Supplemental Figures

Construction of Exosome Non-coding RNA Feature for Non-invasive, Early Detection of Gastric Cancer Patients by Machine Learning: A Multi-Cohort Study

Ze-Rong Cai, Yong-Qiang Zheng, Yan Hu, Meng-Yao Ma, Yi-Jin Wu, Jia-Liu, Lu-Ping Yang, Jia-Bo Zheng, Tian Tian, Pei-shan Hu, Ze-xian Liu, Lin Zhang, Rui-Hua Xu, Huai-Qiang Ju

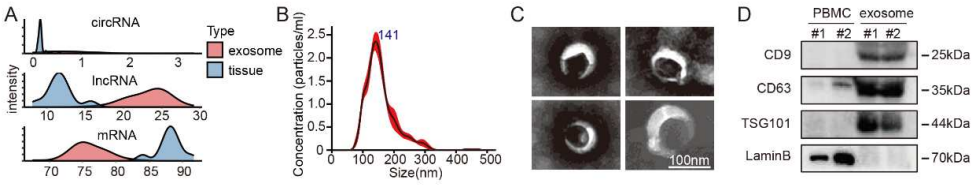

Figure S1. Characteristics of exosomes. (A) Proportions of different ncRNA classes relative to the total ncRNAs in serum exosomes or tissues. (B) Size distribution of serum exosomes measured by NanoSight NS300. (C) TEM images of purified serum exosomes. Scale bar: 100 nm. (D) Western blot analysis showing enrichment of markers (CD9, CD63, and TSG101) in serum exosomes.

Abbreviations: ncRNA: non-coding RNA; TEM: transmission electron microscopy; CD9: Cluster of Differentiation 9; CD63: Cluster of Differentiation 63; TSG101: Tumor Susceptibility Gene 101.

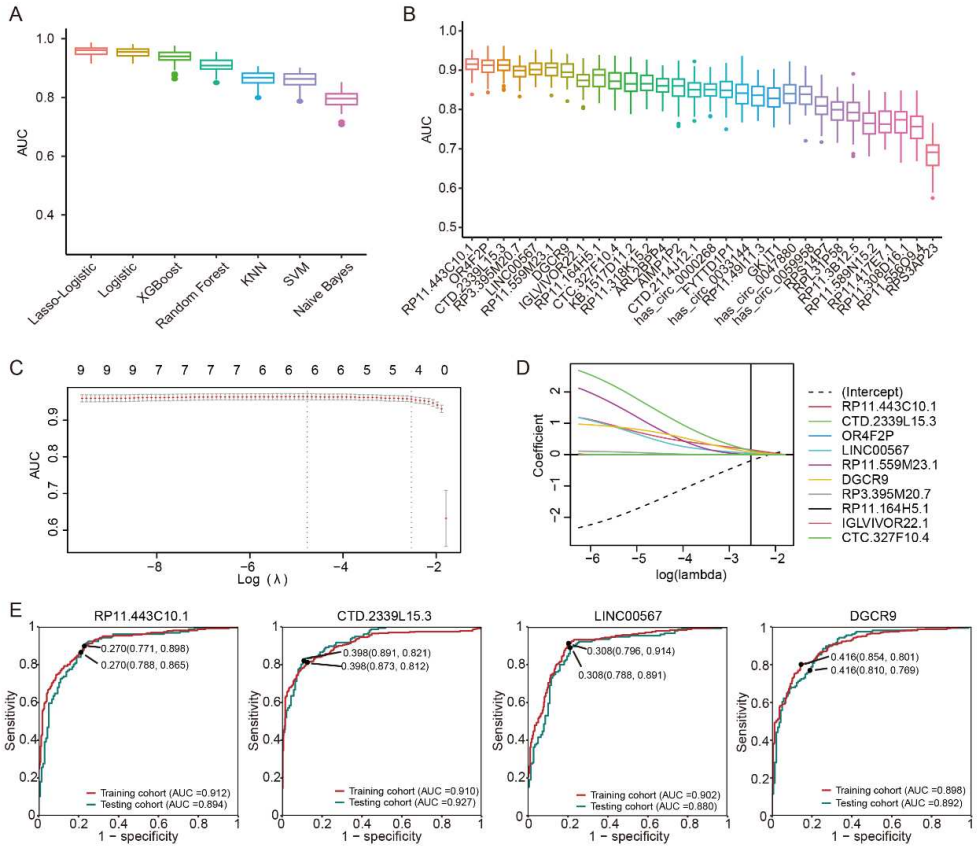

Figure S2. The process of model construction. (A) Application of 100-fold cross validation to compare the performances of seven common machine learning models. (B) Application of 100-fold cross validation to compare the performances of single ncRNA in GC diagnosis. (C) Selection of tuning parameter (lambda) in the LASSO regression using ten-fold cross validation. The two vertical lines were drawn at the optimal scores by minimal criteria and 1-s.e criteria. (D) The LASSO coefficients of 10 ncRNAs during the process of model construction. The vertical line was drawn at the optimal scores by 1-s.e criteria. (E) ROC curves for the four selected ncRNAs in distinguishing gastric cancer patients from healthy donors.

Abbreviations: AUC, area under curve; LASSO, least absolute shrinkage and selection operator; XGBoost, extreme gradient boosting; KNN, K nearest neighbours; SVM, support vector machine; ncRNA, non-coding RNA; GC, gastric cancer; ROC: receiver operating characteristic; DGCR9, DiGeorge syndrome critical region gene 9.

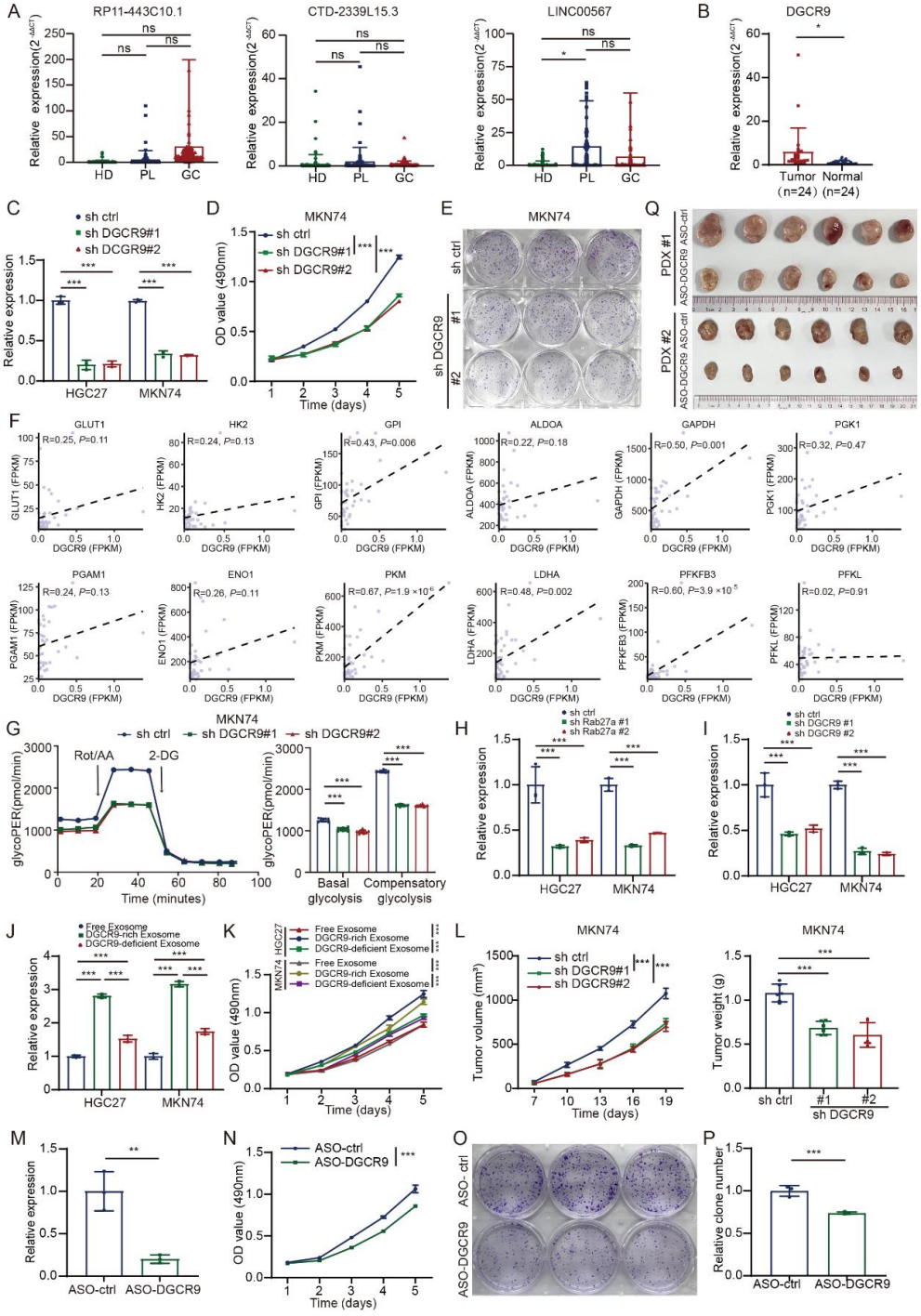

Figure S3. Analysis of DGCR9 in gastric cancer cells. (A) Expression levels of RP11.443C10.1, CTD-2339L15.3 and LINC00567 in serum exosomes from GC patients (n=112), those with precancerous lesions (n=73), and healthy donors (n=100). (B) Expression levels of DGCR9 in GC tissues (n=24) compared to adjacent normal

tissues (n=24). Both tissues were gathered from patients with gastric cancer during surgery. (C) Expression levels of DGCR9 in DGCR9-knockdown HGC27 cells compared to control cells. (D) MTS assay assessing the effect of DGCR9 knockdown on MKN74 cell growth. (E) Clonogenic assay showing colony formation in DGCR9-knockdown MKN74 cells compared to control cells. (F) Correlation analyses of DGCR9 and glycolysis-related genes in our transcriptome data of tissues from 20 patients with gastric cancer. (G) Statistical analysis of the effects of DGCR9 knockdown on glycolytic activity of MKN74 cells. (H) The relative expression of DGCR9 in exosomes from culture media of Rab27a-depleted GC cells; (I) The relative expression of exosome-packaged DGCR9 in GC cells with or without knockdown of DGCR9. (J) The relative expression of DGCR9 in wild-type GC cells after treated with DGCR9-rich or DGCR9-deficient exosomes. (K) The growth curves of wild-type GC cells after treated with DGCR9-rich or DGCR9-deficient exosomes. (L) Growth curves and tumor weights of DGCR9-knockdown MKN74 cells versus control cells in subcutaneous tissues of SCID mice (NCG mice). (M) Expression levels of DGCR9 in HGC27 cells treated with scrambled or DGCR9 inhibitor. (N) MTS assay assessing the growth of HGC27 cells without or with treated by antisense oligonucleotide targeting DGCR9. (O) Clonogenic assay showing colony formation in ASO-DGCR9 HGC27 cells compared to control cells. (P) Statistical analysis of the clonogenic assay. (Q) The image of the PDX model with treated by DGCR9 inhibitor optimized in vivo or control. The numbers of biological replicates were three (C-E, H-K, M-P), four (G) and six (L, Q), respectively. Data in A-D, G-N, P were presented as mean ±S.D. *P* values were determined by one-way ANOVA (A, C, G-J, tumor volumes in L), one-tailed paired Student's t test (B), two-way ANOVA (D, K, tumor volume in L, N) and two-tailed unpaired Student's t test (M, P).

Abbreviations: HD, healthy donor; PL, precancerous lesion; GC, gastric cancer; DGCR9: DiGeorge Critical Region 9; FPKM, fragments per kilobase of transcript per million fragments mapped; GLUT, glucose transporter; HK2, hexokinase 2; GPI, glucose-6-phosphate isomerase; ALDOA, aldolase A; GAPDH, glyceraldehyde 3-

phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGAM, phosphoglycerate mutase; ENO, enolase; PKM, pyruvate kinase M; LDHA, Lactate dehydrogenase A; PFKFB, 6-phosphofructose kinase/fructose-2, 6-bisphosphatase; PFKL, phosphofructokinase liver type; GlycoPER, Glycolytic Proton Efflux Rate; SCID, severe combined immunodeficiency; PDX, patient-derived tumor xenograft; ASO, antisense oligonucleotide; ns, non-significant; *, P <0.05; **, P < 0.01; ***, P < 0.001.