

Figure **S1**. Principal component analysis plot of the first two principal components based on STR profiles from normal samplels per cancer type (from top to bottom: CRC, STAD, UCEC). The popultaion of each patients is indicated by colores in the bottom right legend.

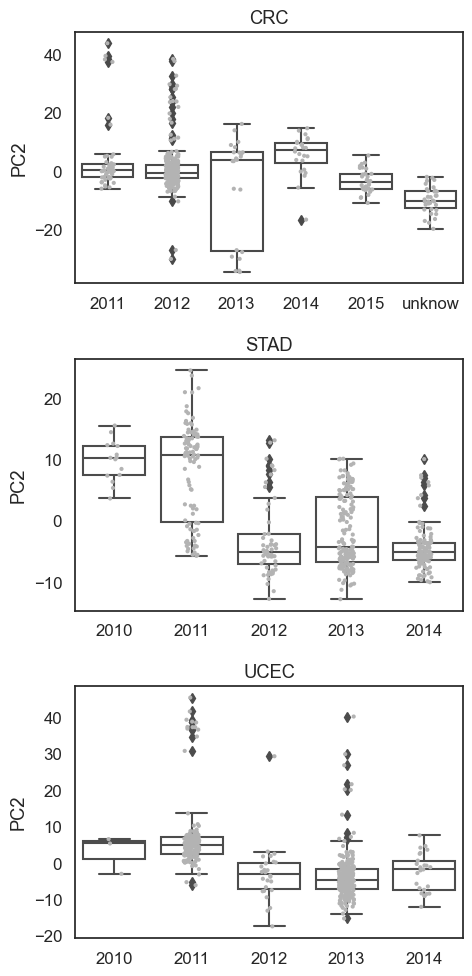


Figure **S2**. Boxplots showing the variations in the second prinicipal component (PC2) acorss samples from different sequencing years. PC2s values were derived from the PCA of STR profiles from normal samples. Sequencing year accounts for a significant portion of the variance, likely due to batch effects introduced by differences in plates, capture kits, and other technical factors.

A graph of different colored objects

Description automatically generated with medium confidence

Figure **S3**. Boxplots showing STR deletion and insertion percentage at STRs in MSI and MSS tumors per cancer type. Significant differences are indicated by asterisks, non-significant differences by n.s.

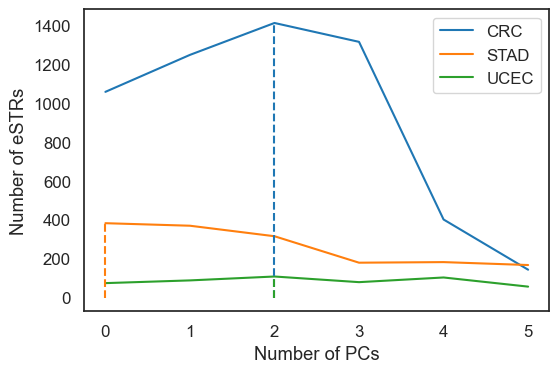


Figure **S4**. Effect of varying numbers of PCs on power to detect eSTRs. eSTRs (FDR < 0.05) were computed for each cancer type after adjusting for a number of PCs ranging from 0 to 5. The x-axis shows the number of PCs adjusted for. The y-axis shows the number of significant eSTRs (see methos). Dashed vertical lines show the number of PCs with most eSTRs detected for each cancer type.

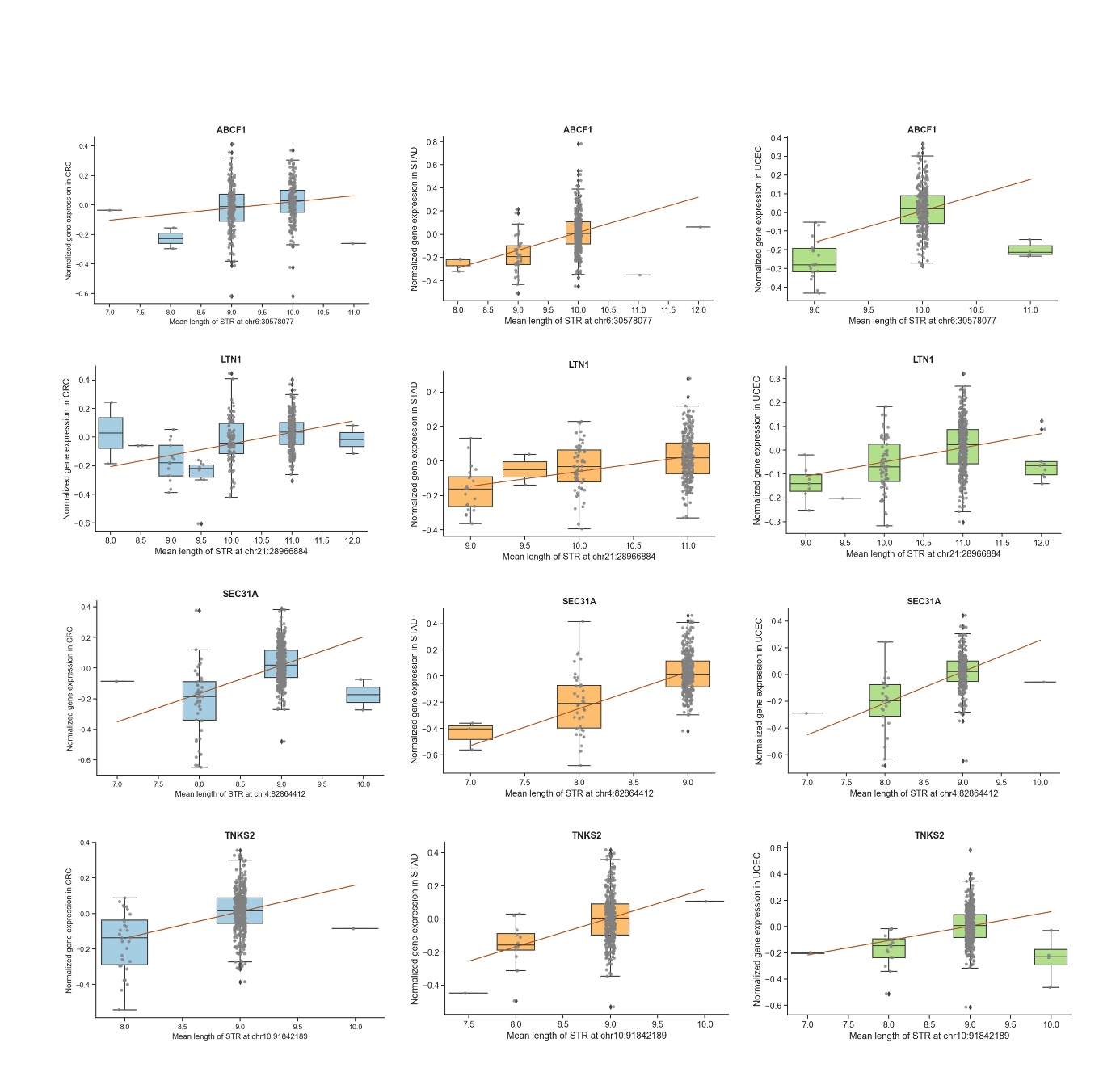


Figure **S5**. Scatterplots showing the associations between Coding-region eSTR length and corresponding gene expression levels (ABCF1, LTN1, SEC31A, TNKS2)

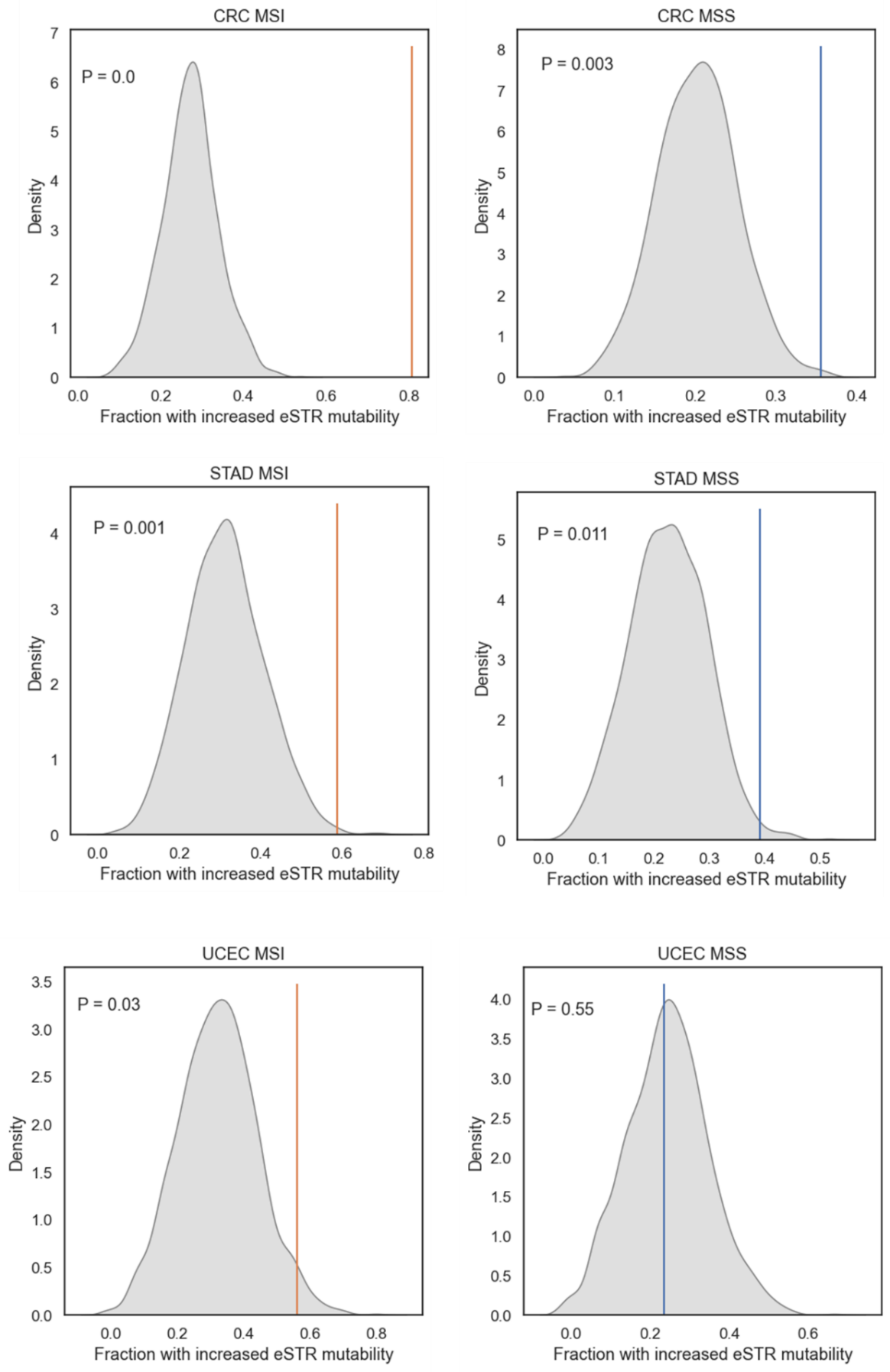


Figure **S6**. Kernel density plots showing the null distributions from permuation tests. For each permutation, the fraction of repeat types for which the eSTRs were more mutable was determined. Vertial colored stripes represent the observed fraction of repeat types where eSTRs were more mutable. P-values obtained from permutation tests are shown in the top left.

A graph of a normal distribution

Description automatically generated

Figure **S7**. Kernel density plots showing the null distributions from permuation tests. The null distributions were generated from permutations of the eSTR effect sizes and the accuracies of correct gene expression direction change were determined. Vertial colored stripes represent the observed accuracy of prediction the direction of gene expression change. P-values obtained from permutation tests are shown in the top left.

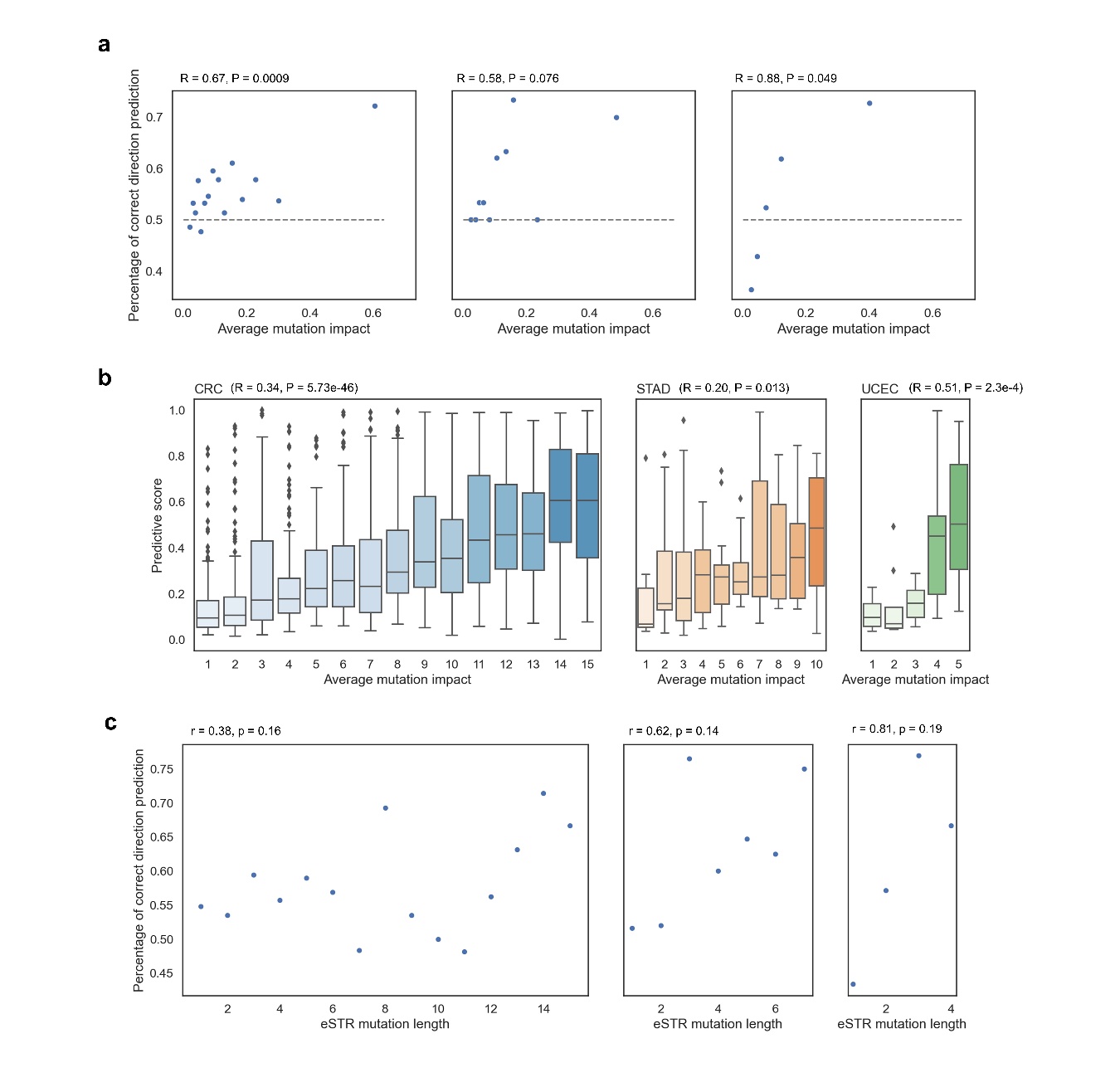


Figure **S8.** (a) Scatterplots showing the relationship between the average mutation impact and the observed accuracy of predicting the direction of gene expression changes for each eSTR mutation group.The x-axis shows the average mutation impact for each group, while the y-axis represents the corresponding accuracy. (b) Boxplots showing the distribution of predictive scores across eSTR mutation groups with increasing average mutation impact. Black points represent single eSTR mutation. Each box corresponds to the same eSTR mutation group in figure (a). The x-axis represents the ranked number of eSTR mutation groups. c) Scatterplots showing the relationship between eSTR mutation length and the observed accuracy of predicting the direction of gene expression changes. Pearson correlation coefficients and associated p-values are shown in the top left of each plot.

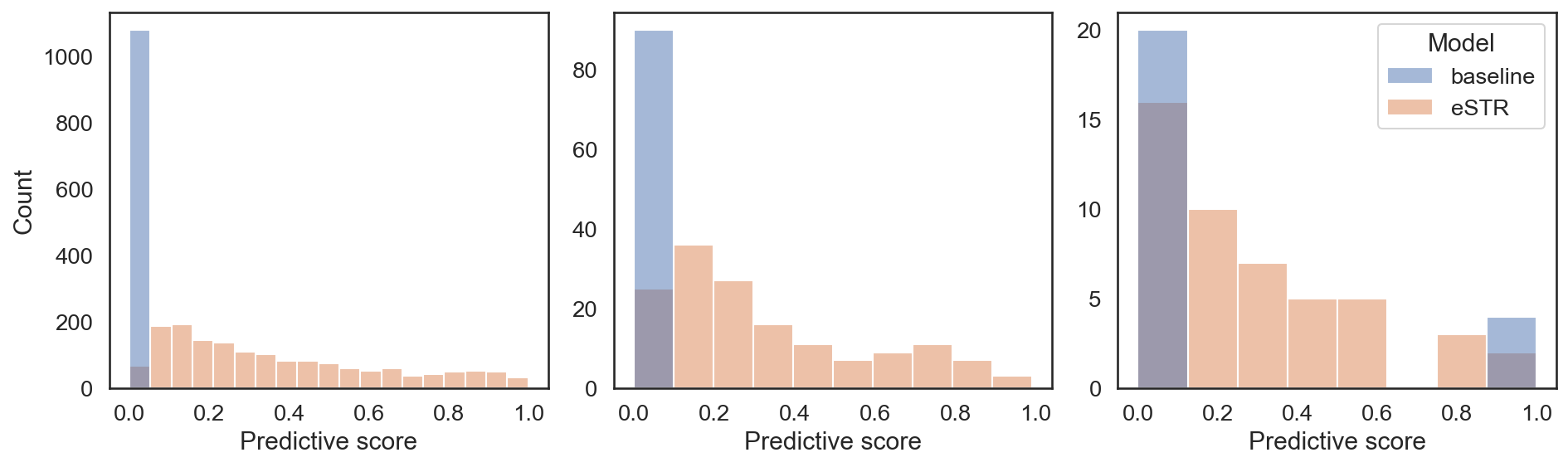


Figure **S9**. The distribution and comparsion of predictice scores between eSTR linear regression model and baseline model (regression model with coefficient zero)