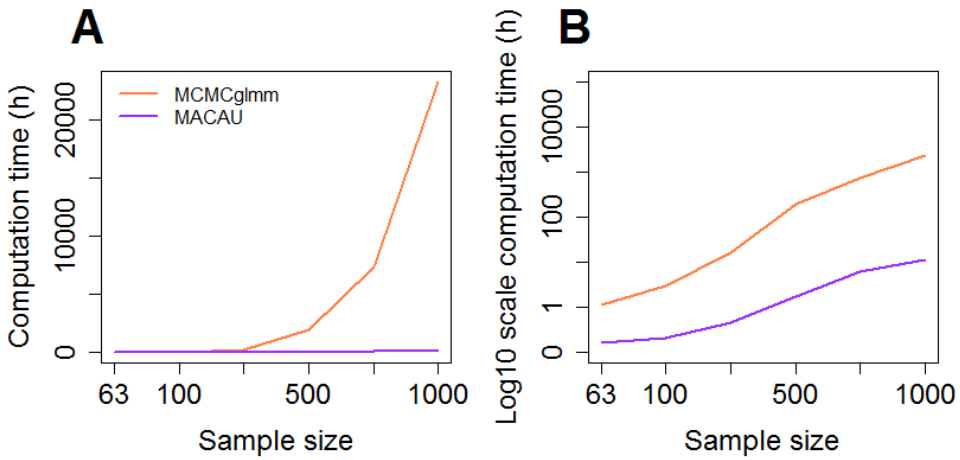
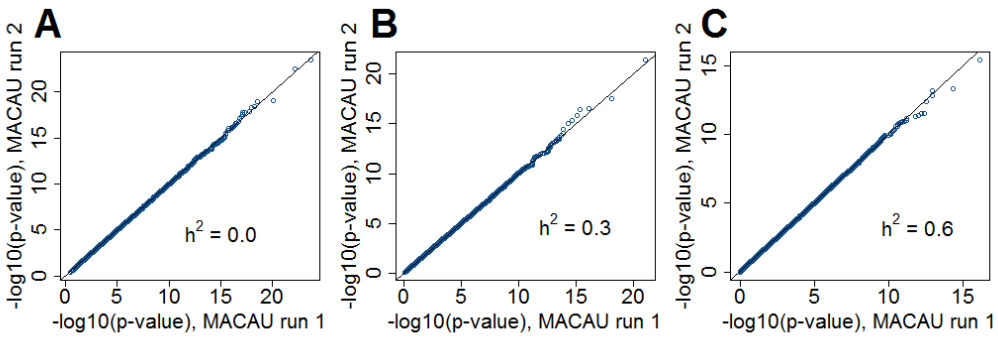
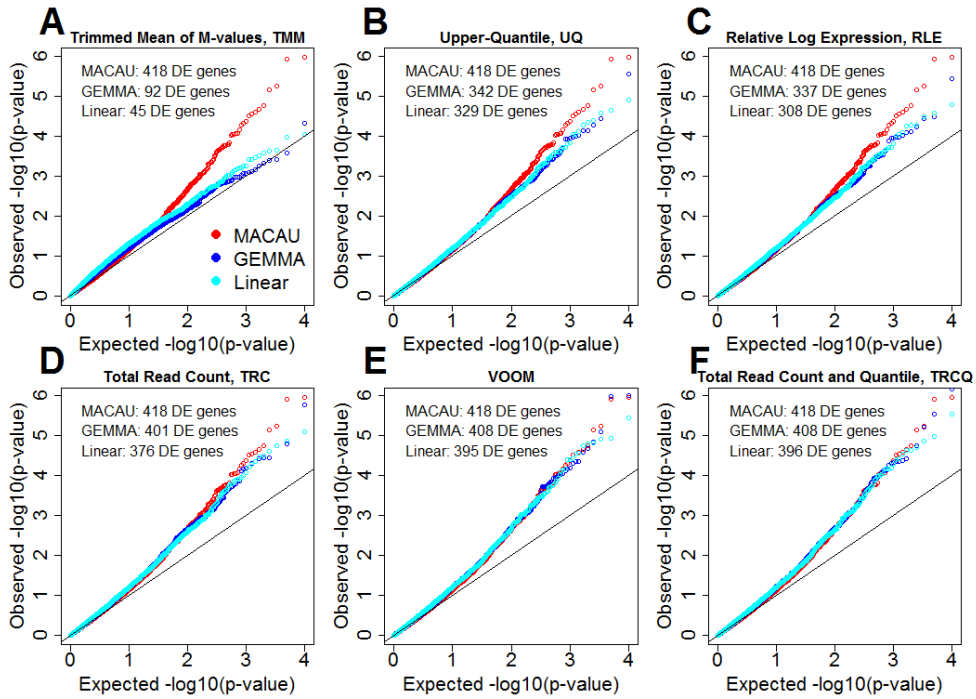
Supplementary Figure 1. A Poisson mixed model (PMM) implemented in MACAU is more computationally efficent than a PMM implemented in the software MCMCglmm. (A) Computation time (in hours) is plotted for datasets containing varying numbers of individuals, each with 10,000 genes (computation time in MACAU scales quadratically with sample size, but linearly with gene number). (B) Computation time (in hours) is plotted on a log scale. All computation was performed on a single core of an Intel Xeon E5-2683 2.00 GHz processor.



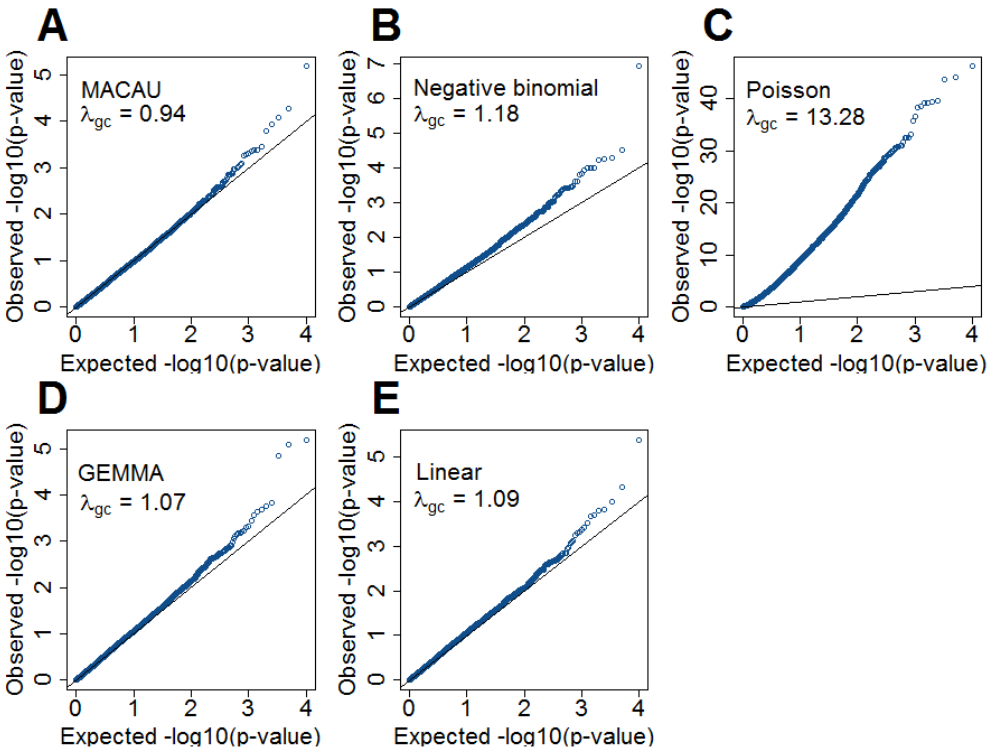
Supplementary Figure 2. MACAU *p*-values are consistent across two independent MCMC runs. QQ-plots comparing the *p*-value distributions for 2 independent runs of MACAU on the same data sets, with different simulated heritability values (A: h^2 = 0; B: h^2 = 0.3; C: h^2 = 0.6). Simulations were performed on m = 10,000 genes (1,000 DE and 9,000 non-DE) and n = 63 individuals, with CV = 0.3, PVE = 0.25, σ^2 = 0.25, and h_x^2 = 0.0.



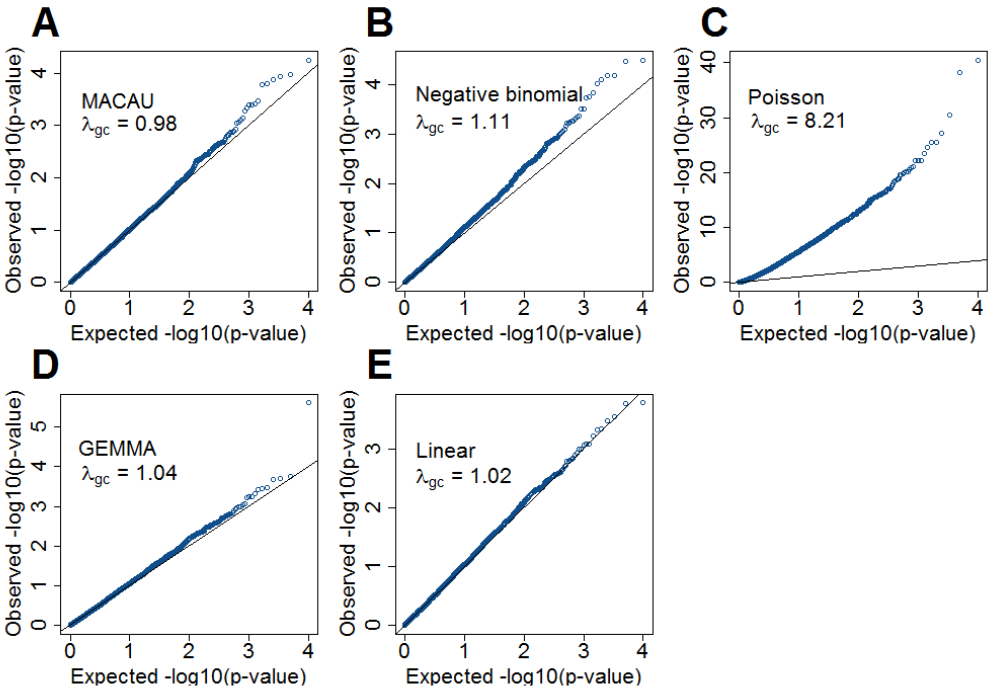
Supplementary Figure 3. QQ-plots compare observed versus expected *p*-value distributions after six data normalization procedures for the linear model (Linear; cyan) and the linear mixed model (GEMMA; blue) in the simulations. Results for MACAU (red) are also included for comparison. Simulations include 1000 DE genes and 9000 non-DE genes, generated with a default set of values ($h^2 = 0.3$, $h_x^2 = 0$, CV = 0.3, $\sigma^2 = 0.25$, PVE = 0.25, and n = 63). Normalization procedures were (A) trimmed mean of M-value (TMM), (B) Upper-quantile (UQ), (C) Relative log expression (RLE), (D) Total read count (TRC), (E) VOOM, and (F) Total read count and quantile (TRCQ). The number of DE genes identified by each method out of the known 1,000 DE genes is also shown on the figures.



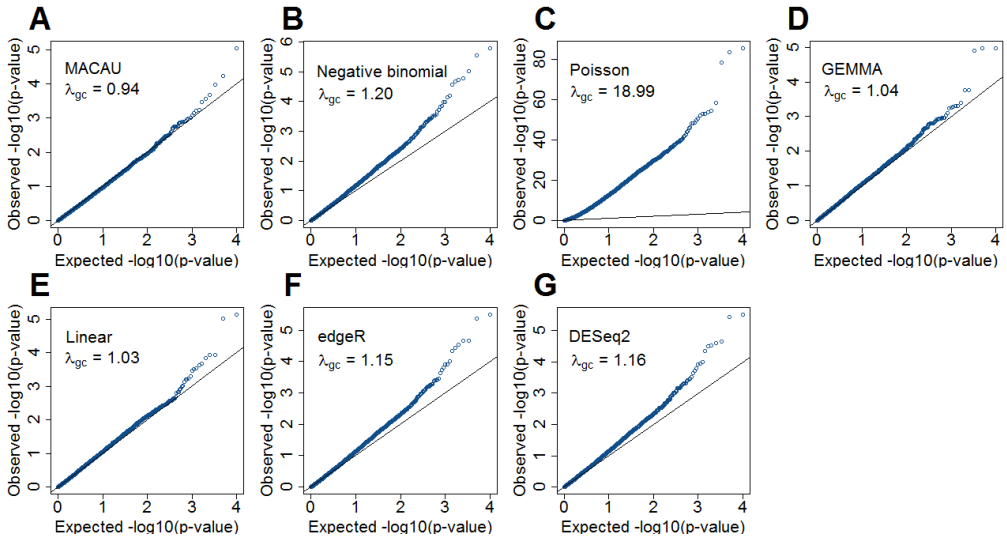
Supplementary Figure 4. QQ-plots comparing observed and expected *p*-value distributions by different methods for the null simulations in the presence of sample non-independence. 10,000 null genes were simulated with $n = 63$, CV = 0.3, $\sigma^2 = 0.25$, $h^2 = 0.3$ and $h_x^2 = 0.4$**.** Methods for comparison include MACAU (A), Negative binomial (B), Poisson (C), GEMMA (D), and Linear (E). Both MACAU and GEMMA properly control for type I error well in the presence of sample non-independence. λ_{gc} is the genomic control factor.



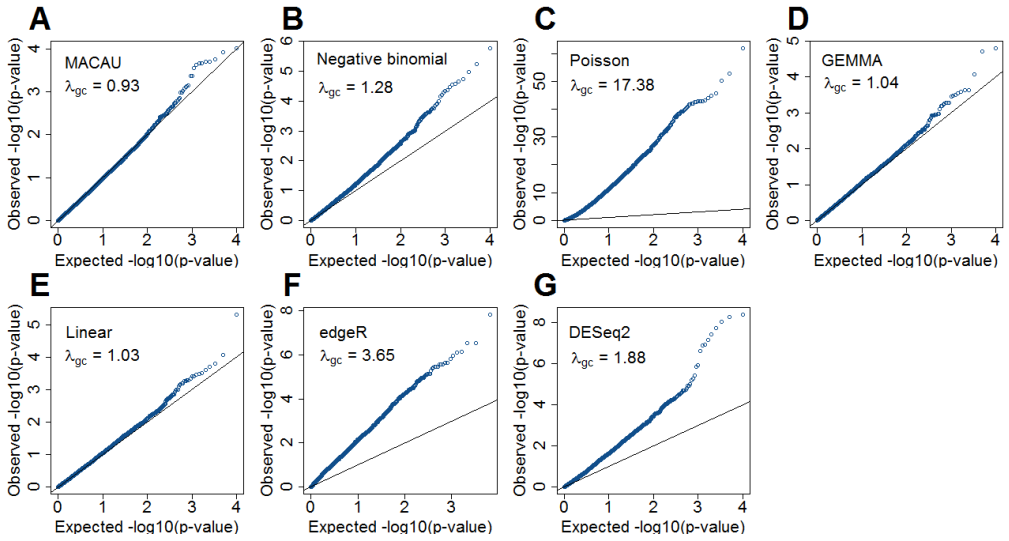
Supplementary Figure 5. QQ-plots comparing observed and expected *p*-value distributions by different methods for the null simulations in the presence of sample non-independence. 10,000 null genes were simulated with n = 63, CV = 0.3, σ^2 = 0.25, h^2 = 0 and h_x^2 = 0.4**.** Methods for comparison include MACAU (A), Negative binomial (B), Poisson (C), GEMMA (D), and Linear (E). λ_{gc} is the genomic control factor.



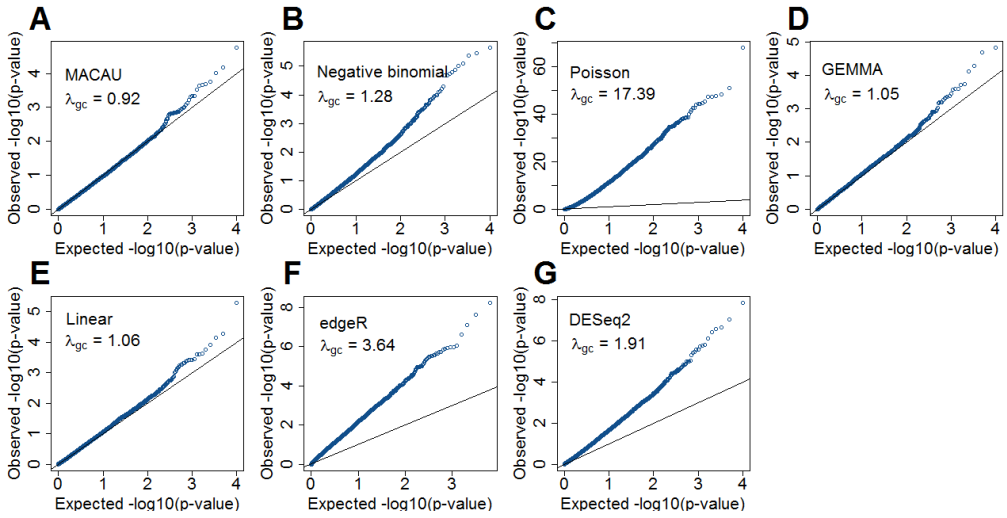
Supplementary Figure 6. QQ-plots comparing observed and expected *p*-value distributions by different methods for the null simulations in the presence of sample non-independence. 10,000 null genes were simulated with $n = 63$, CV = 0.3, $\sigma^2 = 0.25$, $h^2 = 0.6$ and $h_x^2 = 0.4$**.** Methods for comparison include MACAU (A), Negative binomial (B), Poisson (C), GEMMA (D), Linear (E), edgeR (F), and DESeq2 (G). λ_{gc} is the genomic control factor. The variable x is dichotomized.



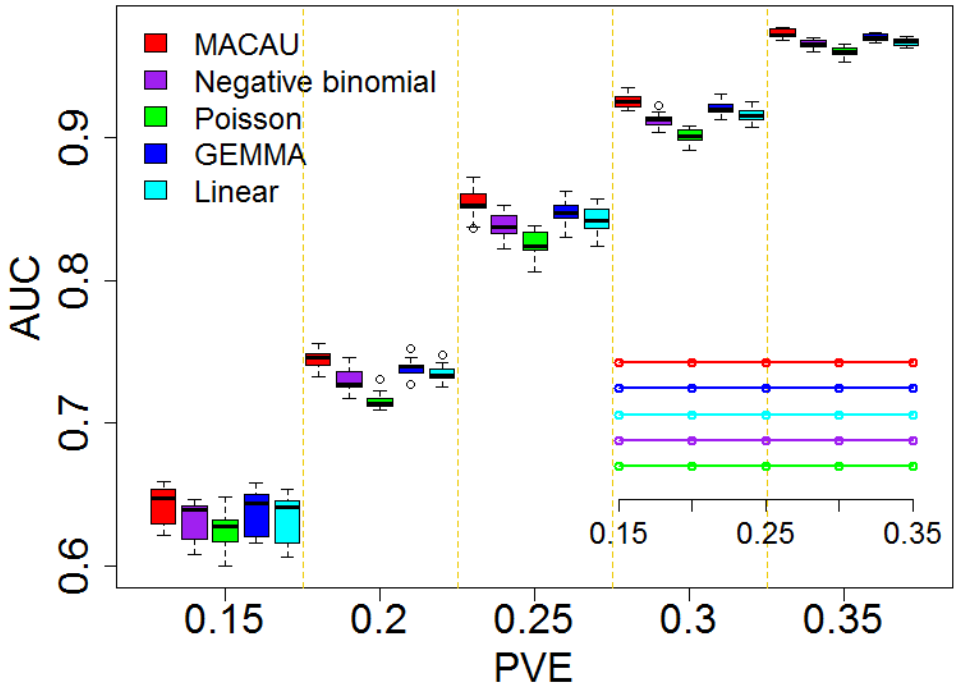
Supplementary Figure 7. QQ-plots comparing observed and expected *p*-value distributions by different methods for the null simulations in the presence of sample non-independence. Top 5 expression PCs are included in all models. 10,000 null genes were simulated with n = 63, CV = 0.3, σ^2 = 0.25, h^2 = 0.6 and h_x^2 = 0.4**.** Methods for comparison include MACAU (A), Negative binomial (B), Poisson (C), GEMMA (D), Linear (E), edgeR (F), and DESeq2 (G). λ_{gc} is the genomic control factor. The variable x is dichotomized.



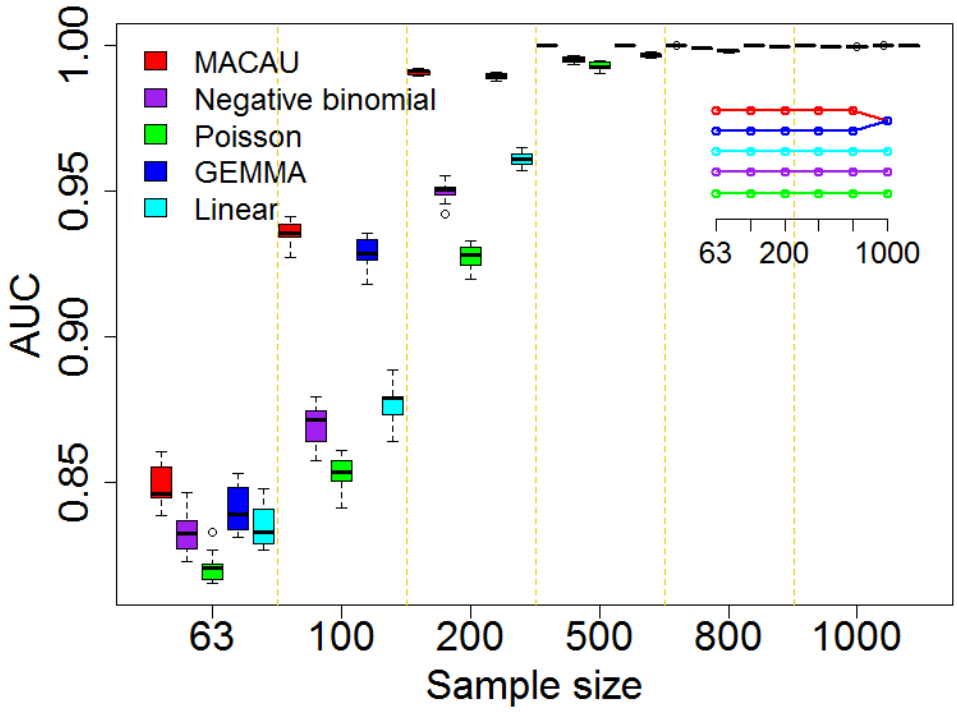
Supplementary Figure 8. QQ-plots comparing observed and expected *p*-value distributions by different methods for the null simulations in the presence of sample non-independence. Top 5 genotype PCs are included in all models. 10,000 null genes were simulated with n = 63, CV = 0.3, σ^2 = 0.25, h^2 = 0.6 and h_x^2 = 0.4**.** Methods for comparison include MACAU (A), Negative binomial (B), Poisson (C), GEMMA (D), Linear (E), edgeR (F), and DESeq2 (G). λ_{gc} is the genomic control factor. The variable x is dichotomized.



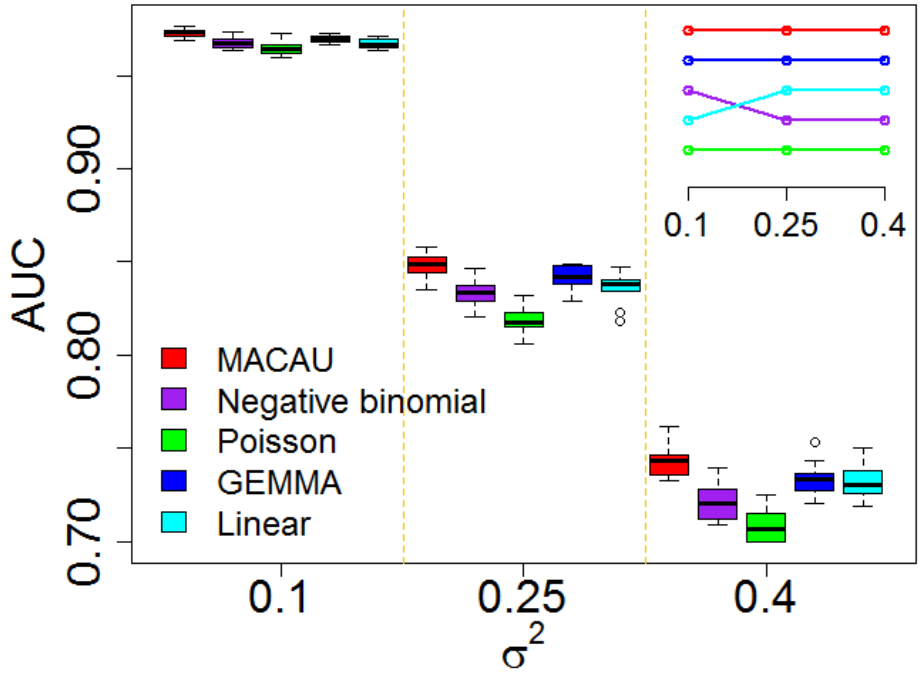
Supplementary Figure 9. Effect size of the predictor variable (PVE) affects power in simulated data sets. Area under the curve (AUC) is used to measure the performance of MACAU (red), Negative binomial (purple), Poisson (green), GEMMA (blue), and Linear (cyan). Each simulation setting consists of 10 simulation replicates, and each replicate includes 10,000 simulated genes, with 1,000 DE and 9,000 non-DE. We used n = 63, CV = 0.3, h_x^2 = 0.0, h^2 = 0.3 and σ^2 = 0.25. Boxplots of AUC across replicates for different methods show that increasing the effect size (PVE) increases power of all methods but does not affect their relative rank. Inset shows the rank of different methods, where the top row represents the highest rank.



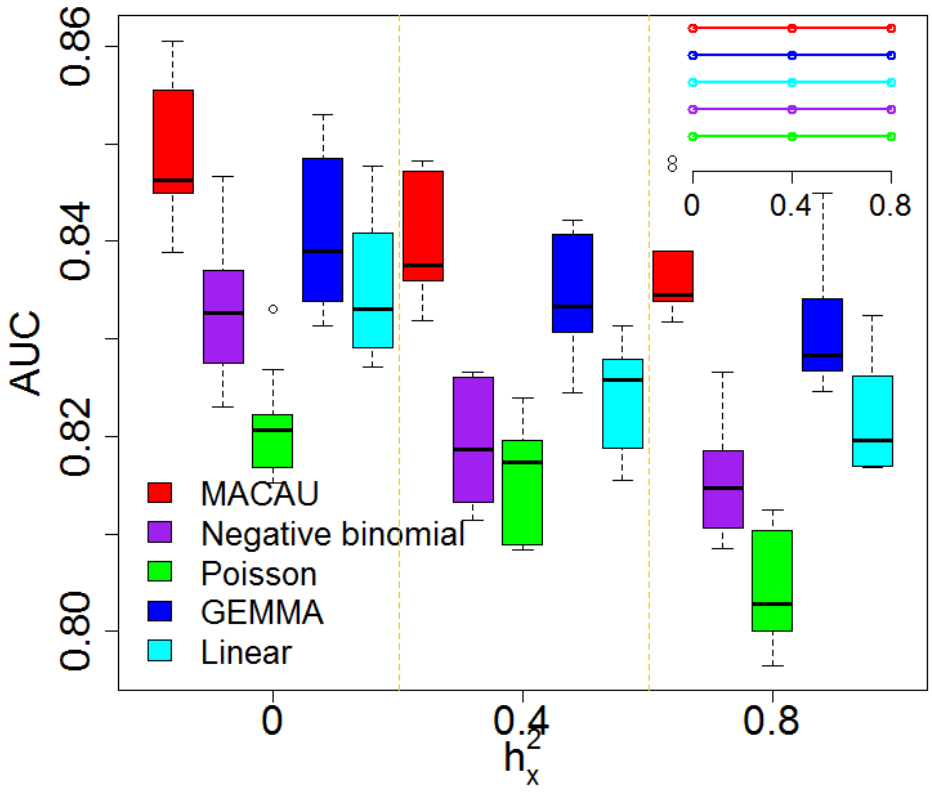
Supplementary Figure 10. Sample size effects the power in simulations. Area under the curve (AUC) is used to measure the performance of MACAU (red), Negative binomial (purple), Poisson (green), GEMMA (blue), and Linear (cyan). Each simulation setting consists of 10 simulation replicates, and each replicate includes 10,000 simulated genes, with 1,000 DE and 9,000 non-DE. We used CV = 0.3, h_x^2 = 0.0, PVE = 0.25, h^2 = 0.3 and σ^2 = 0.25.



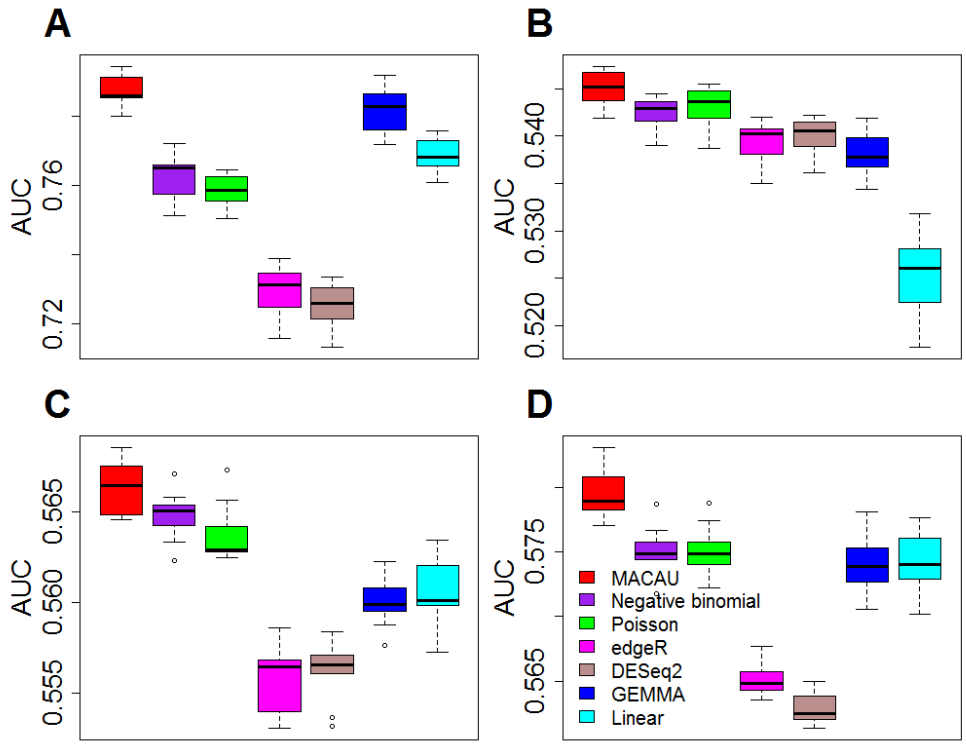
Supplementary Figure 11. Over-dispersion variance (σ^2) affects power in simulations. Area under the curve (AUC) is used to measure the performance of MACAU (red), Negative binomial (purple), Poisson (green), GEMMA (blue), and Linear (cyan). Each simulation setting consists of 10 simulation replicates, and each replicate includes 10,000 simulated genes, with 1,000 DE and 9,000 non-DE. We used n = 63, CV = 0.3, h_x^2 = 0.0, PVE = 0.25 and h^2 = 0.3. Boxplots of AUC across replicates for different methods show that increasing the over-dispersion variance (σ^2) decreases power of all methods. Inset shows the rank of different methods, where the top row represents the highest rank.



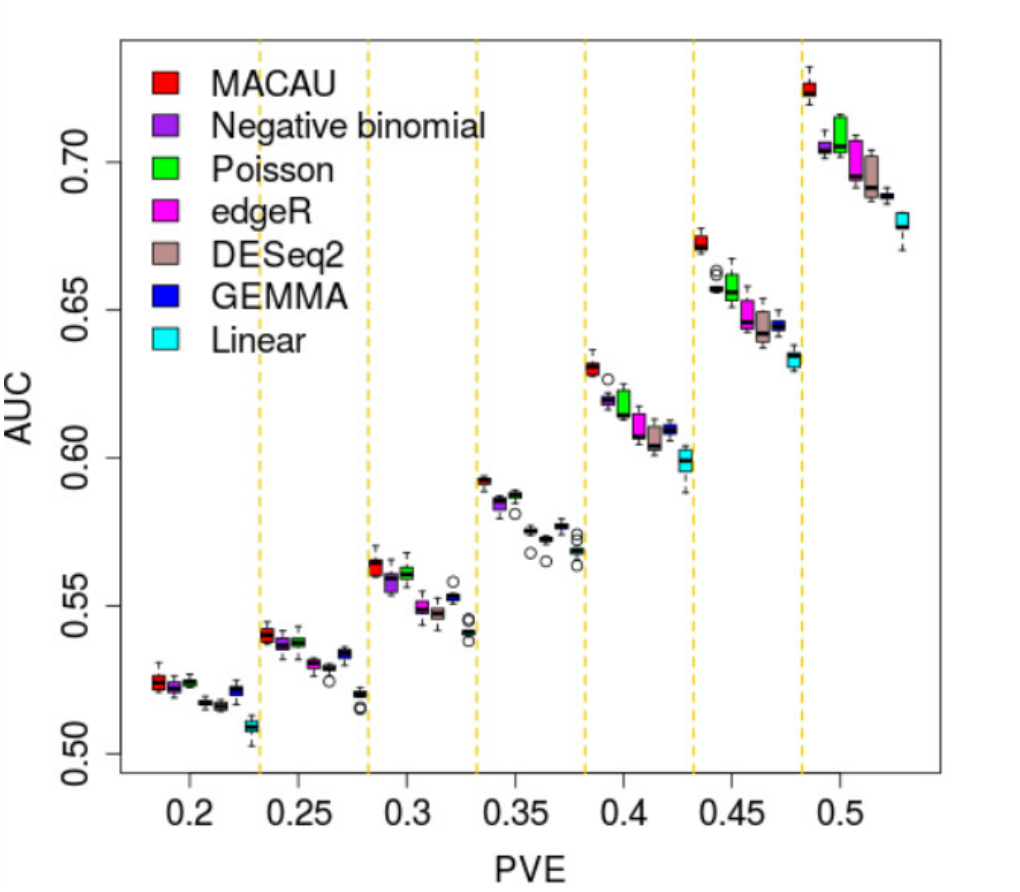
Supplementary Figure 12. Heritability of the predictor variable (h_x^2) affects power in simulations. Area under the curve (AUC) is used to measure the performance of MACAU (red), Negative binomial (purple), Poisson (green), GEMMA (blue), and Linear (cyan). Each simulation setting consists of 10 simulation replicates, and each replicate includes 10,000 simulated genes, with 1,000 DE and 9,000 non-DE. We used n = 63, CV = 0.3, PVE = 0.25, h^2 = 0.3 and σ^2 = 0.25. Boxplots of AUC across replicates for different methods show that increasing heritability of the predictor variable (h_x^2) reduces power of all methods but does not affect their relative rank. Inset shows the rank of different methods, where the top row represents the highest rank.



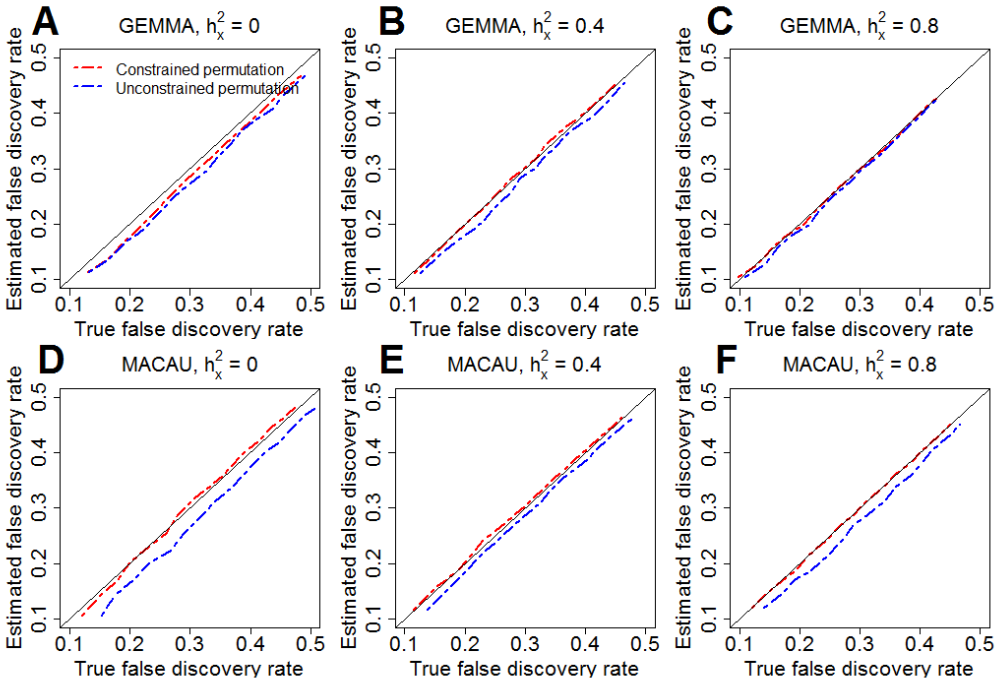
Supplementary Figure 13. Power comparison of different methods when the predictor variable is dichotomized in simulations. Area under the curve (AUC) is used to measure the performance of MACAU (red), Negative binomial (purple), Poisson (green), edgeR (magenta), DESeq2 (rosybrown), GEMMA (blue), and Linear (cyan). Each simulation setting consists of 10 simulation replicates, and each replicate includes 10,000 simulated genes, with 1,000 DE and 9,000 non-DE. (A) Sample size n = 63; (B) Sample size n = 6; (C) Sample size n = 10; (D) Sample size n = 14. For other parameters, we used CV = 0.3, h_x^2 = 0.0, PVE = 0.25, h^2 = 0.3 and σ^2 = 0.25.



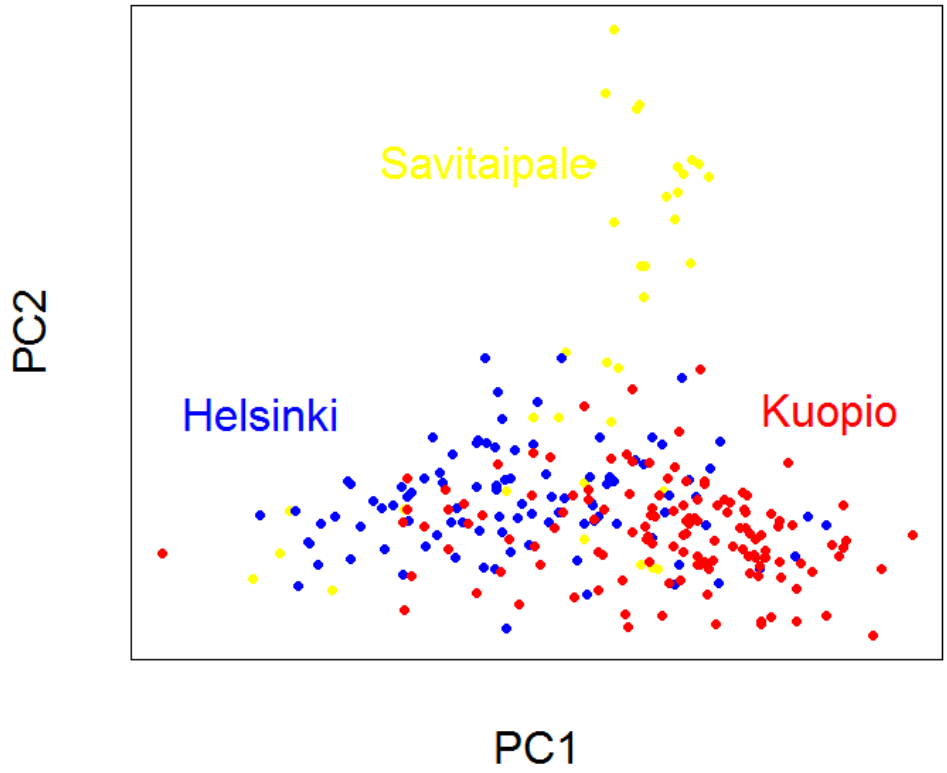
Supplementary Figure 14. Power comparison of different methods in small samples with increasing effect sizes in simulations. The predictor variable is dichotomized to allow comparisons with edgeR and DESeq2. Area under the curve (AUC) is used to measure the performance of MACAU (red), Negative binomial (purple), Poisson (green), edgeR (magenta), DESeq2 (rosybrown), GEMMA (blue), and Linear (cyan). Each simulation setting consists of 10 simulation replicates, and each replicate includes 10,000 simulated genes, with 1,000 DE and 9,000 non-DE. For other parameters, we used n = 6, CV = 0.3, h_x^2 = 0.0, h^2 = 0.3 and σ^2 = 0.25.



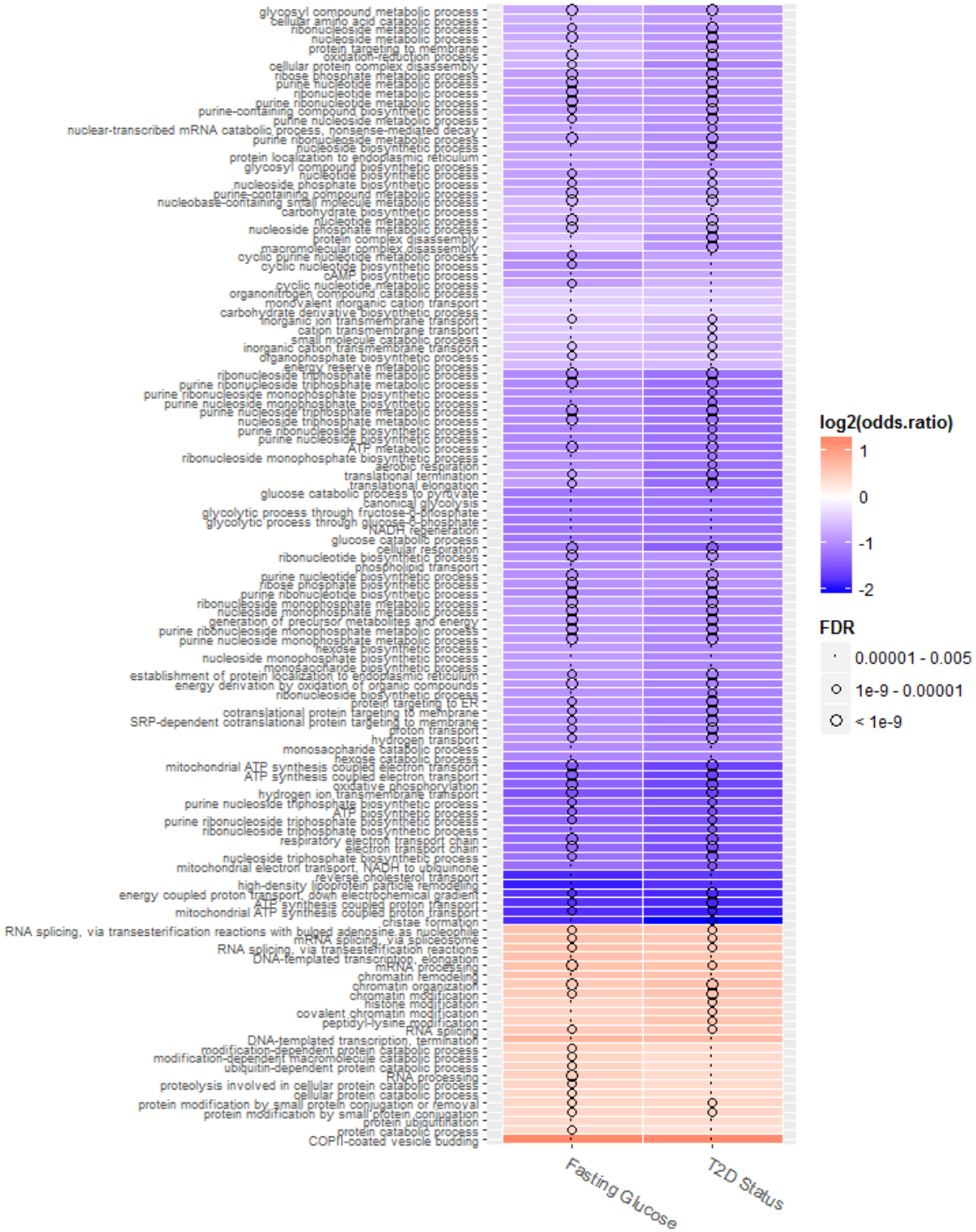
Supplementary Figure 15. The constrained permutation procedure (red) estimates false discovery rate (FDR) more accurately than the unconstrained permutation procedure (blue) in simulations. Panels A, B, and C show the results of GEMMA for $h_x^2 = 0$, $h_x^2 = 0.4$, and $h_x^2 = 0.8$, respectively. Panels D, E, and F show the results of MACAU for $h_x^2 = 0$, $h_x^2 = 0.4$, and $h_x^2 = 0.8$, respectively. Simulations were performed under the null with n = 63, CV = 0.3, h^2 = 0.3 and σ^2 = 0.25.



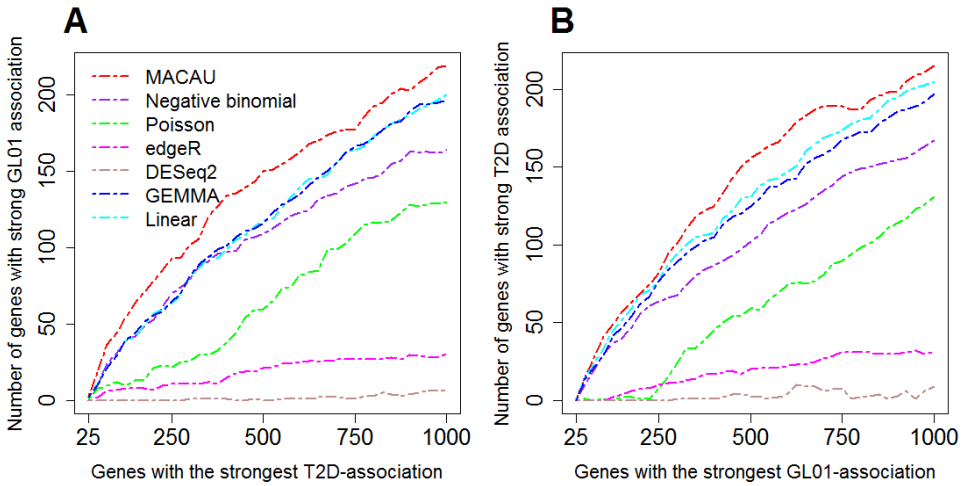
Supplementary Figure 16. Top two genotype principal components (PCs) display the population structure of the FUSION data. Principal component one (PC1) and principal component two (PC2) are generated from the genetic relatedness matrix computed based on genotypes. PC1 explains 0.55% proportion of variance while PC2 explains 0.52% proportion of variance. Individuals are colored according to their origin of municipality: Helsinki (blue), Savitaipale (yellow), and Kuopio (red).



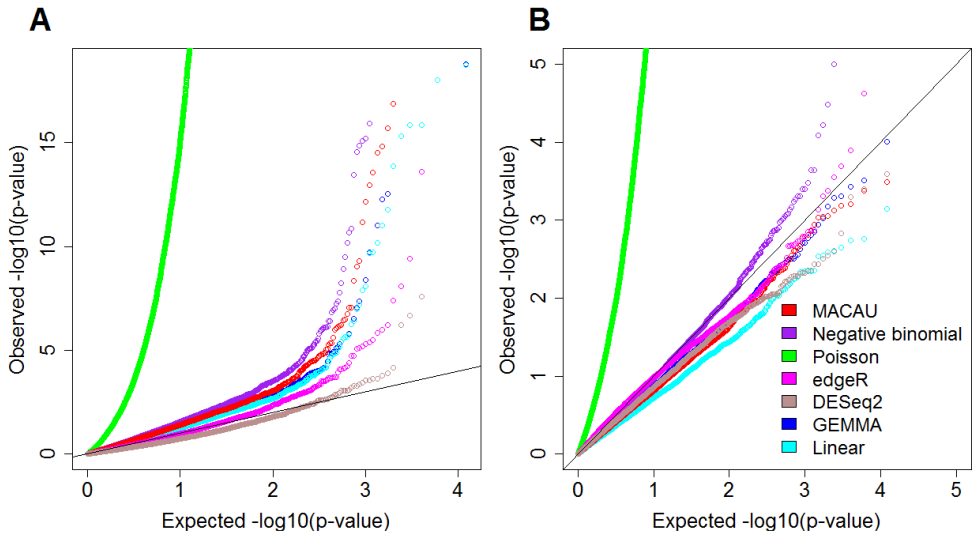
Supplementary Figure 17. Heatmap of GO terms for GL-associated or T2D-associated DE genes. The GO terms are hierarchically clustered. Red represents higher positive gene expression – trait association; Blue represents stronger negative associations. The larger circle sizes represent more significant GO terms.



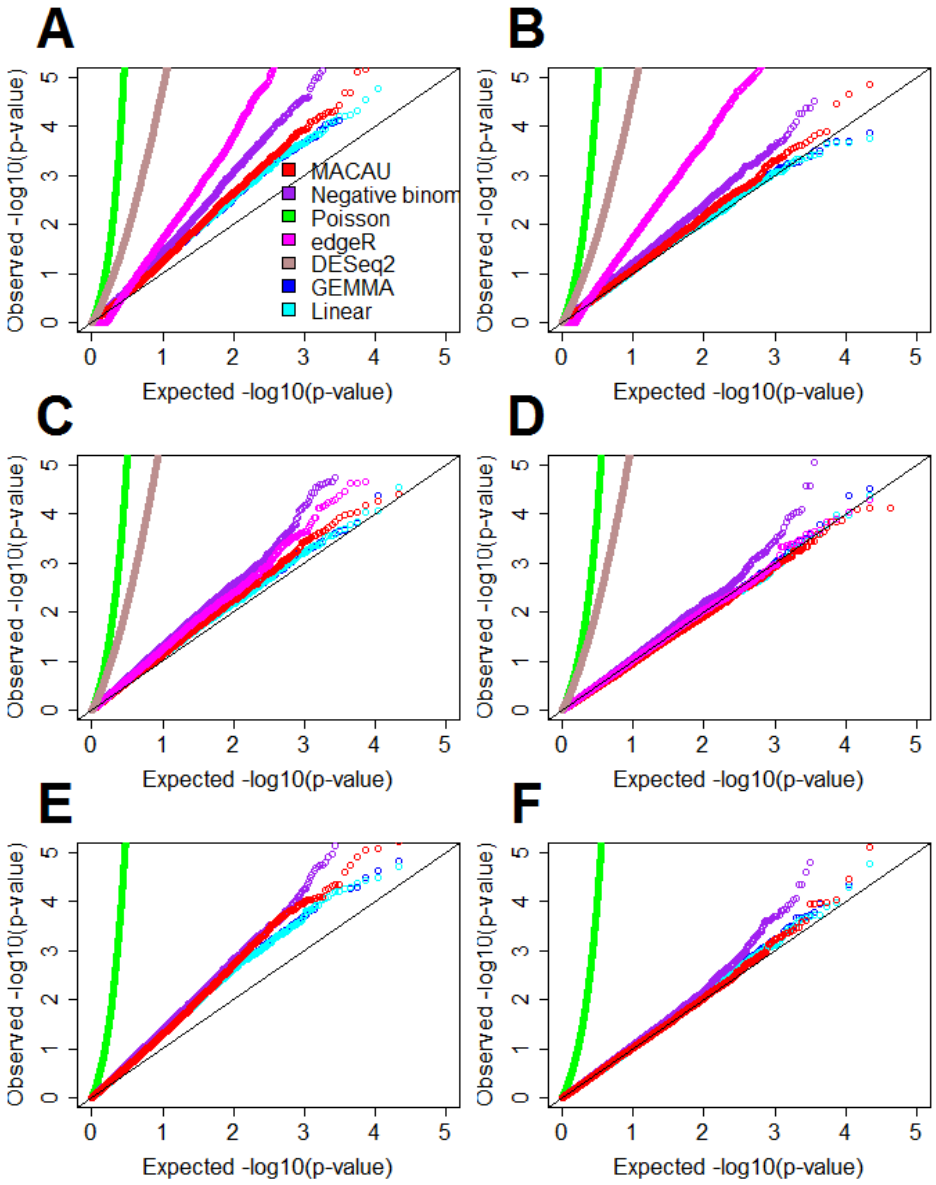
Supplementary Figure 18. Overlap between T2D associated and GL01 associated genes. Methods for comparison include MACAU (red), Negative binomial (purple), Poisson (green), edgeR (magenta), DESeq2 (rosybrown), GEMMA (blue), and Linear (cyan). (A) shows the number of genes that are in the list of top 1,000 genes most significantly associated with GL01 out of the genes that have the strongest association for T2D for each method. For instance, in the top 1,000 genes with the strongest T2D association identified by MACAU, 219 of them are also in the list of top 1,000 genes with the strongest GL01 association identified by the same method. (B) shows the number of genes that are in the list of top 1,000 genes most significantly associated with T2D out of the genes that have the strongest association for GL01 for each method. T2D: type II diabetes; GL01: dichotomized fasting glucose level.



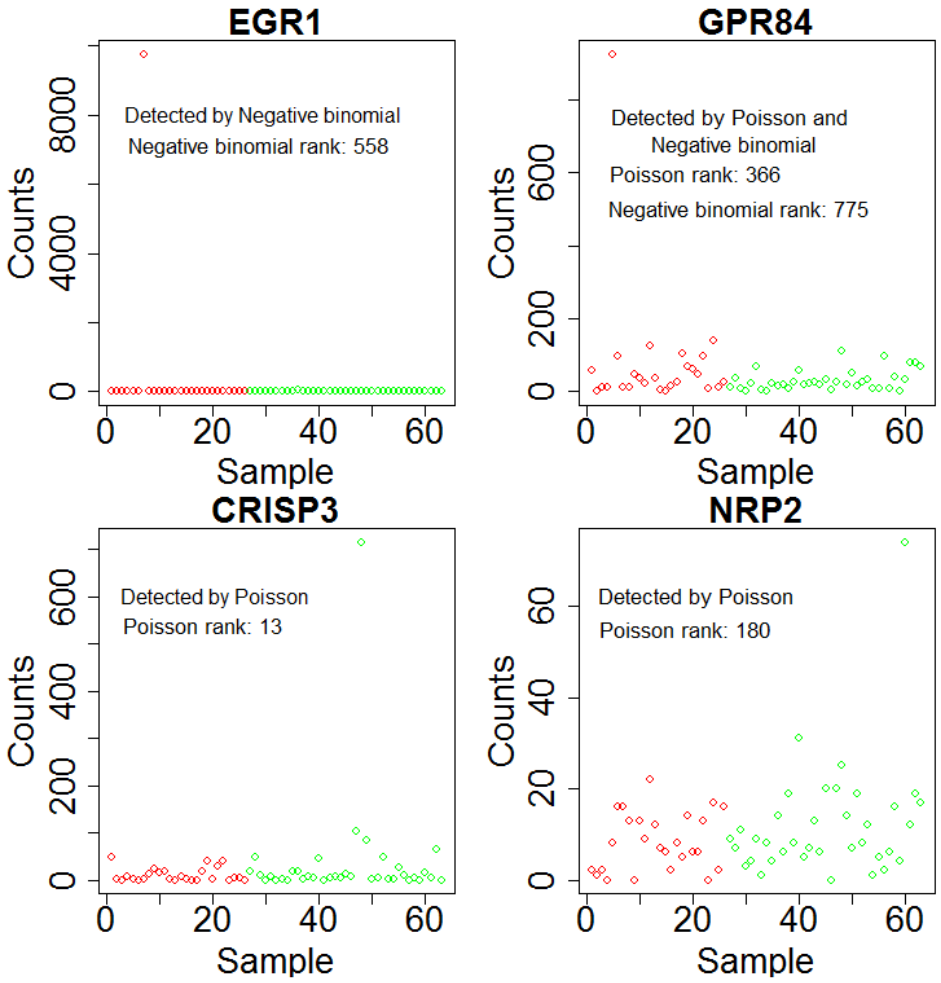
Supplementary Figure 19. QQ-plots for comparing observed versus expected *p*-value distributions from different methods in the baboon data. Methods for comparison include MACAU (red), Negative binomial (purple), Poisson (green), edgeR (magenta), DESeq2 (rosybrown), GEMMA (blue), and Linear (cyan). (A) QQ-plots for identifying sex-associated genes in the real data. (B) QQ-plots for identifying sex-associated genes in the permuted null data. Only *p*-values from MACAU and GEMMA are close to uniformly distributed in the permuted null data.



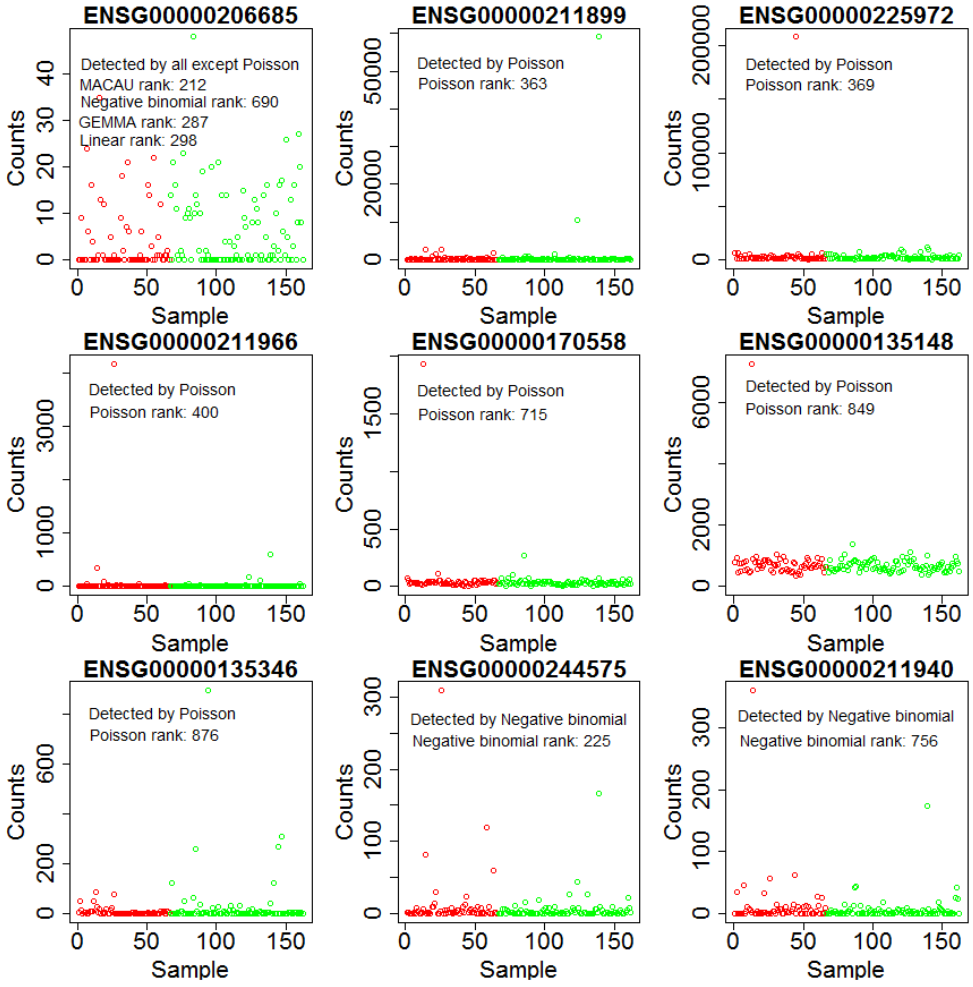
Supplementary Figure 20. QQ-plots for comparing observed versus expected *p*-value distributions by different methods in the FUSION data. Methods for comparison include MACAU (red), Negative binomial (purple), Poisson (green), edgeR (magenta), DESeq2 (rosybrown), GEMMA (blue), and Linear (cyan). Left: QQ-plots for identifying (A) T2D-associated genes; (C) GL-associated genes after dichotomizing the predictor variable; (E) GL-associated genes of original variable in the real data. Right: QQ-plots for identifying (B) T2D-associated genes; (D) GL-associated genes after dichotomizing the predictor variable in the permuted null data; (F) GL-associated genes of original variable in the permuted null data. The *p*-values from MACAU, GEMMA, and the linear model are close to uniformly distributed in the permuted null data.



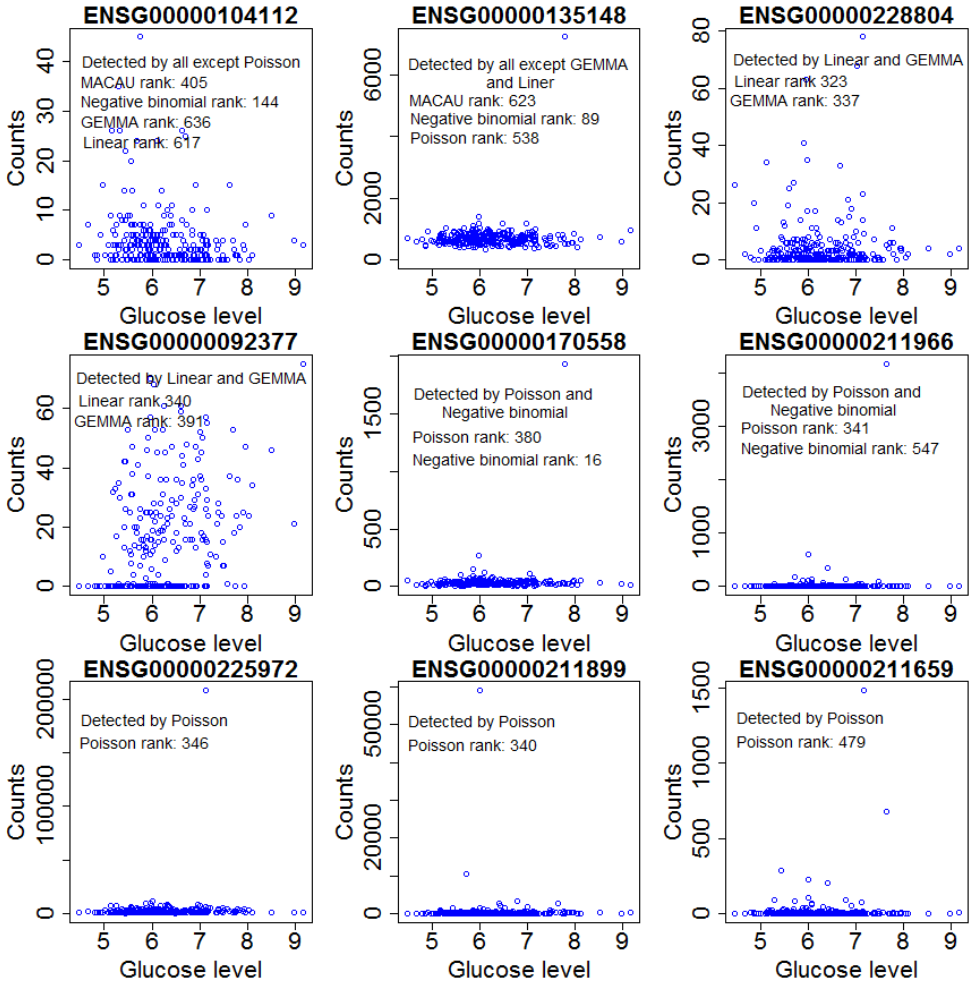
Supplementary Figure 21. Only four genes with potential outliers are in the 1,000 genes with the strongest sex association identified by various methods in the baboon data. For each of the four genes, raw read counts are plotted against samples for females (red) and males (green). These genes are detected by either negative binomial or Poisson model as showing strong sex association, with association rank displayed in the corresponding panels.

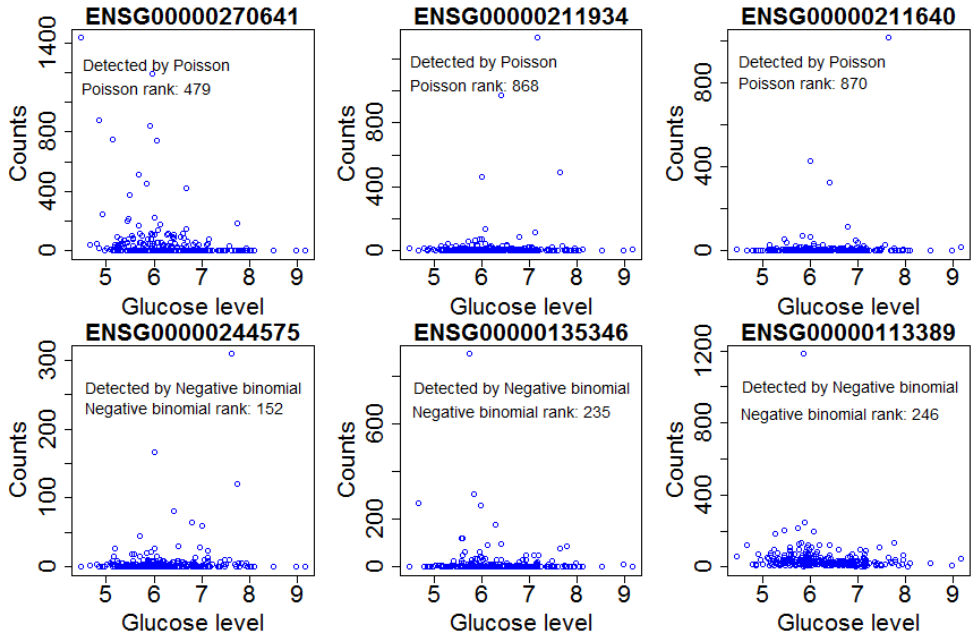


Supplementary Figure 22. Only nine genes with potential outliers are in the 1,000 genes with the strongest T2D association identified by various methods in the FUSION data. For each of the nine genes, read counts are plotted against samples for T2D cases (red) and NGT controls (green). These genes are detected by different methods as showing strong T2D association, with association rank displayed in the corresponding panels.

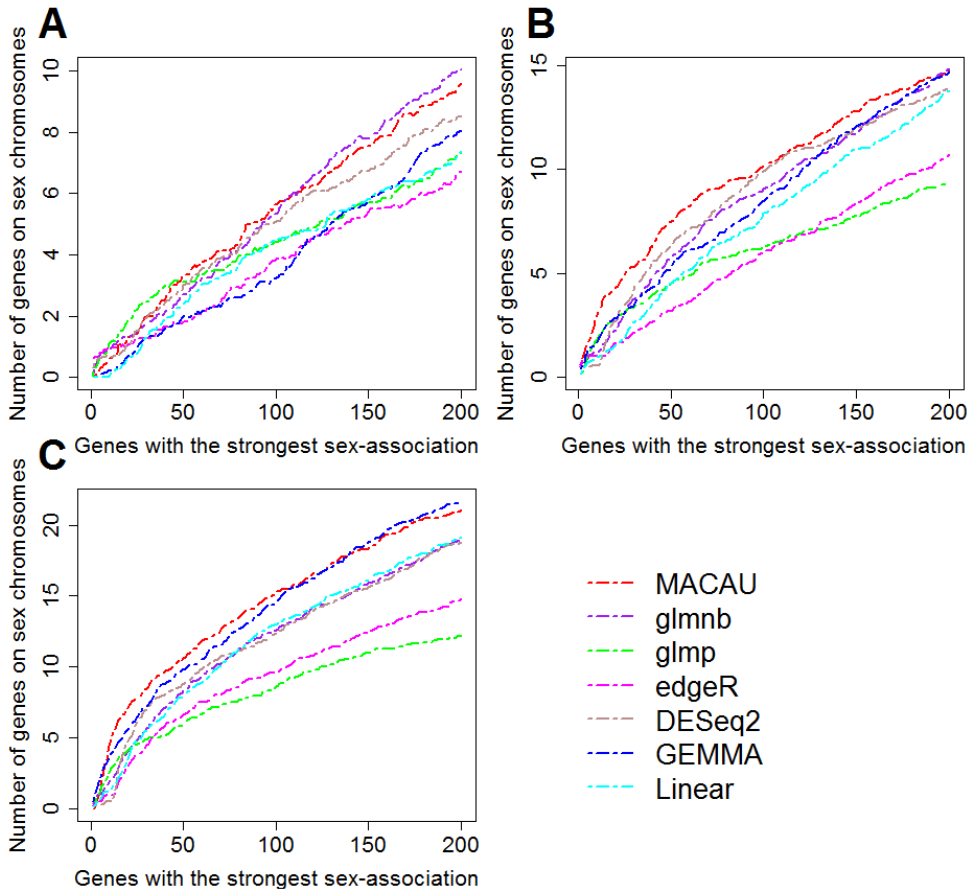


Supplementary Figure 23. Only 15 genes with potential outliers are in the 1,000 genes with the strongest GL association identified by various methods in the FUSION data. For each of the 15 genes, read counts are plotted against glucose levels. These genes are detected by different methods as showing strong GL association, with association rank displayed in the corresponding panels.




ENSG00000270641
Detected by Poisson
Poisson rank: 479
ENSG00000211934
Detected by Poisson
Poisson rank: 868
ENSG00000211640
Detected by Poisson
Poisson rank: 870
ENSG00000244575
Detected by Negative binomial
Negative binomial rank: 152
ENSG00000135346
Detected by Negative binomial
Negative binomial rank: 235
ENSG00000113389
Detected by Negative binomial
Negative binomial rank: 246
Counts
Glucose level

Supplementary Figure 24. Enrichment of genes on sex chromosomes in small data sets. These small data sets are created via subsampling from the YRI data. For each sample size (n = 6, n = 10 or n = 14), we performed 20 replicates. In each replicate, we subsampled an equal number of males and females from the YRI data and applied different methods to identify sex-associated genes. Methods for comparison include MACAU (red), Negative binomial (purple), Poisson (green), edgeR (magenta), DESeq2 (rosybrown), GEMMA (blue), and Linear (cyan). The plots show the average number of genes across the 20 replicates that are on the X chromosome out of the genes that have the strongest sex association for each method (note that the Y chromosome is not used during read alignment and is thus ignored), for (A) n = 6 (3 males vs 3 females); (B) n = 10 (5 males vs 5 females); and (C) n = 14 (7 males vs 7 females). For instance, when n = 6, in the top 50 genes identified by MACAU, an average of 3.3 of them are also on the X chromosome.



Supplementary Table 1. Computation times for each method on two real data sets. All computation was performed on a single core of an Intel Xeon E5-2683 2.00 GHz processor. *n* = number of individuals; *m* = number of genes.

Statistical method	Software	Computation time	
		Baboon (*n* = 63, *m* = 12,018)	*Fusion* (*n* = 267, *m* = 21,753)
Linear model	R (lm)	0.03 min	0.24 min
Linear mixed model	GEMMA	0.11 min	1.02 min
Poisson model	R (glm)	1.2 min	7.47 min
Negative binomial model	R (glm.nb)	2.1 min	12.87 min
	R(edgeR)	1.2 min	9.6 min
	R(DESeq2)	3.1 min	54.6 min
Poisson mixed model	MACAU	1.32 h	19 h
Poisson mixed model	MCMCglmm	13.02 h	9.74 d