Fig S1: Number of SNP-DNAm site pairs in *cis* (A) and *trans* (B).
In the first phase, 22 cohorts performed a mQTL analysis and saved their results for p<1e-5. We found 102,965,711 associations in *cis* and 710,638,230 associations in *trans* at least one cohort. 61,103,065 *cis* associations and 17,246,702 M *trans* associations were found in at least two cohorts.

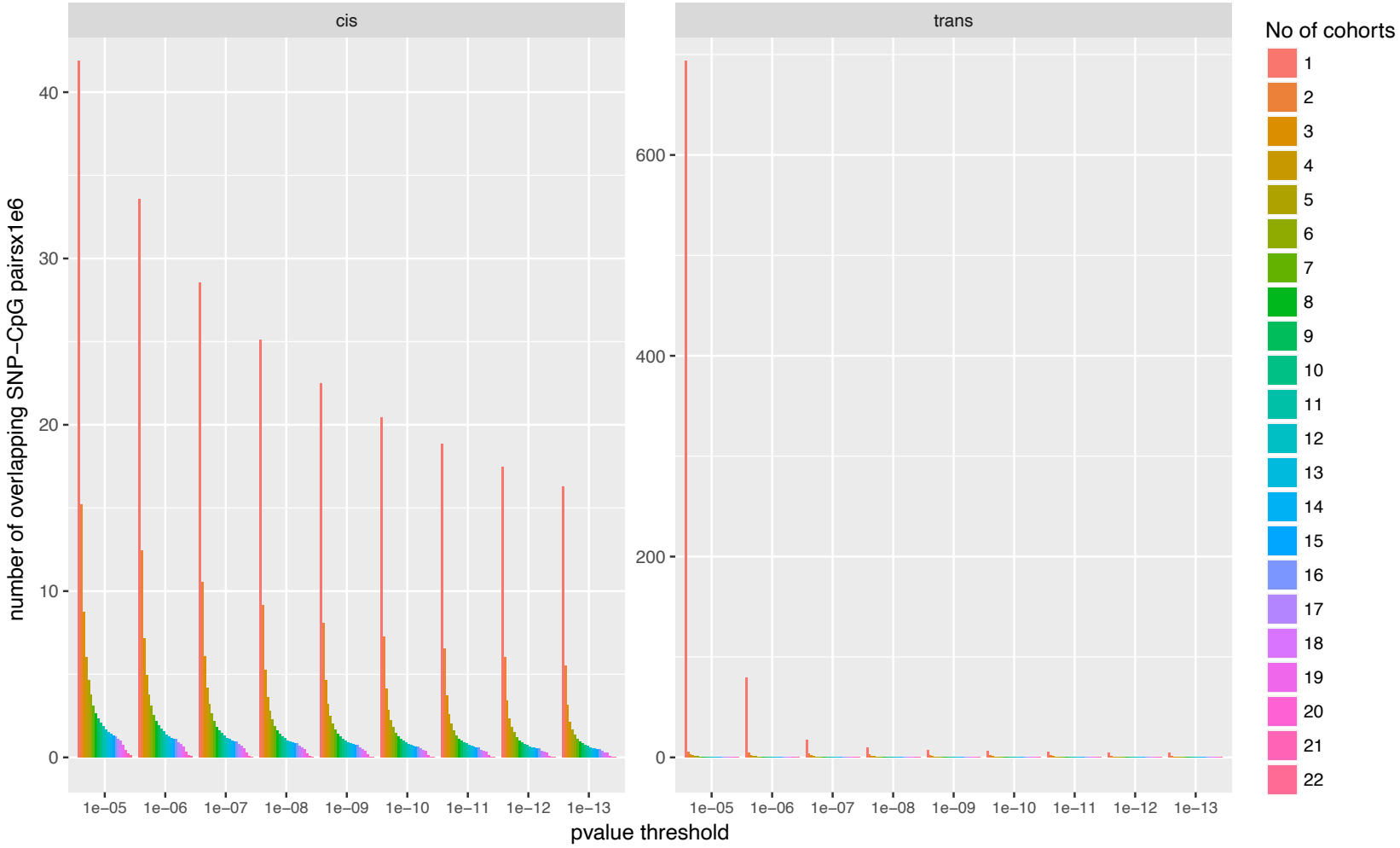
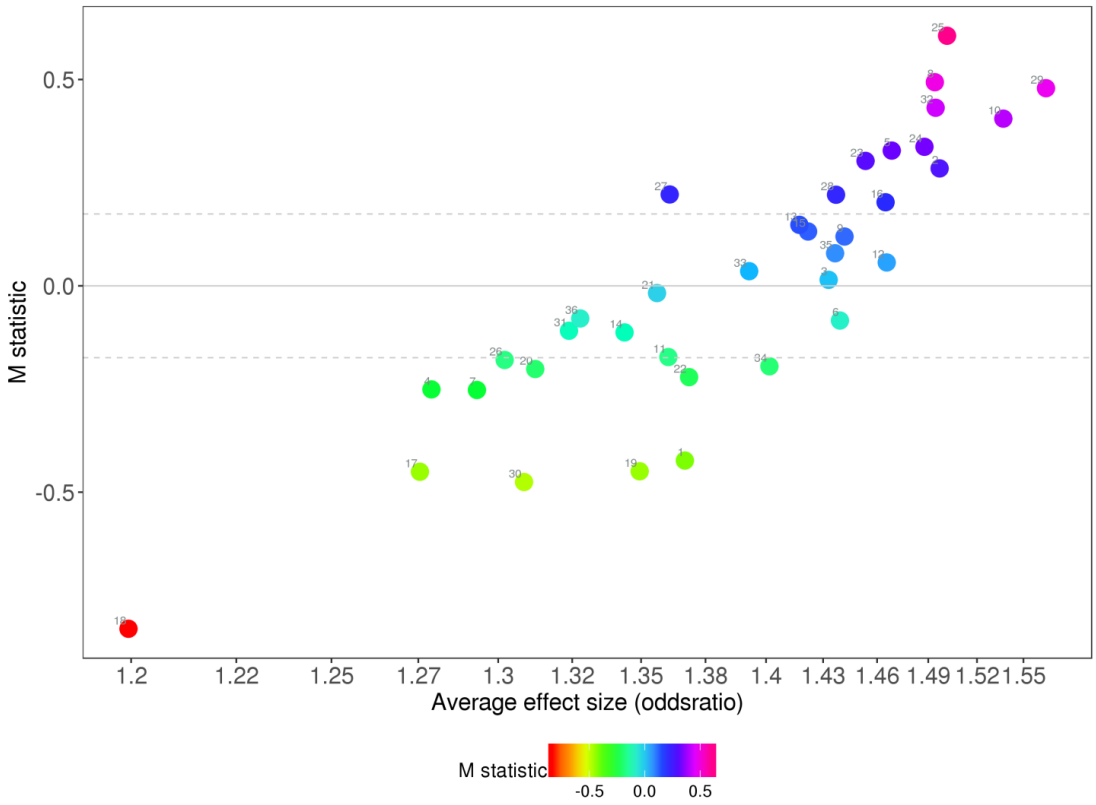


Fig S2: Quality control of 36 studies. We used 337 independent SNPs on chromosome 20 with a pvalue<1e-14 to calculate the Mstatistic[1] for each cohort.



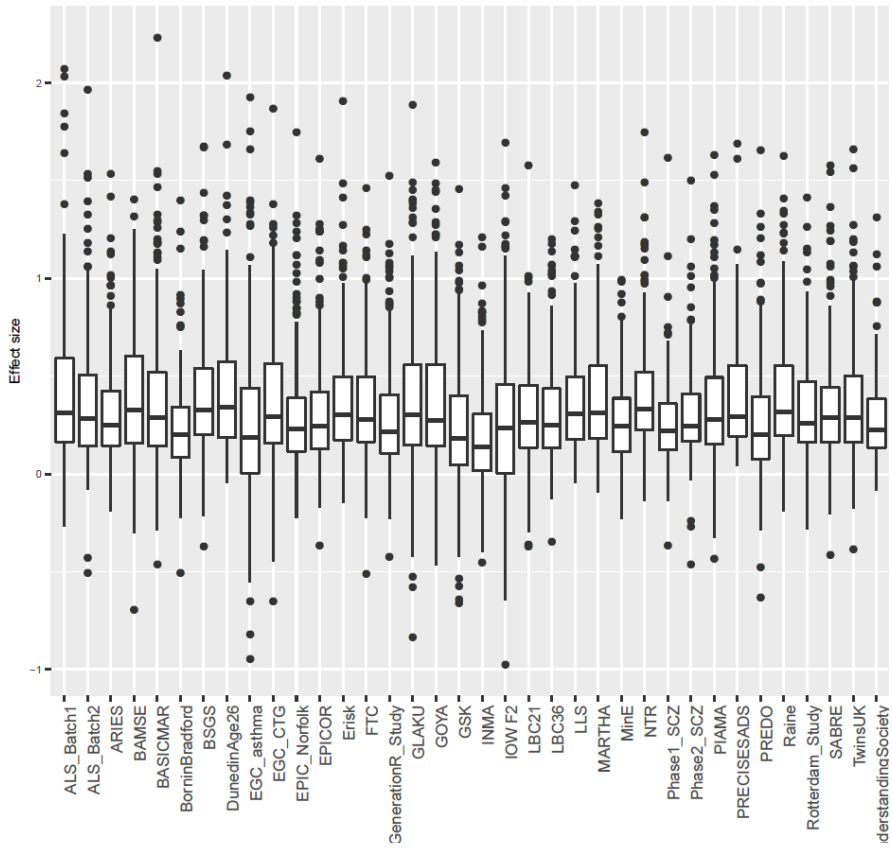


Fig S3: Fixed effects meta analyses A. Distance of SNP from DNAm site against the -$\log_{10}$ Pvalue of the *cis* association. B. mQTL counts vs effect size direction for each study. The counts are high if more studies agree on effect size direction. C. I^2 vs Direction of effect size for each study. D. Regression coefficient vs I^2 category. A correlation has been found between I^2 and the regression coefficient.

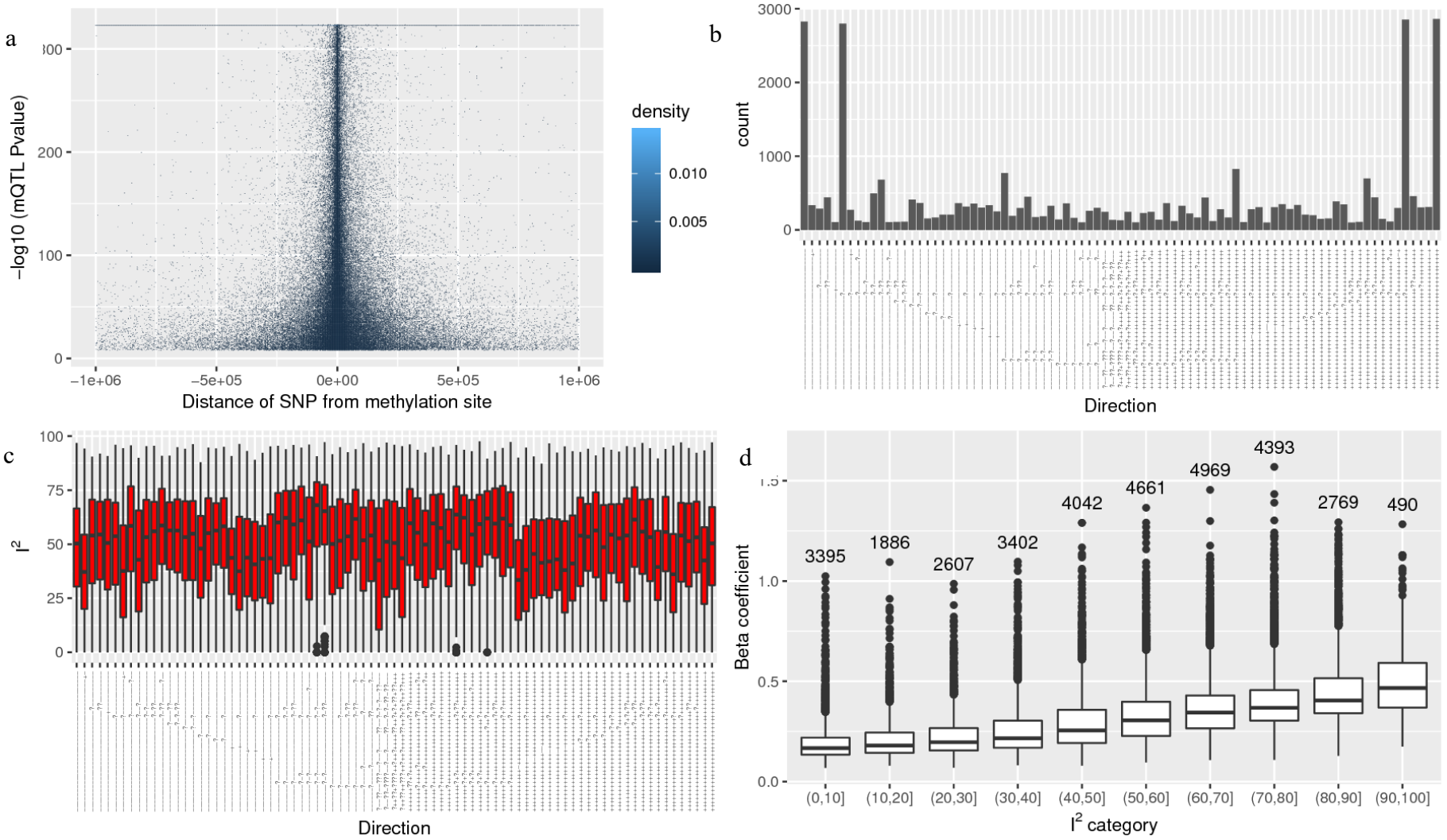


Fig S4: Number of mQTL by pvalue threshold identified in fixed effects meta analyses (Left) and multiplicative random effects analysis (Right) in 36 cohorts.

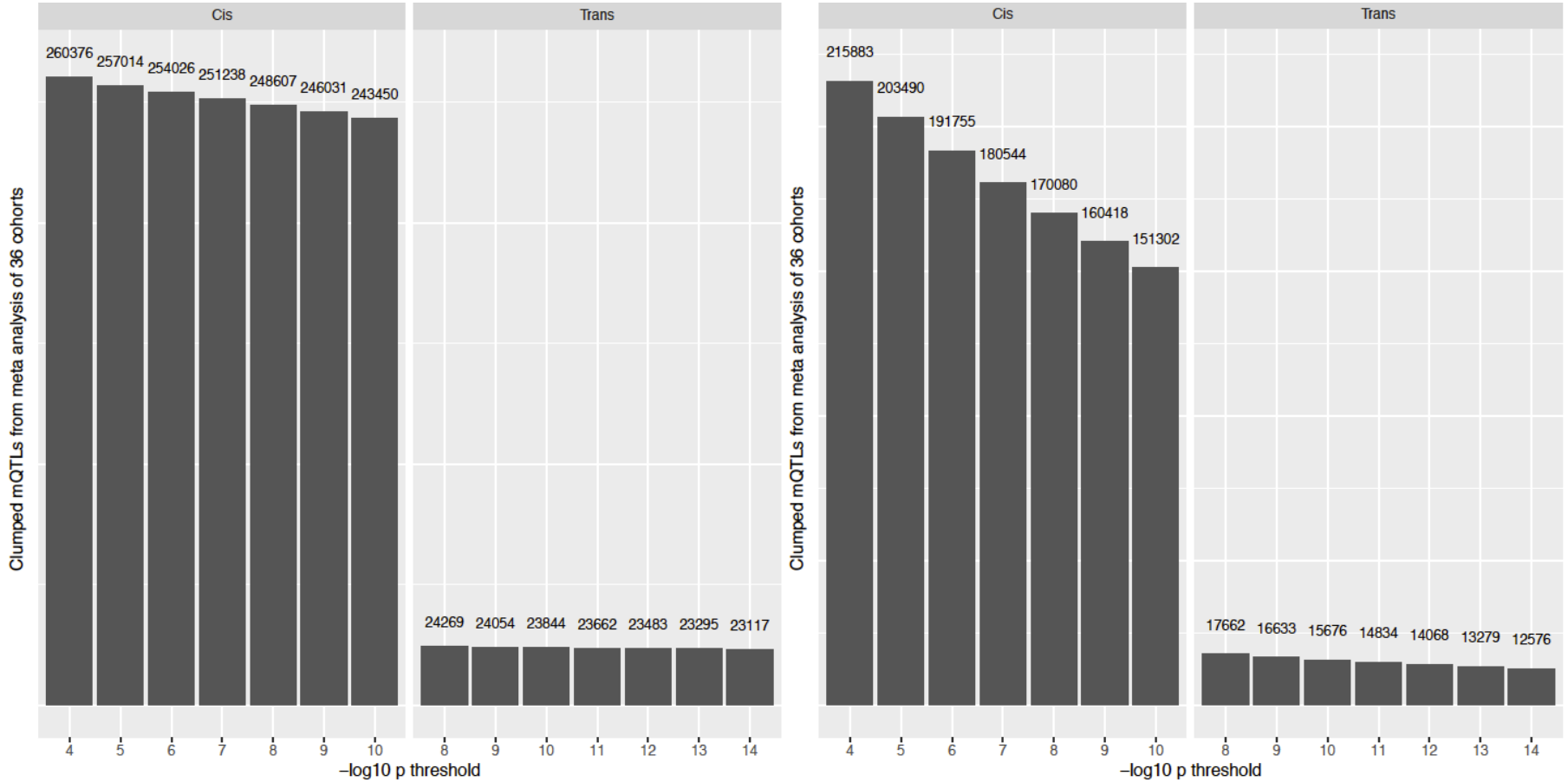


Fig S5: Fixed versus multiplicative random effects meta analyses. Boxplots show statistical properties of associations found with both fixed effects and multiplicative random effects meta analyses (FE+ MRE+) and associations found with fixed effects meta analyses only (FE+ MRE-). Associations with high I^2, tau^2 drop out in MRE. Associations with smaller effect sizes or DNAm site standard deviations or r^2 drop out in MRE.

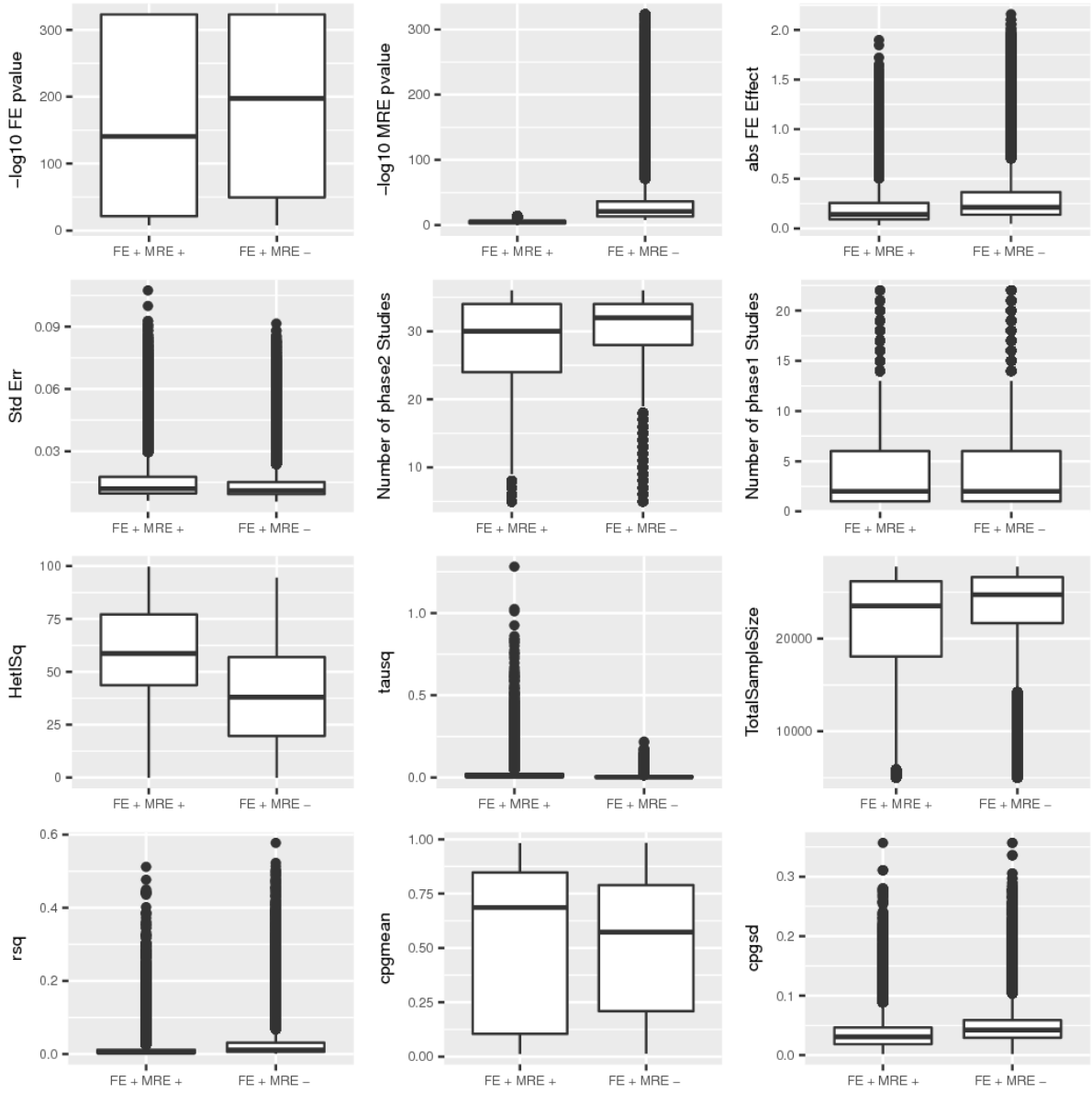


Fig S6: Left: Effect sizes and weighted standard deviation (SD) for each mQTL category. For each DNAm site, the strongest absolute effect size (the maximum absolute additive change in DNAm level measured in SD per allele) was selected**.** Effect sizes were shown for all sites with a mQTL, sites with *cis only* effects, *cis* effects for sites with *cis* and *trans* effects, *trans* effects for sites with *cis* and *trans* effects and sites with *trans only* effects. Comparing the strongest effect size for each site showed that *cis+trans* sites had larger *cis* effect sizes (per allele SD change = 0.05 (s.e.= 0.002), p<2e-16) as compared to *cis only* sites and weaker *trans* effect sizes (per allele SD change = -0.06 (s.e.= 0.002), p<2e-16) as compared to *trans only* sites. To detect these small *trans* effect sizes at sites with both a *cis* and a *trans* association, it is crucial to regress out the *cis* effect to decrease the residual variance and improve power to detect a *trans* effect. **Right. Distribution of SD for each DNAm site category.** For each site the weighted SD across 36 cohorts was taken.

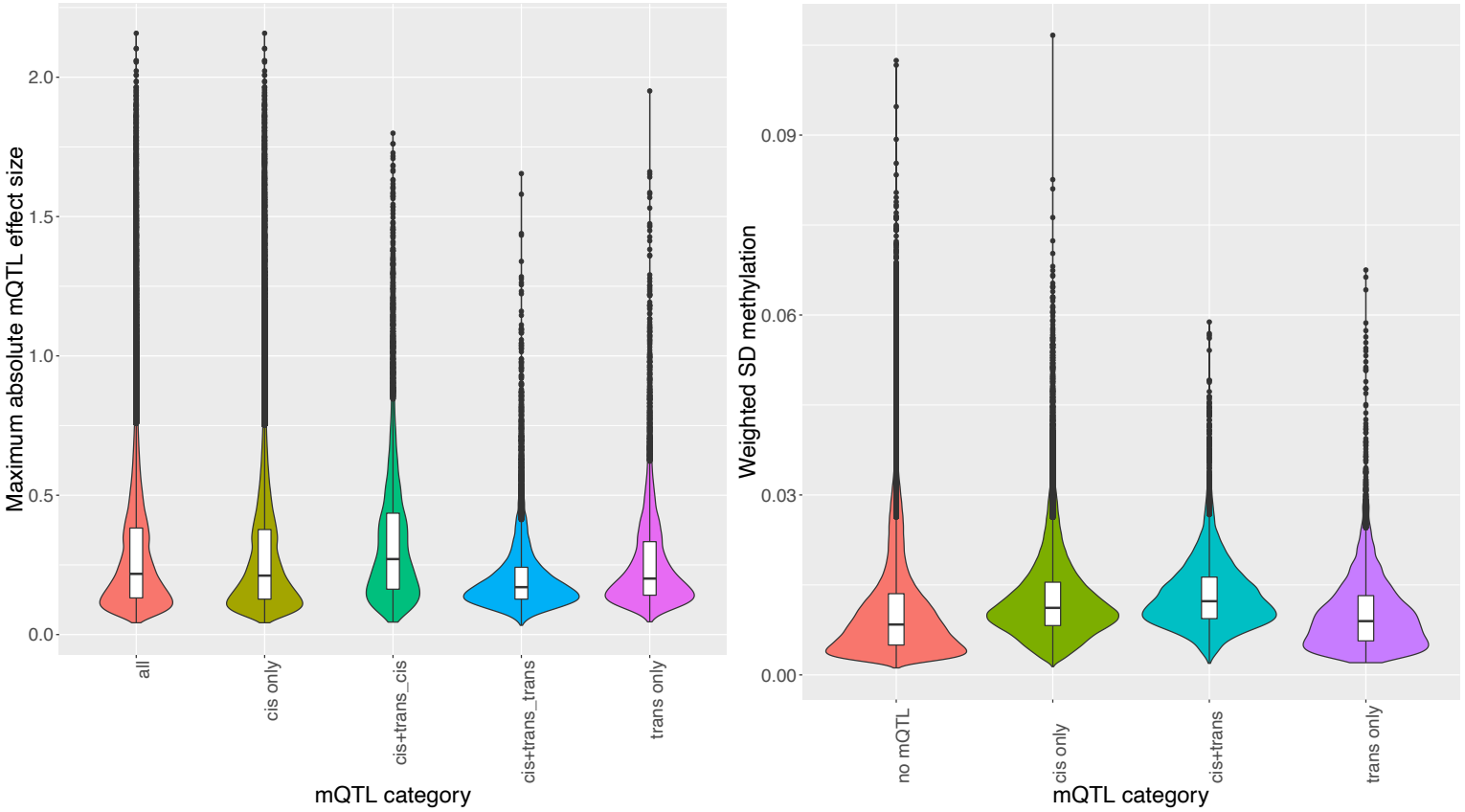


Fig S7: Intrachromosomal distance of SNP and DNAm site
Distance of SNP from DNAm site against the -$\log_{10}$ Pvalue of the mQTL association.

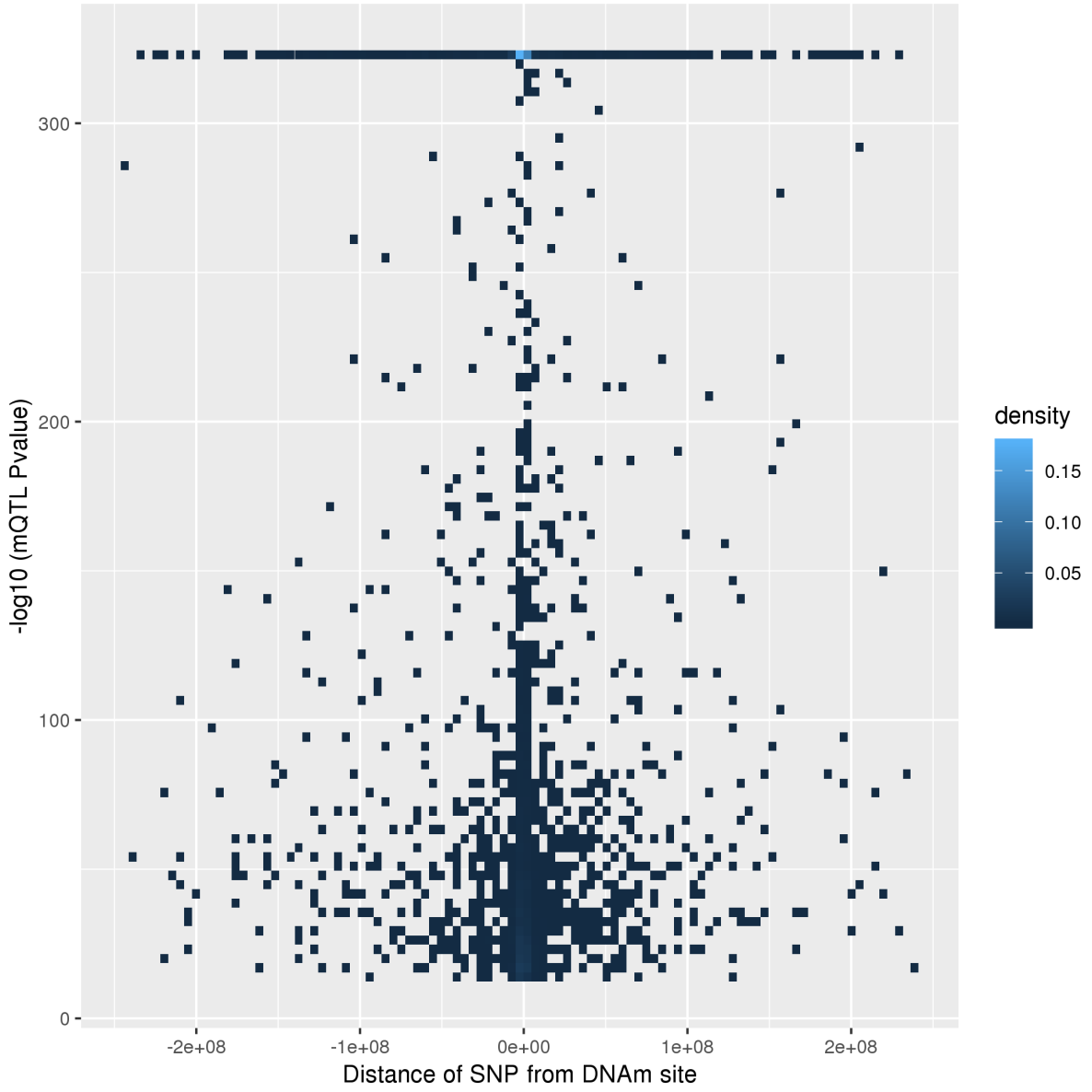


Fig S8: Number of independent *cis* acting SNPs per DNAm site identified in conditional analysis Comparison of the number of independent SNPs for each site for three datasets with variable sample sizes and reference panels. For GoDMC HRC (n=3,984), we meta-analysed three Dutch datasets (LLS, PIAMA and NTR) with similar size as in Bonder *et al*.[2]**.** For GoDMC HRC (n=27,750), we meta-analysed 36 datasets. For Bonder *et al*.[2] we used their reported number of associations. Conditional analysis on Bonder et al[2] and ARIES childhood[3] were performed using individual genotypes as previously reported. An FDR threshold was used as significance threshold for Bonder et al.[2] as for the other datasets a pvalue of 1e-8 was used.

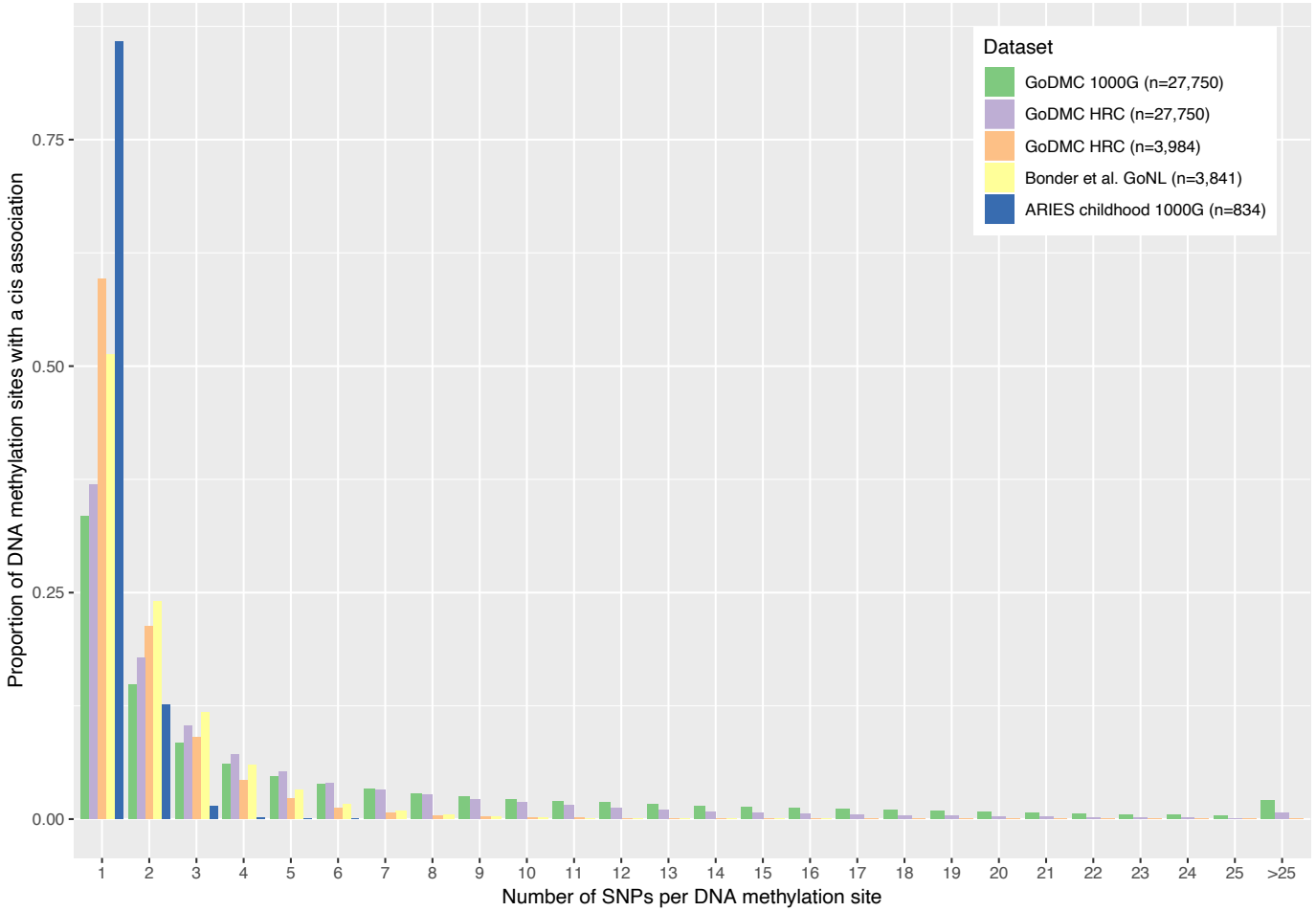


Fig S9: Distribution of DNAm variance explained by mQTL in Generation Scotland. Estimated variance explained obtained from independent replication in Generation Scotland dataset. Left: Histograms of the variance explained estimated in Generation Scotland for 169,656 associations. Right: DNAm variance explained in various proportions of DNAm associations, plotted separately for sites with *cis* and *trans* associations.

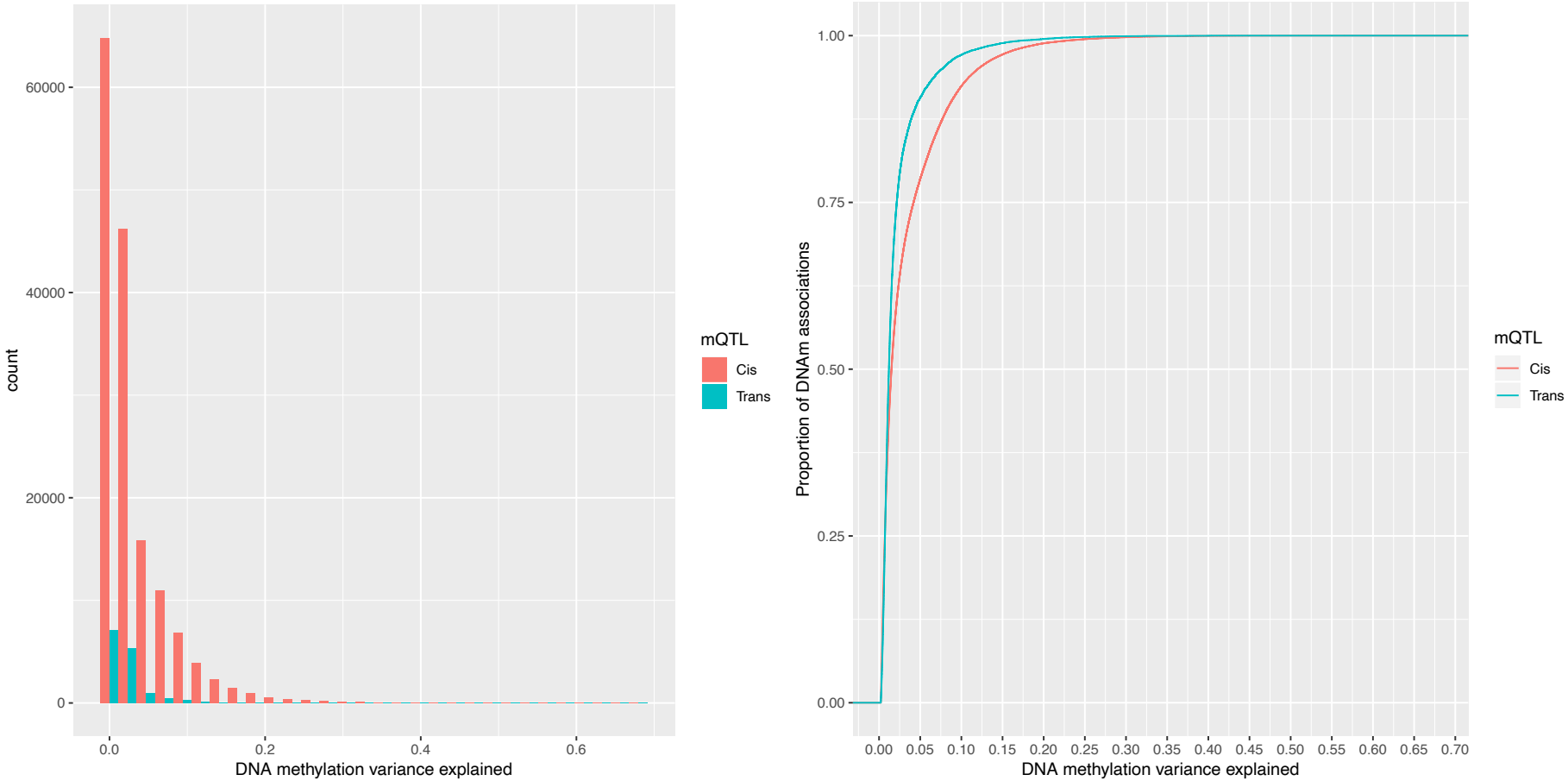


Fig S10: Variance explained by mQTL by DNAm site heritability. Each point represents the average variance explained in DNAm sites binned by heritability. Error bars represent standard errors of the estimate. Heritability estimates are obtained from BSGS twin family study[4] and shown for 115,131 DNAm sites with h2>0.

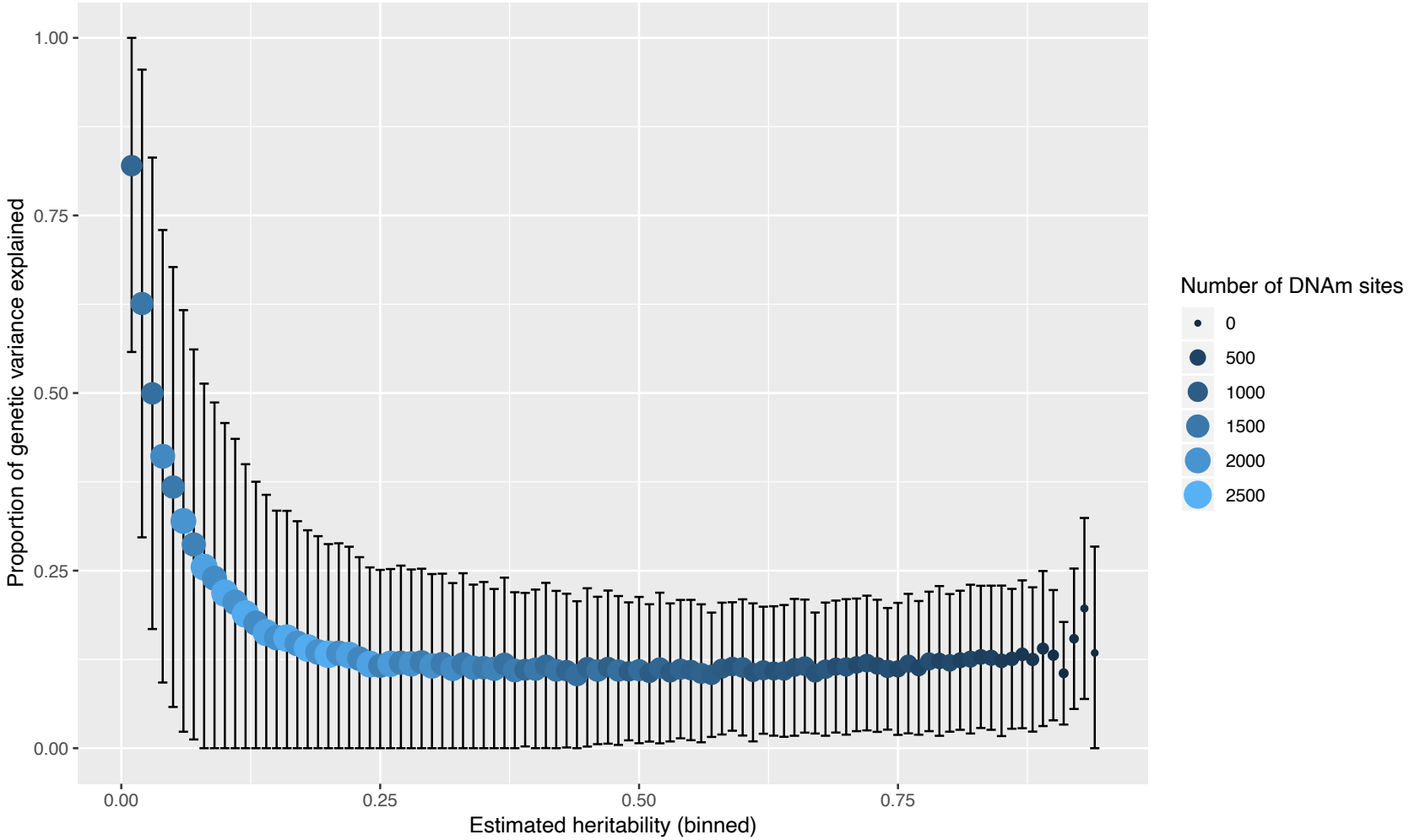


Fig S11: Relationship between heritability of a DNAm site and the number of independent variants identified. Heritability estimates are from a twin study[5] and a twin family study[4]. The independent variants are discovered in GoDMC using LD clumping and replicated in Generation Scotland at p<1e-3. There was a strong correlation between the heritability of a site and the number of identified mQTL (r = 0.69, p < 2.2e-16).

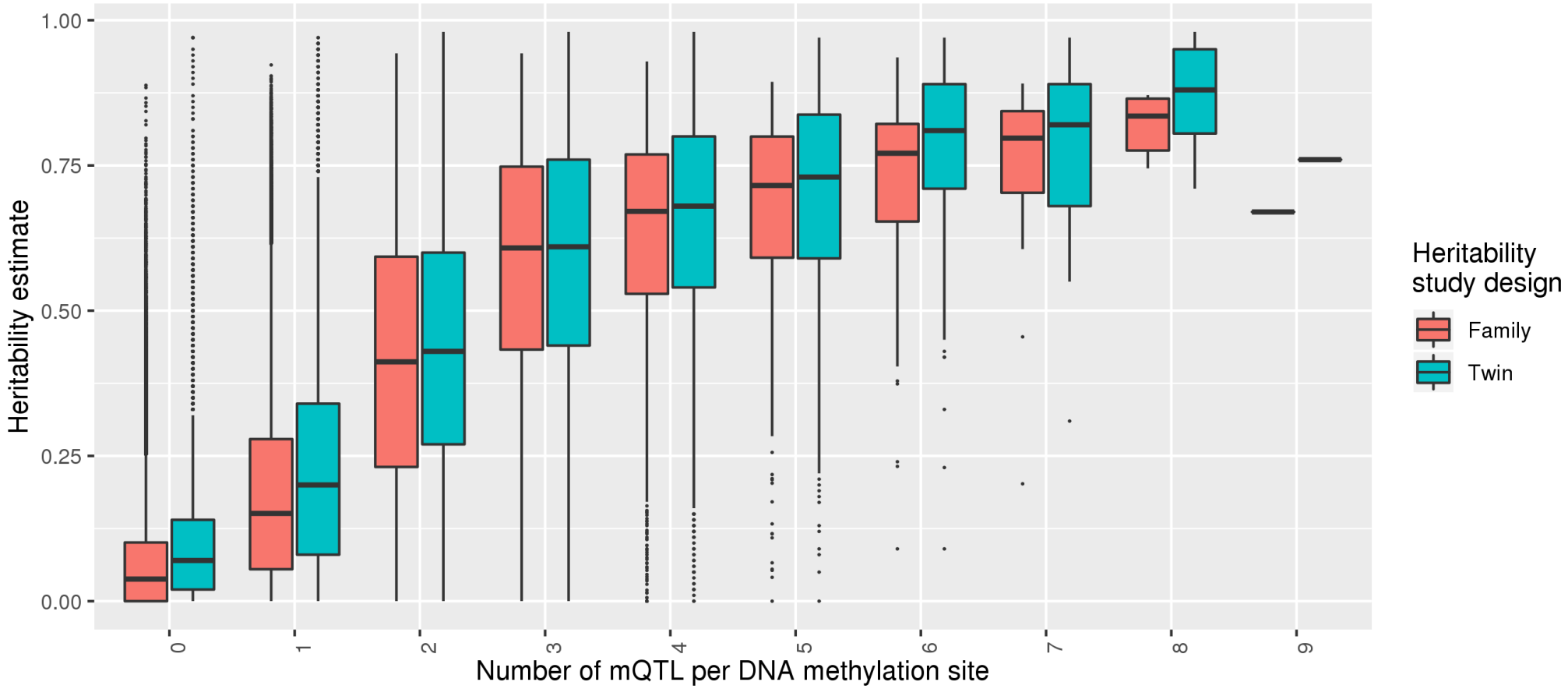


Fig S12: Relationship between heritability of a DNAm site and the maximum effect size of a DNAm site. Heritability estimates are from a twin study[5] and a twin family study[4]. For each DNAm site, the maximum absolute effect size has been plotted. There was a strong correlation between the heritability of a site and the maximum absolute mQTL effect size (r = 0.44, p < 2.2e-16).

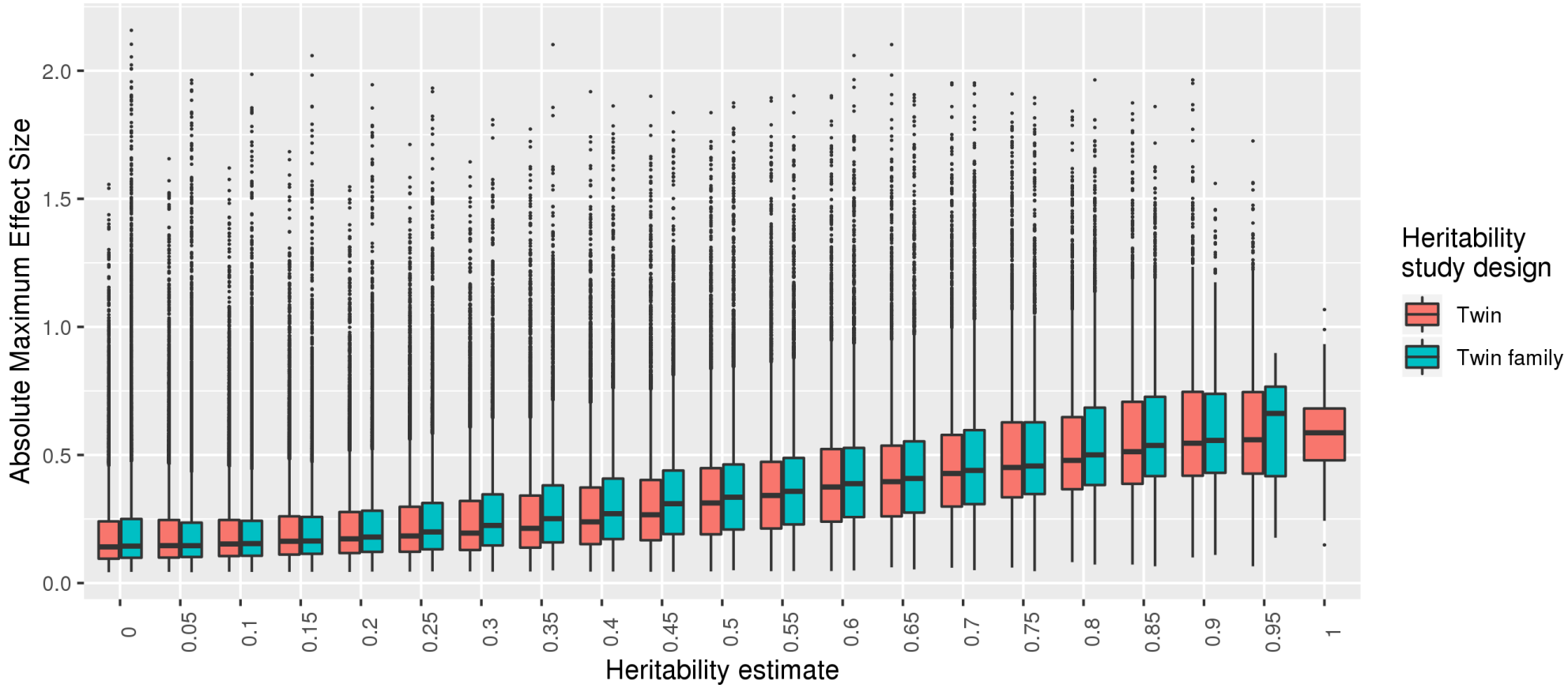


Fig S13: Relationship between variance explained of a DNAm site and the number of independent variants identified. Variance explained was calculated by the sum of r^2 from independent associations for each DNAm site. There was a strong correlation between the variance explained of a site and the number of mQTL identified for it (r = 0.49, p < 2.2e-16).

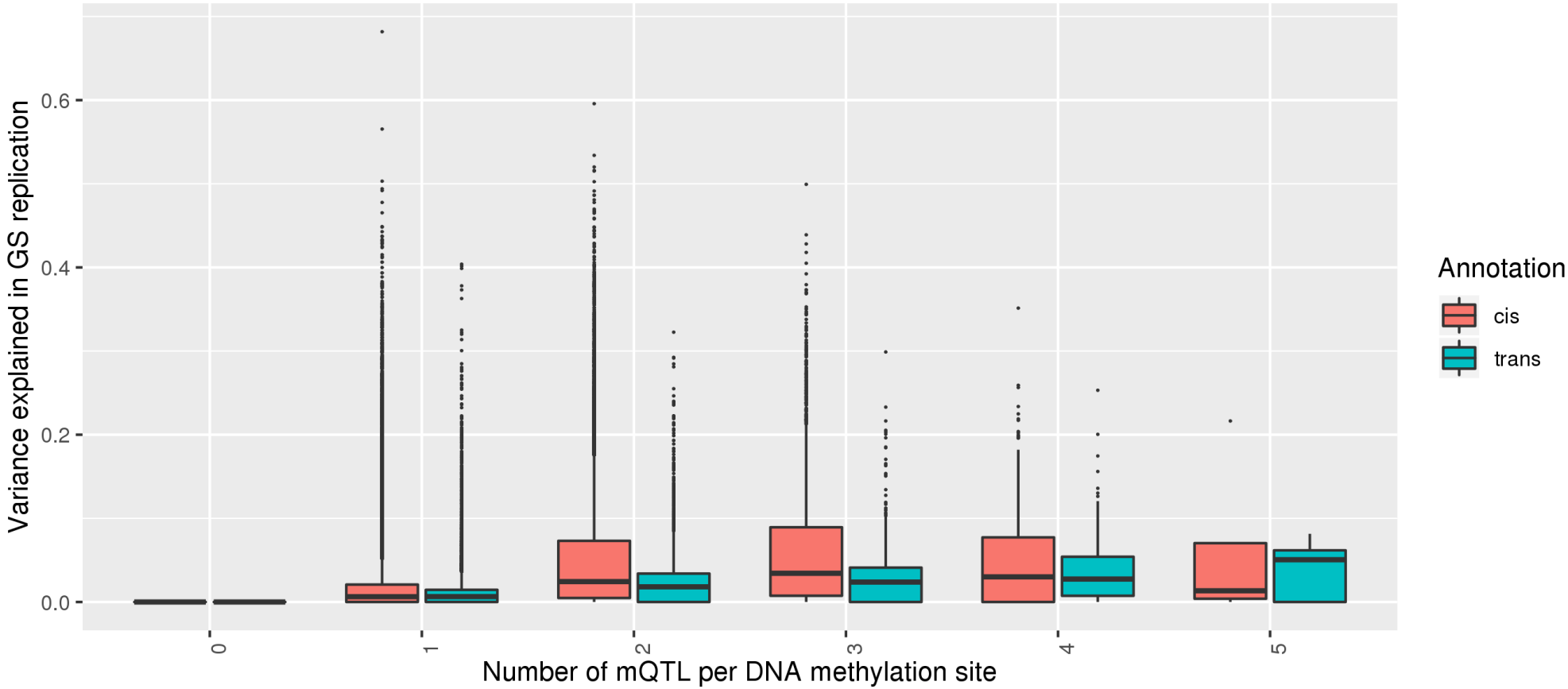


Fig S14: DNAm variance explained by categories of mean DNAm levels. For each DNAm site, the weighted mean in DNAm level across 36 studies was calculated**.** For each site, the sum of r^2 across mQTL associations was calculated. Mean DNAm was categorised as low (0-20%), intermediate (20%-80%) and high (80%-100%) methylation.

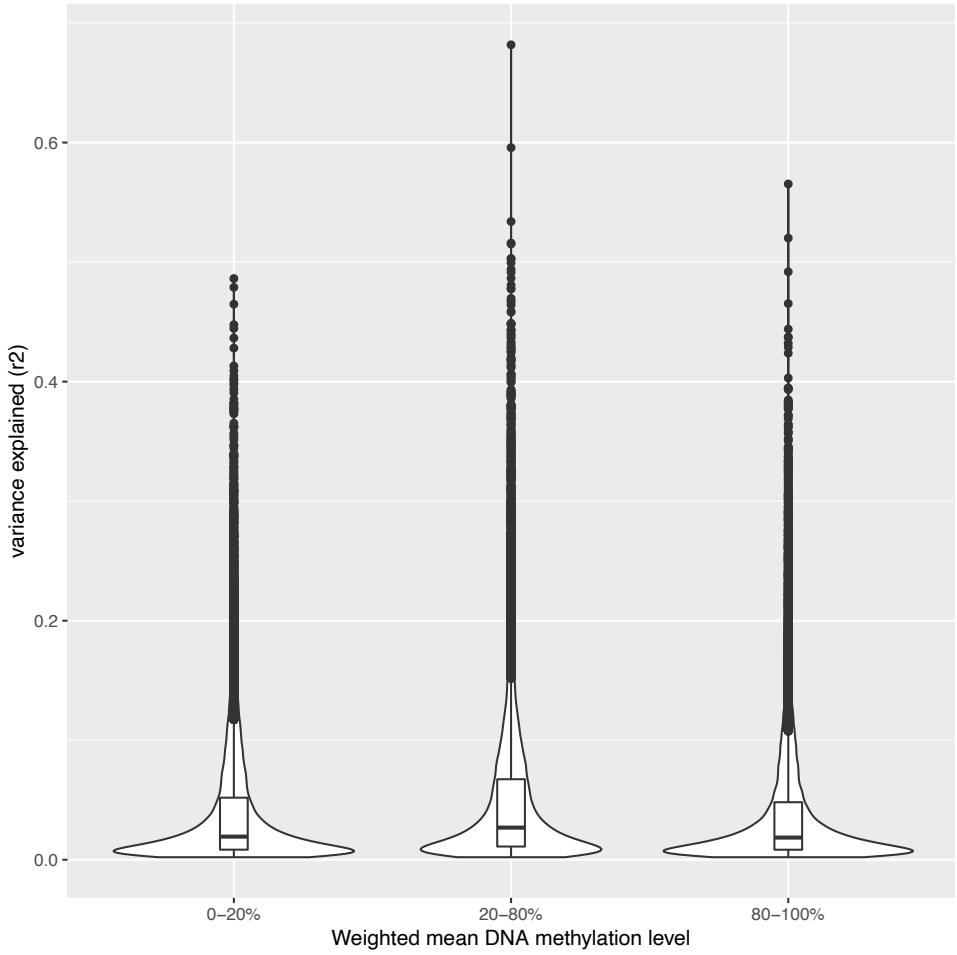


Fig S15: Enrichment or depletion of mQTL sites in predicted chromatin states for *cis+trans*, *cis only* and *trans* only sites.

Plots show odds ratios on a logscale. The size of the circles represents the -log10 Pvalue. Analyses are adjusted for CpG and GC content using LOLA[6]. The x-axis show the 25 states: TssA, Active TSS; PromU, Promoter Upstream TSS; PromD1, Promoter Downstream TSS with DNase; PromD2, Promoter Downstream TSS; Tx5', Transcription 5'; Tx, Transcription; Tx3', Transcription 3'; TxWk, Weak transcription; TxReg, Transcription Regulatory; TxEnh5', Transcription 5' Enhancer; TxEnh3', Transcription 3' Enhancer; TxEnhW, Transcription Weak Enhancer; EnhA1, Active Enhancer 1; EnhA2, Active Enhancer 2; EnhAF, Active Enhancer Flank; EnhW1, Weak Enhancer 1; EnhW2, Weak Enhancer 2; EnhAc, Enhancer Acetylation Only; DNase, DNase only; ZNF/Rpts, ZNF genes & repeats; Het, Heterochromatin; PromP, Poised Promoter; PromBiv, Bivalent Promoter; ReprPC, Repressed PolyComb, Quies Quiescent/Low.

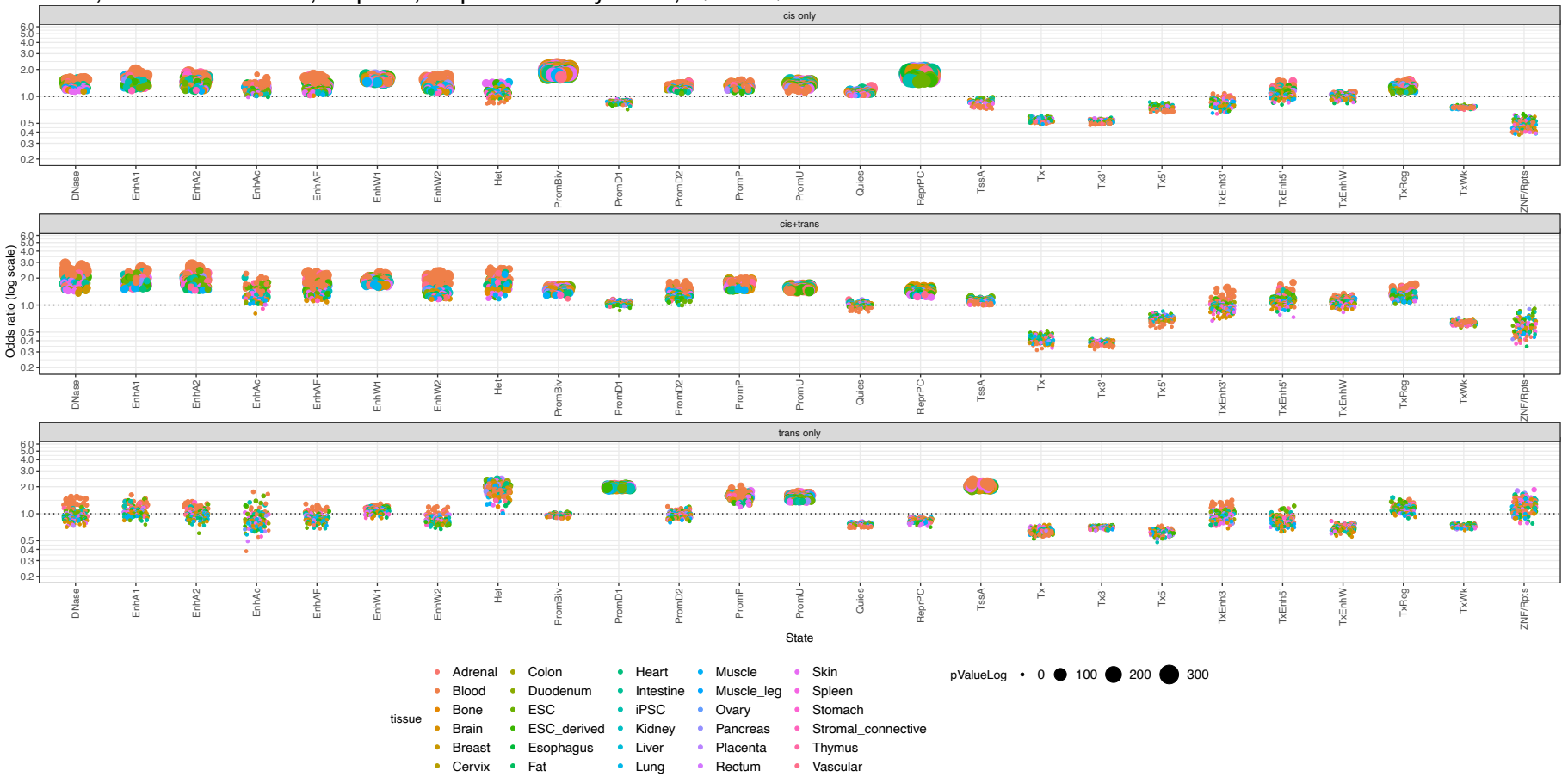


Fig S16: Enrichment or depletion of mQTL sites in gene annotations for *cis+trans*, *cis only* and *trans* only sites. The x-axis shows the gene annotation categories whereas the y-axis shows -log10 P value (top) and log odds ratio with 95% confidence intervals (bottom) for the enrichments. Enrichments are analysed using LOLA[6] and corrected for CpG and CG content.

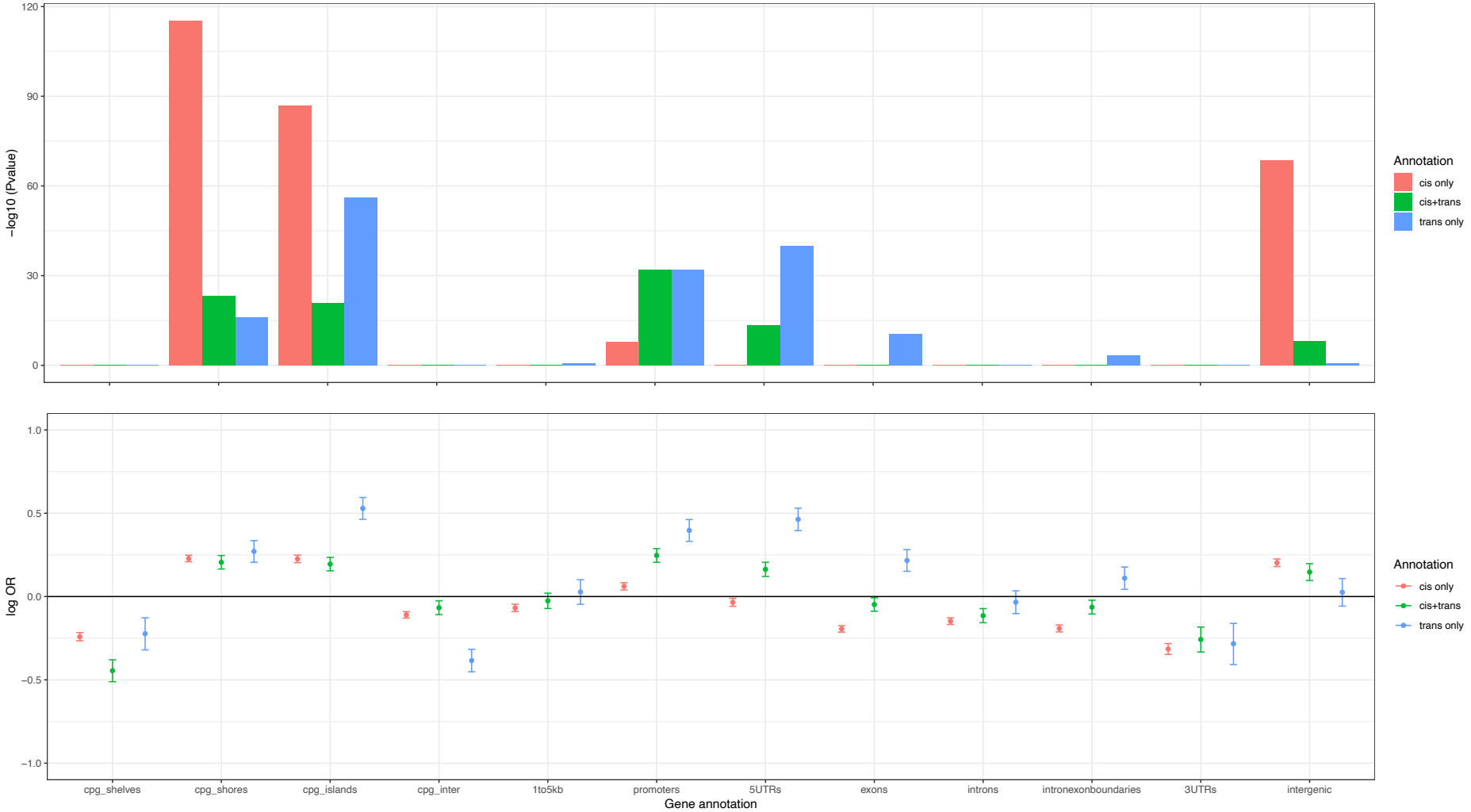


Fig S17: Enrichment or depletion of chromatin states among mQTL SNPs

Plots show odds ratios on a logscale. The size of the circles represents the -log10 Pvalue. Analyses are adjusted for MAF, distance to TSS, number of LD proxies, CpG and GC content using GARFIELD[7]. The x-axis show the 25 states: TssA, Active TSS; PromU, Promoter Upstream TSS; PromD1, Promoter Downstream TSS with DNase; PromD2, Promoter Downstream TSS; Tx5', Transcription 5'; Tx, Transcription; Tx3' , Transcription 3'; TxWk, Weak transcription; TxReg, Transcription Regulatory; TxEnh5', Transcription 5' Enhancer; TxEnh3', Transcription 3' Enhancer; TxEnhW, Transcription Weak Enhancer; EnhA1, Active Enhancer 1; EnhA2, Active Enhancer 2; EnhAF, Active Enhancer Flank; EnhW1, Weak Enhancer 1; EnhW2, Weak Enhancer 2; EnhAc, Enhancer Acetylation Only; DNase, DNase only; ZNF/Rpts, ZNF genes & repeats; Het, Heterochromatin; PromP, Poised Promoter; PromBiv, Bivalent Promoter; ReprPC, Repressed PolyComb, Quies Quiescent/Low.

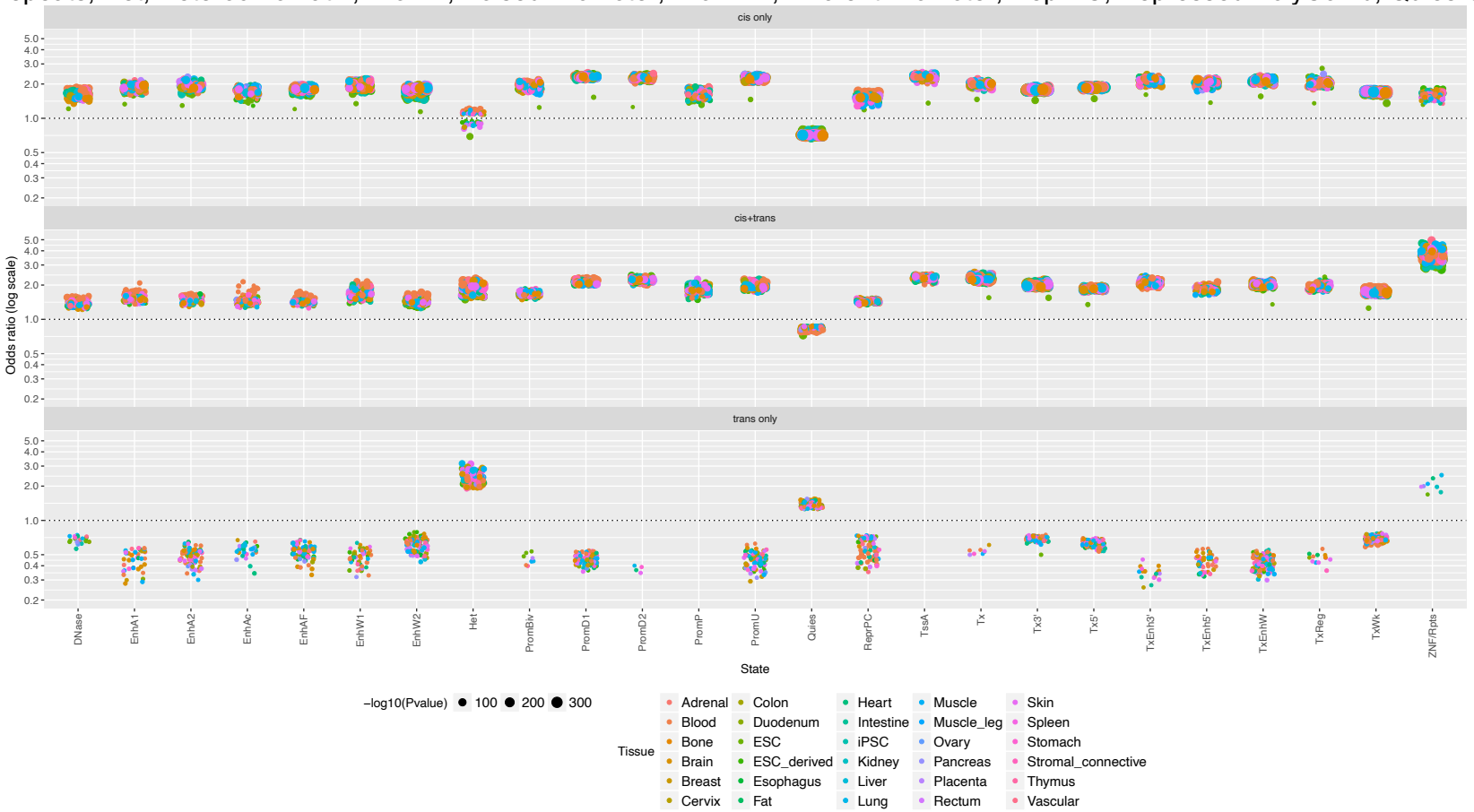


Fig S18: Enrichment or depletion of gene annotations among mQTL SNPs. Results for enrichments where mQTL variants (P < 1e-14), classified into three annotation categories: *cis only*, *cis+trans* and *trans only* are significantly enriched for gene annotation categories. Analyses are adjusted for MAF, distance to TSS, number of LD proxies, CpG and GC content using GARFIELD[7]. A barplot of -log10 P value (top) and log odds ratio with 95% confidence intervals (bottom).

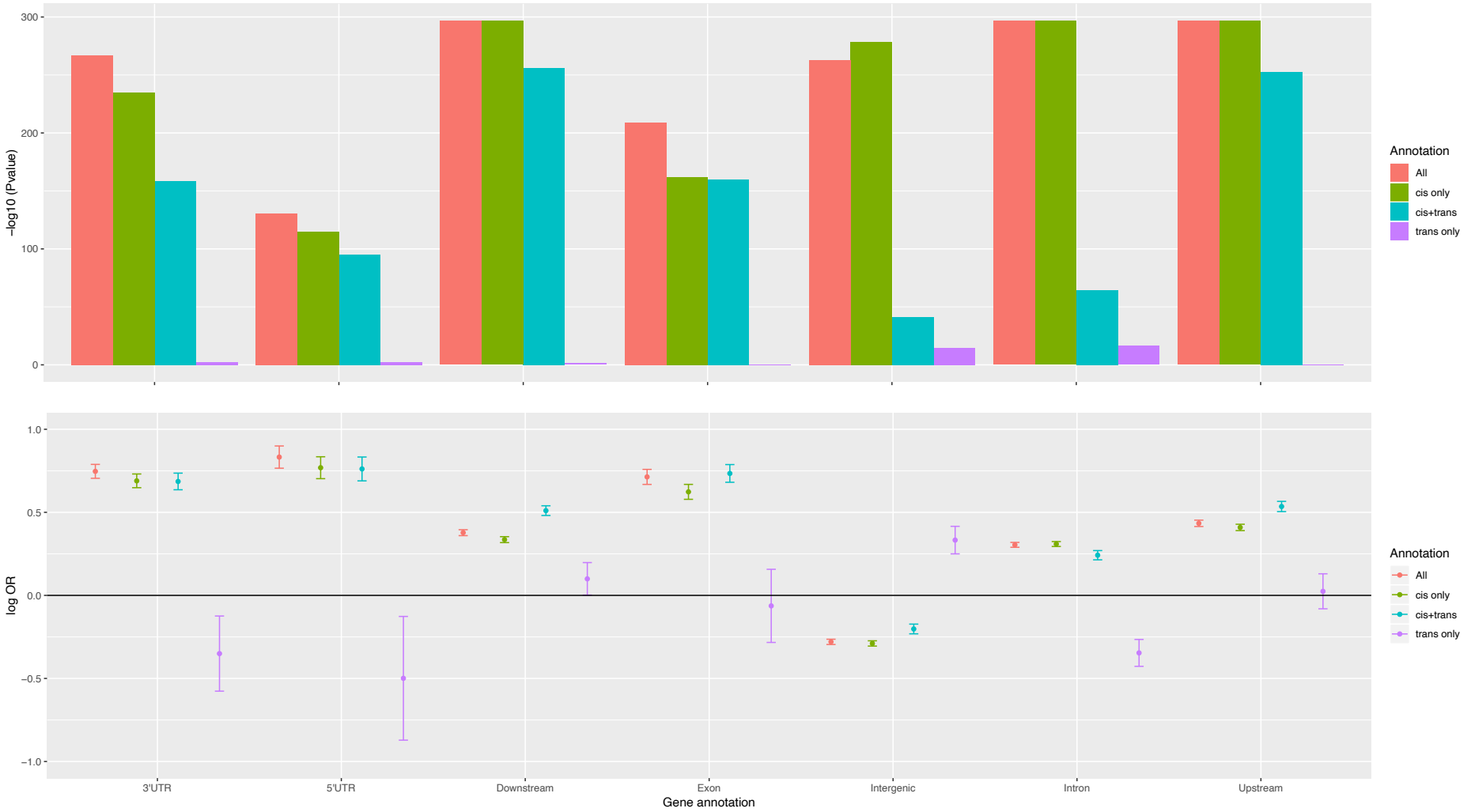


Fig S19: DNAm levels at DNAm sites with a *cis* and a *trans* mQTL association and in two isolated cell subsets (monocytes, Tcells).
The heatmap shows the DNAm levels of *cis+trans* sites in blood (weighted mean across 36 blood studies), monocytes (N=2,002) and T cells (N=214)[8]. Colors represent DNAm level (low = <0.2; intermediate = 0.2-0.8; high >0.8). For 66% of *cis+trans* sites, intermediate DNAm levels (20%-80%) were found. For 74% of *cis+trans* DNAm sites with intermediate DNAm in blood, intermediate DNAm was also found in purified cell types suggesting that cell-to-cell variability rather than differential DNAm between cell types was underlying intermediate DNAm.

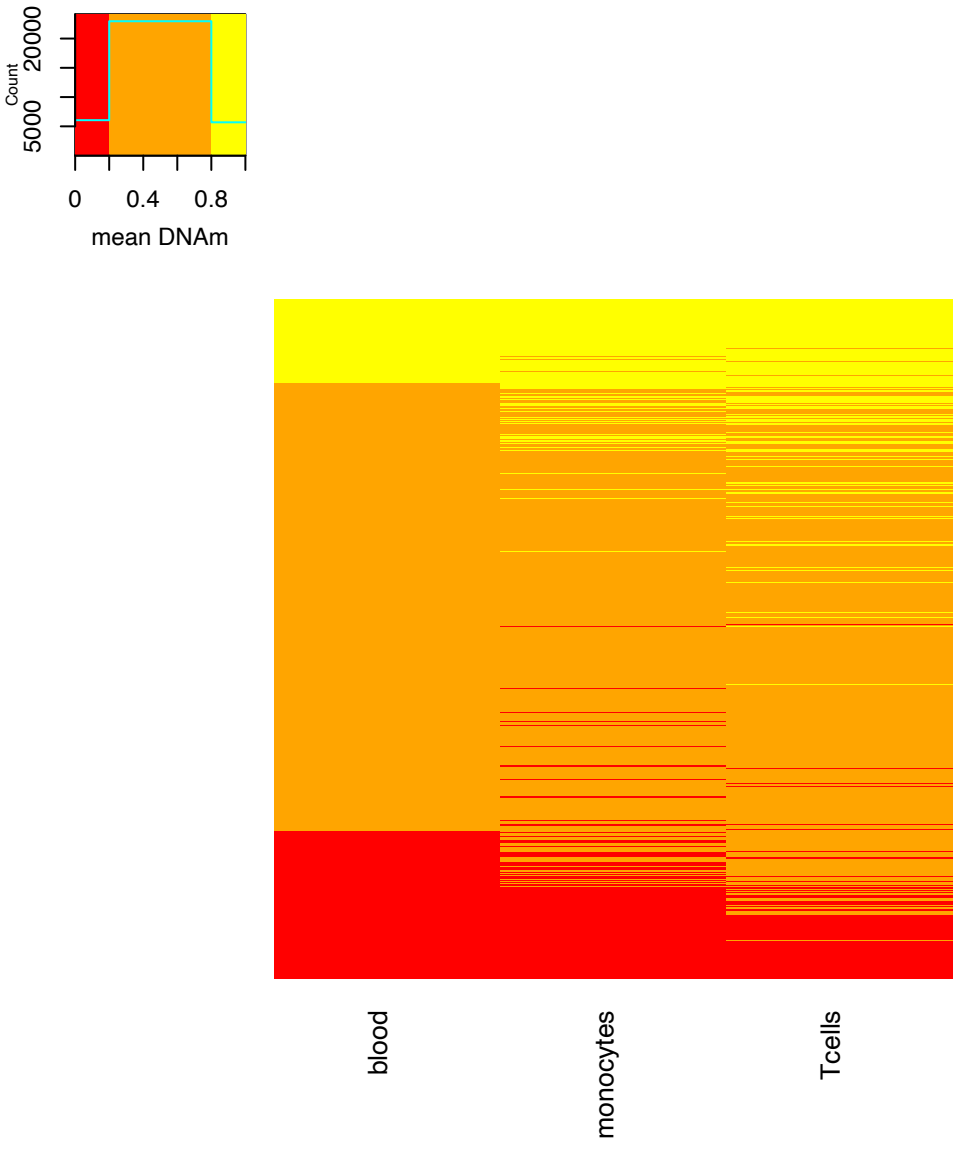


Fig S20: Comparison of enrichments in predicted chromatin states for *cis+trans*, *cis only* and *trans* only sites using >1 Mb versus cross chromosomal to define a *trans* association.
To assess whether these enrichments are driven by long-range *cis* associations, we defined *trans* associations as interchromosomal only and compared the enrichments to the original enrichments where a *trans* association was defined as inter- or intrachromosomal. Using the new *trans* definition, the odds ratios for the heterochromatin state enrichments were decreased for *cis+trans* sites suggesting that long-range *cis* associations contribute to heterochromatin enrichments. Plots show odds ratios of enrichments without chromosome 6. Analyses are adjusted for CpG and GC content using LOLA[6]. Colors represent chromatin state: TssA, Active TSS; PromU, Promoter Upstream TSS; PromD1, Promoter Downstream TSS with DNase; PromD2, Promoter Downstream TSS; Tx5', Transcription 5'; Tx, Transcription; Tx3', Transcription 3'; TxWk, Weak transcription; TxReg, Transcription Regulatory; TxEnh5', Transcription 5' Enhancer; TxEnh3', Transcription 3' Enhancer; TxEnhW, Transcription Weak Enhancer; EnhA1, Active Enhancer 1; EnhA2, Active Enhancer 2; EnhAF, Active Enhancer Flank; EnhW1, Weak Enhancer 1; EnhW2, Weak Enhancer 2; EnhAc, Enhancer Acetylation Only; DNase, DNase only; ZNF/Rpts, ZNF genes & repeats; Het, Heterochromatin; PromP, Poised Promoter; PromBiv, Bivalent Promoter; ReprPC, Repressed PolyComb, Quies Quiescent/Low.

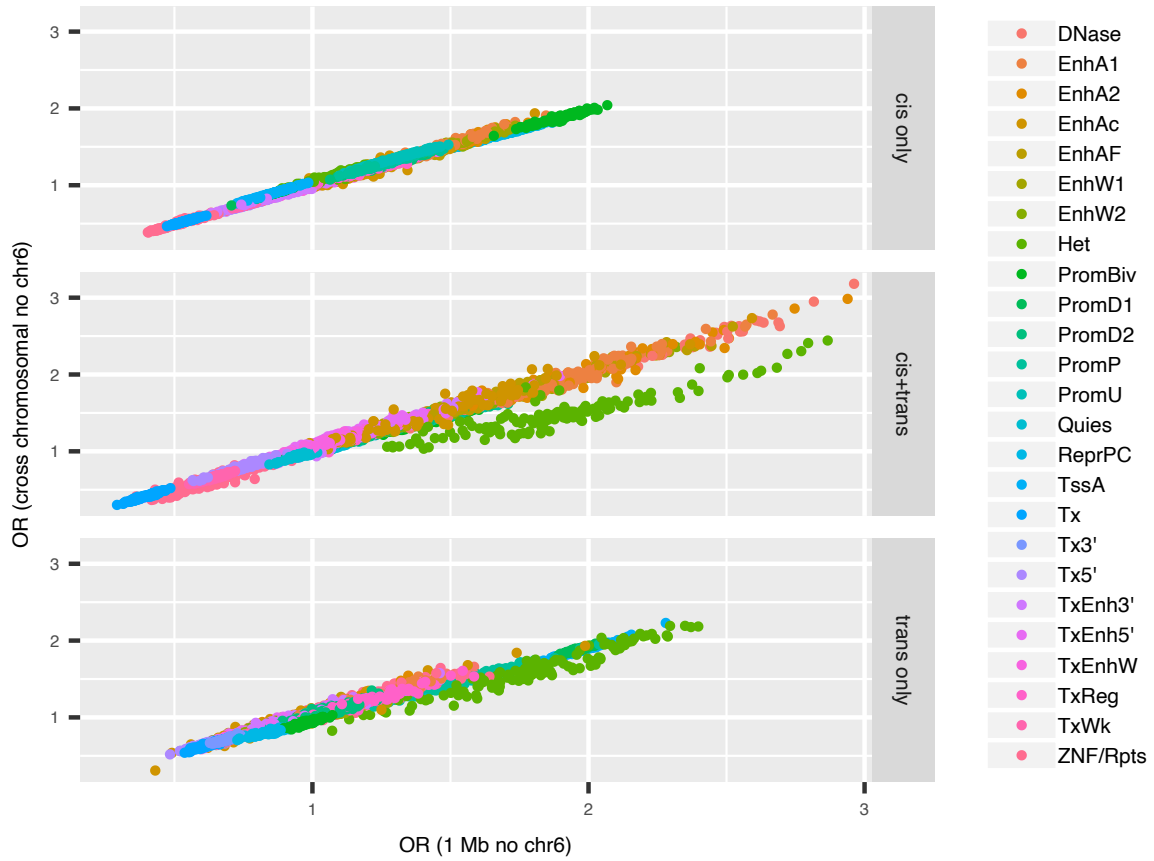


Fig S21: DNAm levels at DNAm sites with a *cis* only mQTL association and in 12 tissues.
The heatmap shows the DNAm levels of *cis only* sites in blood (weighted mean across 36 blood studies) and in 12 tissues (mean across individuals within each tissue type)[9]. Colors represent DNAm level (low <0.2; intermediate 0.2-0.8; high methylation >0.8).

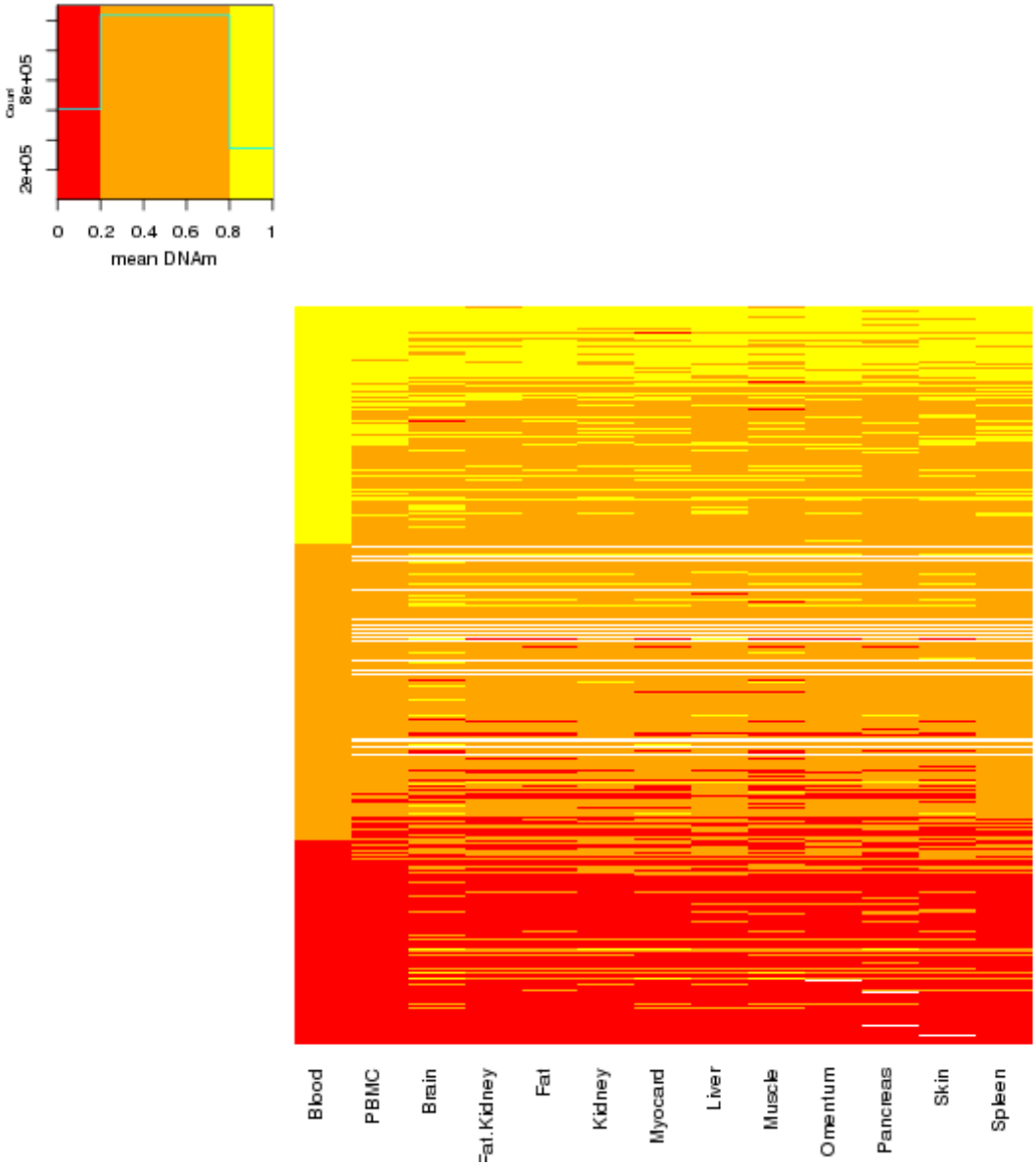


Fig S22: DNAm levels at DNAm sites with a *cis* and a *trans* mQTL association and in 12 tissues.

The heatmap shows the DNAm levels of *cis+trans* sites in blood (weighted mean across 36 blood studies) and in 12 tissues (mean across individuals within each tissue type)[9]. Colors represent DNAm level (low <0.2; intermediate 0.2-0.8; high >0.8).

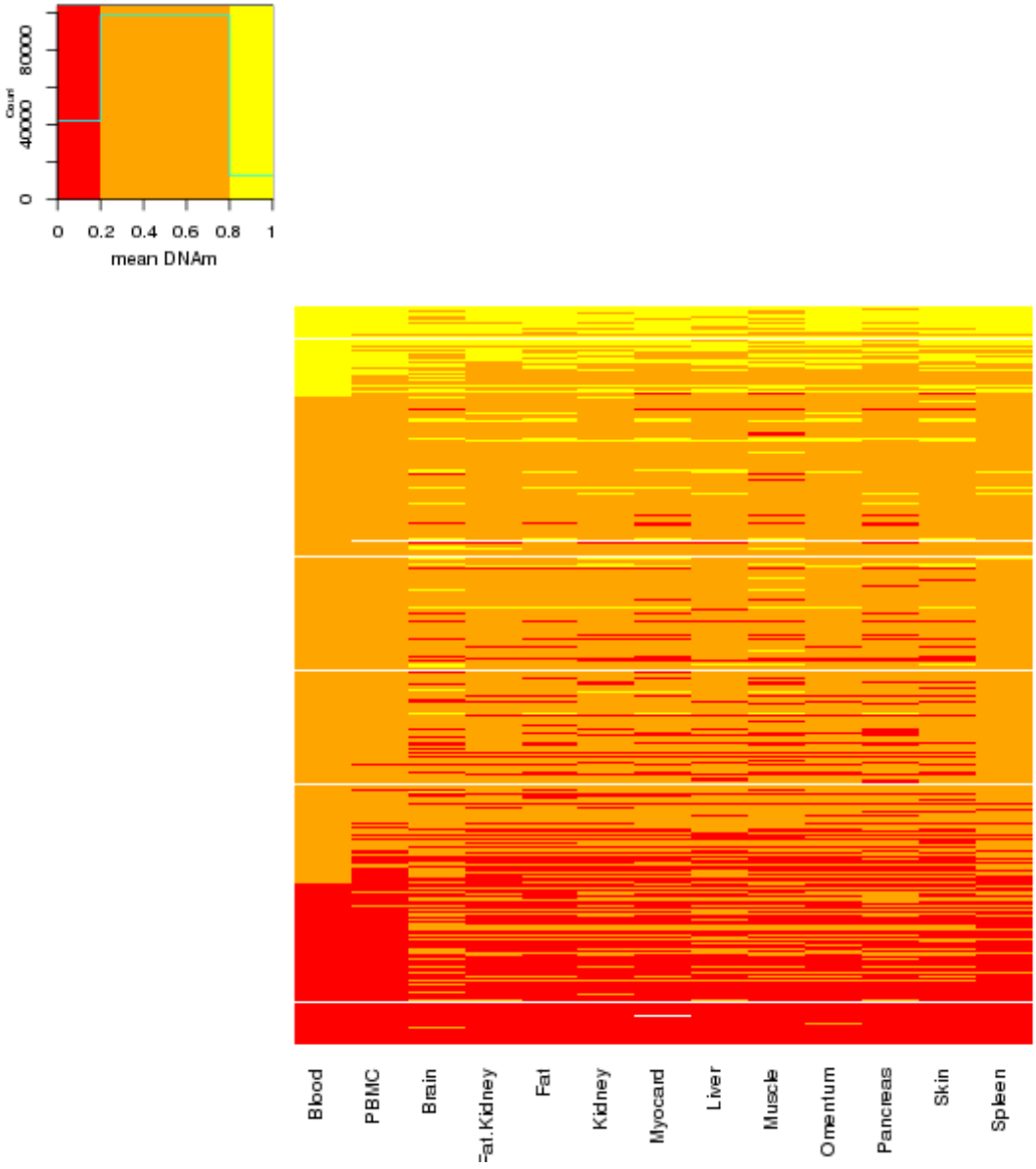


Fig S23: DNAm levels at DNAm sites with a *trans* only mQTL association and in 12 tissues.
The heatmap shows the DNAm levels of *trans only* sites in blood (weighted mean across 36 blood studies) and in 12 tissues (mean across individuals within each tissue type)[9]. Colors represent DNAm level (low <0.2; intermediate 0.2-0.8; high >0.8).

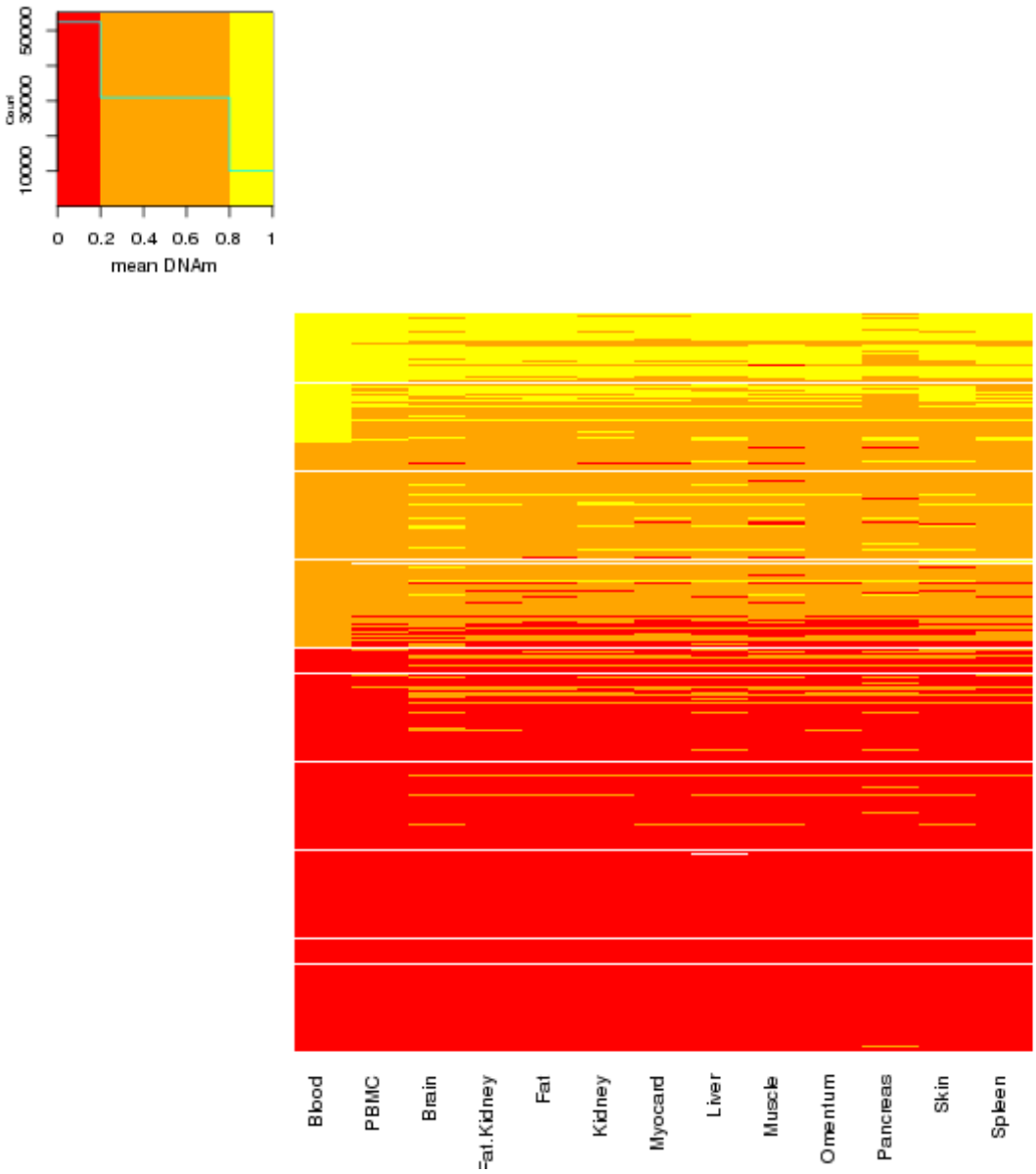
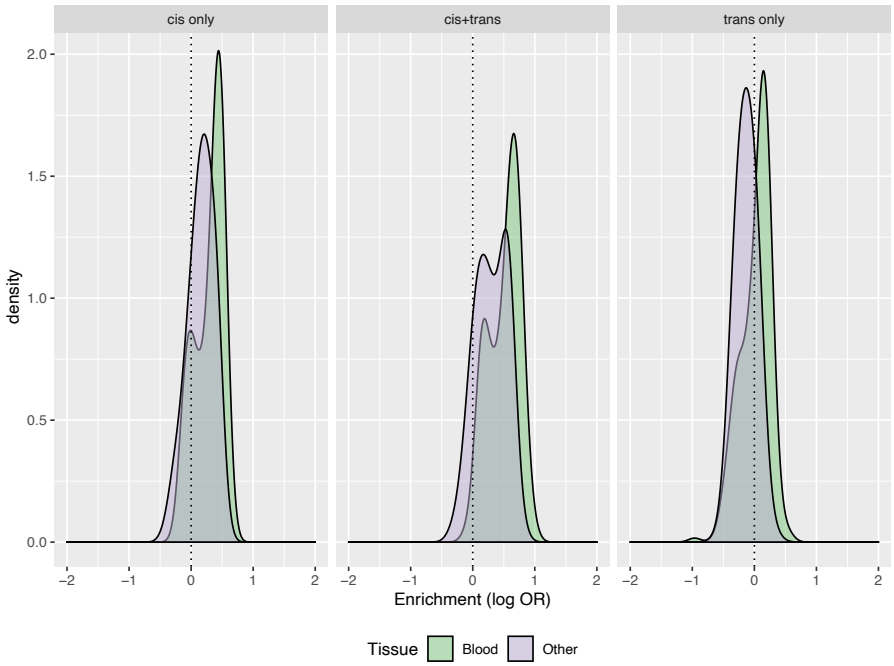


Fig S24: Comparison of enrichments in predicted enhancer and promoter chromatin states for *cis+trans*, *cis only* and *trans* only sites. Top: Enhancer states Bottom: Promoter states



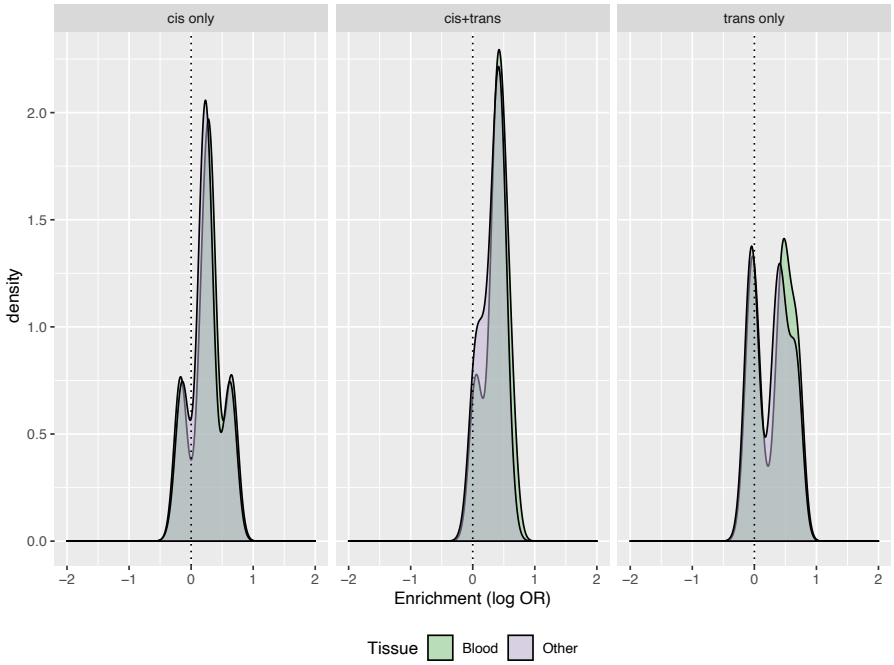


Fig S25: Enrichment of TFBS among mQTL sites

The x-axis shows the TFBS whereas the y-axis shows odds ratios on a logscale. Enrichments are analysed using LOLA[6] and corrected for CpG and CG content. The size of the circles represents the -log10 Pvalue. Different colors represent different tissues.

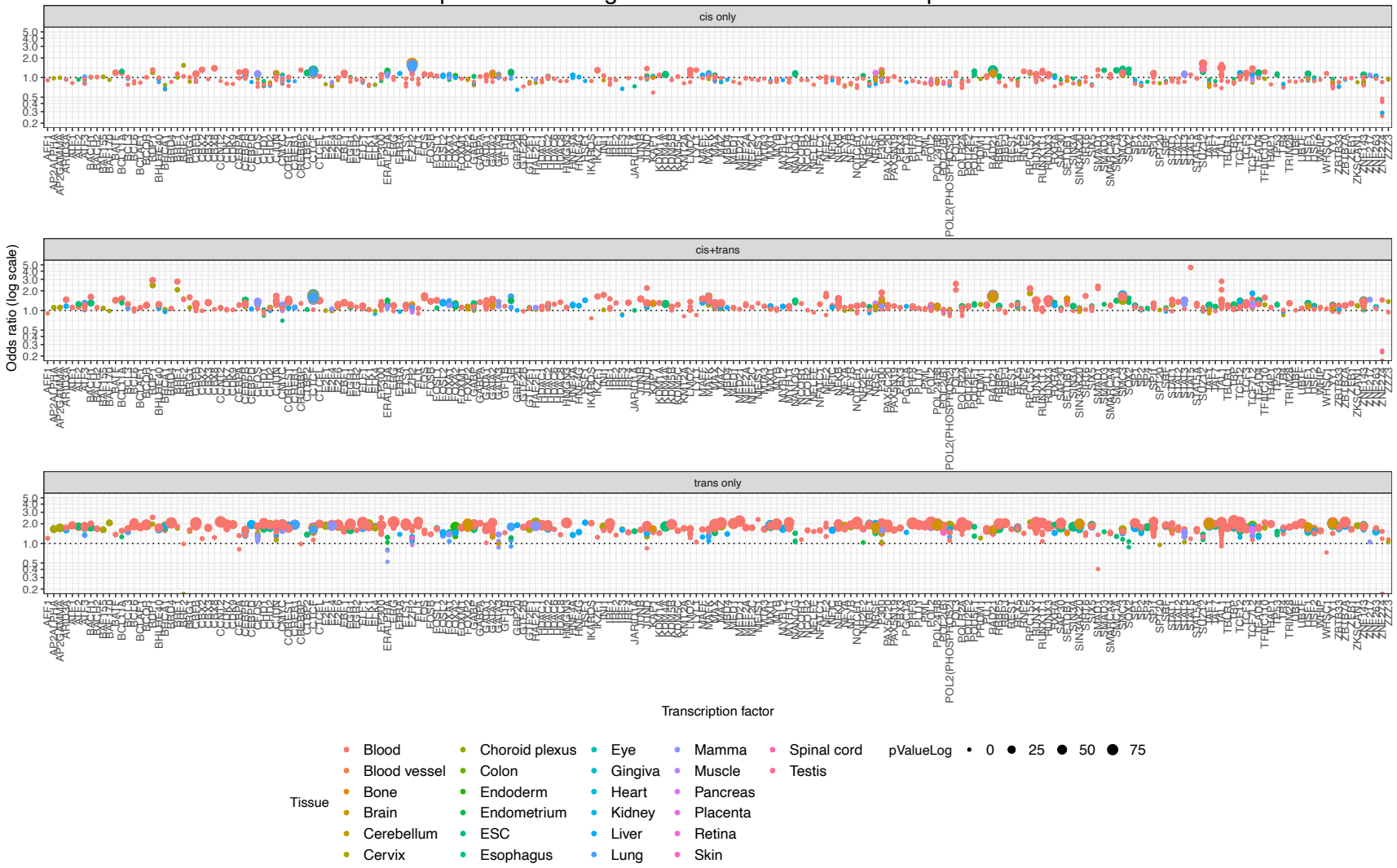


Fig S26: Enrichment of TFBS among mQTL SNPs

The x-axis shows the TFBS whereas the y-axis shows odds ratios on a logscale. Enrichments are analysed using GARFIELD[7] and corrected for MAF, distance to TSS, number of LD proxies, CpG and CG content. The size of the circles represents the -log10 Pvalue. Different colors represent different tissues.

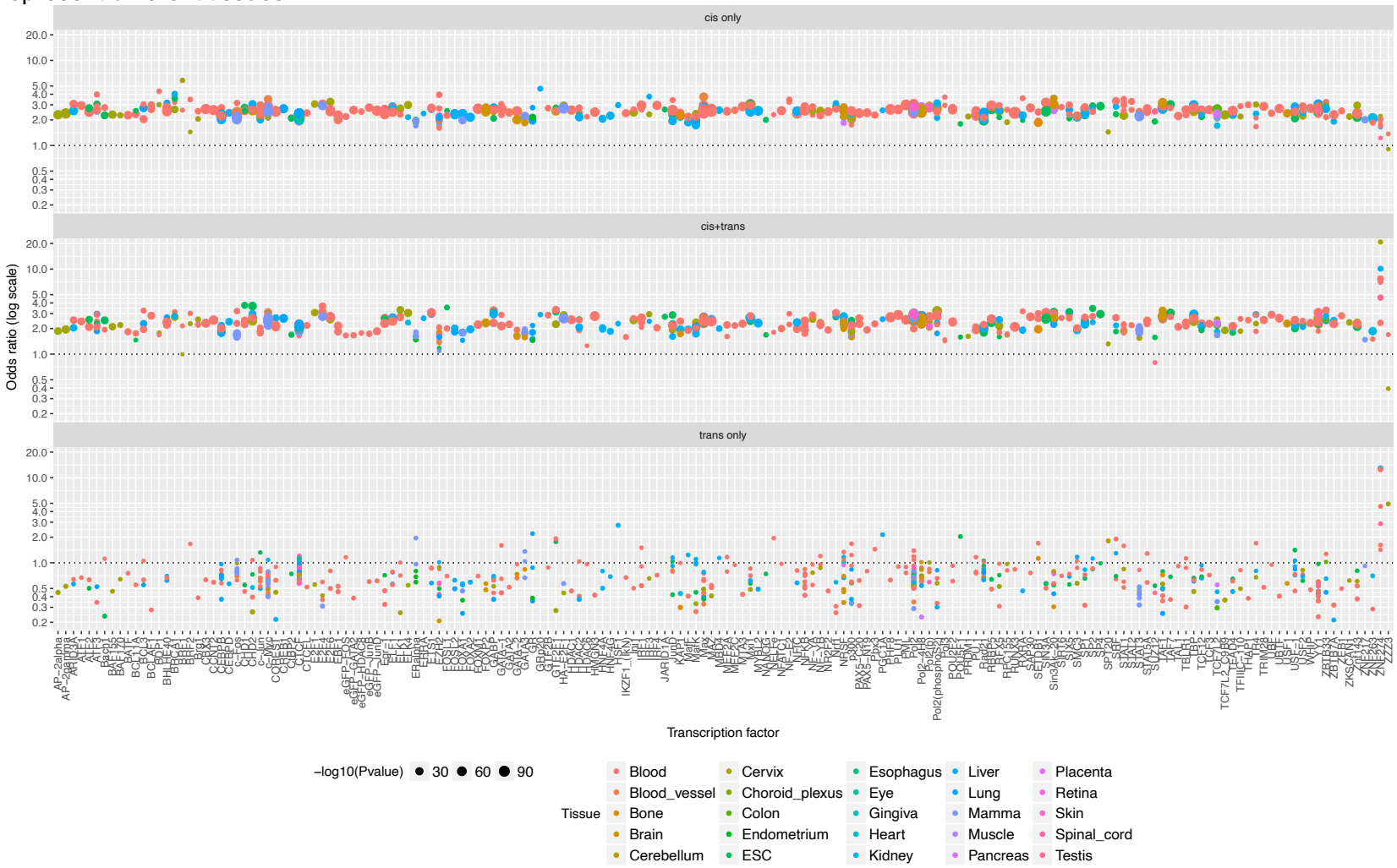


Fig S27: Weighted mean of proportion DNAm by TFBS.
For each TFBS the weighted mean of the proportion of DNAm has been shown. The dashed line shows the weighted mean of the proportion DNAm across all sites (0.52). DNAm sites overlapping TFBS show hypo DNAm.

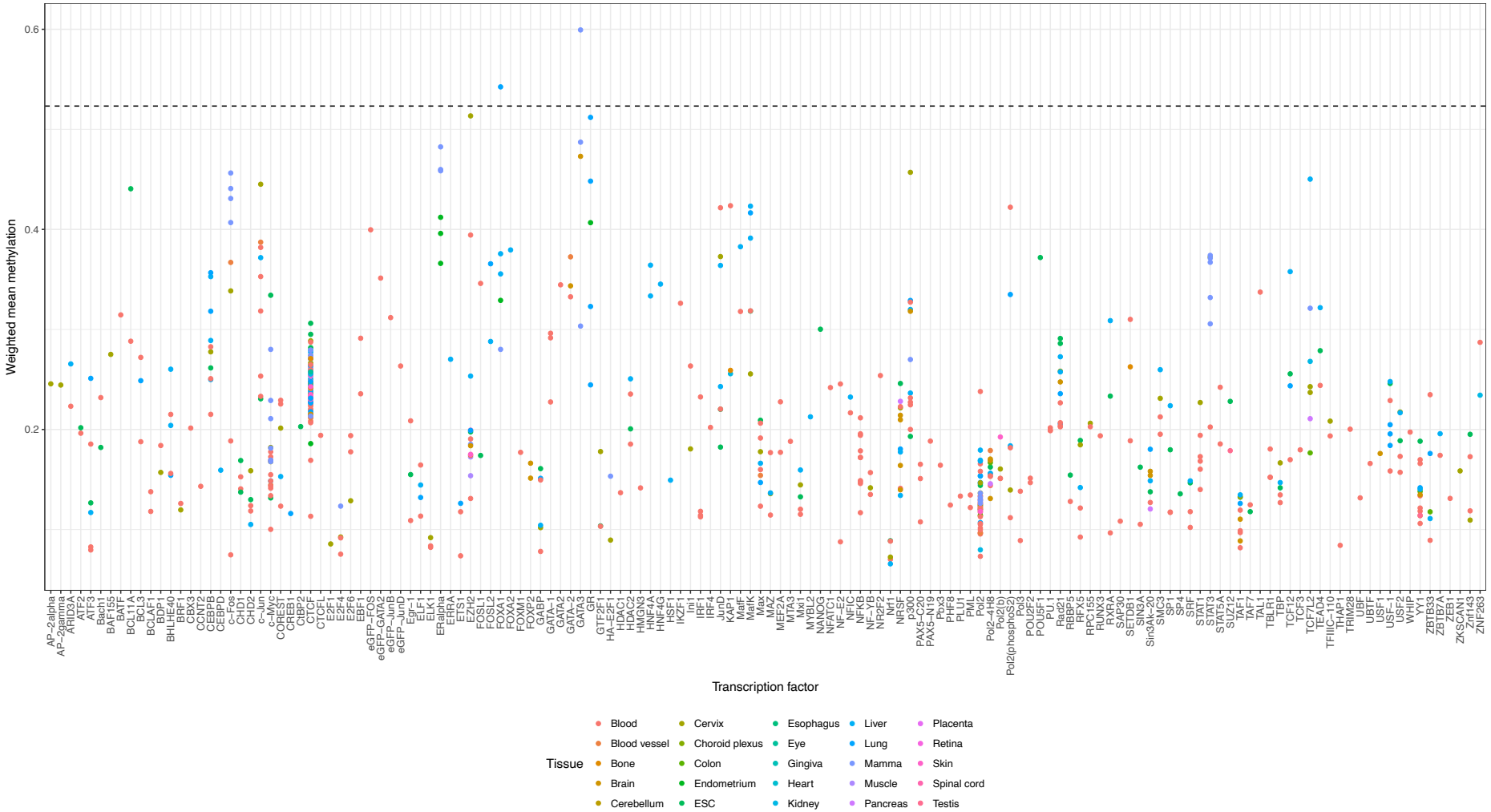


Fig S28: 2D enrichment of SNP and DNAm site TFBS annotation

To test if the annotations of the SNPs involved in *trans*-mQTL were specific to the annotations of the DNAm sites that they influence, we compared the real SNP-DNAm site pairs against permuted SNP- DNAm site pairs, where the biological link between SNP and site is severed whilst maintaining the distribution of annotations for the SNPs and sites. We constructed 100 such permuted datasets b) SNP and site positions were annotated against genomic features, and we quantified how frequently mQTL were found for each pair of SNP-DNAm site annotations. This enabled the construction of 2D-annotation matrices for both the real *trans*-mQTL list and the permuted *trans*-mQTL lists.

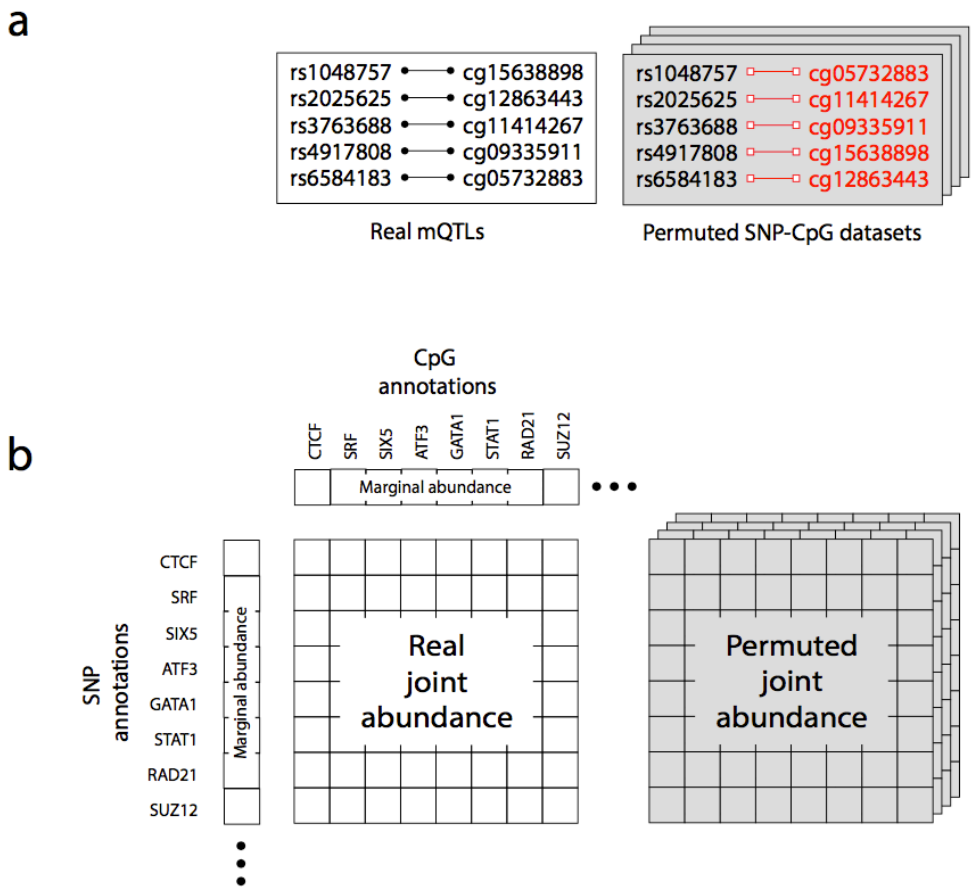


Fig S29: Enrichment of interchromosomal interactions among mQTL.
To test if SNP- DNAm site pairs involved in interchromosomal *trans*-mQTL were enriched for interchromosomal interactions, we compared the real SNP- DNAm site pairs against permuted SNP- DNAm site pairs using HiC data from Rao et al.[10]

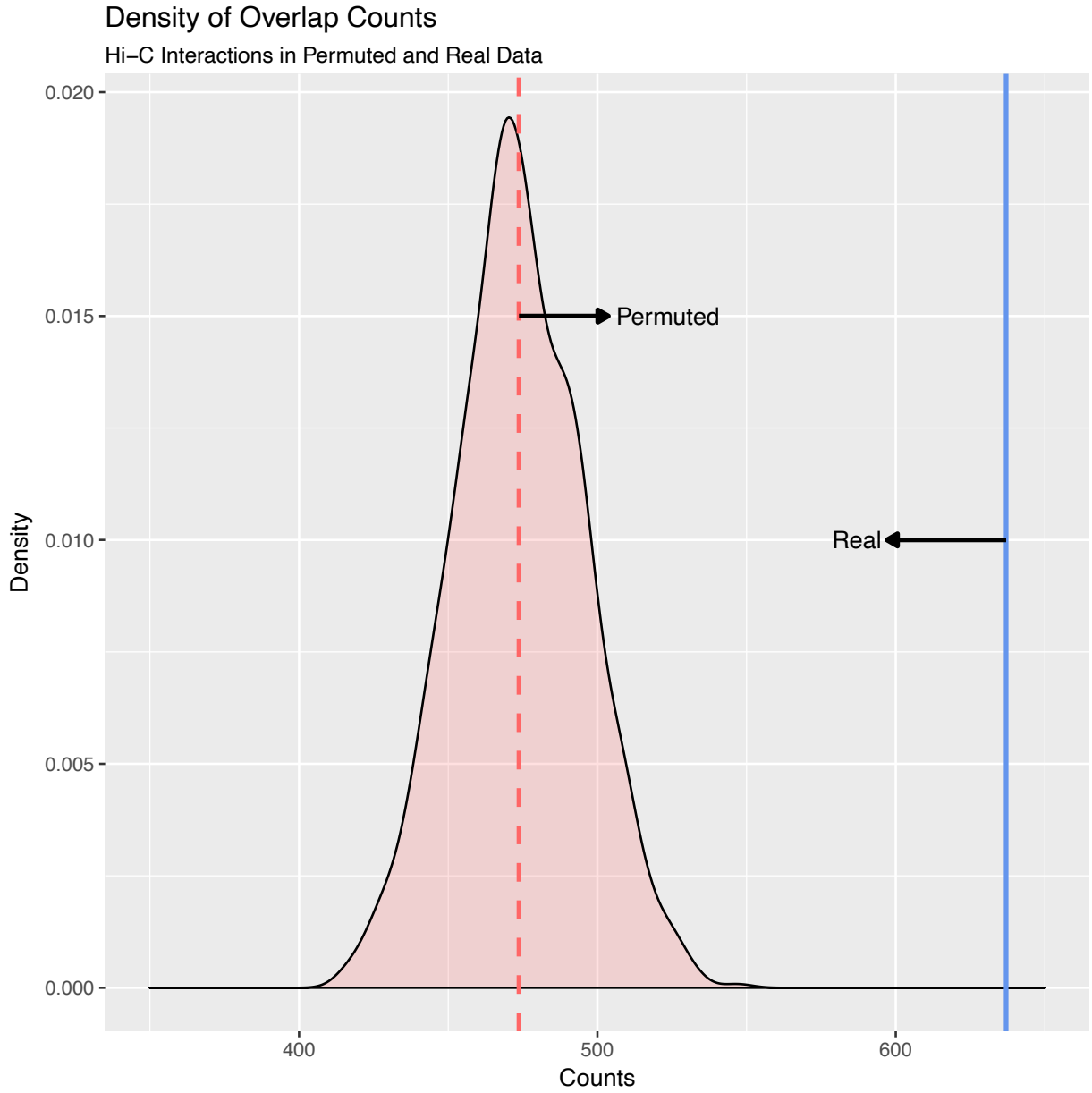


Fig S30: The Wald ratio estimates for *cis*-DNAm sites against *trans*-DNAm sites (y axis) against the estimated correlations (x axis) demonstrates a weak positive relationship, but not one strong enough to suggest that they are related through direct causal relationships.

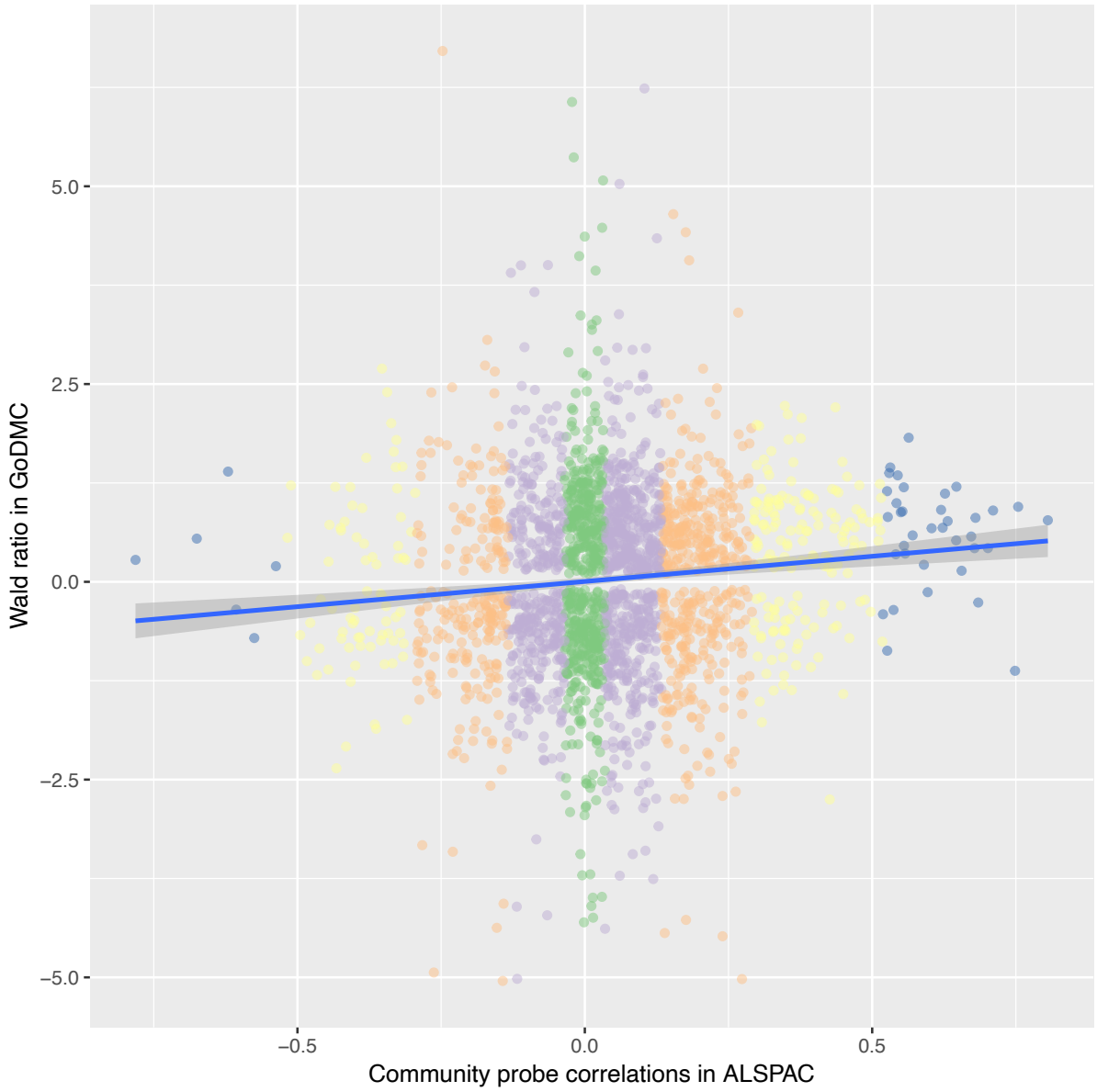


Fig S31: Average correlation amongst DNAm sites within *cis-trans* pairs
The average absolute pairwise observational correlation amongst DNAm sites that are in a *cis-trans* pair, using the ARIES dataset[11], is plotted in green. To evaluate if this is larger than we expect by chance, we permuted the *cis-trans* pairings and estimated the average absolute pairwise correlation amongst the randomly paired sites. Each permutation set is plotted in red.

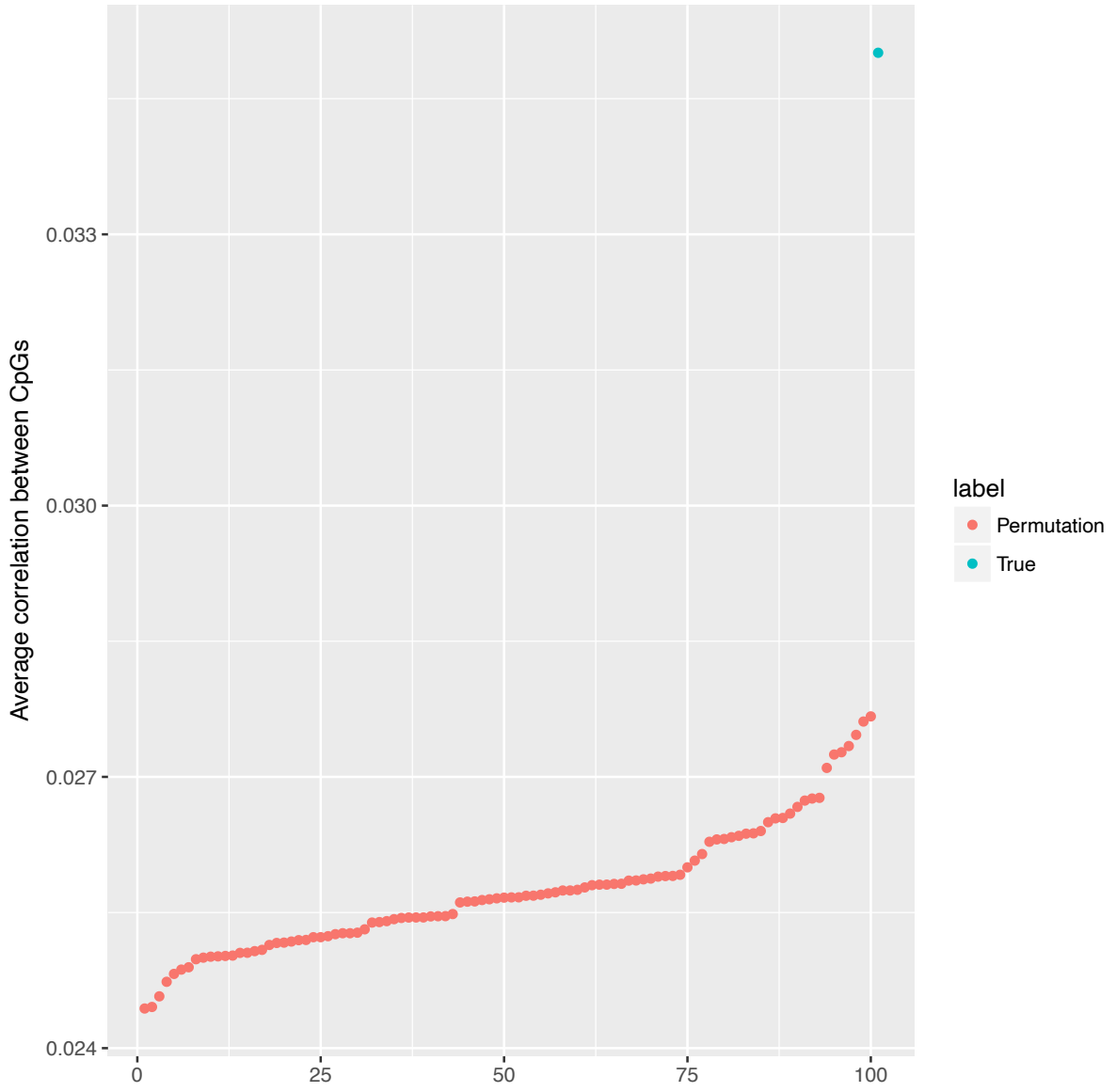


Fig S32: Enrichment of TFBS among mQTL sites by community

The x-axis shows the TFBS whereas the y-axis shows odds ratios on a logscale. Different colors represent different communities. Enrichments are analysed using LOLA[6] and corrected for CpG and CG content. The size of the circles represents the -log10 Pvalue. Three communities (labelled in plot) were enriched for TFBS ($P_{adjusted}$<0.001).

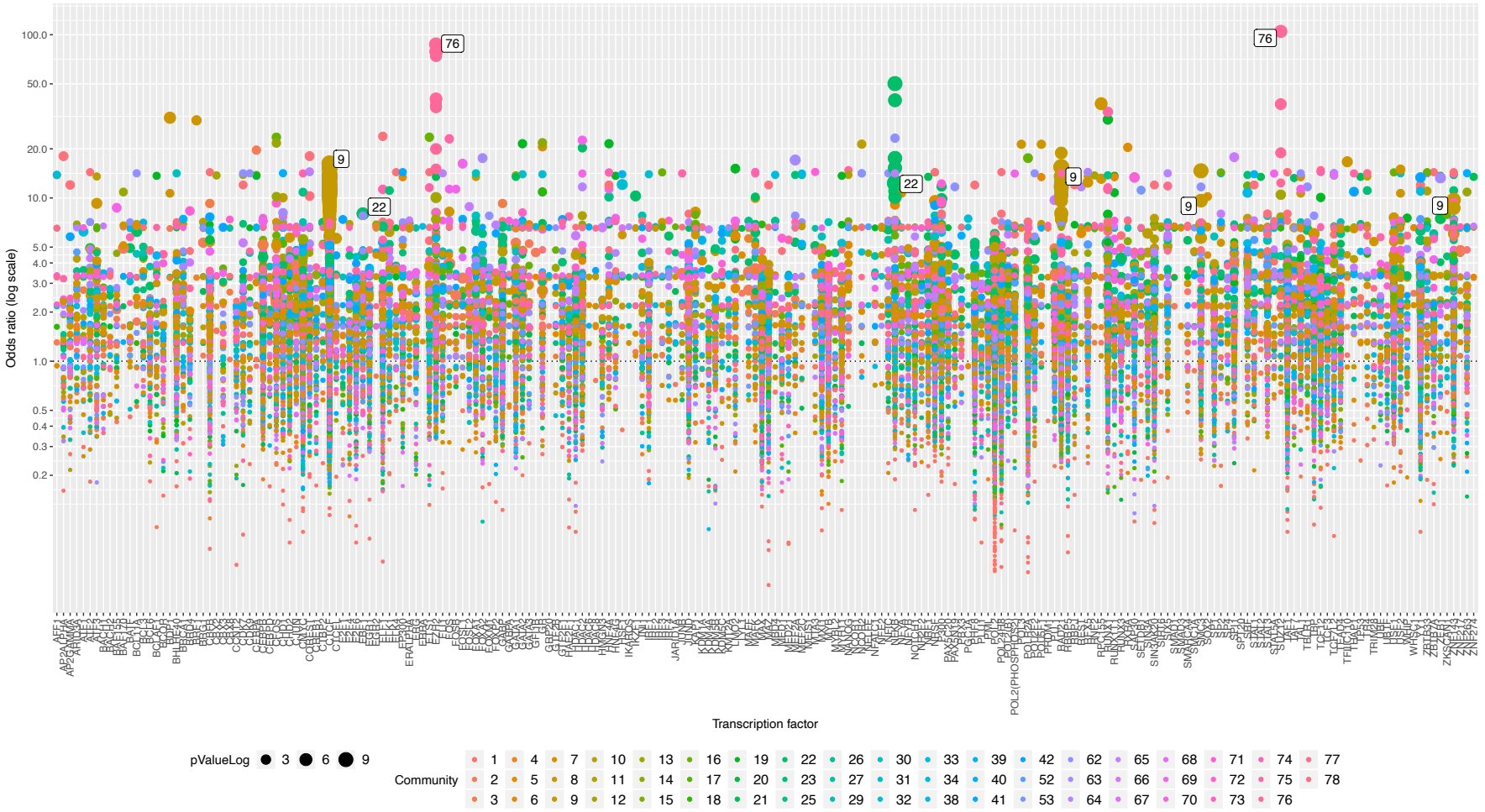


Fig S33: Enrichment of chromatin states among mQTL sites by community

The x-axis shows the chromatin states whereas the y-axis shows odds ratios on a logscale. Different colors represent different tissues. Enrichments are analysed using LOLA[6] and corrected for CpG and CG content. The size of the circles represents the -log10 Pvalue. Communities shown have FDR P<0.001 for at least one chromatin state.

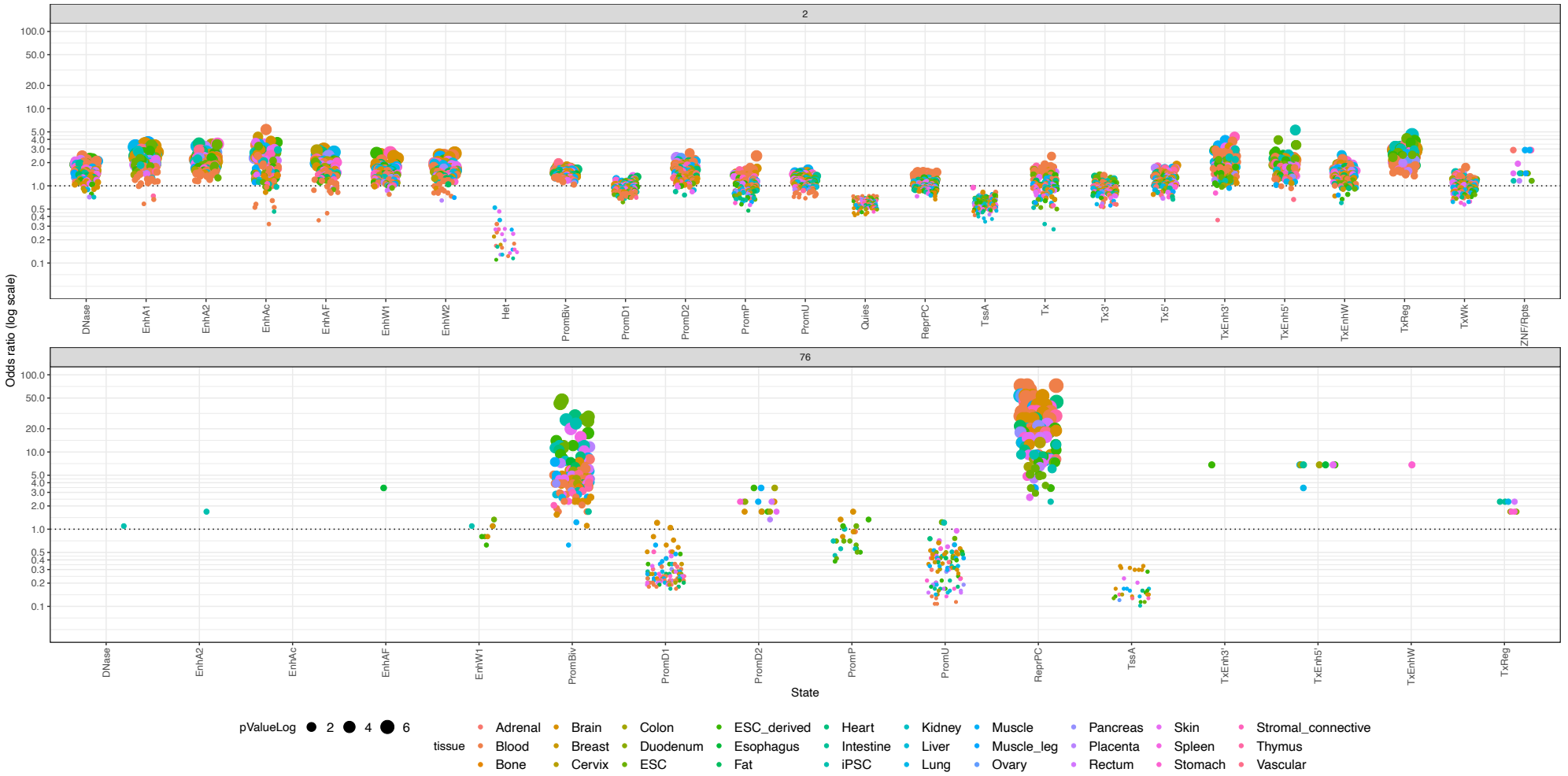


Fig S34: Enrichment of 36 blood GWA traits among community SNPs. The genomic loci for each of the 56 largest communities were tested for enrichment of low p-values in 36 complex trait GWASs[12]. The x-axis shows the blood trait whereas the y-axis shows odds ratios on a logscale. Different colors represent different communities. Enrichments are analysed using a similar method as implemented in GARFIELD[7] and adjusted for the number of proxies, MAF and distance to TSS. The size of the circles represents the -log10 Pvalue. Labelled communities are enriched (FDR P<0.001) for a blood trait GWA.

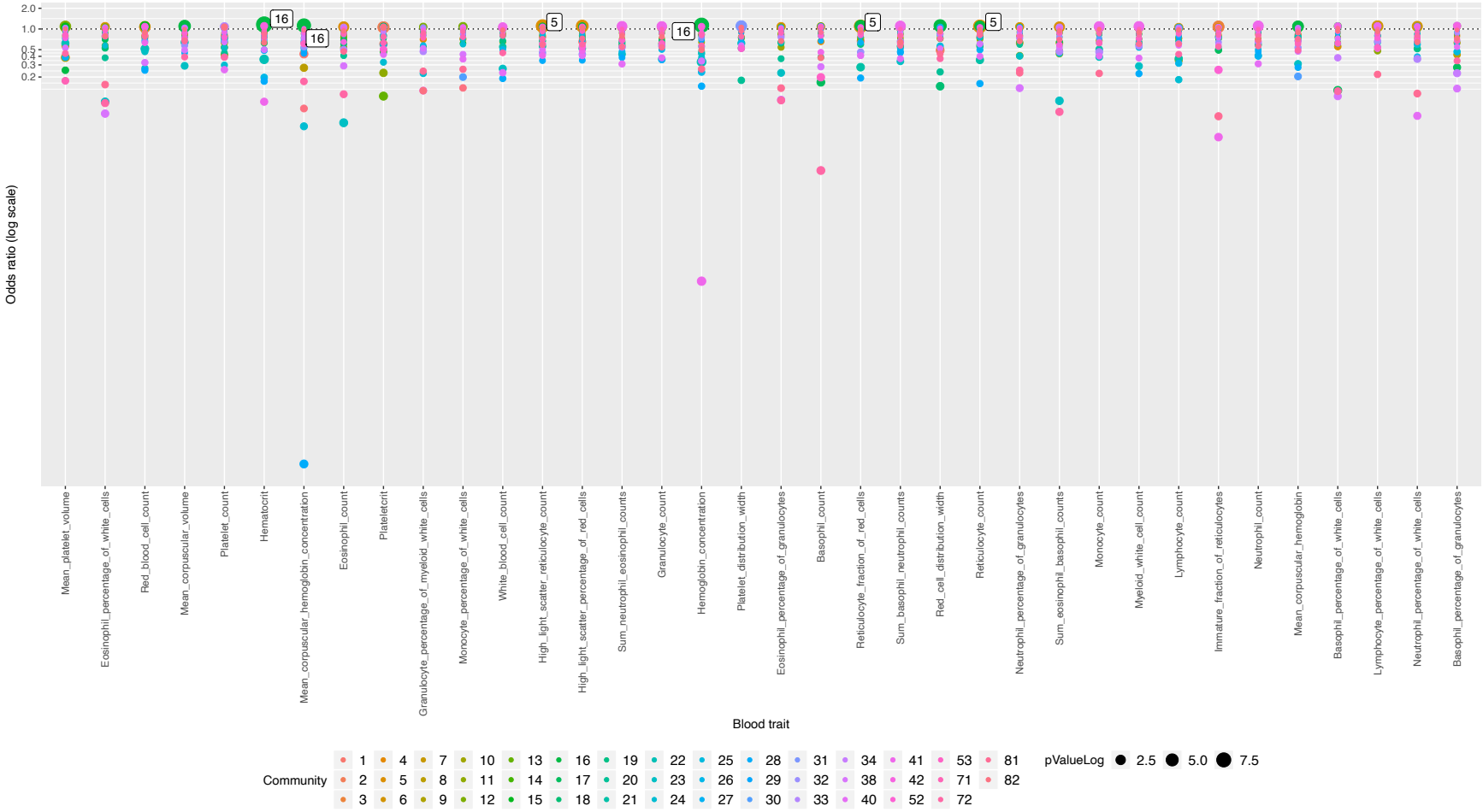


Fig S35: Enrichment of GWA analysis p-values among mQTL SNPs

Results for 41 GWA traits where variants associated with the phenotype (P < 5e-8) are significantly enriched in mQTL (P < 4.5e-4), where mQTL are classified into three annotation categories: *cis only*, *cis+trans* and *trans only*. A barplot of -log10 P value (top) and log odds ratio with 95% confidence intervals (bottom). This analysis was performed using GARFIELD[7]. Some *trans* mQTL enrichments had very large confidence intervals and odds ratios around 0 and were therefore not plotted.

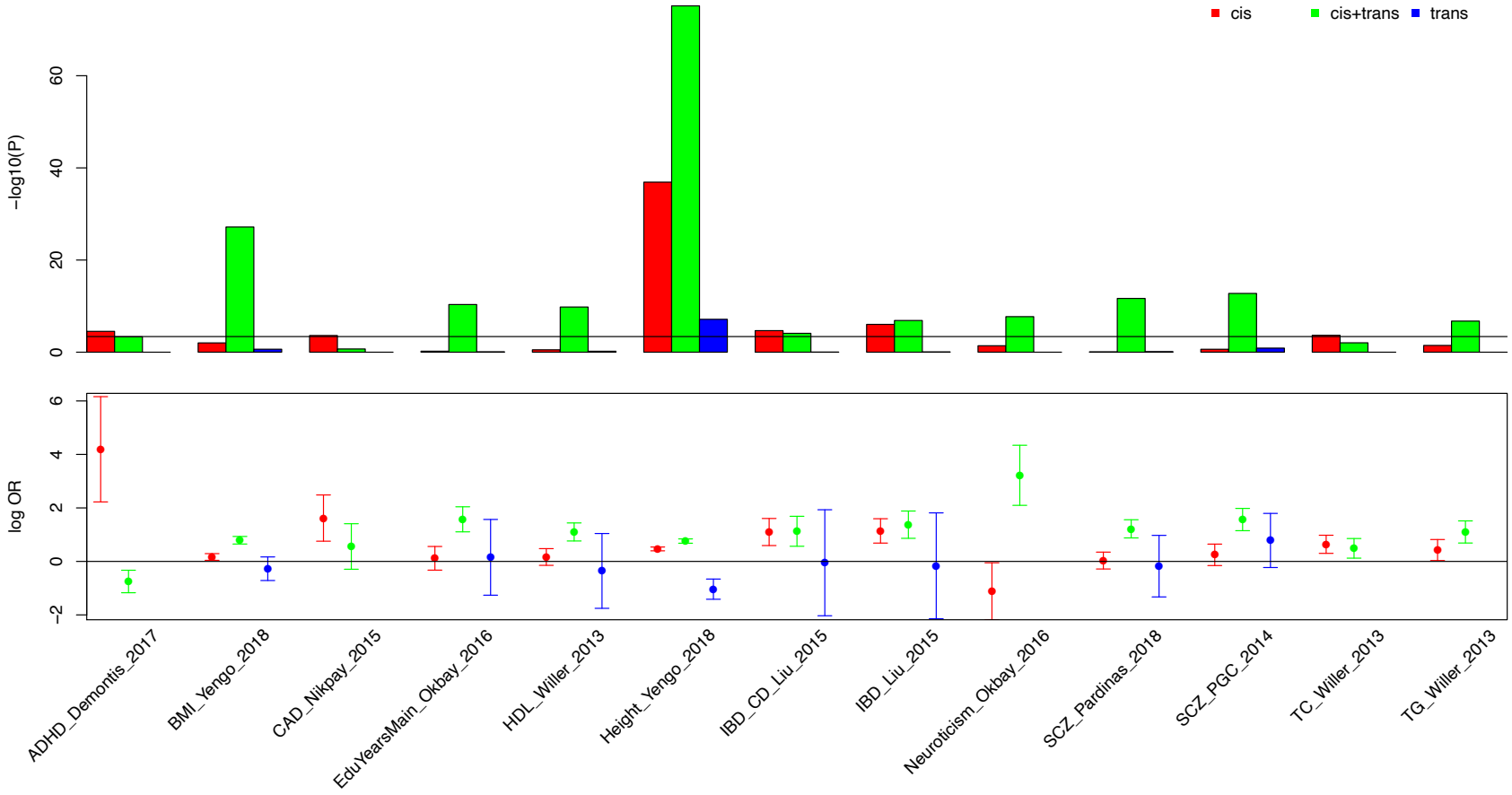


Fig S36: Enrichment of blood trait GWA analyses pvalues amongst mQTL SNPs.
Results for 36 blood traits GWAs[12] where variants associated with the phenotype are significantly enriched in mQTL. mQTL are classified into three annotation categories: *cis only*, *cis+trans* and *trans only*. A barplot of -log10 P value (top) and log odds ratio with 95% confidence intervals (bottom). Some *trans* mQTL enrichments had very large confidence intervals and odds ratios around 0 and were therefore not plotted.

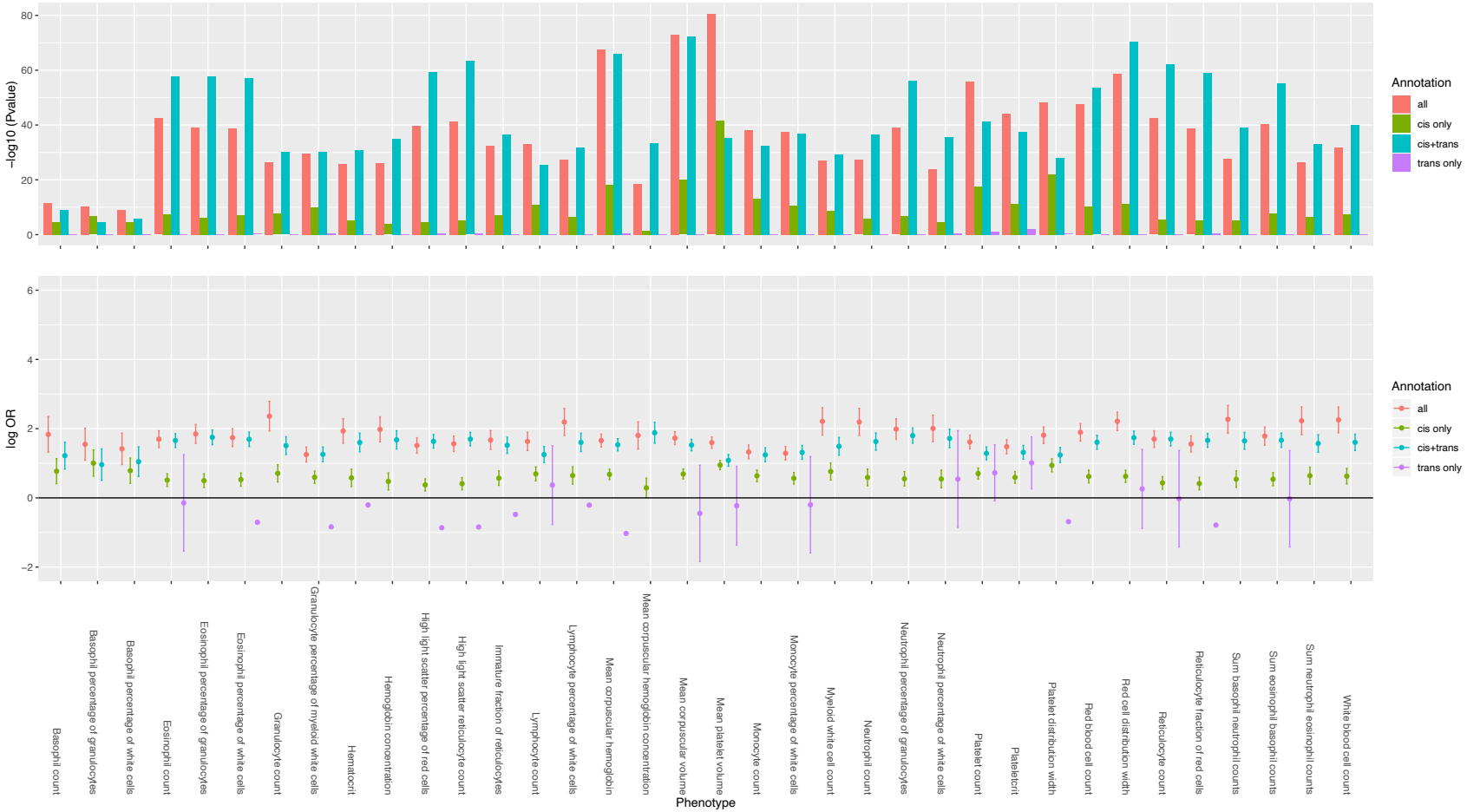


Fig S37: Correspondence of MR estimates amongst multiple independent instruments
To evaluate if a site having a shared causal variant with a trait was potentially due to the site being on the causal pathway to the trait, we reasoned that independent instruments for the site should exhibit consistent effects on the outcome consistent with the original colocalising variant. Amongst the putative colocalising signals, 440 involved a DNAm site that had at least one other independent mQTL. The plot shows the causal effect estimate estimated from the original colocalising signal against the causal effect estimates obtained from the independent variants.

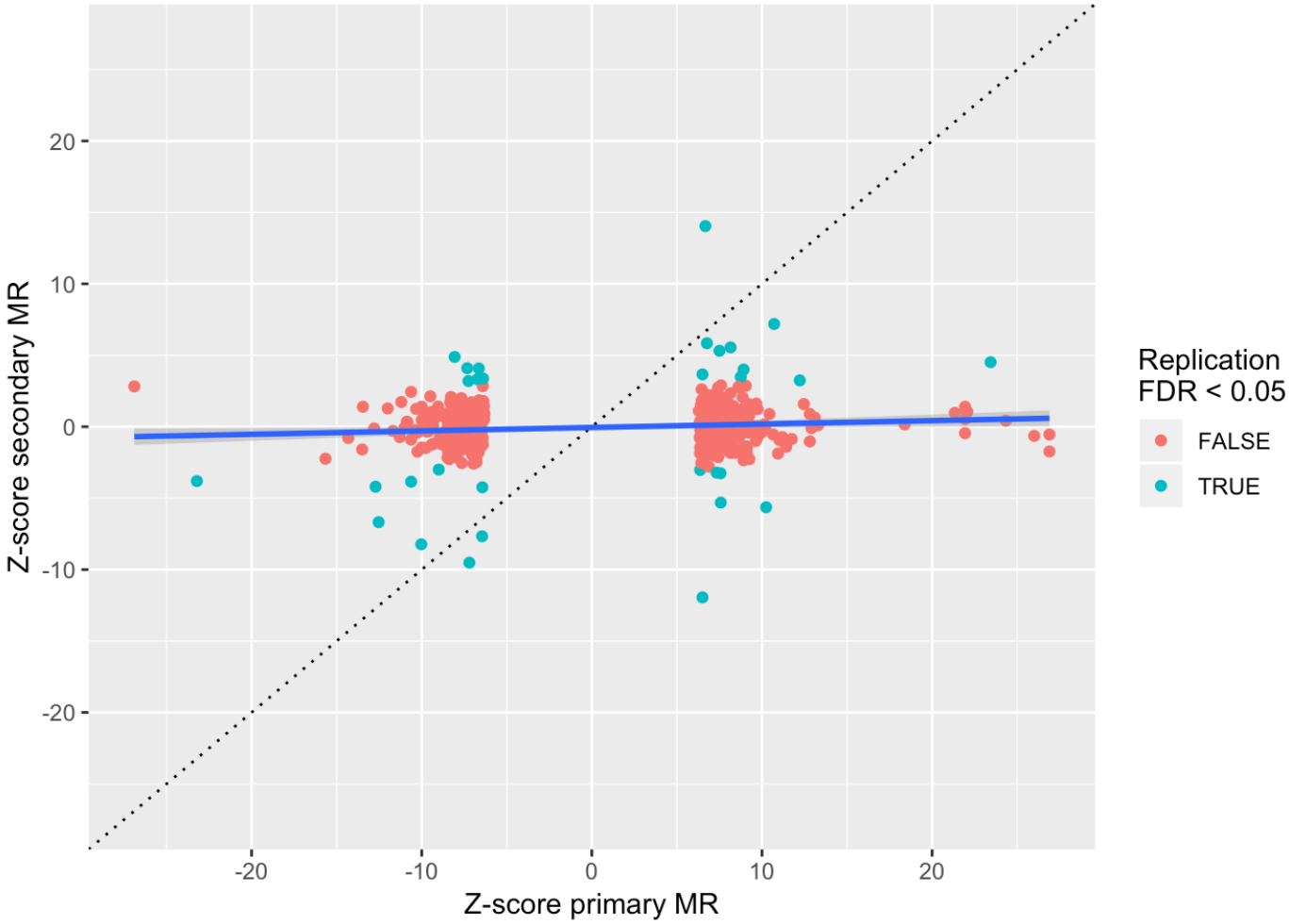


Fig S38: Correspondence of MR estimates amongst multiple independent instruments on 36 blood traits

To evaluate if a site having a shared causal variant with a blood trait[12] was potentially due to the site being on the causal pathway to the trait, we reasoned that independent instruments for the site should exhibit consistent effects on the outcome consistent with the original colocalising variant. Amongst the putative colocalising signals, 30% involved a DNAm site that had at least one other independent mQTL. The plot shows the causal effect estimate estimated from the original colocalising signal against the causal effect estimates obtained from the independent variants. The *HLA* region has been removed and betas are plotted.

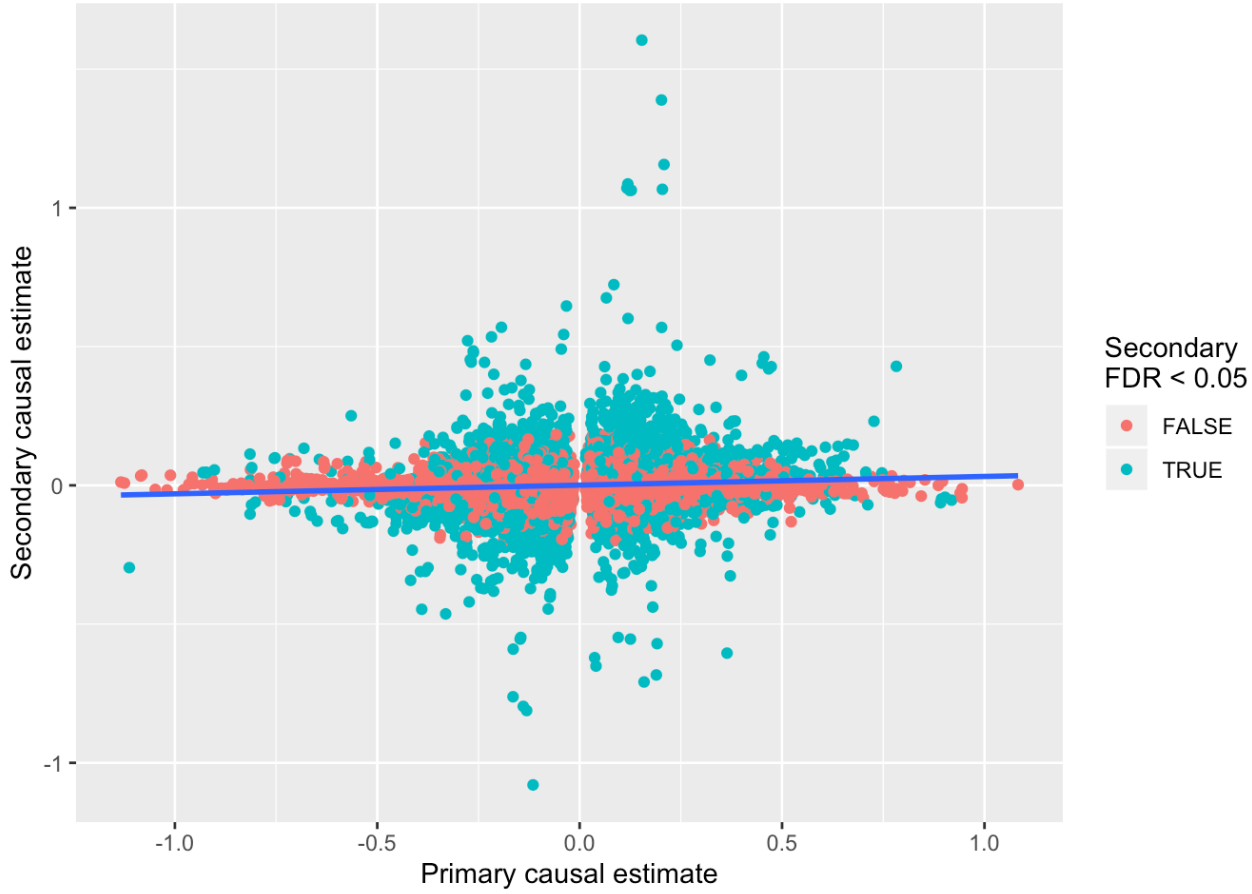


Fig S39: Flow chart illustrating the decision process for filtering results from MR of traits on DNAm sites. The process comprised 10 filtering steps as indicated in brackets.

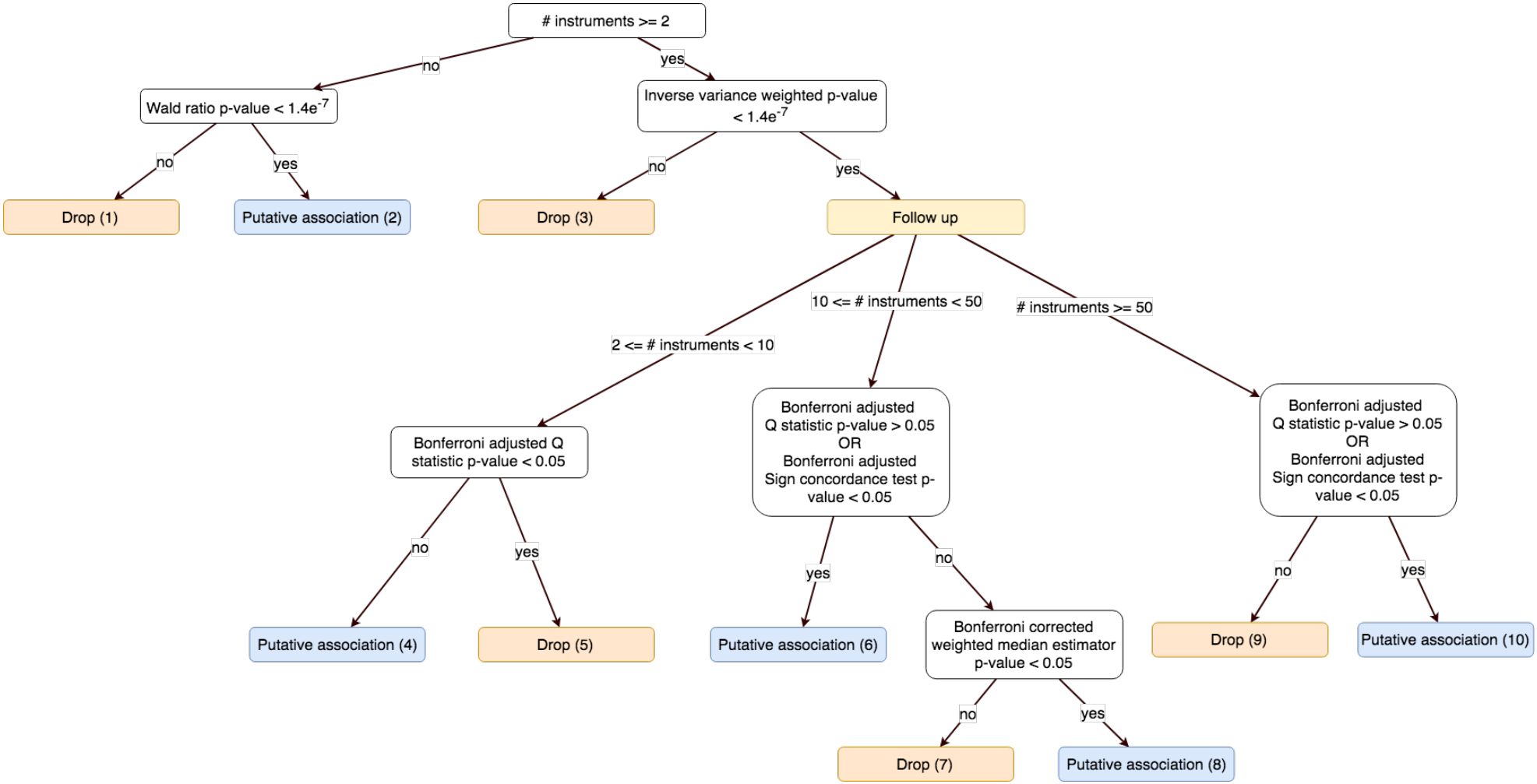


Fig S40: Genomic inflation factors for genome-wide scans of causal effects of traits on DNAm sites
Each trait (x axis) was tested for causal effects against (on average) 317,659 DNAm sites, excluding sites in the *MHC* region. The p-values from IVW MR analysis were used to estimate the genomic inflation for each trait (y-axis). Traits are ordered by genomic inflation factor.

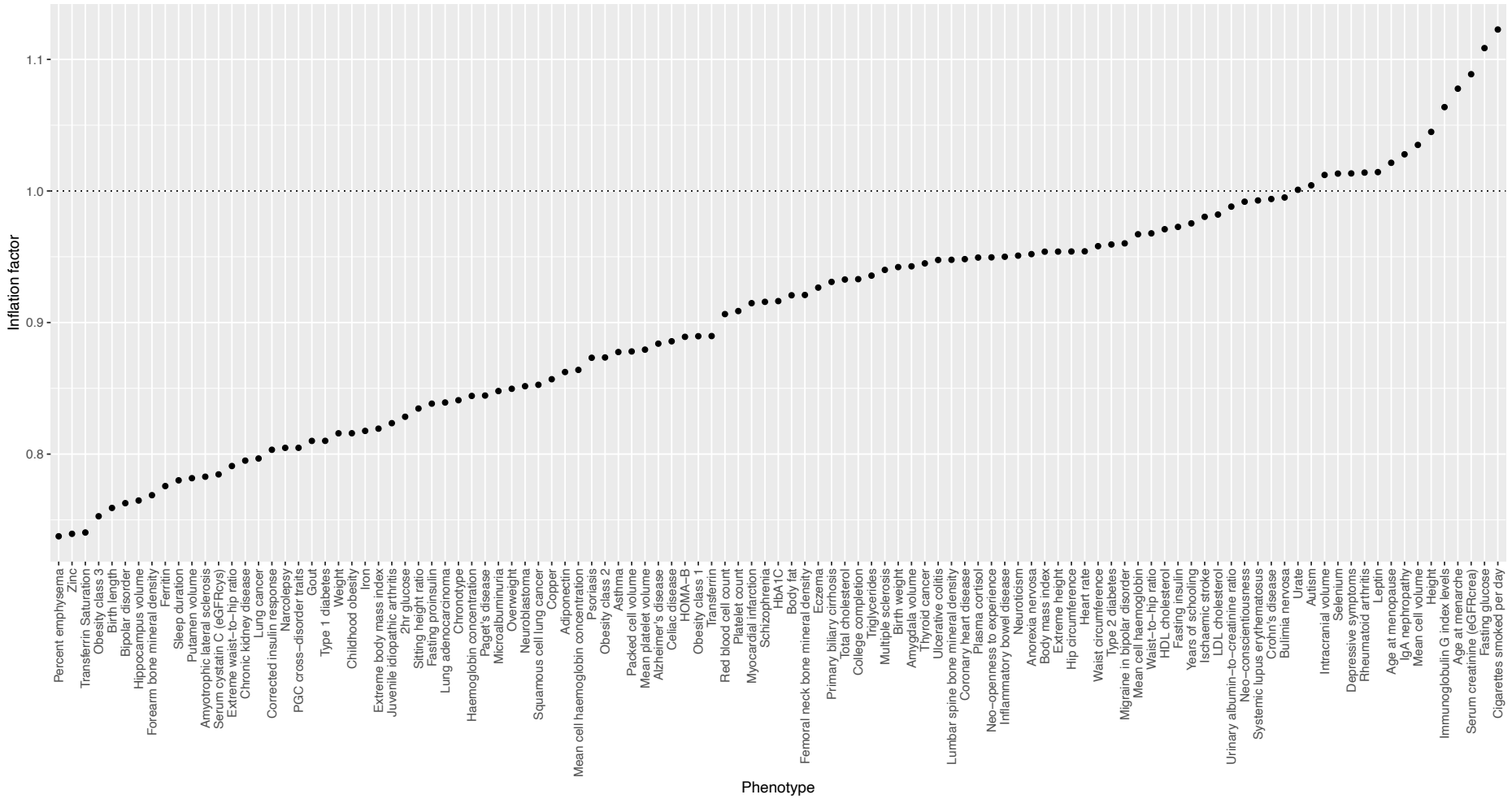


Fig S41: Genomic inflation factors split by chromosome
The five traits with genomic inflation factors higher than 1.05 (red lines, y-axis) are shown with a per-chromosome estimate of the inflation factors.

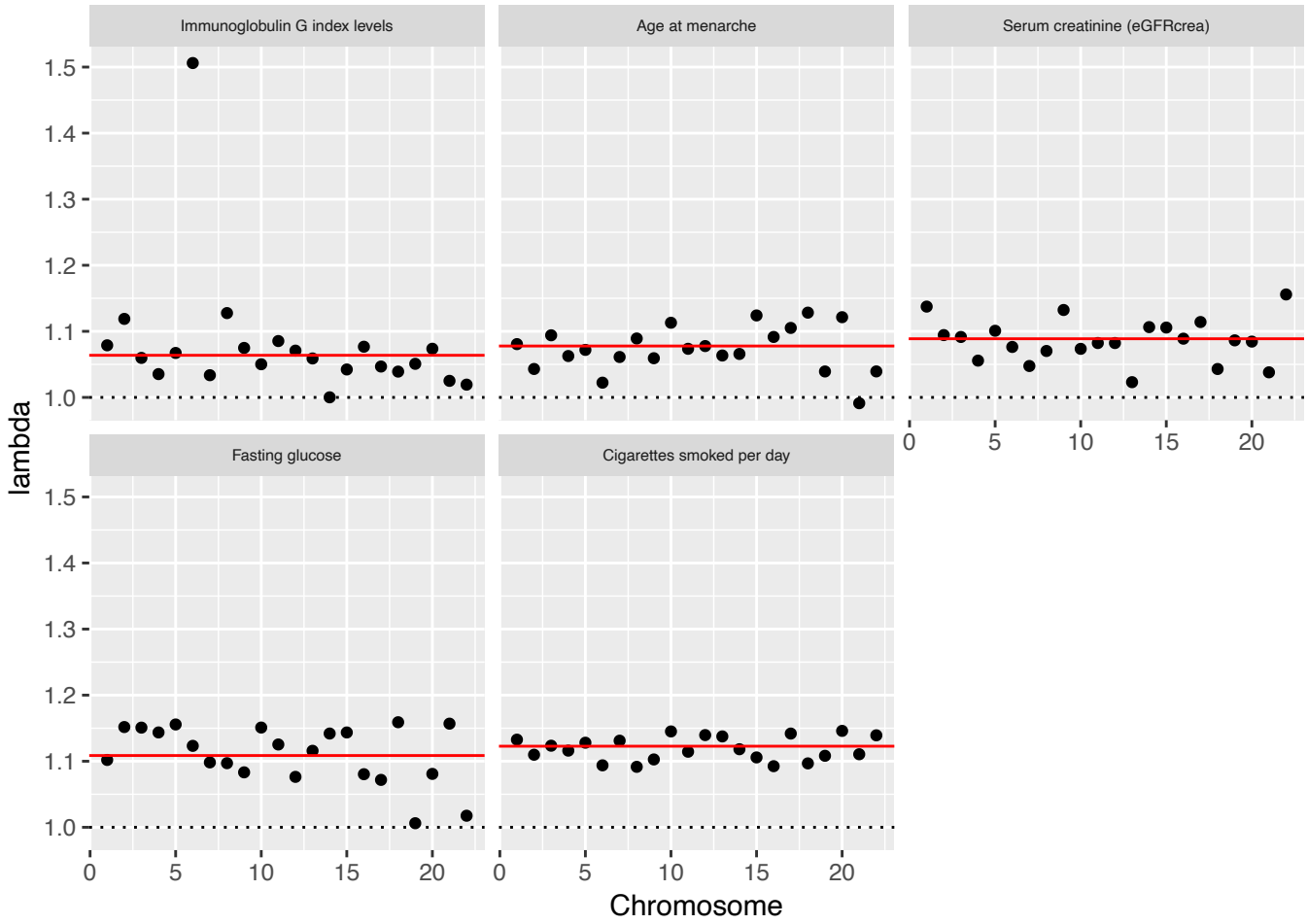


Fig S42: Leave-one-chromosome-out analysis of genomic inflation factors
Each of the five traits with genomic inflation factors higher than 1.05 were re-analysed, sequentially leaving out one chromosome at a time (x-axis) to demonstrate that the inflation was not being driven by a single chromosome.

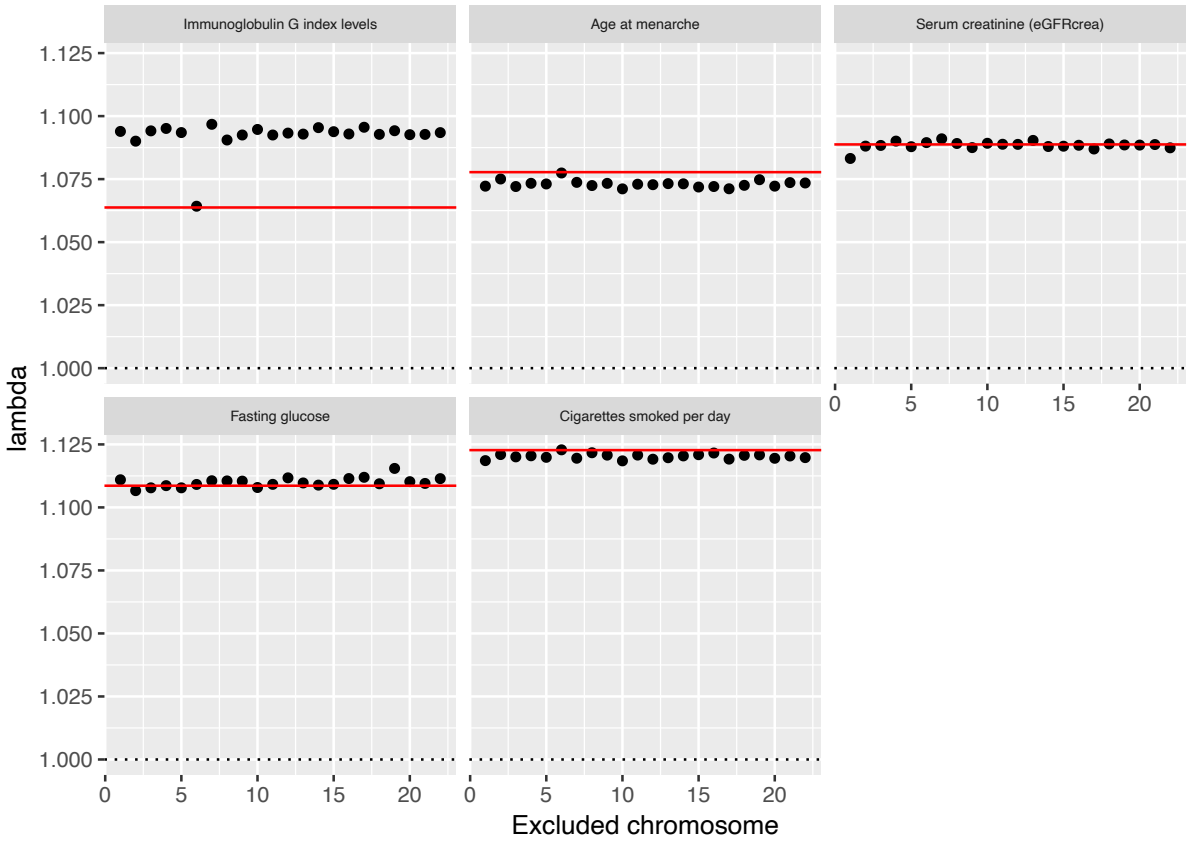


Fig S43: Relationship between strongest effect size for each mQTL SNP and selection signatures

Relationship between maximum absolute effect (MAF adjusted) for *cis* mQTL SNPs, *cis+trans* mQTL SNPs and *trans* mQTL SNPs and selection signals (SDS, F_{st}, iHS, XPEHH in CHB/YRI). Circles represents *cis*, *cis+trans* and *trans* mQTL SNPs. Red and blue lines and p-values represent linear regression and GAM smoothing curve between absolute effect sizes and selection scores adjusted for number of proxies, distance to TSS, GC and CpG frequency.

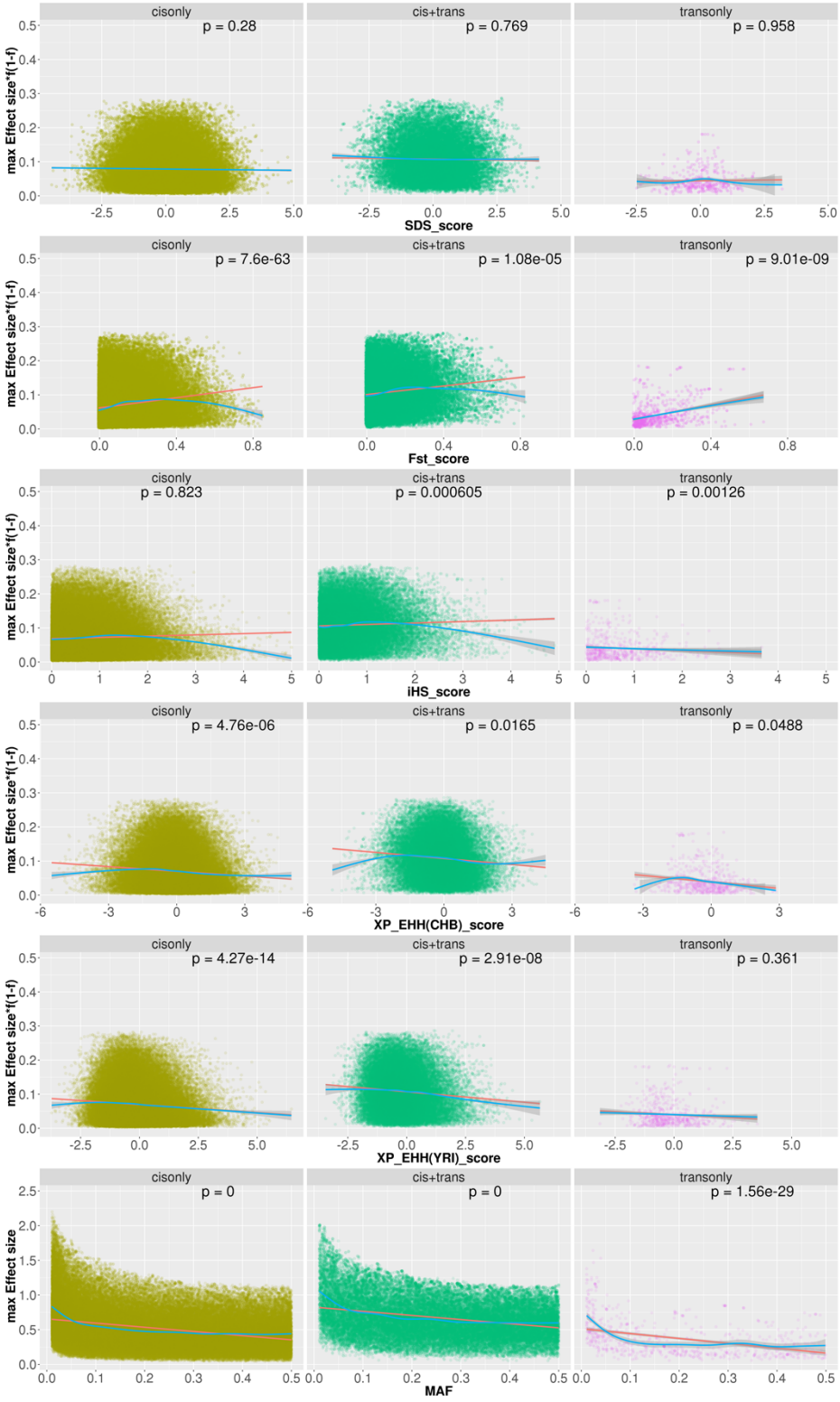


Fig S44: Genetic Architecture of genome-wide DNAm. a-b) show the loss according to the model fit described in Methods, which is based on ensuring that different MAF categories contain the expected amount of variants in the upper tail of the effect-size distribution, given a model of the form $var(\beta) = \sigma^2[f(1-f)]^{-S}$a) shows the loss when comparing how many β's are more than 2 standard deviations above the mean to more than $\sqrt{2}$ above. b) shows this for (1,2) standard deviations respectively. c-e) shows how the model predicts the thresholds will fit on the data, showing c) a "too shallow" value of S, d) the inferred value of S, and e) a "too steep" value of S. f-h) illustrate this by comparing the number of counts above the lower threshold (blue + red) to those above the upper threshold (red), compared to the expected proportion under the model (straight line). In all cases, the best σ is chosen for each S.

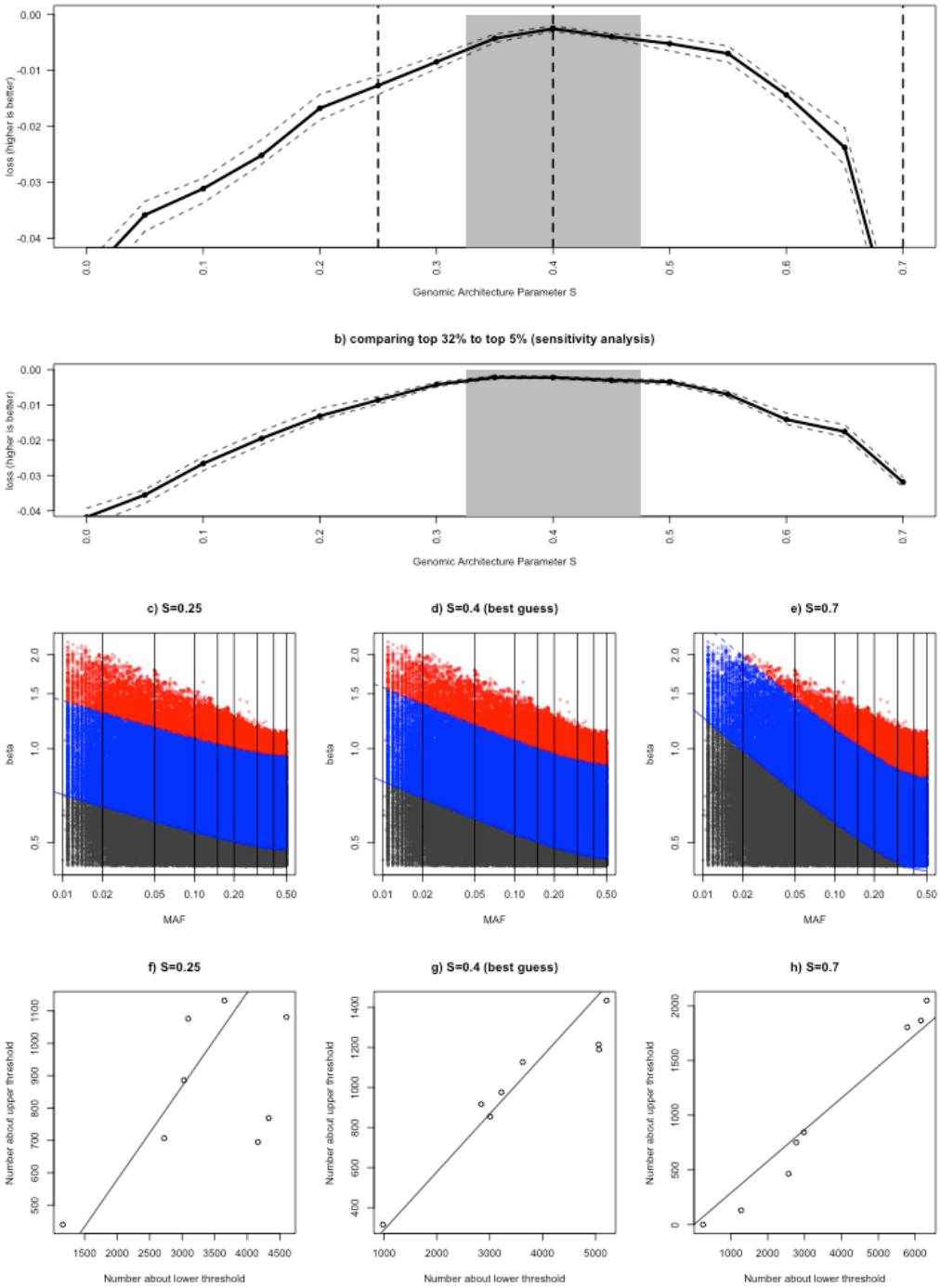
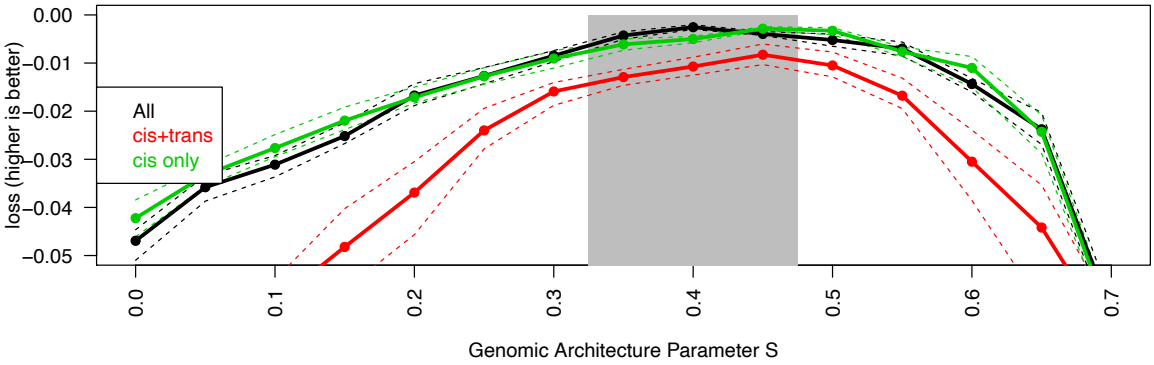


Fig S45: Genetic Architecture of genome-wide DNAm in *cis* and *trans*
Comparison of *cis+trans*, all SNPs, and *cis only* as Figure S33 a-b, but with the genetic architecture of the different subsets of SNPs. In all cases the inferred S is around 0.4, but when either *cis+trans* or *cis only* are considered the inferred S might be slightly larger but still within the confidence interval.



b) comparing top 32% to top 5% (sensitivity analysis)

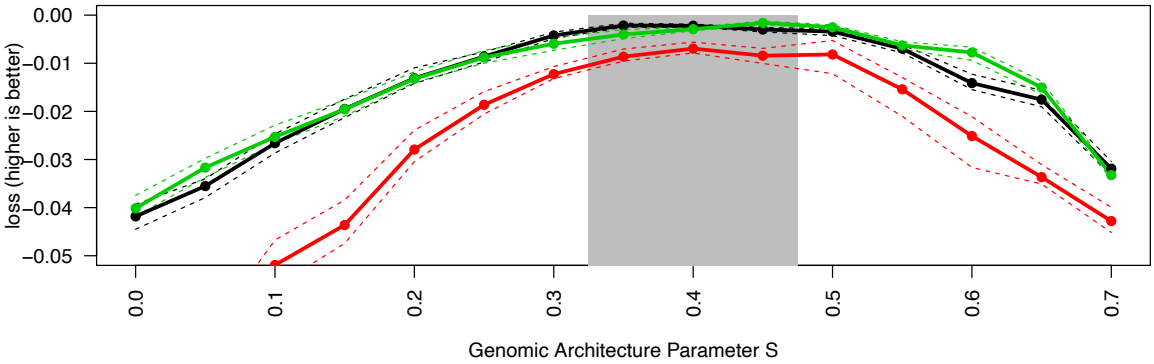


Fig S46: The contribution of GWA SNPs to genetic variance of five complex traits.
Plots show the contribution of all GWA SNPs, GWA SNPs overlapping a mQTL and all GWA SNPs with extreme SDS overlapping a mQTL to genetic variance of extreme height[13], height[14], cardiovascular disease (CVD)[15], Crohn's disease[16] and schizophrenia[17]. SNPs located in the *LCT* and *MHC* regions are removed. All GWA SNPs are extracted from relevant publications; GWA SNPs overlapping a mQTL and GWA SNPs with extreme SDS overlapping a mQTL are extracted using GARFIELD.

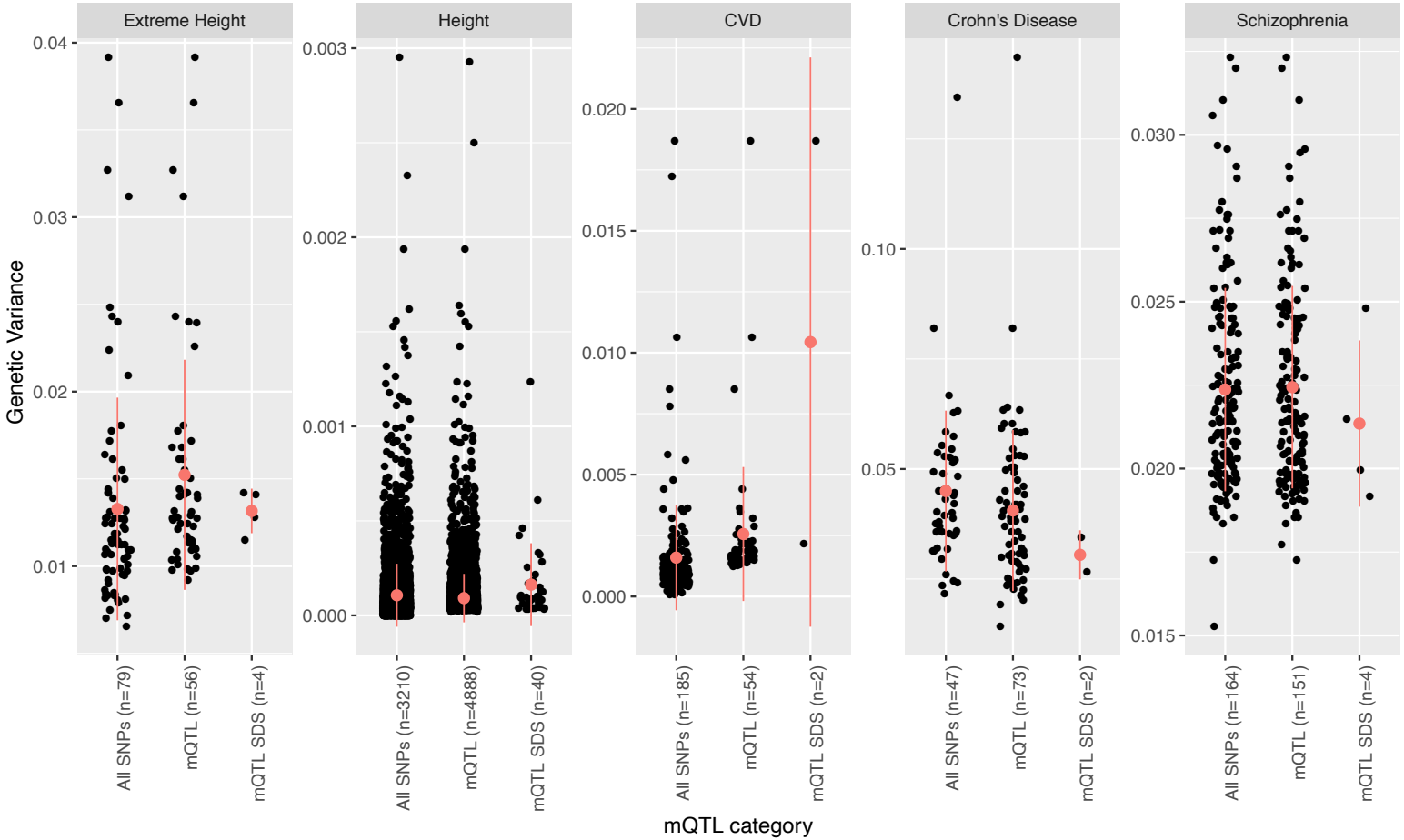


Fig S47: Relationship between number of probes per gene and number of mQTL. 331,884 DNAm sites were annotated to 18,993 protein-coding genes. For each of the genes, we plotted the number of probes on the 450k array (x-axis) against the number of *cis* and *trans* mQTL (y-axis).

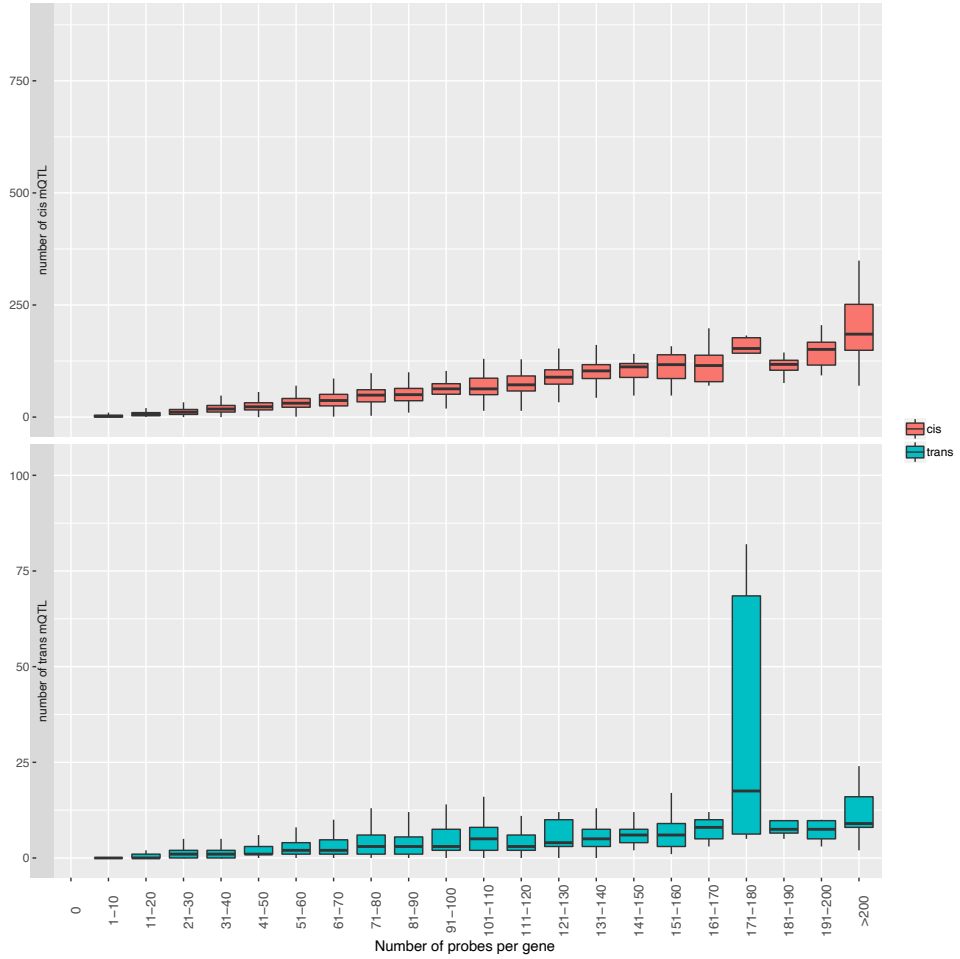


Fig S48: Loss in power in two stage design. We calculated the power of detecting a cis association in at least one of the 22 studies at p<1e-5 or a trans association in at least two of 22 studies at p<1e-5.

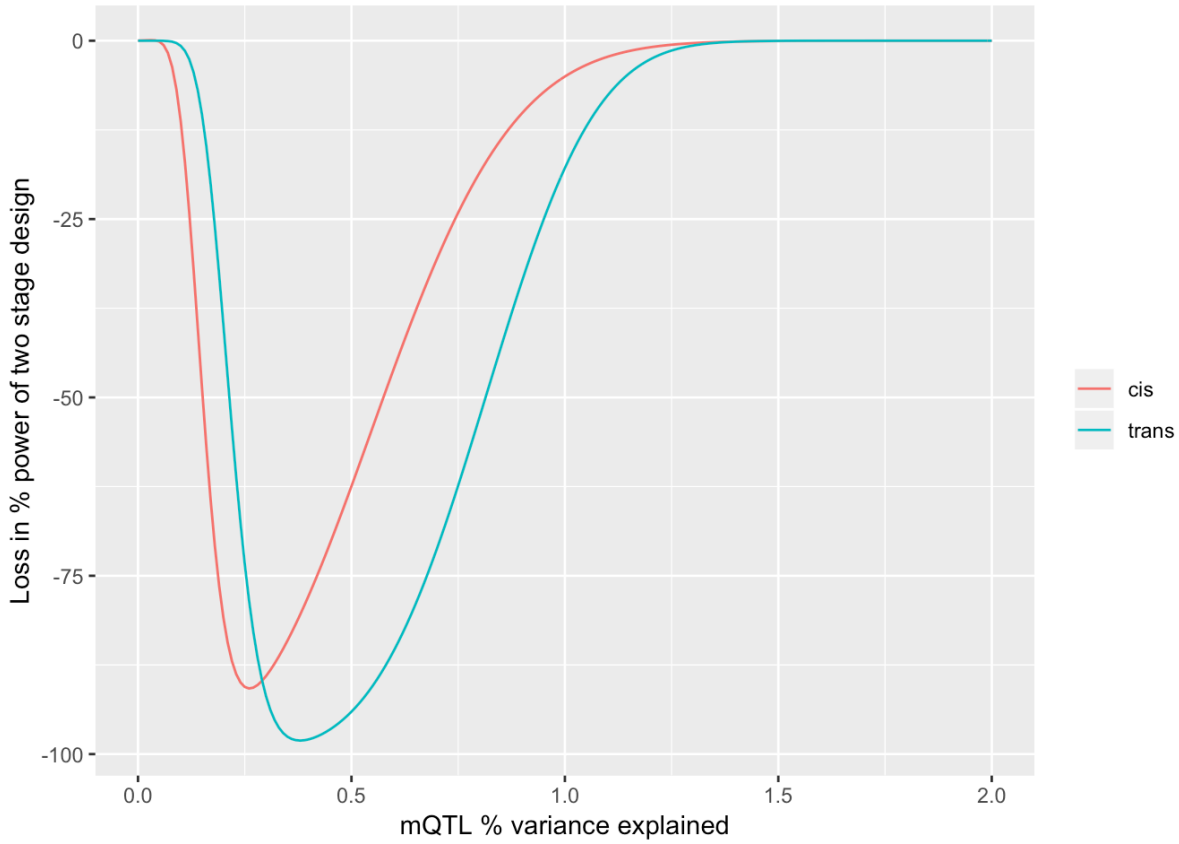


Fig S49: Expected number of mQTL
Using the number of mQTL with a particular r^2 value, and the power of detecting mQTL with that r^2 value, we calculated how many mQTL would expect to exist with that value.

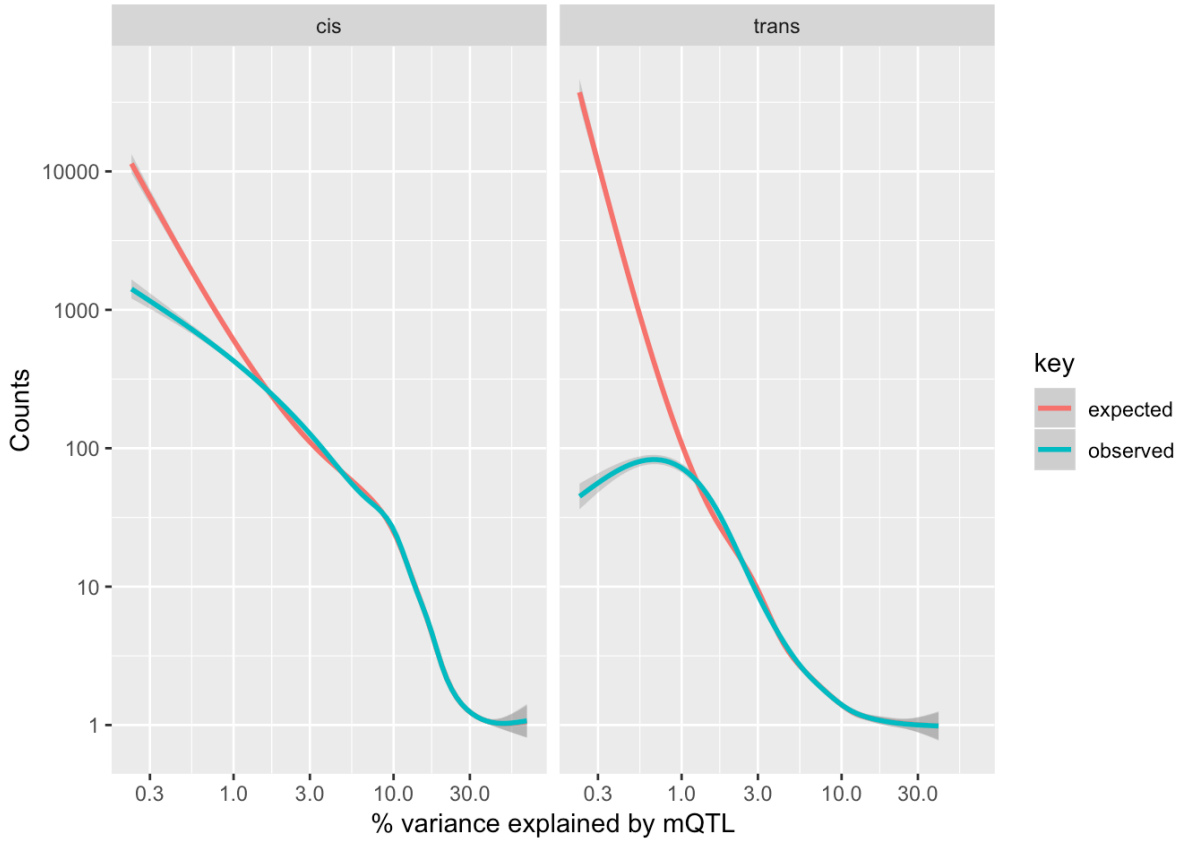


Fig S50: Comparison of enrichments in gene annotations for *cis+trans*, *cis only* and *trans* only sites using >1 Mb versus cross chromosomal to define a *trans* association. Plots show odds ratios of enrichments without chromosome 6. Analyses are adjusted for CpG and GC content using LOLA[6]. Colors represent gene annotations.

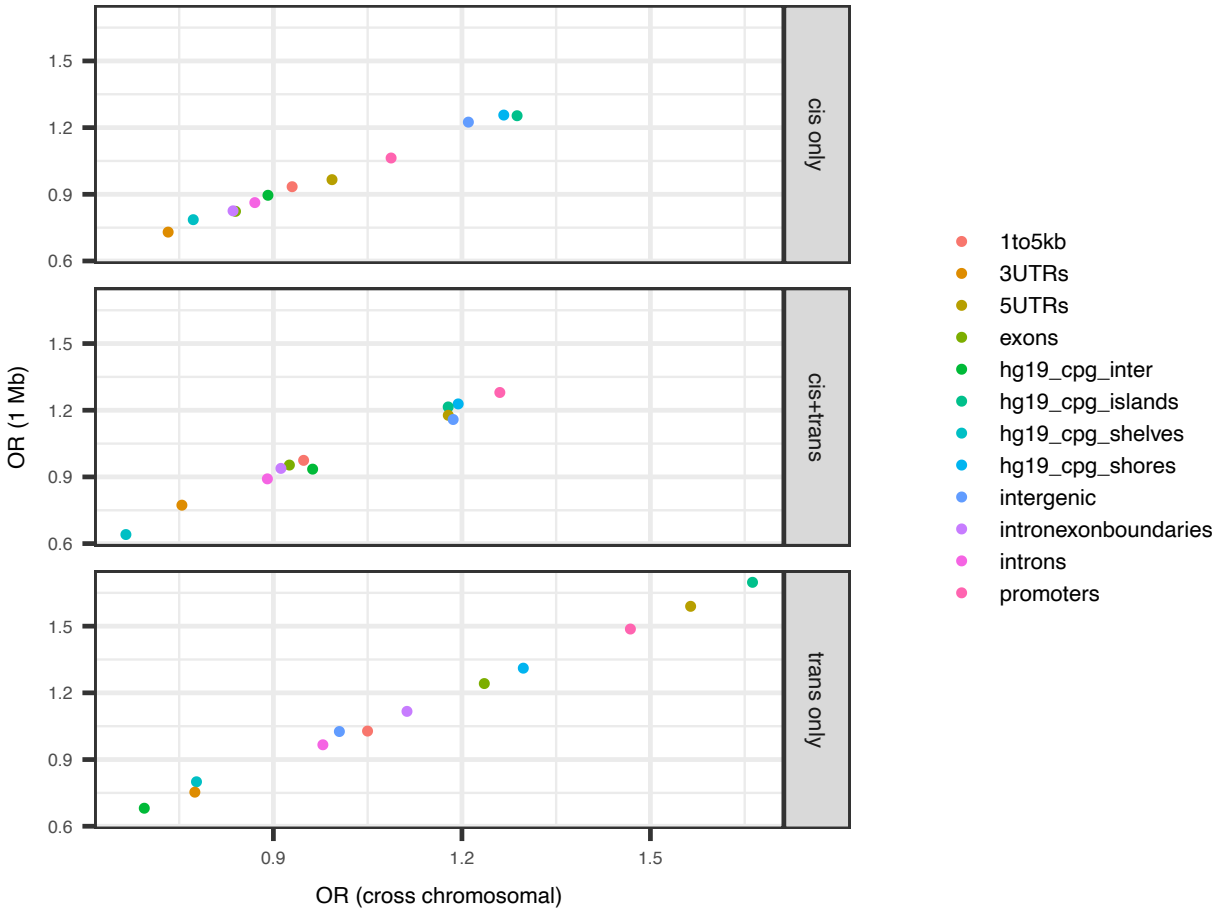


Fig S51: Comparison of enrichments in TFBS for *cis+trans*, *cis only* and *trans* only sites using >1 Mb versus cross chromosomal to define a *trans* association.
Plots show odds ratios of enrichments without chromosome 6. Analyses are adjusted for CpG and GC content using LOLA[6].

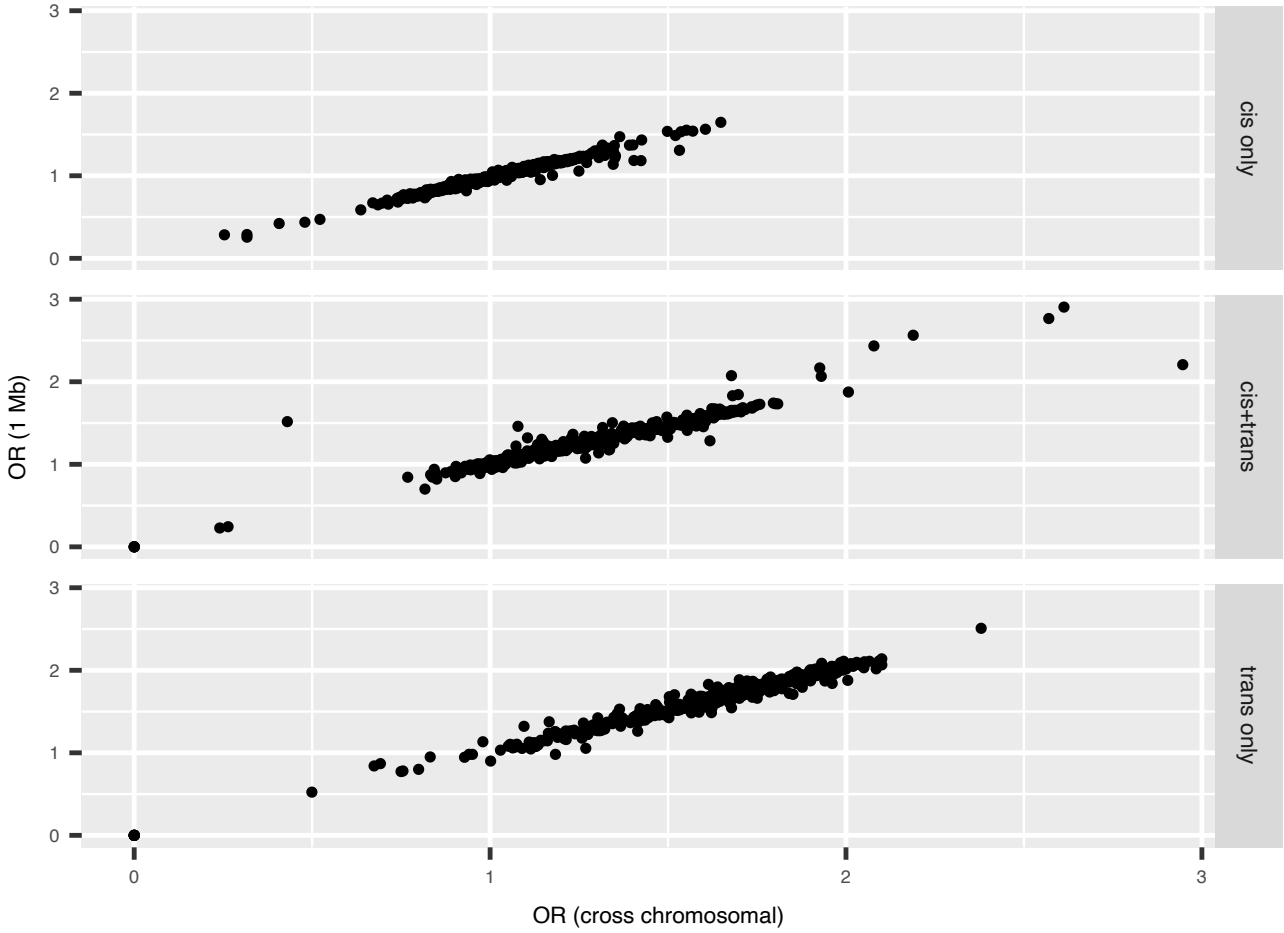


Fig S52: Between study-heterogeneity of 36 blood studies, an adipose and a brain study. We used 337 independent SNPs on chromosome 20 with a pvalue<1e-14 to calculate the Mstatistic[1] for each cohort. Study 9 is the brain study; study 2 is the adipose study.

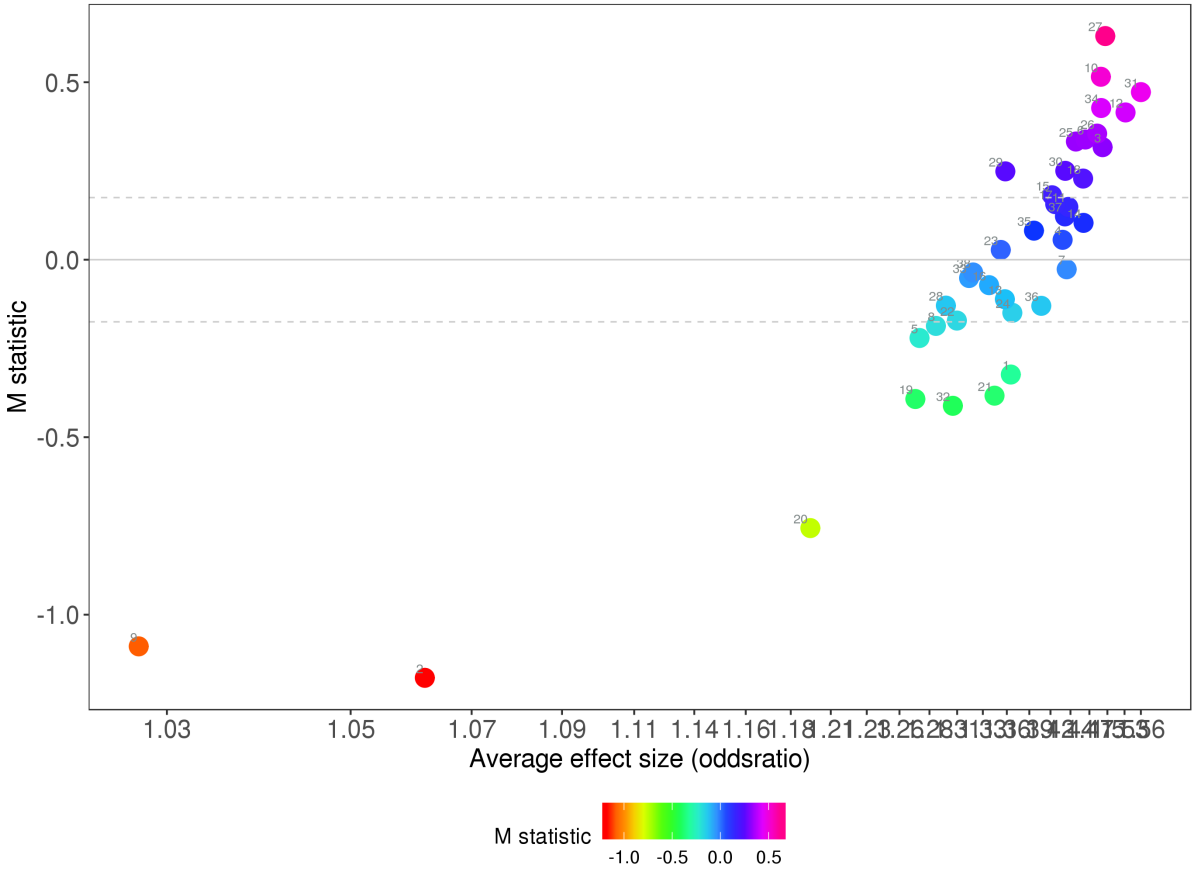


Fig S53: Simulation study recovering genomic architecture. We simulated N SNPs and retained the fraction r that had the largest variance explained. We then fit the architecture parameters S and σ using our method (see text). a) shows the root-mean-squared error for different simulation settings, with lower values being good (x and y locations jittered for illustration). b) shows the mean bias over all S; values close to 0 are good. Larger N or r both result in improved performance. c-d) show two examples for specific simulation settings, showing how inferred S varies with the true simulated S. In c) the inference is accurate and approximately unbiased. In d) there is less data, so inference is usually accurate, but some simulations recover $S{\sim}0$ regardless of the true S. The observed data have significantly larger N and an equivalent r, and therefore are expected to be accurately represented.

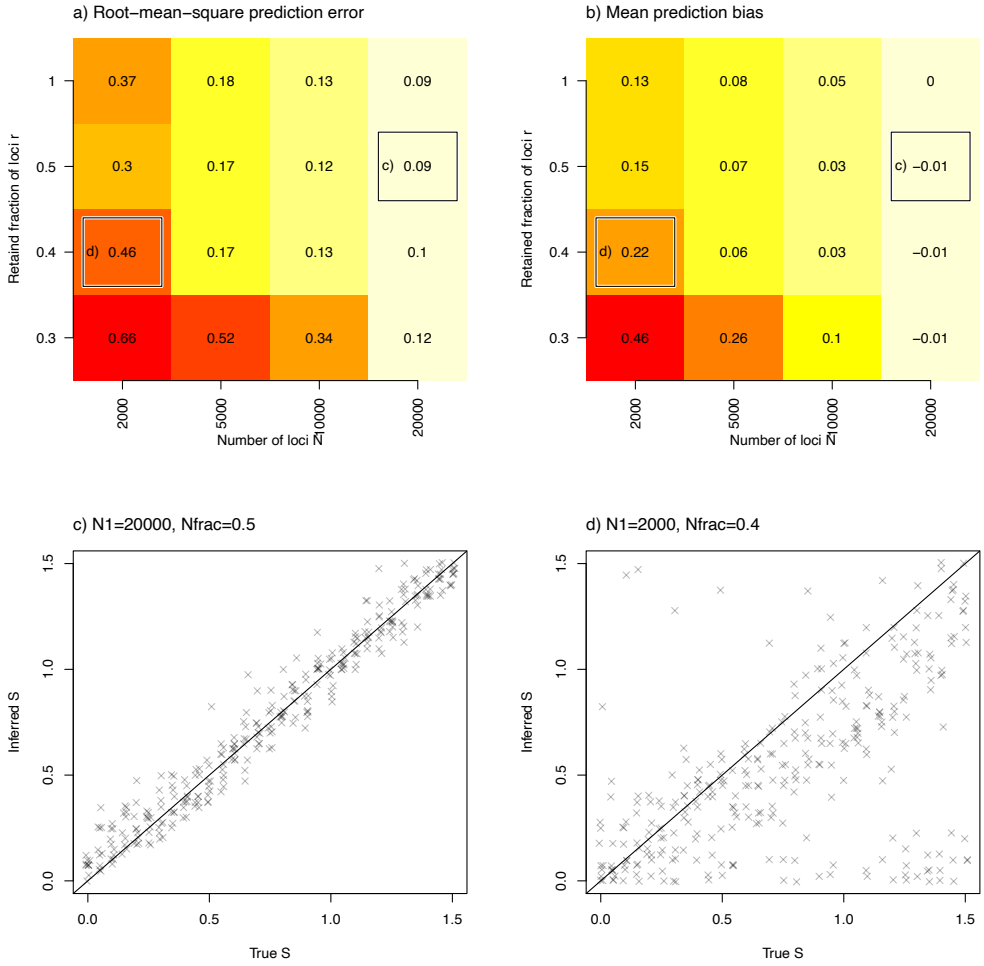
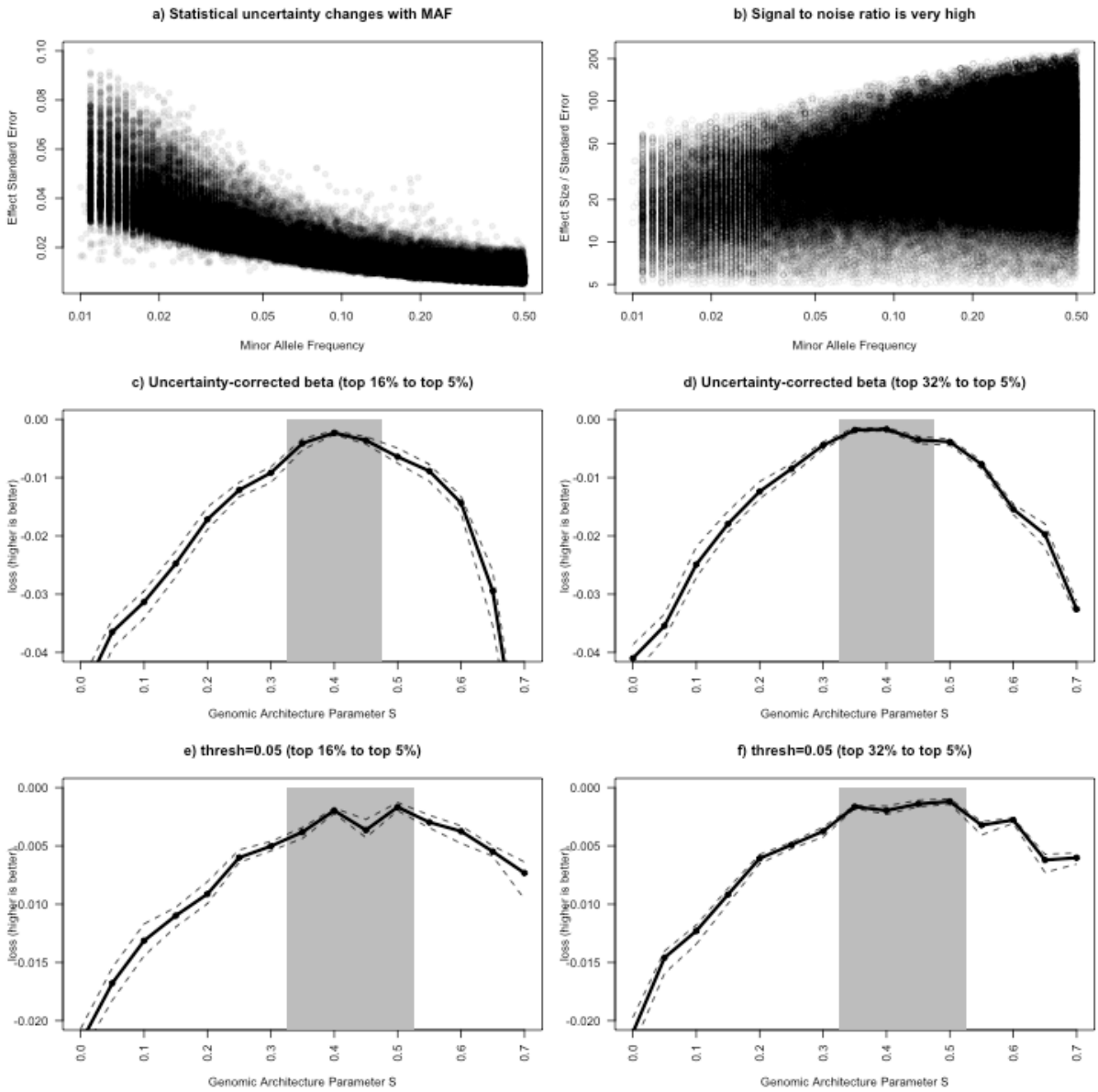


Fig S54: Genetic architecture effects dominate statistical uncertainty. a) Statistical uncertainty (i.e. the standard error of the effect size) is a function of Minor Allele Frequency, which could theoretically confound inference of genomic architecture. However, b) the signal-to-noise ratio is high in our data, at least 5 and averaging around 20. Replacing estimates by a conservative uncertainty coerrected version (c-d), i.e. beta' = sign(beta)*(|beta| - se(beta)) leads to the same inferences of S=0.4 as the raw data (**Figure S44**). Low frequency variants are both the most informative and the most uncertain. Removing these (using a threshold of 0.05, e-f) leads to similar inference (S between 0.375 and 0.525). Figures d and f measure the same quantity as Figures c and e respectively, using different statistical thresholds for sensitivity analysis exactly as in **Figure S44**.



References

1 Magosi, L. E., Goel, A., Hopewell, J. C., Farrall, M. & Consortium, C. D. Identifying systematic heterogeneity patterns in genetic association meta-analysis studies. *PLoS Genet* **13**, e1006755, doi:10.1371/journal.pgen.1006755 (2017).

2 Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nature genetics* **49**, 131-138, doi:10.1038/ng.3721 (2017).

3 Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome biology* **17**, 61, doi:10.1186/s13059-016-0926-z (2016).

4 McRae, A. F. *et al.* Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome biology* **15**, R73, doi:10.1186/gb-2014-15-5-r73 (2014).

5 van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* **7**, 11115, doi:10.1038/ncomms11115 (2016).

6 Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587-589, doi:10.1093/bioinformatics/btv612 (2016).

7 Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature genetics* **51**, 343-353, doi:10.1038/s41588-018-0322-6 (2019).

8 Reynolds, L. M. *et al.* Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun* **5**, 5366, doi:10.1038/ncomms6366 (2014).

9 Slieker, R. C. *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6**, 26, doi:10.1186/1756-8935-6-26 (2013).

10 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

11 Elliott, H. R. *et al.* Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clinical epigenetics* **6**, 4, doi:10.1186/1868-7083-6-4 (2014).

12 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e1419, doi:10.1016/j.cell.2016.10.042 (2016).

13 Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**, 501-512, doi:10.1038/ng.2606 (2013).

14 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649, doi:10.1093/hmg/ddy271 (2018).

15 Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130, doi:10.1038/ng.3396 (2015).

16 Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986, doi:10.1038/ng.3359 (2015).

17 Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).