Figure S1 Expression correlation analysis of CRC DEGs (i.e., *CDKN2A*, *ETS2*, *CXCL1*, *SNAI1*, and *SERPINE1*). TPM, transcripts per million; CRC, colorectal cancer; DEGs, differentially expressed genes.
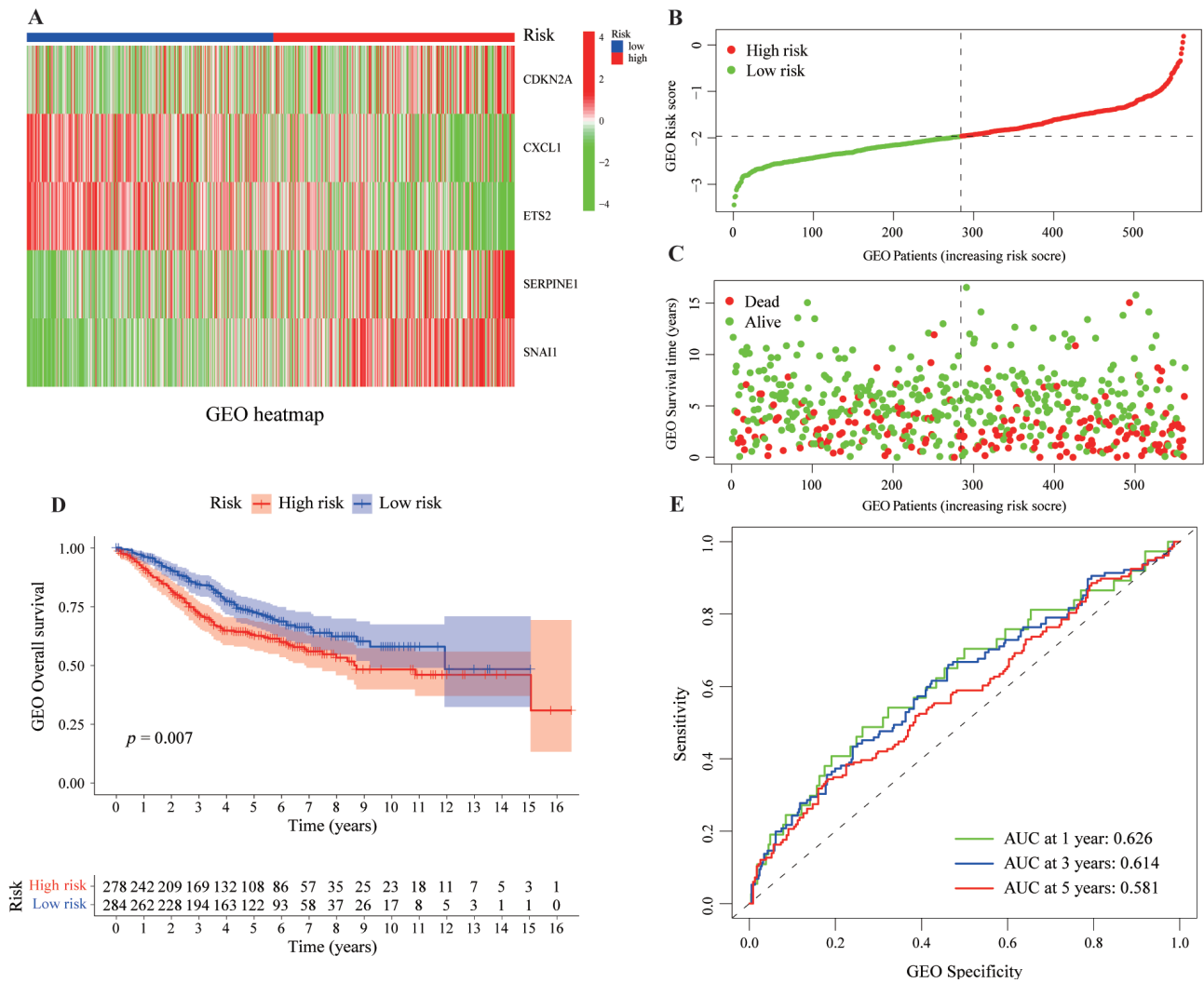
Figure S2 Prognostic analysis of risk-related genes. (A) Heatmap of the distribution of the CRC-COAD prognostic survival genes (i.e., *CDKN2A*, *ETS2*, *CXCL1*, *SNAI1*, and *SERPINE1*) in GEO. The red bars represent the high gene expression, the deeper the color, the higher the expression; the green bars represent the low gene expression, the deeper the color, the lower the expression. (B) The risk score distributions, with 278 patients at high risk and 284 patients at low risk, in the GEO cohort. (C) The correlation between survival time and risk score in the GEO cohort. (D) Survival analysis of the GEO high- and low-risk groups (P=0.007). (E) Risk model AUCs of GEO at 1, 3, and 5 years. GEO, Gene Expression Omnibus; AUC, area under curve; CRC, colorectal cancer; COAD, colon adenocarcinoma.
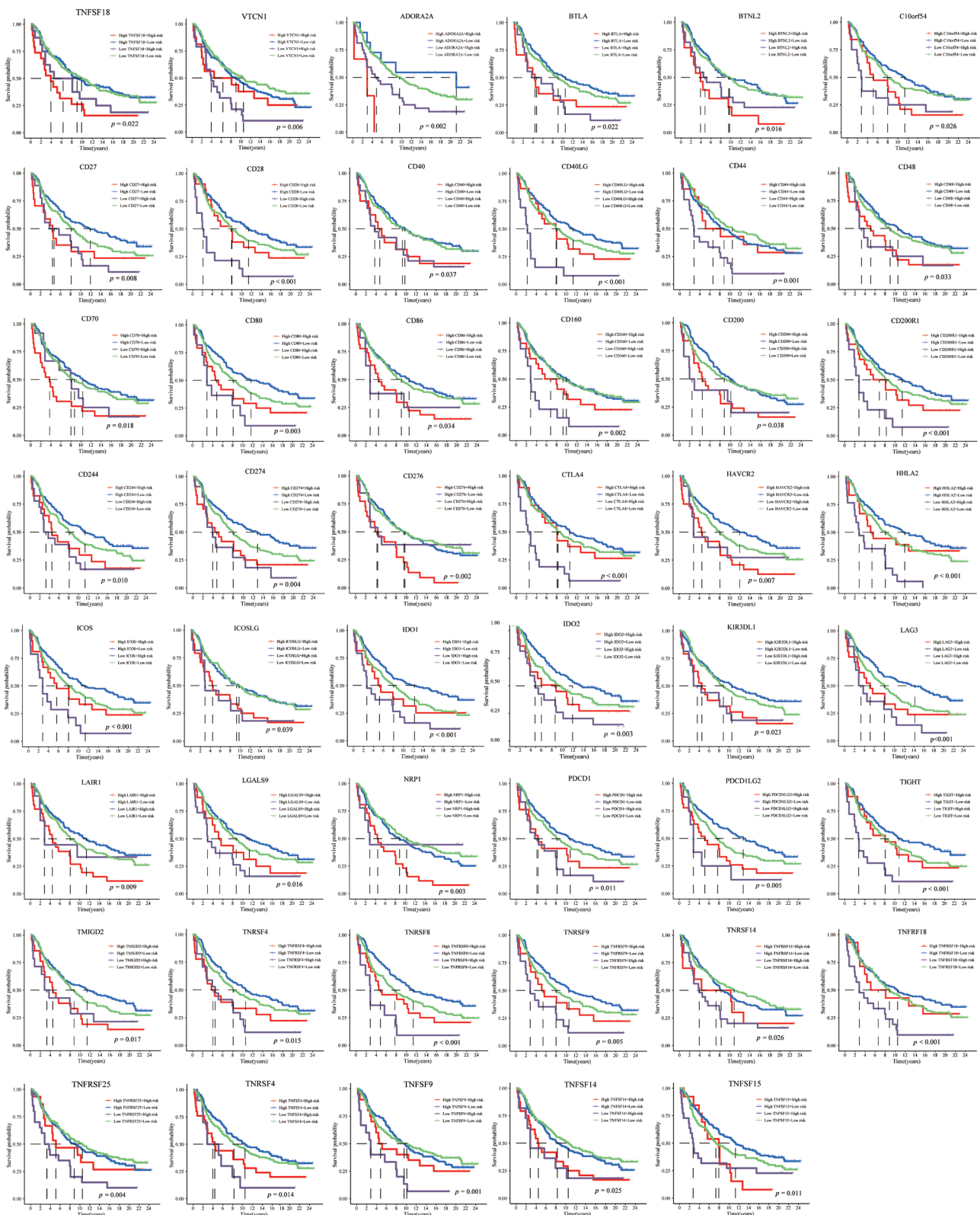
Figure S3 Survival analysis of high- and low-risk groups for different immune checkpoints.
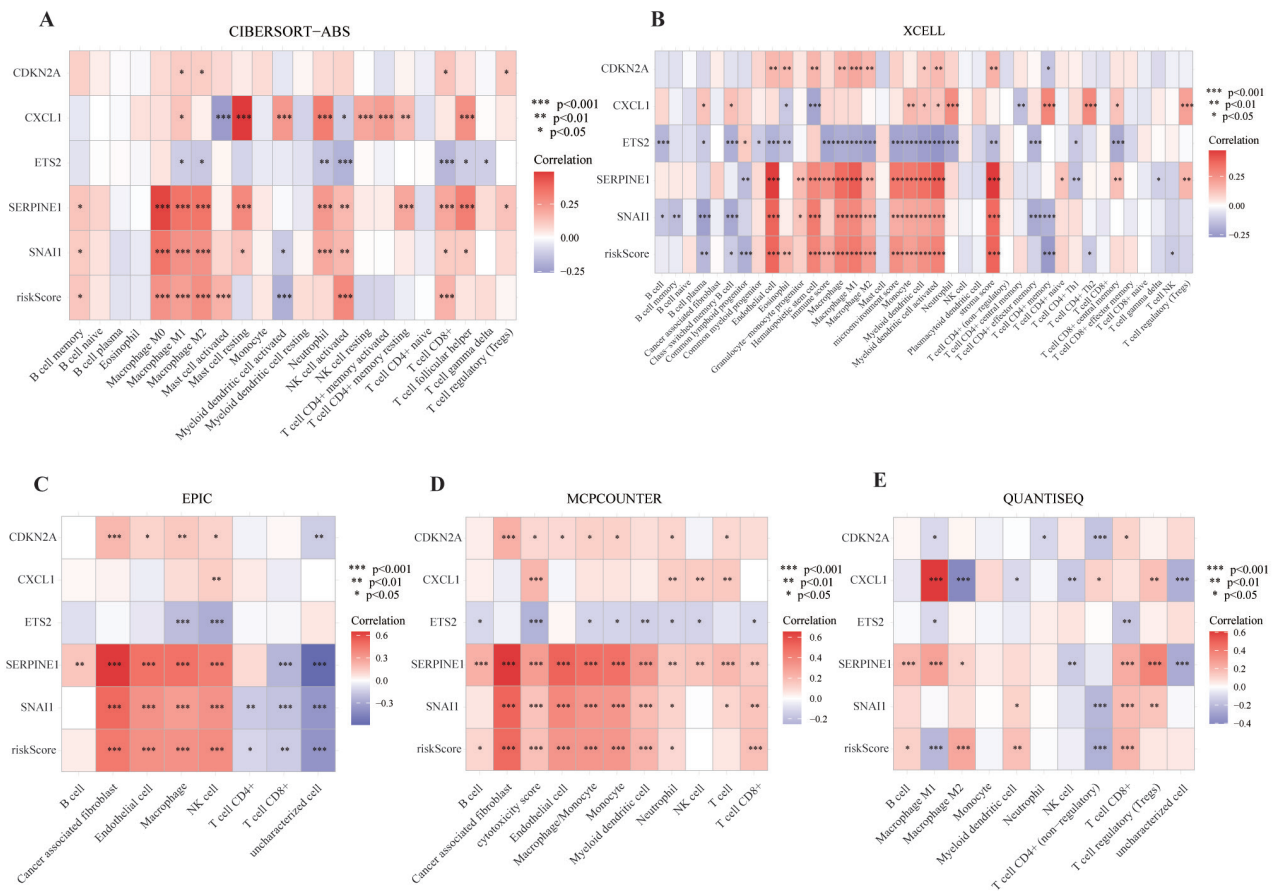
Figure S4 High-risk score cohort and low-risk score cohort enrichment analysis of different immunization databases: (A) CIBERSORT-ABS; (B) XCELL; (C) EPIC; (D) MCPCOUNTER; (E) QUANTISEQ. A stronger red indicates a stronger correlation, and a stronger blue indicates a weaker correlation. *, P<0.05; **, P<0.01; ***, P<0.001. NK, natural killer.

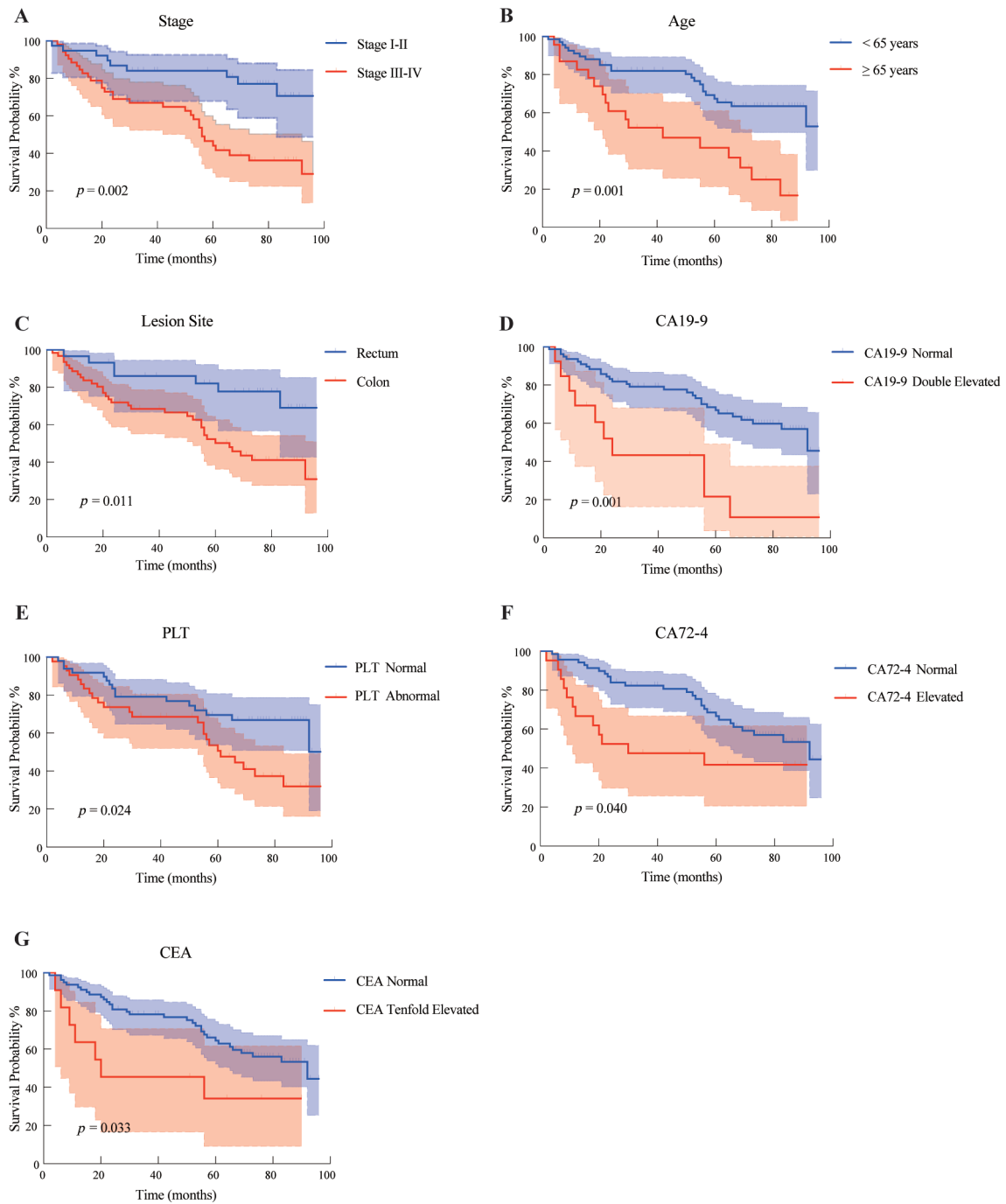**Figure S5** Survival analysis of the clinical parameters: (A) stage; (B) age; (C) lesion site; (D) *CA19-9*; (E) PLT; (F) *CA72-4*; (G) *CEA*. PLT, platelet; CEA, carcinoembryonic antigen.

Table S1 Parameter collection forms for public databases and clinical samples

Variables	TCGA cohort (n=385)	GEO cohort (n=566)	Clinical cohort (n=90)
Age (years)			
≤65	159	223	67
>65	226	342	23
Unknown	0	1	0
Gender			
Female	180	256	40
Male	205	310	50
Stage			
Stage I–II	217	301	38
Stage III–IV	157	265	52
Unknown	11	0	0
T			
T1–2	78	60	13
T3–4	307	486	75
Unknown	0	20	2
N			
N0	231	302	44
N1	88	134	33
N2	66	98	5
N3	0	6	3
Unknown	0	26	5
M			
M0	325	482	69
M1	54	61	20
Unknown	6	23	1
Immunohistochemistry			
CDKN2A tumor positive	–	–	28
CDKN2A tumor negative	–	–	62
ETS2 peritumoral positive	–	–	30
ETS2 peritumoral negative	–	–	60

TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus.

Table S2 Survival analysis of clinical indicators

Variables	Univariate survival analysis			Multivariate survival analysis		
	HR	95% CI	P value	HR	95% CI	P value
Stage	3.224	1.527–6.807	0.002	3.183	1.487–6.841	0.003
Age, 65 years	2.977	1.575–5.627	0.001	3.636	1.890–6.993	<0.001
Sex	1.005	0.539–1.875	0.986	–	–	–
BMI	1.165	0.613–2.212	0.641	–	–	–
Lesion site	2.899	1.280–6.565	0.011	–	–	–
Therapy	0.692	0.357–1.343	0.276	–	–	–
PLT	2.073	1.099–3.910	0.024	–	–	–
WBC	1.091	0.482–2.470	0.834	–	–	–
Neutrophils	1.316	0.643–2.694	0.453	–	–	–
Lymphocyte	1.014	0.505–2.035	0.969	–	–	–
LDH	2.519	0.977–6.493	0.056	2.556	0.947–6.896	0.064
CA19-9	3.41	1.645–7.067	0.001	3.237	1.529–6.850	0.002
CA72-4	2.045	1.033–4.048	0.04	–	–	–
CEA	2.445	1.075–5.561	0.033	–	–	–
CDKN2A	1.962	1.052–3.659	0.034	2.070	1.100–3.894	0.024
ETS2	0.43	0.198–0.933	0.033	–	–	–

HR, hazard ratio; CI, confidence interval; BMI, body mass index; PLT, platelet; WBC, white blood cell; LDH, lactate dehydrogenase; CEA, carcinoembryonic antigen.