# MultiBLUP: improved SNP-based prediction for complex traits. Supplementary Material

Doug Speed and David Balding
UCL Genetics Institute,
University College London

**Contents**

Table S1: Contributions of MHC and non-MHC SNPs to heritability and prediction for WTCCC1 traits.
Table S2: Performance of MultiBLUP for WTCCC1 traits, with regions based on eQTL status.
Table S3: Performance of MultiBLUP for Crohn's Disease, with regions based on pathway information.
Table S4: Performance of methods for WTCCC1 traits.
Table S5: Performance of methods for Celiac Disease and Inflammatory Bowel Disease.
Table S6: Performance of 77 SNP Model for Celiac Disease.
Table S7: Performance for WTCCC1 traits of Adaptive MultiBLUP with alternative settings.

Figure S1: Principal component analysis for WTCCC1 data.
Figure S2: Principal component analysis for Celiac Disease data.
Figure S3: Principal component analysis for Inflammatory Bowel Disease data.
Figure S4: Prediction performance on simulated data, with all SNPs within a region causal.
Figure S5: Genomic Selection Performance on simulated data.
Figure S6: Single-SNP analysis of WTCCC1 traits.
Figure S7: Example region distribution for Adaptive MultiBLUP applied to WTCCC1 traits.
Figure S8: Example region distribution for Celiac Disease and Inflammatory Bowel Disease.
Figure S9: Performance for mouse phenotypes.

# References

Curtis, C., Shah, S., Chin, S., Turashvili, G., Rueda, O., Dunning, M., Speed, D., Lynch, A., Samarajiwa, S., Yuan, H., *et al.*, 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**:346–352.

Dubois, P., Trynka, G., Franke, L., Hunt, K., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G., Ádány, R., and Aromaa, A., *et al.*, 2010. Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.*, **42**:295–302.

Monsuur, A., de Bakker, P., Zhernakova, A., Pinto, D., Verduijn, W., Romanos, J., Auricchio, R., Lopez, A., van Heel, D., Crusius, J., *et al.*, 2008. Effective detection of human leukocyte antigen risk alleles in coeliac disease using tag single nucleotide polymorphisms. *PLoS ONE*, **3**:e2270.

Romanos, J., Rosén, A., Kumar, V., Trynka, G., Franke, L., Szperl, A., Gutierrez-Achury, J., van Diemen, C., Kanninga, R., Jankipersadsing, S., *et al.*, 2014. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. *Gut*, **63**:415–422.

The 1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. *Nature*, **467**:1061–1073.

The International HapMap Consortium, 2003. The International HapMap Project. *Nature*, **426**:789–796.

Yang, J., Benjamin, B., McEvoy, B., Gordon, S., Henders, A., Nyholt, D., Madden, P., Heath, A., Martin, N., Montgomery, G., *et al.*, 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, **42**:565–569.

Yang, J., Manolio, T., Pasquale, L., Boerwinkle, E., Caporaso, N., Cunningham, J., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M., *et al.*, 2011. Genomic partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.*, **43**:519–525.

**Table 1. Contributions of MHC and non-MHC SNPs to heritability and prediction for WTCCC1 traits.**

| | Total | | MHC | | Non-MHC | |
|---|---|---|---|---|---|---|
| Trait | $h^2$ | $r^2$ | $h^2$ | $r^2$ | $h^2$ | $r^2$ |
| Bipolar Disorder | 0.98 | 0.07 | 0.00 | 0.00 | 0.97 | 0.07 |
| Coronary Artery Disease | 0.41 | 0.02 | 0.00 | 0.00 | 0.41 | 0.02 |
| Crohn's Disease | 1.00 | 0.10 | 0.02 | 0.01 | 0.98 | 0.10 |
| Hypertension | 0.48 | 0.02 | 0.00 | 0.00 | 0.48 | 0.02 |
| Rheumatoid Arthritis | 0.57 | 0.12 | 0.13 | 0.11 | 0.44 | 0.01 |
| Type 1 Diabetes | 0.65 | 0.32 | 0.44 | 0.31 | 0.21 | 0.01 |
| Type 2 Diabetes | 0.52 | 0.02 | 0.00 | 0.00 | 0.51 | 0.02 |

For each WTCCC1 trait, the first two values report the total heritability ($h^2$, measured on the observed scale), and prediction accuracy (this time measured by correlation squared between observed and predicted phenotypic values, $r^2$). The next two pairs of values break down total $h^2$ and $r^2$ into contributions from the MHC and non-MHC regions. $h^2$ is converted into $r^2$ more efficiently for the MHC region, as it contains relatively few SNPs and so it easier to accurately estimate effect sizes. By contrast, the non-MHC regions can contribute large $h^2$ but relatively little $r^2$, because across a large region, the assumption of constant effect size variance is inappropriate. Were individuals only distantly related, the total $h^2$ would represent the "chip heritability" (Yang et al. 2010), and each regional $h^2$ value would reflect the variance explained by the SNPs in the corresponding region (the basis of genome partitioning; Yang et al. 2011). However, in general, and as is the case here, the presence of (moderate) relatedness makes these values harder to interpret.

**Table 2. Performance of MultiBLUP for WTCCC1 traits, with regions based on eQTL status.**

| Trait | BLUP | MultiBLUP |
|---|---|---|
| Bipolar Disorder | **0.27** | 0.26 |
| Coronary Artery Disease | **0.13** | 0.12 |
| Crohn's Disease | **0.32** | 0.26 |
| Hypertension | **0.15** | 0.14 |
| Rheumatoid Arthritis | 0.21 | **0.31** |
| Type 1 Diabetes | 0.25 | **0.46** |
| Type 2 Diabetes | **0.16** | 0.16 |

For each WTCCC1 trait, we compare the performance of BLUP with two-region MultiBLUP, where the regions are defined according to eQTL status; we classify a SNP as an eQTL if it shows association ($P < 10^{-10}$) with expression levels for one or more genes. For this we used data from the METABRIC project, a study into Breast Cancer which collected SNP and gene expression data from tumor cells for 1903 patients (data described in Curtis et al. (2012) and available at www.ebi.ac.uk/ega). The values report prediction performance, measured as correlation between predicted and observed phenotypic values, based on ten-fold cross validation. The best performing method for each trait is marked in **bold**.

**Table 3. Performance of MultiBLUP for Crohn's Disease, with regions based on pathway information.**

| Pathway | Genes | $h^2$ | $r^2$ |
|---|---|---|---|
| Interleukin 2 Receptor Beta Chain in T cell Activation | CRKL, E2F1, AKT1, IKZF3, FOS, GRB2, HRAS, FA IL2RA, FASLG, IL2RB, IL2RG, IRS1, JAK1, JAK3, MYC PIK3CA, PIK3CG, PIK3R1, PPIA MAPK1, MAPK, BAD, PTPN6 RAF1, BCL2, BCL2L1, RPS6KB1, SHC1, SOS1, STAT5A, STAT5B SYK, SOCS1, CB, CFLAR, SOCS3, NMI | 0.209 | 0.004 |
| Interleukin 12 Mediated Signalling Pathway | IL12RB2, IL12B, IL12RB1, CD3D, CD3G, CD247, JUN, CD3E IL18R1, JAK2, CCR5, MAPK14, IL18, IFNG, MAP2K6, STAT4 IL12A, TYK2, ETV5, MAPK8 | 0.221 | 0.007 |
| | IL23R | 0.115 | 0.012 |
| | NOD2 | 0.090 | 0.009 |
| Background Region | All SNPs more than 30 kb from exonic regions of above genes | 0.365 | 0.086 |
| Total | | 1.00 | 0.102 |

The first four regions contain SNPs corresponding to each of two gene pathways and two individual genes known to affect susceptibility to Crohn's Disease (based on evidence from independent association studies); the fifth region contains all other SNPs. In each row, the first value reports the heritability contribution from each region ($h^2$), while the second value represents how much each region contributes to the prediction model ($r^2$). As when dividing into MHC and non-MHC, the ability of MultiBLUP to convert heritability into prediction is higher for smaller regions.

**Table 4. Performance of methods for WTCCC1 traits.**

| **Mean Squared Error** | Current methods | | | | MultiBLUP | |
|---|---|---|---|---|---|---|
| Trait | BLUP | Genetic Risk Score | Stepwise Regression | BSLMM | Two-region MHC/non-MHC | Adaptive |
| Bipolar Disorder | **0.21** | 74.56 | 0.23 | 0.22 | 0.22 | 0.22 |
| Coronary Artery Disease | 0.21 | 83.14 | 0.21 | 0.21 | 0.21 | **0.21** |
| Crohn's Disease | 0.21 | 75.91 | 0.22 | 0.21 | 0.22 | **0.21** |
| Hypertension | 0.23 | 52.02 | 0.24 | 0.23 | 0.23 | **0.23** |
| Rheumatoid Arthritis | 0.22 | 95.19 | 0.20 | 0.20 | 0.20 | **0.19** |
| Type 1 Diabetes | 0.22 | 125.2 | 0.17 | 0.16 | 0.16 | **0.15** |
| Type 2 Diabetes | 0.23 | 50.57 | 0.23 | 0.23 | 0.23 | **0.23** |
| Average across 7 traits | 0.22 | 79.51 | 0.21 | 0.21 | 0.21 | **0.20** |

| **Median Absolute Error** | Current methods | | | | MultiBLUP | |
|---|---|---|---|---|---|---|
| Trait | BLUP | Genetic Risk Score | Stepwise Regression | BSLMM | Two-region MHC/non-MHC | Adaptive |
| Bipolar Disorder | **0.41** | 5.44 | 0.45 | 0.41 | 0.42 | 0.41 |
| Coronary Artery Disease | 0.45 | 6.78 | 0.46 | 0.44 | 0.45 | **0.44** |
| Crohn's Disease | 0.40 | 5.53 | 0.43 | 0.39 | 0.41 | **0.38** |
| Hypertension | 0.45 | 4.68 | 0.45 | 0.45 | 0.45 | **0.45** |
| Rheumatoid Arthritis | 0.41 | 6.82 | 0.39 | 0.37 | 0.38 | **0.37** |
| Type 1 Diabetes | 0.44 | 7.23 | 0.33 | 0.32 | 0.32 | **0.31** |
| Type 2 Diabetes | 0.45 | 4.79 | 0.45 | 0.45 | 0.45 | **0.45** |
| Average across 7 traits | 0.43 | 5.89 | 0.42 | 0.41 | 0.41 | **0.40** |

| **Area Under Curve** | Current methods | | | | MultiBLUP | |
|---|---|---|---|---|---|---|
| Trait | BLUP | Genetic Risk Score | Stepwise Regression | BSLMM | Two-region MHC/non-MHC | Adaptive |
| Bipolar Disorder | **0.64** | 0.64 | 0.49 | 0.64 | 0.64 | 0.64 |
| Coronary Artery Disease | 0.58 | 0.57 | 0.53 | 0.59 | 0.58 | **0.59** |
| Crohn's Disease | 0.66 | 0.65 | 0.60 | 0.67 | 0.64 | **0.69** |
| Hypertension | 0.59 | 0.58 | 0.09 | 0.58 | 0.58 | **0.59** |
| Rheumatoid Arthritis | 0.61 | 0.67 | 0.70 | 0.69 | 0.71 | **0.73** |
| Type 1 Diabetes | 0.64 | 0.72 | 0.83 | 0.84 | 0.84 | **0.86** |
| Type 2 Diabetes | 0.59 | 0.58 | 0.55 | 0.60 | 0.59 | **0.60** |
| Average across 7 traits | 0.61 | 0.63 | 0.54 | 0.66 | 0.65 | **0.67** |

We compare MultiBLUP (regions defined according to MHC/non-MHC) and Adaptive MultiBLUP (starting with 75 kb regions), with BLUP, genetic risk scores, stepwise regression and BSLMM (Bayesian Sparse Linear Mixed Models). Phenotypic values were predicted using ten-fold cross-validation. For the main text we measured prediction performance based on correlation. These three tables measure how well these predicted phenotypes match the observed values based on mean squared error, absolute median error and area under curve (AUC). For the genetic risk scores, we consider five $p$-value thresholds (1 to 5 on the $-\log_{10}$ scale), and report the best prediction across these. The best performing method for each trait is marked in **bold**.

**Table 5. Performance of methods for Celiac Disease and Inflammatory Bowel Disease.**

| | BLUP | | Risk Score ($-\log_{10}(P)$) | | Adaptive MultiBLUP | |
|---|---|---|---|---|---|---|
| Trait (Number of Samples) | MSE | MAE | MSE | MAE | MSE | MAE |
| Celiac Disease, All Samples (15283) | 1.84 | **1.16** | 154 (5) | 10.3 (5) | **1.82** | 1.22 |
| Celiac Disease, UK Cohorts Only (10118) | 1.72 | **1.12** | 146 (5) | 10.0 (5) | **1.71** | 1.18 |
| Celiac Disease, UK2 $\rightarrow$ UK1 (6785 $\rightarrow$ 3333) | **1.63** | **1.21** | 147 (5) | 10.1 (5) | 1.65 | 1.40 |
| Inflammatory Bowel Disease (12678) | 0.23 | 0.47 | 0.23 (4) | **0.43** (2) | **0.21** | 0.44 |
| Crohn's Disease (8826) | 0.19 | **0.31** | 0.18 (4) | 0.32 (5) | **0.18** | 0.32 |
| Ulcerative Colitis (9978) | 0.22 | 0.41 | 0.22 (4) | **0.39** (2) | **0.21** | 0.40 |

We compare the performance of BLUP, genetic risk scores, and Adaptive MultiBLUP (starting with 75 kb regions) applied to datasets for Celiac Disease and Inflammatory Bowel Disease. For the main text, we measured prediction performance based on correlation and area under curve. Here we instead report mean squared error (MSE) and median absolute error (MAE) between observed and predicted phenotypes. For Celiac Disease, we consider all samples and separately UK samples only; for Inflammatory Bowel Disease, we consider all samples, only Crohn's Disease cases and only ulcerative colitis cases. The results are based on ten-fold cross-validation, except that for Celiac Disease, we also perform out-of-sample prediction from one UK cohort into the other. The best performing method for each measure is marked in **bold**. Because SNP effect sizes are estimated independently, genetic risk scores will tend to perform poorly for MSE and MAE when SNPs are not first pruned (the case for Celiac Disease); by contrast, the method performs much better when SNPs are pruned (Inflammatory Bowel Disease).

**Table 6. Performance of 77 SNP Model for Celiac Disease.**

| | BLUP | | Risk Score ($-\log_{10}(P)$) | | Adaptive MultiBLUP | | 77 SNP Model | |
|---|---|---|---|---|---|---|---|---|
| Samples | $r$ | AUC | $r$ | AUC | $r$ | AUC | $r$ | AUC |
| All Samples (15283) | 0.46 | 0.79 | 0.45 (1) | 0.79 (1) | **0.57** | **0.86** | 0.40 | 0.76 |
| UK Cohorts Only (10118) | 0.42 | 0.78 | 0.44 (1) | 0.79 (1) | **0.55** | **0.86** | 0.41 | 0.76 |
| UK2 $\rightarrow$ UK1 (6785 $\rightarrow$ 3333) | 0.41 | 0.78 | 0.44 (1) | 0.80 (1) | **0.53** | **0.86** | 0.41 | 0.79 |

For Celiac Disease, we additionally consider the prediction performance of a model based on the work of Romanos et al. (2014). They considered three models, "HLA + 10 SNPs", "HLA + 26 SNPs" and "HLA + 57 SNPs". Each model includes 6 SNPs tagging four human leukocyte antigen (HLA) risk alleles: DQ2.2, DQ2.5, DQ7 and DQ8 (Monsuur et al. 2008). To these they added either 10, 26 or 57 non-HLA SNPs identified as affecting susceptibility through single-SNP tests of association. Only 21 of the SNPs used by Romanos et al. (2014) were present in our dataset, however, we were able to find tags for 77 of the 85 SNPs by studying SNP-SNP correlations across European individuals in the The 1000 Genomes Project Consortium (2010) (the mean and median correlation between target and tagging SNPs were 0.76 and 0.89, respectively). We then constructed a linear prediction model from these 77 SNPs using least squares regression (so effect sizes were fitted jointly). For each of the three subsets of samples considered, the values above report $r$, the correlation between observed and predicted phenotypic values, and area under curve (AUC); for the first two subsets, we used 10-fold cross-validation, while for the third we used out-of-sample prediction from one UK cohort to the other. We found this 77 SNP model performed worse than BLUP, genetic risk scores and Adaptive MultiBLUP. We are keen to emphasize that, on account of the different SNP sets and methodologies used, the model we considered is only an approximation to that of (Romanos et al. 2014); nonetheless, the results suggest that for prediction it is better to include all genome-wide SNPs, rather than just some of the most strongly associated.

**Table 7. Performance for WTCCC1 traits of Adaptive MultiBLUP with alternative settings.**

| Trait | Default Settings | Initial region size 37.5 kb | Initial region size 150 kb | $p$-value threshold $P < 10^{-5}$ | $p$-value threshold $P < 10^{-7}$ | Alternative Correction |
|---|---|---|---|---|---|---|
| Bipolar Disorder | 0.27 | 0.27 | 0.27 | 0.28 | 0.27 | 0.30 |
| Coronary Artery Disease | 0.16 | 0.16 | 0.16 | 0.16 | 0.15 | 0.17 |
| Crohn's Disease | 0.36 | 0.36 | 0.35 | 0.37 | 0.36 | 0.38 |
| Hypertension | 0.17 | 0.17 | 0.16 | 0.18 | 0.15 | 0.19 |
| Rheumatoid Arthritis | 0.37 | 0.37 | 0.37 | 0.37 | 0.37 | 0.38 |
| Type 1 Diabetes | 0.59 | 0.59 | 0.59 | 0.60 | 0.58 | 0.60 |
| Type 2 Diabetes | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.20 |
| Average over 7 traits | 0.30 | 0.30 | 0.30 | 0.30 | 0.30 | 0.32 |

The two main parameter settings in Adaptive MultiBLUP are the initial region size and the $p$-value threshold used for determining which of these are associated with the phenotype. The default settings are 75 kb and $10^{-6}$. Here we see the effect when analyzing the WTCCC1 traits of instead starting with an initial region size of 37.5 kb or 150 kb, or of using a $p$-value threshold of $10^{-5}$ or $10^{-7}$. These changes have little effect, although prediction slightly improves (average correlation 0.304 vs 0.301) when the initial region size is reduced. For the analyses of human data, we first regressed case/control status on sex plus the first 20 axes from principal component analysis. This approach could be criticized as over-conservative, as it might remove true signal from the data which could be used for prediction. As an alternative, we consider instead regressing status on sex and two population axes derived independently from the HapMap reference panel (The International HapMap Consortium 2003). This change results in improved prediction for all traits. Some of this improvement may be due to residual population structure; although in some cases this form of relatedness is a legitimate source of information for prediction, it was desirable to eliminate it from our analyses to allow fairer comparison with stepwise regression, a sparse method that is less likely to benefit from this population structure signal, and also so that our results are a better reflection of out-of-sample prediction.

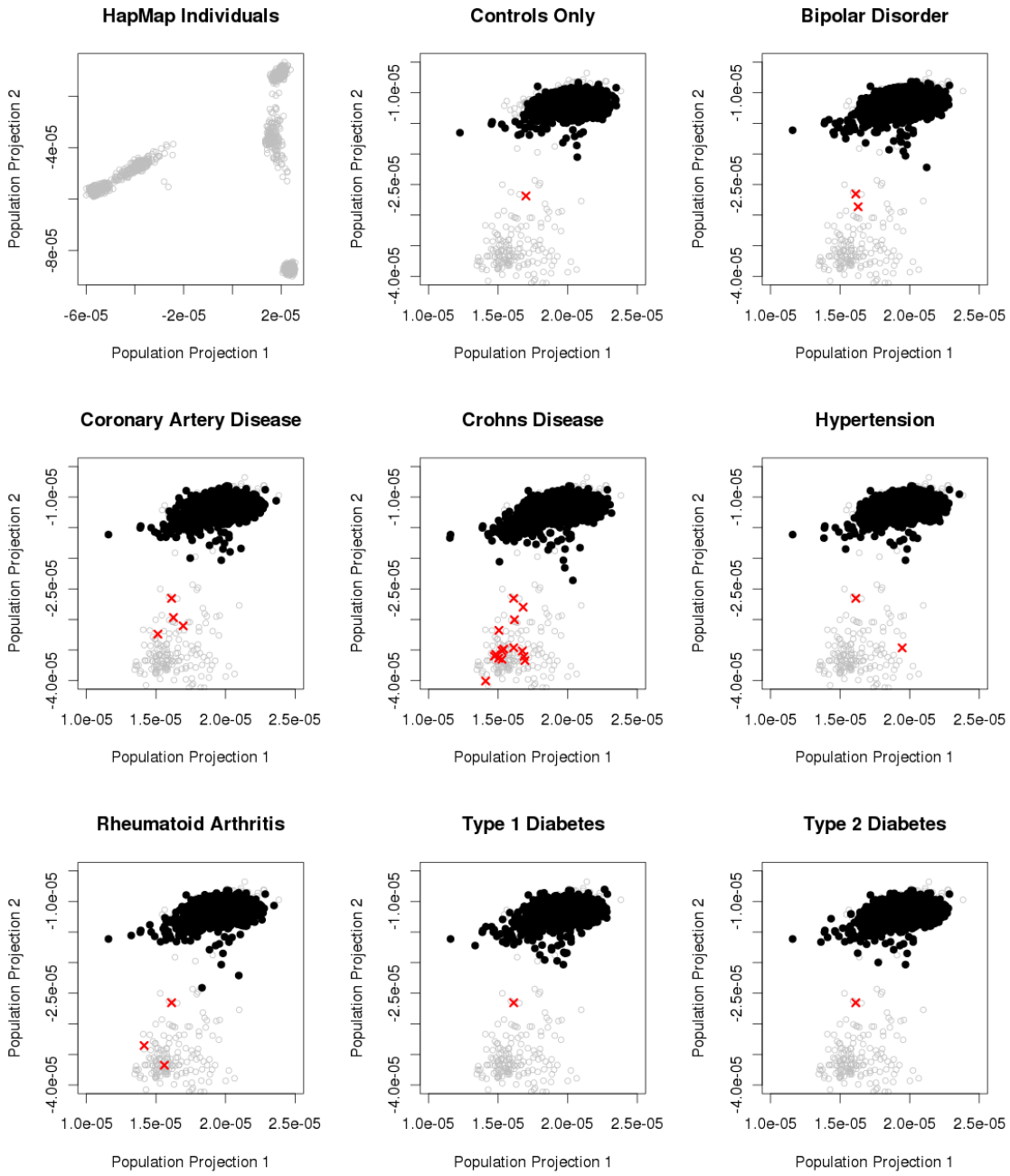


**Figure 1. Principal component analysis for WTCCC1 data.** We obtained the leading two principal component axes from the HapMap data upon which we projected all WTCCC1 individuals to examine ethnicity. The top left plot plot shows the projection for the HapMap individuals; the cluster near the top right corner corresponds to European-ancestry individuals. The remaining eight plots (zoomed in on the European Cluster) show the projections for first the control individuals, then the control and case individuals used in each association study. Red crosses indicate individuals considered population outliers and so removed from further analyses.

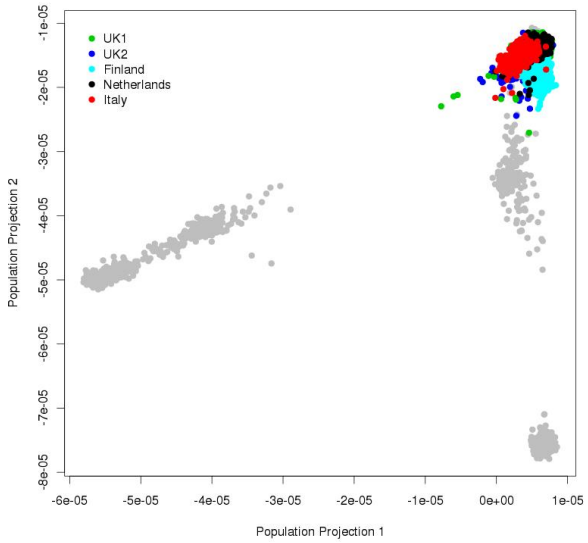


**Figure 2. Principal component analysis for Celiac Disease data.** Again, we obtained two principal component axes from the HapMap reference panel, upon which we projected all individuals to examine ethnicity. Individuals are colored according to the cohort from which they originate. We conclude that the sample quality control performed by Dubois et al. (2010) was sufficient for our purposes so removed no further individuals; although we noted that the UK cohorts appear more ethnically homogeneous, hence our decision to perform a UK-only analysis.

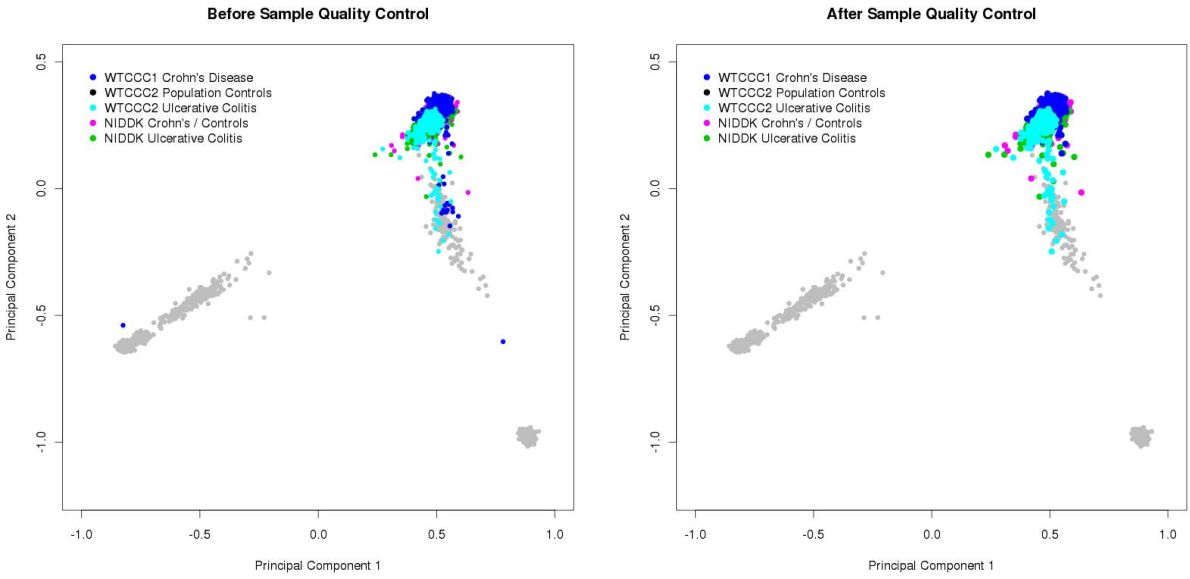


**Figure 3. Principal component analysis for Inflammatory Bowel Disease data.** To examine ancestry, we project individuals onto two population axes derived from the HapMap reference panel. Individuals are colored according to the study from which they were obtained. The left plot contains all individuals, while the right plot shows only those that survived sample quality control.

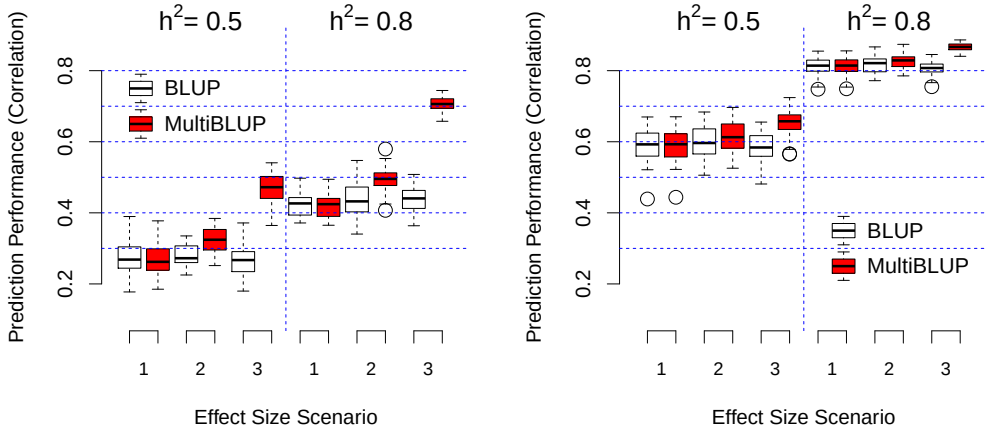


**Figure 4. Prediction performance on simulated data, with all SNPs within a region causal.** Both BLUP and MultiBLUP assume that every SNP is causal, although the effect-size variance can be very small. For the main text simulations, phenotypes were generated using 20 SNPs per contributing region; here, we instead allow every SNP within the region to contribute. The two plots correspond to unrelated humans (left) and related mice (right). They show across 50 repetitions the correlation between predicted and observed phenotypes in the test set, for BLUP (white boxes) and MultiBLUP (red boxes). The $x$-axis indexes the simulation scenarios, with increasing heterogeneity of effect sizes across the five regions. MultiBLUP uses five GSMs, one from computed from each region used in the simulations. Within each plot, the true (simulated) heritability is 0.5 (left half) or 0.8 (right half).

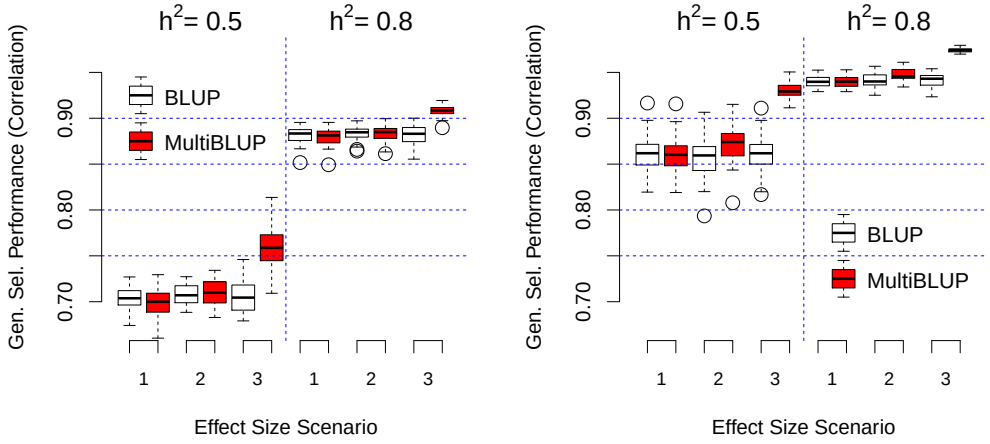


**Figure 5. Genomic Selection Performance on simulated data.** For the simulations in the main text (20 causal SNPs per contributing region), as well as measuring prediction performance for test individuals, we can also consider genomic selection performance for training individuals, the ability to estimate phenotypic values with environmental noise discounted (these are referred to as "breeding values" by animal and plant geneticists). The two plots correspond to unrelated humans (left) and related mice (right). They show across 50 repetitions the correlation between predicted and true breeding values in the training set, for BLUP (white boxes) and MultiBLUP (red boxes). The $x$-axis indexes the simulation scenarios, with increasing heterogeneity of effect sizes across the five regions. MultiBLUP uses five GSMs, one computed from each region used in the simulations. Within each plot, the true (simulated) heritability is 0.5 (left half) or 0.8 (right half).

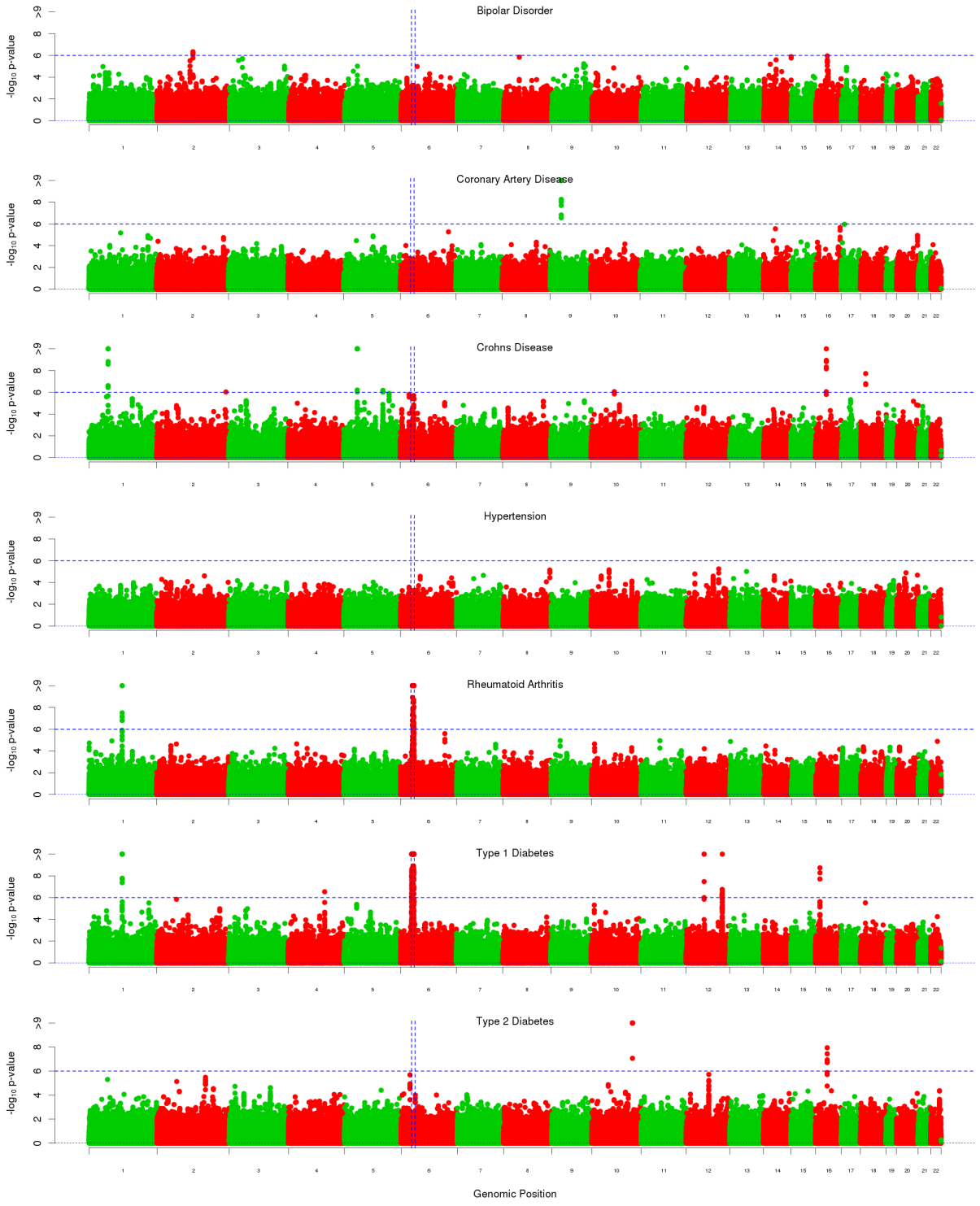


**Figure 6. Single-SNP analysis of WTCCC1 traits.** The seven plots present $-\log_{10} P$ for each of the seven traits (values above 9 have been truncated). The horizontal blue line marks $P = 10^{-6}$, the cut-off used for stepwise regression, while the vertical blue lines delineate the major histocompatibility complex (MHC) on Chromosome 6.

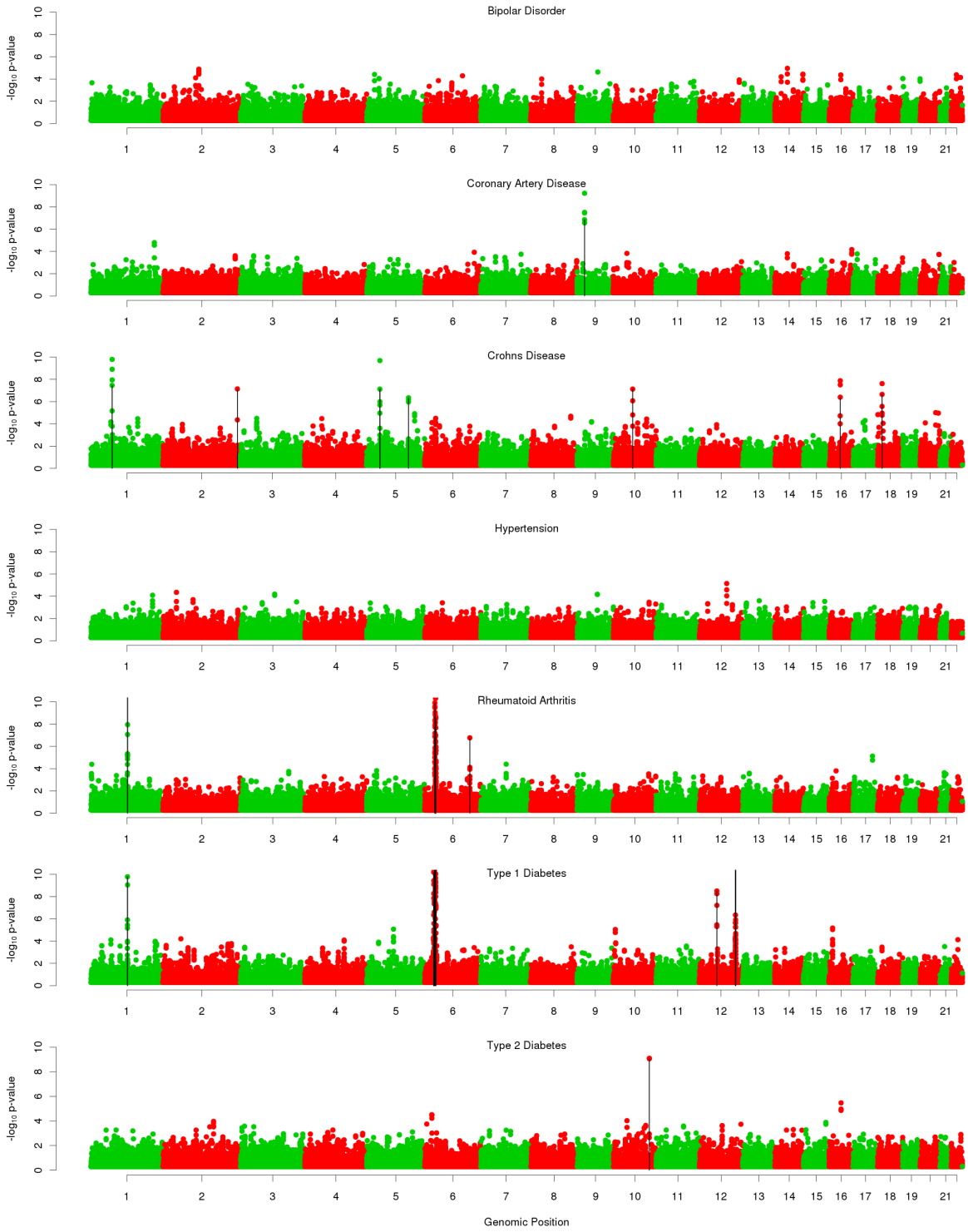


**Figure 7. Example region distribution for Adaptive MultiBLUP applied to WTCCC1 traits.** Adaptive MultiBLUP was run starting with about 68 000 regions of size 75 kb, with an overlap of 37.5 kb. The plots show for each trait the final regions for the first fold of the ten-fold cross-validation. Each vertical blue line signifies a distinct local region, while SNPs outside these blue lines formed the background region.

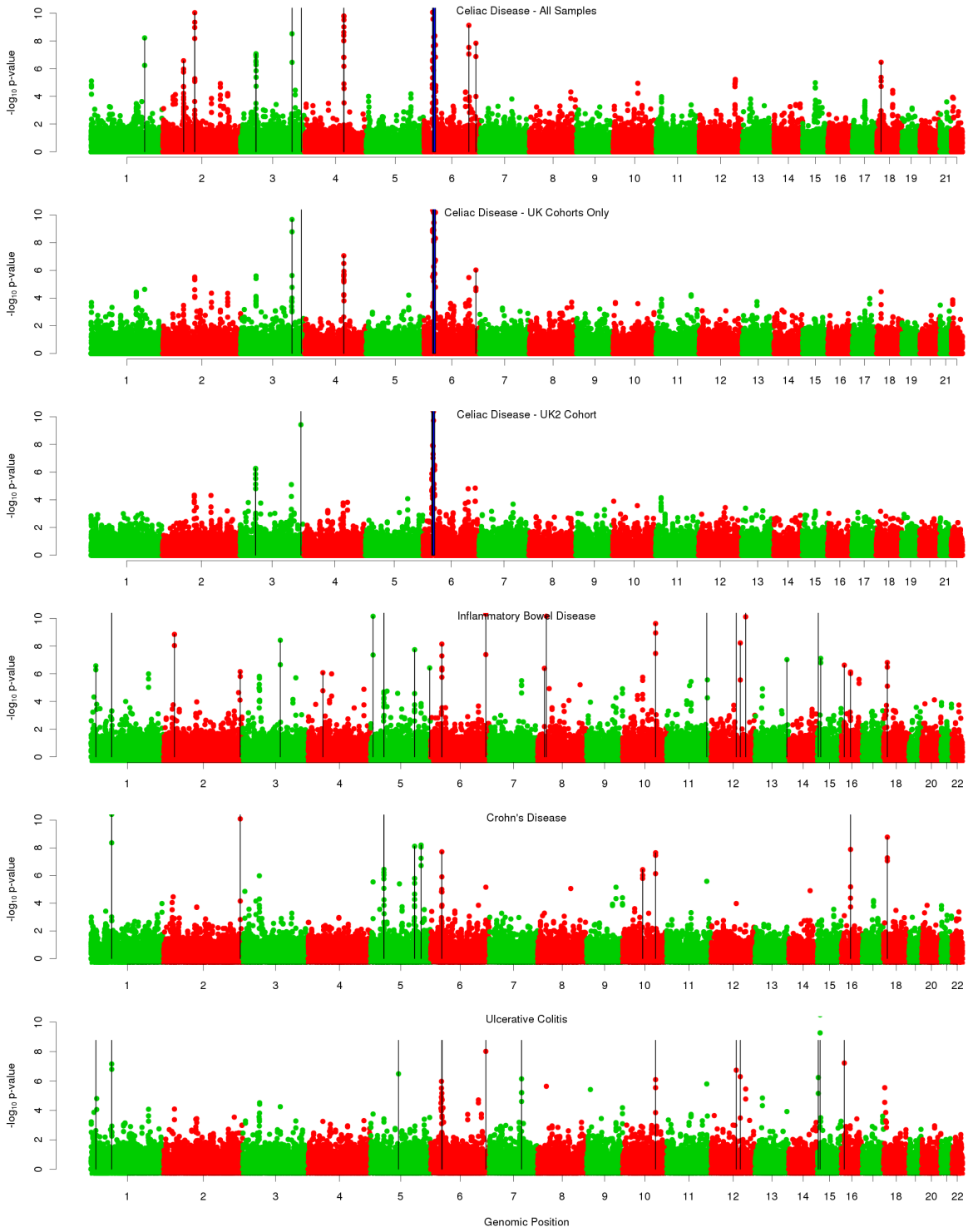


**Figure 8. Example region distribution for Adaptive MultiBLUP applied to Celiac Disease Inflammatory Bowel Disease data.** Adaptive MultiBLUP was run starting with either 67 000 (Celiac Disease) or 60 000 regions (Inflammatory Bowel Disease) regions of size 75 kb, with an overlap of 37.5 kb. The plots show for each analysis the final regions for the first fold of the ten-fold cross-validation (except for the plot entitled "Celiac Disease – UK2 Cohort," where the $p$-values are based on all UK2 Cohort samples). Each vertical blue line signifies a distinct local region, while SNPs outside these blue lines formed the background region.

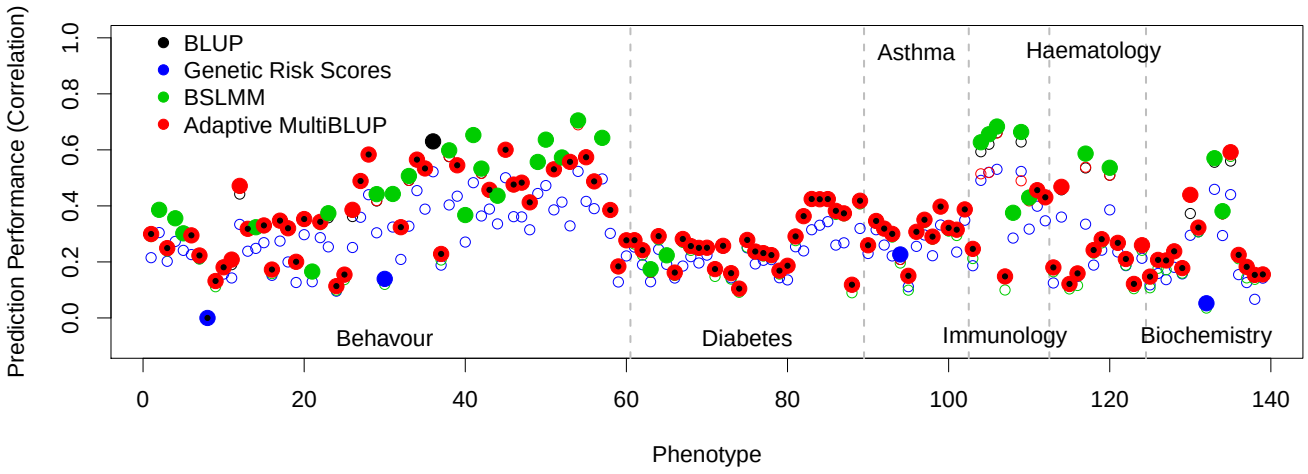


**Figure 9. Performance for mouse phenotypes.** In total, we considered 139 mouse phenotypes, spanning five broad categories: behavior, diabetes, asthma, immunology, haematology or biochemistry. For each phenotype, we report the performance for BLUP, genetic risk scores, BSLMM and Adaptive MultiBLUP, measured by correlation between predicted and observed phenotypic values based on ten-fold cross-validation. For genetic risk scores, we considered five $p$-value thresholds (1 to 5 on the $-\log_{10}$ scale), and for each trait report only the best performing. The most accurate method is marked by a solid circle. For many phenotypes, Adaptive MultiBLUP found no significant regions, so was identical to BLUP; a black spot within a red circle indicates BLUP and Adaptive MultiBLUP tied as the best performing method. The only clear conclusion is that genetic risk scores copes poorly when the dataset contains high levels of relatedness (average correlation 0.265); otherwise, which of BLUP, BSLMM and Adaptive MultiBLUP performs slightly better than the other two appears to depend on which phenotype is being tested (average correlations 0.335, 0.336 and 0.336, respectively).