RESEARCH ARTICLE

Figures and figure supplements

Circulating small extracellular vesicle RNA profiling for the detection of T1a stage colorectal cancer and precancerous advanced adenoma

Li Min *et al*.

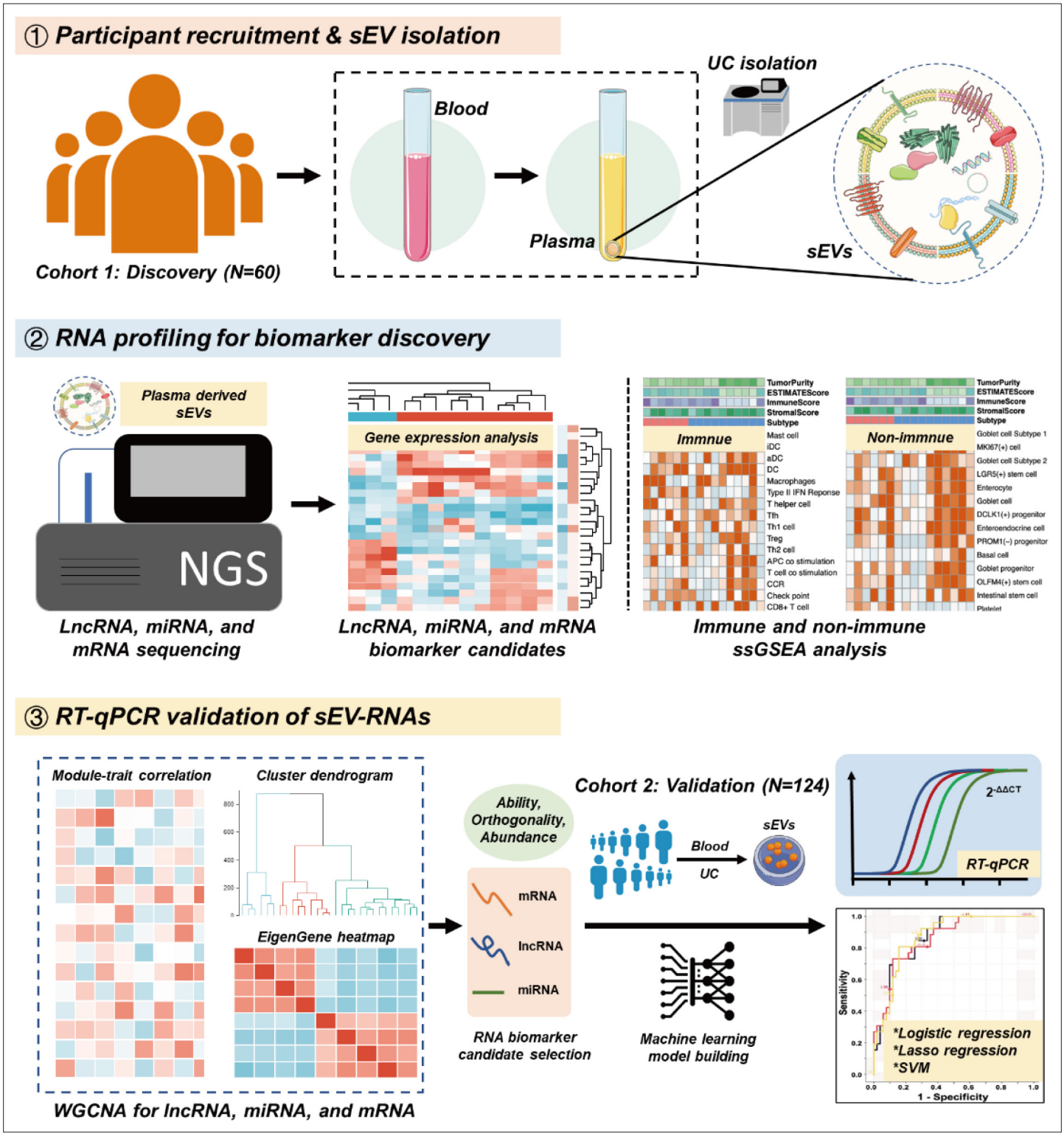

Figure 1. Schematic overview of the study design.

Figure 2. Transcriptome profiling of circulating sEVs. (**a**) TEM images of circulating sEVs isolated from human plasma. (**b**) NTA results of circulating sEVs enriched from plasma. (**c**) WB results of sEV positive (Alix, TSG101, CD9) and negative (Calnexin) markers. (**d**) The numbers of detected RNA species in different groups. (**e**) The hierarchical clustering results of top 100 miRNAs (left panel), mRNAs (middle panel), and lncRNAs (right panel). (**f**) t-SNE clustering by those candidate RNAs. (**g**) A Venn diagram showed DEGs shared between different comparisons (CRC vs NC, AA vs NC, CRC vs AA).

Figure 2 continued on next page

Figure 2 continued

(**h**) KEGG enrichment of all those DEGs identified. (**i–k**) potential core regulatory networks between miRNAs and mRNAs in DEGs identified in three (**i**), two (**j**), and one (**k**) of all comparisons.

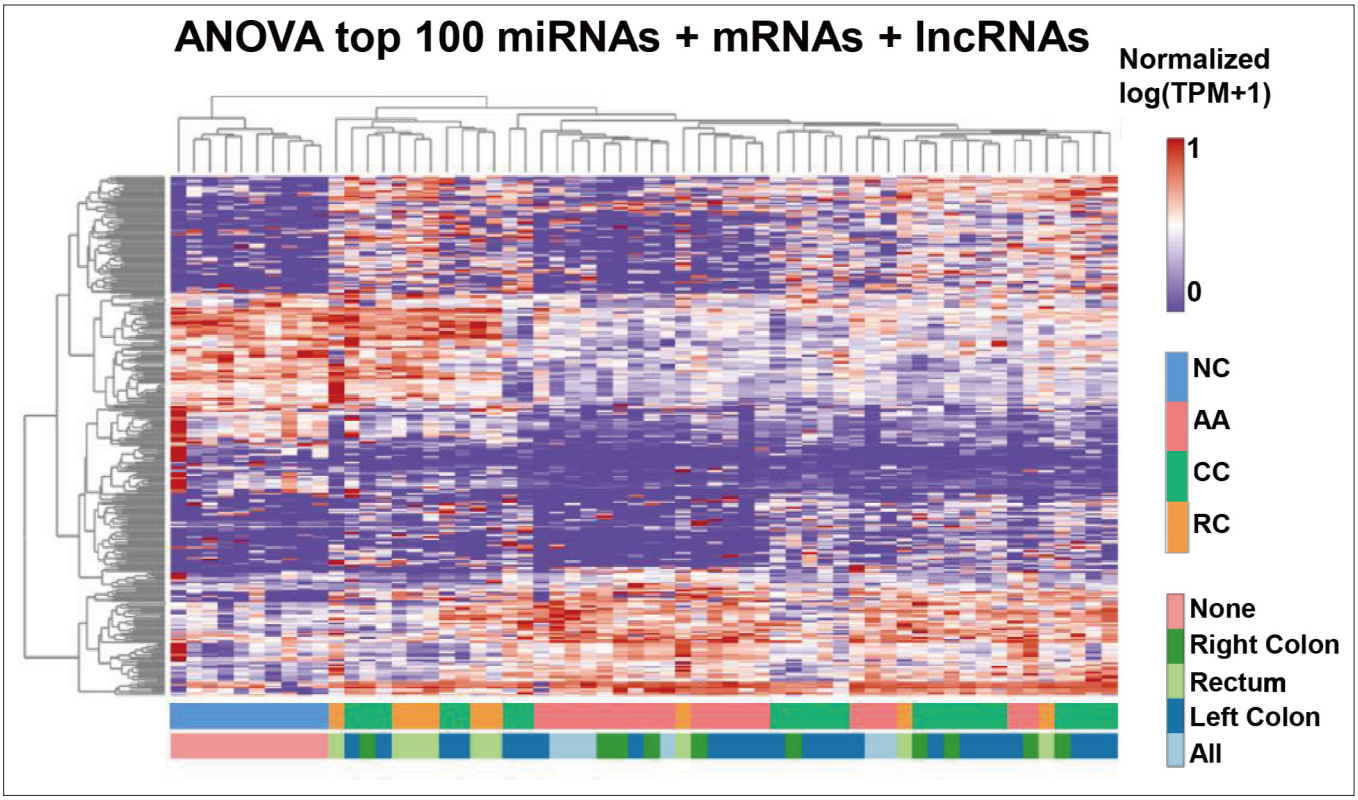

Figure 2—figure supplement 1. The hierarchical clustering results of Top 100 miRNAs/mRNAs/lncRNAs.

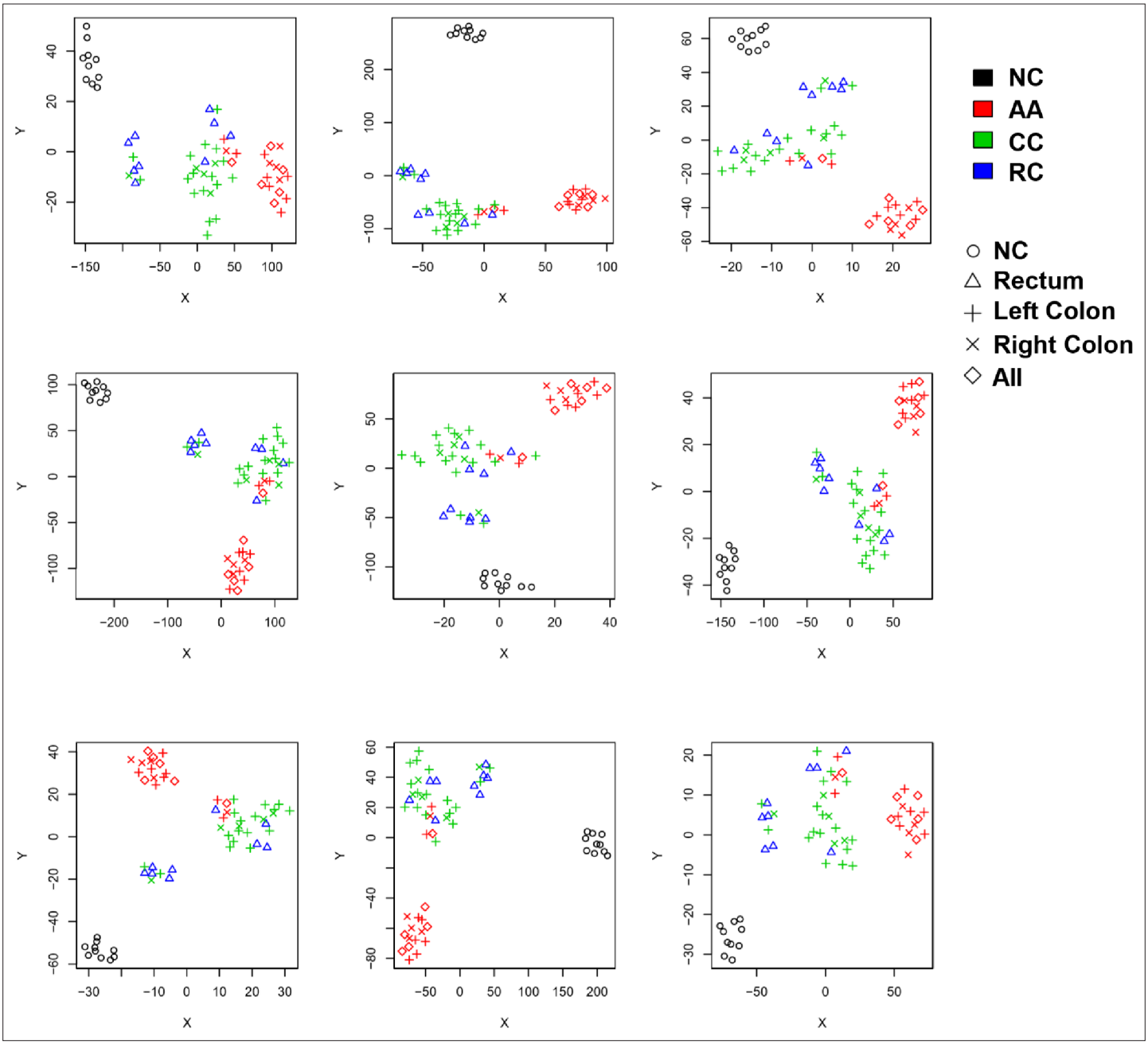

Figure 2—figure supplement 2. Unsupervised t-SNE clustering by those 200 RNAs with nine repeats.

Figure 3. Cell-specific features of the sEV-RNA profile. (**a**) The hierarchical clustering heatmap of immune cell-specific features of each sample. (**b**) The hierarchical clustering heatmap of stromal-related features of each sample. (**c**) Boxplot of cell-specific features overexpressed in CC and RC patients (*p<0.05, **p<0.01, ***p<0.001). (**d**) Boxplot of cell-specific features overexpressed in NC participants (*p<0.05, **p<0.01, ***p<0.001). (**e**) The violinplot of the microenvironmental scores in different subgroups (*p<0.05, **p<0.01, ***p<0.001).(f) Correlation among cell-specific features differentially enriched among different groups.

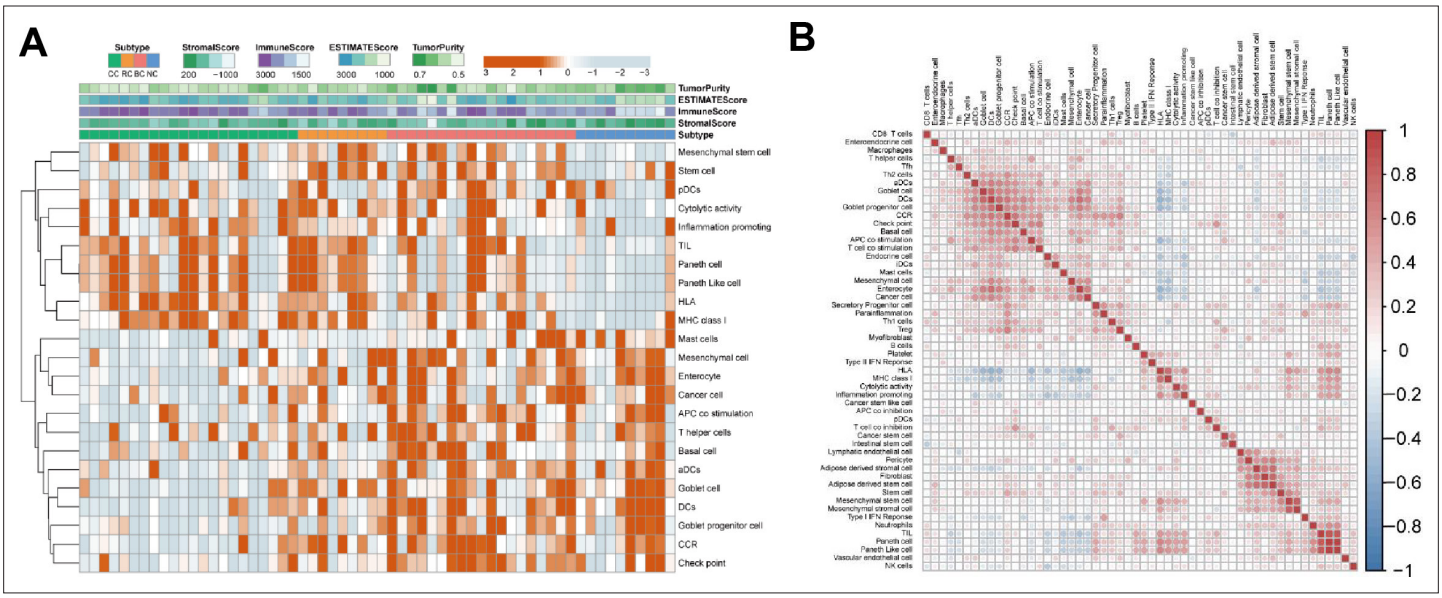

Figure 3—figure supplement 1. Cell-specific features of the sEV-RNA profile. (**A**) The hierarchical clustering heatmap of different cell features in all sEV samples. (**B**) Correlation among all cell-specific features.

Figure 4. WGCNA analysis of sEV-RNAs. (**a**) Gene coexpression module construction of all DEGs identified in sEV-RNAs. (**b**) The heatmap exhibited Pearson correlations among different modules. (**c**) Bar plot of module composition of different modules (all DEGs). (**d**) Percentage bar plot of the RNA composition of different modules (all DEGs). (**e**) A heatmap exhibited the expression levels of the top 10 DEGs in each module. (**f**) t-SNE clustering by the top 10 DEGs in each module. (**g**) t-SNE clustering by the top 5 DEGs in each module. (**h**) t-SNE clustering by the top1 DEGs in each module.

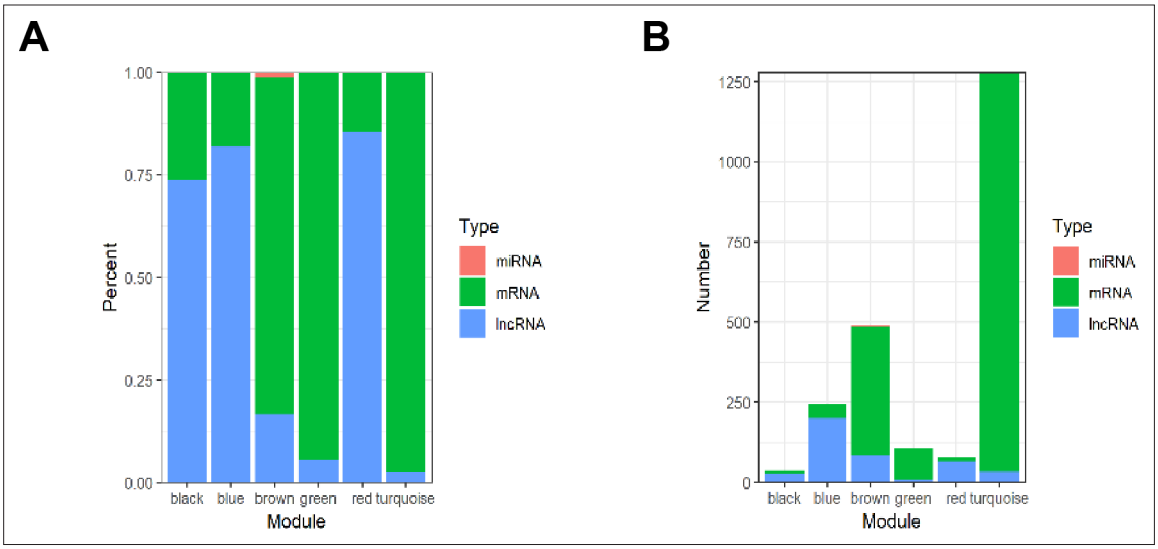

Figure 4—figure supplement 1. Proportions and numbers of RNA species in different modules. (**A**) Percentage barplot of the RNA composition of different modules (only DEGs with kME >0.7). (**B**) Barplot of module composition of different modules (only DEGs with kME >0.7).

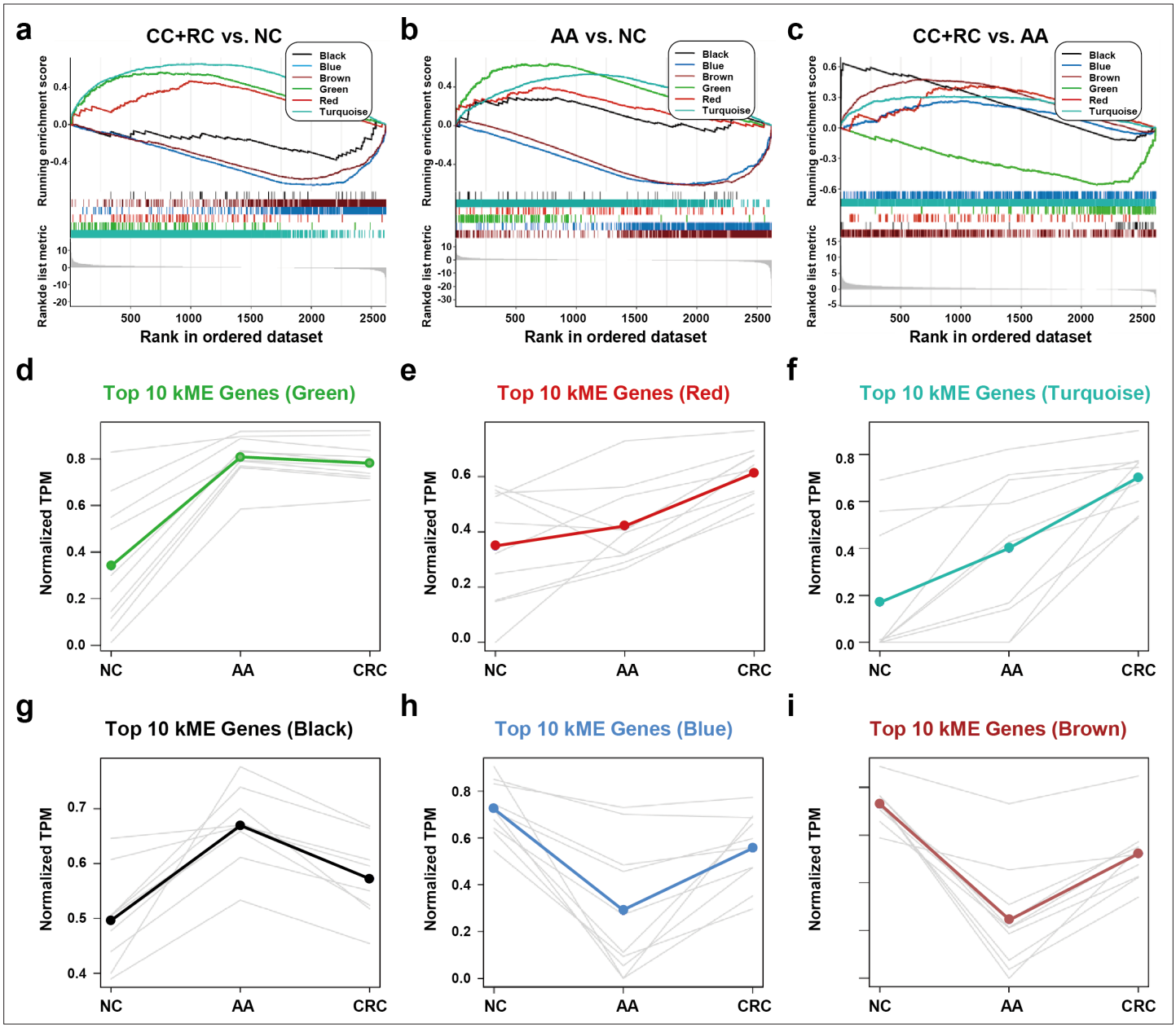

Figure 5. The expression trends of sEV-RNA modules. (**a–c**) GSEA analysis of DEGs in different modules (a: CRC vs. NC; b: AA vs. NC; c: CRC vs. AA). (**d-i**) The expression trends of the Top 10 DEGs of each module among NC, AA, and CRC (d: green module; e: red module; f: turquoise module; g: black module; h: blue module; i: brown module).

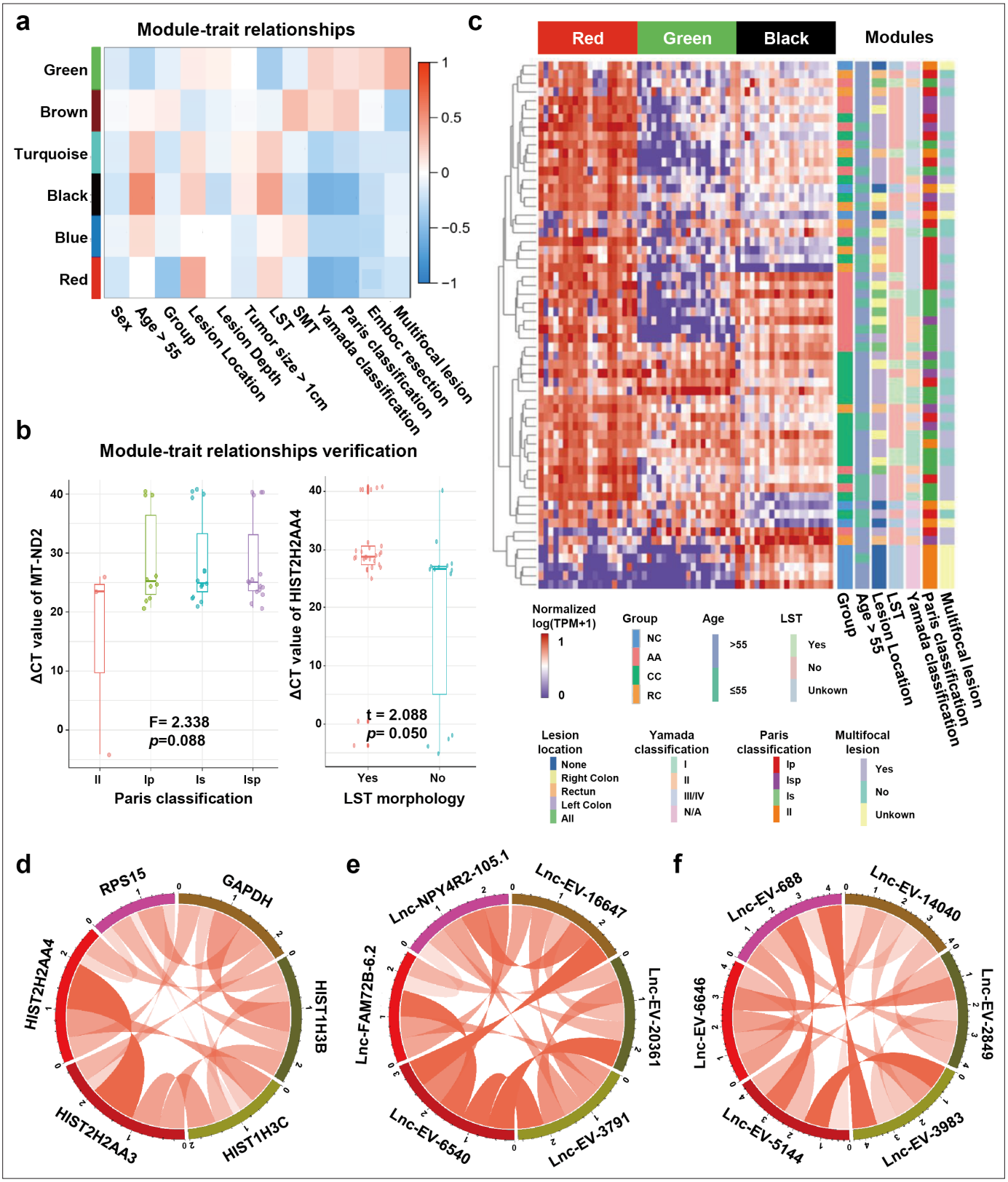

Figure 6. Module-trait correlation analysis of sEV-RNA modules. (**a**) The heatmap exhibited the correlation between modules and clinical traits. (**b**) The RT-qPCR validation of representive module-trait correlation (left panel: correlation between MT-ND2 and Paris classification; right panel: correlation between HIST2H2AA4 and LST morphology). (**c**) The heatmap exhibited the sEV-RNA expression levels of red, black, and green modules. (**d-f**) Circos plot showed the inner correlations among sEV-RNAs in the module green (**d**), red (**e**), and black (**f**).

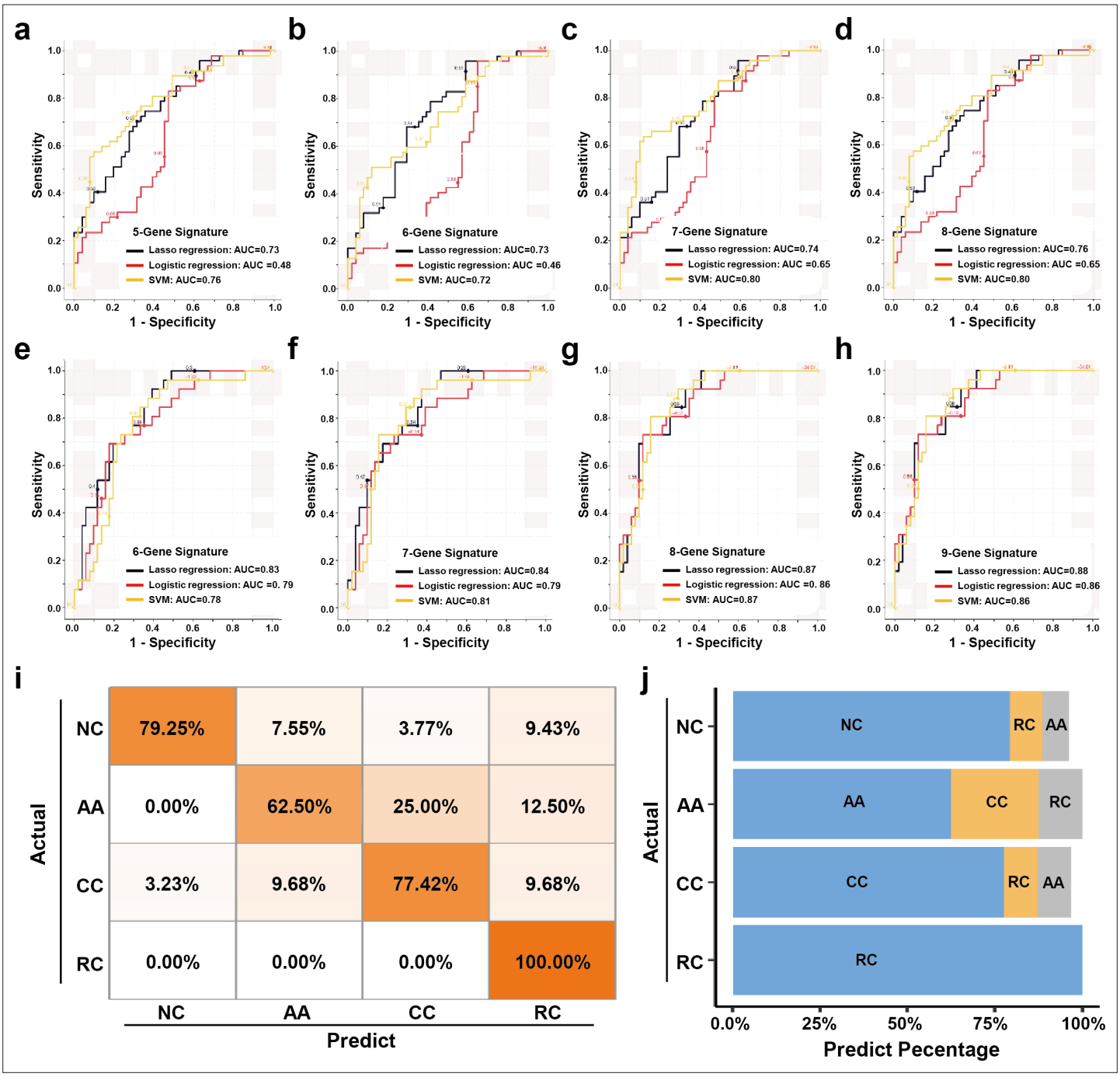

Figure 7. The plasma sEVs-RNA signature to detect early CRC and AA. (**a–d**) The ROC analysis of different sEV-RNA signatures in the prediction of CRC patients by different algorithms (a: 5-gene panel; b: 6-gene panel; c: 7-gene panel; d: 8-gene panel). (**e–h**) The ROC analysis of different sEV-RNA signatures in the prediction of AA patients by different algorithms (e: 6-gene panel; f: 7-gene panel; g: 8-gene panel; h: 9-gene panel). (**i**) The QDA results of all 13 sEV-RNAs in classifying all samples. (**j**) Statistical summary of QDA performance in each sample group.

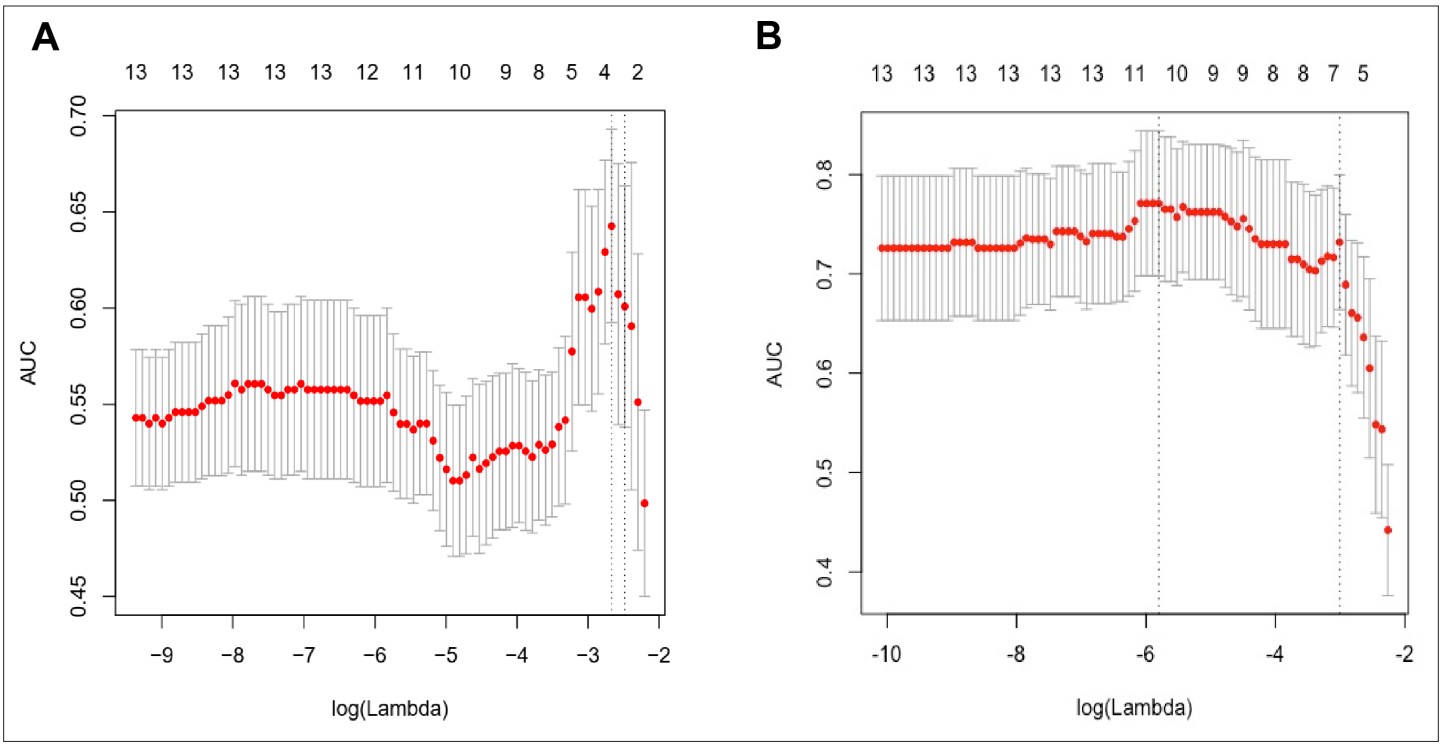

Figure 7—figure supplement 1. Lasso regression to construct multivariate prediction models. (**A**) Performance of Lasso regression in variable selection to identify CRC. (**B**) Performance of Lasso regression in variable selection to identify AA.

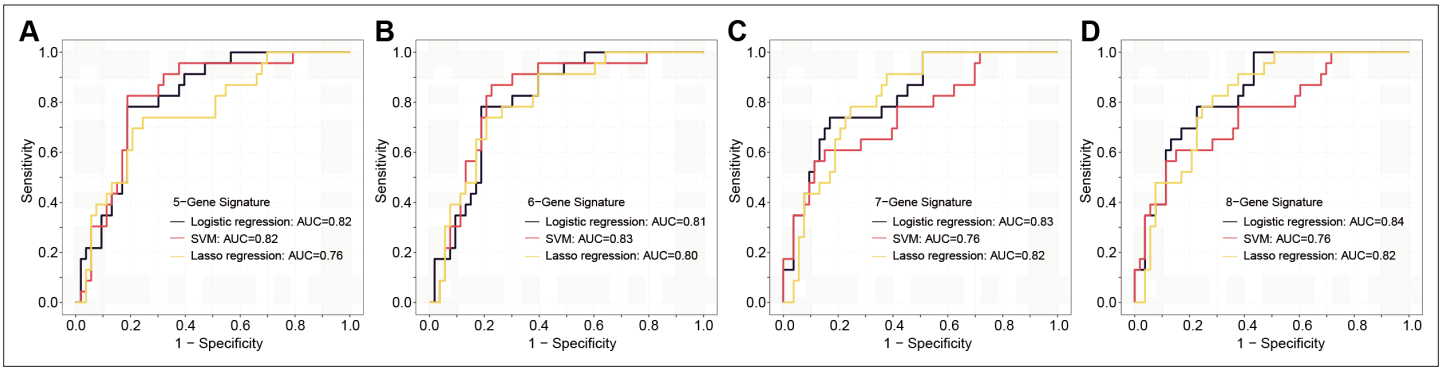

Figure 7—figure supplement 2. The ROC analysis of different sEV-RNA signatures in the prediction of stage I CRC patients by different algorithms (**a**: 6-gene panel; **b**: 7-gene panel; **c**: 8-gene panel; **d**: 9-gene panel).

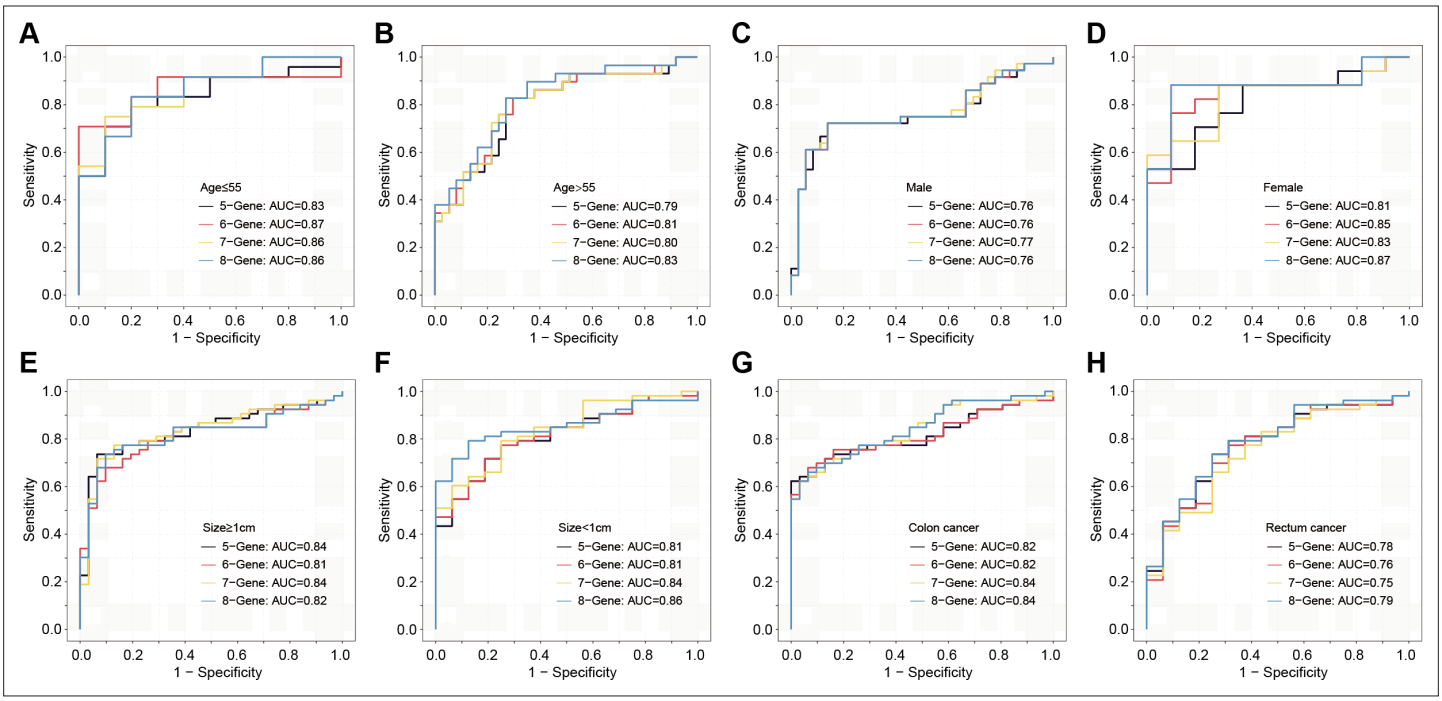

Figure 7—figure supplement 3. The ROC analysis of different sEV-RNA signatures for predicting CRC patients using the Lasso regression algorithm in different clinical parameters (ab: age; cd: gender; ef: tumor size; gh: anatomical position).