

Discovery Dataset
TCGA

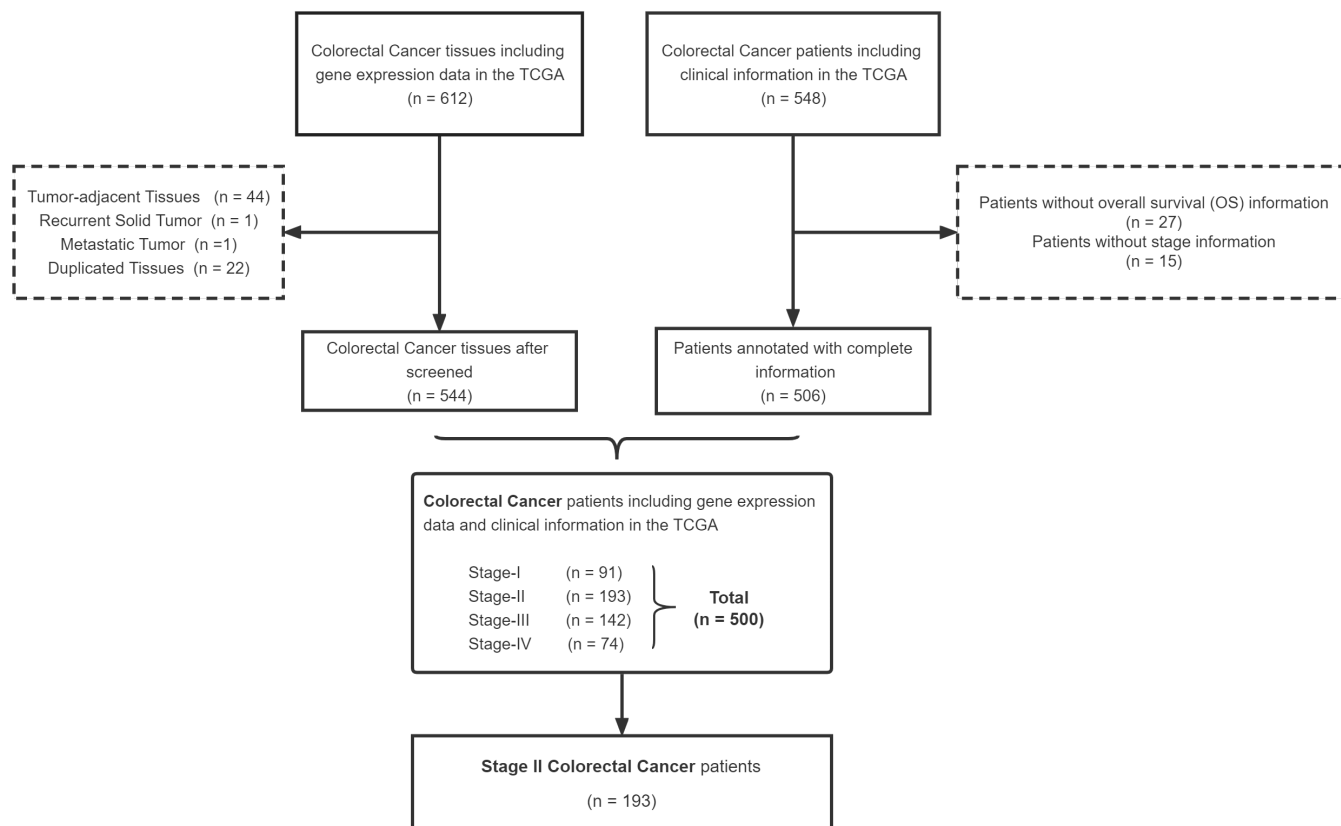


Figure S1 Patients composition of the “Discovery Dataset” TCGA.

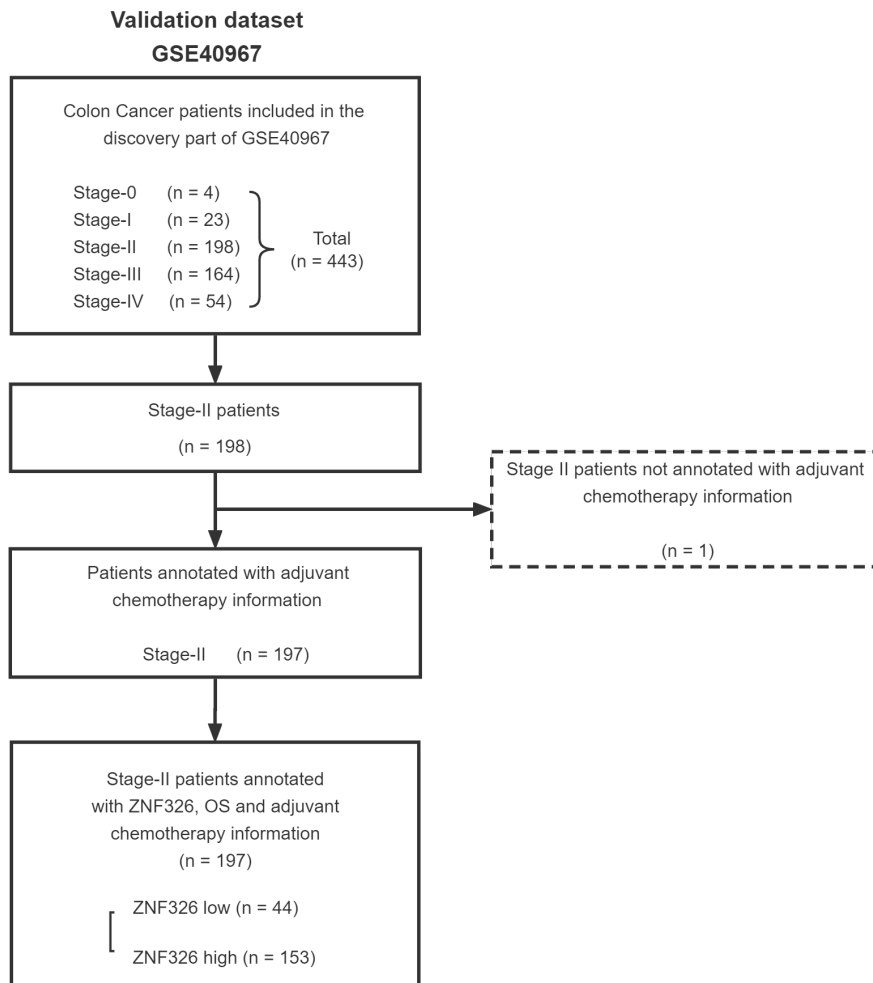


Figure S2 Patients composition of the “Validation Dataset” GSE40967.

**Expansion Dataset #1
GSE29623**

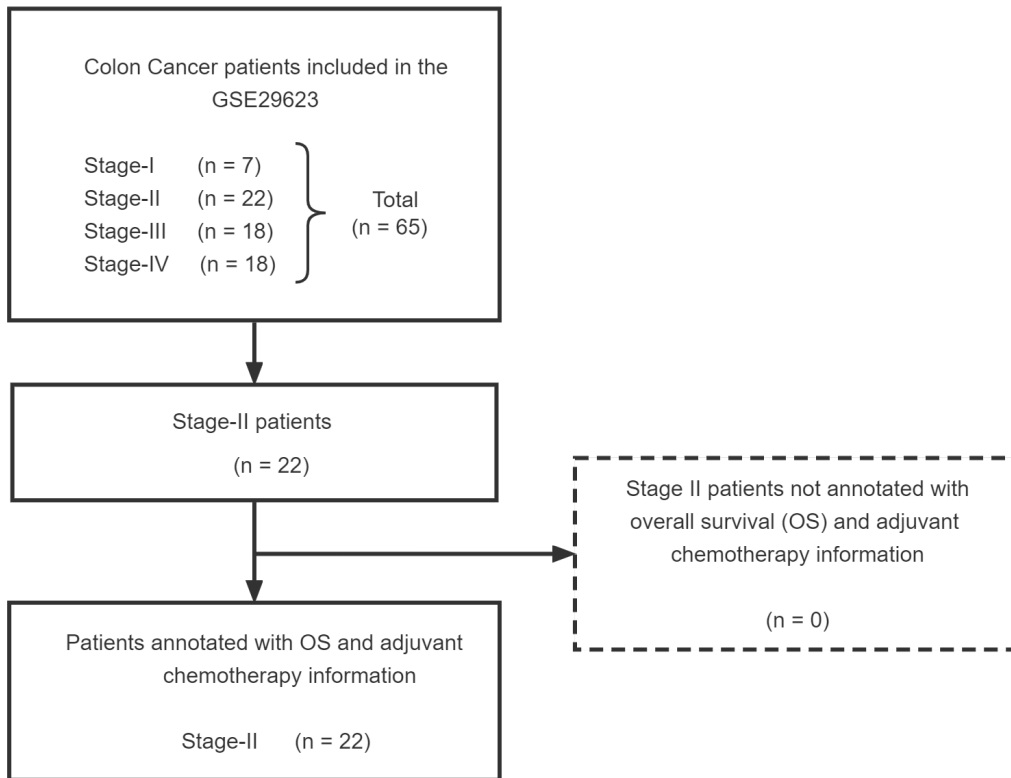


Figure S3 Patients composition of the “Expansion Dataset #1” GSE29623.

**Expansion dataset #2
GSE103479**

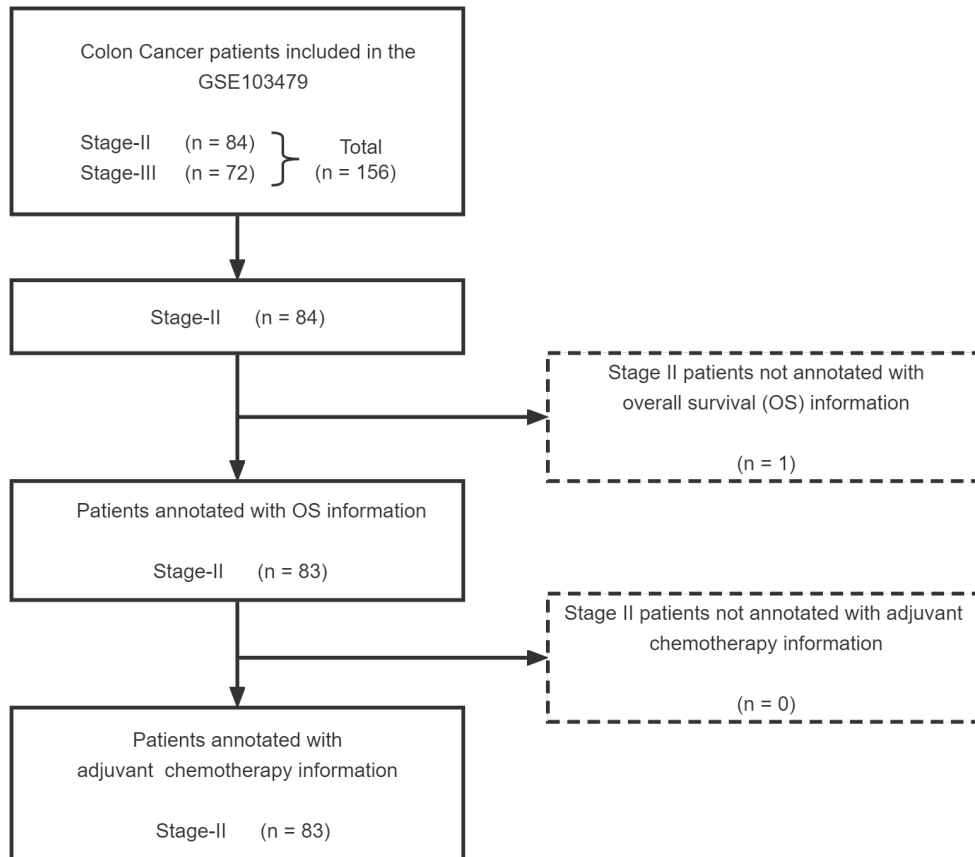


Figure S4 Patients composition of the “Expansion Dataset #2” GSE103479.

Table S1 Characteristics of stage II patients at baseline and follow-up in TCGA and GEO datasets

Variable	TCGA (n=193)	GSE40967 (n=197)	GSE29623 (n=22)	GSE103479 (n=83)
Age, year	68.0 [60.0, 77.0]	70.0 [61.0, 77.0]	N/A ^a	70.6 [61.9, 77.8]
Male	109 (56.5)	111 (56.3)	11 (50.0)	48 (57.8)
T stage				
T2	0 (0.0)	4 (2.0)	0 (0.0)	0 (0.0)
T3	180 (93.3)	142 (72.1)	19 (86.4)	70 (84.3)
T4	13 (6.7)	44 (22.3)	3 (13.6)	13 (15.7)
N/A ^a	0 (0.0)	7 (3.6)	0 (0.0)	0 (0.0)
N stage				
N0	193 (100.0)	190 (96.4)	21 (95.5)	83 (100.0)
NX	0 (0.0)	0 (0.0)	1 (4.5)	0 (0.0)
N/A ^a	0 (0.0)	7 (3.6)	0 (0.0)	0 (0.0)
M stage				
M0	179 (92.7)	190 (96.4)	21 (95.5)	49 (59.0)
MX	12 (6.2)	0 (0.0)	1 (4.5)	34 (41.0)
N/A ^a	0 (0.0)	7 (3.6)	0 (0.0)	0 (0.0)
Adjuvant chemotherapy	N/A ^a	40 (20.3)	9 (40.9)	25 (30.1)
Follow-up				
Overall survival	172 (89.1)	136 (69.0)	19 (86.4)	59 (71.1)
Overall survival time, year	2.0 [1.1, 3.3]	4.6 [2.8, 6.9]	4.3 [3.0, 5.1]	4.7 [3.5, 7.2]

Data are presented as median [IQR] or n (%). ^a, missing value.

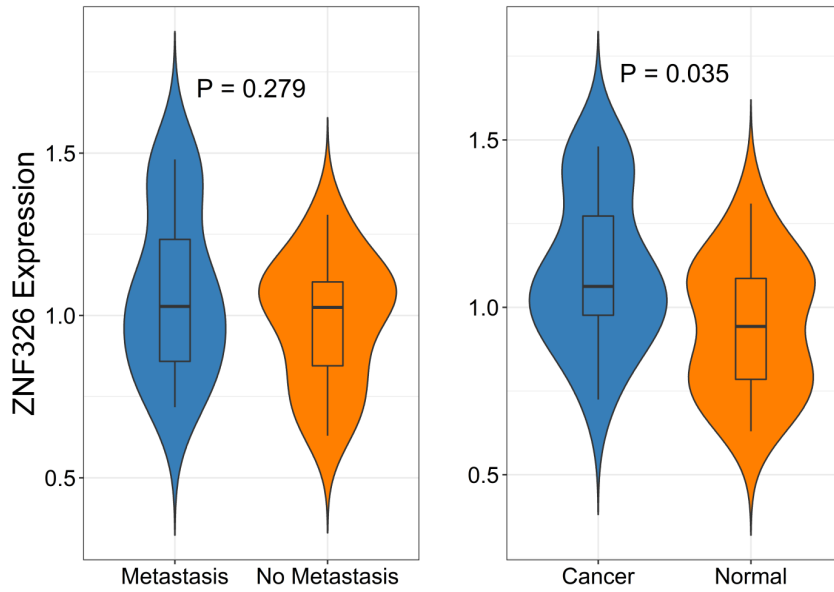
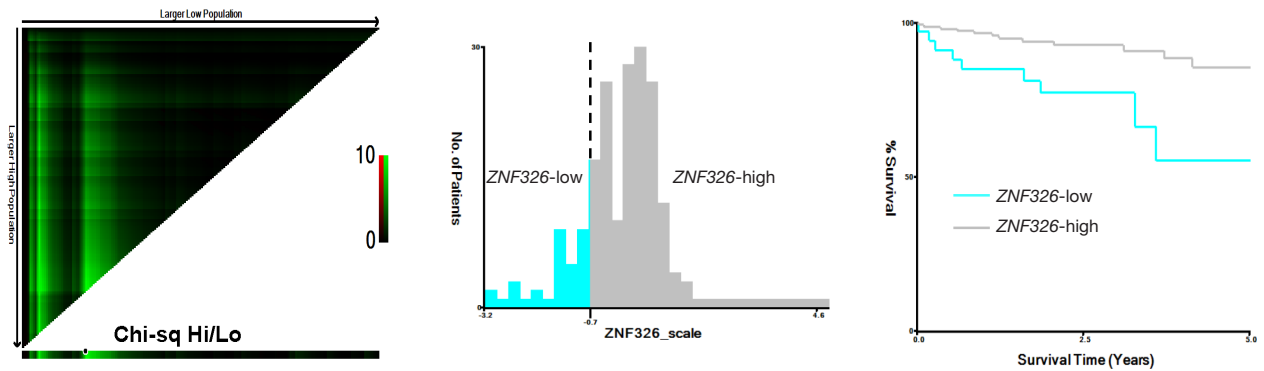


Figure S5 Violin diagram showing the distribution of ZNF326 protein expression. Left: the distribution of ZNF326 protein expression between metastatic and non-metastatic groups. Right: tumor and adjacent tissues.

Survival Analysis: ZNF326



Subpopulation Cutpoints:

Pt No	% Total	Events	Rate	Rank	Range
35	18.13	9	25.71	0 to 34	-3.18 thru -0.72
158	81.87	12	7.59	35 to 192	-0.72 thru 4.62
193	100.00	21	10.88	0 to 192	-3.18 thru 4.62

Figure S6 X-tile plots of ZNF326 gene expression in stage II patients, using TCGA dataset.

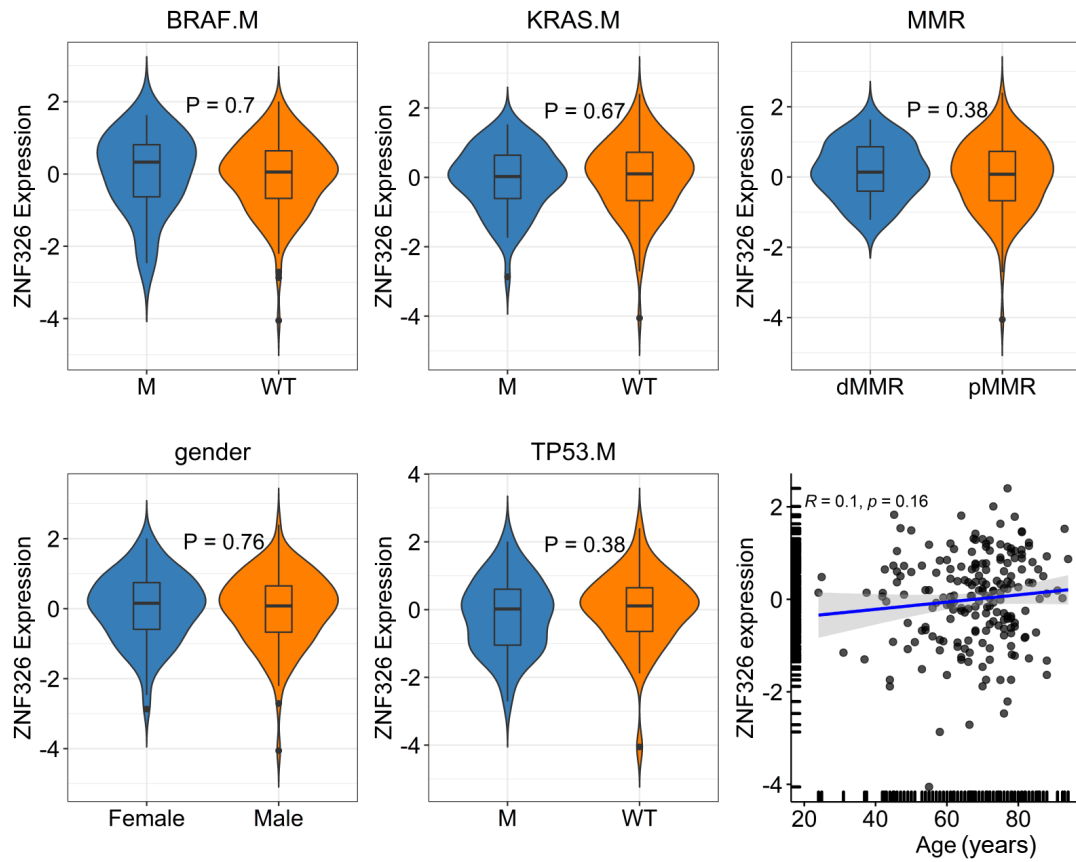


Figure S7 Violin-plots and *t*-test, scatterplot and Pearson's correlation analysis showing the relationships between *ZNF326* gene expression and existing prognostic factors (age, gender, genetic mutations in the *BRAF*, *KRAS*, and *TP53* genes, and MMR), using GSE40967 dataset.

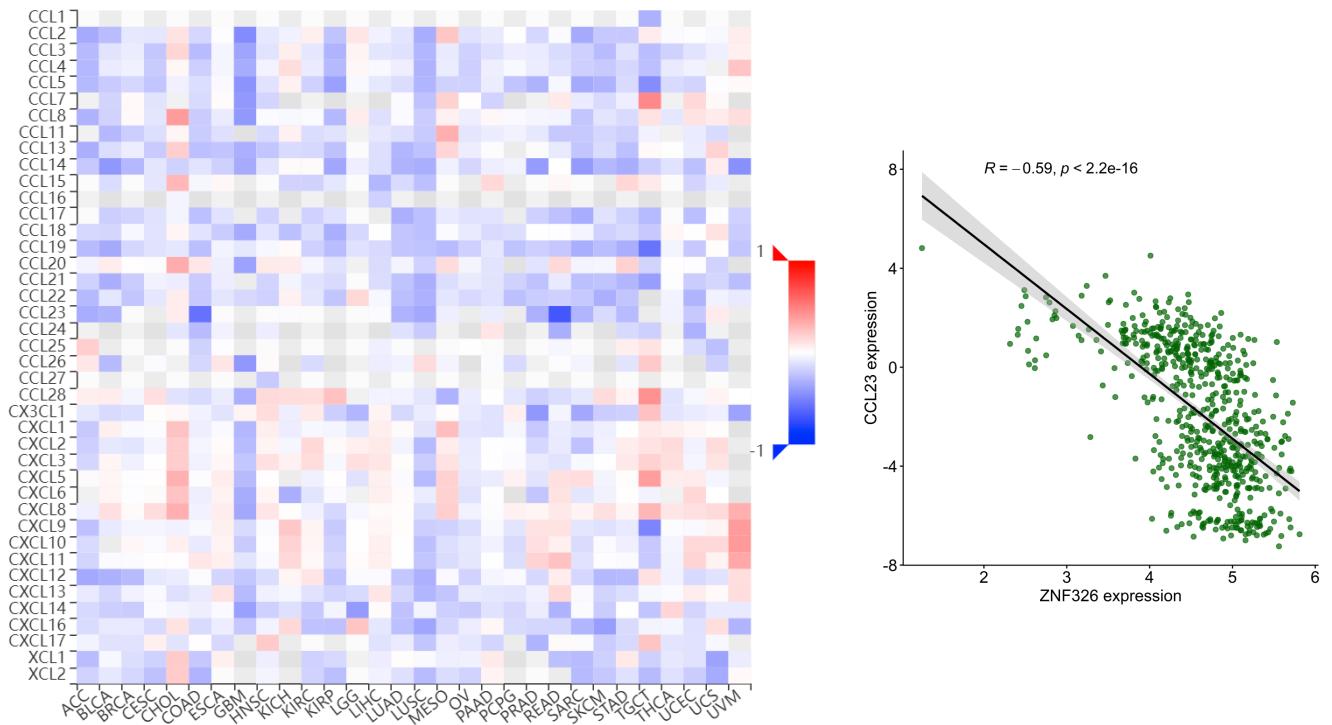


Figure S8 Spearman correlations between *ZNF326* expression and chemokines (left) across human cancers (X axis) and CCL23 in human colorectal cancer (right).