46.02	rs6964415	chip	5.49E-05	2.71E-04				
7	68.56	rs2851504	chip	9.18E-01	6.29E-05				
7	85.62	rs7804971	chip	3.87E-05	2.07E-04				
7	147.03	rs2710107	chip	2.67E-05	1.41E-04				
8	3.80	rs17068216	chip	1.84E-04	4.20E-05				
8	123.90	rs10095188	chip	3.30E-02	7.52E-05				
8	133.28	rs17654436	chip	7.57E-05	1.51E-04				
8	140.23	rs4074554	chip	1.39E-05	6.75E-05				
10	10.32	rs2151595	chip	2.64E-05	1.00E-00				
10	16.32	rs2991895	chip	5.68E-05	3.72E-05				
10	21.36	rs604251	chip	3.72E-05	1.57E-04				
10	27.72	rs567829	chip	1.21E-04	6.58E-05				
10	49.96	rs12269023	chip	1.80E-05	8.37E-05				
11	16.29	rs297367	chip	8.70E-05	4.25E-04				
11	21.38	rs10833525	chip	2.39E-05	8.62E-05				
11	30.38	rs12575085	chip	8.14E-05	4.25E-04				
11	73.80	rs633568	chip	1.04E-02	4.29E-05				
12	63.02	rs10784404	chip	1.75E-04	8.87E-05				
12	83.68	rs7300456	chip	1.16E-05	4.85E-05				
12	100.53	rs1727091	chip	2.21E-05	1.20E-04				

13	28.05	rs3812868	chip	7.90E-05	4.06E-04
13	30.03	rs1556428	chip	5.09E-04	5.86E-05
13	37.72	rs610642	chip	1.11E-04	8.79E-05
13	42.82	rs9316014	chip	9.74E-05	1.36E-04
13	60.33	rs167272	chip	1.81E-01	3.66E-05
14	54.32	rs709939	chip	1.85E-05	7.76E-05
14	60.77	rs4902035	chip	4.83E-01	1.11E-05
16	77.21	rs2548876	chip	8.37E-01	6.23E-05
18	63.55	rs1373365	chip	2.45E-04	9.81E-05
19	43.87	rs973009	chip	3.51E-01	3.14E-05
19	48.99	rs349045	chip	6.59E-05	2.68E-04
20	22.25	rs2424430	chip	2.85E-05	1.41E-04
20	59.07	rs6129032	chip	4.37E-05	2.25E-04

e) Rheumatoid Arthritis

Strong or moderate association

Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	Sex-differentiated: Trend p-value	Sex-differentiated: Genotypic p-value
1	2.44 - 2.77	rs6684865	chip	5.37E-06	3.14E-05	3.47	2.97	2.14E-05	1.71E-04
1	80.16 - 80.36	rs11162922	chip	1.80E-06	1.03E-05	**4.11**	3.80	2.17E-05	2.39E-04
1	113.54 – 114.16	rs6679677	chip	**4.90E-26**	**5.55E-25**	**22.36**	**21.99**	**3.85E-24**	**1.57E-22**
4	24.99 - 25.13	rs3816587	chip	7.65E-03	9.25E-06	0.50	2.64	4.38E-02	5.64E-05
6	MHC	rs6457617	chip	**3.44E-76**	**5.18E-75**	74.84	73.1754	5.65E-70	5.22E-68
6	138.00 - 138.06	rs6920220	chip	4.99E-06	1.58E-05	3.49	3.17	1.76E-05	1.13E-04
7	130.80 - 130.84	rs11761231	chip	1.74E-06	2.65E-06	3.92	3.42	**3.91E-07**	1.37E-06
10	6.07 - 6.16	rs2104286	chip	7.03E-06	2.52E-05	3.37	2.57	9.82E-05	7.64E-04
12	56.26 – 56.29	rs775251	imputed	3.44E-06	2.03E-05	3.91	3.24	3.74E-05	3.52E-04
13	19.845 - 19.855	rs9550642	chip	8.44E-06	3.90E-05	3.35	3.02	5.03E-05	3.22E-04
21	41.430 - 41.465	rs2837960	chip	3.45E-02	1.68E-06	0.05	2.70	2.97E-01	4.30E-05
22	35.870 - 35.885	rs743777	chip	7.92E-06	1.15E-06	3.29	3.52	8.93E-06	6.33E-07
				$1x10^{-5}$ < p-value < $1x10^{-4}$					
1	3.61	rs12027041	chip	1.08E-05	6.21E-05				
1	62.61	rs626787	chip	8.48E-04	8.51E-05				
1	113.60	rs12723859	chip	2.98E-05	1.21E-04				
2	40.03	rs7601303	chip	9.36E-05	1.55E-04				
2	241.24	rs2953175	chip	3.77E-05	1.48E-04				
3	65.79	rs17073902	chip	9.33E-05	1.00E-00				
3	150.56	rs11718592	chip	3.03E-05	1.00E-00				
3	195.97	rs4677742	chip	9.42E-04	9.94E-05				
4	182.12	rs6831911	chip	4.87E-04	9.01E-05				

5	79.13	rs7343	chip	8.29E-05	1.35E-04
5	95.20	rs17085170	chip	2.17E-05	1.00E-00
6	29.65	rs1233400	chip	8.64E-05	3.71E-04
6	30.77	rs1075496	chip	2.44E-05	2.90E-05
6	33.72	rs12205634	chip	5.86E-05	2.20E-04
6	46.07	rs3777612	chip	5.41E-03	7.37E-05
6	46.97	rs220704	chip	4.06E-05	1.76E-04
6	93.61	rs6909753	chip	4.29E-01	5.67E-05
7	83.74	rs2715038	chip	8.63E-05	1.05E-04
7	109.19	rs3114834	chip	1.21E-05	2.59E-05
7	128.17	rs3807306	chip	3.03E-05	1.27E-04
7	134.46	rs17236136	chip	8.08E-05	1.28E-04
8	21.05	rs12549890	chip	2.32E-05	6.31E-05
8	34.25	rs16881910	chip	2.32E-04	9.46E-05
8	85.02	rs7009279	chip	8.41E-05	2.32E-04
9	18.13	rs983230	chip	4.62E-05	1.18E-04
9	36.14	rs10814339	chip	2.07E-03	6.77E-05
10	6.43	rs4750316	chip	5.55E-05	1.96E-04
10	102.83	rs10786617	chip	3.56E-03	4.71E-05
11	15.43	rs7949682	chip	6.33E-05	2.67E-04
11	57.70	rs2514189	chip	2.59E-05	1.44E-04
11	122.70	rs7119209	chip	9.36E-05	3.93E-04
12	10.05	rs1447888	chip	5.96E-05	9.39E-05
12	56.25	rs1678542	chip	3.51E-05	1.32E-04
12	116.49	rs6490130	chip	3.47E-03	1.18E-05
12	129.43	rs11060878	chip	7.80E-03	8.55E-05
13	60.38	rs17223208	chip	4.49E-05	2.13E-04
13	74.11	rs9318297	chip	1.38E-02	8.57E-05
14	24.32	rs854350	chip	8.11E-05	1.65E-04
14	102.75	rs2771369	chip	3.84E-01	7.57E-05
16	26.78	rs2188776	chip	1.35E-02	9.71E-05
16	82.60	rs17724230	chip	4.62E-05	1.63E-04
17	28.98	rs17836884	chip	2.84E-01	9.52E-05
17	29.95	rs11080287	chip	2.15E-02	9.53E-05
17	35.90	rs896136	chip	9.01E-05	3.05E-04
18	69.43	rs4892117	chip	8.57E-03	3.21E-05
18	70.31	rs11876710	chip	7.33E-01	4.13E-05
19	24.00	rs7257520	chip	2.99E-05	1.00E-00
22	35.87	rs3218253	chip	1.53E-04	6.98E-05
22	43.52	rs1076933	chip	3.04E-02	4.21E-05

f) Type 1 Diabetes

Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	Sex-differentiated: Trend p-value	Sex-differentiated: Genotypic p-value
Strong or moderate association									
1	113.54 – 114.16	rs6679677	chip	**1.17E-26**	**5.43E-26**	**23.07**	**22.83**	**1.27E-25**	**2.13E-24**
1	221.92 - 222.17	rs2639703	chip	8.46E-06	1.74E-05	3.25	3.06	4.81E-05	1.65E-04
2	100.29 – 100.30	rs2309837	imputed	3.57E-06	1.90E-05	3.58	3.09	9.21E-06	8.16E-05
4	123.02 - 123.92	rs6534347	imputed	**4.48E-07**	1.83E-06	5.15	4.69	2.26E-06	2.47E-06
		rs17388568	chip	5.00E-07	3.27E-06	**4.42**	3.89	9.38E-07	7.67E-06
5	86.20 - 86.50	rs2544677	chip	8.23E-06	4.43E-05	3.32	2.70	4.63E-05	4.21E-04
5	132.64 - 132.67	rs17166496	chip	6.06E-01	5.20E-06	-0.97	3.25	7.05E-01	1.02E-05
6	MHC	rs9272346	chip	**2.42E-134**	**5.47E-134**	141.93	139.77	9.88E-133	1.63E-131
10	6.07 - 6.18	rs2104286	chip	7.97E-06	4.32E-05	3.31	2.88	1.01E-05	1.15E-04
10	33.47	rs2383983	imputed	2.72E-06	8.10E-06	3.74	3.47	7.64E-06	4.74E-05
12	9.71 - 9.80	rs3764021	imputed	7.19E-05	**5.08E-08**	2.12	4.60	2.86E-04	7.36E-07
		rs11052552	chip	1.02E-04	7.24E-07	2.22	3.80	4.24E-04	7.85E-06
12	54.64 – 55.09	rs11171739	chip	**1.14E-11**	**9.71E-11**	**8.89**	**8.24**	**5.91E-11**	**1.42E-09**
12	109.82 – 111.49	rs17696736	chip	**2.17E-15**	**1.51E-14**	**12.53**	**11.56**	**2.13E-14**	**2.90E-13**
16	10.93 – 11.37	rs12708716	chip	**9.24E-08**	**4.92E-07**	5.15	4.71	5.33E-07	4.10E-06
		rs9746695	chip	8.19E-09	4.85E-08	**6.19**	**5.71**	**4.45E-08**	6.17E-07
18	12.76 - 12.91	rs2542151	chip	1.89E-06	1.16E-05	3.91	3.52	1.12E-05	3.29E-05
$1x10^{-5}$ < p-value < $1x10^{-4}$									
1	19.44	rs214321	chip	5.71E-05	2.12E-04				
1	63.82	rs2269241	chip	1.23E-05	6.69E-05				
1	111.15	rs2070748	chip	1.93E-02	6.54E-05				
1	120.00	rs17258425	chip	5.51E-05	1.31E-04				
1	201.12	rs12061474	chip	1.27E-05	2.66E-05				
1	218.40	rs4579763	chip	7.53E-05	3.11E-04				
1	221.98	rs9804142	chip	8.43E-05	6.57E-05				
1	226.22	rs640333	chip	7.47E-05	3.40E-04				
1	227.44	rs10864649	chip	6.30E-05	1.52E-04				
1	233.13	rs577193	chip	8.91E-05	3.87E-04				
2	43.50	rs6732426	chip	7.96E-05	3.76E-04				
2	74.60	rs1063588	chip	6.93E-05	2.21E-04				
2	100.28	rs9653442	chip	3.19E-05	1.75E-04				
2	175.01	rs13035792	chip	3.98E-01	4.19E-05				
2	184.60	rs826140	chip	5.85E-05	2.56E-04				
2	204.53	rs11571304	chip	4.35E-05	1.61E-04				
2	215.45	rs10221582	chip	2.36E-05	9.63E-05				
3	12.49	rs709165	chip	4.97E-05	2.47E-04				

3	22.48	rs1450349	chip	9.11E-05	1.81E-04
3	46.26	rs13096142	chip	2.35E-05	2.46E-05
3	55.44	rs1520703	chip	1.08E-05	6.16E-05
3	97.03	rs1368515	chip	8.85E-05	2.12E-04
3	179.98	rs11924694	chip	4.89E-01	1.96E-05
4	57.69	rs1718886	chip	8.24E-05	4.07E-04
5	7.04	rs6873965	chip	5.45E-05	2.90E-04
5	35.94	rs1025039	chip	1.57E-05	8.90E-05
5	86.33	rs7722135	chip	1.70E-05	7.89E-05
5	167.89	rs244895	chip	9.63E-05	3.34E-04
6	18.29	rs365237	chip	1.44E-04	8.75E-05
6	25.53	rs9295657	chip	2.63E-05	2.71E-05
6	92.88	rs2452941	chip	9.98E-05	4.98E-04
7	29.10	rs2214570	chip	5.79E-02	8.61E-05
8	134.27	rs3739262	chip	3.84E-05	1.75E-04
9	22.01	rs1292136	chip	3.71E-05	1.88E-04
9	124.09	rs4838140	chip	6.86E-04	4.93E-05
10	33.46	rs2666236	chip	2.13E-05	6.78E-05
10	130.27	rs12358786	chip	5.31E-05	2.15E-04
11	2.23	rs6578252	chip	5.87E-05	2.82E-04
11	6.64	rs10500664	chip	5.93E-01	9.96E-05
11	7.55	rs7120154	chip	8.52E-05	1.00E-00
11	116.86	rs4938390	chip	9.41E-05	3.72E-05
12	9.71	rs2114870	chip	2.04E-04	1.84E-05
12	54.65	rs2069408	chip	1.45E-04	6.53E-05
12	67.98	rs11177587	chip	4.88E-01	7.63E-05
13	22.54	rs9634385	chip	1.56E-03	8.17E-05
13	23.20	rs4238171	chip	3.39E-03	6.58E-05
13	29.05	rs9579410	chip	7.57E-05	3.64E-04
15	84.95	rs2346733	chip	1.30E-04	8.83E-05
15	91.02	rs285753	chip	2.96E-04	8.33E-05
16	10.98	rs12708713	chip	5.66E-05	1.33E-04
16	88.12	rs3803676	chip	1.83E-03	9.38E-05
17	36.02	rs7221109	chip	2.20E-05	9.78E-05
18	40.56	rs12958322	chip	1.27E-04	5.14E-05
20	6.11	rs6133296	chip	4.91E-01	2.34E-05
22	35.87	rs3218253	chip	6.79E-05	3.19E-04

g) Type 2 Diabetes

Strong or moderate association									
Chromosome	Region / Position (Mb)	SNP	Type	Trend p-value	Genotypic p-value	log_{10} BF additive	log_{10} BF general	Sex-differentiated Trend p-value	Sex-differentiated Genotypic p-value
1	66.04 - 66.36	rs4655595	chip	2.68E-06	1.33E-05	3.81	3.47	8.91E-06	7.41E-05
2	160.90 - 161.17	rs6718526	chip	2.40E-06	1.15E-05	3.86	3.35	2.14E-05	7.68E-05
2	205.87	rs7587983	imputed	9.98E-06	5.68E-05	3.39	2.91	5.76E-05	5.29E-04
3	55.24 - 55.32	rs358806	chip	4.77E-01	3.05E-06	-0.83	2.72	7.22E-01	3.30E-05
4	122.92 - 123.02	rs7659604	chip	2.10E-02	9.42E-06	0.13	2.74	5.70E-02	9.64E-05
5	65.87	rs4583845	imputed	8.87E-06	4.73E-05	3.20	2.77	4.14E-05	4.30E-04
6	20.63 – 20.84	rs9465871	chip	1.02E-06	**3.34E-07**	**4.15**	3.98	2.90E-06	3.08E-06
10	43.43 - 43.63	rs9326506	chip	7.78E-06	2.99E-05	3.27	2.92	4.95E-05	3.91E-04
10	81.90 – 81.91	rs2789686	imputed	8.47E-07	2.64E-06	4.35	4.09	1.61E-06	5.95E-06
10	114.71 – 114.81	rs4506565	chip	**5.68E-13**	**5.05E-12**	**10.14**	**9.43**	**7.92E-12**	**1.53E-10**
12	49.50 - 49.87	rs12304921	chip	5.37E-02	7.07E-06	-0.09	2.68	1.09E-01	4.28E-06
12	69.58 - 69.96	rs1495377	chip	1.31E-06	6.52E-06	**4.01**	3.15	4.56E-06	4.15E-05
15	72.24 - 72.50	rs2930291	chip	7.72E-06	4.40E-05	3.30	2.42	2.03E-05	1.23E-04
15	78.12 - 78.36	rs2903265	chip	9.57E-06	4.98E-05	3.24	2.53	3.57E-05	2.42E-04
16	9.29	rs2099106	imputed	8.08E-06	1.52E-05	3.26	3.17	3.95E-05	2.58E-05
16	52.36 – 52.41	rs7193144	chip	1.44E-08	4.78E-08	**5.89**	**5.64**	**8.52E-08**	**6.90E-07**
		rs9939609	chip	**5.24E-08**	**1.91E-07**	5.35	5.05	2.98E-07	2.39E-06
$1x10^{-5}$ < p-value < $1x10^{-4}$									
1	48.93	rs12086219	chip	6.95E-05	1.00E-00				
1	73.98	rs1340430	chip	6.68E-05	2.70E-04				
1	205.98	rs6691406	chip	3.36E-05	1.83E-04				
1	219.68	rs1341987	chip	1.45E-04	4.34E-05				
2	30.78	rs7583600	chip	8.83E-05	3.38E-04				
2	60.46	rs9309324	chip	1.52E-04	1.93E-05				
2	160.97	rs1020731	chip	9.42E-05	3.57E-04				
2	189.00	rs11688935	chip	2.13E-05	8.35E-05				
2	205.86	rs17248501	chip	1.37E-05	7.62E-05				
3	11.27	rs440646	chip	4.42E-05	2.23E-04				
3	134.46	rs769097	chip	7.31E-05	3.11E-04				
3	150.03	rs16861027	chip	2.71E-05	1.00E-00				
3	154.52	rs10513440	chip	6.23E-04	9.01E-05				
4	17.12	rs1852749	chip	9.84E-05	1.00E-00				
4	123.02	rs6815973	chip	1.81E-01	1.61E-05				
4	161.74	rs1371251	chip	9.72E-01	7.59E-05				
4	178.39	rs6846031	chip	5.64E-05	9.84E-05				
5	72.74	rs4292434	chip	2.75E-05	1.00E-00				

5	122.49	rs6872465	chip	3.42E-05	1.00E-00
5	153.56	rs4958711	chip	8.85E-05	2.34E-04
6	2.40	rs9391949	chip	6.34E-05	2.46E-04
6	55.33	rs7452656	chip	1.35E-04	3.15E-05
6	107.54	rs1665901	chip	2.46E-05	1.37E-04
8	15.75	rs2736010	chip	5.42E-02	5.83E-05
8	98.43	rs2679765	chip	2.27E-03	3.37E-05
9	88.03	rs7019589	chip	4.73E-01	7.49E-05
9	114.58	rs2185935	chip	9.63E-05	1.51E-04
9	135.59	rs2590504	chip	1.17E-03	8.84E-05
10	7.78	rs7474871	chip	2.34E-01	6.59E-05
10	53.47	rs11000542	chip	7.04E-05	9.72E-05
10	104.07	rs17780667	chip	4.90E-05	2.08E-04
10	130.48	rs10829494	chip	1.02E-04	3.37E-05
11	94.53	rs11021059	chip	1.35E-03	2.22E-05
11	119.32	rs657317	chip	3.80E-05	1.00E-00
12	12.55	rs16908188	chip	9.48E-05	4.45E-04
12	18.47	rs12581163	chip	1.16E-01	6.07E-05
12	49.61	rs17125088	chip	3.60E-02	4.90E-05
12	69.69	rs11178531	chip	1.01E-05	3.75E-05
13	71.99	rs4053550	chip	3.91E-05	1.00E-00
14	83.09	rs1007383	chip	3.31E-01	7.59E-05
14	98.08	rs8012854	chip	3.98E-05	2.13E-04
14	98.29	rs4343209	chip	9.40E-05	1.45E-04
16	9.29	Rs2099106	chip	1.74E-05	2.98E-05
18	62.34	Rs508987	chip	1.50E-03	6.28E-05
18	75.56	Rs70198	chip	6.91E-03	9.84E-05
20	40.25	Rs7262414	chip	9.94E-05	1.58E-04
22	45.01	Rs739164	chip	7.12E-05	3.58E-04

Supplementary Table 7 | Association results by disease. Regions with at least one SNP with either a strong or moderate association (from Tables 3 & 4 and Supplementary Tables 8, 9 & 10) or a p-value in the range $1x10^{-5}$ to $1x10^{-4}$ and within 200 Kb of at least one other SNP with a p-value less than $1x10^{-3}$. Numbers in bold represent strong associations that were displayed in a previous table. Positions are in NCBI build 35 coordinates

Collection	Gene	Chromosome	SNP	Add vs. Gen p-value	Dom/Rec vs. Gen p-value
BD	*PALB2*	16p2	rs420259	$5.41 \cdot 10^{-6}$	0.958 (rec)
CAD		9p21	rs1333049	0.273	
CD		1p31	rs11805303	0.750	
CD		2q37	rs10210302	0.039	0.028 (dom)
CD	*BSN*	3p21	rs9858542	0.003	0.148 (dom)
CD		5p13	rs17234657	0.930	
CD	*IRGM*	5q33	rs1000113	0.591	
CD		10q21	rs10761659	0.924	
CD		10q24	rs10883365	0.344	
CD		16q12	rs17221417	0.296	
CD		18p11	rs2542151	0.389	
RA	*PTPN22*	1p13	rs6679677	0.928	
T1D	*PTPN22*	1p13	rs6679677	0.117	
T1D		12q13	rs11171739	0.745	
T1D		12q24	rs17696736	0.490	
T1D		16p13	rs12708716	0.522	
T2D	*CDKAL1*	6p22	rs9465871	0.019	0.008 (dom)
T2D	*TCF7L2*	10q25	rs4506565	0.896	
T2D	*FTO*	16q12	rs9939609	0.219	

Supplementary Table 8 | Comparision of disease models. For each of the 19 SNPs outside the MHC region that are listed in Main Table 3 we tested for departures from an additive model on the log-odds scale (a model with multiplicative odds ratios). We compared a 1-df additive model with a 2-df general model, the unadjusted p-values are reported in the column "Add vs. Gen p-value". Two of the 19 SNPs show strong evidence for departure from additivity. At the SNP rs420259 associated with BD the best fitting model is a recessive model and at the SNP rs9858542 associated with CD a dominant model fits best. Another two SNPs showed moderate evidence for departure from additivity, rs10210302 associated with CD and rs9465871 associated with T2D. The best fitting model for both of these is a general 2-df model.

Collection	Chromosome	SNP	Position	Expanded reference group: Trend p-value	Expanded reference group: Genotypic p-value	58C+UKBS controls: Trend p-value	58C+UKBS controls: Genotypic p-value	Exp. ref. group frequency	Case frequency
BD	1p31	rs2989476	60771280	1.71E-07	2.60E-07	2.27E-05	1.24E-05	0.425	0.470
BD	2q31	rs12465451	176719036	1.18E-03	3.30E-07	5.06E-03	3.50E-04	0.130	0.149
BD	12q21	rs1526805	73670904	6.80E-02	2.18E-07	3.37E-02	1.26E-04	0.050	0.057
BD	22q12	rs8138016	35026649	6.41E-02	1.61E-07	5.04E-02	2.20E-02	0.012	0.016
CAD	9p21	rs6475606	22071850	1.64E-17	1.75E-16	4.39E-14	2.67E-13	0.480	0.554
CAD	15q25	rs1994016	76867289	4.93E-07	2.68E-06	1.06E-04	3.94E-04	0.422	0.379
CD	1p31	rs11805303	67387537	3.03E-23	2.90E-22	6.45E-13	5.85E-12	0.306	0.391
CD	1q24	rs12037606	169630059	3.37E-07	2.15E-06	1.79E-06	1.08E-05	0.392	0.438
CD	2p23	rs906805	28516530	5.98E-08	3.57E-07	9.34E-05	4.13E-04	0.468	0.419
CD	2q37	rs6431654	233943769	1.39E-16	5.98E-16	8.68E-14	9.06E-14	0.476	0.401
CD	3p21	rs9858542	49676987	8.55E-08	1.20E-08	7.71E-07	3.58E-08	0.286	0.331
CD	5p13	rs17234657	40437266	1.04E-16	4.00E-16	2.13E-13	1.99E-12	0.128	0.181
CD	5q23	rs10077785	131829057	2.25E-09	1.55E-08	1.81E-06	1.11E-05	0.238	0.192
CD	5q33	rs1000113	150220269	4.31E-07	1.68E-06	5.10E-08	3.15E-07	0.073	0.098
CD	10q21	rs10761659	64115570	2.36E-08	1.55E-07	2.68E-07	1.75E-06	0.457	0.406
CD	10q24	rs10883371	101282445	1.23E-09	4.20E-09	1.43E-08	4.93E-08	0.482	0.537
CD	14q13	rs17496932	33007950	8.49E-06	3.14E-07	5.28E-04	2.90E-04	0.037	0.053
CD	16q12	rs17221417	49297083	2.90E-14	1.46E-14	9.36E-12	3.98E-11	0.292	0.356
CD	18p11	rs2542151	12769947	1.04E-10	8.46E-11	4.56E-08	2.03E-07	0.164	0.208
CD	21q22	rs2836754	39213610	1.07E-07	5.94E-07	1.07E-05	6.11E-05	0.353	0.399
HT	14q11	rs11158632	23839503	1.93E-08	1.16E-07	2.57E-04	1.20E-03	0.208	0.247
RA	1p13	rs6679677	114015850	1.33E-37	2.00E-36	4.90E-26	5.55E-25	0.098	0.168
RA	6q23	rs5029939	138237416	1.21E-08	3.24E-08	5.42E-06	2.09E-05	0.036	0.055
RA	10p15	rs10795791	6148346	5.17E-08	1.40E-07	2.62E-05	6.74E-06	0.407	0.455
RA	11p15	rs10500889	20994596	7.87E-13	1.00E-00	1.63E-04	1.63E-04	0.000	0.002
T1D	1p13	rs6679677	114015850	3.98E-40	1.40E-41	1.17E-26	5.43E-26	0.098	0.169
T1D	2q33	rs3087243	204564425	2.07E-07	1.14E-06	3.27E-05	1.41E-04	0.448	0.404
T1D	3p21	rs6441961	46327388	4.50E-07	5.58E-07	1.17E-05	2.20E-05	0.295	0.335
T1D	10p15	rs10795791	6148346	1.26E-08	6.95E-08	1.36E-05	5.94E-05	0.407	0.456
T1D	12p13	rs10772079	9765661	4.14E-08	1.79E-07	1.70E-04	2.29E-04	0.354	0.308
T1D	12q13	rs11171739	54756892	2.04E-14	1.98E-13	1.14E-11	9.71E-11	0.426	0.493
T1D	12q24	rs17696736	110949538	2.62E-15	2.23E-14	2.17E-15	1.51E-14	0.437	0.506
T1D	16p13	rs12708716	11087374	3.73E-11	1.91E-10	9.24E-08	4.92E-07	0.352	0.297
T1D	17q21	rs7221109	36023812	2.00E-07	1.27E-06	2.20E-05	9.78E-05	0.355	0.311
T1D	18p11	rs2542151	12769947	1.67E-08	1.04E-07	1.89E-06	1.16E-05	0.164	0.201
T1D	22q13	rs229541	35915818	3.39E-07	2.18E-06	7.68E-04	1.29E-03	0.423	0.466
T2D	2q24	rs7593730	160996961	2.02E-07	1.24E-06	3.58E-06	1.87E-05	0.225	0.188
T2D	5q14	rs6865544	82933140	1.38E-07	1.01E-07	7.09E-06	3.91E-05	0.008	0.016
T2D	10q25	rs4506565	114746031	2.72E-24	3.30E-23	5.68E-13	5.05E-12	0.313	0.395
T2D	12q15	rs7132840	69697828	1.93E-08	9.42E-08	8.29E-06	2.97E-05	0.454	0.502
T2D	16q12	rs8050136	52373776	6.46E-11	4.20E-10	2.00E-08	7.04E-08	0.400	0.455

Supplementary Table 9 | Regions of the genome exhibiting strong signals using expanded reference group analysis. Regions containing p-values less than $5x10^{-7}$ on trend and/or genotypic tests of association for analyses that used an expanded reference group. For each of BD, CAD, HT and T2D, the expanded reference group comprised the 58BC and UKBS controls supplemented by the other six disease sample sets (See Main text). For CD, RA, and T1D, the expanded reference group included 58BC and UKBS controls augmented with the 7670 cases from the non-autoimmune disease sets. For result on the X chromosome see Table 16. Allele frequencies are shown for the allele that is minor in the expanded reference group. Positions are in NCBI build 35 coordinates.

Collection	Gene	Chromosome	Strongest signal of association in region					Evidence at previously reported SNPs				
			SNP	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	Reported SNP	WTCCC SNP	r^2	Trend p-value	Genotypic p-value
BD	*BDNF*	11p14	rs4923460	9.82E-02	1.80E-01	-0.36	-0.55	rs6265	rs6265	-	-	0.17
BD	*DAOA*	13q33	rs1981272	3.15E-02	2.71E-02	-0.04	-0.32	-	-	-	-	-
BD	*DTNBP1*	6p22	rs742206	1.14E-03	1.48E-03	1.41	1.31	-	-	-	-	-
BD	*DISC1*	1q42	rs1407601	3.22E-02	9.33E-02	0.09	-0.23	-	-	-	-	-
BD	*NRG1*	8p12	rs1487152	3.48E-03	1.34E-02	0.81	-1.59	-	-	-	-	-
CAD	*ALOX5AP*	13q12	rs4075131	1.82E-01	1.78E-01	-0.59	-0.72	-	-	-	-	-
CAD	*KIAA0992*	4q32	rs17054463	6.55E-02	1.81E-01	-0.08	-0.35	rs12510359	rs1039386	-	0.49	0.37
CAD	*ROS1*	6q22	rs9320600	6.07E-03	2.14E-02	0.75	0.49	rs619203	rs501109	0.07	0.92	0.97
CAD	*TAS2R50*	12p13	rs10772414	4.74E-02	8.09E-02	-0.07	-0.23	rs1376251	-	-	-	-
CAD	*OR13G1*	1q44	rs1144812	5.72E-02	9.72E-03	-0.24	0.28	rs1151640	rs880143	-	0.83	0.98
CAD	*PCSK9*	1p32	rs594226	1.56E-01	3.65E-01	-0.53	-0.67	-	-	-	-	-
CAD	*FactorV*	1q24	rs7474070	1.57E-01	3.68E-01	-0.30	-0.31	rs6025	rs6020	-	0.68	1
CAD	*Prothrombin*	11p11	rs3136439	1.40E-01	3.34E-01	-0.31	-0.43	-	-	-	-	-
CAD	*PAI-1*	7q22	rs3807513	8.50E-02	1.35E-01	-0.32	-0.49	-	-	-	-	-
CD	*CARD15*	16q12	rs17221417	9.36E-12	3.98E-11	8.93	8.47	rs2066844	rs17221417	0.23	9.40E-12	4.00E-11
CD	*IL23R*	1p31	rs11805303	6.45E-13	5.85E-12	10.07	9.19	rs11805303	rs11805303	-	6.50E-13	5.90E-12
CD	*ATG16L1*	2q37	rs10210302	7.10E-14	5.26E-14	11.11	11.07	rs2241880	rs10210302	0.97	7.10E-14	5.30E-14
HT	*ADD1*	4p16	rs7665452	2.84E-02	7.04E-02	0.27	0.29	rs4961	rs2239728	0.95	0.39	0.69
HT	*GNB3*	12p13	rs11064432	3.35E-01	1.62E-01	-0.45	-0.40	rs5443	rs10849538	0.77	0.66	0.91
HT	*ADRB2*	5q32	rs11959615	1.58E-02	3.69E-02	0.22	-0.36	rs1042713	rs17778257	0.96	0.089	0.23
HT								rs1042714	rs2400707	0.97	0.034	0.05
HT	*AGT*	1q42	rs2479131	1.82E-02	6.01E-02	0.17	-0.12	rs4762	rs2479131	0.52	0.019	0.06
HT								rs699	rs11122577	0.69	0.16	0.26
HT	*WNK1*	12p13	rs11611246	5.06E-02	1.41E-01	-0.13	-0.33	-	-	-	-	-
RA	*PADI4*	1p36	rs1748041	1.56E-01	3.59E-01	-0.56	-0.70	-	-	-	-	-
RA	*MHC2TA*	16p13	rs7201430	2.71E-01	3.70E-01	-0.59	-0.66	-	-	-	-	-
RA	*FCRL3*	1q23	rs17676026	5.86E-01	2.76E-01	-0.88	-0.84	-	-	-	-	-
RA	*SLC22A4*	5q23	rs4705938	4.56E-01	1.12E-01	-0.91	-0.67	-	-	-	-	-
RA	*CTLA4*	2q33	rs11571300	3.38E-02	5.16E-03	0.08	0.34	rs3087243	rs3087243	-	0.085	0.22

Supplementary Table 10 | Evidence for signal of association at other genes of previous interest. For all genomic regions of prior interest we scanned the data for signals of association in the relevant collection of the WTCCC. Two approaches were taken. First we looked at the reported putative disease gene and surrounding 20kb, reporting the lowest p-value for the trend and genotypic test, the highest $\log_{10}$ Bayes factor for the additive and general models, and the rsID of the SNP (columns 4-8). Second, where information on the strength of association at a particular SNP had been previously published we tabulated the p-value of both the trend and genotype test at the same SNP (if in our study), or the best tag SNP (defined to be the SNP with highest r^2 with the reported SNP, calculated in the CEU population of the HapMap project). This information is given in columns 9-13. Positions are in NCBI build 35 coordinates.

Combination of Collections	Chromosome	SNP	Position	Trend p-value	Genotypic p-value	In Main Table 3	In Main Table 4
T1D+RA	1p13	rs6679677	114015850	2.54E-34	3.90E-33	y	
T1D+RA	6	MHC				y	
T1D+RA	10p15	rs2104286	6139051	5.92E-08	2.53E-07	y	y
T1D+RA	12q24	rs17696736	110949538	5.72E-11	3.29E-10	y	
T1D+RA+CD	1p13	rs6679677	114015850	2.96E-16	2.55E-15	y	
T1D+RA+CD	6	MHC				y	
T1D+RA+CD	12q24	rs17696736	110949538	9.33E-10	2.62E-09	y	
T1D+RA+CD	18p11	rs2542151	12769947	9.29E-08	6.17E-07	y	

Supplementary Table 11 | Strongest association signals in combinations of diseases with putatively similar aetiology. Regions with at least one SNP giving a p-value of less than $5x10^{-7}$ in either the trend or the genotypic test when the given cases are pooled together, with a representative SNP for each (chosen to have the smallest p-value and a good cluster plot). Also included are whether these regions showed a strong or moderate association in a previous analysis. Positions are in NCBI build 35 coordinates.

Collection	Chromosome	SNP	Position	Trend p-value	Genotypic p-value
p-value < $5x10^{-7}$					
T1D	4q27	rs6534347	123556040	4.48E-07	1.83E-06
T1D	12p13	rs3764021	9724895	7.19E-05	5.08E-08
$5x10^{-7}$ < p-value < $1x10^{-5}$					
BD	1p31	rs2989476	60771280	1.61E-05	7.47E-06
BD	2q31	rs11888446	181196401	7.01E-07	2.96E-06
BD	2q33	rs4673905	200989289	9.72E-06	5.44E-05
BD	3p22	rs9834970	36831034	1.21E-06	7.00E-06
BD	6q22	rs6901299	123817025	3.13E-06	1.08E-05
BD	7p21	rs1405318	11484132	4.54E-06	2.72E-05
CAD	15q25	rs7173512	76636969	4.58E-06	1.57E-05
CD	9q32	rs7869487	114660468	3.25E-05	4.72E-06
CD	12p11	rs11610584	30076780	8.06E-06	3.70E-05
CD	17q21	rs744166	37767727	7.19E-06	3.85E-05
CD	17q25	rs4362447	73545157	6.46E-06	3.69E-05
CD	21q22	rs2836753	39213057	8.28E-06	4.74E-05
HT	1p31	rs1577396	71813438	0.124169	5.11E-06

HT	10q11	rs7897289	49938769	3.31E-06	1.82E-05
HT	19q13	rs10426528	48991138	3.80E-06	2.23E-05
T1D	2q11	rs2309837	100296085	3.57E-06	1.90E-05
T1D	10p11	rs2383983	33466820	2.72E-06	8.10E-06
T2D	2q33	rs7587983	205865210	9.98E-06	5.68E-05
T2D	5q12	rs4583845	65866380	8.87E-06	4.73E-05
T2D	16p13	rs2099106	9288209	8.08E-06	1.52E-05

Supplementary Table 12 | Regions of strong and moderate association identified by imputing HapMap SNPs, but not near regions already found. This table gives SNPs that represent possible regions of association found through imputation. These are separated into two classes: the top of the table shows strong signals of association (p-value < $5x10^{-7}$) and the bottom of the table shows moderate signals of association ($5x10^{-7}$ < p-value < $1x10^{-5}$). Note that regions with strong association are reported only when there is no strong signal at a SNP directly typed in the study within 500kb (the results are robust to this figure), and SNPs showing moderate association are reported only if there is no moderate or stronger association at SNPs directly typed in the study within 500kb. This is to identify regions of association not found in the single SNP analyses. See Methods for details of QC for these analyses. Positions are in NCBI build 35 coordinates.

	Number of samples	Proportion missing
58C	1480	0.0030
UKBS	1458	0.0037
BD	1868	0.0039
CAD	1926	0.0034
CD	1748	0.0033
HT	1952	0.0036
RA	1860	0.0035
T1D	1963	0.0034
T2D	1924	0.0038
Total	**16179**	**0.0035**

Supplementary Table 13 | Missing data rates by collection. For each collection, the number of non-excluded samples and the proportion of missing data in the clean dataset (after excluding bad samples and bad SNPs).

Collection	Chromosome	SNP	Position	Trend p-value	Genotypic p-value	$\log_{10}$ BF additive	$\log_{10}$ BF general	In Main Table 3	In Main Table 4
BD	16p12	rs420259	23541527	2.19E-04	6.29E-08	1.96	4.78	y	
CAD	9p21	rs1333049	22115503	1.79E-14	1.16E-13	11.66	11.03	y	
CD	1p31	rs11805303	67387537	6.45E-13	5.85E-12	10.07	9.19	y	
CD	2q37	rs10210302	233940839	7.10E-14	5.26E-14	11.11	11.07	y	
CD	3p21	rs9858542	49676987	7.71E-07	3.58E-08	4.24	5.22	y	
CD	5p13	rs17234657	40437266	2.13E-13	1.99E-12	10.41	9.89	y	
CD	5q23	rs6596075	131770127	5.40E-07	3.19E-06	4.54	4.01		y
CD	5q33	rs11747270	150239060	4.26E-08	2.03E-07	5.43	5.07	y	
CD	6p21	rs9469220	32766288	8.65E-07	2.28E-06	4.19	3.92		y
CD	10q21	rs10761659	64115570	2.68E-07	1.75E-06	4.69	3.80	y	
CD	10q24	rs10883365	101277754	1.41E-08	5.82E-08	5.91	5.13	y	
CD	16q12	rs2066843	49302700	1.16E-12	1.79E-12	9.79	9.67	y	
CD	18p11	rs2542151	12769947	4.56E-08	2.03E-07	5.42	5.00	y	
RA	1p31	rs11162922	80284079	1.80E-06	3.60E-06	4.11	3.80		y
RA	1p13	rs6679677	114015850	4.90E-26	5.55E-25	22.36	21.99	y	
RA	6	MHC						y	
T1D	1p13	rs6679677	114015850	1.17E-26	5.43E-26	23.07	22.83	y	
T1D	4q27	rs17388568	123686967	5.00E-07	3.27E-06	4.42	3.89		y
T1D	6	MHC						y	
T1D	12q13	rs11171739	54756892	1.14E-11	9.71E-11	8.89	8.24	y	
T1D	12q24	rs17696736	110949538	2.17E-15	1.51E-14	12.53	11.56	y	
T1D	16p13	rs9746695	11115395	8.19E-09	4.85E-08	6.19	5.71	y	
T2D	6p22	rs9465871	20825234	1.02E-06	3.34E-07	4.15	3.98	y	
T2D	10q25	rs4506565	114746031	5.68E-13	5.05E-12	10.14	9.43	y	
T2D	12q15	rs1495377	69863368	1.31E-06	6.52E-06	4.01	3.15		y
T2D	16q12	rs7193144	52368187	1.44E-08	4.78E-08	5.89	5.64	y	

Supplementary Table 14 | Strongest association signals using Bayesian Analysis. Regions with at least one SNP with a -$\log_{10}$ Bayes factor of greater than 4. Also included are the p-values for the frequentist tests and whether these SNPs feature in Main Tables 3 & 4. Positions are in NCBI build 35 coordinates.

Combination	Chromosome	SNP	Position	Trend p-value (Sex as effect modifier)	Genotypic p-value (Sex as effect modifier)	In Main Table 3	In Main Table 4
CAD	9p21	rs1333049	22115503	8.86E-12	3.67E-11	Y	
CD	1p31	rs11805303	67387537	6.79E-12	1.36E-10	Y	
CD	2q12	rs3792048	105360191	3.80E-02	2.13E-07		
CD	2q37	rs10210302	233940839	2.36E-14	5.27E-14	Y	
CD	3p21	rs9858542	49676987	1.76E-06	3.25E-07	Y	
CD	5p13	rs17234657	40437266	4.16E-12	5.66E-11	Y	
CD	5q33	rs11747270	150239060	2.26E-07	3.12E-06	Y	
CD	10q24	rs10883365	101277754	7.29E-08	4.09E-07	Y	
CD	16q12	rs2066843	49302700	1.48E-11	7.29E-11	Y	
CD	18p11	rs2542151	12769947	2.28E-07	1.73E-06	Y	
RA	1p13	rs6679677	114015850	3.85E-24	1.57E-22	Y	
RA	6	MHC				Y	
RA	7q32	rs11761231	130827294	3.91E-07	1.37E-06	Y	y
T1D	1p13	rs6679677	114015850	1.27E-25	2.13E-24	Y	
T1D	6	MHC				Y	
T1D	12q13	rs11171739	54756892	5.91E-11	1.42E-09	Y	
T1D	12q24	rs17696736	110949538	2.13E-14	2.90E-13	Y	
T1D	16p13	rs9746695	11115395	4.45E-08	6.17E-07	Y	
T2D	10q25	rs4506565	114746031	7.92E-12	1.53E-10	Y	
T2D	16q12	rs7193144	52368187	8.52E-08	6.90E-07	Y	

Supplementary Table 15 | Strong association at SNPs using a sex differentiated test. Regions with at least one SNP giving a p-value of less than $5x10^{-7}$ in either the trend or the genotypic test using a sex differentiated test (see Methods), with a representative SNP for each (chosen to have the smallest p-value and a good cluster plot). Also included are whether these regions showed a strong or moderate evidence of association in a previous analysis. As noted in the main text, this test is still sensitive to SNPs whose effects are the same in both sexes, but with some loss of power compared to the earlier analyses. Positions are in NCBI build 35 coordinates.

Collection	SNP	Position	Trend p-value	Genotypic p-value
BD	rs975687	110318150	2.09E-06	9.99E-06
Expanded reference group				
CD	rs2807261	135387319	1.32E-07	3.95E-07
HT	rs5938070	74450099	1.36E-08	8.60E-08
HT	rs5932296	126859352	8.49E-06	4.48E-05

Supplementary Table 16 | Regions with strong or moderate association on the X Chromosome. SNPs showing strong signal of association on the X chromosome. We used a modified version of the trend and genotypic test which controls for the hemizygosity of males (see Methods). Expanded references groups explained above. Positions are in NCBI build 35 coordinates.