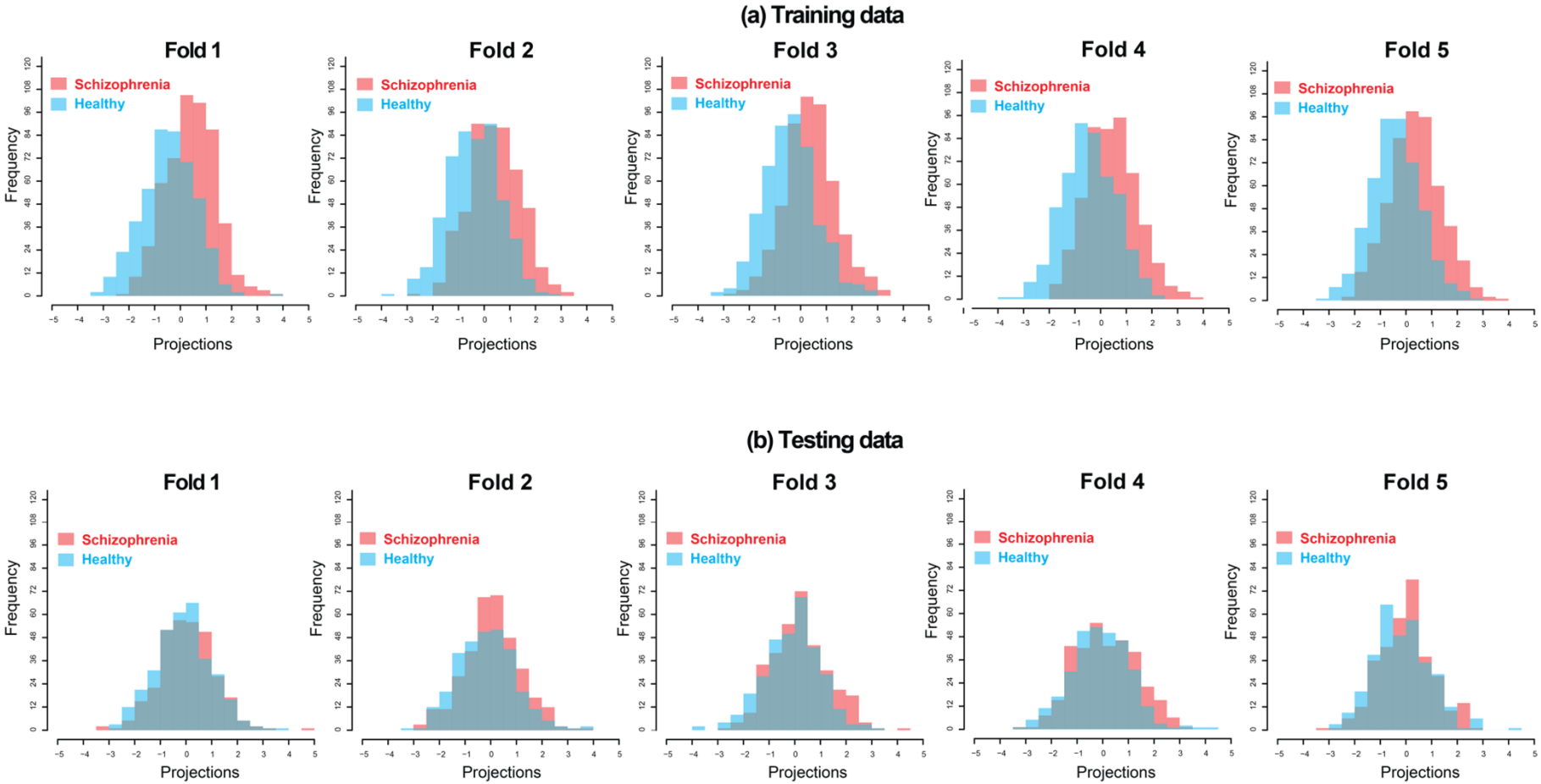
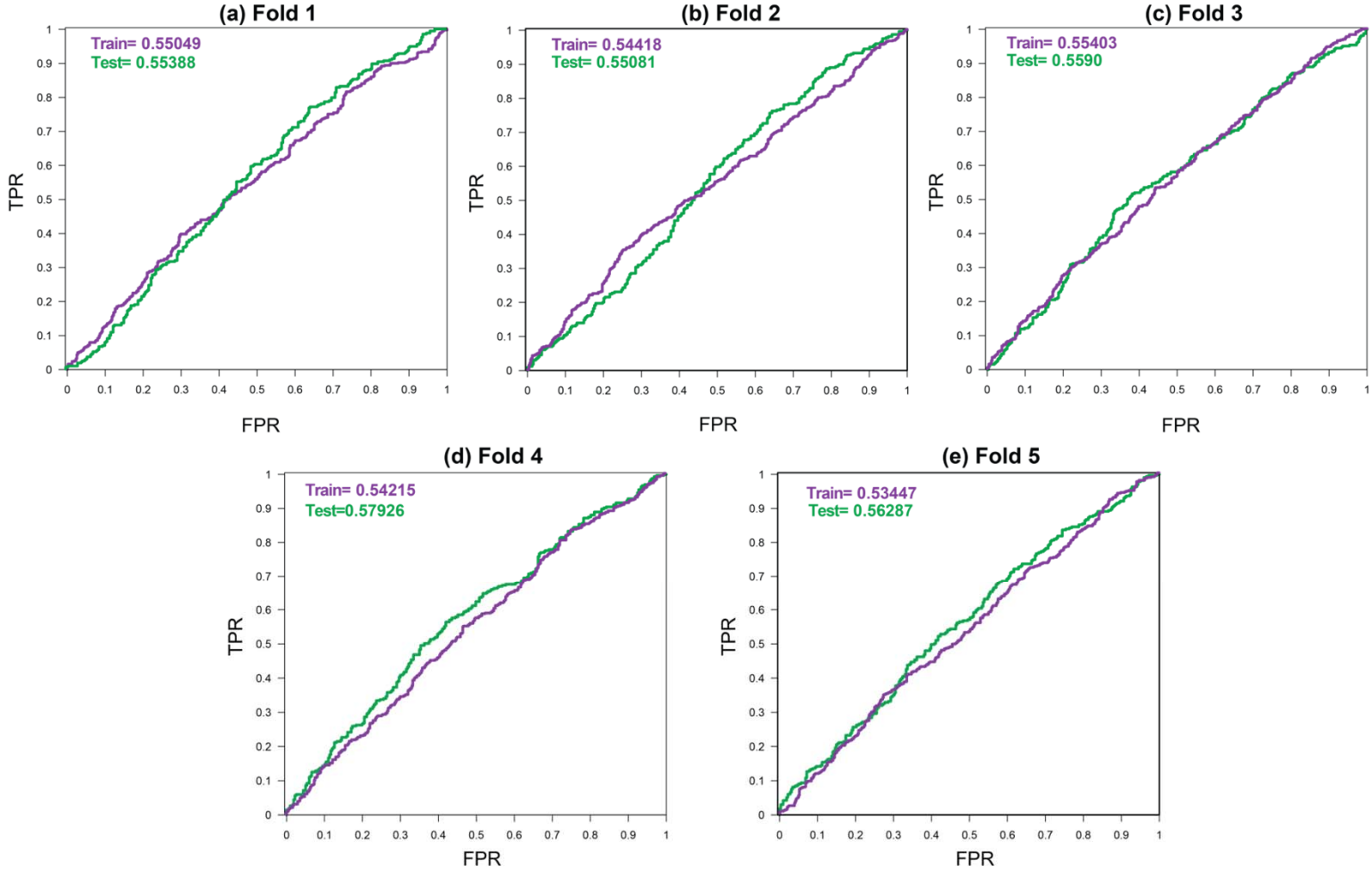
Ehrenreich et al Supplementary Figure 1



Classification of schizophrenia and healthy subjects using a linear discriminant that incorporates homo- and heterozygous risk allele information, a-b.
Histograms over the leading direction of discrimination for 5 cross-validation folds, classifying schizophrenia (red) and healthy (blue) individuals based on their risk allele (SNP) information, using linear discriminant analysis [Fisher, 1936]. Homozygous and heterozygous risk allele information has been used to train the model. For each fold (1-5), histograms for training (**a**) and testing (**b**) data are shown. As illustrated, class separation is minimal between schizophrenia and healthy groups in each of the testing folds, indicating that disease allele-based classification with a linear projection procedure does not have enough representational power to separate the data. Non-separability quantification with McNemar test (null hypothesis that the classification of the linear discriminant is equal to that of a random classifier) yielded in every fold p-values >0.39.

Ehrenreich et al Supplementary Figure 2



Classification of schizophrenia and healthy subjects using a random forest machine learner that incorporates homo- and heterozygous risk allele information, **a-e**.
Receiver operating characteristic (ROC) analysis for 5 cross-validation folds classifying schizophrenia and healthy individuals based on their risk allele (SNP) information using a random forest machine learner [Breiman, 2001]. Homozygous and heterozygous risk allele information has been used to train the model. Random splits of training (purple) and testing (green) sets were created to measure the ability of the classifier to separate *unseen* test observations. The area under the curve (AUC) for training and testing sets is always close to 0.5 (random classification), indicating that disease allele-based class separation is not possible using a highly non-linear method such as the random forest. Non-separability quantification with McNemar test (null hypothesis that the classification of the random forest is equal to that of a random classifier) yielded in every fold p-values >0.25.