A chemoresistance lncRNA signature for recurrence risk stratification of colon cancer patients with chemotherapy

Hao Wang,[1,4,6] Yuzhen Gao,[2,3,6] Somayeh Vafaei,[5] Qiaoyan Yu,[1,4] Jun Zhang,[2,3] and Liangjing Wang[1,4]

[1]Department of Gastroenterology, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang 310009, China; [2]Department of Clinical Laboratory, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310016, China; [3]Biomedical Research Center, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310013, China; [4]Institution of Gastroenterology, Zhejiang University, Hangzhou, Zhejiang, China; [5]Department of Molecular Medicine, Faculty of advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran

Chemotherapy is considered the nonsurgical treatment of choice for colon cancer patients. However, no precise molecular markers are available to determine which patients can actually benefit from it. In this study, we identified 55 chemotherapy-specific long non-coding RNAs (lncRNAs) of colon cancer patients through a systematic assessment of lncRNA expression profiles from a public database. These were taken from multiple cohorts of colon cancer patients who had received chemotherapy, or not. Based on these data, a chemoresistance lncRNA signature, named CRLSig, was constructed and successfully applied to divide chemotherapy patients into two groups with different recurrence-free survival (RFS) rates. Gene set enrichment analysis revealed that patients with low CRLSig had more infiltrating CD8+ T cells and macrophages, while those with high CRLSig had more infiltrating natural killer T cells. KEGG pathway analysis revealed that the low CRLSig group had more activated metabolic pathways compared with those in the high CRLSig group, indicating better response to chemotherapy. Single-cell sequencing analysis revealed that stromal cells and epithelial cells had higher CRLSig. Thus, we have constructed an auxiliary prognostic tool, CRLSig, able to discriminate patients at high risk of RFS, despite having received standard adjuvant chemotherapy treatment.

INTRODUCTION

Colon cancer is one of the most common malignancies of the gastro-intestinal tract; it ranks third in terms of incidence, while second in terms of mortality, worldwide.[1] Currently, colectomy, when combined with adjuvant chemotherapy and radiotherapy, is recognized as the standard treatment for colon cancer. In addition, biologics and immunotherapy are reported to benefit patients with metastatic colon cancer, such as anti-VEGF monoclonal antibody targeting angiogenesis, anti-EGFR therapies, PD-1 blockade, and CTLA-4 inhibitor.[2–6] Although chemotherapy is beneficial, outcomes vary widely. Moreover, no clinical predictors have been developed to determine which colon cancer patients will benefit from chemotherapy, indicating the importance of proper patient stratification.

Based on current guidelines, stage II colon cancer patients with high-level microsatellite instability (MSI-H) or defective DNA mismatch repair (dMMR) are not likely to have successful chemotherapeutic outcomes in clinical practice.[7,8] Consequently, MMR checks are routinely performed in the clinic. However, the practice is imprecise because of the large gap between microsatellite status and accurate identification of patients who will benefit from adjuvant chemotherapy in primary colon cancer.[9] In addition, tumor-tissue DNA mutation profiling and blood-derived circulating tumor DNA, as well as the expression profiles of protein-coding genes, have all been reported as predictors of chemotherapy response.[10–12] Here, we focus on long non-coding RNAs (lncRNAs) as a predictor of chemotherapy in colon cancer patients to address these gaps and provide better patient stratification resulting in personalized chemotherapy treatment that is more effective and less futile.

The lncRNAs belong to a class of transcripts that are not translated into functional proteins and that are longer than 200 nucleotides.[13,14] They can modulate gene expression on pre-transcriptional, transcriptional, and post-transcriptional levels by interacting with DNA, mRNA, and proteins.[15,16] In addition, as competitive endogenous RNAs of microRNAs (miRNAs), lncRNAs can also modulate gene expression by regulating miRNAs to target mRNAs.[17,18] In recent years, lncRNAs have been associated with the development and progression of cancer.[19] For colon cancer, several lncRNAs have been associated with cell proliferation and apoptosis, cell metastasis and invasion, epithelial-mesenchymal transition, drug resistance, and
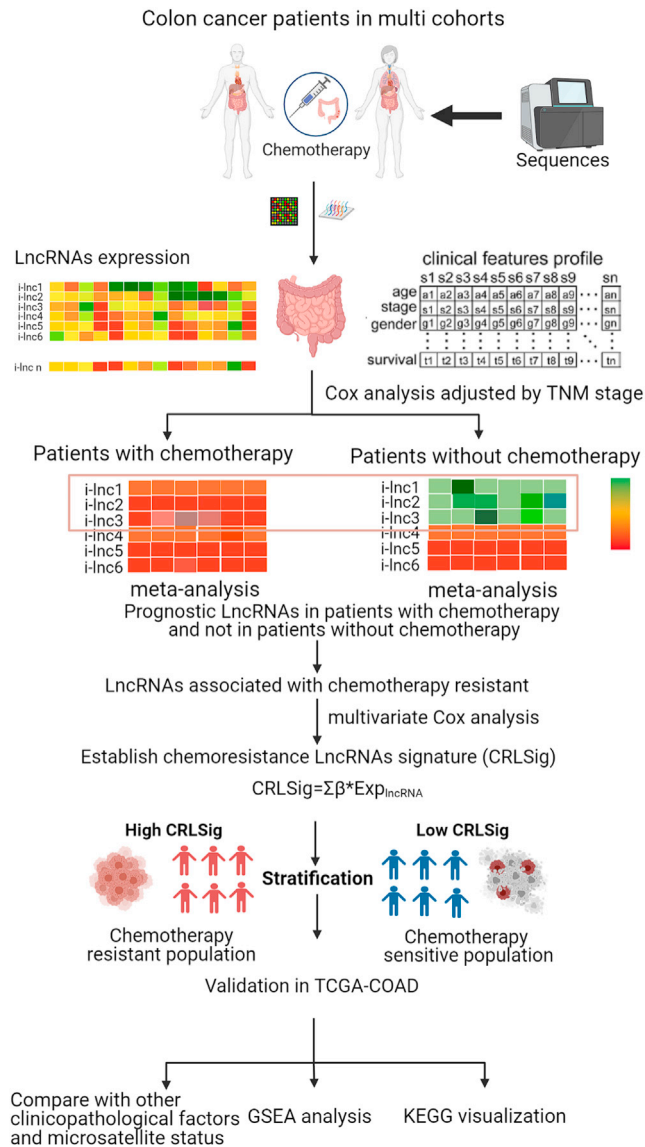
Figure 1. Workflow for constructing CRLSig in this study

cancer stem cell regulation.[20] Nowadays, accumulating evidence has shown that the construction of an lncRNA expression profile signature could effectively predict overall survival (OS) and determine the recurrence risk of patients with colon cancer. For example, Zhou et al.[21] identified a six-lncRNA signature to determine the postoperative recurrence risk of colon cancer. In addition, another nine-lncRNA signature was revealed to predict OS of patients with colon cancer.[22] However, the expression characteristics of lncRNA in relation to chemotherapy resistance have still not been established.

Herein, we systematically evaluate lncRNA expression profiles from public datasets. These consisted of multiple cohorts of colon cancer patients with and without chemotherapy. In total, 55 chemo-

therapy-specific lncRNAs were identified in our study. Furthermore, based on these data, a chemoresistance lncRNA signature (CRLSig) was constructed to predict the clinical outcomes of chemotherapy patients. This novel tool provides reliable and effective prognosis in colon cancer patients with chemotherapy, differentiating those patients most likely to benefit from chemotherapy treatment from those less likely.

RESULTS

Identification of chemotherapy-related lncRNAs in colon cancer patients using systematic meta-analysis

Our study flowchart is shown in Figure 1. Clinical data with bulk sequence of colon cancer from three Gene Expression Omnibus (GEO) cohorts, GSE103479, GSE39582, and GSE72970, and a TCGA_COAD (colon adenocarcinoma) cohort, are listed in Tables 1 and S1. Colon cancer patients were divided into two groups according to chemotherapy treatments. Comparison of the prognosis of colon cancer patients, with or without chemotherapy in these three cohorts, shows significant difference (Figure S1, p < 0.05).

Notably, our study has two phages: discovery (training) and validation. First, by matching GENCODE (release 25) and RefSeq (release 79), we obtained a total of 2,456 unique lncRNAs from these microarray and sequence data from patients included in GSE103479, GSE39582, and GSE72970. Then, we performed Cox proportional-hazards regression analysis to detect recurrence-free survival (RFS)-related hazard ratios (HRs) with 95% CI of 2,456 lncRNAs, adjusted by TNM stage of patients with or without chemotherapy in these populations, respectively. To retain stable RFS-related lncRNAs in patients with chemotherapy, we used the fixed-effect model of systematic meta-analysis to pool the HR values of 2,456 lncRNAs in the subgroup population with chemotherapy (3 cohorts of 383 colon cancer patients with chemotherapy from GSE39582 [n = 240], GSE72970 [n = 83], and GSE103479 [n = 60]) and the other subgroup population without chemotherapy (2 cohorts of 396 patients without chemotherapy from GSE39582 [n = 326] and GSE103479 [n = 70]). Among them, 177 lncRNAs were significantly related to RFS in patients without chemotherapy (Figure 2A, p < 0.05), and 268 lncRNAs were significantly related to RFS in patients with chemotherapy (Figure 2A, p < 0.05). However, 49 lncRNAs were significantly related to RFS in patients with and without chemotherapy (Figure 2A, p < 0.05). Finally, 55 lncRNAs were recognized as stable chemotherapy-resistant lncRNAs in colon cancer patients who had undergone chemotherapy (meta-HR > 1, p < 0.001).

Among these 55 lncRNAs, we found significant positive relationships in patients with chemotherapy (Figure 2B; Table S2, p < 0.0001). Both prognostic values and β coefficients of chemotherapy-related lncRNAs calculated by the meta-analysis are shown in Figure 2C and Table S3.

Construction and validation of CRLSig

Next, based on the Cox regression formula, as described in Materials and methods, and the β values, we initially established the CRLSig for

Table 1. Clinical features of colon cancer patients with chemotherapy of four cohorts

Clinical variables	Level	GSE39582 (n = 240)	GSE72970 (n = 83)	GSE103479 (n = 60)	TCGA_COAD (n = 71)	p value
Age, mean (SD)		62.98 (11.88)	61.89 (11.51)	62.89 (9.95)	63.15 (12.35)	0.872
Sex (%)	female	107 (44.6)	35 (42.2)	28 (46.7)	34 (47.9)	0.937
	male	133 (55.4)	48 (57.8)	32 (53.3)	37 (52.1)	
T (%)	T1	3 (1.2)	0 (0.0)	1 (1.7)	0 (0.0)	<0.001
	T2	10 (4.2)	3 (3.6)	2 (3.3)	5 (7.0)	
	T3	166 (69.2)	36 (43.4)	44 (73.3)	50 (70.4)	
	T4	51 (21.2)	29 (34.9)	13 (21.7)	16 (22.5)	
	TX	10 (4.2)	15 (18.1)	0 (0.0)	0 (0.0)	
N (%)	N0	66 (27.5)	10 (12.0)	24 (40.0)	16 (22.5)	<0.001
	N1	91 (37.9)	20 (24.1)	27 (45.0)	33 (46.5)	
	N2	68 (28.3)	38 (45.8)	9 (15.0)	22 (31.0)	
	N3	5 (2.1)	0 (0.0)	0 (0.0)	0 (0.0)	
	NX	10 (4.2)	15 (18.1)	0 (0.0)	0 (0.0)	
M (%)	M0	198 (82.5)	12 (14.5)	35 (58.3)	44 (62.0)	<0.001
	M1	31 (12.9)	71 (85.5)	0 (0.0)	17 (23.9)	
	Mx	11 (4.6)	0 (0.0)	25 (41.7)	10 (14.1)	
TNM stage (%)	I	0 (0.0)	1 (1.2)	0 (0.0)	0 (0.0)	<0.001
	II	58 (24.2)	3 (3.6)	24 (40.0)	15 (21.1)	
	III	152 (63.3)	8 (9.6)	36 (60.0)	39 (54.9)	
	IV	30 (12.5)	71 (85.5)	0 (0.0)	17 (23.9)	
MMR/MSI status (%)	dMMR	15 (6.2)	NA	NA	3 (4.2)	NA
	pMMR	210 (87.5)	NA	NA	7 (9.9)	
	NA	15 (6.2)	83 (100.0)	60 (100.0)	61 (85.9)	
OS (%)	alive	164 (68.3)	22 (26.5)	45 (75.0)	54 (76.1)	<0.001
	death	76 (31.7)	61 (73.5)	15 (25.0)	17 (23.9)	
RFS (%)	no	148 (61.7)	7 (8.4)	40 (66.7)	51 (71.8)	<0.001
	RE	91 (37.9)	76 (91.6)	20 (33.3)	20 (28.2)	
	NA	1 (0.4)				

MMR, mismatch repair; MSI, microsatellite instability; OS, overall survival; RFS, recurrence-free survival.

all colon cancer patients. The distributions of CRLSig were not significantly different between patients with and without chemotherapy in these cohorts (Figure S2). The CRLSig values of all patients with chemotherapy from four cohorts were listed in Table S4. We set the median values as the cutoff for CRLSig to divide two groups of patients (chemotherapy and no chemotherapy) into three independent cohorts, namely 1.2, 1.2, and 1.5, respectively. Prognosis analysis showed that the RFS rate of colon cancer patients in the chemotherapy group could be significantly distinguished by CRLSig in these cohorts (Figures 3A–3C), including GSE39582 (p < 0.001), GSE72970 (p = 0.004), and GSE103479 (p = 0.033). Furthermore, in the validation set, the significant difference between high CRLSig and low CRLSig groups could also be observed (Figure 3D, TCGA_COAD, n = 71, p = 0.005). To evaluate the prognostic performance of CRLSig in patients without chemotherapy, we also performed the log rank test and found that CRLSig failed to divide patients into two groups with different RFS rate in both training and validation sets (Figure S3). In

addition, because CRLSig could differentiate between high and low groups, this tool could be used identify patients with better prognosis by undergoing chemotherapy or patients with high risk of RFS, despite receiving the standard adjuvant chemotherapy treatment in these cohorts (Figure S4). By using time-receiver operating characteristic (ROC) analysis, CRLSig showed a range of AUCs (GSE39582 [0.58–0.65], GSE72790 [0.73–0.75], GSE103479 [0.46–0.65], and TCGA_COAD [0.71–0.83]) for predicting the prognosis of colon cancer patients with chemotherapy in these cohorts (Figures 3E–3H). A histogram showing the status of recurrence along with increasing CRLSig of the patients with chemotherapy also revealed the same significant trend (Figures 3I–3L).

Subgroup analysis of CRLSig in colon cancer patients with chemotherapy

Here, we further tested the prognosis values of CRLSig at different TNM stages in the GSE39582, GSE72970 and TCGA_COAD
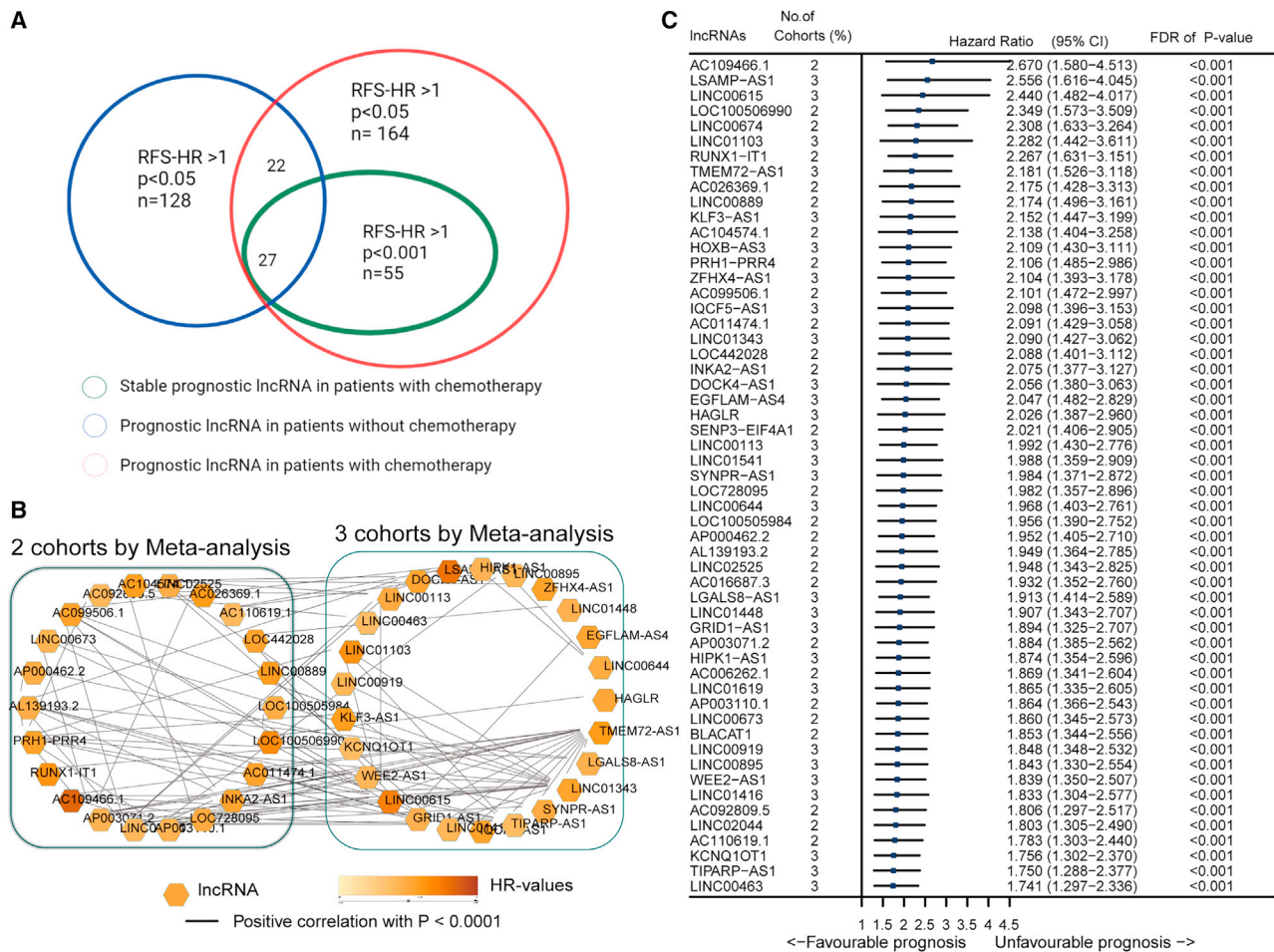
Figure 2. Identification of chemotherapy-related lncRNAs of patients with colon cancer by using systematic meta-analysis

(A) Identification of prognostic lncRNAs in colon cancer patients with chemotherapy and non-chemotherapy to obtain chemotherapy-related lncRNAs by using systematic meta-analysis. (B) Relationships among 55 chemotherapy-related lncRNAs in patients with chemotherapy of the GSE39582 cohort. Left square (27 lncRNAs) and right square (28 lncRNAs) were identified by meta-analysis in two and three frequencies, respectively. (C) Detailed prognostic value of 55 chemotherapy-related lncRNAs.

cohorts (patients of GSE103479 were all stage II + III, no stage IV) to evaluate the accuracy and effectiveness of CRLSig, although we had constructed CRLSig based on the adjustment for TNM stage. We found significant differences in RFS rate between high and low CRLSig groups in patients with stage II + III or stage IV in both training cohorts (GSE72970 and GSE39582) and validation cohort (TCGA_COAD) (Figures 4A–4F). In addition, chemotherapy is currently recommended according to MMR status at different stages. We tested the prognosis values of MMR status at different TNM stages. In contrast, MMR status could not distinguish between patients with stage II and stage III relative to chemotherapy prognosis (Figure S5).

Prediction of CRLSig for treatment response in chemotherapy

Since the 55 lncRNAs may reduce the effectiveness of chemotherapy, we validated the prediction value of CRLSig for the treatment responses of the patients in The Cancer Genome Atlas (TCGA) cohort (n = 71). The bar plot shows the CRLSig score as increasing along with the incidence of SD/PD in response to chemotherapy (Figure 5A). Chi-square test of the histogram also found a significant association between CRLSig and chemotherapy response (Figure 5B, p < 0.05). The CRLSig score of chemotherapy patients with SD/PD status was higher than that of patients with PR/CR status (Figure 5C, p = 0.01). ROC analysis showed that the AUC with 95% CI of CRLSig to predict the PR/CR was 0.731 (0.554–0.878) (Figure 5D).

Pathway enrichment analysis between low and high CRLSig in TCGA_COAD

According to Liebermeister et al.,[23] proteome quantitative data can be visualized using the graphical tool Proteomaps, which shows "the composition of proteomes with a focus on protein abundances and functions." In Proteomaps, each protein is shown as a polygon-shaped tile, with an area representing protein abundance.
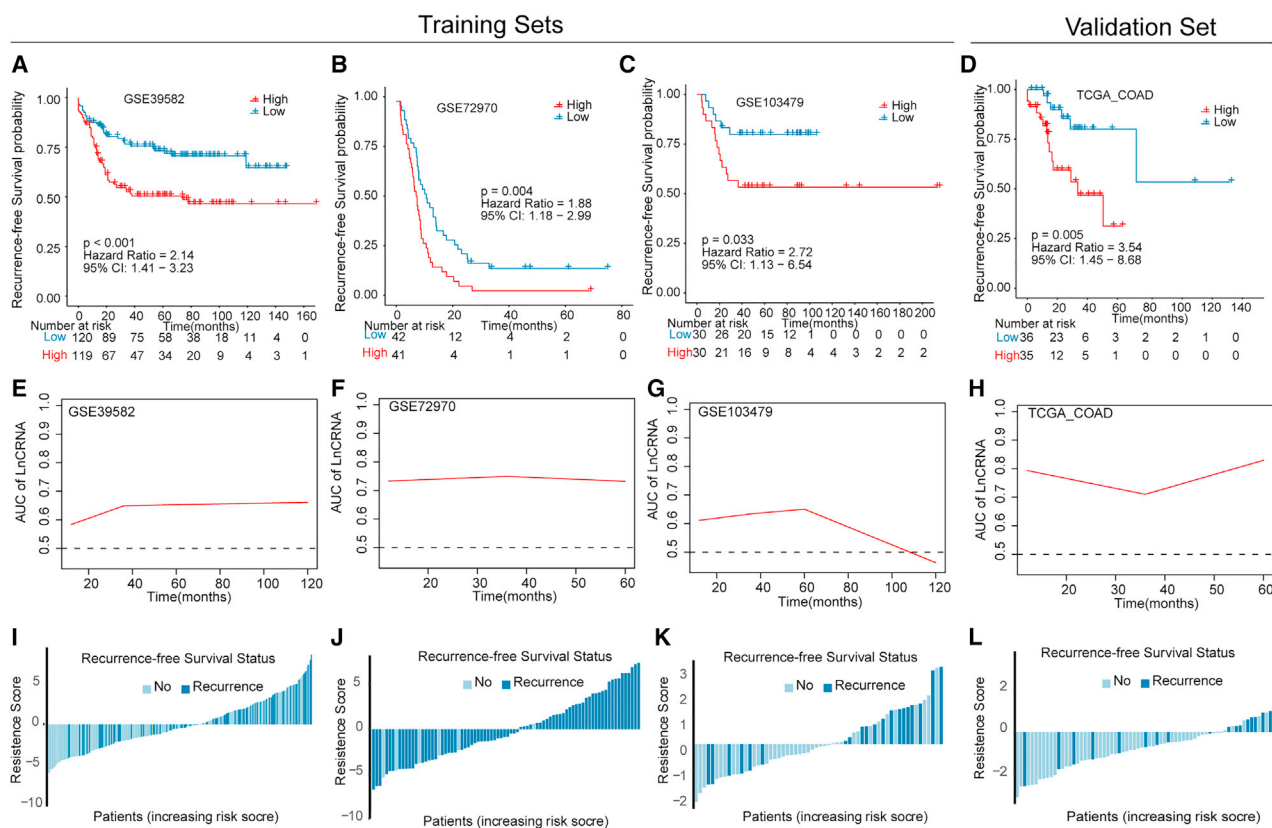
Figure 3. CRLSig was associated with clinical outcome in three training sets and one validation set

(A–D) Kaplan-Meier survival curves of recurrence-free survival (RFS) probability between patients with high CRLSig and low CRLSig. (E–H) Time-dependent receiver operating characteristic curve of CRLSig for predicting the prognosis of the colon cancer patients with chemotherapy. (I–L) The histogram of recurrence status along with the increasing CRLSig score of the patients with chemotherapy from GSE39582, GSE72970, GSE103479, and TCGA_COAD cohorts.

Functionally related proteins appear in adjacent regions. Here, we performed the online Proteomaps (https://bionic-vis.biologie.uni-greifswald.de/) to observe dynamic change in the proportion of their Kyoto Encyclopedia Genes and Genomes (KEGG) pathway between different patient groups. Patients who responded to chemotherapy had a higher proportion of metabolic genes, according to KEGG pathway analysis, than unresponsive patients (Figures 5E and 5F). Also, according to KEGG pathway analysis, patients with low CRLSig showed a proportion of metabolic genes similar to patients who responded to chemotherapy (Figures 5G and 5H). Thus, in either the responsive group or the low CRLSig group, higher levels of metabolic RNAs dominated.

To observe the metabolic characteristics of these groups, we further collected 113 metabolic pathways from a previous study[24] and performed gene set variation analysis (GSVA) for TCGA_COAD patients with chemotherapy in our study. By comparing high/low CRLSig groups or CR, PR/PD, and SD groups among patients with chemotherapy, we found that 27 and 25 metabolic pathways were activated in patients with PR/CR and patients with low CRLSig, respectively (Tables S5 and S6). Among them, 17 metabolic pathways,

such as urea cycle, pyruvate metabolism, lipoic acid metabolism, and others, were activated both in PR/CR and low CRLSig groups (Figures 5I and 5J).

Independence of CRLSig from other clinicopathological factors and its clinical features

To determine whether the prognostic value of CRLSig is independent of other clinicopathological factors in the training and validation cohorts, we used univariate Cox regression analysis to test the performance of CRLSig, and Table S7 shows the results. Then, multivariate Cox regression analysis showed that CRLSig was significantly related to RFS in the training cohorts (GSE103479: HR = 2.626; 95% CI, 1.007–6.852; GSE72970: HR = 1.928; 95% CI, 1.217–3.056; GSE39582: HR = 2.129; 95% CI, 1.386–3.271) and the validation cohort (TCGA_COAD: HR = 3.95; 95% CI, 1.418–11.003), respectively (Figure 6A). We also tested the relationship between CRLSig and other clinicopathological factors in TCGA cohort. Besides OS and RFS, CRLSig was also significantly related to TNM stage in colon cancer patients with chemotherapy (Figure 6B, TCGA_COAD). CRLSig also significantly distinguished the OS rate by the log rank test in TCGA validation cohort (Figure 6C).
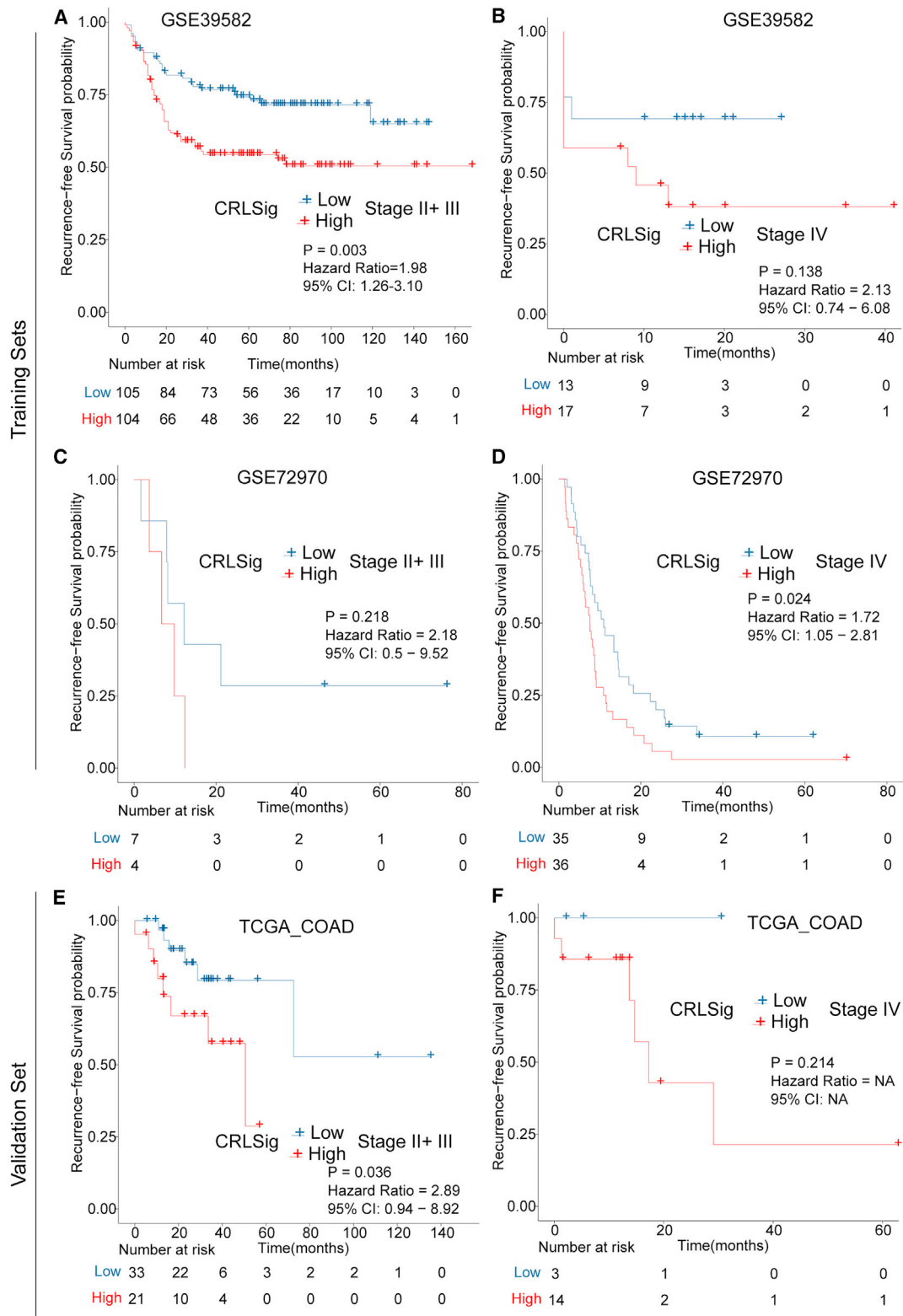
Figure 4. Subgroup analysis of Kaplan-Meier survival curves of CRLSig in predicting RFS prognosis of colon cancer patients after chemotherapy

(A) GSE39582 stage II + III; (B) GSE39582 stage IV; (C) GSE72970 stage II + III; (D) GSE72970 stage IV; (E) TCGA_COAD stage II + III; (F) TCGA_COAD stage IV.
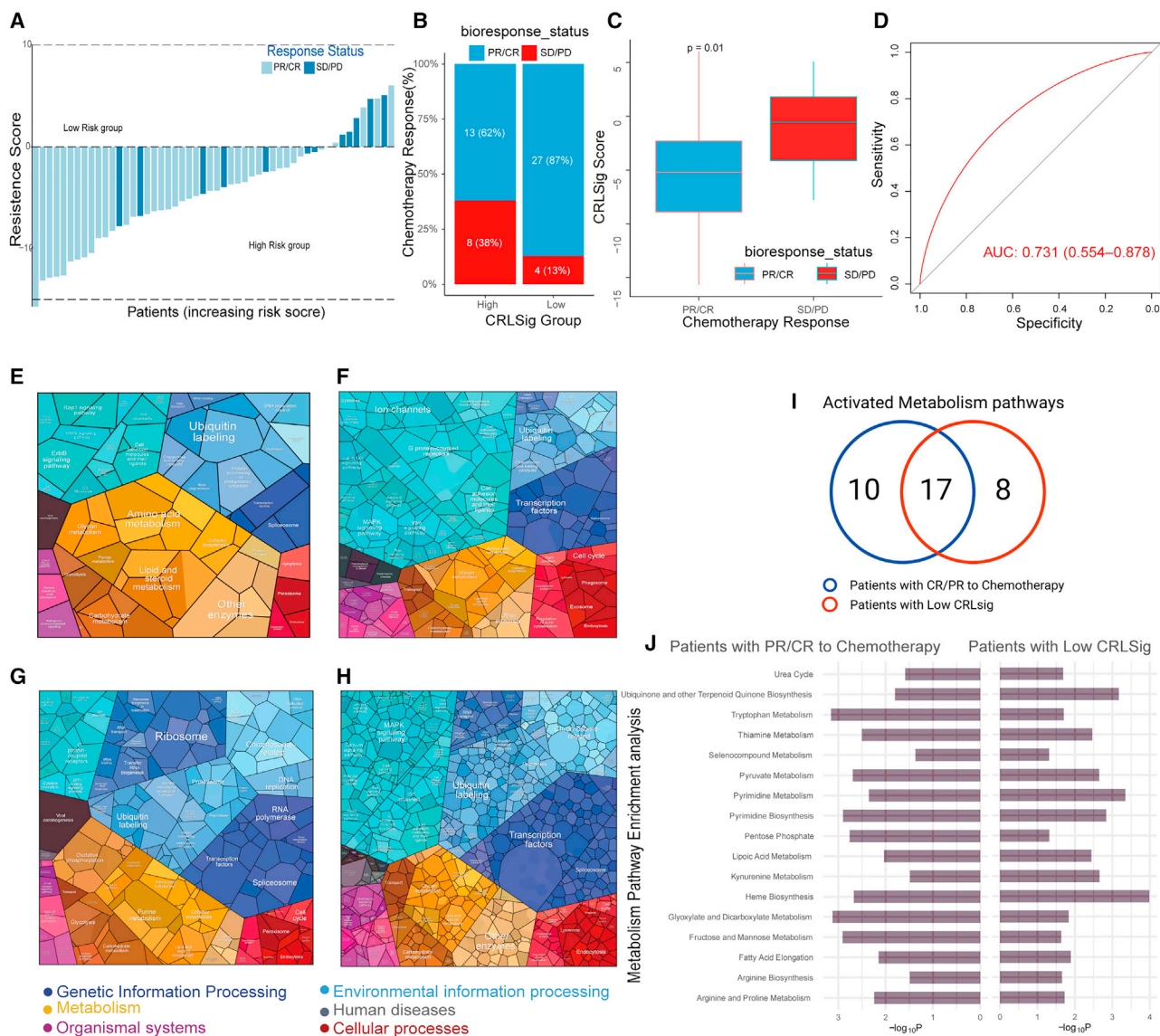
Figure 5. Prediction of CRLSig for treatment response and Proteomaps pathway analysis of chemotherapy patients

(A–C) The relationship between CRLSig and disease state in response to chemotherapy are shown as a bar plot, a histogram, and a boxplot. (D) Receiver operating characteristic curve of CRLSig for predicting the PR/CR status in response to chemotherapy. (E–H) Online Proteomaps pathway analysis of patients responding to chemotherapy (top left) and not responding to chemotherapy (top right). Low CRLSig (bottom-left) and high CRLSig (bottom right). Each small polygon corresponds to a single KEGG pathway, and the size correlates with the ratio between responder or low CRLSig (left) and non-responder or high CRLSig (right) subgroups. (I–J) Metabolism pathways activated in the patients with PR/CR to chemotherapy and the patients with low CRLSig, respectively.

Pan-cancer analysis for CRLSig in 18 TCGA cohorts with chemotherapy

To validate the application value of CRLSig in other cancer patients undergoing chemotherapy, we performed the Cox regression analysis for CRLSig in 18 types of pan-cancer patients with chemotherapy, all of whom were selected according to treatment data, including COAD. Patients with PAAD (pancreatic adenocarcinoma) (HR = 2.13; 95% CI, 1.04–4.37) and STAD (stomach adenocarcinoma) (HR = 2.86; 95% CI, 1.21–6.76) could also be divided into two groups with signif-

icantly different prognosis by CRLSig, respectively (Figures S6A–S6C). The prognostic values of CRLSig of patients from pan-cancer cohorts are shown in Tables S8 and S9.

Microenvironment characteristics for CRLSig

We applied the xCell R package to TCGA cohort and generated an abundance of immune-infiltrating cells of colon cancer patients with chemotherapy. Natural killer T (NKT) cells were significantly related to CRLSig (logFC > 1, p < 0.05). Other immune cells, including
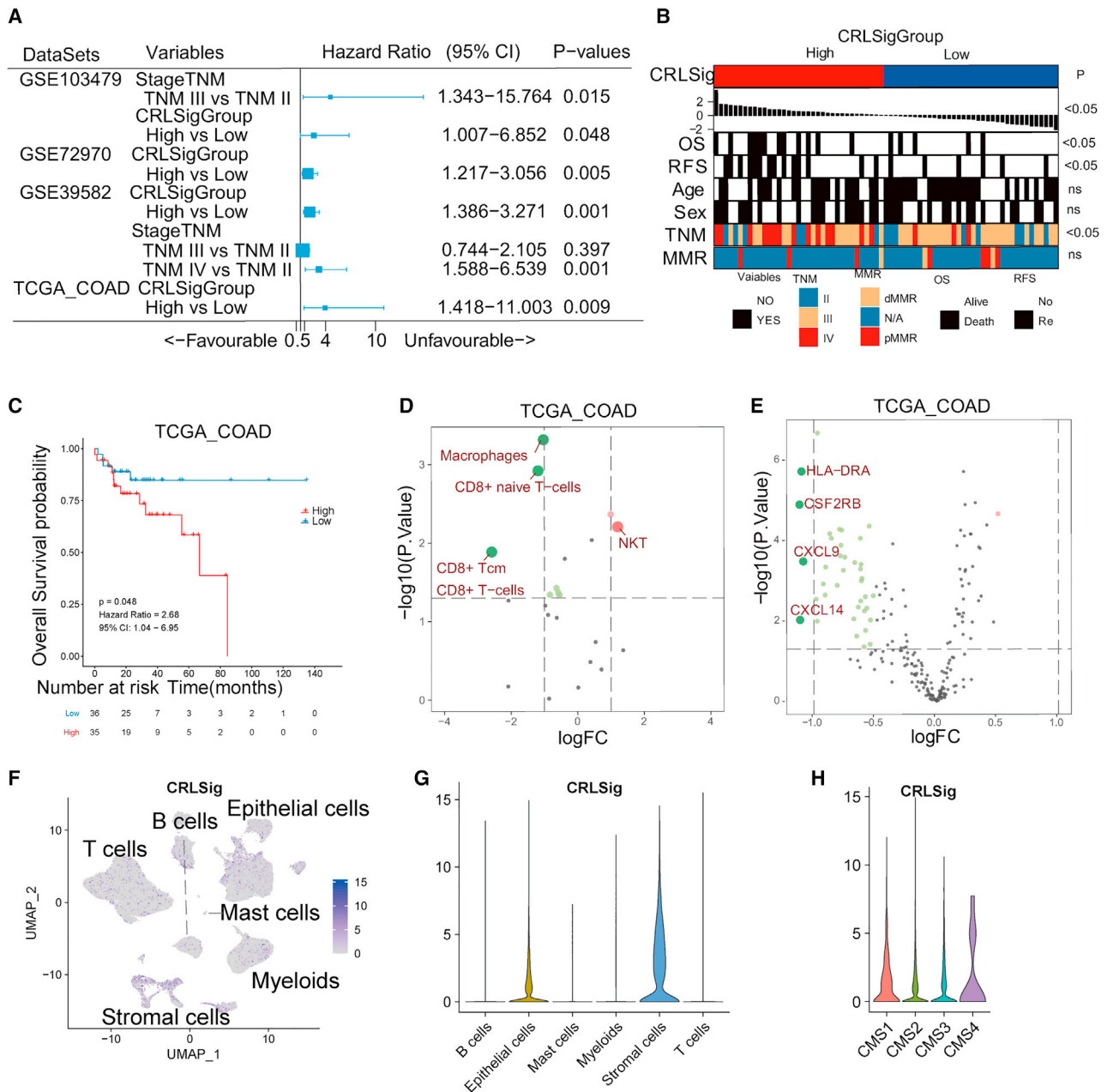
Figure 6. Independence of CRLSig from other clinicopathological factors and its clinical features

(A) The prognostic value of CRLSig and TNM stage in these cohorts. (B) The relationship of CRLSig with other clinicopathological factors in the TGCA cohort. (C) Kaplan-Meier survival curves of overall survival probability of CRLSig in TCGA validation cohort. (D–E) Gene set enrichment analysis for immune-infiltrating cells, chemokines, and cytokines of colon cancer chemotherapy patients with low and high CRLSig in TCGA cohort. (F–G) Single-cell sequencing analysis for CRLSig in different cell types (GSE132465). (H) Consensus molecular subtypes (CMS) subtypes analysis of colon cancer in tumor epithelial cells (GSE132465).

macrophages, CD8+ naive T cells, CD8+ T cells, and CD8+ Tcm were negatively related with CRLSig (logFC < −1, p < 0.05) (Figure 6D). Some chemokine and cytokine receptor genes, such as CSF2RB, CXCL14, and CXCL9, were also negatively related to CRLSig (Figure 6E, logFC < −0.5, p < 0.05). Detailed results were shown in Table S10.

Furthermore, we calculated the CRLSig in 65,362 cells in the single-cell sequencing data from GSE132465 and found that CRLSig values were higher in stromal cells compared with other cells (Figures 6F and 6G). Consensus molecular subtypes (CMS) provided a biologically rational stratification. To explore the relationships between CRLSig with CMS subtypes in single-cell levels, we used the previous

scRNA-seq data[25] to find that the CMS1 and CMS4 subtypes had higher CRLSig than other CMS groups in tumor epithelial cells (Figure 6H).

DISCUSSION

The standard of colon cancer care for most patients currently consists of surgery followed by systemic chemotherapy. In recent years, owing to the rapid development of high-throughput technology, the number of molecular characteristics correlated to patients with colon cancer has continuously increased. At the same time, a parallel increase in the application of molecular biomarkers to guide prognostication and clinical strategies management has occurred, including MSI, BRAF mutations, RAS mutations, HER2 overexpression, and kinase fusions.[9] The complexity of molecular profiles is challenging for the stratification of patients and biomarker-guided therapy. To solve the current dilemma of molecular biomarkers guiding patients' prognosis for receiving chemotherapy, we identified and constructed the expression profiles of chemotherapy-related lncRNAs in patients with colon cancer. In addition, we established an auxiliary prognostic tool, named CRLSig, and constructed a risk scoring system. Furthermore, we found that chemotherapy patients with high CRLSig scores had a poor prognosis compared with patients with low scores, which indicated that chemotherapy patients with high scores did not really benefit from chemotherapy. In fact, these patients may undertake additional physical and economic burdens. The construction of CRLSig may identify a patient subset that manifests high risk of RFS, despite completing standard adjuvant chemotherapy treatment, thus providing clinicians with strategies for the precise screening of patients able to benefit from the delivery of personalized cancer medicine. Our novel CRLSig tool uses lncRNA data to predict chemotherapy outcomes for precise optimization of drug selection and treatment regimen for colon cancer patients.

Consequently, in our study, we identified and validated an lncRNA signature, consisting of 55 lncRNAs that could distinguish between high- and low-risk groups in colon cancer patients with adjuvant chemotherapy. First, we integrated three GEO training datasets to determine the prognostic lncRNAs of patients with and without adjuvant chemotherapy using meta-analysis. Second, we further recognized prognostic lncRNAs only identified in patients with adjuvant chemotherapy as candidate chemotherapy-related lncRNAs. Then, CRLSig was constructed to identify patients who would benefit from adjuvant chemotherapy based on the 55 chemotherapy-related lncRNAs. Finally, we validated the prognostic value of CRLSig in TCGA cohort and compared it with other clinicopathological factors and MSI status.

Owing to a low risk of recurrence and a lack of treatment benefit, colon cancer patients with stage II and MSI-H or dMMR were not recommend for administration of 5-FU-based adjuvant chemotherapy.[7,8] At the same time, for patients with proficient DNA mismatch repair (pMMR) and stage II, a study performed by Sargent et al. revealed that no treatment benefit was present for surgery combined with adjuvant chemotherapy versus surgery alone.[26] In addi-tion, with the higher risk of recurrence, patients with stage III should receive adjuvant chemotherapy regardless of MSI status. Previous studies revealed a significant benefit from chemotherapy for MSS tumors,[26–28] while no clear conclusion was reached for MSI tumors owing to a substantial degree of heterogeneity.[29] However, owing to the heterogeneity of patients, all MSS or MSI patients have good and poor prognosis after chemotherapy. Unfortunately, few studies have been performed to identify the prognostic molecular characteristics of chemoresistance among patients with chemotherapy based on microsatellite status. Our study revealed that no significant prognostic difference was observed between pMMR and dMMR chemotherapy patients with stage III, while CRLSig could effectively stratify these patients. Patients with high CRLSig had a high risk of recurrence, while patients with low CRLSig had a prolonged RFS. For patients with stage II, neither MMR status nor CRLSig could identify them as having good prognosis following chemotherapy. When analyzing patients with stage II–III, we could achieve the same results as those with stage III patients. The above results suggested that CRLSig was superior to microsatellite status relative to the prediction of chemotherapy prognosis. In addition, our multivariate Cox analysis revealed that CRLSig was independent of other clinicopathological factors.

The lncRNA, as a regulator, can modulate gene expression on multiple levels and play a crucial role in cancer prognosis and tumorigenesis,[30] showing the possibility for building a prognostic signature in patients with colon cancer. Although emerging research has identified a substantial fraction of functional lncRNAs, a huge number of lncRNAs still have unknown functions.[31] Among 55 lncRNAs used to build CRLSig, 17 were demonstrated to be related to various cancers. Furthermore, 5 lncRNAs, including KCNQ1OT1, BLACAT1, HAGLR, HOXB-AS3, and RUNX1-IT1, had been previously reported to be linked with colon cancer. Li et al.[32] reported that KCNQ1OT1 could directly combine with miR-34a, thereby upregulating the expression of Atg4B and enhancing the chemoresistance of oxaliplatin in colon cancer. A previous study found that BLACAT1, as a cell-cycle regulator, could repress p15 expression by binding to EZH2 and further contribute to the cell proliferation of colorectal cancer.[33] HAGLR (Hoxd antisense growth-associated lncRNA) was reported to play a crucial role in gut development.[34] Moreover, Sun et al.[35] reported that HAGLR could sponge miR-185-5p and activate CDK4 and CDK6 to promote the growth, invasion, and migration of colon cancer. Interestingly, previous research found that the lncRNA HOXB-AS3 could encode a conserved 53-amino acid peptide and that the loss of this peptide played a critical role in oncogenic events and metabolic reprogramming in patients with colon cancer.[36] Shi et al.[37] found that RUNX1-IT1 played a tumor-suppressive role by inhibiting proliferation and migration and promoting apoptosis of colorectal cancer cells.

Gene set enrichment analysis for immune cells revealed that low CRLSig had more infiltrating CD8+ T cells and macrophages, while high CRLSig had more infiltrating NKT cells. In addition, chemokine and cytokine receptors, such as CXCL14, CXCL9, and CSF2RB, were

mainly expressed in patients with low CRLSig. 5-FU-based chemotherapeutics have a greater effect on tumor cells in a state of high metabolism and proliferation. Similarly, pathway analysis revealed that the low CRLSig group had more activation of metabolism-related pathways compared with the high CRLSig group, indicating a better response to chemotherapy. CMS defined four groups of colorectal cancers: CMS1-MSI/immune; CMS2-epithelial/canonical; CMS3-epithelial/metabolic; and CMS4-mesenchymal/stromal.[9] Single-cell sequencing analysis revealed that stromal cells had the highest CRLSig compared with other cell types and that higher CRLSig was associated with CMS1-MSI/immune and CMS4-mesenchymal/stromal cells, all indicating the possible pathogenic role of stromal cells in chemotherapy resistance. Furthermore, the CMS2-epithelial/canonical and CMS3-epithelial/metabolic groups of colon cancers related to metabolic pathways had low CRLSig and, thus, might lead to low risk of recurrence and better response to chemotherapy.

Limitations presented in our study should be considered. First, as our study was performed based on public datasets, some clinical information of patients was unavailable and thus could have led to potential bias. Second, owing to the discrepancy and limitation of the sequencing platform, the lncRNA signature identified in training datasets may not exactly match the validation set, potentially leading to inevitable evaluation bias. Finally, as mentioned above, the identified lncRNAs still required experimental verification and further validation in more cohorts.

In conclusion, we constructed a CRLSig to provide reliable and effective prognosis in colon cancer patients and guidance for clinical decision making with respect to patients benefiting the most from adjuvant chemotherapy. Moreover, CRLSig is an independent predictive marker of RFS and showed superior predictive ability compared with that of microsatellite status. In the future, more patients would benefit from the clinical use of CRLSig.

MATERIALS AND METHODS
Study design and main outcome
In brief, to construct the CRLSig of patients with colon cancer we performed a two-phase design. For the initial phase of lncRNAs identification, we systematically used the fixed-model of meta-analysis to confirm the relationships of lncRNAs expression with the prognosis of the colon cancer patients with or without chemotherapy from multicenter cohorts. Then we obtained two prognostic lncRNA expression sets, one in patients with chemotherapy and the other in patients without chemotherapy, respectively. Next, we performed a cross-selection process whereby we selected significantly poor prognostic lncRNAs identified in patients with chemotherapy, but not in non-chemotherapy patients. Based on these data, we constructed CRLSig to differentiate between chemoresistant and chemosensitive populations among colon cancer patients with chemotherapy. A validation phase followed in which we applied CRLSig to validate predictive ability in the validation set. To feature the grouped patients, we also matched and compared using characteristics of other clinicopathological factors, such as microsatellite status, immune cells, and

some specific expression of chemokine or cytokine receptors. RFS rate was regarded as the main outcome in our study.

Data collection
All transcription profiles and detailed clinical information were downloaded from TCGA (https://portal.gdc.cancer.gov) and GEO (http://www.ncbi.nlm.nih.gov/geo), using the following selection criteria: (1) have the overall/RFS data of each cohort; (2) have the information of chemotherapy and other therapies of each cohort; (3) have sample size of more than 50 patients; and (4) have clinical data, such as AJCC TNM stage, age, and sex. A total of 1,171 patients with colon cancer were enrolled in our study, including 715 patients without chemotherapy and 456 patients with chemotherapy by filtering. Patients with chemotherapy from GSE103479 (n = 60), GSE39582 (n = 240), and GSE72970 (n = 83) were selected as the training cohorts to identify the prognostic lncRNAs of chemotherapy. Then, 71 patients with only chemotherapy from the TCGA_COAD cohort were recruited as the external validation cohort. Patients without chemotherapy from GSE103479 (n = 70), GSE39582 (n = 326), and TCGA_COAD (n = 319) were used as comparison groups. The pan-TCGA cohorts were downloaded from the UCSC Xena website (https://xenabrowser.net/), and all the chemotherapy patients were selected according to the treatment information of their clinical features. In addition, the 65,362 cells with single-cell sequencing data of the SMC cohort (GSE132465) in colon cancer were analyzed by the "Seurat" packages in R. The CMS subtypes for epithelial cells in scRNA-seq data also obtained from SMC cohort.[25] All data generated or analyzed during this study are freely available in previous publications or the public domain.

Preprocessing of extraction of lncRNAs from the transcription profiles
All transcripts of microarray datasets from different platforms were matched with each GPL annotation file. For GSE39582 and GSE72970 the probes were from the Affymetrix HG-U133_Plus 2.0 microarray, while for GSE103479 the probes were from the Almac Diagnostics Custom Xcel array. Finally, by matching the GENCODE (release 25) and the RefSeq (release 79), we obtained a total of 2,456 unique lncRNAs in the GEO and TCGA cohorts, which were similar to those in the previous publication.[38] For the GEO cohort, expression of all lncRNAs was log2-transformed for further analysis. For TCGA cohort, we used the FPKM format of the sequenced lncRNA expression profiles (Illumina HiSeq platform) on TCGA website.

Construction of CRLSig
To better distinguish among chemoresistant populations, we constructed CRLSig based on the coefficients of each chemotherapy-related lncRNA identified in our study, as

$$CRLSig = \sum_{i}^{j} \beta i * \exp i,$$

where i and j represent the sequence and total number of chemotherapy-related lncRNAs, βi is the coefficient of corresponding

lncRNA, and exp*i* is the normalized expression of the lncRNA in the corresponding cohort.

Gene set enrichment analysis for immune cells

To evaluate microenvironment of immune-infiltrating cells of patients with colon cancer, gene set enrichment analysis was carried out with a total of 10,783 genes, called "xCell," to assess 24 tumor-infiltrating cell types in the specific cohort, according to normalized data.

Pathway enrichment analysis and Proteomaps visualization

Significant differentially expressed genes (DEGs) were calculated using the "limma" package in R with adjusted p values between the two groups in our study. Here, Proteomaps, a bionic visualization method of all bulk pathways for *Homo sapiens* (https://bionic-vis.biologie.uni-greifswald.de/), was generated by DEGs. Then, we set up Proteomaps for different groups in our study, such as the high/low CRLSig group and the CR/PD group to assess the dynamic change of pathways. The Proteomaps could be divided into six parts, including "Genetic information processing," "Metabolism," "Organismal systems," "Environmental information processing," "Human Disease," and "Cellular processes," presenting different proportions of each part of specific patients. We further collected 113 metabolic pathways from a previous study[24] and performed GSVA for them in specific patients.

Statistical analysis

Statistical analysis of all clinical data was performed in R 4.0. Standard tests included Student's t test, Wilcoxon rank-sum test, and Fisher's exact test. The Benjamini-Hochberg method (FDR) was used to adjust the p values for multiple comparisons. The relationship between CRLSig and other continuous variables was calculated using the Spearman method. The log rank test, and univariate and multivariate Cox proportional-hazards regression were used to analyze any related independent predictors of prognosis in colon cancer. Time-dependent ROC curve and ROC analysis were used to detect the prognostic value and the chemoresistance of CRLSig for colon cancer patients with chemotherapy. All reported p values were two-sided, and statistical significance was set at 0.05.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2021.12.015.

AUTHOR CONTRIBUTIONS

Conception and design, H.W., Y.G., L.W., and J.Z.; financial support, L.W.; provision of study materials, H.W. and Y.G.; collection and assembly of data, all authors; data analysis and interpretation, Y.G., H.W., S.V., and Q.Y.; manuscript writing, H.W. and Y.G.; manuscript supervision, L.W. and J.Z.; final approval of manuscript, all authors.

REFERENCES

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *68*, 394–424.

2. Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., et al. (2004). Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. New Engl. J. Med. *350*, 2335–2342.

3. Heinemann, V., von Weikersthal, L.F., Decker, T., Kiani, A., Vehling-Kaiser, U., Al-Batran, S.E., Heintges, T., Lerchenmuller, C., Kahl, C., Seipelt, G., et al. (2014). FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. Lancet Oncol. *15*, 1065–1075.

4. Venook, A.P., Niedzwiecki, D., Lenz, H.J., Innocenti, F., Fruth, B., Meyerhardt, J.A., Schrag, D., Greene, C., O'Neil, B.H., Atkins, J.N., et al. (2017). Effect of first-line chemotherapy combined with cetuximab or bevacizumab on overall survival in patients with KRAS wild-type advanced or metastatic colorectal cancer: a randomized clinical trial. Jama *317*, 2392–2401.

5. Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. N. Engl. J. Med. *372*, 2509–2520.

6. Overman, M.J., Lonardi, S., Wong, K.Y.M., Lenz, H.J., Gelsomino, F., Aglietta, M., Morse, M.A., Van Cutsem, E., McDermott, R., Hill, A., et al. (2018). Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. J. Clin. Oncol. *36*, 773–779.

7. Labianca, R., Nordlinger, B., Beretta, G.D., Mosconi, S., Mandala, M., Cervantes, A., Arnold, D., and Group, E.G.W. (2013). Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. *24*, vi64–72.

8. Sepulveda, A.R., Hamilton, S.R., Allegra, C.J., Grody, W., Cushman-Vokoun, A.M., Funkhouser, W.K., Kopetz, S.E., Lieu, C., Lindor, N.M., Minsky, B.D., et al. (2017). Molecular biomarkers for the evaluation of colorectal cancer: guideline from the American Society for clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and the American Society of clinical Oncology. J. Clin. Oncol. *35*, 1453–1486.

9. Sveen, A., Kopetz, S., and Lothe, R.A. (2020). Biomarker-guided therapy for colorectal cancer: strength in complexity. Nat. Rev. Clin. Oncol. *17*, 11–32.

10. Yothers, G., O'Connell, M.J., Lee, M., Lopatin, M., Clark-Langone, K.M., Millward, C., Paik, S., Sharif, S., Shak, S., and Wolmark, N. (2013). Validation of the 12-gene colon cancer recurrence score in NSABP C-07 as a predictor of recurrence in patients with stage II and III colon cancer treated with fluorouracil and leucovorin (FU/LV) and FU/LV plus oxaliplatin. J. Clin. Oncol. *31*, 4512–4519.

11. Tie, J., Cohen, J.D., Wang, Y., Christie, M., Simons, K., Lee, M., Wong, R., Kosmider, S., Ananda, S., McKendrick, J., et al. (2019). Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. JAMA Oncol. *5*, 1710–1717.

12. Sinicrope, F.A., Shi, Q., Allegra, C.J., Smyrk, T.C., Thibodeau, S.N., Goldberg, R.M., Meyers, J.P., Pogue-Geile, K.L., Yothers, G., Sargent, D.J., et al. (2017). Association of DNA mismatch repair and mutations in BRAF and KRAS with survival after recurrence in stage III colon cancers: a secondary analysis of 2 randomized clinical trials. JAMA Oncol. *3*, 472–480.

13. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

14. Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

15. Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K.M., Chang, A.N., Fahl, S.P., et al. (2017). Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. Cell *171*, 103–119.e118.

16. Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2020). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol 22, 96–118.

17. Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147, 358–369.

18. Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146, 353–358.

19. Bhan, A., Soleimani, M., and Mandal, S.S. (2017). Long noncoding RNA and cancer: a new paradigm. Cancer Res. 77, 3965–3981.

20. Chen, S., and Shen, X. (2020). Long noncoding RNAs: functions and mechanisms in colon cancer. Mol. Cancer 19, 167.

21. Zhou, M., Hu, L., Zhang, Z., Wu, N., Sun, J., and Su, J. (2018). Recurrence-associated long non-coding RNA signature for determining the risk of recurrence in patients with colon cancer. Mol. Ther. Nucleic Acids 12, 518–529.

22. Lin, Y., Pan, X., Chen, Z., Lin, S., and Chen, S. (2020). Identification of an immune-related nine-lncRNA signature predictive of overall survival in colon cancer. Front. Genet. 11, 318.

23. Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., and Milo, R. (2014). Visual account of protein investment in cellular functions. Proc. Natl. Acad. Sci. U S A 111, 8488–8493.

24. Rosario, S.R., Long, M.D., Affronti, H.C., Rowsam, A.M., Eng, K.H., and Smiraglia, D.J. (2018). Pan-cancer analysis of transcriptional metabolic dysregulation using the Cancer Genome Atlas. Nat. Commun. 9, 5330.

25. Lee, H.O., Hong, Y., Etlioglu, H.E., Cho, Y.B., Pomella, V., Van den Bosch, B., Vanhecke, J., Verbandt, S., Hong, H., Min, J.W., et al. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat. Genet. 52, 594–603.

26. Sargent, D.J., Marsoni, S., Monges, G., Thibodeau, S.N., Labianca, R., Hamilton, S.R., French, A.J., Kabat, B., Foster, N.R., Torri, V., et al. (2010). Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. J. Clin. Oncol. 28, 3219–3226.

27. Jover, R., Zapater, P., Castells, A., Llor, X., Andreu, M., Cubiella, J., Balaguer, F., Sempere, L., Xicola, R.M., Bujanda, L., et al. (2009). The efficacy of adjuvant chemotherapy with 5-fluorouracil in colorectal cancer depends on the mismatch repair status. Eur. J. Cancer 45, 365–373.

28. Ribic, C.M., Sargent, D.J., Moore, M.J., Thibodeau, S.N., French, A.J., Goldberg, R.M., Hamilton, S.R., Laurent-Puig, P., Gryfe, R., Shepherd, L.E., et al. (2003). Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. New Engl. J. Med. 349, 247–257.

29. Guastadisegni, C., Colafranceschi, M., Ottini, L., and Dogliotti, E. (2010). Microsatellite instability as a marker of prognosis and response to therapy: a meta-analysis of colorectal cancer survival data. Eur. J. Cancer 46, 2788–2798.

30. Goodall, G.J., and Wickramasinghe, V.O. (2021). RNA in cancer. Nat. Rev. Cancer 21, 22–36.

31. Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. 22, 96–118.

32. Li, Y., Li, C., Li, D., Yang, L., Jin, J., and Zhang, B. (2019). lncRNA KCNQ1OT1 enhances the chemoresistance of oxaliplatin in colon cancer by targeting the miR-34a/ATG4B pathway. OncoTargets Ther. 12, 2649–2660.

33. Su, J., Zhang, E., Han, L., Yin, D., Liu, Z., He, X., Zhang, Y., Lin, F., Lin, Q., Mao, P., et al. (2017). Long noncoding RNA BLACAT1 indicates a poor prognosis of colorectal cancer and affects cell proliferation by epigenetically silencing of p15. Cell Death Dis. 8, e2665.

34. Zakany, J., Darbellay, F., Mascrez, B., Necsulea, A., and Duboule, D. (2017). Control of growth and gut maturation by HoxD genes and the associated lncRNA Haglr. Proc. Natl. Acad. Sci. U S A 114, E9290–E9299.

35. Sun, W., Nie, W., Wang, Z., Zhang, H., Li, Y., and Fang, X. (2020). Lnc HAGLR promotes colon cancer progression through sponging miR-185-5p and activating CDK4 and CDK6 in vitro and in vivo. OncoTargets Ther. 13, 5913–5925.

36. Huang, J.Z., Chen, M., Chen, D.E., Gao, X.C., Zhu, S., Huang, H., Hu, M., Zhu, H., and Yan, G.R. (2017). A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. Mol. Cell 68, 171–184.e176.

37. Shi, J., Zhong, X., Song, Y., Wu, Z., Gao, P., Zhao, J., Sun, J., Wang, J., Liu, J., and Wang, Z. (2019). Long non-coding RNA RUNX1-IT1 plays a tumour-suppressive role in colorectal cancer by inhibiting cell proliferation and migration. Cell Biochem. Funct. 37, 11–20.

38. Sun, J., Zhang, Z., Bao, S., Yan, C., Hou, P., Wu, N., Su, J., Xu, L., and Zhou, M. (2020). Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. J. Immunother. Cancer 8, e000110.