



FIGURE 5 | Feature selection.(A,B) Venn diagram of DEGs in GSE179265(NP1), GSE136825 (NP2) and GSE98895 (MS). (C) The effect of the number of decision trees on the error rate. x-axis represents the number of decision trees and y-axis represents the error rate. The error rate is relatively stable when the number of decision trees is approximately 400. (D) Results of the Gini coefficient in the random forest classifier. y-axis represents the genetic variables and x-axis represents the importance index. (E) The coefficients of CRSwNP and MS common DEGs with predictive value are shown by the lambda parameter. (F) Partial likelihood deviation with log(lambda) derived by LASSO algorithm and tenfold cross-validation. (G,H) Signature genes are expressed in the external dataset GSE194282, and in NP1, NP2. (I) AUC and AUPR plots of diagnostic efficacy scores for the XGBoost model based on feature genes for the nasal polyp training set, validation set and MS dataset.