The American Journal of Human Genetics, *91*

Supplemental Data

Improved Heritability Estimation from Genome-wide SNPs

Doug Speed, Gibran Hemani, Michael R. Johnson, and David J. Balding

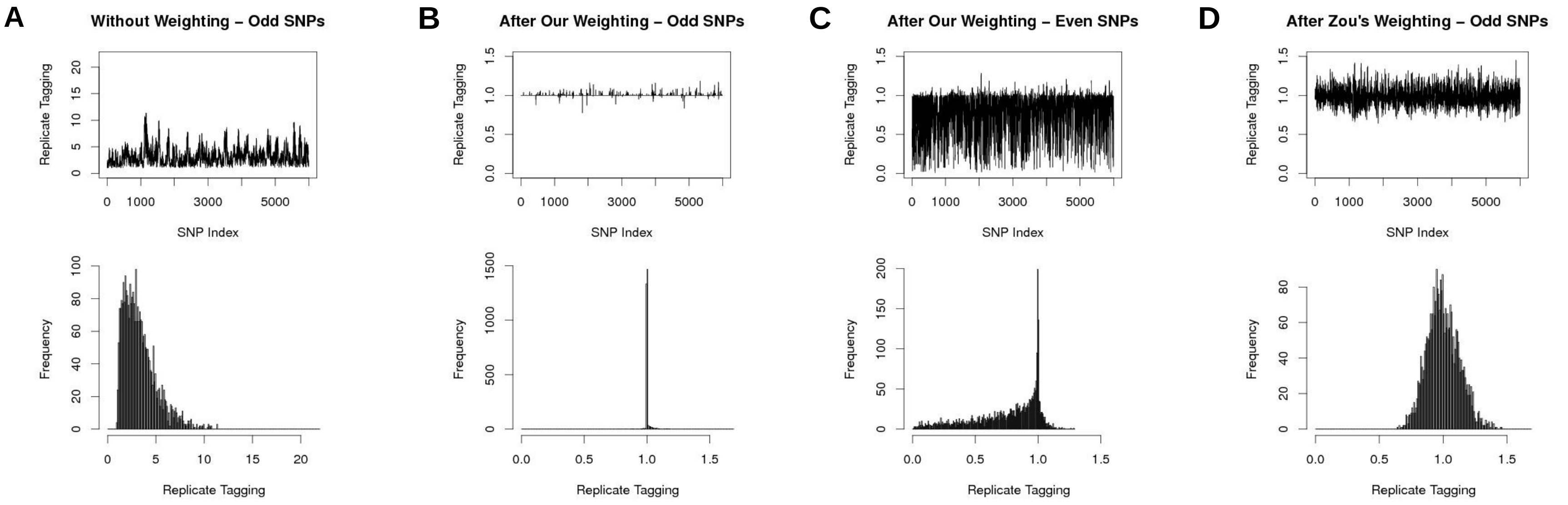

Figure S1. Equalizing the Tagging of SNPs through the Introduction of Weightings

(A) The uneven nature of tagging is demonstrated. The *y*-axis indicates the tagging (defined by (5) in the main text) for the first 3000 odd-numbered SNPs. When using the standard kinship matrix, these values also correspond to the variance effectively assigned to each SNP's signal – showing how the standard method assumes more tagged variants have on average larger effect sizes.

(B) The aim of the weightings is to equalize this tagging / variance. The weightings here have been calculated based on the odd-numbered SNPs, so equalize their variance almost perfectly.

(C) Even though the weightings are calculated agnostic of the even-numbered SNPs, they none-the-less equalize their variance fairly well too.

(D) By comparison, the weighting proposed by Zou et al.[1] is less effective.

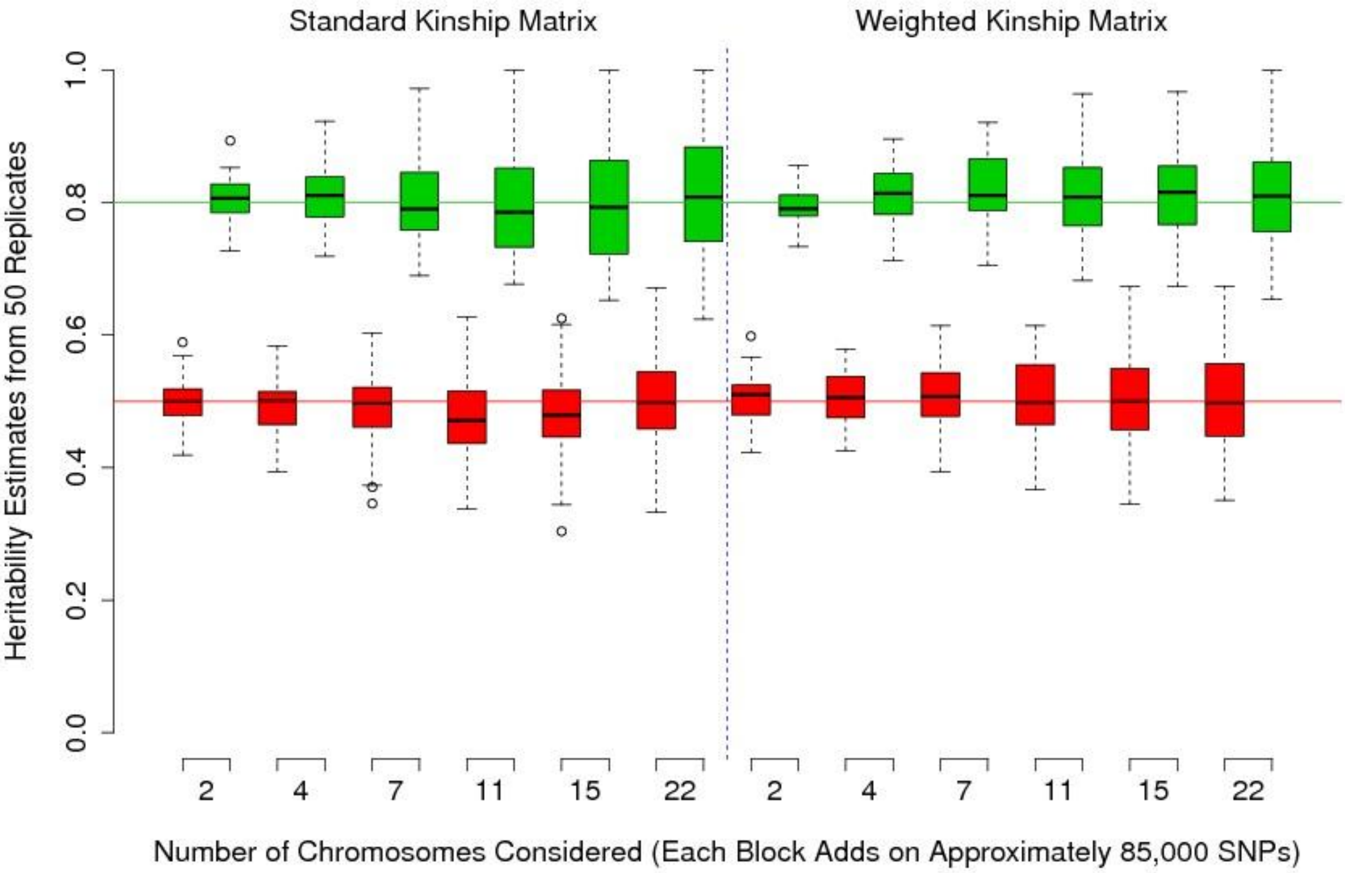

Figure S2. Increasing the Number of (Redundant) SNPs When Calculating Heritability

All heritability estimates presented in the main text were based on a kinship matrix calculated across only the 81,327 SNPs on Chromosomes 1 and 2. Here, we increased the number of SNPs across which allelic correlations were calculated. The boxes show the spread of h^2 estimates across 50 replicates for each scenario; their colors indicate the simulated h^2 (red: 0.5; green: 0.8). The *x*-axis indicates the number of chromosomes considered; each block corresponds to including approximately 85,000 extra SNPs, until the final one for which all 22 autosomal chromosomes (507,444 SNPs) were considered. With causal variants chosen only from the first two chromosomes, each block increases the number of redundant SNPs, and thus lowers the accuracy of using correlations over all SNPs as an estimate of correlations over just causal SNPs. Even so, the effect on the precision of estimates appears modest; even when the number of SNPs is increased six-fold, meaning that over 425,000 SNPs in no way tag any causal variation, mixed model analysis still provides reasonable accuracy, with the standard deviation only about twice that observed when just the first two chromosomes are considered.

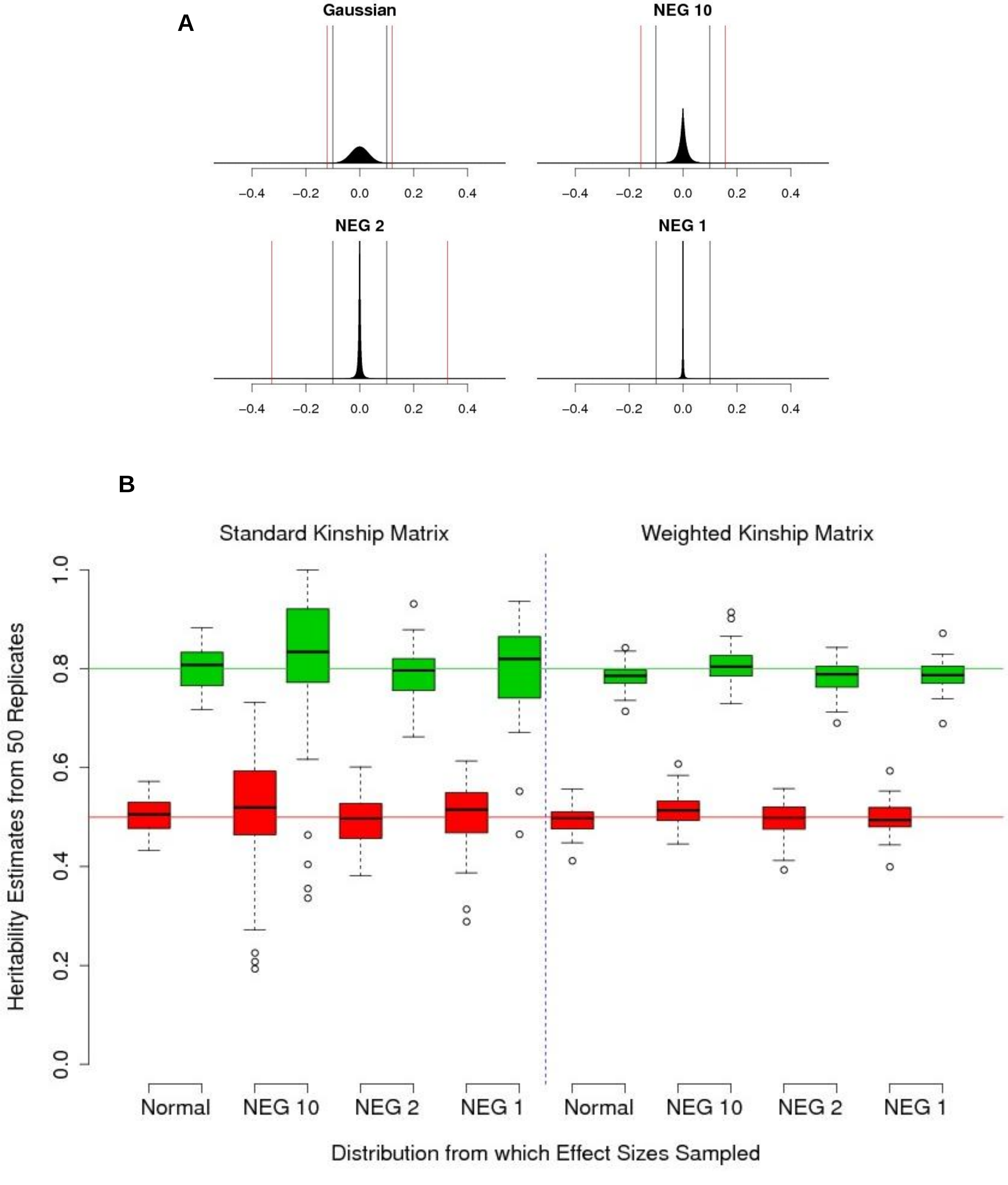

Figure S3. Distribution of Effect Sizes

We test the effect on estimation of h^2 when effect sizes of causal variants are sampled from distributions other than the Gaussian, that assumed by the standard linear mixed model. The Normal Exponential Gamma (NEG) can be viewed as a Laplace (exponential) distribution with rate drawn from a gamma distribution with fixed shape and scale parameters; decreasing the shape parameter increases the thickness of the tails (i.e. makes large magnitude effect sizes more likely). We considered three NEG distributions, with shape parameters 10, 2 and 1.

(A) How the densities of the three alternative distributions compare to the Gaussian. The black and red vertical lines indicate, respectively, the 0.1 & 99.9 percentiles and the 0.01 & 99.99 percentiles; for these plots the scale parameters were chosen so the black lines align across distributions. The distance between the black and red lines is a measure of the thickness of the tails (for NEG with shape 1, the tails are so thick that the red line cannot be shown on the *x*-axis presented here, and also the *y*-axis is truncated).

(B) The spread of h^2 estimates for each distribution; colors indicate simulated h^2 (red: 0.5; green: 0.8). Their precision seems only modestly affected by wrongly assuming a Gaussian distribution of effect sizes, and the weighted kinship matrix provides greater precision.

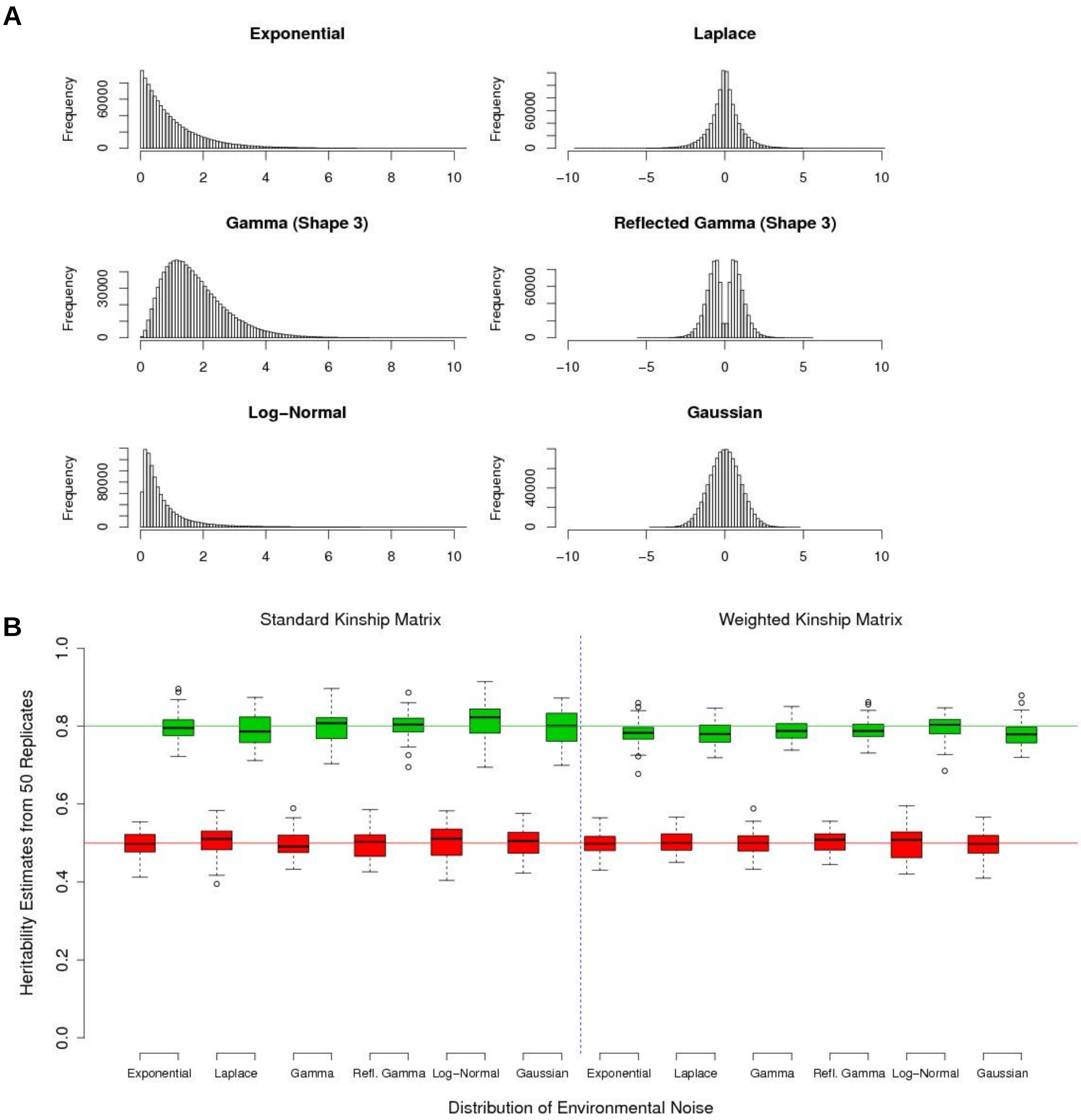

Figure S4. Distribution of Noise Terms

(A) How the densities of the five alternative distributions we used for generating noise terms compare to that assumed by the mixed model, the Gaussian distribution. In particular, the Exponential and Gamma distributions have heavier tails than the Gaussian, making larger noise terms more likely. It should not matter than three of the distributions generate positive values only, as it would make no difference if a constant value was subtracted off all noise terms to produce negative values as well.

(B) The results of mixed model analysis using each of the six noise distributions, where once again colors indicate the simulated h^2 (red: 0.5; green: 0.8). Despite the marked differences between the shapes of the five alternative distributions and the Gaussian, the estimates of h^2 still appear reliable, with once more, our weighted kinship matrix providing slightly more precision.

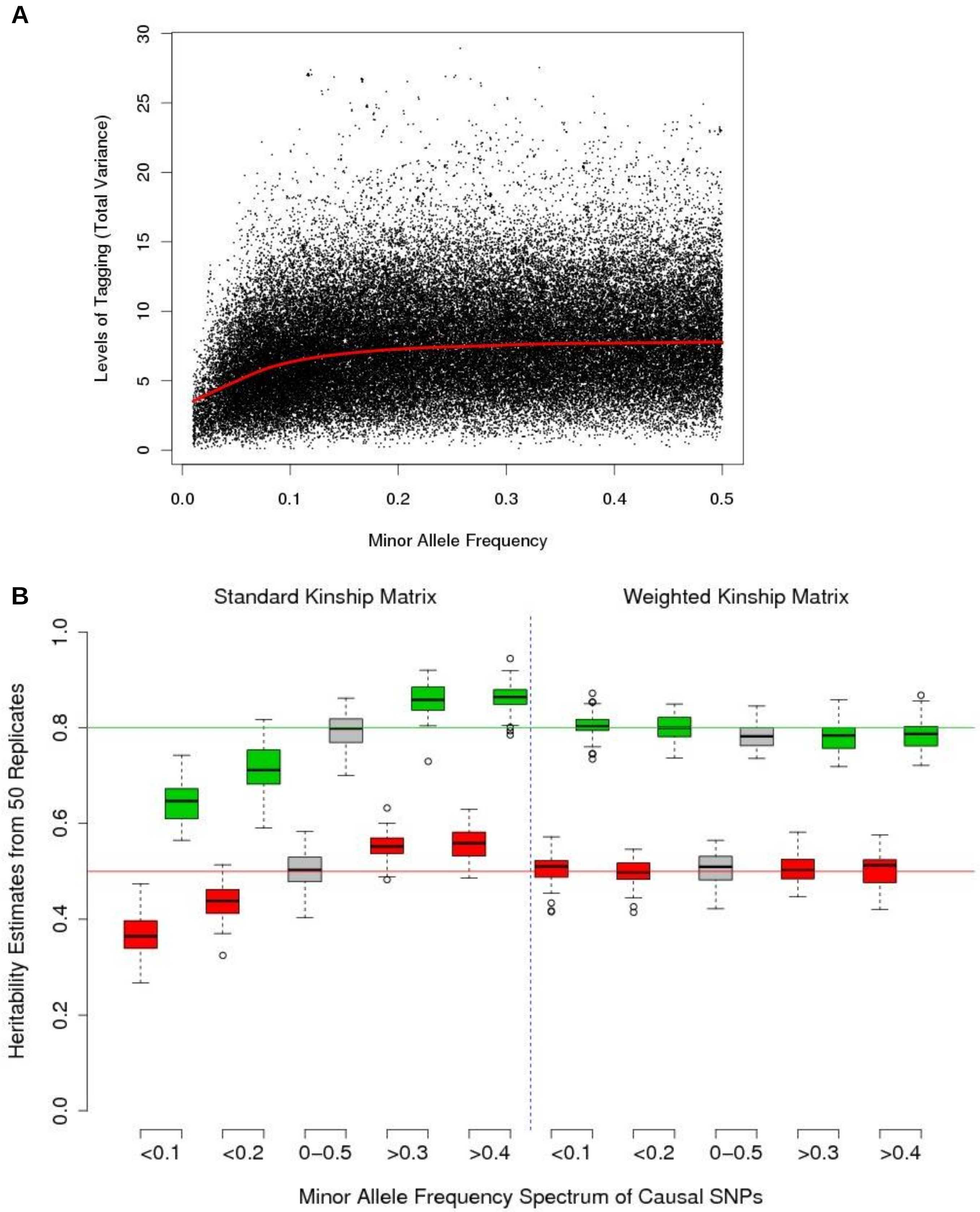

Figure S5. Relationship between Tagging and MAF

(A) The trend for lower frequency variants is more poorly tagged (to avoid overcrowding, only 2000 of the 507,444 points have been plotted). The red line corresponds to LOWESS regression across all

507,444 SNPs, its positive gradient, especially apparent for SNPs with MAF<0.1, indicates that more common variants experience, on average, higher levels of LD with their neighbors.

(B) The distribution of h^2 estimates; colors indicating the simulated h^2 (red: 0.5; green: 0.8). As a consequence of the relationship between tagging and LD, when causal variants are of either higher or lower than average frequency, mixed model analysis using the standard kinship matrix will tend to under- or over-estimate h^2 (left half). However, this bias is neutralized using our LD-adjusted matrix (right half).

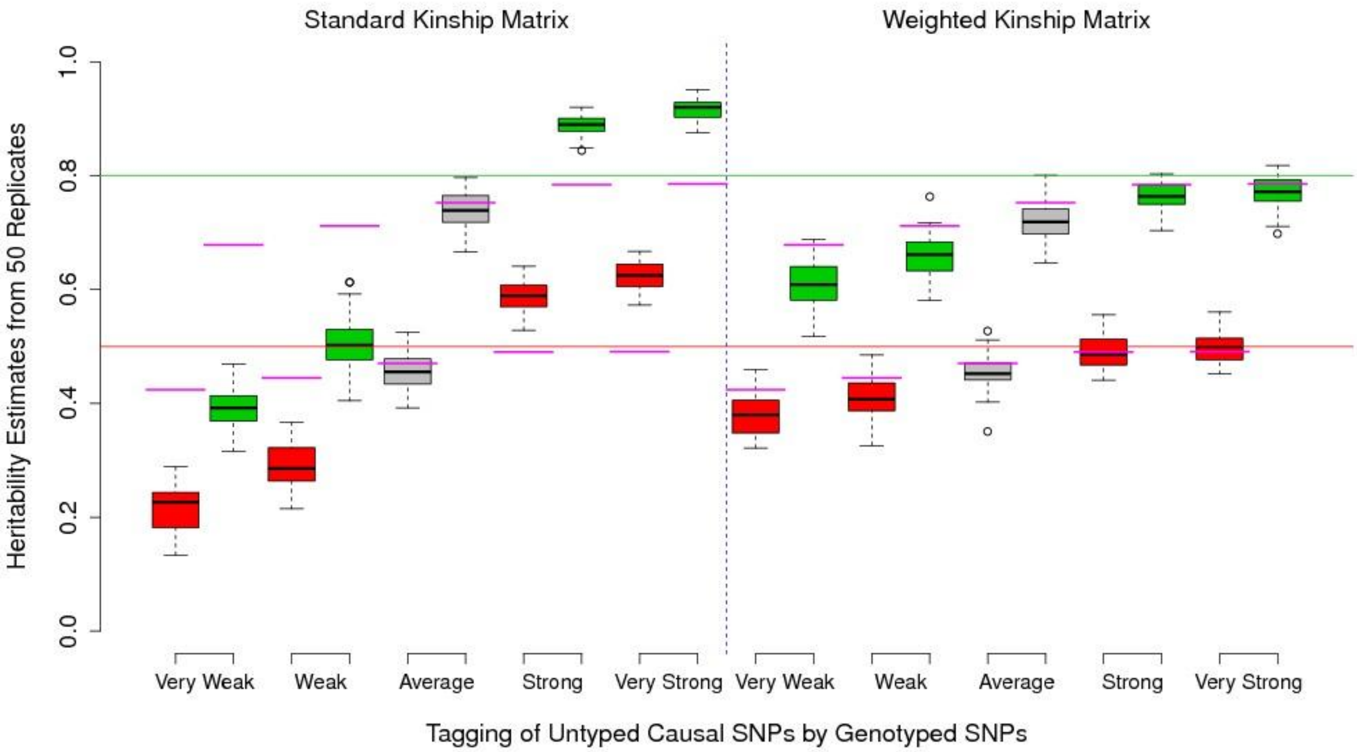

Figure S6. Recognizing Heritability Contributions from Untyped Causal Variants

By masking every alternate SNP in our simulation data, we considered scenarios where the causal variants were untyped, so were only partially tagged by the genotyped SNPs. For each untyped SNP, we calculated the proportion of its variance that could be explained by a linear combination of all genotyped SNPs within 1Mbp. We then divided SNPs according to these proportions: weakly (very weakly) tagged SNPs were those with values in the bottom 40% (20%), while strongly (very strongly) tagged SNPs were those with values in the top 40% (20%). Intuitively, we would expect the proportion of heritability from an untyped causal variant that is recognized by mixed model analysis, to depend on the proportion of that SNP's variance which is tagged by the genotyped SNPs. The boxes show the spread of h^2 estimates, where color indicates the simulated h^2 (red: 0.5; green: 0.8). For each box, the horizontal purple line indicates how much heritability we would expect to recognize based on the average tagging of the SNPs considered causal. For example, the gray box corresponds to causal variants picked at random from all untyped SNPs; on average, an untyped variant will have 88% of its variance tagged by the genotyped SNPs, so we would expect 88% of the total heritability to be captured (0.44 if h^2=0.5 or 0.70 if h^2=0.8). We see that when using the standard kinship matrix (left half) the average heritability captured can be noticeably higher or lower than these values. This is because, for example, when the causal SNPs are poorly tagged, the genotyped SNPs tagging these are themselves on average poorly tagged relative to other genotyped SNPs, so mixed model analysis using standard kinships is liable to under-estimate their heritability contribution. By contrast, when using our LD-adjusted kinship matrix (right half) the proportion of total heritability recognized more closely matches the proportion we would expect. We believe that the reason why the realized heritability is slightly less than the expected heritability owes to the fact that each untyped SNP will have different components of its variation tagged by different genotyped SNPs. When we carried out a study where each untyped variant was tagged by only one genotyped SNP, the realized heritability matched exactly the value we expected (results not shown).

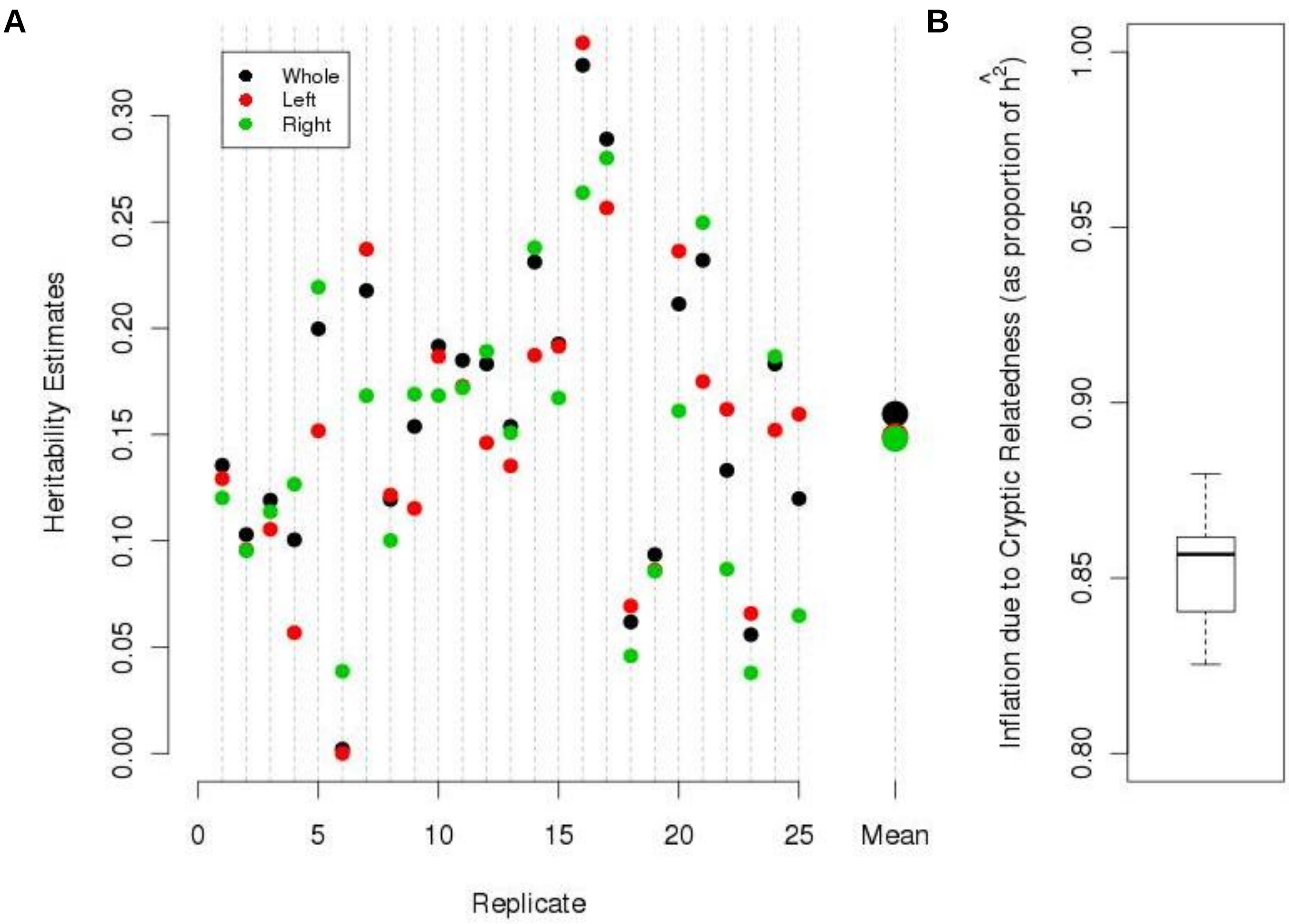

Figure S7. Testing for Inflation Due to Cryptic Relatedness

Browning and Browning[2] constructed a toy example, derived from the WTCCC control datasets,[3] to demonstrate that heritability estimates could be dramatically affected by population structure. We recreate their example, but show how genomic partitioning[4] can be used as a simple test to assess the inflation of h^2 estimates caused by population structure (or more accurately, cryptic relatedness, as residual relatedness between individuals can also contribute to this inflation). The test involves estimating first h^2_L, the heritability from the "left half" of the genome (say, Chromosomes 1-8), then h^2_R, the heritability from the remaining chromosomes, then the total heritability. If the estimates are accurate (i.e. there is no inflation), we would expect the sum of the first two estimates to equal the third, as the estimate from a particular region should only pick up heritability contributions from the causal variants it contains. However, the presence of cryptic relatedness will induce long-range genome sharing, so that causal variants from, say, the left half, will be correlated with variants on the right half, and thus the estimate of h^2_L will include more than just the contribution of left half causal variants. As both the left and right halves of the genome should be sufficiently long to capture the full effects of cryptic relatedness (as these effects should be relatively strong), the three estimates should be inflated equally. Therefore, it is possible to estimate inflation by subtracting the estimate for h^2 from the sum of the estimates for h^2_L and h^2_R.

(A) Plots of the three estimates for 25 replicates (black points for total; red points for left half; green points for right half). As found by Browning and Browning[2] we also get large heritability estimates, which considering how the dataset was constructed, must result entirely from the population-specific ascertainment bias rather than true causal variation. However, it is also clear that there is very little difference between the estimates from the whole genome and those from either half.

(B) The derived estimate of inflation (“left + right - whole”) as a proportion of the total heritability estimate. If we were to obtain an estimate of inflation close to zero, we could be satisfied that cryptic relatedness was not significantly contributing to the heritability estimates. By contrast, here it is clear that inflation is responsible for almost all the observed heritability, signaling that the experimental design was flawed and the estimates were not to be trusted. We make use of this test when analyzing the WTCCC data.

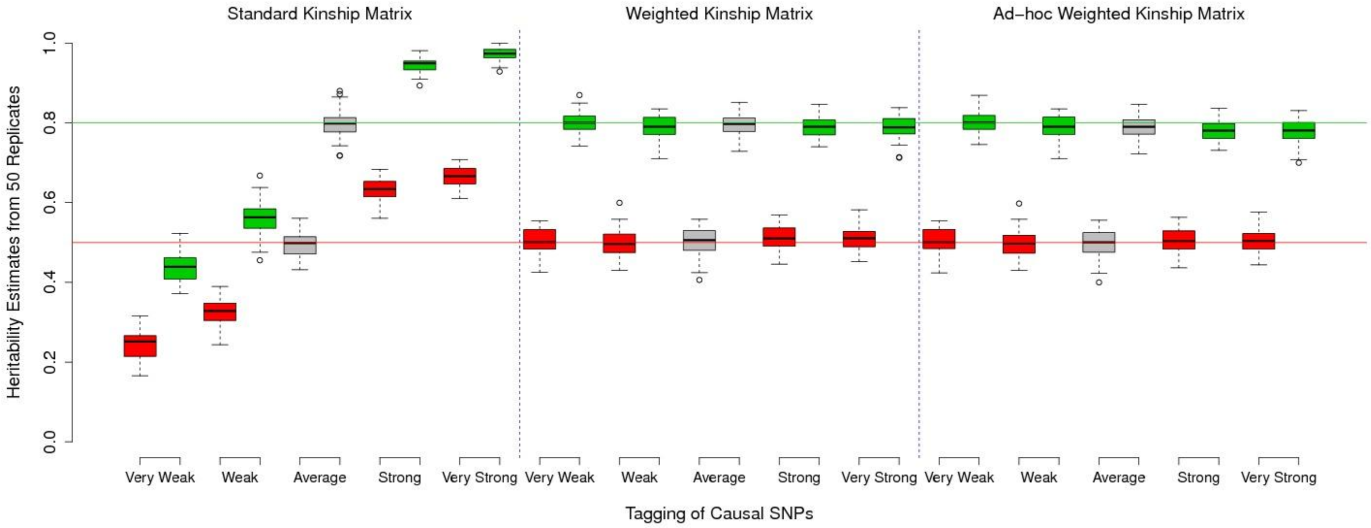

Figure S8. Guarding against Genotyping Errors

Heritability estimates will be inflated by differential genotyping errors. This can be especially problematic when analyzing binary outcomes when cases and controls have been genotyped separately. Our weightings have the potential to increase this inflation, as a poorly genotyped SNP will typically display lower levels of LD with its neighbors, so receive a higher weighting in calculation of the LD-adjusted kinship matrix. To protect against this occurrence, when computing $C(j,j')^2 = e^{-\lambda d_{jj'}} r^2_{jj'}$, the weighted correlation between a pair of SNPs, we suggest calculating $r^2_{jj'}$ across cases and controls separately, then setting $C(j,j')^2$ equal to the higher of the two values observed. In this way, if a genotyping error in, say, the case samples has caused SNP *j* to be poorly tagged by its neighbors, provided this genotyping error has not affected the control samples also, a realistic correlation squared value can be obtained from those and the weighting should not be adversely affected. The first two blocks of this figure are identical to those in the main text, demonstrating how use of our LD-adjusted kinship matrix corrects for the biases introduced when causal variants come from areas of lower or higher than average tagging. There does not appear to be a drop in performance when using the ad-hoc LD-adjusted kinship matrix described above (third block). Therefore, when analyzing a binary outcome where subsets of samples have been genotyped separately, we recommend use of this fix. An even more conservative approach, if resources are available, is to calculate the weightings using an entirely independent genotype dataset. Assuming once more that it is unlikely genotyping errors in the independent dataset will coincide with those in the dataset under consideration, then the weightings for unreliable SNPs should not artificially be inflated. None-the-less, even if using one of these fixes, it remains crucial that thorough quality control is performed before analysis.

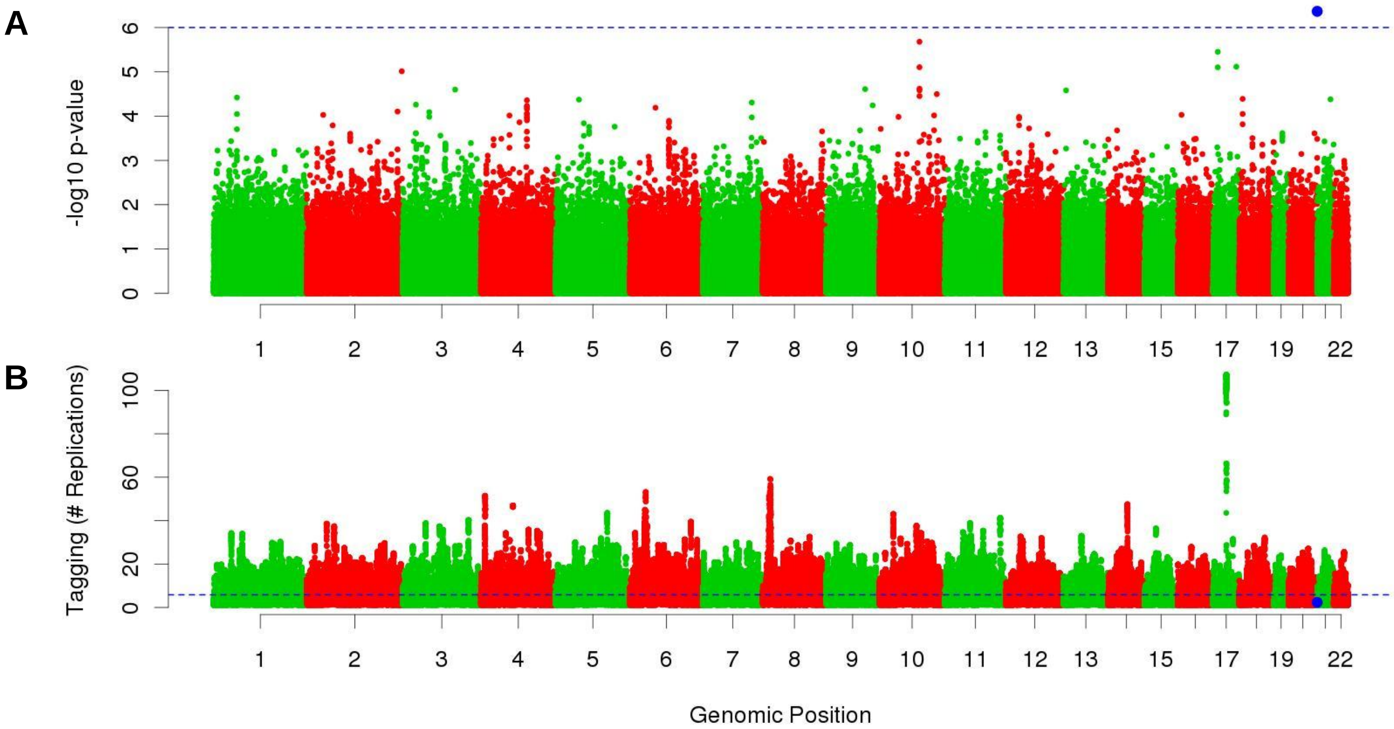

Figure S9. Pseudo Case-Control Study – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

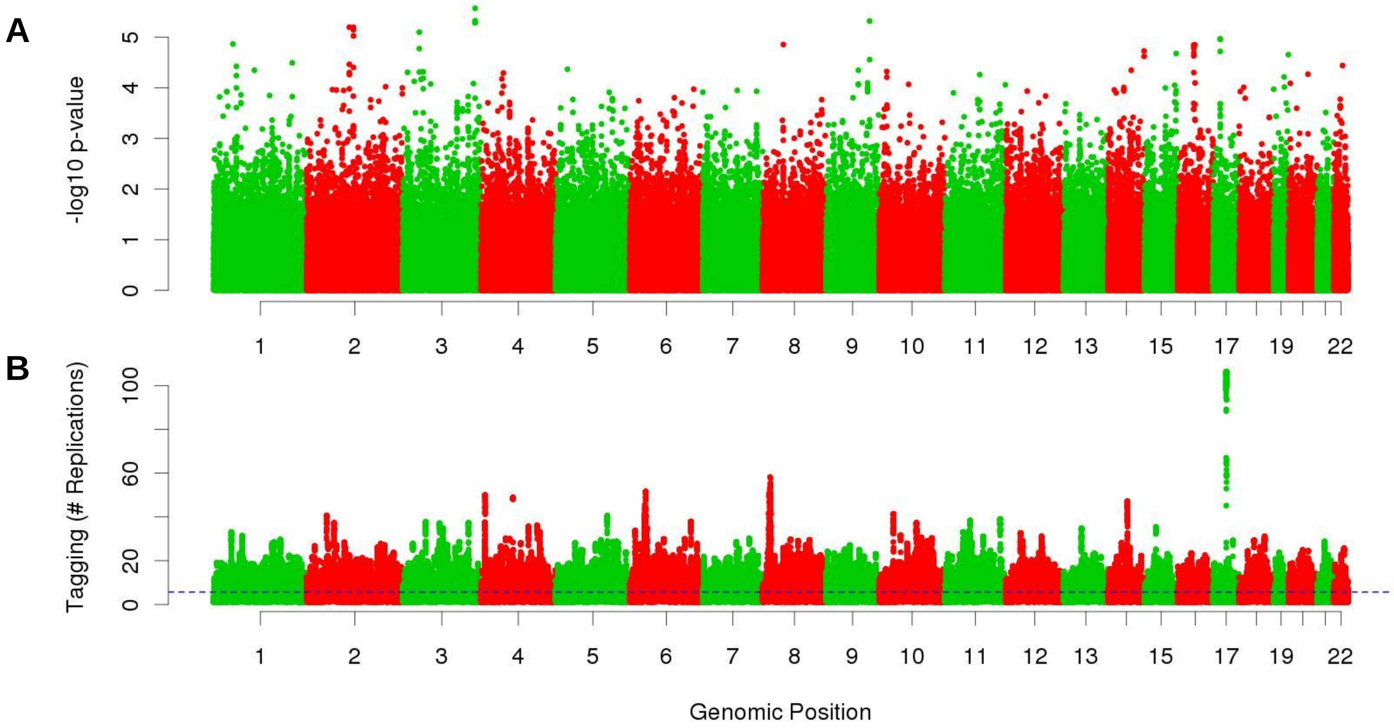

Figure S10. Bipolar Disorder – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

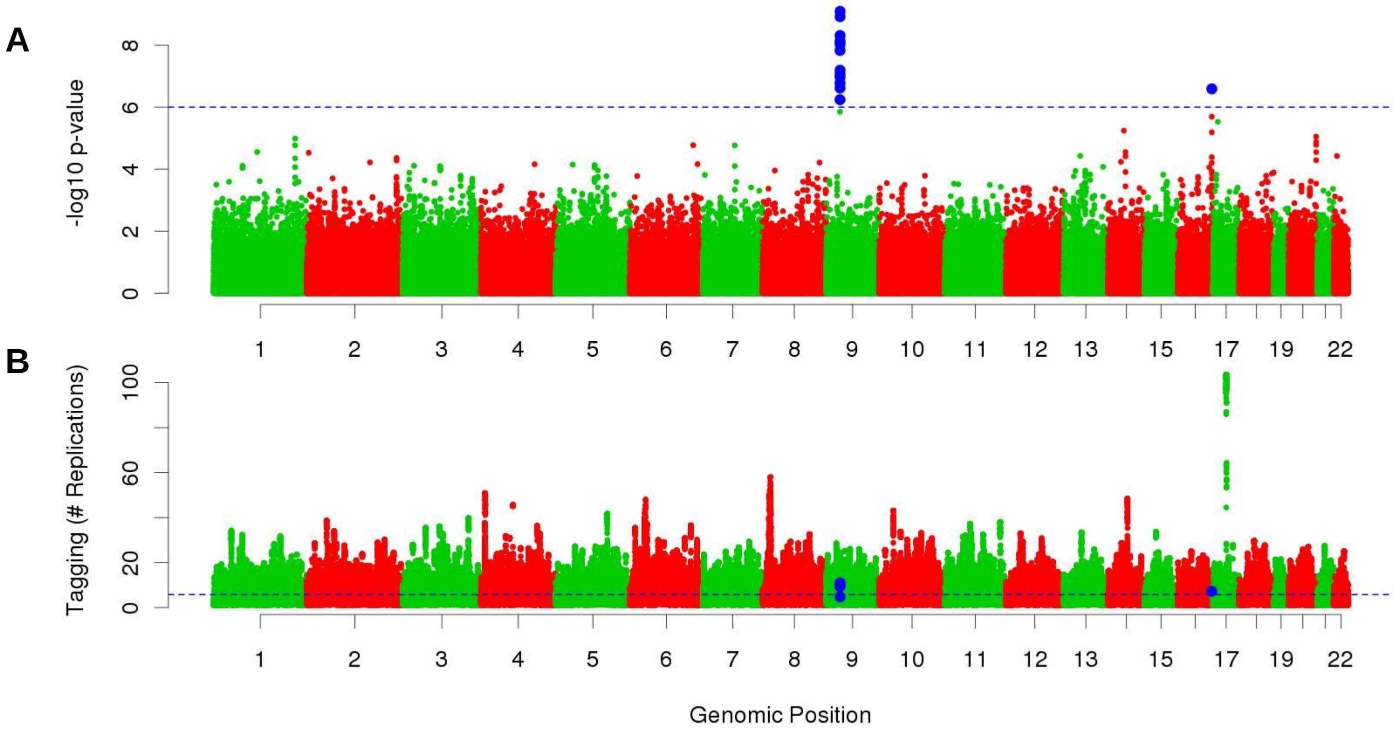

Figure S11. Coronary Artery Disease – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

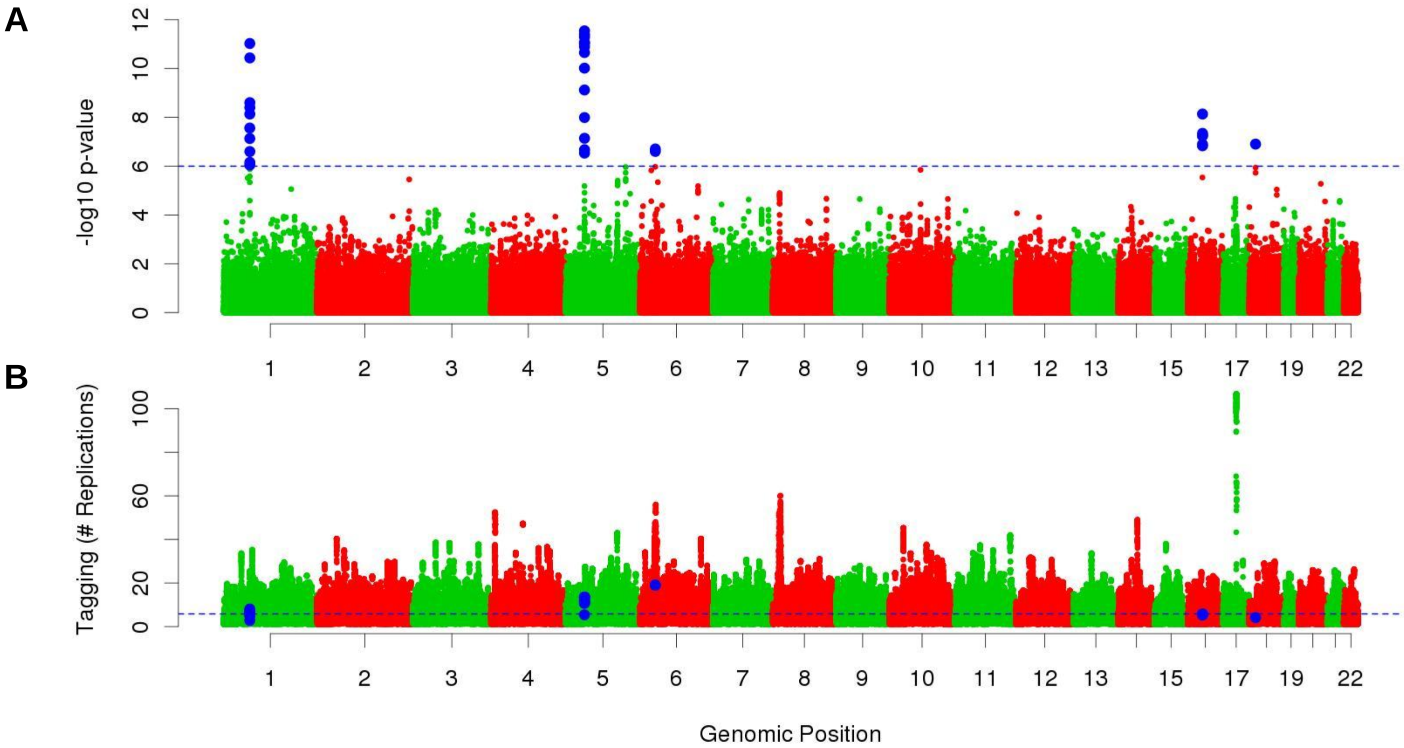

Figure S12. Crohn Disease – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

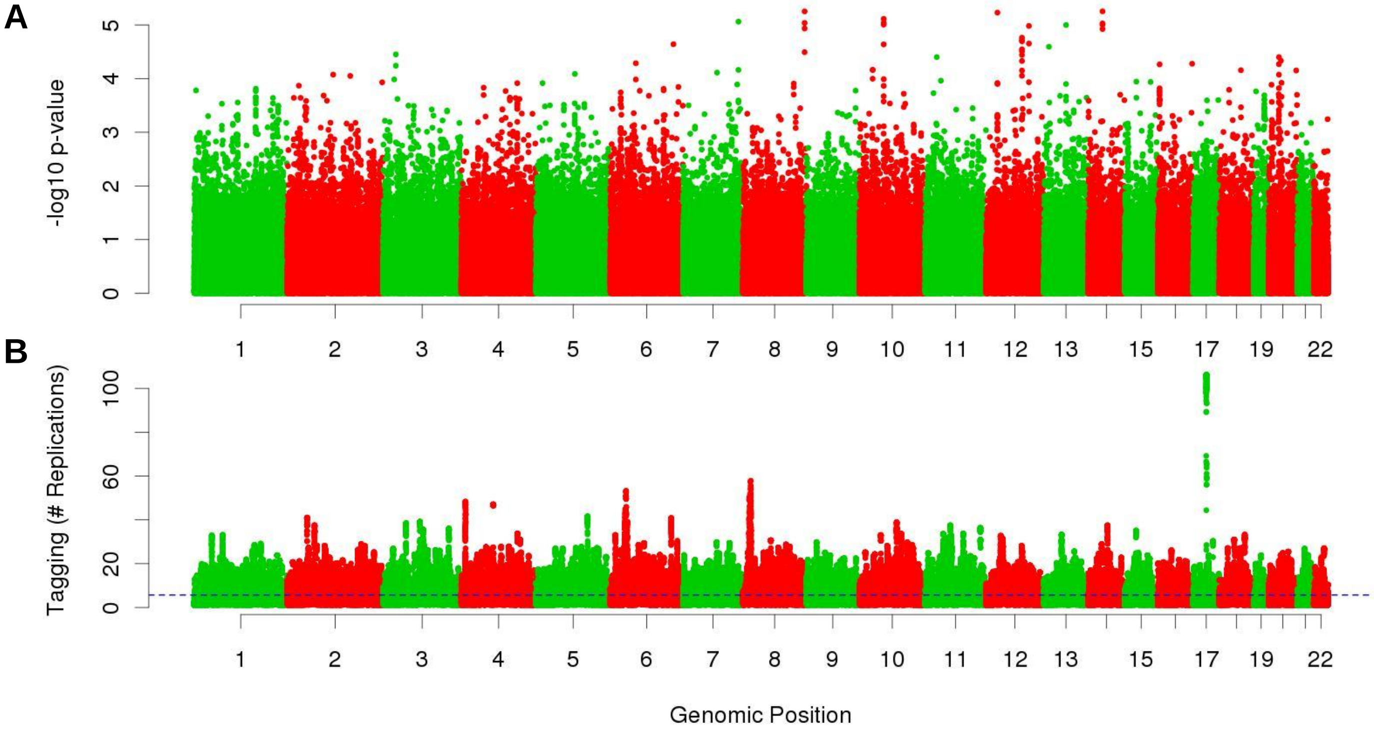

Figure S13. Hypertension – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

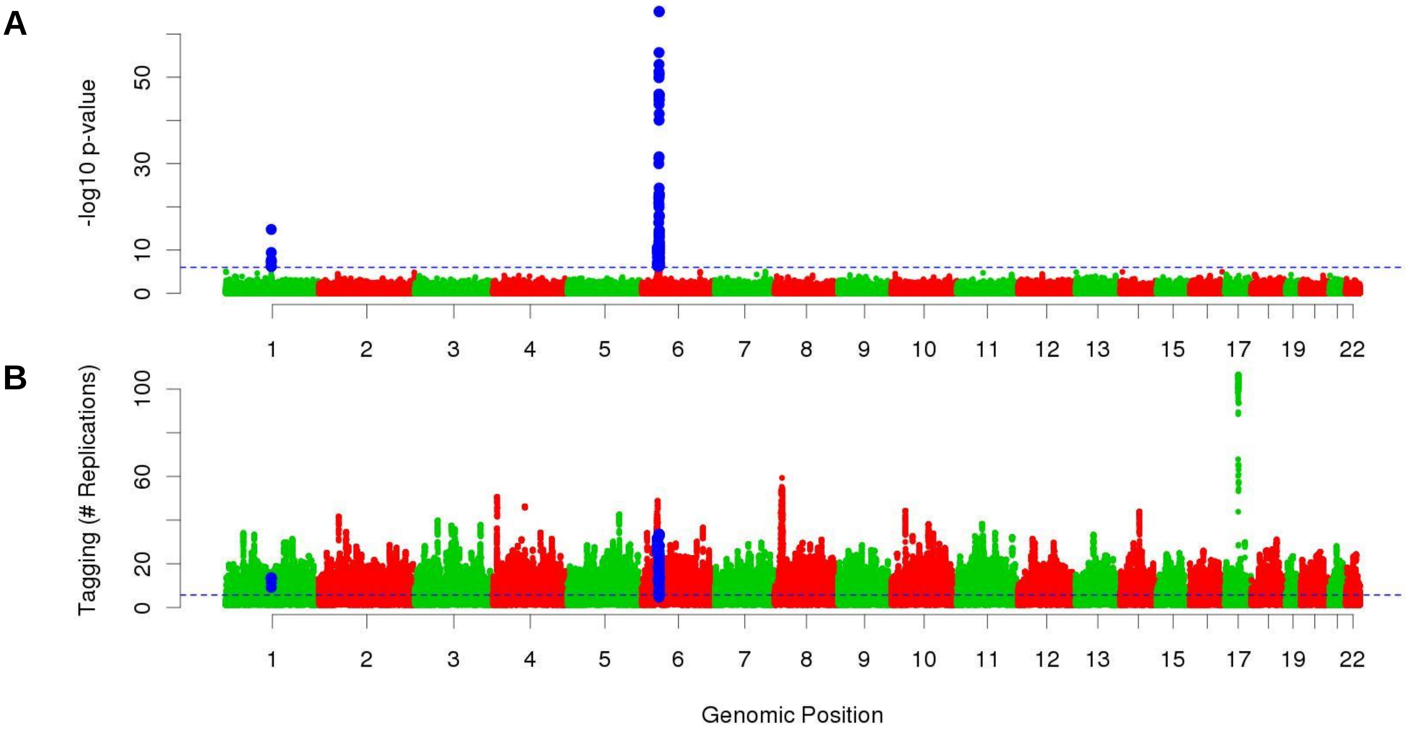

Figure S14. Rheumatoid Arthritis – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

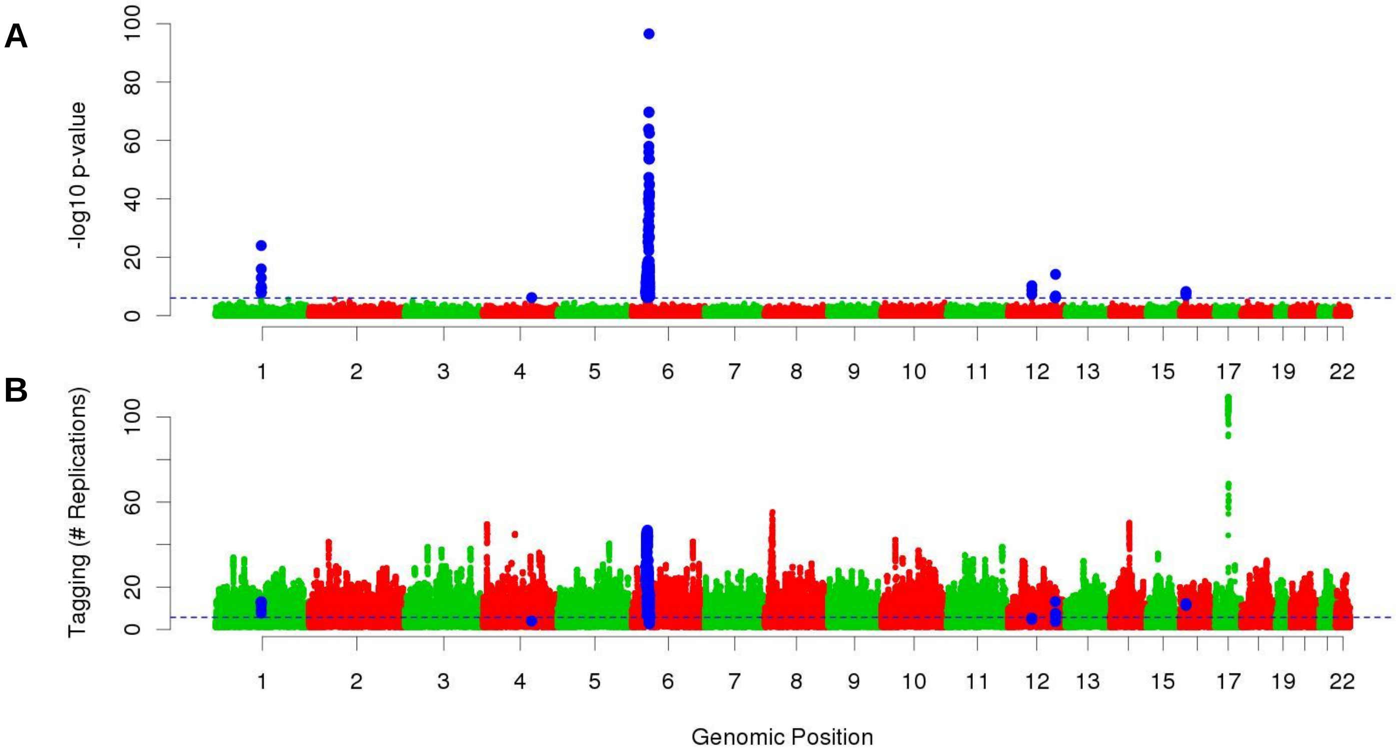

Figure S15. Type 1 Diabetes – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

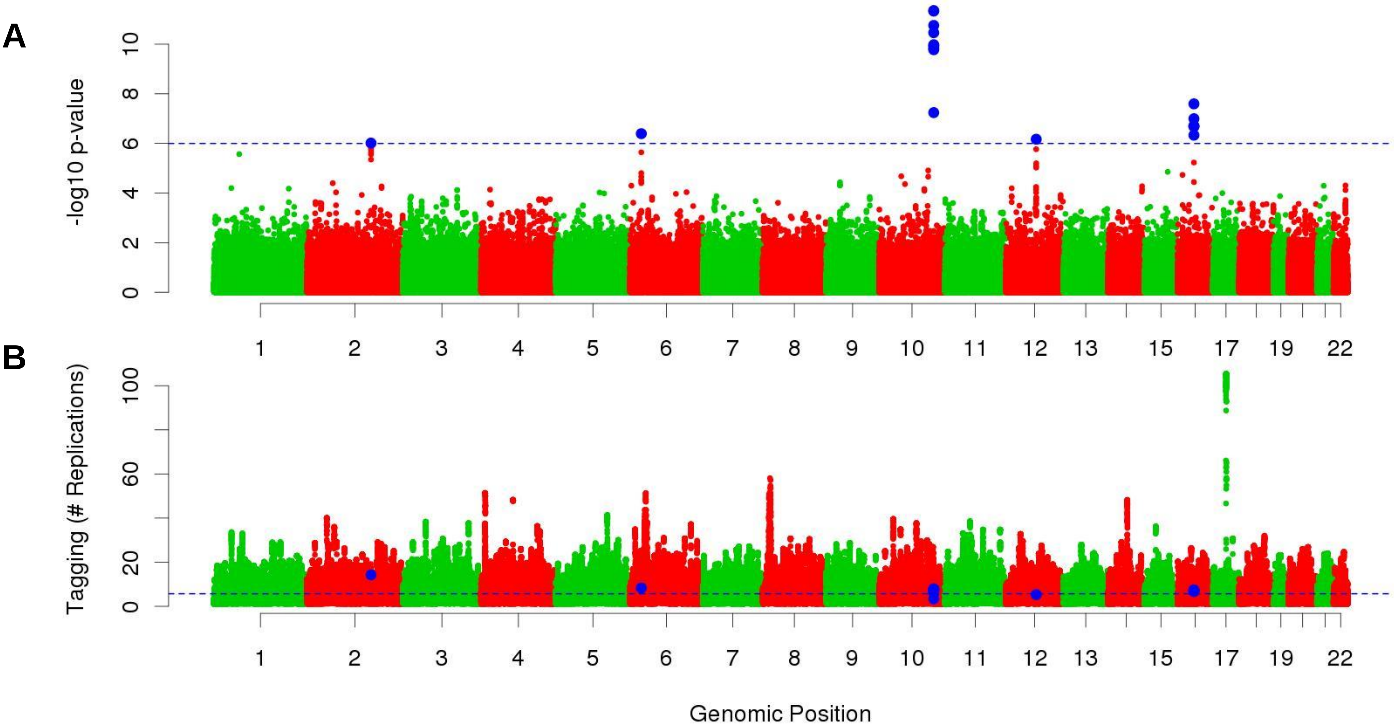

Figure S16. Type 2 Diabetes – Tagging of Top Hits from Association Study

Shown in (A) are the –log10 p-values from testing each SNP marginally for association. SNPs significant at 10^{-6} are marked in blue. The coloring carries over to (B), which shows the variable tagging of SNPs throughout the genome and indicates the tagging of the top hits.

Table S1. Full Results from Analysis of WTCCC[3] Data

Trait	n	m’	Total Heritability (SD)			Chromosome 6 Heritability (SD)		
			Standard	Weighted	Difference	Standard	Weighted	Difference
Pseudo Case-Control	2947	297,894	11 (10)	7 (14)	-4	0 (2)	1 (3)	+1
Bipolar Disorder	4802	278,772	59 (6)	69 (8)	+10	5 (2)	6 (2)	+1
Coronary Artery Disease	4858	282,170	39 (6)	41 (8)	+3	2 (2)	1 (1)	-1
Crohn’s Disease	4709	285,989	54 (6)	58 (8)	+5	5 (2)	6 (2)	+1
Hypertension	4878	280,461	42 (6)	52 (8)	+10	5 (2)	8 (2)	+4
Rheumatoid Arthritis	4777	281,401	57 (6)	52 (8)	-6	19 (2)	17 (2)	-2
Type 1 Diabetes	4913	282,212	73 (6)	74 (8)	0	37 (2)	35 (2)	-2
Type 2 Diabetes	4843	280,653	35 (6)	44 (8)	+9	4 (2)	5 (2)	+1

Trait	Left Half Heritability	Right Half Heritability	Cryptic Rel. Inflation, P	Heritability Based on Call Rate Tranches			
				Quarter 1	Quarter 2	Quarter 3	Quarter 4
Pseudo Case-Control	0	13	2	1	13	1	17
Bipolar Disorder	30	29	1	33	45	23	43
Coronary Artery Disease	17	22	1	25	20	29	32
Crohn’s Disease	33	24	4	31	40	41	42
Hypertension	21	23	3	26	31	30	33
Rheumatoid Arthritis	41	16	0	43	41	41	41
Type 1 Diabetes	59	16	1	56	46	47	56
Type 2 Diabetes	20	16	1	19	21	29	28

n and m’ denote the numbers of individuals and SNPs, respectively. To calculate P, the extent by which the estimate of total heritability is inflated due to cryptic relatedness, we separately computed heritability from “Left Half” SNPs (Chromosomes 1-8) and “Right Half” SNPs (Chromsomes 9-22). We would expect each of these two estimates to also be inflated by an amount P. Therefore, we can estimate P as the Left Half heritability plus the Right Half heritability, minus the Total Heritability. The figures would suggest this inflation is modest.

To assess the inflation due to genotyping errors, we divided the SNPs into four quarters according to call rate – which statistic we treated as a proxy for SNP reliability – then estimated the heritability from each subset of SNPs separately. If genotyping errors were cause for concern, we would expect the heritability estimates from the less reliable SNPs (Quarters 3 and 4) to be typically greater than those from the more reliable SNPs (Quarters 1 and 2). As no trend is noticeable across studies, this suggests the inflation caused by genotyping errors is at most slight.

References

1. Zou, F., Lee, S., Knowles, M., and Wright, F. (2010). Quantification of population structure using correlated SNPs by shrinkage principal components. Hum. Hered. *70*, 9-22.
2. Browning, S. and Browning, B. (2011). Population structure can inflate SNP-based heritability estimates. Am. J. Hum. Genet. 89, 191-193.
3. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661-678.
4. Yang, J., Manolio, T., Pasquale, L., Boerwinkle, E., Caporaso, N., Cunningham, J., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M., et. al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519-525.