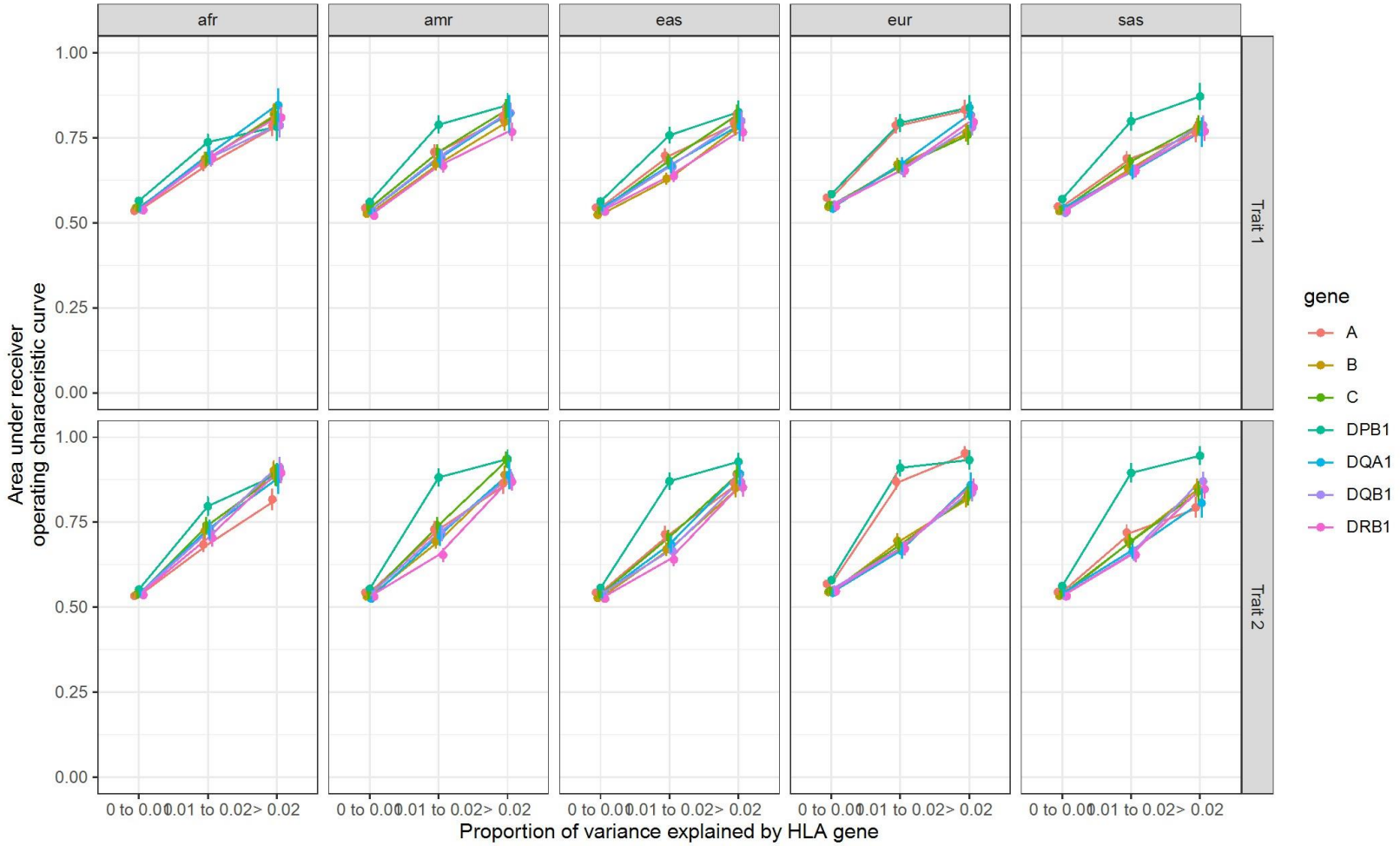
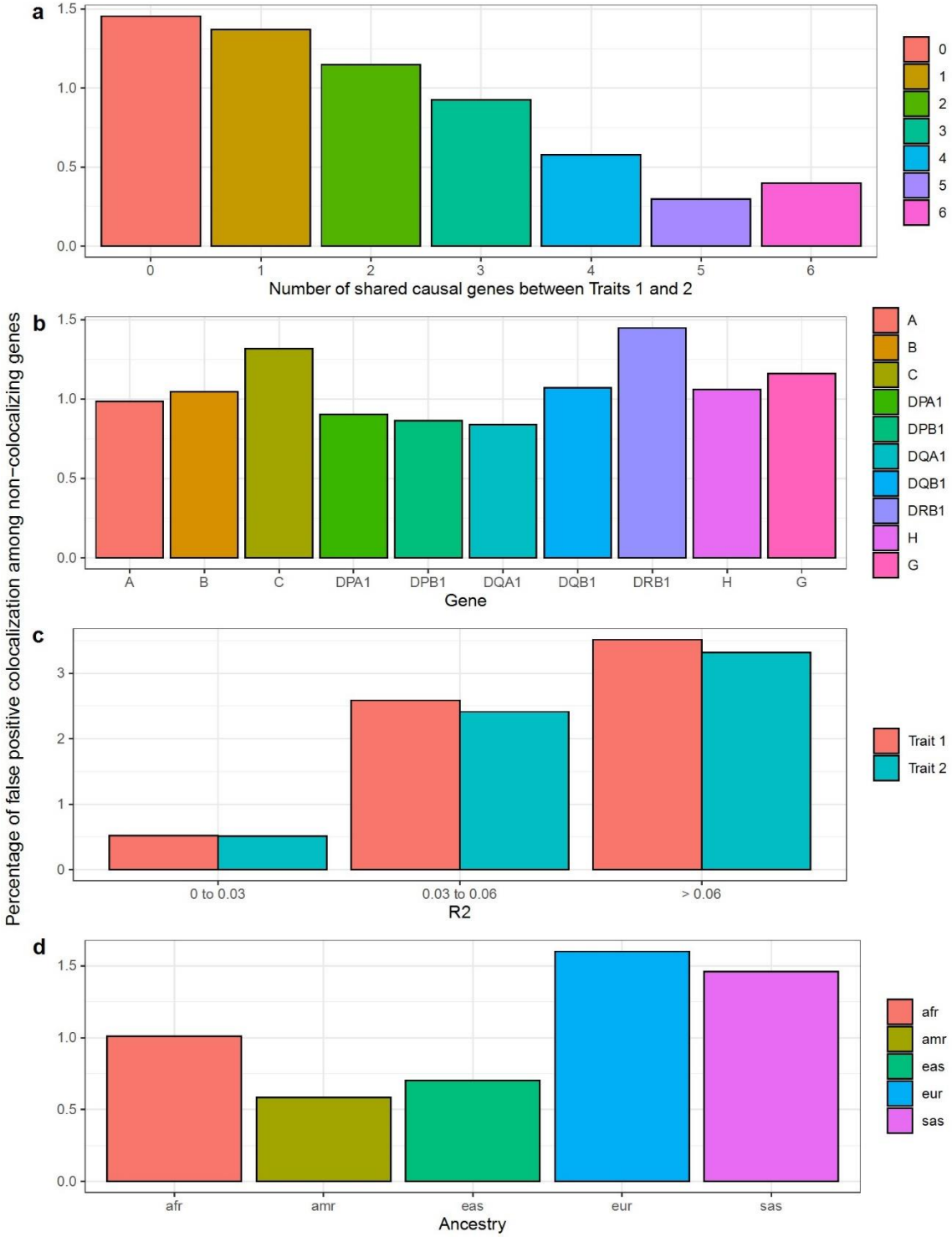
Supplementary Figure 1: Per ancestry ROC area under the curves for simulations of quantitative traits



Area under the ROC curves of HLA-colocalization PIPs for different variance explained per gene and genetic ancestries for the simulation of the quantitative traits. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry.

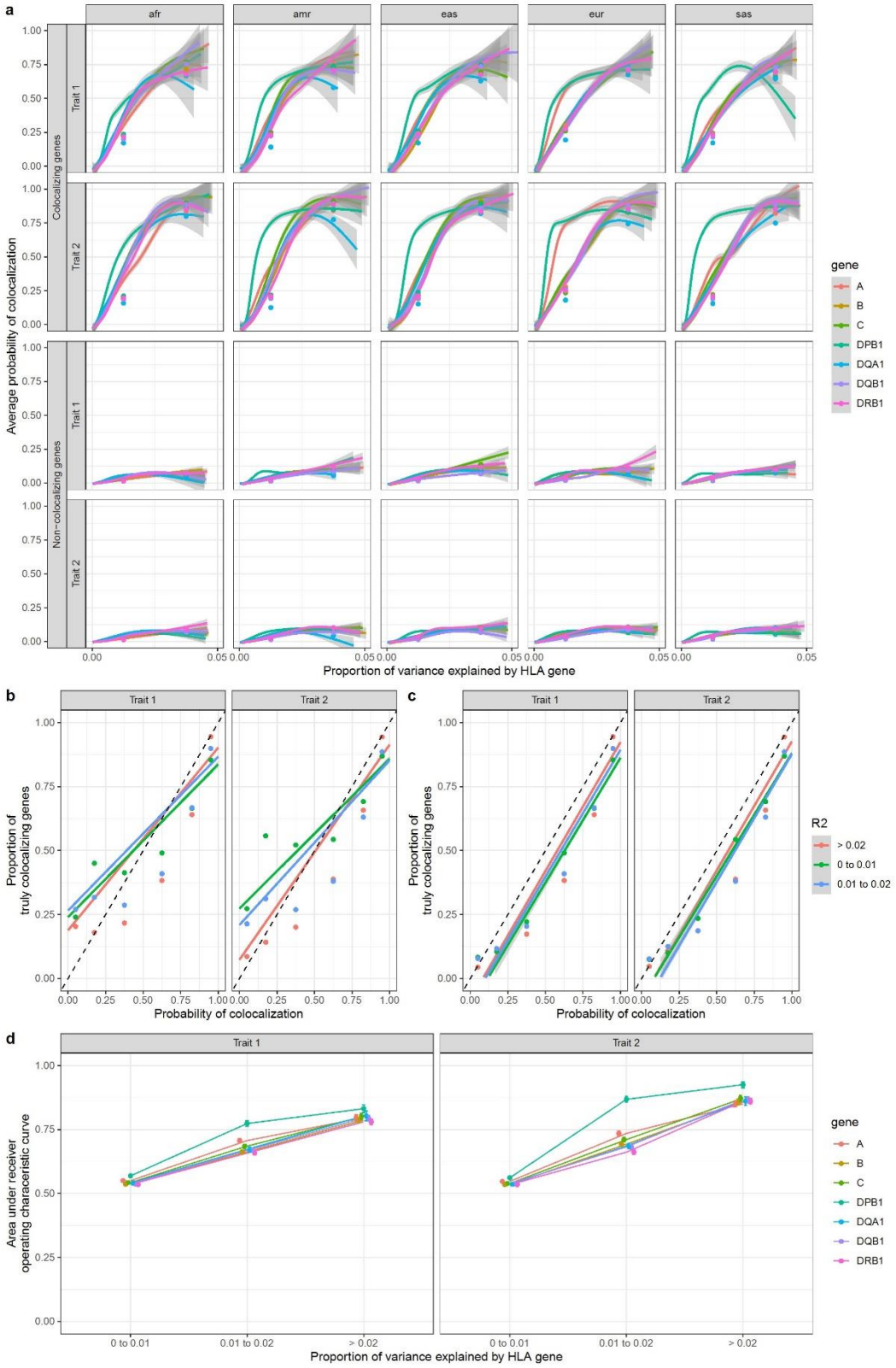
Supplementary Figure 2: False colocalizing genes.



We calculated the proportion of non-colocalizing genes with a probability of colocalization above 80% under varying circumstances. **a)** The proportion was higher in simulations where less genes were truly colocalizing then in simulations with more, since more truly colocalizing genes decreases the chance of a random false colocalization. **b)** The proportion was higher at *HLA-DRB1* (1.45%) and *HLA-C* (1.32%) than other genes (e.g. *HLA-DQA1* 0.84%). **c)** The proportion was higher

in gene which explained a larger portion of the traits, since genes with lower R2 are less likely to colocalize. **d)** The proportion was slightly higher in Europeans (1.6%) compared to other ancestries (e.g. admixed Americans at 0.58%). Overall, only 1.1% of non-colocalizing genes were given a probability of colocalization above 80%.

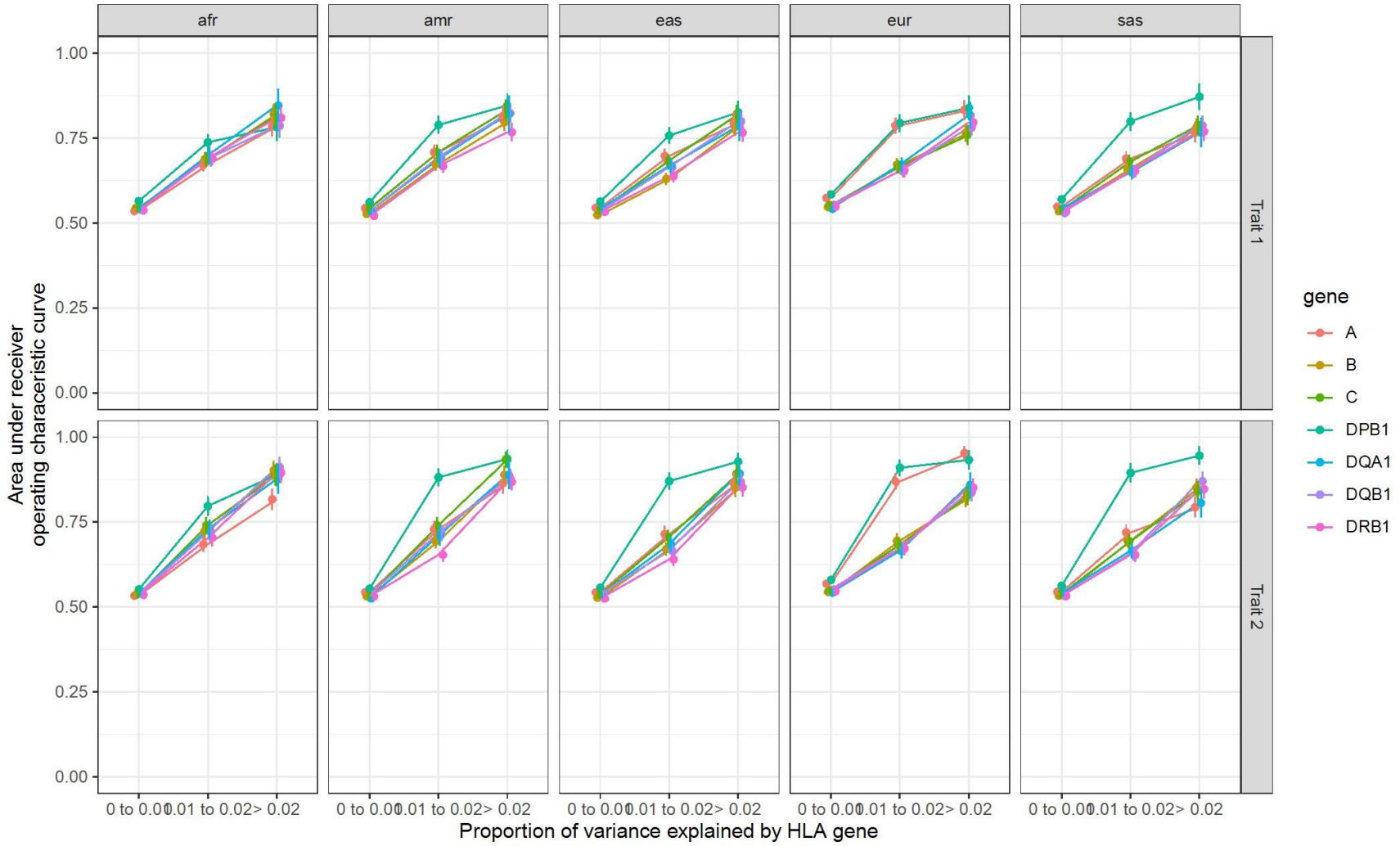
Supplementary Figure 3: HLA allele HLA-colocalization simulation results for binary traits



Pairs of quantitative traits were simulated having either true overlap, or no true overlap between causal HLA alleles, using a bivariate normal model as described in Methods. In each simulation a

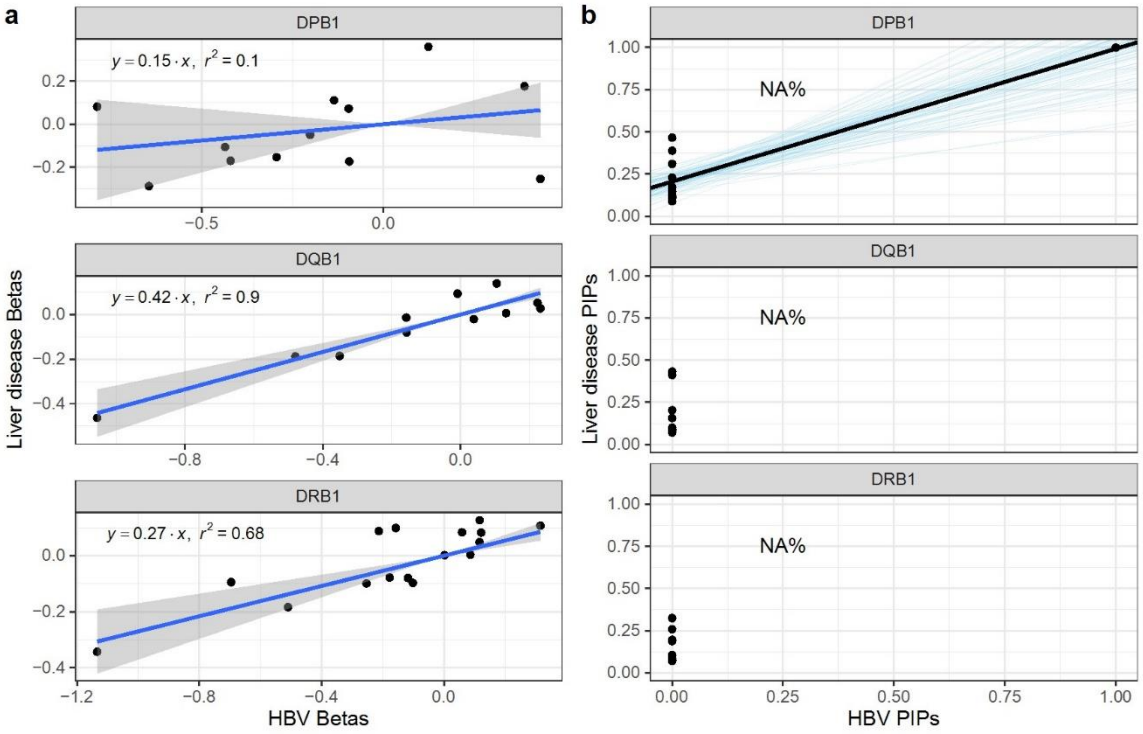
total proportion of trait variance explained was assumed. A total of 50,000 simulations (10,000 per ancestry group) were performed covering different parameter values (**Methods**). HLA allele distributions were simulated using UK Biobank participants. **a)** The average posterior probability of colocalization in truly colocalizing increases with the amount of phenotype variance explained by each gene, as expected. The average posterior probability of colocalization in truly non-colocalizing genes remains stable with increasing variance explained. The lines were drawn using a generalized additive model with *geom_smooth* in R. The grey area represents 95% confidence intervals. The individual dots represent the average in the corresponding variance bins. **b)** The proportion of simulated genes that were truly colocalizing shown as a function of the probability of colocalization. This is close to the identity line, though errs on the more conservative side for genes with lower R2. **c)** The deviation from the identity line is largely due to situations where SuSiE is unable to assign a PIP larger than 50% in at least one allele at a gene. When we restrict to genes with a minimal PIP of 50%, the method is almost perfectly calibrated. **d)** Average area under the curve as a function of variance explained for each gene. For this plot, average ROC area under the curve across ancestry was shown. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry.

Supplementary Figure 4: Per ancestry ROC area under the curves for simulations of binary traits



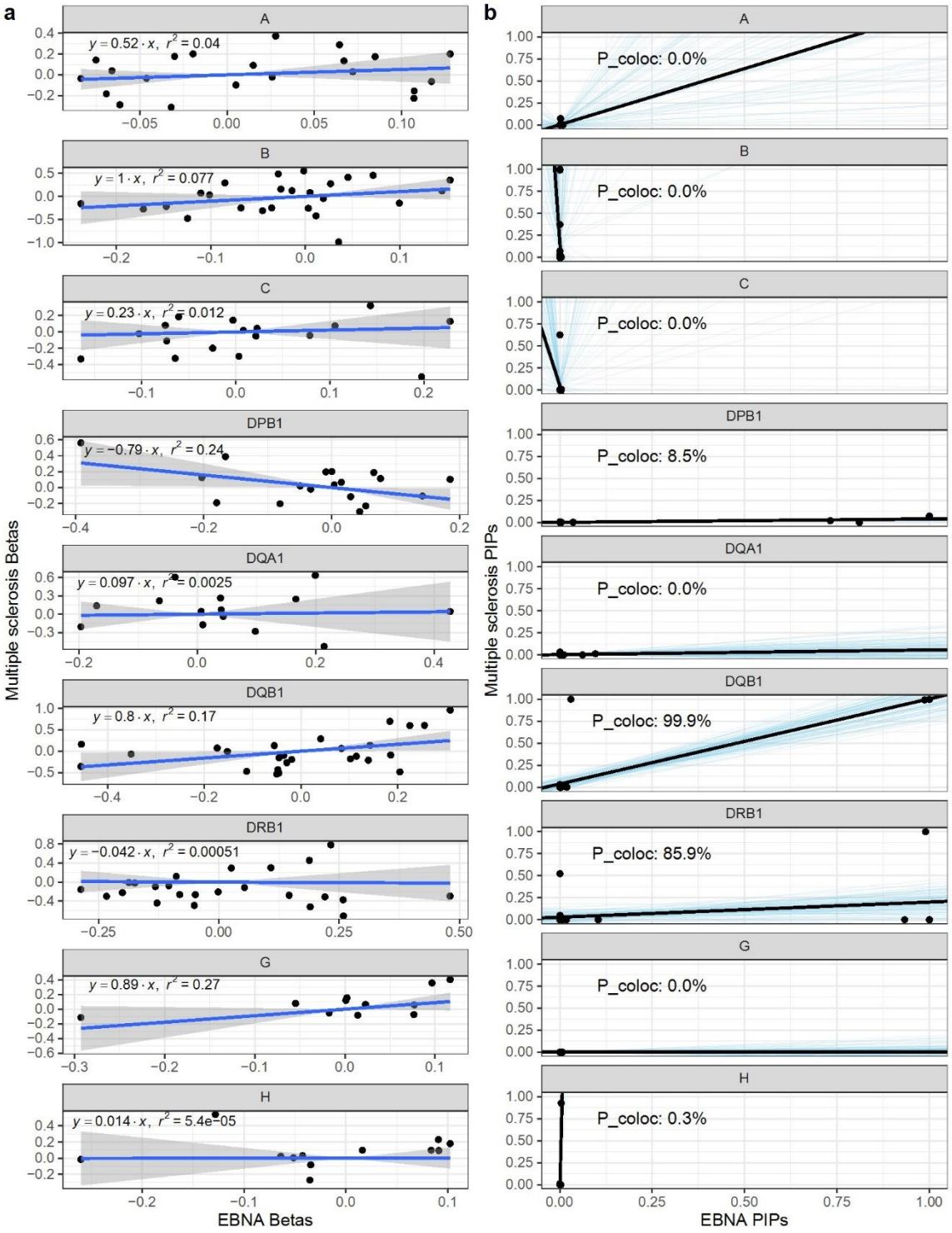
Area under the ROC curves of HLA-colocalization PIPs for different variance explained per gene and genetic ancestries for the simulation of the binary traits. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry.

Supplementary Figure 5: Hepatitis B (HBV) and liver disease HLA-colocalization in the Taiwan Biobank



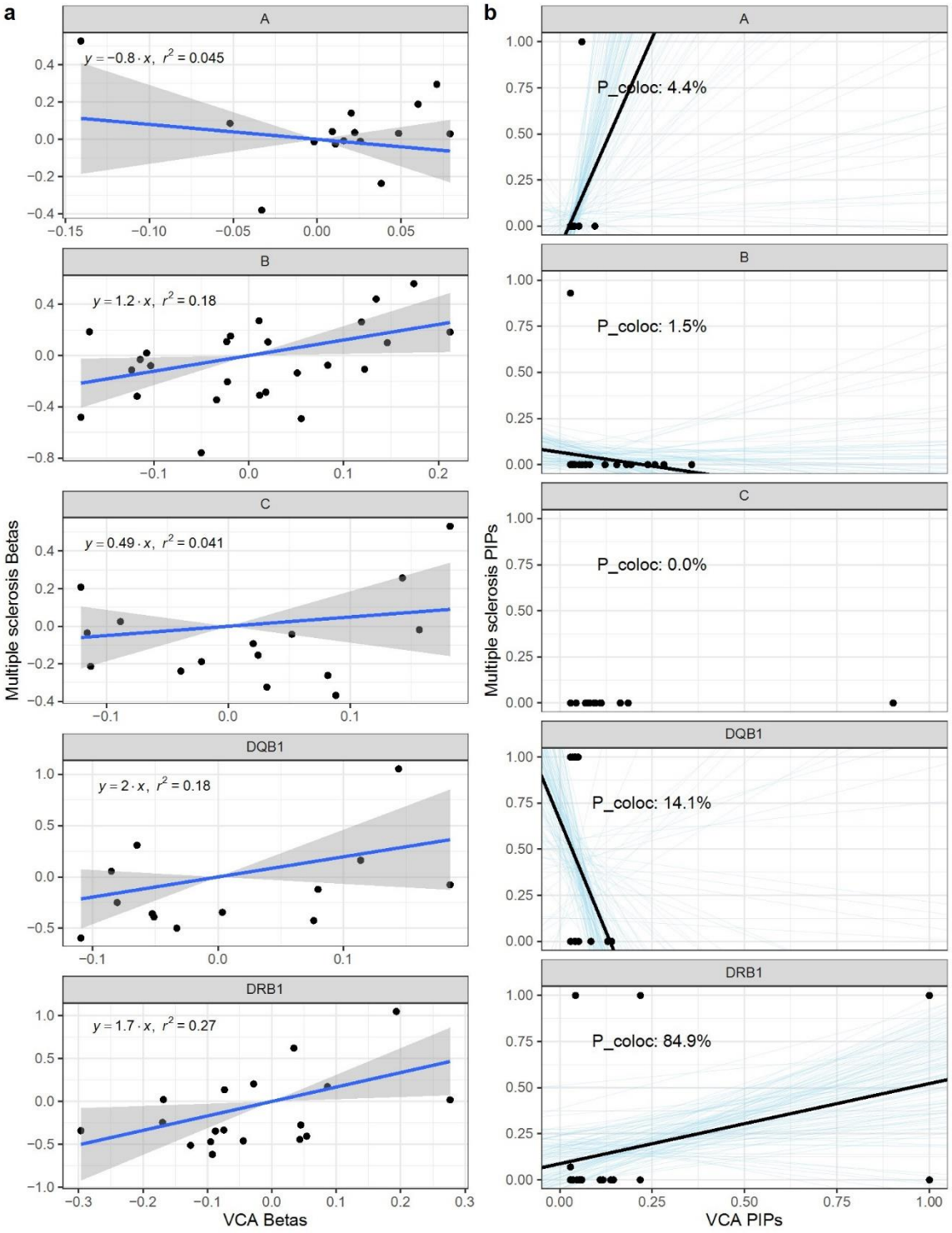
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of liver disease PIPs on HBV PIPs causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease. Once again, we observe HLA-colocalization at *HLA-DPB1*.

Supplementary Figure 6: EBNA and multiple sclerosis HLA-colocalization in the UK Biobank



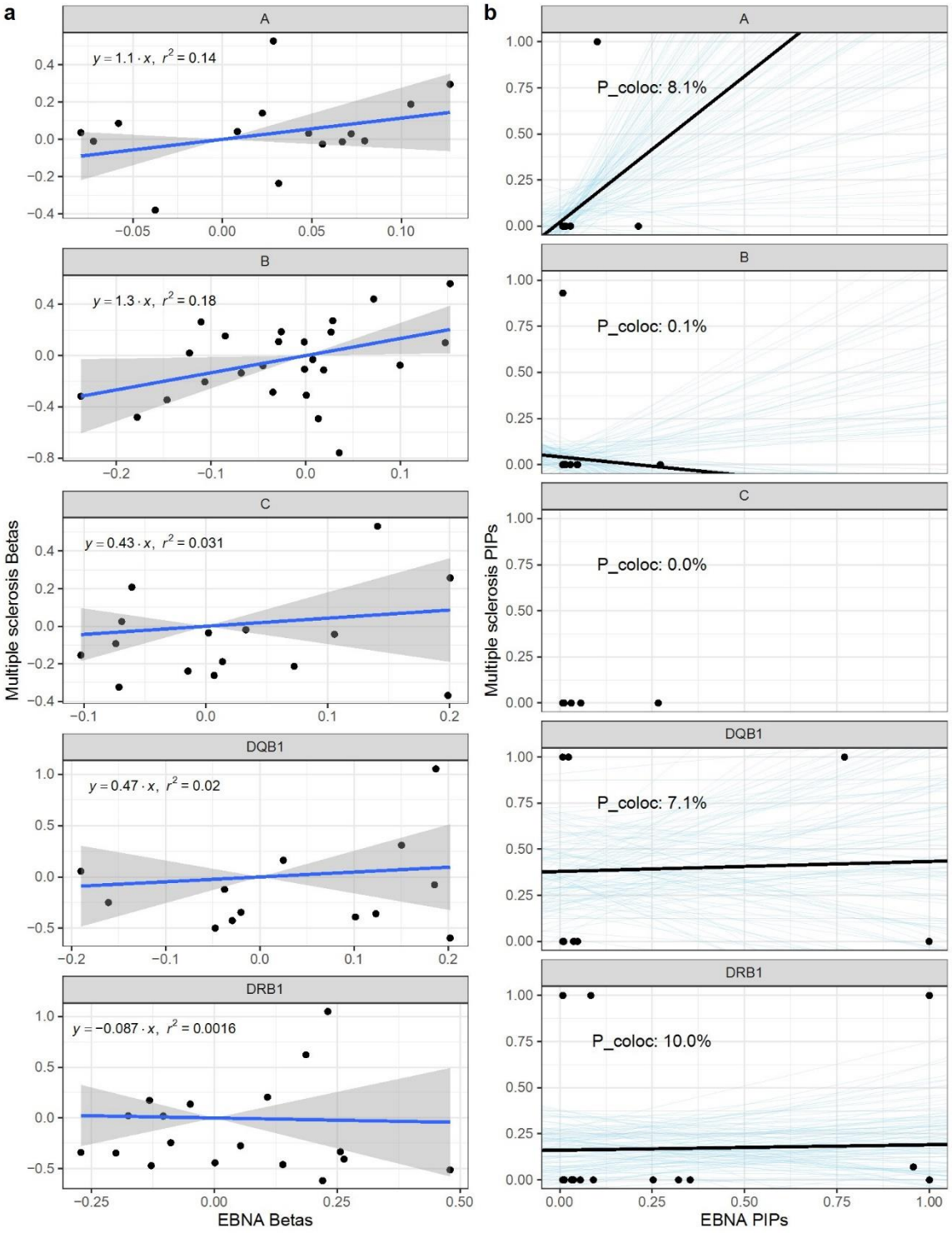
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on EBNA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.

Supplementary Figure 7: VCA and multiple sclerosis HLA-colocalization in the IMSGC



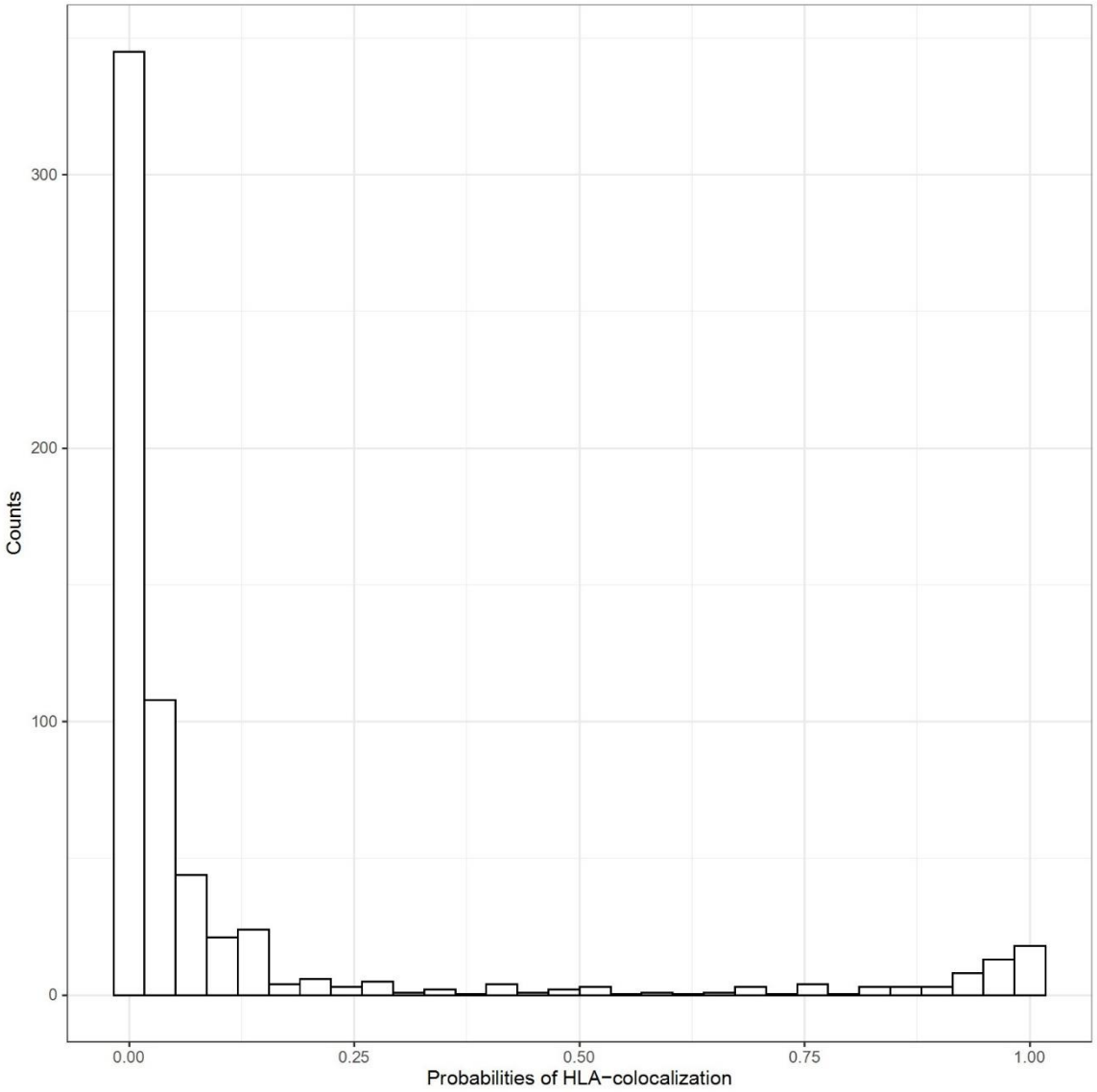
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on VCA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.

Supplementary Figure 8: EBNA and multiple sclerosis HLA-colocalization in the IMSGC



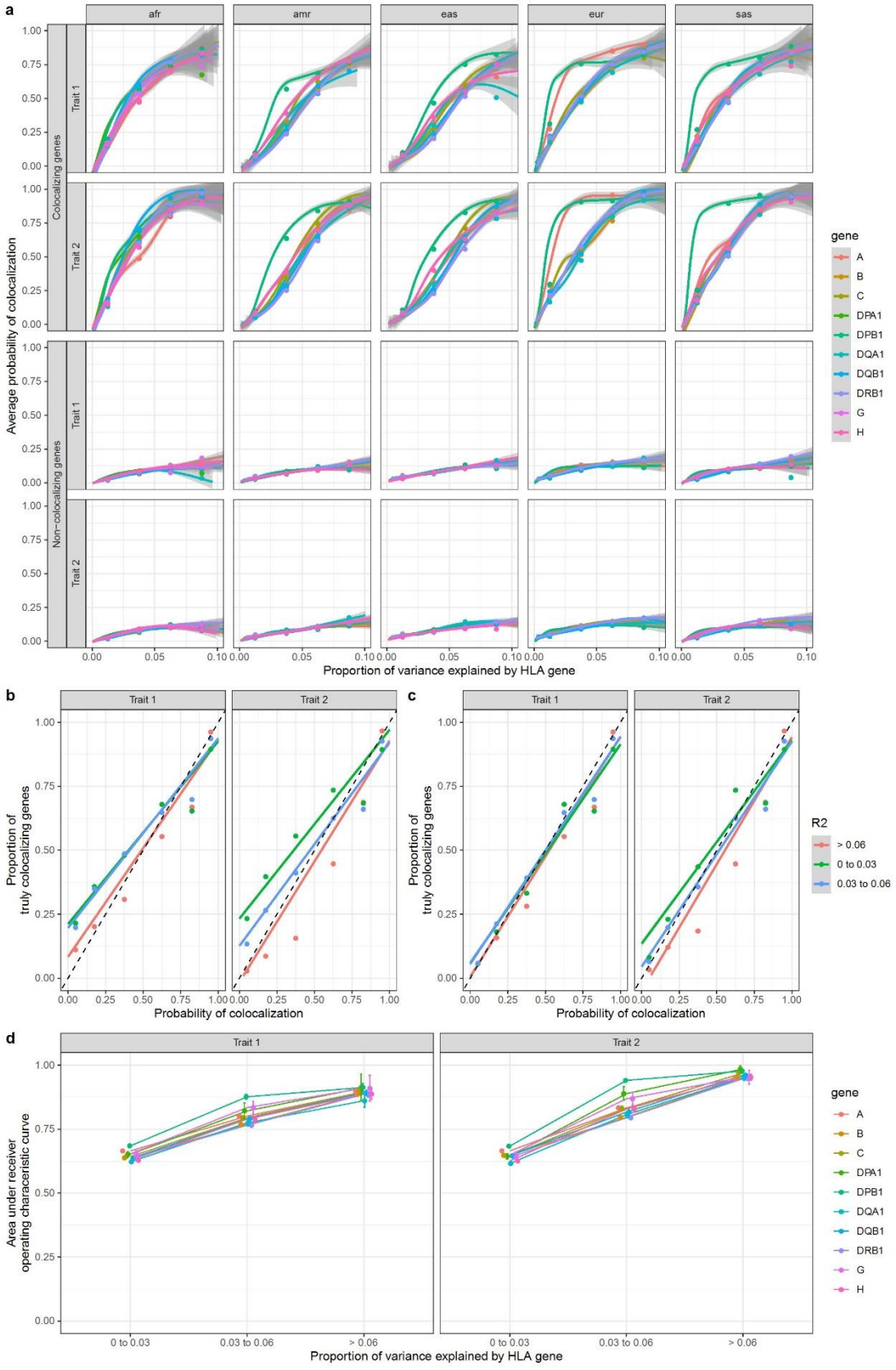
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on EBNA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.

Supplementary Figure 9: pathogen and auto-immune traits colocalization results



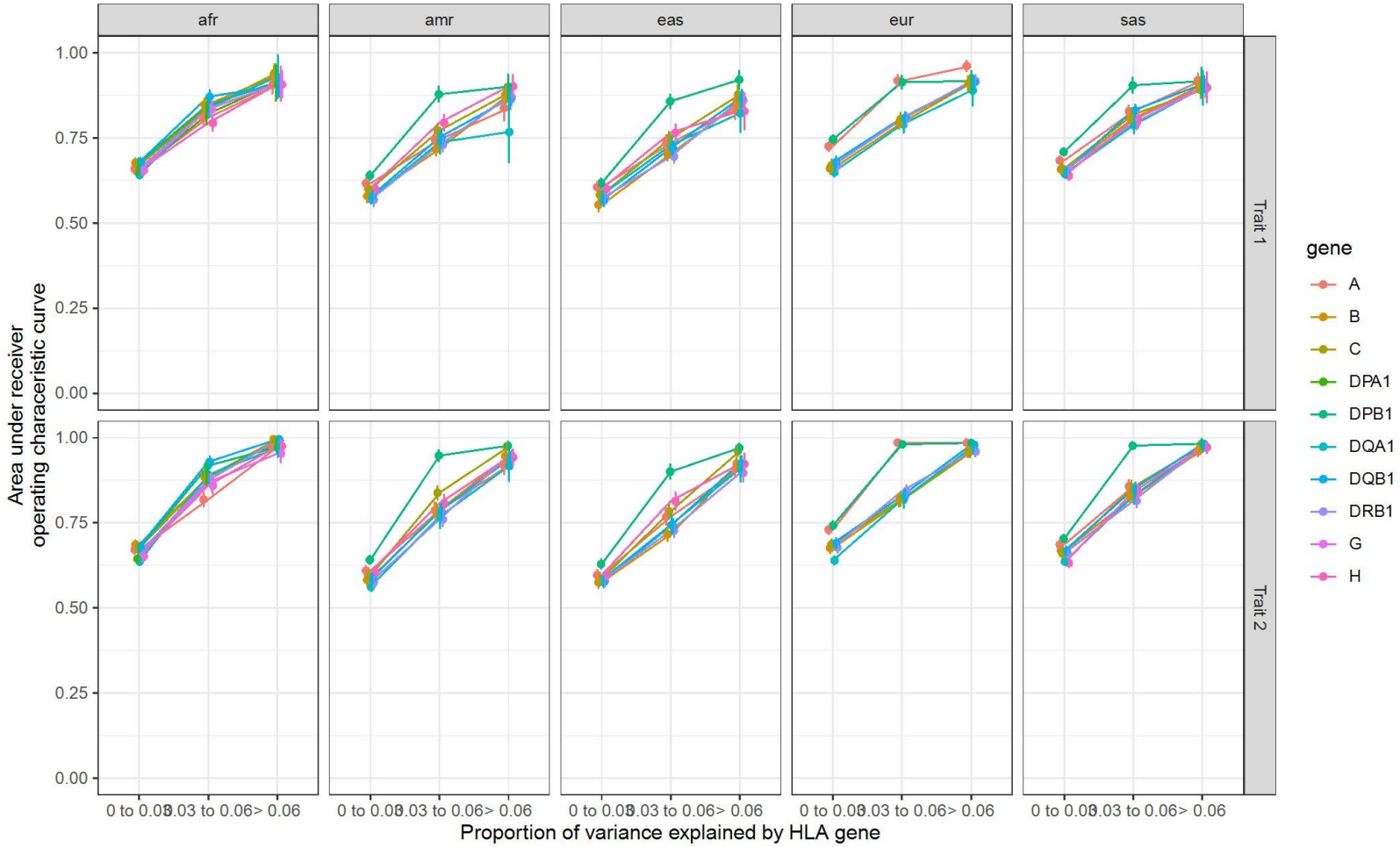
Distribution of HLA-colocalization probabilities for all pairs of pathogen serology and auto-immune diseases traits (n = 630 pathogen to autoimmune diseases pairs). As can be seen, most pairs of traits do not colocalize, which is expected and suggest that our method is well calibrated to complex real-world data. See **Supp. Data 2** for the full results.

Figure 10: HLA allele HLA-colocalization simulation results for quantitative traits with L = 20



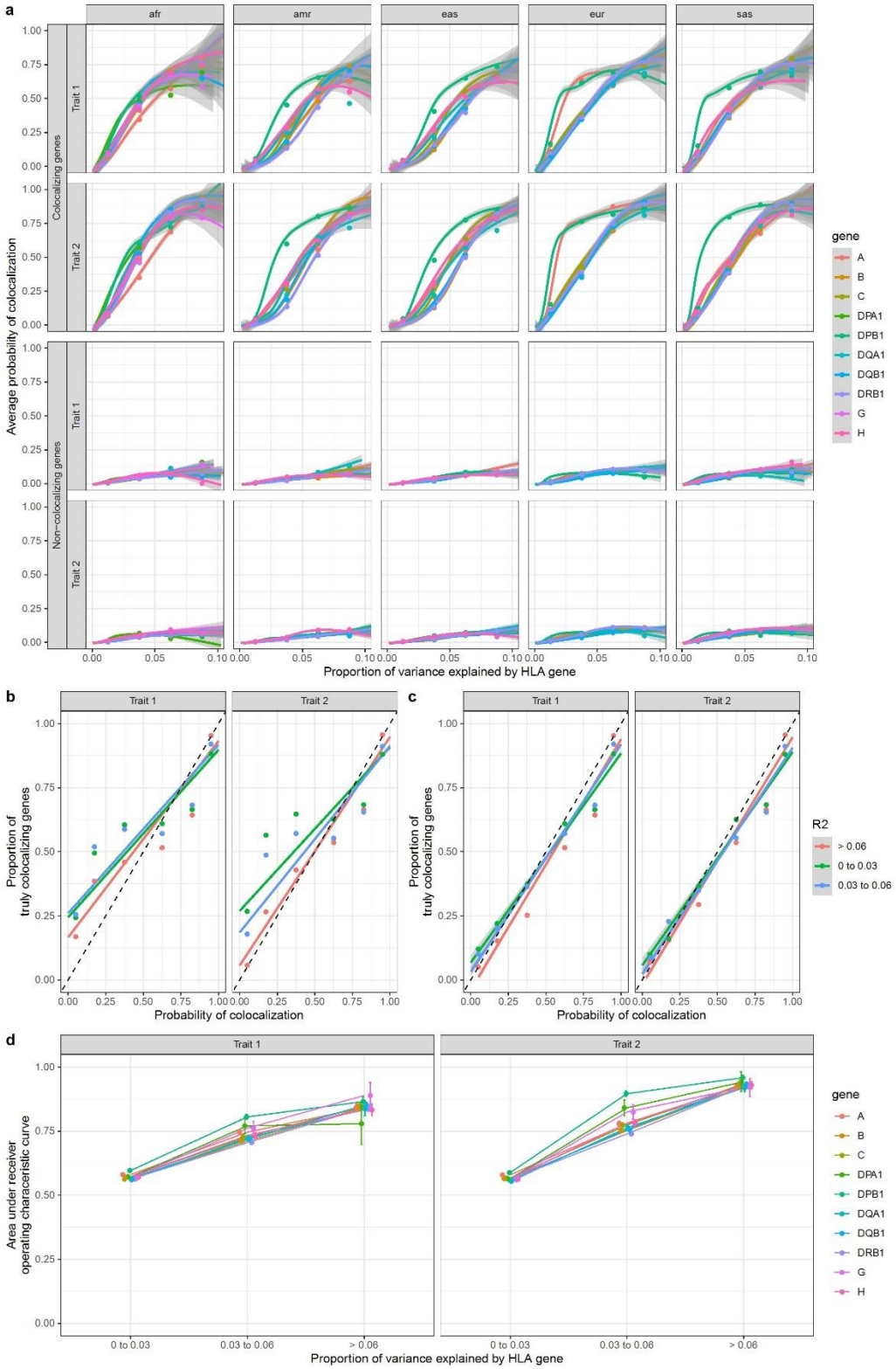
This simulation was done using the number of single effect option of SuSiE at 20 i.e. (L = 20). **a)** The average posterior probability of colocalization in truly colocalizing increases with the amount of phenotype variance explained by each gene, as expected. The average posterior probability of colocalization in truly non-colocalizing genes remains stable with increasing variance explained. The lines were drawn using a generalized additive model with *geom_smooth* in R. The grey area represents 95% confidence intervals. The individual dots represent the average in the corresponding variance bins. **b)** The proportion of simulated genes that were truly colocalizing shown as a function of the probability of colocalization. This is close to the identity line, though errs on the more conservative side for genes with lower R2. **c)** The deviation from the identity line is largely due to situations where SuSiE is unable to assign a PIP larger than 50% in at least one allele at a gene. When we restrict to genes with a minimal PIP of 50%, the method is almost perfectly calibrated. **d)** Average area under the curve as a function of variance explained for each gene. For this plot, average ROC area under the curve across ancestry was shown. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry.

Supplementary Figure 11: Per ancestry ROC area under the curves for simulations of quantitative traits with L = 20



This simulation was done using the number of single effect option of SuSiE at 20 i.e. (L = 20). Area under the ROC curves of HLA-colocalization PIPs for different variance explained per gene and genetic ancestries for the simulation of the quantitative traits. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry.

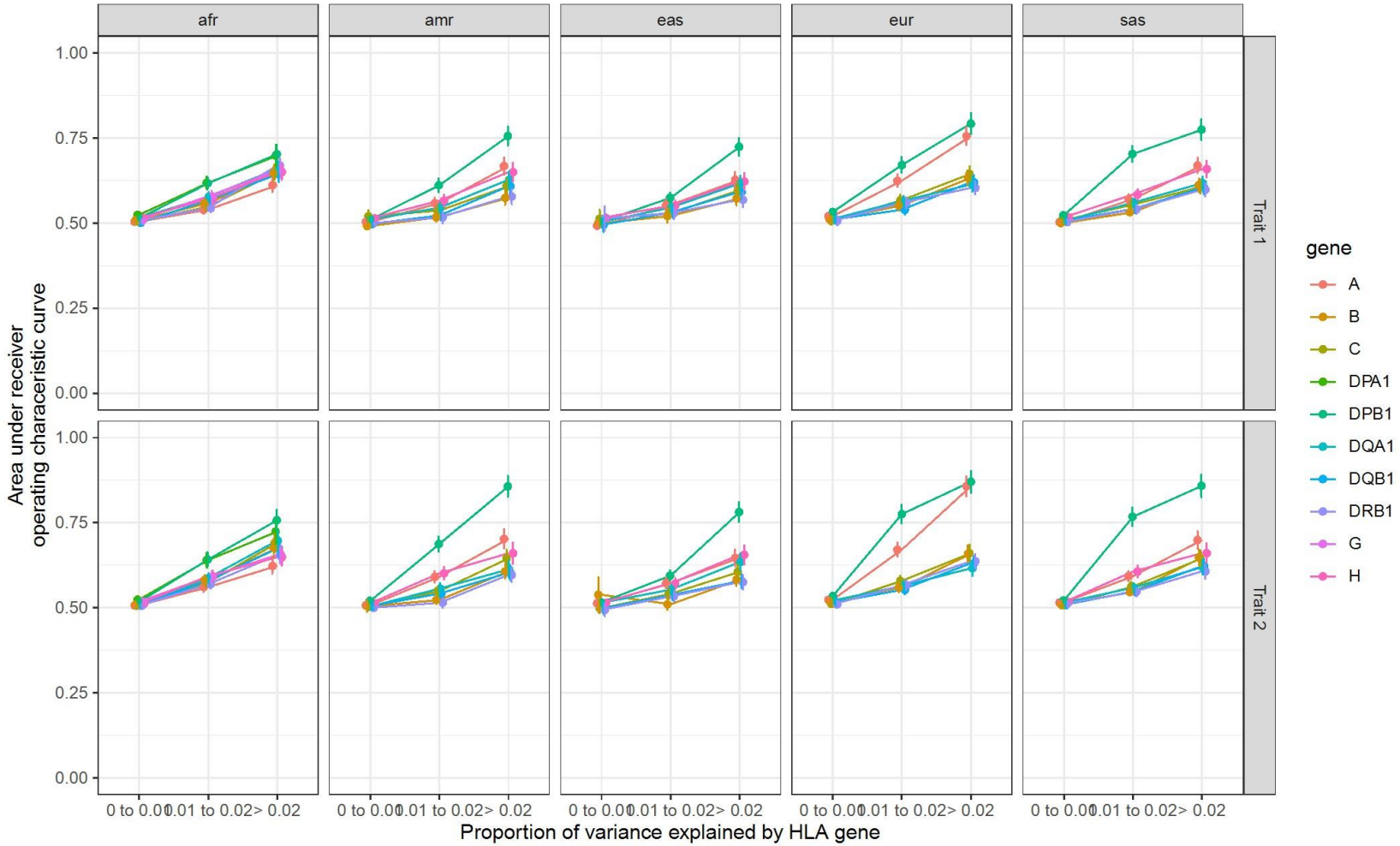
Supplementary Figure 12: HLA allele HLA-colocalization simulation results for binary traits with L = 20



This simulation was done using the number of single effect option of SuSiE at 20 i.e. (L = 20). **a)** The average posterior probability of colocalization in truly colocalizing increases with the amount

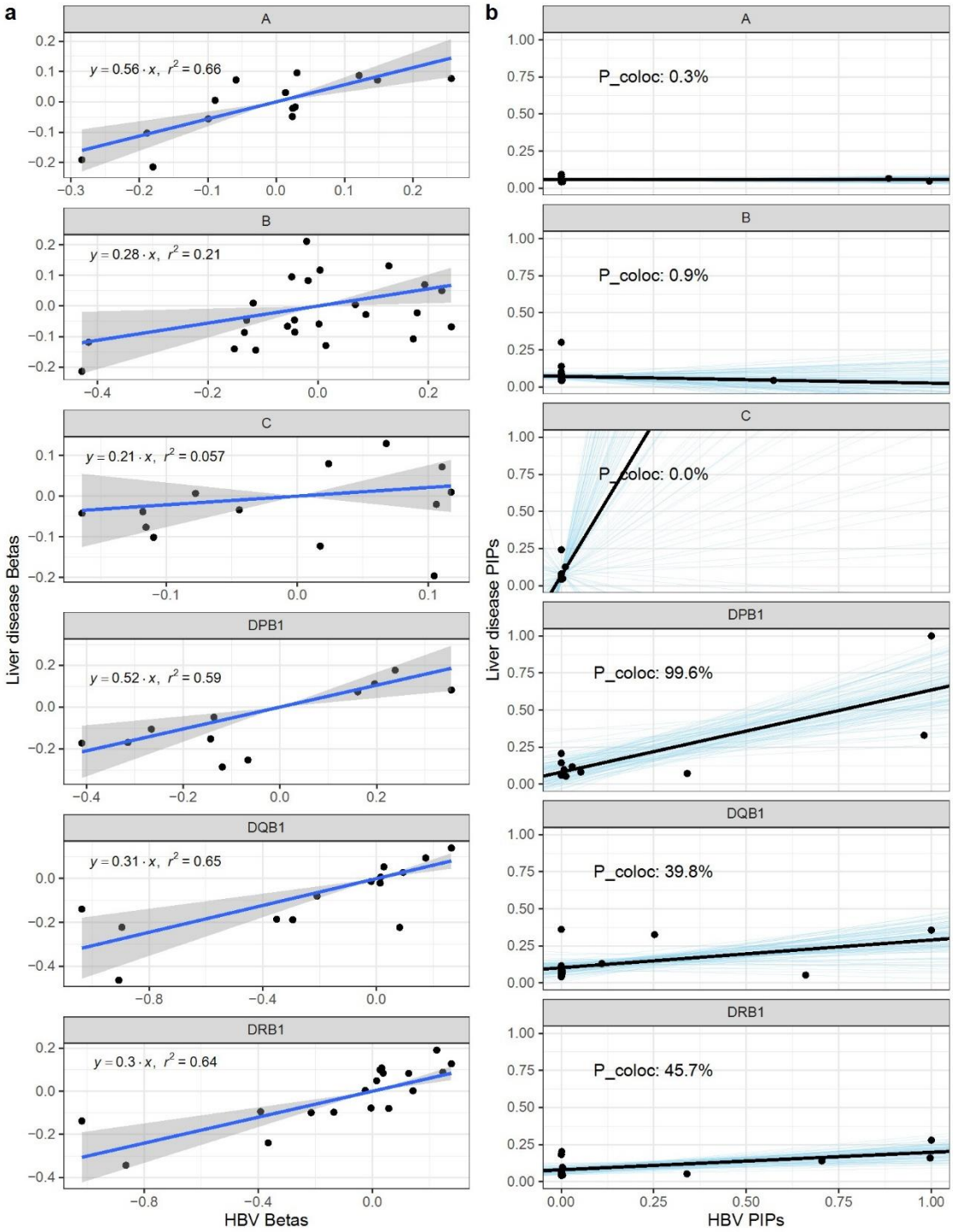
of phenotype variance explained by each gene, as expected. The average posterior probability of colocalization in truly non-colocalizing genes remains stable with increasing variance explained. The lines were drawn using a generalized additive model with *geom_smooth* in R. The grey area represents 95% confidence intervals. The individual dots represent the average in the corresponding variance bins. **b)** The proportion of simulated genes that were truly colocalizing shown as a function of the probability of colocalization. This is close to the identity line, though errs on the more conservative side for genes with lower R2. **c)** The deviation from the identity line is largely due to situations where SuSiE is unable to assign a PIP larger than 50% in at least one allele at a gene. When we restrict to genes with a minimal PIP of 50%, the method is almost perfectly calibrated. **d)** Average area under the curve as a function of variance explained for each gene. For this plot, average ROC area under the curve across ancestry was shown. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry.

Supplementary Figure 13: Per ancestry ROC area under the curves for simulations of binary traits with L = 20



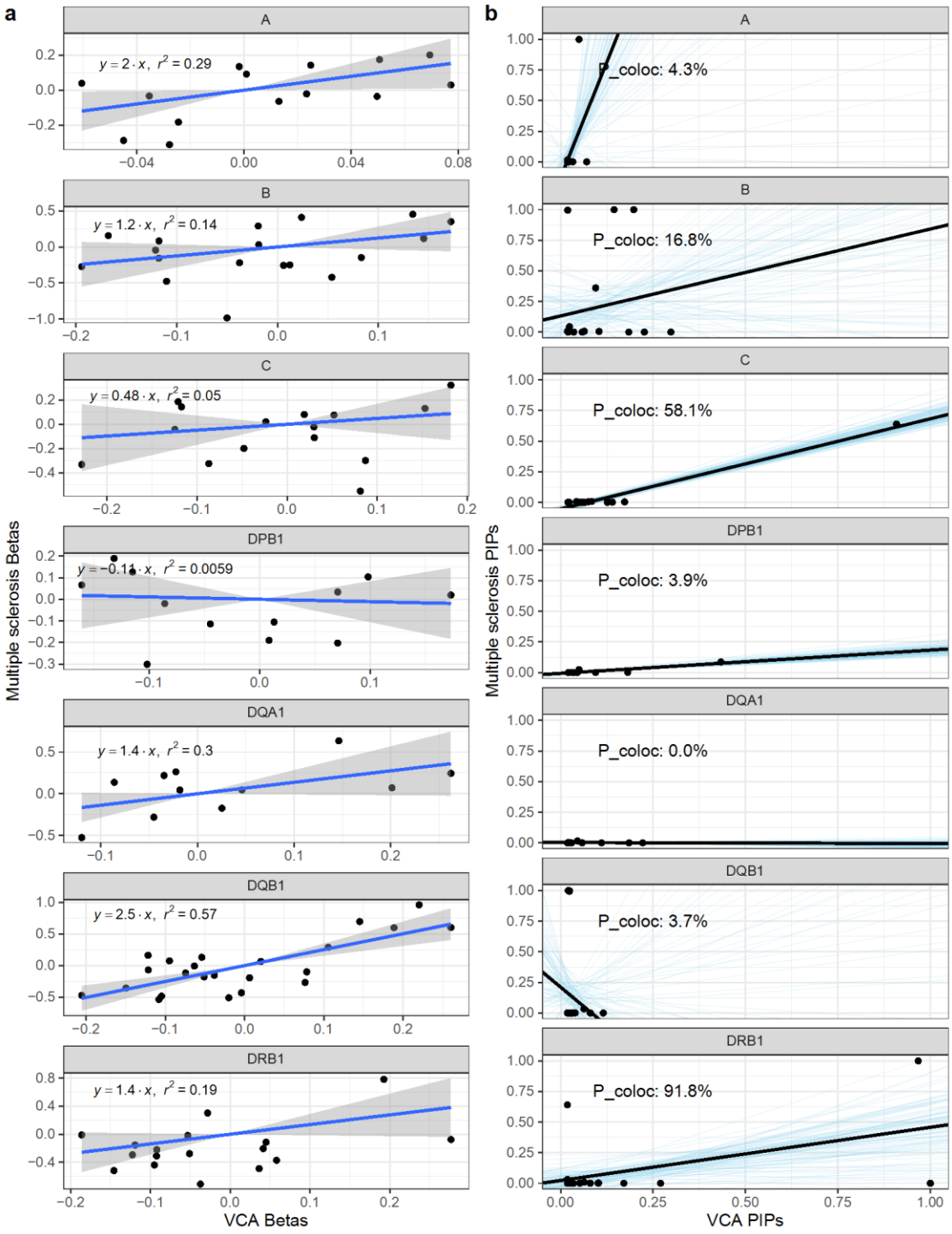
This simulation was done using the number of single effect option of SuSiE at 20 i.e. (L = 20). Area under the ROC curves of HLA-colocalization PIPs for different variance explained per gene and genetic ancestries for the simulation of the binary traits. Legend: afr: African genetic ancestry, amr: Admixed American genetic ancestry, eas: East Asian genetic ancestry, eur: European genetic ancestry, sas: South Asian genetic ancestry

Supplementary Figure 14: Liver disease and HBV antigenemia HLA-colocalization with allele frequencies > 1%



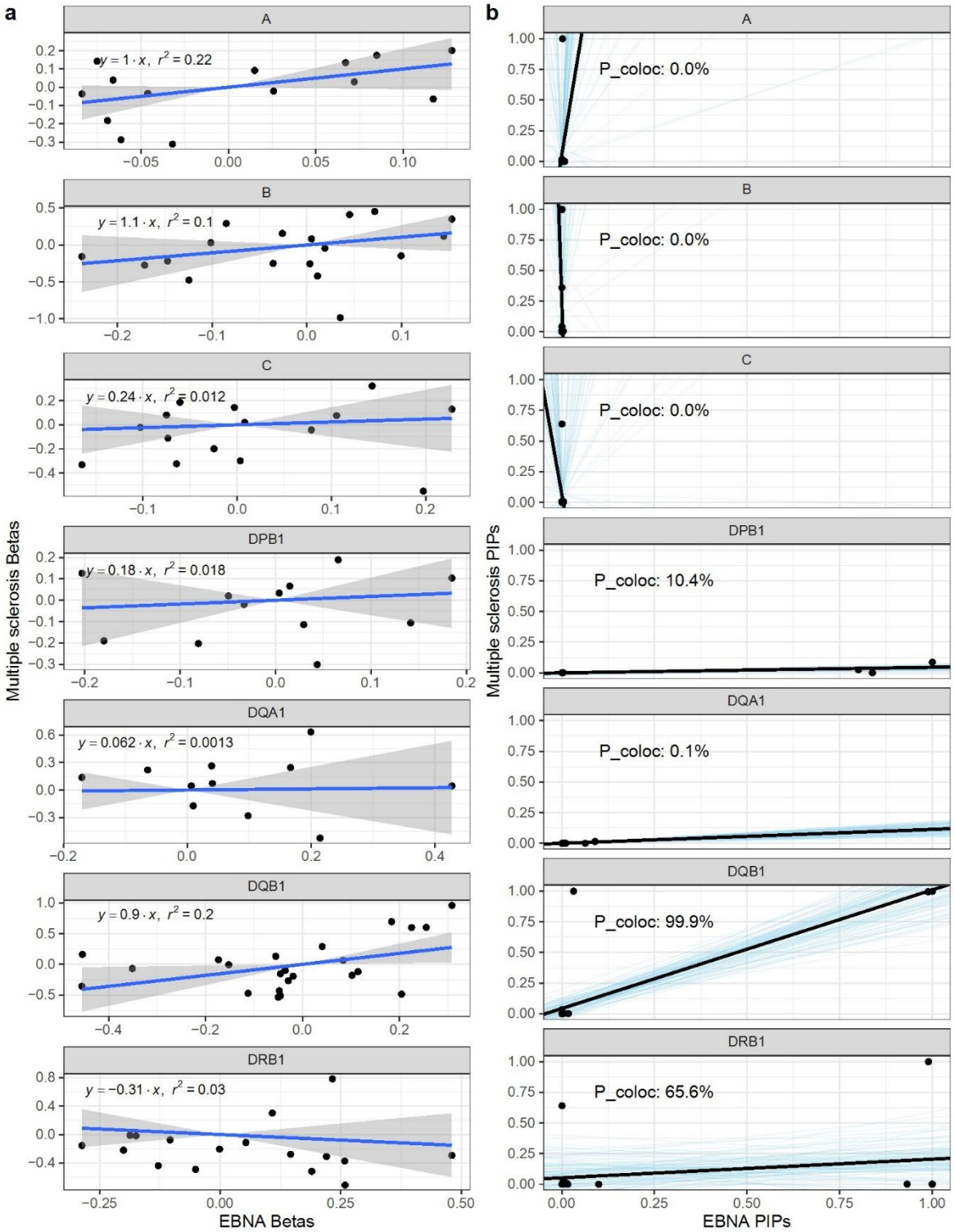
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of HBV and liver disease PIP causal signature. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease. Hence, after Bayesian variable selection at the HLA locus, *HLA-DPB1* shows evidence of shared liver disease and HBV genetic architecture.

Supplementary Figure 15: VCA and multiple sclerosis HLA-colocalization in the UK Biobank with allele frequencies > 1%



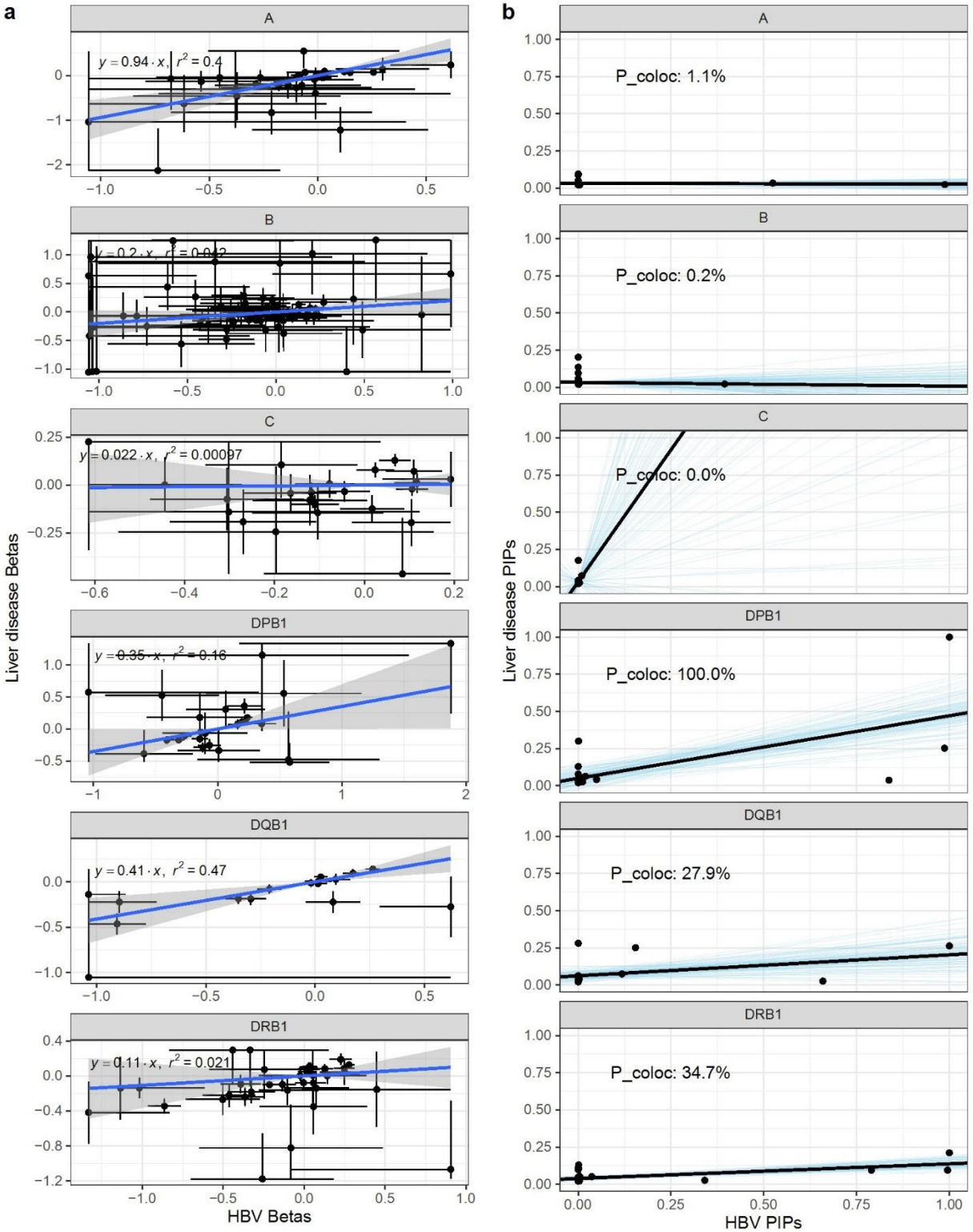
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on VCA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.

Supplementary Figure 16: EBNA and multiple sclerosis HLA-colocalization in the UK Biobank with allele frequencies > 1%



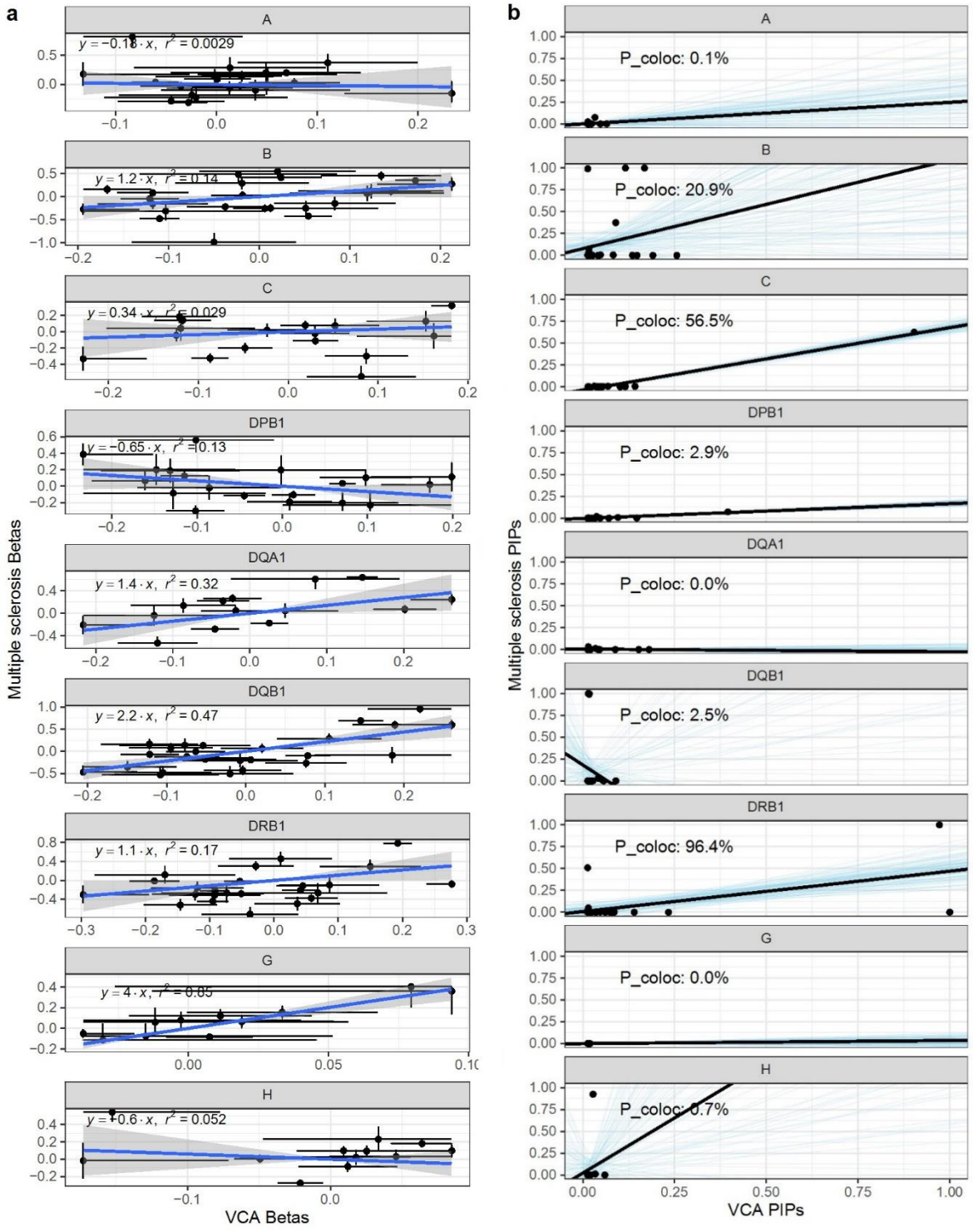
a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on EBNA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.

Supplementary Figure 17: Liver disease and HBV antigenemia HLA-colocalization with error bars



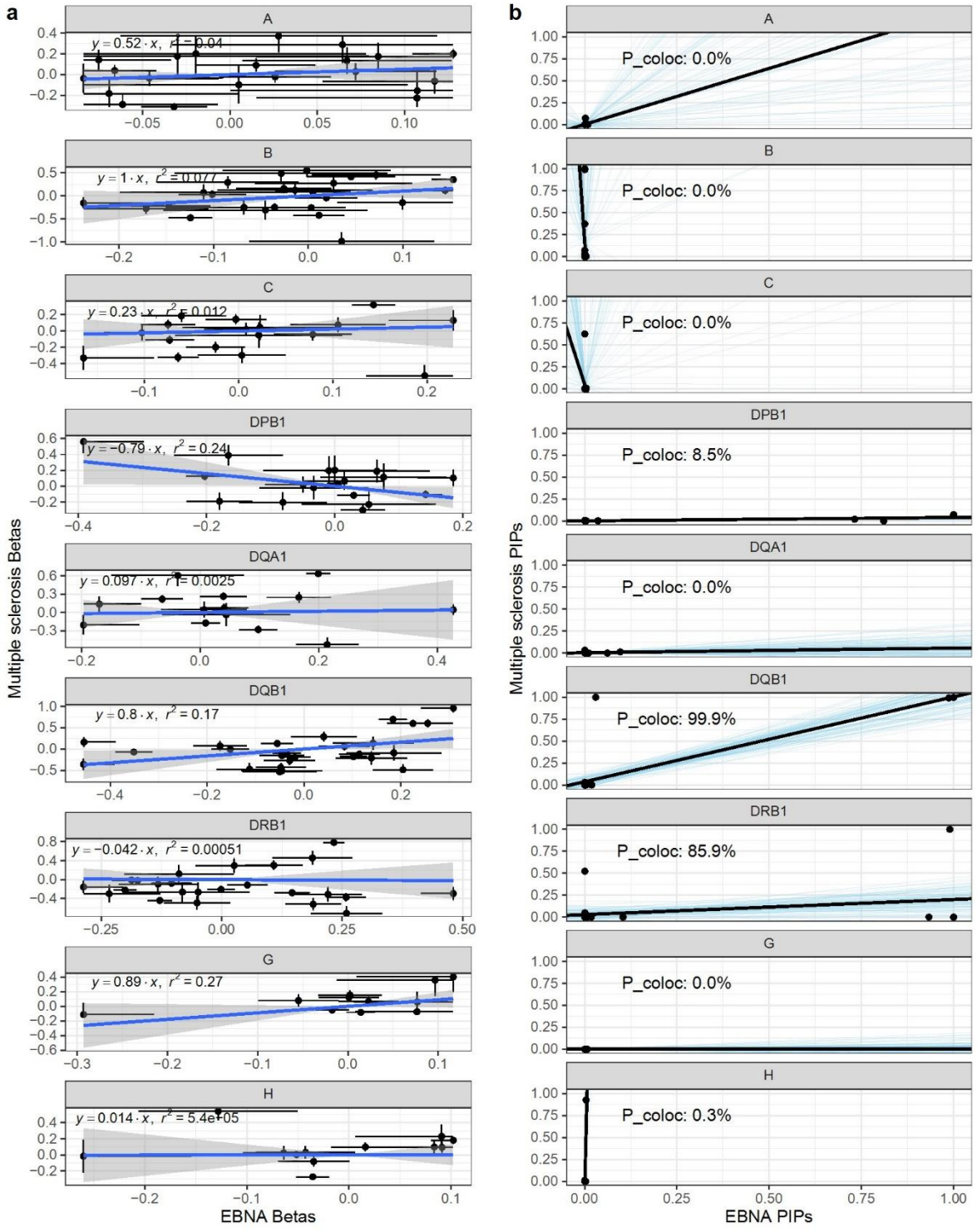
a) linear regression (with 95% confidence intervals) and 1 standard error bars of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of HBV and liver disease PIP causal signature. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease. Hence, after Bayesian variable selection at the HLA locus, *HLA-DPB1* shows evidence of shared liver disease and HBV genetic architecture.

Supplementary Figure 18: VCA and multiple sclerosis HLA-colocalization in the UK Biobank with error bars



a) linear regression (with 95% confidence intervals) and 1 standard error bars of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on VCA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.

Supplementary Figure 19: EBNA and multiple sclerosis HLA-colocalization in the UK Biobank with error bars



a) linear regression (with 95% confidence intervals) of beta coefficients from the additive HLA allele association studies. **b)** Bayesian regression of multiple sclerosis PIPs on EBNA PIP causal signatures. The black lines show the regression fit, while the blue lines show 100 random draws from the posterior distributions. The resulting probabilities of HLA-colocalization (P_coloc) are also written for ease.