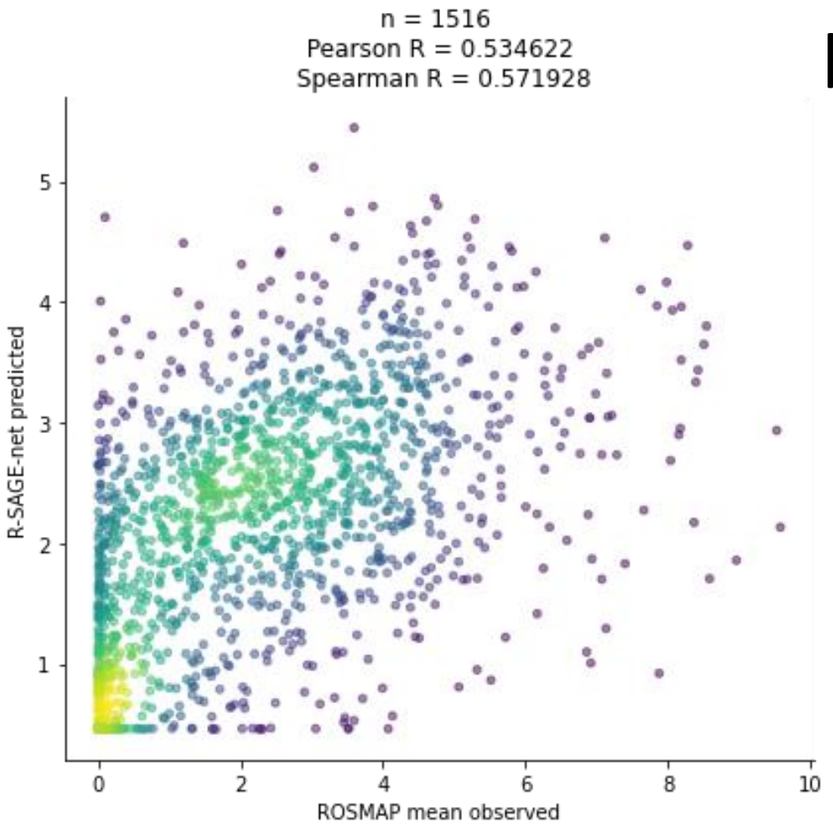
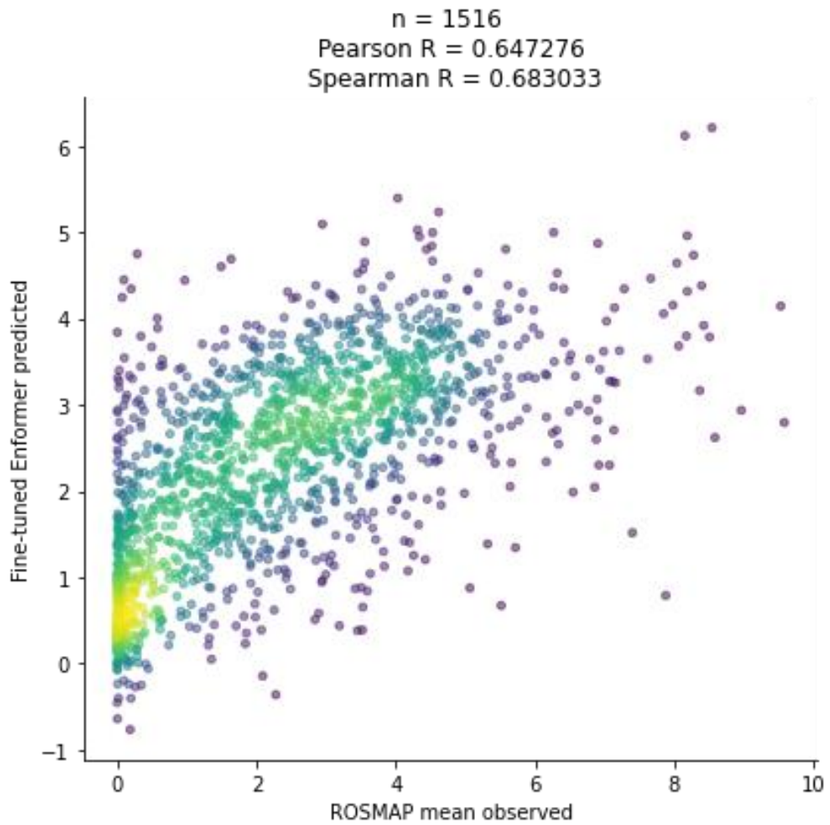
Figure S1

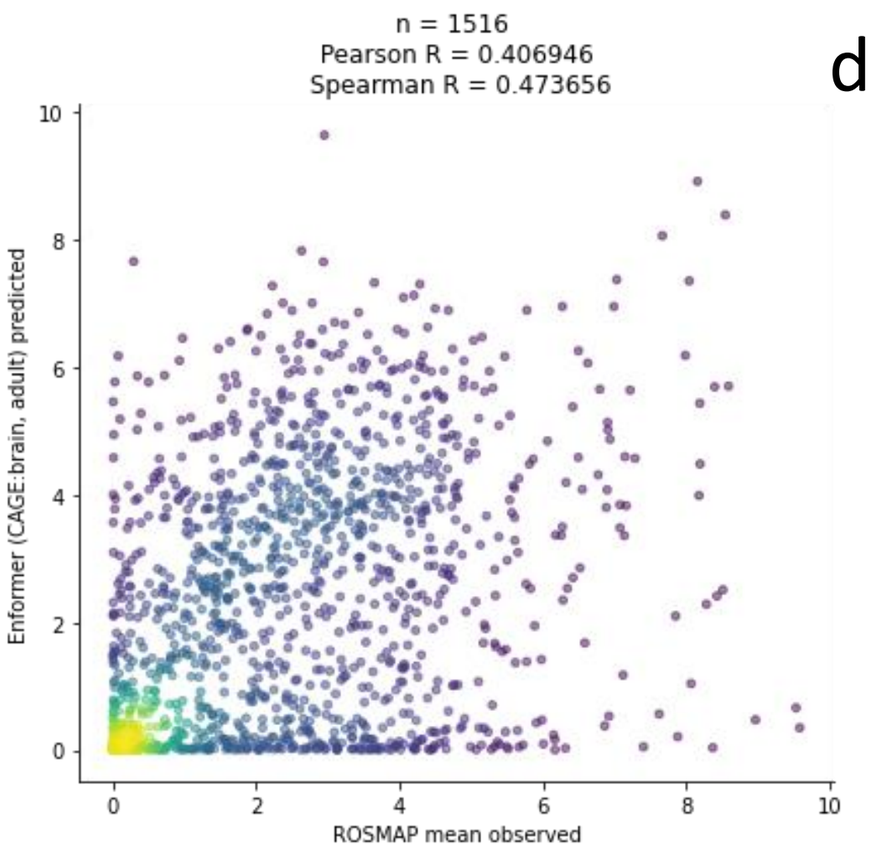
a



b



c



d

Model	N Trainable parameters	Inference time (1 gene, 859 individuals)
r-SAGE-net	2.7M	~20s
Enformer	249M	~1392s

Figure S1. Comparisons between r-SAGE-net and Enformer. (a) R-SAGE-net trained on genes in Enformer's training set (n=15335) and evaluated on genes in Enformer's test set (n=1516). R-SAGE-net is trained and evaluated on 40 kb reference sequence and average log-transformed gene expression across ROSMAP individuals. Scatterplot is colored by density. (b) Fine-tuned Enformer evaluated on Enformer's test set genes. Fine-tuning follows the procedure outlined in our previous work (9) (Methods). Evaluation is done with the same length input as in Enformer training (196,608 bp). (c) Enformer evaluated on Enformer's test set genes without fine-tuning. Insead of fine-tuning, we select the representative track ("CAGE:brain, adult"), as done previously (9). (d) Comparison between r-SAGE-net and Enformer model sizes and inference times. Both models are evaluated on one NVIDIA RTX A4000 GPU.

Figure S2

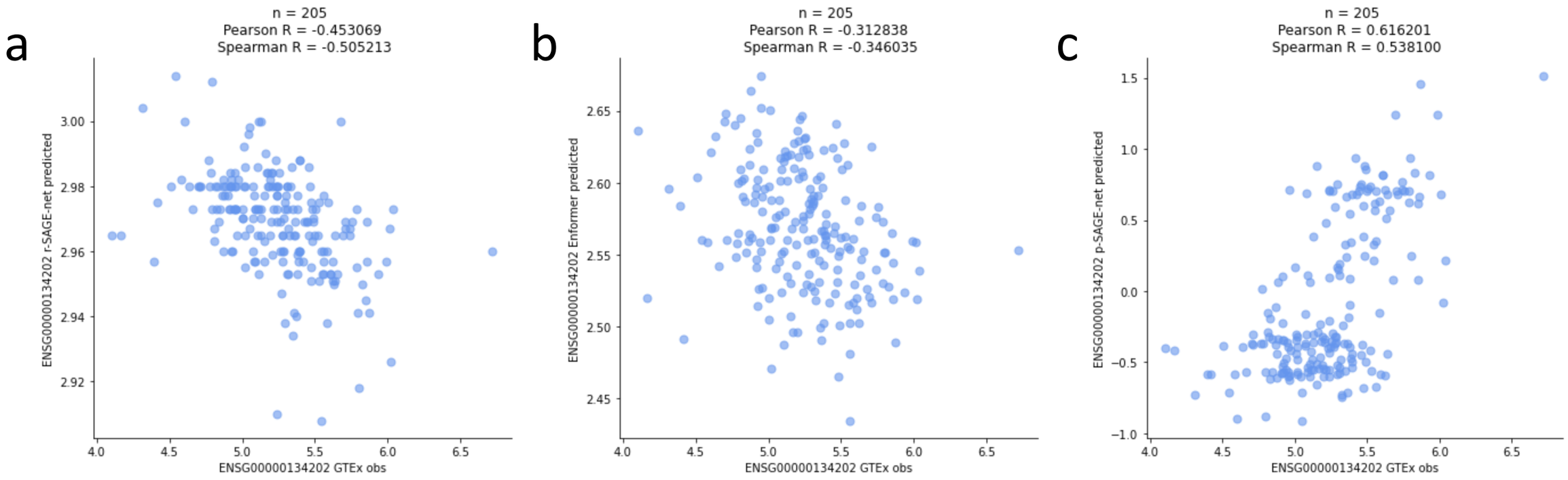


Figure S2. Model performance for GSTM3. (a) R-SAGE-net GTEx predicted vs. observed. (b) Fine-tuned Enformer GTEx predicted vs. observed. (c) P-SAGE-net GTEx predicted vs. observed. Note that the p-SAGE-net model used here was trained on the top 1000 gene set, which includes GSTM3 (but the GTEx individuals shown here were not seen in model training).

Figure S3

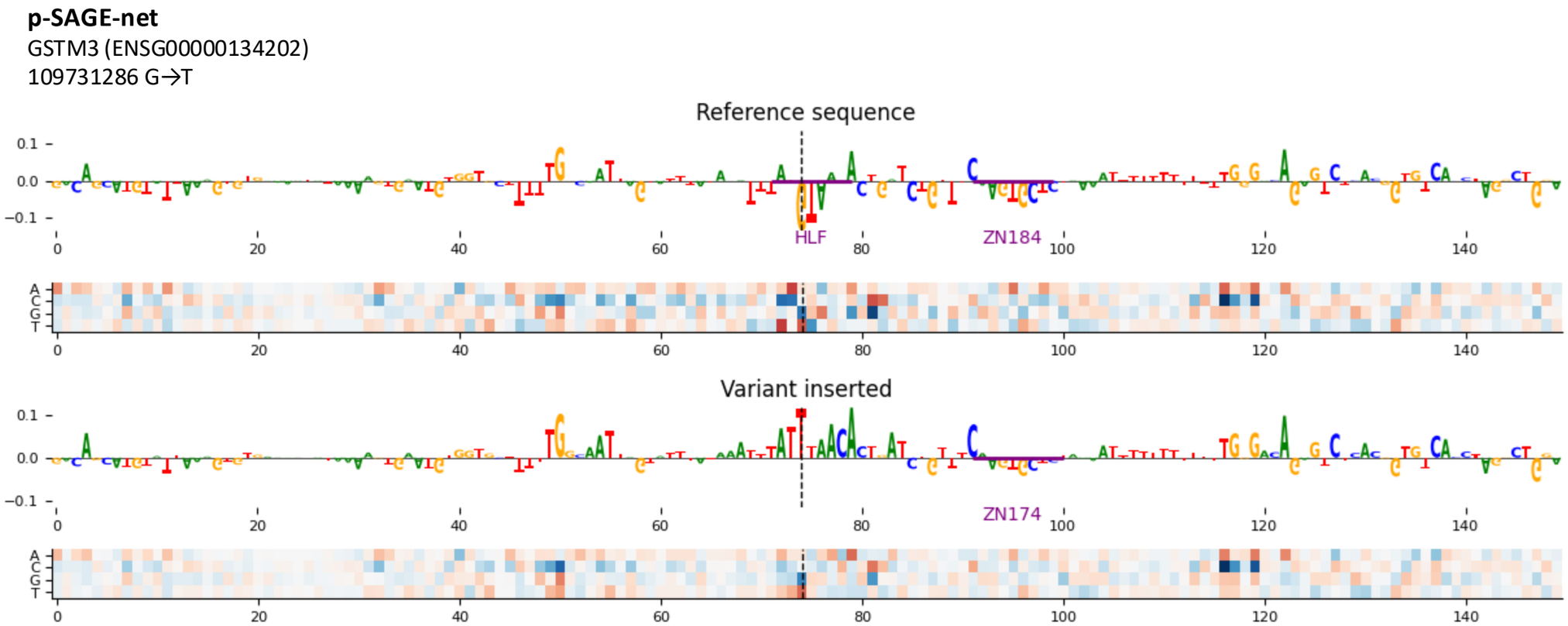


Figure S3. P-SAGE-net ISM for GSTM3. (a) Zero-centered ISM for a 150bp window around the variant of interest (chromosome 1, hg38 position 109731286 G → T), shown for the reference sequence. Motifs shown are from seqlets matched to the HOCOMOCO v12 database (18) by TOMTOM (19) with p<0.05. (b) Same ISM as in (a), but shown for all bases instead of only reference sequence. White represents value=0. (c) Same ISM as in (a), but done for reference sequence with variant inserted (T instead of G at 109731286). (d) Same ISM as in (c), but shown for all bases instead of only reference sequence.

Figure S4

r-SAGE-net
GSTM3 (ENSG00000134202)
109731286 G→T

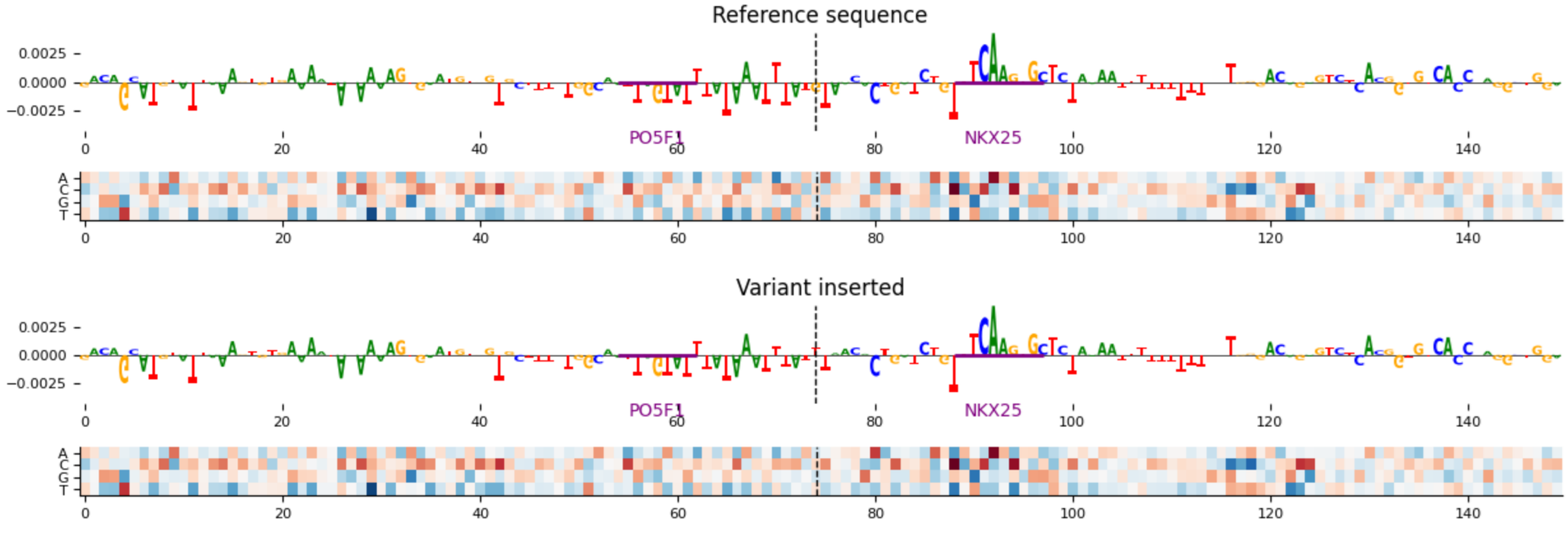


Figure S4. R-SAGE-net ISM for GSTM3. Same analysis as Fig. S3, but using r-SAGE-net instead of p-SAGE-net.

Figure S5

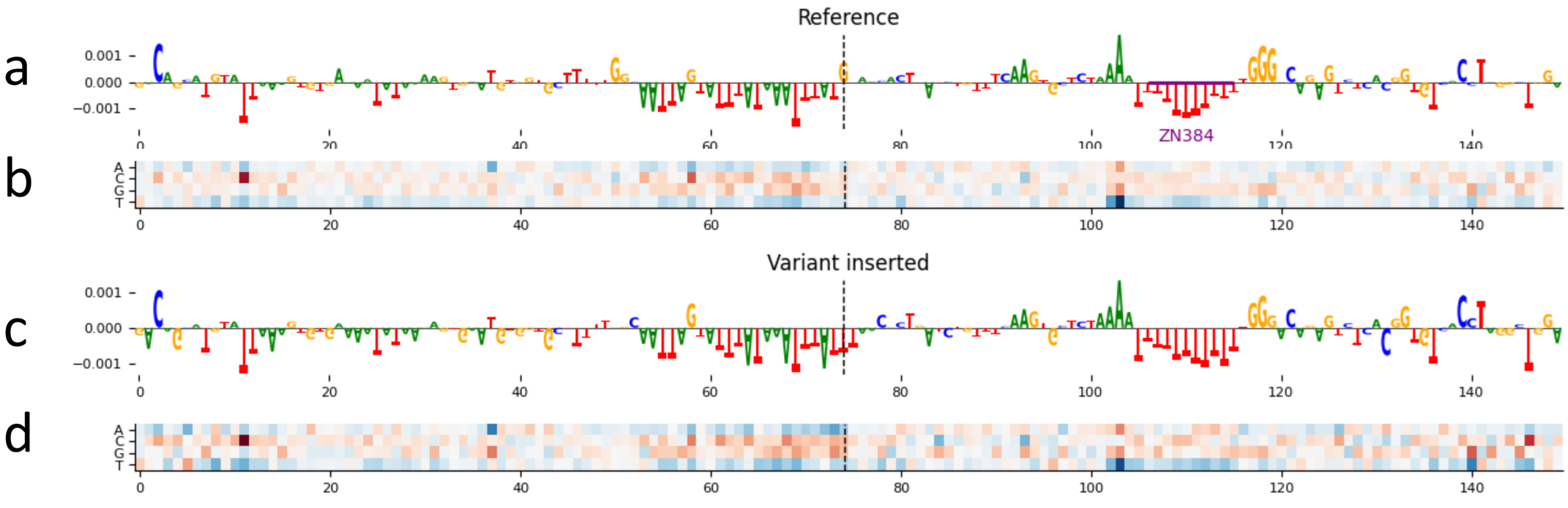
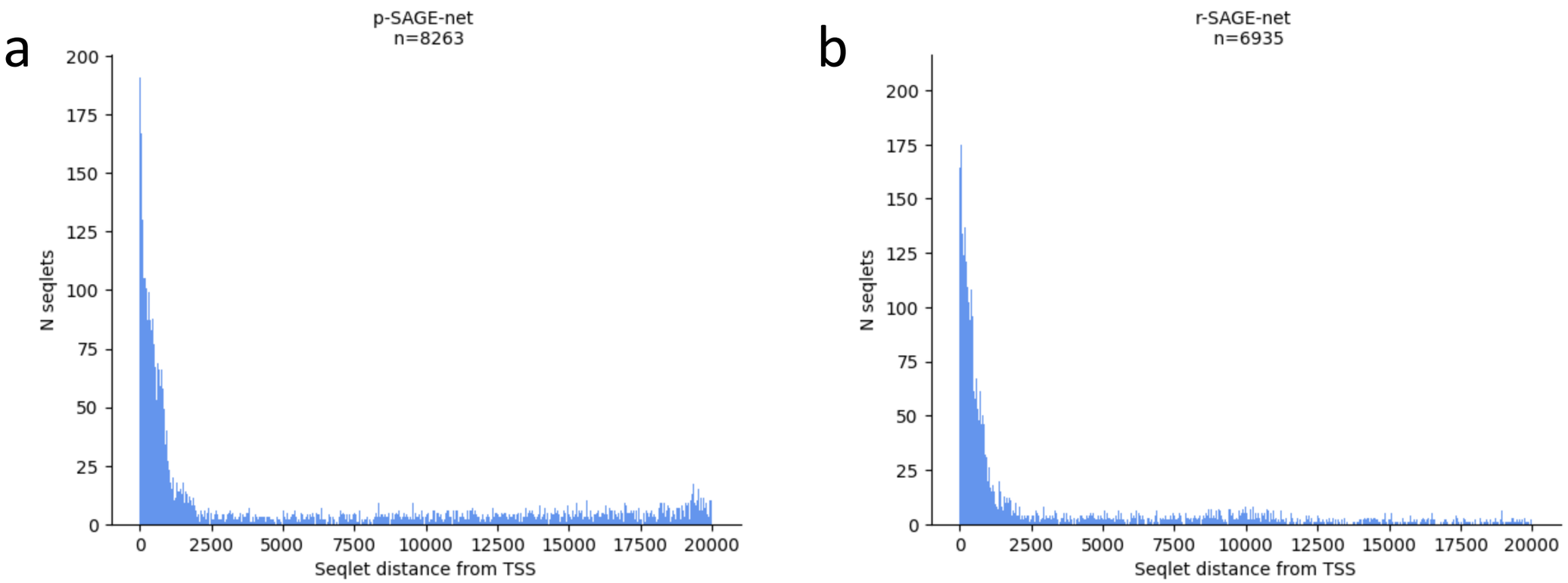


Figure S5. Enformer ISM for GSTM3. Same analysis as Fig. S3, but using fine-tuned Enformer instead of p-SAGE-net.

Figure S6



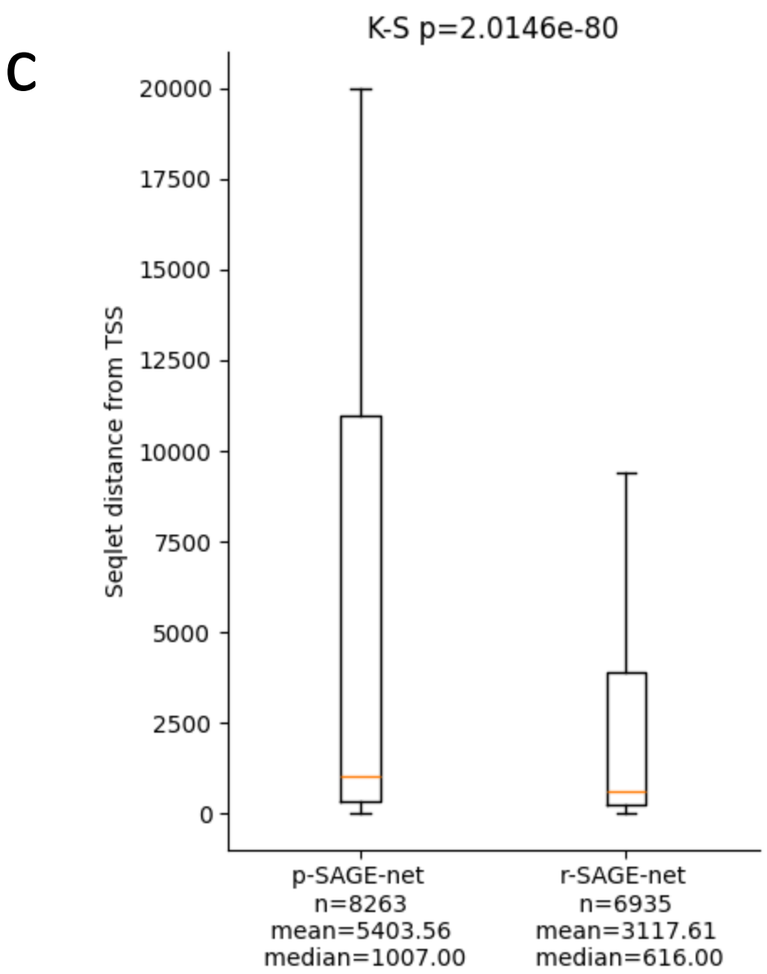


Figure S6. Seqlet distances from TSS for p-SAGE-net vs. r-SAGE-net. (a) Distribution of p-SAGE-net seqlet distances from TSS. Analysis shown (combined) for high performance genes in model training set (n=466 genes with GTEx per-gene correlation>0.3). We approximate ISM values using gradients (20) and then identify seqlets with p<0.005 (see Methods). (b) Same analysis as (a), but with r-SAGE-net instead of p-SAGE-net. (c) Comparison between seqlet distances from gene TSS for p-SAGE-net and r-SAGE-net, p-values from the two-sample Kolmogorov-Smirnov test for goodness of fit.

Figure S7

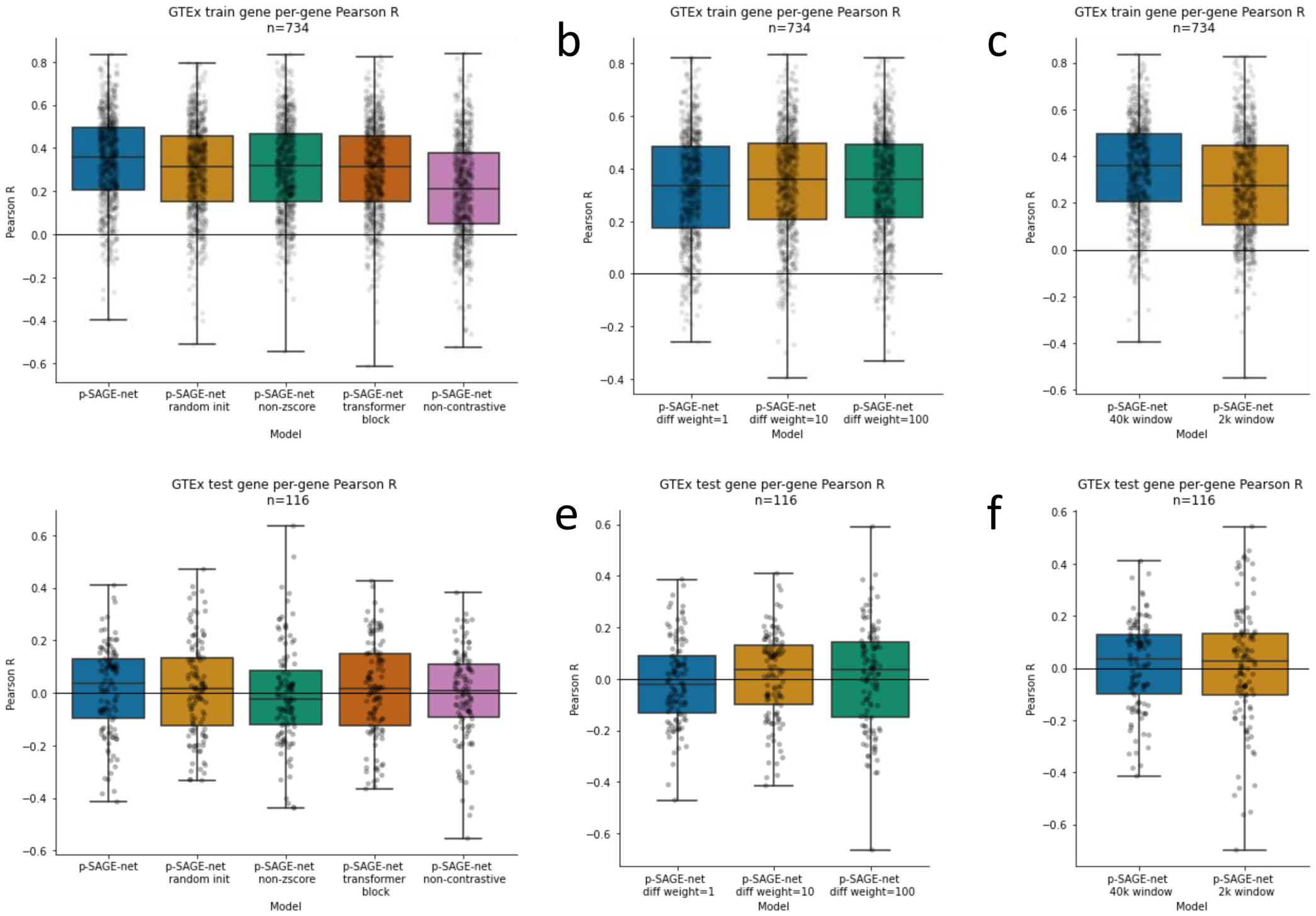


Figure S7. Model ablation analyses. (a) Comparison between best model (p-SAGE-net) and variations: randomly initializing model weights instead of initializing from r-SAGE-net ("random init"), using personal expression difference instead of z-score for model difference output ("non-zscore"), replacing the last convolutional block with a transformer block ("transformer block"), and predicting personal gene expression from personal sequence, without the contrastive approach ("non-contrastive"). See Methods for details on each variation. Each model is trained on 689 ROSMAP training individuals for 734 training genes from the top 1000 gene set and evaluated on 205 GTEx individuals for the same gene set. Each dot is one gene, boxplots show interquartile range with whiskers extending to minimum and maximum. (b) Modifications to loss function hyperparameters. For each weight on the "difference" portion of the loss function (diff weight = 1, 10, 100), the weight on the "mean expression" portion of the loss function = 1 – so when diff weight = 1, the two portions of the loss function are equally weighted. The difference weight selected for all other analyses is 10. Model training and evaluation gene and individual sets are the same as in (a). (c) Comparison between models with 40k vs. 20k input window. Window size applies to both model training and evaluation. Model training and evaluation gene and individual sets are the same as in (a). (d) Same analysis as (a), but shown for unseen test genes (n=116). (e) Same analysis as (b), but shown for unseen test genes (n=116). (f) Same analysis as (c), but shown for unseen test genes (n=116).

Figure S8

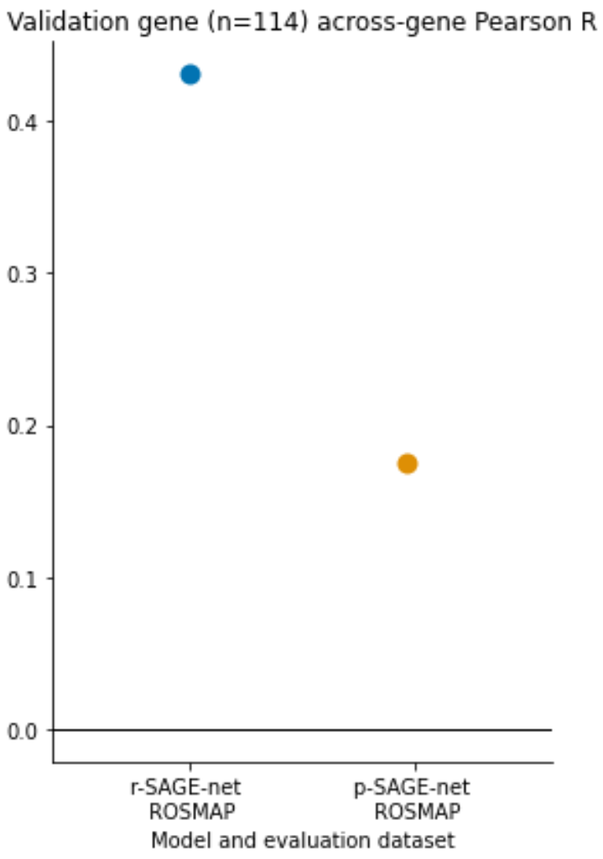


Figure S8. Validation gene across-gene performance for p-SAGE-net and r-SAGE-net. 114 validation genes are from the top 1000 gene set used in Fig. 1c,d. P-SAGE-net and r-SAGE-net models are the same as in Fig. 1c,d.

Figure S9

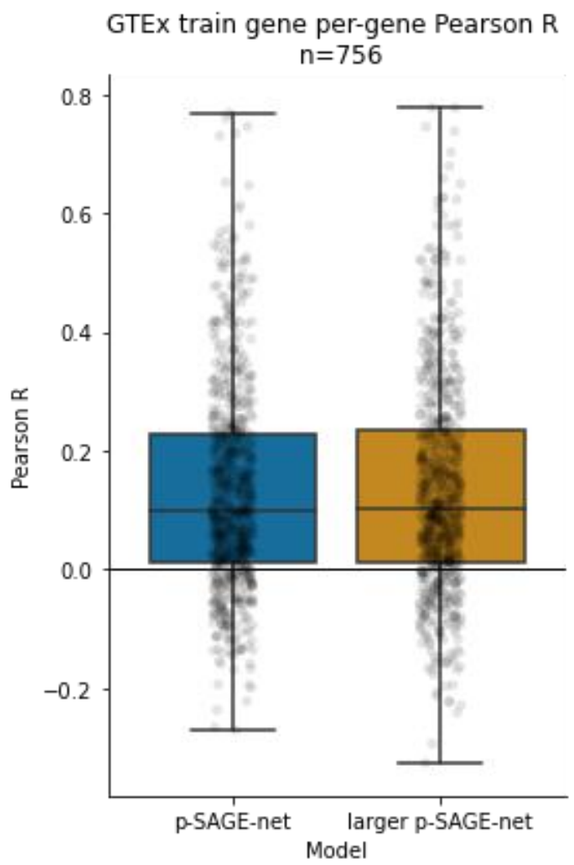


Figure S9. Training on a larger gene set with increased model capacity. The evaluation shown for p-SAGE-net is the same as Fig. 2d, “Random genes”, gene set size = 3000. The evaluation shown for “larger p-SAGE-net” is the same but for a model with double the number of convolutional kernels in all convolutional layers after the first convolutional layer (512 instead of 256).

Figure S10

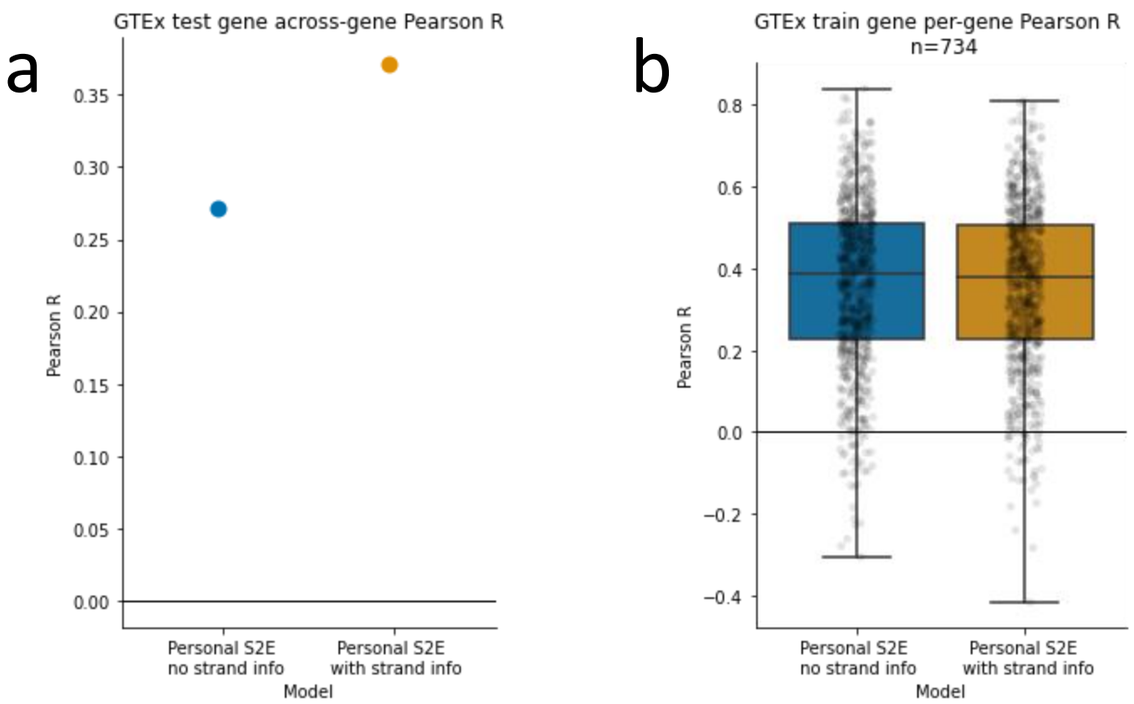


Figure S10. Effect of using gene strand information. Genes are from the top 1000 gene set. For “with strand info”, we take the reverse complement of the sequence for genes on the negative strand, while for “no strand info” we do not.

Figure S11

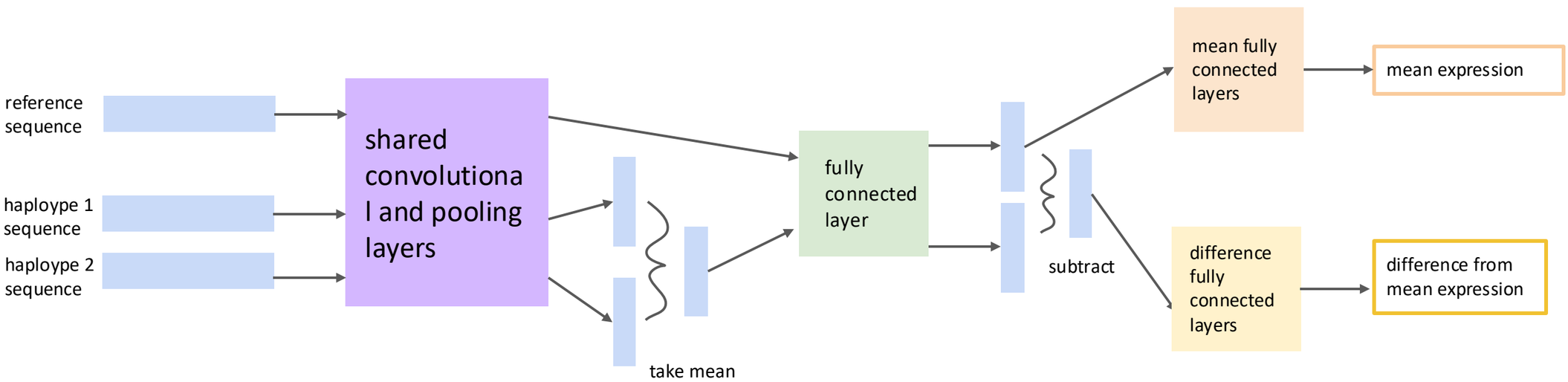


Figure S11. p-SAGE-net model architecture. Model input is reference sequence for a given gene as well as an individual's two haplotypes for that gene. These three sequences pass through the same shared convolutional layers, after which the two haplotypes are averaged. The averaged personal tensor and the reference tensor then pass through the same fully connected layer and different output heads to produce mean expression and difference from mean expression. Mean expression is predicted from reference sequence alone, while difference from mean expression is predicted using all three sequences, specifically by subtracting the "reference" tensor intermediate output from the "personal". See Methods for model layer specifics.

Figure S12

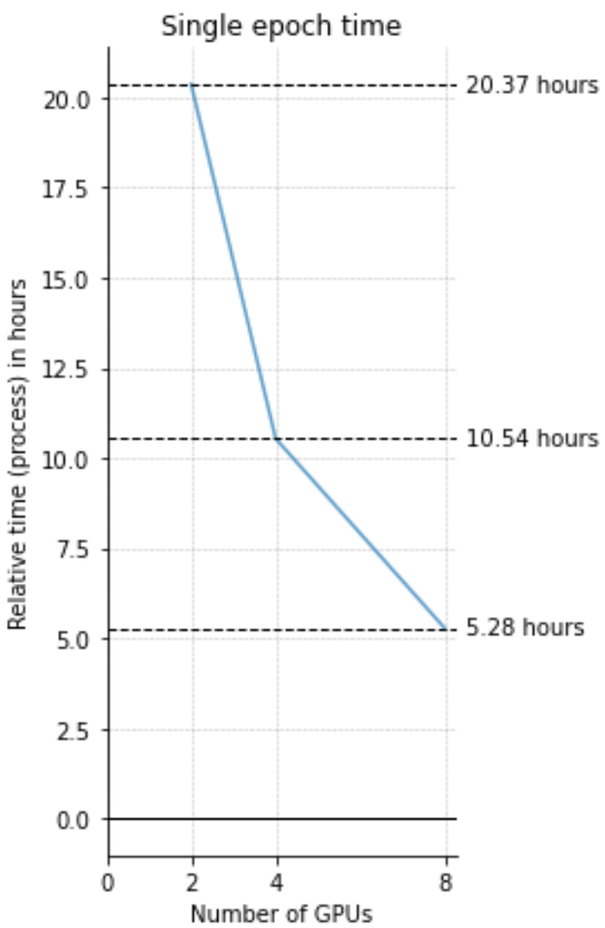


Figure S12. p-SAGE-net training time. P-SAGE-net single epoch training time, shown for training parallelized over 2, 4, 8 NVIDIA A40 GPUs. For one epoch, the model is trained on 689 individuals x 741 genes, evaluated on 85 individuals x 114 genes.

Figure S13

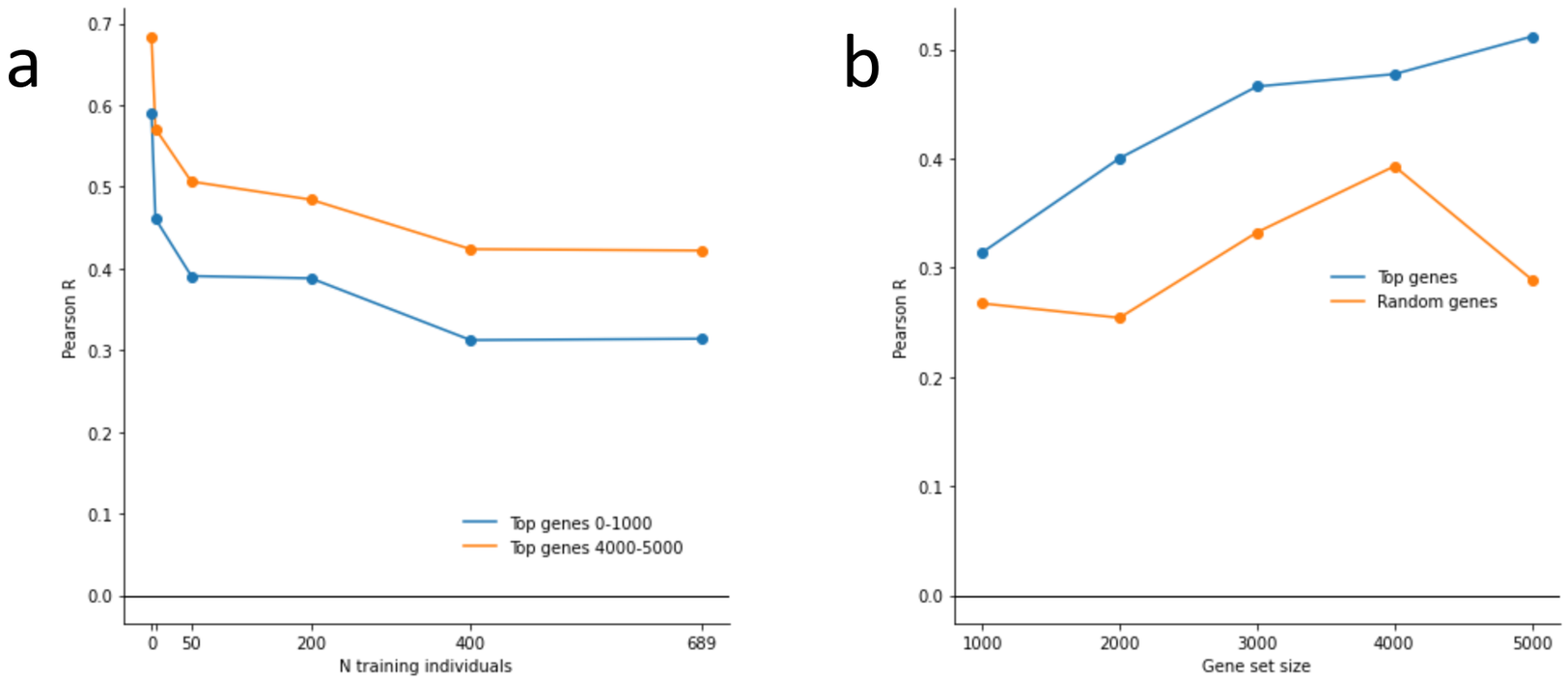


Figure S13. Unseen gene across-gene performance across number of training individuals, gene set size. (a) Same analysis as in Fig. 2c, but shown for unseen genes, across-gene correlation. (b) Same analysis as in Fig. 2d, but shown for unseen genes, across-gene correlation.