Noninvasive urinary protein signatures associated with colorectal cancer diagnosis and metastasis

Yulin Sun
Cancer Hospital, Chinese Academy of Medical Sciences https://orcid.org/0000-0001-5232-710X

Zhengguang Guo
Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College

Zongpan Jing
Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

Jun Li
Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

Lijun Yang
Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

Xiaoyan Liu
Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences

Meng Cai
Cancer Hospital of CAMS

Zhaoxu Zheng
Cancer Hospital of CAMS

Chen Shao
State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Lifeomics, National Center for Protein Sciences (The PHOENIX Center, Beijing)

Yefan Zhang
Cancer Hospital of CAMS

Haidan Sun
Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/School of Basic Medicine, Peking Union Medical College

Li Wang
Cancer Hospital of CAMS

Yue Han
Cancer Hospital of CAMS

Shuangmei Zou
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Jiajia Gao

Cancer Hospital of CAMS
Yan Sun
Cancer Hospital of CAMS
Yan Zhao
Cancer Hospital of CAMS
Peng Nan
Chinese Academy of Medical Sciences and Peking Union Medical College
Xiufeng Xie
Cancer Hospital of CAMS
Lusong Tian
Cancer Hospital of CAMS
Fang Liu
Cancer Hospital of CAMS
Lanping Zhou
Cancer Hospital of CAMS
Wei Sun
Institute of Basic Medical Sciences, CAMS
Xiaohang Zhao (zhaoxh@cicams.ac.cn)
Cancer Hospital of CAMS https://orcid.org/0000-0003-3668-2712

Article

Keywords: colorectal cancer (CRC), biomarkers, metastasis, diagnosis

Posted Date: December 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-124929/v1

Noninvasive urinary protein signatures associated with colorectal cancer diagnosis and metastasis

Short title: Urinary protein signatures of colorectal cancer

Yulin Sun[1]†, Zhengguang Guo[2]†, Zongpan Jing[1], Jun Li[1], Lijun Yang[1], Xiaoyan Liu[2], Meng Cai[1], Zhaoxu Zheng[3],Chen Shao[4], Yefan Zhang[5], Haidan Sun[2], Li Wang[6], Yue Han[7], Shuangmei Zou[8], Jiajia Gao[1], Yan Sun[1], Yan Zhao[1], Peng Nan[1], Xiufeng Xie[1], Lusong Tian[1], Fang Liu[1], Lanping Zhou[1], Wei Sun[2]*, Xiaohang. Zhao[1]*

[1]State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.
[2]Core Facility of Instruments, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.
[3]Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.
[4]State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing, China.
[5]Department of Hepatobiliary Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Beijing 100021, China.
[6]Department of Clinical Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.
[7]Department of Interventional Therapies, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, China.
[8]Department of Pathology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, China.

* Correspondence and requests for materials should be addressed to W. S. (email: sunwei@ibms.pumc.edu.cn) and X. Z. (email: zhaoxh@cicams.ac.cn).

†These authors contributed equally to this work.

Abstract

Currently, imaging and serum carcinoembryonic antigen (CEA) tests are not sufficient for early detection and evaluation of metastasis and recurrence in colorectal cancer (CRC). To comprehensively discover and validate more accurate noninvasive biomarkers in urine, we adopted a staged discovery-verification-validation pipeline in 657 urine and 665 tissue samples from healthy controls and CRC patients with a distinct metastatic risk. The diagnostic signature generated combined with serum CEA levels revealed a significantly increased sensitivity (+34.8%) compared to CEA alone. Moreover, over 50% of CEA-negative metastatic patients were correctly predicted by the metastatic signature generated for metastatic risk. The tissue validation and functional study showed that CORO1C was associated with distant metastasis and enhanced the invasion and metastasis of CRC cells via a novel integrin/FAK/SRC and relevant pathways. Our findings provide novel urinary protein biomarkers and potential interventional targets to reliably detect CRC, especially in patients with metastatic CRC.

Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer death globally, accounting for approximately 1 in 10 cancer cases and deaths[1]. It was estimated that the global burden of CRC will increase by 60% to more than 2.2 million new cases and 1.1 million cancer deaths by 2030[2].

Notably, stage at diagnosis which is mainly defined by regional lymph node metastasis and distant metastasis is the prognostic factor most directly related to survival and recurrence of CRC. For example, the 5-year relative survival of CRC patients was 90% for patients with localized disease, 71% for patients with regional disease, and 14% for patients with distant spread[3]. However, approximately 35% and 20% of patients with newly diagnosed CRC present with regional and synchronous distant metastases, respectively[3]. Moreover, approximately 25% of patients will develop distant liver metastases 6 months after the primary tumor is diagnosed (metachronous). Therefore, metastasis, including distant and high-risk lymph node spread, is the most important prognostic factor for survival in patients with CRC.

In the clinic, imaging and serum carcinoembryonic antigen (CEA) testing play central roles in monitoring the recurrence and metastasis of CRC. Computed tomography (CT) has a sensitivity and specificity of 51% and 85% for local lymph nodes, 62% and 92% for distant lymph nodes and approximately 71-73.5% and 96% for liver metastases detection[4, 5]. Magnetic resonance imaging (MRI) is superior to CT and has a sensitivity of 39-95% for lymph node metastases and 91-97% for liver metastases[5]. However, these mainstream imaging modalities also have some pitfalls, including high cost, poor detection of small lymph nodes or lesions (< 1 cm), and unsuitable for patients with implants or impaired renal function[5] By contrast, the sensitivity of serum CEA for patients with stage I-IV disease and recurrence is 4-11%, 25-30%, 38-44%, 65% and 50-71%, respectively[6], with an overall specificity of approximately 70%[7]. Therefore, the current single surveillance strategy is not sufficient to evaluate metastasis and recurrence. The urgent need for more accurate and noninvasive biomarkers remains in the clinic.

Urine is a source for discovering early and sensitive biomarkers, because it can rapidly reflect changes in the body[8]. Moreover, its protein composition is significantly less complex than that of serum or plasma, thus urine is a good sample for biomarker analyses[8]. Our and other previous studies identified more than 8000 proteins in human urine[9]. Approximately 40% of urinary proteins originate from plasma proteins, and over 1800 proteins that are highly expressed in the colon can be detected in the urine of healthy individuals[9]. Therefore, it is feasible to identify noninvasive biomarkers for CRC in urine.

In previous studies, fragments of fibrinogen, hepcidin-20 and β2-microglobulin were found to be discriminative in urine between 76 cancer patients and 72 noncancer patients using SELDI-TOF and MALDI-TOF methods[10]. Urinary cysteine-rich protein 61 (Cyr61) and trefoil factor 3 (TFF3) could yield a diagnostic capacity with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.75 using ELISA in 176 CRC patients and healthy controls[11]. A series of studies discovered and validated naturally occurring peptides (NOPs) in urine to discriminate CRC liver metastases from healthy controls using an LC-MS/MS method[12]. Hydroxylated collagen peptide (AGP) and two additional NOPs derived from collagen alpha 1(I) and collagen alpha-1(III) that were measured by multiple reaction monitoring and parallel reactive monitoring (PRM) approaches, respectively, were found to complement serum CEA to improve the detection of CRC liver metastases (15-20% increase in sensitivity)[13, 14]. Previous

study provided useful information on urinary biomarkers for detecting CRC. However, a comprehensive study, including discovery, validation and verification in a large-scale cohort, is still unavailable until now.

In the present study, we adopted a staged pipeline to develop urinary protein signatures of CRC for diagnosis and metastatic risk prediction in large-scale cohorts using urine and tumor tissue samples (Fig. 1). The performance of the signatures was evaluated and compared with that of serum CEA. Finally, the expression of key urinary protein was validated in tissue specimens, and the function was investigated *in vivo* and *in vitro*. This stepwise study yielded highly accurate noninvasive urinary protein signatures and will benefit the application of urinary proteomics to CRC research in the future.

Results

Clinical characteristics of urine specimens

A total of 657 subjects, including healthy controls (HCs) and CRC patients without metastases (NM), with lymph node metastasis (LNM), and with distant liver metastasis (DM), were recruited in this study. After removing 105 individuals (Fig. 1), 552 qualified individuals were included for subsequent analyses. The detailed clinical information is shown in Table 1. The age and sex distributions were balanced among the HCs and three cohorts of CRC patients. Histological differentiation grade, tumor location, CEA level and TNM staging were significantly different among the three CRC patient cohorts. Notably, there were basically no statistically significant differences in these clinical parameters among the three subgroups within each cohort.

Discovery stage

Differential urinary proteomic analysis using the TMT approach

First, CRC urinary protein candidates were discovered by the TMT labeling-2D-LC-MS/MS approach. By the criteria with a 1% false discovery rate (FDR) at both the peptide and protein levels, 2291, 2642, 3364 and 3182 proteins were identified in the four groups, respectively . The

median technical and interindividual CV was 7.3% and 22.2% in the four groups, respectively. By excluding the proteins with a technical CV>30%, and with an interindividual CV>60% in each group, a total of 1976, 2151, 2635 and 2772 proteins were quantified in the four groups, respectively (Supplementary Table 1A-1D).

Unsupervised principal component analysis (PCA) analysis of 995 common proteins in the four groups was performed to visualize the urinary protein profiling differences among the HCs and patients with different stages of CRC. The results suggested apparent discrimination between the HC and CRC groups; however, the three CRC groups could not be clearly distinguished (Supplementary Fig. 1B). In the orthogonal partial least squares discriminant analysis (OPLS-DA) model, the four groups could be clearly separated (Fig. 2A). One hundred permutation tests indicated no overfitting of the models (Supplementary Fig. 1C).

Furthermore, to show the urinary proteomic pattern of the four groups, unsupervised clustering of the average protein quantitation value for each group was performed. The results indicated that HCs clustered into one branch, while within the three CRC groups, NM showed a different pattern from that of LNM and DM (Fig. 2B). The above clustering revealed distinct urinary proteomic patterns of the four groups.

Functional annotation of the differential urinary proteins

The pairwise differential urinary proteins between NM, LNM or DM and HC were defined using a criterion of fold change≥1.5; thus, a total of 273, 337 and 355 proteins were found, respectively (Supplementary Table 1E). By Ingenuity Pathway Analysis (IPA) of differential proteins, a pathway-pathway interaction diagram was generated by connecting all pairs of interacting pathways/diseases/functions. These differential proteins were enriched in tumor-related pathways, including tumor growth, tumor invasion, immune response, metabolism and signaling (RAC, FAK, CDC42, and RhoA) pathways (Fig. 2C, Supplementary Table 1F-1G).

Using the same criteria, the differential proteins within three stages of CRC were defined in each pairwise comparison, and 93, 69 and 114 differential proteins were found, respectively (Supplementary Table 1H). IPA showed that tumor invasion-, immune response-, hemostasis-,

angiogenesis-, and metabolism-related pathways/functions were enriched and tightly connected (Fig. 2D, Supplementary Table 1I-1J).

Furthermore, in IPA of disease and biofunction analysis, cell death- and apoptosis-related proteins were increasingly inhibited, whereas the proteins in tumor proliferation, migration and protein metabolism modules were gradually activated along with the development and progression of CRC. The immune response module was only activated only in the early stage of CRC (NM group), while the tumor invasion-related proteins were activated only in the late stage of CRC (DM group) (Fig. 2E, Supplementary Table 1K).

The canonical pathway analysis revealed that the tumor metabolism (glycolysis I)-related pathways and the tumor survival-related pathways (AKT/PI3K pathway) were activated in CRC. Tumor invasion-related pathways, such as RAC, FAK, CDC42, and RhoA pathways, were increasingly activated along with the progression of CRC (Fig. 2F, Supplementary Table 1L, detailed pathway shown in Fig. 2G). Collectively, the above results indicated that urinary proteomics could reflect the enhanced tumor growth and malignancy status of CRC as well as the tumor invasion status in metastatic CRC.

Verification stage

Verification of differential urinary proteins using PRM-based targeted MS

The key differentially expressed proteins (41 proteins) were verified by the PRM approach in 82 independent samples (Fig. 1). A pooled urine mixture was used as a quality control (QC) to evaluate the system stability during the experimental process. The average Pearson's correlation coefficient of QC samples was 0.99, indicating the repeatability of the QC samples and the stability of the MS platform (Supplementary Fig. 2A).

As a result of the PRM assay, 66 peptides from 41 differential proteins were successfully validated with consistent trends with the TMT approach in the four groups (Supplementary Table 2A and 2B). The heatmap and scatter plots of the 41 proteins (Fig. 3A, Supplementary Fig. 3 and 4) showed that 18 proteins were remarkably downregulated and 23 proteins were upregulated in

CRC. Twelve upregulated proteins exhibited gradually increasing trends with CRC progression (Supplementary Table 2B).

Generating urinary protein signatures for CRC diagnosis and metastatic risk prediction

Multilevel machine learning analysis was used to define urinary protein signatures for CRC based on 23 upregulated proteins. First, to obtain complementary biomarker combinations, 23 proteins were evaluated by Spearman's rank correlation. Nine proteins with a moderately high correlation with more than five other proteins (rho $\geq$ 0.6) were excluded, and the 14 remaining proteins with less interdependency (median correlation coefficient of 0.34) were selected for subsequent analyses (Supplementary Fig. 2B).

Next, we evaluated these 14 proteins as input variables and identified the most important features in the diagnostic model and metastatic model using the random forest algorithm (Fig. 3B). Meanwhile, we measured the classification performance of each protein in the diagnosis and metastatic risk prediction using ROC analysis (Fig. 3C). The proteins with the top 10 highest values in the two models were chosen as candidate classifiers. Therein, 8 common proteins showed good performance for CRC diagnosis or metastatic risk prediction. Furthermore, the complementary performance of any two proteins was evaluated by comparing the combined AUC value for the diagnosis and metastatic risk prediction (Fig. 3D, 3E).

Finally, a urinary protein signature for CRC diagnosis consisted of CORO1C, APRC5 and RAD23B, and a classifier for CRC metastasis consisted of CORO1C, RAD23B, GSPT2 and NDN. IPA showed that these five proteins formed a tightly functional protein interaction network, and are involved in digestive system cancer, tumorigenesis and tumor migration (Supplementary Fig. 2C).

Urinary protein signature performances for CRC diagnosis and metastatic risk prediction

The above diagnostic signature achieved 80.0% specificity and 86.0% sensitivity with an AUC of 0.89 (Fig. 3F, Supplementary Table 2C). In addition, the metastatic signature yielded a sensitivity, specificity and AUC of 81.1%, 70.0% and 0.78, respectively, in discriminating between CRC patients with and without metastases (Fig. 3G, Supplementary Table 2D).

To evaluate diagnostic signature performance in discriminating the NM, LNM and DM groups from the HC group, three pairwise comparisons produced AUCs of 0.800, 0.948, and 0.935, respectively (Supplementary Table 2C, Supplementary Fig. 2D), showing better power for metastatic stages.

The performance of the metastatic signature in discriminating the three CRC groups was also analyzed in pairwise comparisons. The signature differentiated the NM group from the LNM and DM groups with AUCs of 0.723 and 0.827, respectively (Supplementary Table 2D, Supplementary Fig. 2E).

Validation stage

Independent validation of urinary protein signatures using an immunoassay

To validate urinary signatures to distinguish HCs and CRC patients on a large scale, we developed a quantitative dot blot detection system using urine[15] according to previous methods used with serum [15, 16]. The urinary protein amount was quantified by standard curves of each protein and then calibrated by the corresponding urine creatinine measurement.

A total of 434 urine samples were recruited in the validation stage. The concentrations of CORO1C, APRC5, RAD23B, GSPT2 and NDN were significantly higher in the urine of patients with CRC than in the urine of HCs (P<0.0001 for all; Fig. 4A-4E). In the three CRC groups, the levels of these five urinary proteins showed a gradient with an increasing trend that correlated with disease progression, achieving the highest levels in DM (P<0.01 for all; Fig. 4A-4E). Moreover, RAD23B and GSPT2 concentrations were significantly higher in patients with LNM than in patients with NM (P<0.05 for both).

Furthermore, we found that the abundance of these five proteins in urine displayed significantly positive correlations with TNM staging and M staging. Moreover, the abundance of GSPT2 and NDN was higher in patients with lymph node metastasis. Intriguingly, urine CORO1C and APRC5 levels were significantly greater in the patients with the BRAF-V600E mutation than in those without this mutation (Supplementary Fig. 5A, 5B). In addition, the

abundances of CORO1C, GSPT2 and NDN were not age- and sex-dependent in HCs; however, APRC5 and RAD23B were age-dependent in HCs (P<0.05; Supplementary Fig. 5C).

Performance of urinary protein diagnostic and metastatic signatures

In the 434 urine samples detected by immunoassay, serum CEA measurements were available in 312 samples, including samples from 154 HCs and 158 CRC patients. To facilitate the comparison of CEA results, these 312 samples were used to evaluate the performance of our urinary protein diagnostic and metastatic signatures.

For the diagnostic model, in the training set (CRC vs. HCs, n=208), the signature (CORO1C, APRC5, and RAD23B) had the AUC, sensitivity, specificity and accuracy of 0.83, 63.8%, 92.2% and 77.9%, respectively (Supplementary Fig. 6A, Supplementary Table 3A). In the validation set (51 HCs vs. 53 CRC, total n=104), the signature achieved the AUC, sensitivity, specificity, and accuracy of 0.87, 79.3%, 86.3% and 82.7%, respectively (Supplementary Fig. 6B, Supplementary Table 3A). Moreover, the discriminative capacity of the diagnostic signature was strengthened from the NM, LNM or DM group versus the HC group, yielding the AUCs of 0.80, 0.81, and 0.91, respectively (Supplementary Fig. 6C-6E).

For the metastatic model, the signature consisting of CORO1C, RAD23B, GSPT2 and NDN was used to distinguish metastatic and nonmetastatic CRC. Due to the small sample size of the NM group, 27 NM samples were duplicated twice, and a total of 81 NM and 78 metastatic CRC samples were used to train the model. The other 14 NM and 39 metastatic CRC samples were used as the validation set. Finally, the signature yielded the AUCs, sensitivity, specificity and accuracy of 0.77, 57.7%, 92.6% and 75.5% for the training set and 0.68, 48.7%, 92.9% and 60.4% for the validation set, respectively (Supplementary Fig. 6F-6G, Supplementary Table 3B). Moreover, the discriminative power of the metastatic signature was higher for DM versus NM (0.83) than for LNM versus NM (0.71) (Supplementary Fig. 6H-6I).

Urinary protein signatures complemented serum CEA

For comparison, the clinical CRC biomarker, CEA, was also measured in 312 serum samples. At the most commonly used threshold of 5 ng/mL in the clinic, its sensitivity, specificity and

accuracy were 48.7%, 95.4% and 71.8%, respectively, for diagnosing CRC (n=312). When it was used to predict metastatic risk, the sensitivity, specificity and accuracy were 58.1%, 78.0% and 63.3%, respectively (n=158).

In the diagnostic model, the combination of serum CEA and urinary proteins in the diagnostic signature achieved better diagnostic capability with AUCs, sensitivity, specificity and accuracy values of 0.93, 81.0%, 89.3% and 85.1% in the training set and 0.89, 88.7%, 80.4% and 84.6% in the validation set, respectively (Fig. 4F, Supplementary Fig. 6J, Supplementary Table 3A). In addition, in the stratified discrimination of NM, LNM, and DM patients from HCs, the combination yielded AUCs of 0.83, 0.93, and 0.96, respectively (Supplementary Table 3A, Supplementary Fig. 6K-6M). In the CRC patients in the training and validation cohorts, CEA was positive in 56 (53.3%) and 21 (39.6%) subjects, and the diagnostic signature increased the diagnostic power in an additional 30 (28.6%) and 18 (34.0%) subjects, respectively. For CEA-negative (CEA<5 ng/ml) patients, 61.2% (30 out 49 CRC patients) of patients in the training set and 56.3% (18 out of 32 CRC patients) of patients in the validation set were correctly diagnosed by the diagnostic signature (Fig. 4H).

In the metastatic model, similar to the results of the diagnostic model, the combination of the urinary protein signature with serum CEA had a better prediction with AUC, sensitivity, specificity and accuracy values of 0.85, 73.1%, 92.6% and 83.0% in the training set and 0.74, 82.1%, 64.3% and 77.4% in the validation set, respectively (Fig. 4G, Supplementary Fig. 6N, Supplementary Table 3B). Moreover, in the stratified discrimination of LNM and DM from NM, the combination yielded AUCs of 0.78 and 0.91, respectively (Supplementary Table 3B, Supplementary Fig. 6O-6P). In the patients with metastatic CRC in the training and validation cohorts, CEA was positive in 45 (57.7%) and 20 (51.3%) patients, and the metastatic signature increased the diagnostic power in an additional 19 (24.4%) and 13 (33.3%) patients, respectively. Moreover, 57.6% (19 out of 33) and 58.4% (13 out of 19) of the CEA-negative patients with metastatic CRC were correctly predicted to have metastases using the panel in the training and validation sets, respectively (Fig. 4I).

To visualize the urinary protein performance on CRC diagnosis and metastatic risk prediction, with a specificity of 90%, the cutoff values of each protein and signatures were used

in the training and validation urine samples. In the diagnostic model, the combined signature with serum CEA led to a sensitivity of 77.1% in the training set and 77.4% in the validation set, respectively (Supplementary Table 3C, Fig. 4J, Supplementary Fig 6Q). In the metastatic model, the combined signature with serum CEA yielded a sensitivity of 75.6% in the training set and 82.1% in the validation set, respectively (Supplementary Table 3D, Fig. 4K, Supplementary Fig. 6R). The panel could achieve higher sensitivity than a single protein. In addition, the panel combined with serum CEA provided a higher positive rate in the patients with advanced stage CRC (DM group) than in the patients with intermediate stage CRC (LNM group) in both models.

CORO1C is overexpressed in CRC tissues and associated with advanced stages

As the shared urinary protein in both the diagnostic and metastatic signatures, CORO1C expression was assayed in 658 tumor and 591 adjacent nontumor tissues from 665 subjects with colon adenocarcinoma using immunohistochemistry (Fig. 4L).

CORO1C was barely detectable in the nontumor colon tissues, and the positive rate was only 2.72% (17/625). However, positive cytoplasmic immunostaining for CORO1C was observed in 41.64% (274/658) of tumors. The weak and strong staining was found in 66.79% (183/274) and 33.21% (91/274) of cases, respectively. Overall, CORO1C was dramatically overexpressed in CRC (P<0.0001; Fig. 4M).

Subsequent clinical significance analysis of CORO1C expression showed that higher CORO1C levels were associated with microvascular invasion/perineural invasion (P<0.0001), tumor depth of invasion (T staging, P=0.0416), distinct metastasis (M staging, P=0.0097) and advanced AJCC staging (P=0.0383; Fig. 4N, 4O).

CORO1C enhances the migration and invasion capacities of CRC cells *in vitro*

We further investigated the roles of CORO1C in the tumorigenesis. First, the expression levels of CORO1C were assessed in a panel of seven CRC cell lines. CORO1C was found to vary across different cell lines, with HCT8 showing a relatively high expression of CORO1C and the rest showing relatively low levels of CORO1C (Supplementary Fig. 7A). Subsequently, we used *in vitro* systems to ectopically express CORO1C in DLD1 or HCT116 cells and to silence

CORO1C in HCT8 cells (Supplementary Fig. 7B and 7C). CORO1C had no significant influence on colony formation, cell cycle progression or cell apoptosis (Supplementary Fig. 7D-7I). However, its overexpression dramatically enhanced the migration, invasion and adhesion capacities of CRC cells (Fig. 5A-5C, Supplementary Fig. 7J-7L), while the depletion of CORO1C suppressed these phenotypes (Fig. 5D-5F). Taken together, these results indicated that CORO1C functionally enhances the invasive and metastatic potential of CRC cells.

CORO1C promotes tumor growth of CRC *in vivo*

Subsequently, we explored the *in vivo* roles of CORO1C in mouse xenograft models of CRC. CORO1C overexpression significantly strengthened tumor proliferation in mice bearing subcutaneous DLD1 tumors. By the study end point, both tumor volume and tumor weight in the CORO1C overexpression group had at least a two-fold increase compared with those in the control group (Fig. 5G-5I). By contrast, in the HCT8 xenograft model, CORO1C knockdown resulted in an approximately 50% reduction in tumor weight and tumor volume by the study endpoint compared with those in the control group (Fig. 5J-5L).

Tumor tissues in the CORO1C overexpression group or knockdown group showed an obvious upregulation or downregulation of CORO1C compared to the expression of CORO1C in the corresponding control groups (Fig. 5M and 5N). Moreover, the percentage of Ki67 stained cells in tumor tissues showed similar trend to CORO1C expression.

CORO1C is elicited by the activation of integrin/FAK/SRC and downstream pathways

To further investigate the mechanism of CORO1C facilitating cell migration and invasion as well as *in vivo* proliferation in CRC, CORO1C-dependent downstream signaling molecules were explored using Western blot analysis (Fig. 5O). We observed that the expression of integrin αV, integrin β1, ARHGEF7, RCC2 and RAC1 and phosphorylated FAK, SRC, AKT, p70S6K, 4EBP1, JNK and c-Jun levels were markedly increased along with CORO1C overexpression, whereas the above molecules were obviously reduced after CORO1C depletion. Consequently, the activities of MMP9 and MMP2, which are the target genes of JNK and c-Jun, significantly increased or decreased along with the expression levels of CORO1C in cells (Fig. 5P).

Furthermore, we proposed a model to explain how CORO1C provoked cell migration, invasion and *in vivo* proliferation in CRC (Fig. 5Q). CORO1C overexpression stimulates integrin/FAK/SRC signaling, leading to the activation of RAC1 and PI3K/AKT, which results in the activation of downstream effectors, including JNK/c-Jun, p70S6K and 4EBP1. Eventually, malignant phenotypes, such as increased invasive and metastatic potential, occur in CRC.

Discussion

Colorectal cancer is a highly heterogeneous disease. Recent genomic and transcriptomic analyses indicate that CRC metastasis, including lymph node metastasis and distant metastasis, is partly mediated through a polyclonal seeding mechanism in at least one-third of patients [17, 18]. Therefore, regional spread and distant metastases of CRC reflect the inherent characteristics of primary tumors, and both share a similar genetic basis but show significant differences.

Since the acquisition of metastatic capacity is an early event in tumorigenesis, it provides a possibility for predicting regional or distant dissemination based on primary tumor tissues or human body fluids. However, high heterogeneity of primary tumors may result in inaccurate results based on transcriptomic and proteomic profiling of primary tumors, especially for biopsy specimens. Thus, body fluids that reflect the general change in pathophysiological status may serve as good sources for finding metastasis-associated biomarkers.

In this study, we systematically analyzed urinary proteins for the diagnostic and prognostic prediction of CRC in combination with quantitative proteomics, targeted proteomics and immunoassays. To date, this is the largest and most comprehensive study to determine noninvasive biomarkers of CRC. As for the specificity of our diagnostic and metastatic signatures for CRC rather than other tumors, it needs further evaluation.

In the discovery stage, a series of CRC related differential proteins were identified. Functional analysis showed that the urinary proteome could reflect the features that are according to the hallmarks of cancer, including sustaining cell proliferation, resisting cell death, reprogramming energy metabolism etc.[19]. In addition, the urinary proteome also reflected the characteristics of CRC in different stages. For example, we found that the immune response module was activated only in early stage CRC (NM group), which is in line with the results of a previous study at the tissue level[20]. Importantly, tumor invasion-related proteins and pathways represented by RAC,

FAK, CDC42 and RhoA signaling were enriched only in the DM group, demonstrating unique characteristics of metastatic CRC.

Previous studies (Supplementary Table 1M-1O) identified 5 diagnostic and 2 metastatic proteins of CRC in urine [10, 11, 12, 13, 14]. In our study, using high-throughput TMT methods, a total of 583 diagnostic and 223 metastatic proteins were identified, 4 and 1 of which were consistent with those identified in previous studies, respectively. For CRC serum/plasma proteomics studies, 55 diagnostic and 13 metastatic proteins were identified in both this study and previous works. Notably, by comparing the 31 cancer-associated proteins identified by the CPTAC CRC project[21] with proteins founds in our study, 3 proteins showed similar expression trends, including S100P, CTHRC1 and S100A11. Therefore, our urinary proteomic analysis reflected the changes in CRC tumor tissues.

At the verification and validation stages of this study, a panel of three urinary biomarkers (CORO1C, APRC5, and RAD23B) for CRC diagnosis, and a panel of four urinary biomarkers (CORO1C, RAD23B, GSPT5, and NDN) for CRC metastatic risk prediction were defined. Among these proteins, CORO1C and ARPC5 are known regulators of actin cytoskeleton dynamics, and we will discuss them in-depth later. RAD23B is involved in the nucleotide excision repair of damaged DNA, and its abnormal expression has been found in breast cancer[22]. GSPT2 is a GTPase that mediates translation termination and has been reported to be a biomarker for hepatocellular carcinoma and CRC liver metastasis in serum[23, 24]. NDN is a member of the melanoma-associated antigen family and serves as a candidate tumor suppressor gene to facilitate the entry of the cell into cell cycle arrest in multiple tumors, including CRC[25].

Notably, our urinary diagnostic and metastatic signatures showed superior performance to the conventional biomarker CEA used in the clinic. First, our diagnostic signature showed superior performance for patients with early-stage CRC (NM group). In the training cohort, our diagnostic signature was positive in 63.0% of patients (17/27) versus serum CEA at 29.6% (8/27), whereas in the validation set, they were 50.0% (7/14) versus 7.1% (1/14), respectively. Further analyses with total early-stage, intermediate-stage (LNM group), advanced-stage (DM group) and all CRC cases from all cohorts revealed that the diagnostic signature combined with serum CEA increased the diagnostic sensitivity to 70.7% (29/41), 82.3% (51/62), 94.5% (52/55) and 83.5% (132/158), respectively. Thus, our urinary diagnostic signature is a potent biomarker panel for the detection of early-stage CRC with better accuracy than serum CEA alone.

Next, the main goal of the urinary metastatic model was to increase the sensitivity of CEA to identify lymph node metastasis and distant metastasis more accurately in patients with a medical record of CRC. The combination of CEA and urinary metastatic signature compared with CEA alone increased the sensitivity from 47.6% to 61.9% (+14.3%) for LNM patients and from 69.4% to 86.1% (+16.7%) for DM patients in the training set. Moreover, in the validation set, the sensitivity increased from 42.9% to 81.0% (+38.1%) for the LNM group and from 61.1% to 83.3% (+22.2%) for the DM group. Therefore, the urinary metastatic signature combined with CEA identified lymph node metastasis, distant metastasis and all metastasis at 68.3% (43/63), 85.2% (46/54) and 76.1% (89/117), respectively. Further follow-up and prospective cohort studies are needed to evaluate whether our urinary metastatic signature could provide a lead time for the detection of recurrent CRC.

CORO1C belongs to the short coronins subfamily and is expressed ubiquitously. It can bind to actin, inactive RAC1 and the Arp2/3 complex[26, 27]. The Arp2/3 complex is a central actin nucleator that promotes branched filament formation and creates a complex cortical membrane actin network to generate the force necessary for protrusion[28]. Notably, the Arp2/3 complex contains seven protein subunits, and ARPC5 is one of them. In addition, CORO1C is necessary for the release of inactive RAC1 from the non-protrusive membrane and activation and redistribution of RAC1 to a protrusive tip; accordingly, the activation of RAC1 induces membrane ruffling and lamellipodia formation at the leading edge[26, 29]. CORO1C is recruited to endosome buds to facilitate the timing of endoplasmic reticulum (ER) recruitment for ER-associated endosome fission[30]. Consequently, CORO1C plays roles in wound healing, protrusion formation, cell proliferation, cytokinesis, endocytosis, axonal growth, secretion, migration and invasion[27, 31]. Overexpression of CORO1C has been found in multiple types of aggressive tumors, including glioma, gastric cancer, triple-negative breast cancer and nasopharyngeal carcinoma[32, 33, 34, 35]. High expression of CORO1C is an unfavorable prognostic factor in gastric cancer, triple-negative breast cancer and nasopharyngeal carcinoma[32, 33, 34]. To date, there have been no reports of CORO1C in CRC. We first found that CORO1C was overexpressed in colon adenocarcinoma but not in mucinous carcinoma (n=45, P=0.0741, data not shown), indicating distinct biological characteristics between mucous tumor cells and adenocarcinoma cells.

Mechanistically, our findings are in agreement with previous studies of other tumor types and show that CORO1C promoted the migration, invasion and *in vivo* growth of CRC cells. However,

we addressed a new mechanism of CORO1C in cell migration and invasion. We found that CORO1C upregulated the expression of integrin αv and β1, which are indicators of metastasis (Fig. 5O). This agrees well with the observation that CORO1C promoted CRC cell adhesion to fibronectin (Fig. 5C and 5F). Subsequently, high levels of integrins bound to fibronectin and activated FAK/SRC signaling via phosphorylation. FAK phosphorylated ARHGEF7 and thereby facilitated the targeting of activated Rac1 to focal adhesions. The downstream effectors, including AKT, JNK and mTOR, were also phosphorylated. Moreover, activated RAC1 is also an efficient activator of the JNK/c-Jun signaling cascade. The transcription factor c-Jun translocates to the nucleus and activates the transcription of target genes, such as MMP9. Furthermore, CORO1C can release RAC1 from the lateral membrane and traffic it to the leading edge[26, 29]. In this process, it also bound to RCC2, a competitive inhibitor of RAC1 activation, to regulate the redistribution of RAC1 by preventing off-axial protrusion[26]. Thus, CORO1C boots the invasion and metastasis of CRC cells via the newly discovered integrins/FAK/SRC relative signaling pathways (Fig. 5Q). Notably, we also observed that a high level of CORO1C in urine was associated with the BRAF-V600E mutation in CRC. Actually, as reported by studies in a mouse model, inhibition of FAK increased the efficacy of BRAF inhibitors in BRAF-mutant melanoma, while FAK inhibition alone had no effect on tumor growth[36]. Thus, integrin/FAK/SRC signaling was presumably activated in BRAF inhibitor resistant cells. CORO1C may contribute to the activation of cell adhesion-mediated pathways in BRAF-dependent tumors. Our study hereby provides novel therapeutic regimens and targets for metastatic CRC.

In conclusion, this study is currently the largest and most comprehensive urinary biomarker discovery study in CRC utilizing a discovery-verification-validation pipeline. The urinary diagnostic and metastatic signatures combined with serum CEA (≥5 ng/ml) improved the sensitivity by 34.8% (83.5% vs. 48.7%) and 20.6% (76.1% vs. 55.5%) compared with CEA alone, respectively. We also presented the mechanism of one signature protein, CORO1C, in CRC by *in vitro* and *in vivo* analyses. The above results showed that the urinary proteome could comprehensively reflect the pathophysiological changes in different CRC stages. Our findings provide not only novel, very promising urinary protein biomarkers to reliably diagnose and detect CRC, whether or not in combination with serum CEA but also, potential interventional targets for metastatic CRC.

Materials and Methods

Experimental Design

The objectives of the present study were to systemically identify and validate potential noninvasive diagnostic and metastatic predictive markers for CRC in urine. Therefore, three groups of CRC patients with distinct metastatic risk were recruited: CRC patients without any metastases, with lymph node metastasis and with distant metastasis. In addition, a four-stage workflow consisting of a series of mass spectrometry (MS) and immunoassay-based approaches including a TMT labeling-2D-LC-MS/MS quantitative proteomic strategy, PRM-based targeted proteomic method, quantitative dot blot analysis and tissue immunohistochemistry method, was used to build a diagnostic signature and a metastatic signature (Fig. 1). In addition, the functions of one candidate, CORO1C, were investigated *in vivo* and *in vitro*.

Patients and healthy controls

A total of 359 CRC patients were recruited from the Cancer Hospital, Chinese Academy of Medical Sciences from January 2015 to October 2018. All patients were pathologically diagnosed by two senior pathologists, and random morning midstream urine samples were collected prior to surgical operations or chemotherapy/radiotherapy. We excluded 26.7% of patients with post-operation; a pathological diagnosis with nontubular adenocarcinoma (mucinous adenocarcinoma, melanoma, signet ring carcinoma, neuroendocrine carcinoma); accompanied with other benign or malignant tumors; abnormal renal functions; receiving chemoradiotherapy; a failure of quality control of PRM (without signals in more than 40% of peptides) or dot blot analysis (CV>20%). The 263 qualified patients were divided into the CRC patients without metastases (NM, n=76), with lymph node metastasis (LNM, n=97) and with distant metastasis (DM, n=90) according to pathology report (Fig. 1).

In addition, 298 urine samples from healthy controls were obtained from the Health Medical Center of the Cancer Hospital and PLA General Hospital from August 2014 to October 2018. The enrollment criteria for healthy control subjects were as follows: (1) the absence of benign or malignant tumors; (2) a qualified physical examination finding no dysfunction of vital organs and (3) normal renal function and without albuminuria. Nine HCs were excluded for quality control of PRM (without signals in more than 40% of peptides) or dot blot analysis (CV>20%) (Fig. 1). Table 1 lists the demographic and clinical characteristics of the 552 patients and healthy

controls. There was no statistically significant difference in age or sex distribution among the four groups. After collection, urine samples were stored at -80°C. This study was approved by the Ethics Committee of Institute of Basic Medical Sciences and Cancer Hospital, Chinese Academy of Medical Sciences (#047-2019, Beijing, China) and was performed according to the Declaration of Helsinki Principles.

Sample preparation

Before use, the thawed urine samples were centrifuged in a thermostatic centrifuge for 45 min at 5000×g and 4°C and the supernatant was collected. For isobaric tandem mass tags (TMT) and parallel reaction monitoring (PRM) assays, urinary proteins were enriched via our developed urimem method with some modifications[37]. Briefly, 40 ml urine was diluted with 40 ml of 0.2 M Na_2HPO_4 buffer, and the mixture was passed through a nitrocellulose membrane (0.22 μm), which was placed onto a vacuum suction filter bottle (10 cm^2 filter area). After drying at 56°C in an oven, the protein-bound membrane was cut into small pieces and placed in a 2 mL tube, to which 1.7 mL of acetone and 250 μL of 0.5% NH_4HCO_3 were added. After 10 min of intense vortexing, the tube was then incubated at 4°C for 1 h, followed by centrifugation at 14,000×g for 15 min. The precipitates were collected and air-dried (30 min) at room temperature. Afterwards, 400 μL of lysis buffer (7 M urea, 2 M thiourea, 65 mM DTT and 82 mM Tris) was added to resuspend the pellets, followed by intense vortexing for 10 min. The sample was centrifuged at 3,500×g for 30 min at 4°C and the supernatant was collected and then quantified by the Bradford assay.

The protein was digested by the filter-aided sample prep (FASP) method. After digestion, peptides were extracted by a C18 extraction column and dried under vacuum. The peptide concentration was further quantified by the BCA method.

TMT labeling and 1D off-line separation

Nine randomly selected samples from the healthy control, NM, LNM and DM groups were individually labeled with the 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, and 130C 10-plex TMT reagent according to the manufacturer's protocol (Thermo Fisher Scientific, Waltham, MA, USA). A mixed sample from all four groups was labeled with 131 TMT reagents. The labeled samples from each group were mixed individually. The pooled mixture of

TMT-labeled samples was fractionated using a high-pH reversed-phase liquid chromatography (RPLC) column from Waters (4.6 mm × 250 mm, Xbridge C18, 3 μm). The samples were loaded onto the column in buffer A1 (H_2O, pH = 10). The elution gradient was 5-25% buffer B1 (90% acetonitrile, pH = 10; flow rate: 0.8 mL/min) for 48 min. The eluted peptides were collected at one fraction per minute. The dried 48 fractions were resuspended in 0.1% formic acid and pooled into 24 samples by combining fractions 1 and 25, 2 and 26, and so on.

LC-MS/MS analysis

The fractionated labeled samples were analyzed using a self-packed RP C18 capillary LC column (75 μm×100 mm, 1.9 μm). A total of 96 fractions from urinary peptide mixtures in four groups were analyzed by LC-MS/MS. The gradient was eluted in 5-30% buffer B1 (0.1% formic acid, 99.9% H_2O; flow rate: 0.3 μL/min) for 45 min. Each sample was run 3 times. LTQ Orbitrap Fusion Lumos MS (Thermo Fisher Scientific) was used to acquire raw data. MS data were acquired using the following parameters: top speed data dependent mode (3 s) per full scan, full scans acquired in Orbitrap at a resolution of 60,000, MS/MS scans were with 32% normalized collision energy in HCD mode at a resolution of 15,000, charge state screening (excluding precursors with unknown charge state or +1 charge state) and dynamic exclusion (exclusion size list 500, exclusion duration 30 s).

Database searching

The MS/MS spectra were searched against the SwissProt human database from the UniProt website (www.uniprot.org) using the Proteome Discoverer software suite (v2.1, Thermo Fisher Scientific). Trypsin was chosen as the cleavage specificity with a maximum number of allowed missed cleavages of two. Carbamidomethylation of cysteine and TMT 10-plex labels were set as the fixed modifications, and the oxidation of methionine, deamidation of asparagine and glutamine, carbamyl of lysine and the peptide N-terminus were set as the dynamic modifications. The searches were performed using a peptide tolerance of 20 ppm and a product ion tolerance of 0.05 Da. As a filter, a 1% false positive rate at the protein level was used, and each protein contained at least 1 unique peptide.

After filtering the results as described above, the peptide abundances in different reporter ion channels of the MS/MS scan were normalized. The protein abundance ratio was based on unique

peptide results. Proteins with a technical CV over 0.3 or an interindividual CV over 0.6 within each group were excluded. Proteins with a fold change ≥ 1.5 between the NM, LNM, or DM group and the control group were considered differential proteins.

PRM Analysis

The selected differential proteins were verified in 82 samples by PRM, and each sample was analyzed by schedule mode. To ensure the quality of the data, the analysis of the mixed sample is used as a QC to observe the stability of the instrument signal during the whole analysis process, before and after all samples, and among every 8-10 samples. The iRT standard peptide analysis was added to each sample, and the stability of the chromatographic retention time was observed during the analysis. Two technical repeats were run for each sample. To avoid system errors, different groups of samples were analyzed in random orders for mass spectrometry analysis.

Each sample was analyzed with a C18 RP self-packed capillary LC column (75 μm×100 mm). The eluted gradient was 5-30% buffer B2 (0.1% formic acid, and 99.9% ACN; flow rate: 0.3 μL/min) for 40 min. A TripleTOF 5600 mass spectrometer was used to analyze eluted peptides from the LC. The MS data were acquired using the high-sensitivity mode with the following parameters: PRM mode, full scans acquired at a resolution of 40,000 and MS/MS scans at a resolution of 20,000, rolling collision energy, charge state screening (including precursors with +2 to +4 charge state), MS/MS scan range of 100-1800 m/z, and scan time of 100 ms. Each sample was run three times.

PRM data analysis

PRM data processing was performed with Skyline 3.6 software. All the results were imported into Skyline, the correct peaks were selected manually, and all the peptide results in all samples were exported. The total ionic chromatography (TIC) of the +2-+5 charges of each sample was extracted by Progenesis software. The abundance of each peptide of each sample was normalized with the TIC of the respective sample to correct the sample loading amount and MS signal intensity. The PRM results including protein names and peak areas, were exported for further analysis and the differential proteins between different groups were screened and compared using the TMT results.

Generating predictors for colorectal cancer based on PRM data

To generate urinary protein biomarker panels that distinguish CRC patients from healthy controls (diagnostic model) as well as CRC patients with metastasis (including regional lymph node metastatic and distant metastatic patients) from those without metastases (metastatic model), the protein expression abundance obtained from PRM data was normalized following the methodology described in a previous study[38].

First, we calculated the correlation matrix of protein expression abundance by Spearman's rank correlation to measure the intercorrelation between peptides. To screen the classifiers, the proteins with low-expression similarity were used to calculate the importance of discriminating two classes in the diagnostic and metastatic models using the random forest algorithm. To minimize randomness, 100 random forests consisting of 150,000 trees were computed to generate averaged mean decrease accuracy values for each protein. Mean decrease accuracy values were averaged for each protein among the 100 random forest replicates. Next, the area under the receiver operating characteristic (ROC) curve (AUC) was calculated as a performance measure. The overlapping proteins with the top 10 highest mean decrease accuracy values and AUC values were used for subsequent analyses. That is, the AUC of the combination of any two proteins was computed. The representative proteins that showed the highest AUC and the strongest complementarity with other proteins were chosen as the most relevant features. The above analyses were performed using the R statistical environment.

Quantitative dot blot analysis of urinary proteins

Dot blot analysis of urinary protein was performed using a Whatman Minifold I 96-well dot-blot array system (GE Healthcare, Chicago, IL, USA) according to the manufacturer's instructions. Briefly, the PVDF membrane was first immersed in methanol for 20 min and then in PBS for 10 min. Then, the dot blot apparatus was assembled, and 500 μl of diluted urine samples or standards in PBS were loaded into each well. The recombinant human proteins for CORO1C (Cat. No. RY-02857), APRC5 (Cat. No. H00010092-P01), RAD23B (Cat. No. H00005887-P01), GSPT2 (Cat. No. H00023708-P01) and NDN (Cat. No. H00004692-P01) were purchased from RunYu BioTech. Inc. (Shanghai, China) and Abnova (Taiwan, China). Next, vacuum was applied to filter the sample through the PVDF membrane. Thereafter, the membrane was blocked with 10% skim milk in PBS and probed with primary antibodies against CORO1C (Cat. No.

H00023603-M02, Abnova), ARPC5 (Cat. No. sc-1666766, Santa Cruz Biotech., Dallas, TX, USA), RAD23B (Cat. No. A1034, ABclonal Technology, Woburn, MA, USA), GSPT2 (Cat. No. 12989-1-AP, Proteintech Group Inc., Rosemont, IL, USA) and NDN (Cat. No. sc-1001224, Santa Cruz Biotech). Following intensive washing, the membranes were developed using an enhanced chemiluminescence detection reagent (Thermo Fisher Scientific, Waltham, MA, USA) and visualized with the ImageQuant LAS4000 system (GE Healthcare, Chicago, IL, USA) with the intensity adjusted to avoid saturation of the spots.

Spot intensities were measured and corrected to the background with ImageJ software. The raw concentration of each sample was calculated by standard curves and then corrected by several samples that are common to each study and run on each blot. Additionally, urinary creatinine concentration was quantified using the Creatinine Parameter Assay Kit (R&D Systems, Minneapolis, MN, USA). The relative absorbance units of each protein were normalized to that of urinary creatinine excretion.

For performance analysis, samples used for dot blot analysis were divided into a training set (67% of data set) and validation set (33% of data set). For the diagnosis model, 208 patients were enrolled in the training set (103 HC vs. 105 CRC) and 104 in the validation set (51 HCs vs. 53 CRC patients). For the metastasis model, 105 patients were enrolled in the training set (27 NM vs. 78 LNM or DM CRC) and 53 in the validation set (14 NM vs. 39 LNM or DM CRC). To confirm the robustness of the metastasis model, the sample size of NM in the training set was duplicated twice (increased to 81 NM). The models were constructed using logistic regression by MedCalc 15. The ROC curves were plotted using SigmaPlot 14.0.

Cell lines, plasmids and antibodies

The human colorectal cancer cell lines DLD1, HCT116, SW480, HT29 and LS174T were purchased from the National Infrastructure of Cell Line Resource (Beijing, China). HCT8 and HCT15 were purchased from the American Type Culture Collection (Rockville, MD, USA). All cells were maintained in the recommended media at 37°C with 5% CO_2.

The constructs pReceiver-Lv157-CORO1C (NM_001276471), pReceiver-Lv157-control, psi-nU6-CORO1C shRNA and psi-nU6 shRNA scramble control were purchased from GeneCopoeia (Rockville, MD, USA). The nucleotide sequences of CORO1C shRNA and

scramble control were 5’-GCAAGTTTCGCATGTATT-3’ and 5’-GCTTCGCGCCGTAGTCTTA -3’, respectively.

Primary antibodies including integrin αV, integrin β1, FAK, SRC, AKT, ARHGEF7, p70S6K, 4EBP1, c-Jun, Ki67, phospho-FAK (Tyr397), phosphor-SRC (Tyr416), phosphor-AKT (Ser473), phosphor-p70S6K (Thr389), phosphor-4EBP1 (Thr37/46), phosphor-c-Jun (Ser63) and HA-tag were purchased from Cell Signaling Technology (Boston, MA, USA). Anti-Rac1 and anti-RCC2 were purchased from Abcam (Cambridge, MA, USA). Anti-JNK and anti-phospho-JNK were purchased from Santa Cruz Biotech (Dallas, TX, USA). Anti-CORO1C was purchased from Abnova. Anti-β-actin was purchased from Proteintech Groups Inc. (Chicago, IL, USA).

Transfection and Western blot analysis

The packaged lentivirus particles that included pReceiver-Lv157-CORO1C or pReceiver-Lv157-control plasmid were used to infect DLD1 and HCT116 cells, respectively. Seventy-two hours later, stable clones with significant CORO1C ectopic expression were selected by 0.2 or 0.8 mg/mL G418 for further studies. For CORO1C knockdown, psi-nU6-CORO1C and scramble control vectors were transfected into HCT8 cells with Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Forty-eight hours after transfection, stable clones were selected by 6 μg/ml puromycin.

Cells were lysed using a lysis buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 4% SDS, protease inhibitor cocktail (Roche, Germany) and phosphatase inhibitor cocktail (Bimake, Houston, TX, USA). Protein samples were separated by SDS-PAGE and Western blotting was performed according to the protocol of the dot blot mentioned above.

Colony formation assays

The colony formation assay was performed in 6-well plates in which 1,000 cells were seeded per well and cultured for 10-12 days. After fixation in 4% paraformaldehyde followed by staining with 0.5% crystal violet, colonies were counted manually.

Cell cycle and apoptosis assays

For the cell cycle assay, DLD1 and HCT8 cells were washed with PBS and incubated with serum-free medium for 24 h. Cell-cycle entry was stimulated by adding fresh complete medium with 10% FBS and incubated for 24 h. Then, harvested cells were slowly added to 4 ml of ice-cold 95% ethanol and incubated overnight at -20°C. The next day, the cells were centrifuged and 15 μl of propidium iodide (PI) solution was added. After incubation at 37°C in the dark for 30 min, samples were analyzed on a BD LSRII flow cytometer (Franklin Lakes, NJ, USA). The DNA content of the cells was analyzed for the percentage of cells in the G1, S and G2/M phase.

The apoptotic assay was carried out using the Annexin V-FITC/PI Apoptosis Detection Kit (NanJing KeyGen Biotech Co., Nanjing, China). In brief, DLD1 and HCT8 cells were trypsinized and centrifuged at 800 rpm for 5 min. After suspension in ice cold binding solution, 1×10^5 cells were stained with Annexin V-FITC for 5 min and then PI dye for 5 min in the dark, according to the manufacturer's instrument. Apoptosis events were analyzed by a BD LSRII flow cytometer.

Wound healing assay

Cells were seeded in a 6-well plate and cultured until confluent. Artificial wound tracks were generated by scraping the cells with a pipette tip. After removal of the detached cells by gently washing with PBS 3 times, the medium was replaced with serum-free medium. Images were taken at 0 and 48 h after scratching using a 10× objective. Four randomly selected wounded areas were measured for each condition.

Migration and invasion assays

For both assays, 30,000 DLD1 cells, 50,000 HCT116 cells or 10,000 HCT8 cells in serum-free media were added to the upper chamber; then, complete medium containing 10% FBS was added to the lower chambers as a chemoattractant. After 48 h of incubation, the upper chambers were stained with 0.5% crystal violet. The cells that invaded through the membrane were observed under a microscope and manually counted within five random 100× fields. The invasion assays were performed in the same protocol as the migration assay with the exception that the inserts were precoated with 30 μg of Matrigel (Corning Incorporated, NY, USA).

Cell adhesion assay

A 96-well plate was coated with 20 μg/ml human fibronectin (Millipore, Temecula, CA, USA) for 1 h at room temperature and then blocked with 10 mg/ml heat-inactivated bovine serum albumin (BSA) at room temperature for 30 min. DLD1, HCT116 and HCT8 cells were trypsinized and allowed to recover for 10 min. Next, they were seeded at a density of 10,000 cells/well and cultured for 30, 60 and 90 min. After removing unattached cells, Cell Counting Kit-8 (CCK-8) reagents were added to each well and allowed to react for 3 h at 37°C. The absorbance was detected at 450 nm in a Synergy H1 Multi-Mode Reader (Bio-Rad).

Immunohistochemistry staining

Ten tissue microarrays (TMAs) of colon cancer were purchased from Shanghai Outdo Biotech Co., Ltd (Shanghai, China) and SuperBiotek Co., Ltd (Shanghai, China). Among them, six tissue microarrays that contained 538 patients had recorded follow-up information.

After deparaffinization and rehydration, the TMAs were immersed in methanol containing 0.3% hydrogen peroxide for 10 min to block endogenous peroxidase. Heat-induced antigen retrieval was performed in a water bath for 30 min in 0.1 M sodium citrate buffer (pH 6.0). After washing, TMAs were incubated overnight with the anti-CORO1C (Cat No. TA349821; OriGene Technologies, Inc, Rockville, MD, USA) or anti-RAD23B antibody (Cat No. A1034; Abcolonal Inc., Wuhan, China) at 4°C. The staining was performed using the Prolink-2 Plus HRP rabbit polymer detection kit (Golden Bridge International Inc., Bothell, WA, USA) according to the manufacturer's instructions. The images were captured using Aperio ScanScope CS software (Vista, CA).

The results were evaluated separately by two independent pathologists. The staining intensity and area were quantified as described previously[39]. A staining index between 0 and 12 was achieved by multiplying of the extent of positivity and intensity. For CORO1C, a staining index was used in which 0-4 is considered negative, 5-8 is weakly positive and 9-12 is strongly positive.

Xenograft Tumor-Formation Assays

Five to six-week-old female BALB/c nude mice were purchased from Beijing Huafukang Bioscience Inc. (Beijing, China). A total of 5×10^6 DLD1 cells that stably expressed CORO1C or empty vector, or 3.5×10^6 HCT8 cells that stably knocked down CORO1C or scramble control

were suspended in 100 μl of PBS and subcutaneously injected into the back flank of nude mice (n=8 for each group). The mice were then monitored for tumor volume every 3-4 days and calculated by length×width^2×0.5. After approximately 4 weeks, animals were euthanized, and tumors were excised by dissection. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Cancer Hospital, Chinese Academy of Medical Sciences.

For immunohistochemistry staining, the xenografts were fixed in 10% formalin and embedded in paraffin. The slides were stained using antibodies against CORO1C and Ki67 as previously described. Images were captured using Aperio ScanScope CS software (Vista, CA, USA).

Zymography

The activities of gelatinases (MMP2 and MMP9) were assessed by gelatin zymography according to the standard protocol[40]. Briefly, serum-free media were collected from DLD1 and HCT8 cells transfected 24 h post-culture with either the pReceiver-Lv157-CORO1C or psi-nU6-CORO1C shRNA plasmid and then centrifuged at 1000 ×g for 10 min. The proteins in the medium were separated by 10% SDS-PAGE with 0.5 mg/mL gelatin. After electrophoresis, the gel was soaked in renaturing buffer for 30 min and then replaced with developing buffer for 30 min incubation. Then, the gel was transferred to a new developing buffer and incubated for 12 h at 37°C. After staining with Coomassie Brilliant Blue for 4 h with agitation, the gel was destained until bands could be seen.

Statistical analyses

Pattern recognition analysis (PCA and OPLS-DA) was performed using SIMCA 14.0 (Umetrics, Sweden) software. The unsupervised clustering was performed using the MetaboAnalyst tool (www.metaboanalyst.ca). Complete clustering with the Euclidean distance was performed on the group average protein quantitation data. Student's *t*-test, Mann-Whitney rank test and one-way ANOVA were used for statistical analyses of quantitative data with GraphPad Prism software (v7; San Diego, CA, USA). The other analyses previously mentioned are described above. Statistical significance was defined as a two-sided *P* value of less than 0.05.

References

1. Bray F., et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018).

2. Arnold M., et al. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683-691 (2017).

3. Siegel R. L., et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin* **67**, 177-193 (2017).

4. Kekelidze M., et al. Colorectal cancer: current imaging methods and future perspectives for the diagnosis, staging and therapeutic response evaluation. *World J Gastroenterol* **19**, 8502-8514 (2013).

5. Sahani D. V., et al. Current status of imaging and emerging techniques to evaluate liver metastases from colorectal carcinoma. *Ann Surg* **259**, 861-872 (2014).

6. Nicholson B. D., Shinkins B. & Mant D. Blood Measurement of Carcinoembryonic Antigen Level for Detecting Recurrence of Colorectal Cancer. *JAMA* **316**, 1310-1311 (2016).

7. Collins J. J. & Black P. H. Specificity of the carcinoembryonic antigen (CEA). *N Engl J Med* **285**, 175-176 (1971).

8. Casado-Vela J., et al. Human urine proteomics: building a list of human urine cancer biomarkers. *Expert Rev Proteomics* **8**, 347-360 (2011).

9. Zhao M., et al. A comprehensive analysis and annotation of human normal urinary proteome. *Sci Rep* **7**, 3024 (2017).

10. Ward D. G., et al. Proteomic profiling of urine for the detection of colon cancer. *Proteome Sci* **6**, 19 (2008).

11. Shimura T., et al. Urinary Cysteine-Rich Protein 61 and Trefoil Factor 3 as Diagnostic Biomarkers for Colorectal Cancer. *Translational oncology* **12**, 539-544 (2019).

12. Broker M. E., et al. Collagen peptides in urine: a new promising biomarker for the detection of colorectal liver metastases. *PLoS One* **8**, e70918 (2013).

13. Lalmahomed Z. S., et al. Hydroxylated collagen peptide in urine as biomarker for detecting colorectal liver metastases. *Am J Cancer Res* **6**, 321-330 (2016).

14. van Huizen N. A., et al. Identification of a Collagen Marker in Urine Improves the Detection of Colorectal Liver Metastases. *J Proteome Res* **19**, 153-160 (2020).

15. Yang L., et al. Effects of a 4-hour staying at room temperature on urinary proteome analysis. *Basic & Clinical Medicine* **39**, 1330-1334 (2019).

16. Tan L. D., et al. Serum HER2 level measured by dot blot: a valid and inexpensive assay for monitoring breast cancer progression. *PLoS One* **6**, e18764 (2011).

17. Ulintz P. J., et al. Lymph Node Metastases in Colon Cancer Are Polyclonal. *Clin Cancer Res* **24**, 2214-2224 (2018).

18. Naxerova K., et al. Origins of lymphatic and distant metastases in human colorectal cancer. *Science* **357**, 55-60 (2017).

19. Hanahan D. & Weinberg R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

20. Pancione M., et al. Immune escape mechanisms in colorectal cancer pathogenesis and liver metastasis. *J Immunol Res* **2014**, 686879 (2014).

21. Vasaikar S., et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **177**, 1035-1049 e1019 (2019).

22. Linge A., et al. Identification and functional validation of RAD23B as a potential protein in human breast cancer progression. *J Proteome Res* **13**, 3212-3222 (2014).

23. Li M., et al. eRF3b, a biomarker for hepatocellular carcinoma, influences cell cycle and phosphoralation status of 4E-BP1. *PLoS One* **9**, e86371 (2014).

24. Zhu D., et al. Predicting metachronous liver metastasis from colorectal cancer using serum proteomic fingerprinting. *J Surg Res* **184**, 861-866 (2013).

25. De Faveri L. E., et al. Putative tumour suppressor gene necdin is hypermethylated and mutated in human cancer. *British Journal of Cancer* **108**, 1368-1377 (2013).

26. Williamson R. C., et al. Coronin-1C and RCC2 guide mesenchymal migration by trafficking Rac1 and controlling GEF exposure. *J Cell Sci* **127**, 4292-4307 (2014).

27. Rosentreter A., et al. Coronin 3 involvement in F-actin-dependent processes at the cell cortex. *Exp Cell Res* **313**, 878-895 (2007).

28. Suetsugu S. & Takenawa T. Regulation of cortical actin networks in cell migration. *Int Rev Cytol* **229**, 245-286 (2003).

29. Williamson R. C., et al. Coronin-1C Protein and Caveolin Protein Provide Constitutive and Inducible Mechanisms of Rac1 Protein Trafficking. *J Biol Chem* **290**, 15437-15449 (2015).

30. Hoyer M. J., et al. A Novel Class of ER Membrane Proteins Regulates ER-Associated Endosome Fission. *Cell* **175**, 254-265 e214 (2018).

31. Xavier C. P., et al. Phosphorylation of CRN2 by CK2 regulates F-actin and Arp2/3 interaction

and inhibits cell migration. *Sci Rep* **2**, 241 (2012).

32. Castagnino A., et al. Coronin 1C promotes triple-negative breast cancer invasiveness through regulation of MT1-MMP traffic and invadopodia function. *Oncogene* **37**, 6425-6441 (2018).

33. Cheng X., et al. CORO1C expression is associated with poor survival rates in gastric cancer and promotes metastasis in vitro. *FEBS Open Bio* **9**, 1097-1108 (2019).

34. Fan L., Wei Y., Ding X. & Li B. Coronin3 Promotes Nasopharyngeal Carcinoma Migration And Invasion By Induction Of Epithelial-To-Mesenchymal Transition. *Onco Targets Ther* **12**, 9585-9598 (2019).

35. Thal D., et al. Expression of coronin-3 (coronin-1C) in diffuse gliomas is related to malignancy. *J Pathol* **214**, 415-424 (2008).

36. Hirata E., et al. Intravital imaging reveals how BRAF inhibition generates drug-tolerant microenvironments with high integrin beta1/FAK signaling. *Cancer Cell* **27**, 574-588 (2015).

37. Jia L., et al. Urimem, a membrane that can store urinary proteins simply and economically, makes the large-scale storage of clinical samples possible. *Sci China Life Sci* **57**, 336-339 (2014).

38. Kim Y., et al. Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer. *Nat Commun* **7**, 11906 (2016).

39. Sun Y., et al. ABCC4 copy number variation is associated with susceptibility to esophageal squamous cell carcinoma. *Carcinogenesis* **35**, 1941-1950 (2014).

40. Hu X. & Beeton C. Detection of functional matrix metalloproteinases by zymography. *J Vis Exp*, (2010).

Acknowledgments

We thank Professor Y. Gao of Beijing Normal University for technical support regarding urine sample collection and helpful suggestions for this study; and Professors W. Cui and Z. Zhou of Cancer Hospital of CAMS for administrative support for sample collection. This work was supported by grants from the National Key R & D Program (2017YFC0906603, 2018YFC1313101 and 2016YFC0901403), the State Key Project for Infectious Diseases (2017ZX10203205-003 to X.Z.), the National Natural Science Foundation of China (81872033 to X.Z.) and (82073327 to Y.S.), the Non-profit Central Research Institute Fund of CAMS (2018RC310011 to W.S.) and the CAMS Innovation Fund for Medical Sciences (2016-I2M-1-001 and 2019-I2M-1-003 to X.Z.) and (2017-I2M-3-005 to Y.S.).

Additional information

Supplementary Information accompanies this paper at Nature Communication website.

Author contributions

X.Z., W.S., Y.S., and Z.G. designed all studies and discussed the results. M.C., L.Y., J.L., Y.Z., Y.S., J.G., X.X. L.T. and L.W. collected and prepared urine samples. Z.G., W.S., H.S. and X.L. performed urinary proteomics analysis including TMT labeling and LC-MS/MS and PRM analyses. C.S., W.S., Y.S., Z.G., X.L., Z.J. and L.Y. performed database searching and bioinformatics and statistical analyses. F.L. and S.Z. performed immunohistochemistry staining and pathological analysis. J.L., L.Y., P.N. and L.Z. performed xenograft tumor-formation assays. L.Y. and J.L. performed quantitative dot blot analysis of urinary proteins and CORO1C functional study. Z.J. performed and analyzed the activation of the integrin/FAK/SRC signaling pathways by Western blotting and zymography assays. Z.Z., Y.Z. and Y.H. selected patients and collected their clinical information. Under the direction of X.Z. and W.S., Y.S. wrote the manuscript with input from all authors.

Competing interests

All authors declare that they have no competing interests.

Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Figures legends

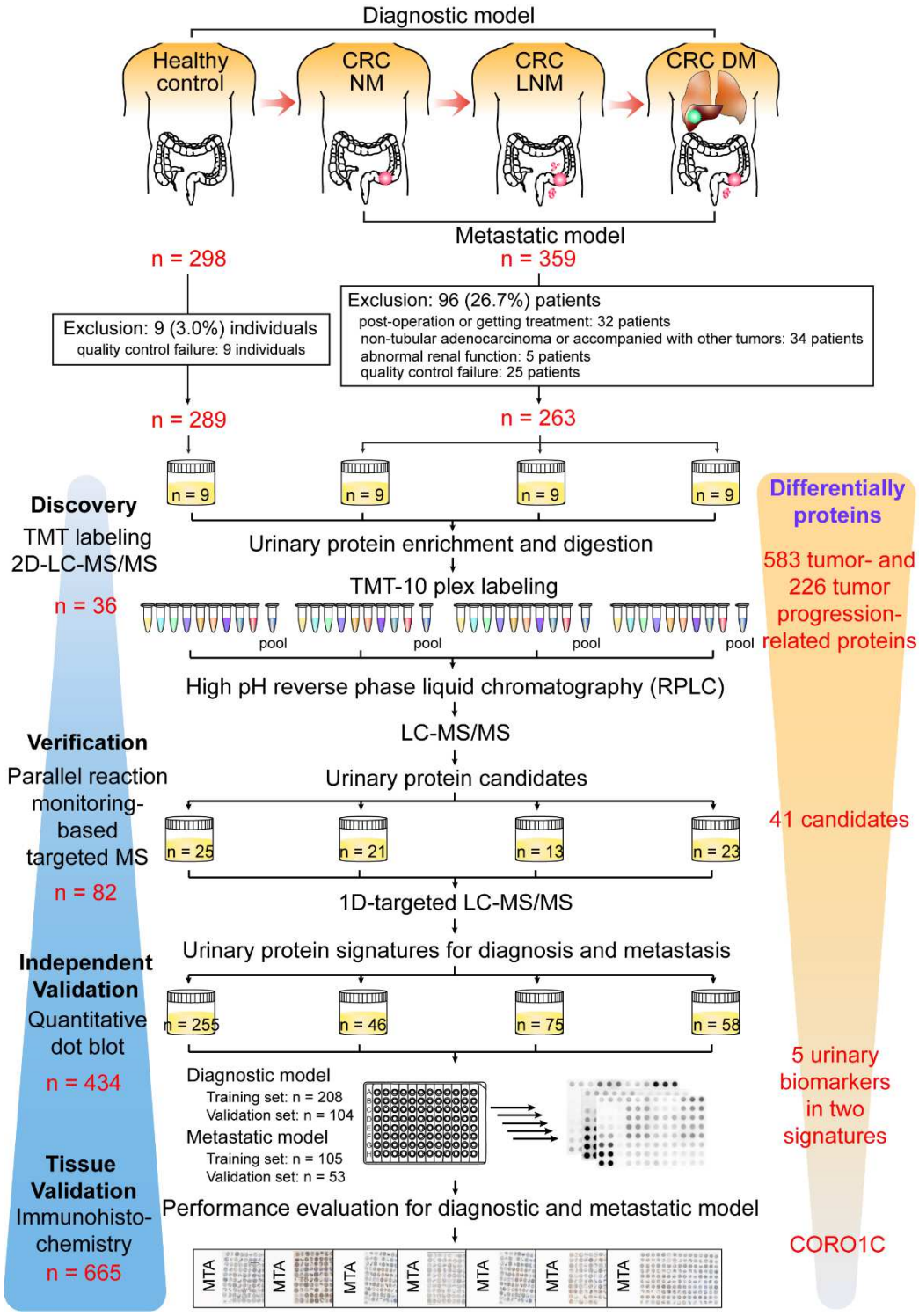
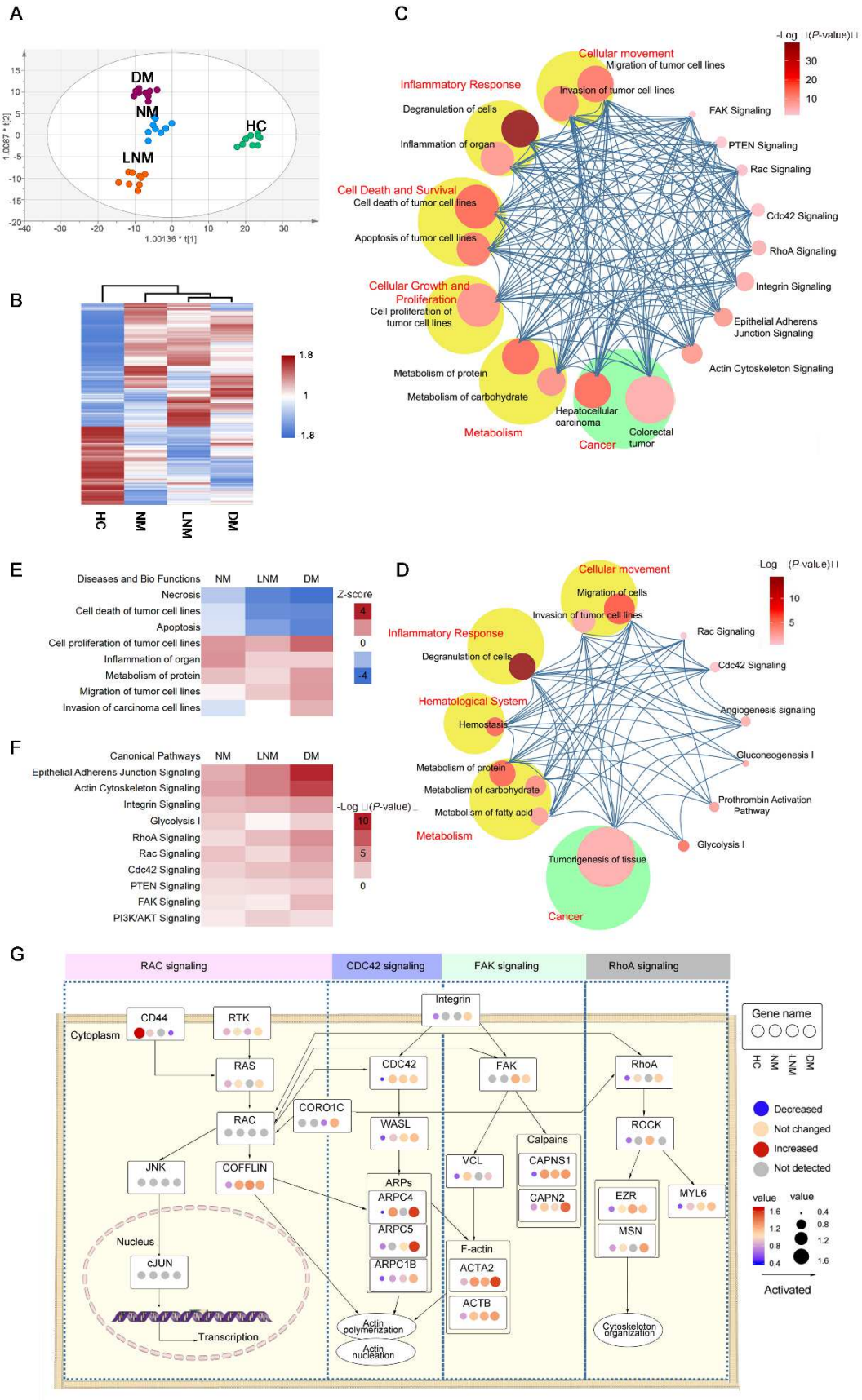


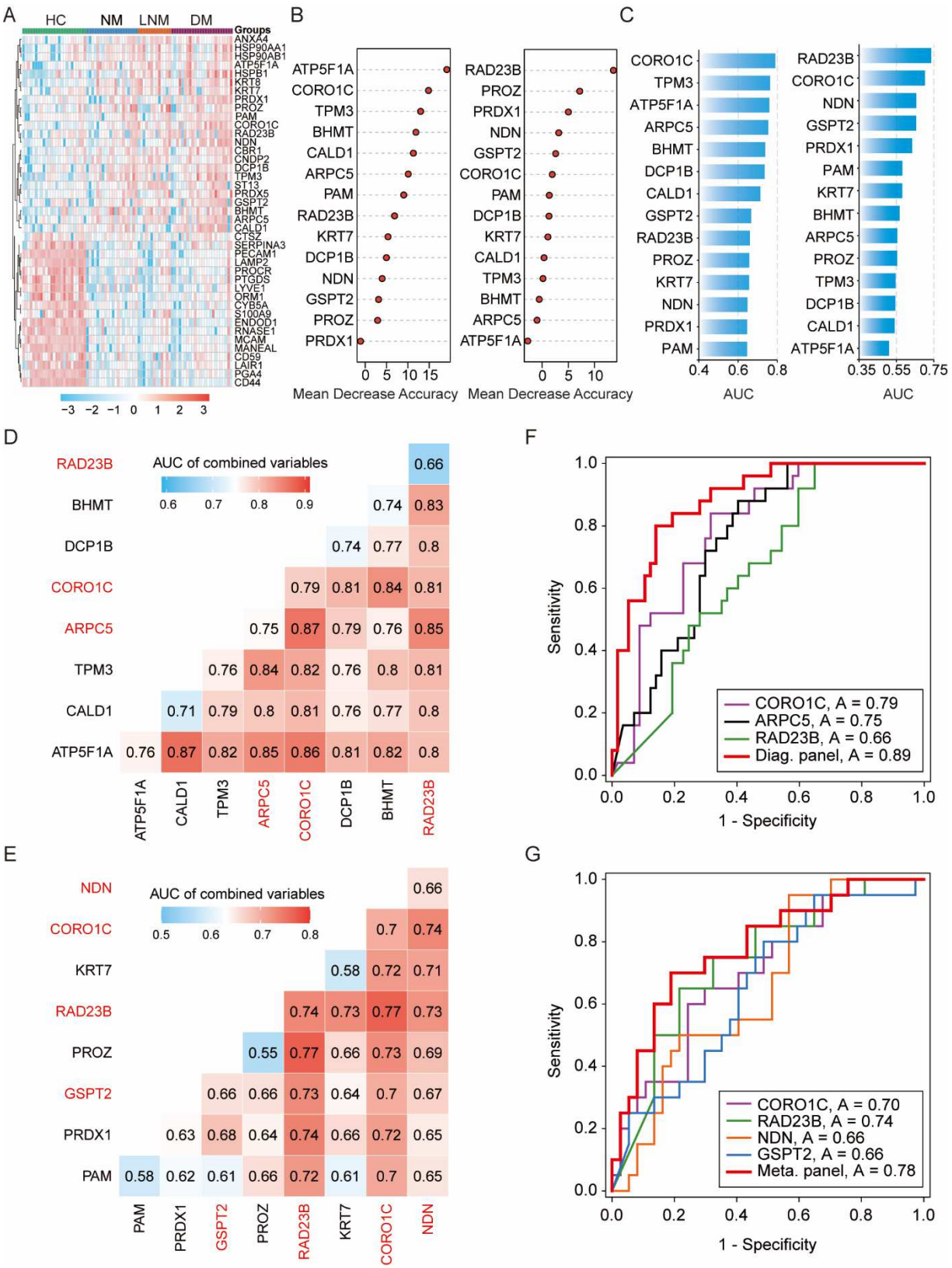
Fig. 1. The overall workflow of study sample inclusion and exclusion criteria as well as the discovery, verification, independent validation and tissue validation for CRC urine

biomarkers. The detailed inclusion and exclusion criteria of the samples are shown. The CRC patients were divided into three groups by metastatic status, including patients without metastases (NM), with lymph node metastasis (LNM) and with distant metastasis (DM) groups. The four-stage workflow consisted of a series of mass spectrometry (MS) and immunoassay-based approaches, including TMT labeling-2D-LC-MS/MS quantitative proteomic strategy, parallel reaction monitoring-based targeted proteomic method, quantitative dot blot analysis and tissue immunohistochemistry, to construct a coherent and high-throughput cancer biomarker method in urine.


A
DM
NM
HC
LNM
1.00136 * t[1]
1.0097 * t[2]
B
HC
NM
LNM
DM
1.8
1
-1.8
C
-Log (P-value)
30
20
10
Cellular movement
Migration of tumor cell lines
Invasion of tumor cell lines
Inflammatory Response
Degranulation of cells
Inflammation of organ
FAK Signaling
PTEN Signaling
Rac Signaling
Cell Death and Survival
Cell death of tumor cell lines
Apoptosis of tumor cell lines
Cdc42 Signaling
RhoA Signaling
Integrin Signaling
Cellular Growth and Proliferation
Cell proliferation of tumor cell lines
Epithelial Adherens Junction Signaling
Actin Cytoskeleton Signaling
Metabolism of protein
Metabolism of carbohydrate
Hepatocellular carcinoma
Colorectal tumor
Metabolism
Cancer
D
-Log (P-value)
10
5
Cellular movement
Migration of cells
Invasion of tumor cell lines
Inflammatory Response
Degranulation of cells
Rac Signaling
Cdc42 Signaling
Hematological System
Hemostasis
Angiogenesis signaling
Gluconeogenesis I
Metabolism of protein
Metabolism of carbohydrate
Metabolism of fatty acid
Prothrombin Activation Pathway
Metabolism
Glycolysis I
Tumorigenesis of tissue
Cancer
E
Diseases and Bio Functions
NM
LNM
DM
Necrosis
Cell death of tumor cell lines
Apoptosis
Cell proliferation of tumor cell lines
Inflammation of organ
Metabolism of protein
Migration of tumor cell lines
Invasion of carcinoma cell lines
Z-score
4
0
-4
F
Canonical Pathways
NM
LNM
DM
Epithelial Adherens Junction Signaling
Actin Cytoskeleton Signaling
Integrin Signaling
Glycolysis I
RhoA Signaling
Rac Signaling
Cdc42 Signaling
PTEN Signaling
FAK Signaling
PI3K/AKT Signaling
-Log (P-value)
10
5
0
G
RAC signaling
CDC42 signaling
FAK signaling
RhoA signaling
Integrin
CD44
RTK
Cytoplasm
RAS
CDC42
FAK
RhoA
CORO1C
RAC
WASL
ROCK
Calpains
CAPNS1
CAPN2
VCL
JNK
COFFLIN
ARPs
ARPC4
ARPC5
ARPC1B
EZR
MSN
MYL6
F-actin
ACTA2
ACTB
Nucleus
cJUN
Transcription
Actin polymerization
Actin nucleation
Cytoskeleton organization
Gene name
HC
NM
LNM
DM
Decreased
Not changed
Increased
Not detected
value
value
1.6
1.2
0.8
0.4
0.4
0.8
1.2
1.6
Activated

944

Fig. 2. Quantitative urinary proteomics analysis in CRC at the discovery stage. (**A**) Score plot of orthogonal partial least squares discriminant analysis (OPLS-DA) model among the four groups. (**B**) Relative abundance of differential proteins among the four groups by unsupervised clustering. (**C** and **D**) CRC tumor-related (C) and tumor progression-related (D) pathway networks. Pathways are grouped vertically into three classes: disease, function and canonical pathways. The color of each node represents the -log10 (*P* value) of that pathway. The size of each node represents the number of differential proteins in that pathway/disease/function. Interactions between pairs of pathways are indicated by curves. (**E**) Heatmap of the deregulated biofunctions in the three CRC patient groups depicted by IPA. Red: Z-score>0, activated; Blue: Z-score<0, inhibited. (**F**) Heatmap of the deregulated canonical pathways in the three CRC groups depicted by IPA. The color represents the -log10 (*P* value) of that pathway. (**G**) Schematic diagram of tumor progression-related pathways, including RAC, CDC42, FAK and RhoA signaling pathways. The protein levels in the HC, NM, LNM, and DM groups are shown. The color and the size of the circle within each gene represent the expression levels of different stages of CRC of each gene.


A
HC
NM
LNM
DM
Groups
ANXA4
HSP90AA1
HSP90AB1
ATP5F1A
HSPB1
KRT8
KRT7
PRDX1
PROZ
PAM
CORO1C
RAD23B
NDN
CBR1
CNDP2
DCP1B
TPM3
ST13
PRDX5
GSPT2
BHMT
ARPC5
CALD1
CTSZ
SERPINA3
PECAM1
LAMP2
PROCR
PTGDS
LYVE1
ORM1
CYB5A
S100A9
ENDOD1
RNASE1
MCAM
MANEAL
CD59
LAIR1
PGA4
CD44
−3 −2 −1 0 1 2 3
B
ATP5F1A
CORO1C
TPM3
BHMT
CALD1
ARPC5
PAM
RAD23B
KRT7
DCP1B
NDN
GSPT2
PROZ
PRDX1
0 5 10 15
Mean Decrease Accuracy
RAD23B
PROZ
PRDX1
NDN
GSPT2
CORO1C
PAM
DCP1B
KRT7
CALD1
TPM3
BHMT
ARPC5
ATP5F1A
0 5 10
Mean Decrease Accuracy
C
CORO1C
TPM3
ATP5F1A
ARPC5
BHMT
DCP1B
CALD1
GSPT2
RAD23B
PROZ
KRT7
NDN
PRDX1
PAM
0.4 0.6 0.8
AUC
RAD23B
CORO1C
NDN
GSPT2
PRDX1
PAM
KRT7
BHMT
ARPC5
PROZ
TPM3
DCP1B
CALD1
ATP5F1A
0.35 0.55 0.75
AUC
D
RAD23B
BHMT
DCP1B
CORO1C
ARPC5
TPM3
CALD1
ATP5F1A
AUC of combined variables
0.6 0.7 0.8 0.9
0.66
0.74 0.83
0.74 0.77 0.8
0.79 0.81 0.84 0.81
0.75 0.87 0.79 0.76 0.85
0.76 0.84 0.82 0.76 0.8 0.81
0.71 0.79 0.8 0.81 0.76 0.77 0.8
0.76 0.87 0.82 0.85 0.86 0.81 0.82 0.8
ATP5F1A
CALD1
TPM3
ARPC5
CORO1C
DCP1B
BHMT
RAD23B
E
NDN
CORO1C
KRT7
RAD23B
PROZ
GSPT2
PRDX1
PAM
AUC of combined variables
0.5 0.6 0.7 0.8
0.66
0.7 0.74
0.58 0.72 0.71
0.74 0.73 0.77 0.73
0.55 0.77 0.66 0.73 0.69
0.66 0.66 0.73 0.64 0.7 0.67
0.63 0.68 0.64 0.74 0.66 0.72 0.65
0.58 0.62 0.61 0.66 0.72 0.61 0.7 0.65
PAM
PRDX1
GSPT2
PROZ
RAD23B
KRT7
CORO1C
NDN
F
Sensitivity
1 - Specificity
CORO1C, A = 0.79
ARPC5, A = 0.75
RAD23B, A = 0.66
Diag. panel, A = 0.89
G
Sensitivity
1 - Specificity
CORO1C, A = 0.70
RAD23B, A = 0.74
NDN, A = 0.66
GSPT2, A = 0.66
Meta. panel, A = 0.78

960

Fig. 3. The generation of the CRC urinary protein biomarker signature. (**A**) The unsupervised clustering analysis of 41 deregulated proteins in the four groups (healthy controls and CRC patients) based on PRM data. (**B**) Variable importance plots produced by the random forest algorithm measured as each variable's mean decrease in accuracy. The most important predictors have the highest mean decrease accuracy values. Left panel, for the class of CRC patients vs. healthy controls (diagnostic model); right panel, for the class of patients with metastasis (lymph node metastasis and distant metastasis) vs. those without metastasis (metastatic model). (**C**) The AUC was used to evaluate the ability of individual proteins to distinguish between CRC patients and healthy controls (left) as well as between patients with metastasis and those without metastasis (right). (**D** and **E**) The AUC of combining any two variables was calculated and shown as matrix plots for the diagnostic model (D) and metastatic model (E). The proteins that show superior discrimination and complementarity are marked in red. (**F** and **G**) ROC curves for the diagnostic model (F) and metastatic model (G). The performance for the selected protein signature and individual proteins are compared.

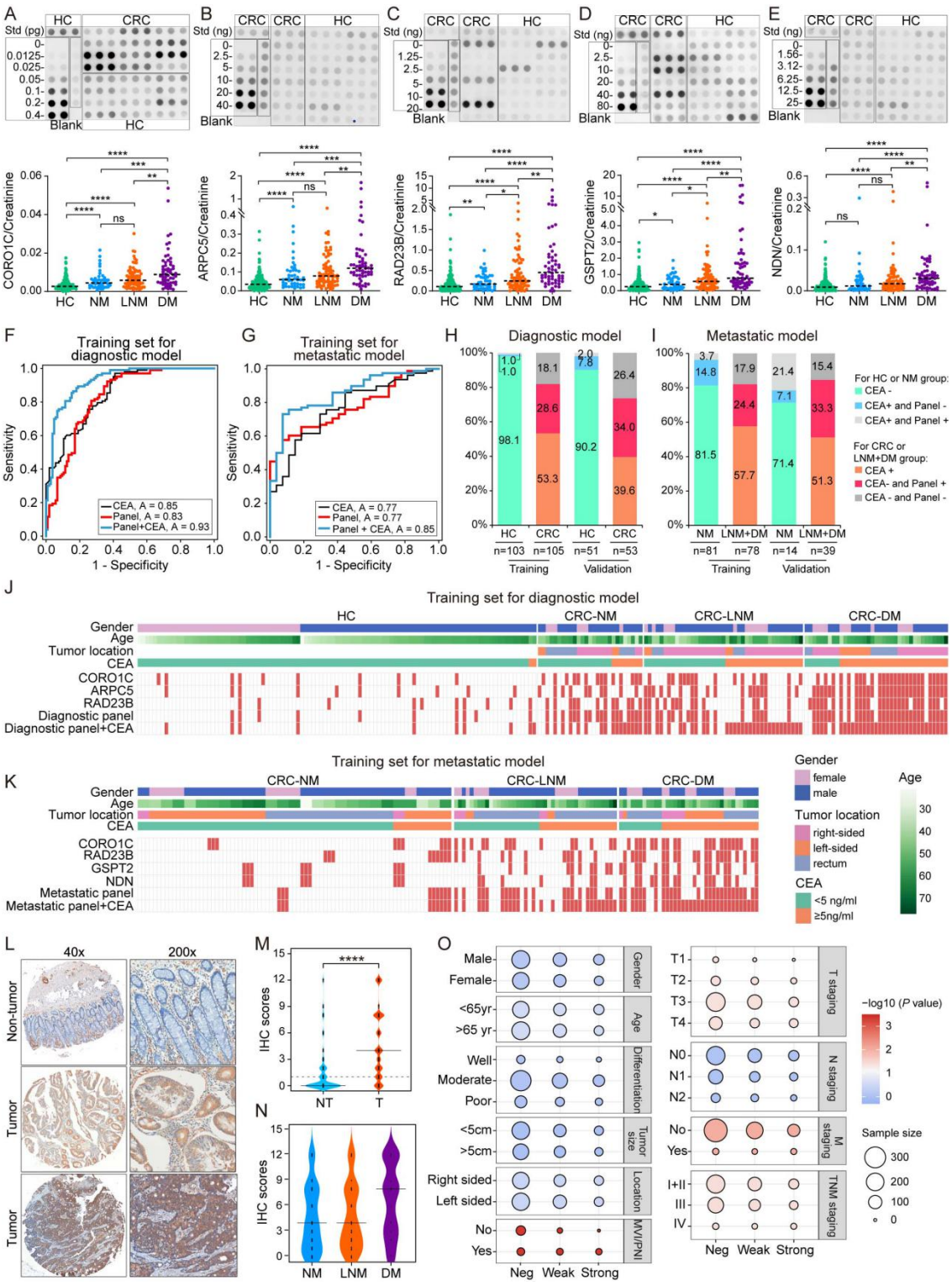
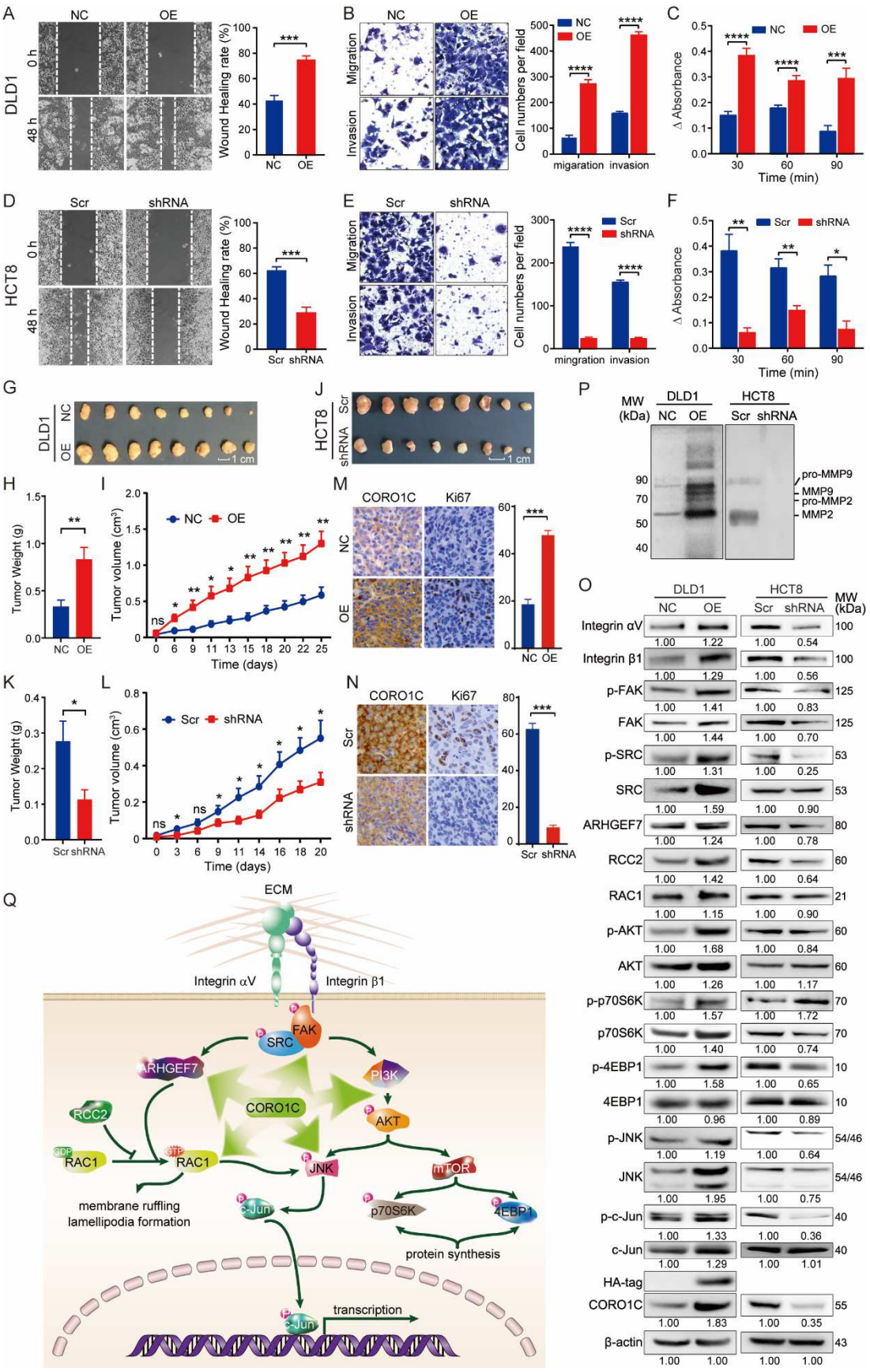


Fig. 4. Independent urine and tissue validation of the urinary protein signature using dot blot analysis and immunohistochemical staining assay. (**A**-**E**) Representative dot blot and scatter plot for CORO1C (**A**), ARPC5 (**B**), RAD23B (**C**), GSPT2 (**D**) and NDN (**E**) in 255 healthy

controls (HCs) and 179 CRC patients, including NM (n=46), LNM (n=75) and DM (n=58). Std, standards. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$, ns, not significant. (**F**) ROC curve of serum CEA, diagnostic panel and the combination of the diagnostic panel and CEA for the diagnostic model in the training set. (**G**) ROC curve of serum CEA, metastatic panel and the combination of the metastatic panel and CEA for metastatic model in the training set. (**H**) Diagnostic power of the diagnostic signature in the individuals who were misdiagnosed by serum CEA in the training and validation groups. (**I**) Metastatic predictive power of the metastatic signature in the individuals who were misdiagnosed by serum CEA in the training and validation groups. (**J**) Heatmap of dot plot data for CORO1C, APRC5, RAD23B, the diagnostic panel, and the diagnostic panel combined with serum CEA (Diagnostic panel+CEA) for the diagnostic model in the training samples. (**K**) Heatmap of dot plot data for CORO1C, RAD23B, GSPT2, NDN, the metastatic panel, and the metastatic panel combined with CEA (Metastatic panel+CEA) for the metastatic model in the training samples. Red: positive using the cutoff value indicated in Table S6. The serum CEA, tumor location, sex and age are indicated by color coding (right side). (**L**) Representative immunohistochemical staining of CORO1C in nontumor (upper panel) and CRC specimens (middle panel: weak positive; lower panel: strong positive). (**M**) Bean plot of immunohistochemical staining scores of CORO1C in the nontumor (NT) and tumor (T) tissues of colon adenocarcinoma (n=665). The median in each group of individuals is shown as long black lines. (**N**) Bean plot of immunohistochemical scores of CORO1C in samples of NM, LNM and DM patients with colon adenocarcinoma. (**O**) The balloon plot for the clinical significance of CORO1C in colon adenocarcinoma patients with distinct staining intensities.


A
NC
OE
DLD1
0 h
48 h
Wound Healing rate (%)
B
Migration
Invasion
Cell numbers per field
migaration
invasion
C
Δ Absorbance
Time (min)
D
Scr
shRNA
HCT8
E
mingration
F
G
1 cm
J
H
Tumor Weight (g)
I
Tumor volume (cm³)
Time (days)
M
CORO1C
Ki67
K
L
N
P
MW (kDa)
pro-MMP9
MMP9
pro-MMP2
MMP2
O
Integrin αV
Integrin β1
p-FAK
FAK
p-SRC
SRC
ARHGEF7
RCC2
RAC1
p-AKT
AKT
p-p70S6K
p70S6K
p-4EBP1
4EBP1
p-JNK
JNK
p-c-Jun
c-Jun
HA-tag
CORO1C
β-actin
Q
ECM
FAK
SRC
PI3K
CORO1C
AKT
mTOR
JNK
c-Jun
p70S6K
4EBP1
membrane ruffling
lamellipodia formation
protein synthesis
transcription

1004

Fig. 5. CORO1C promotes the migration and invasion of CRC cells via integrin/FAK/SRC downstream signaling pathways. (**A** and **D**) Wound healing assay was used to measure cell front migration at 48 h in CORO1C-transfected (OE) DLD1 (**A**) and CORO1C shRNA transfected (shRNA) HCT8 cells (**D**) compared with that in the relative negative control cells (NC or Scr). (**B** and **E**) The migration and invasion capacities were measured by Transwell assays in the cells as demonstrated above. (**C** and **F**) Cell adhesion assay of the cells as demonstrated above the fibronectin-coated wells. The absorbance at 450 nm is represented for the indicated progressive culture times. Macroscopic images of xenograft tumors (**G** and **J**), tumor weights of extirpated tumor xenografts (**H** and **K**) and growth curve of subcutaneous implantation tumor volume (**I** and **L**) of CORO1C-transfected (OE) DLD1 and negative control (NC) cells (**G**-**I**) or CORO1C shRNA transfected (shRNA) HCT8 and scramble negative control (Scr) cells (**J**-**L**). (**M** and **N**) Immunohistochemical staining of CORO1C and Ki67 in xenograft tumor tissues with DLD1 (**M**) and HCT8 (**N**) cells. For A-F, the data shown are done in triplicate and displayed as the mean ± SEM. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant. (**O**) Western blot analysis of integrin, FAK and SRC activation and their downstream signaling molecules in DLD1 and HCT8 cells with the overexpression or silencing of CORO1C. β-actin was used as a loading control. (**P**) Gelatin zymography analysis of pro and active forms of MMP9 and MMP2 in the conditional media of DLD1 and HCT8 cells transfected with the CORO1C or shRNA vector. (**Q**) Schematic diagram of CORO1C-regulated pathways in CRC cells. CORO1C overexpression upregulates the expression of integrin αV, integrin β1, ARHGEF7, RCC2 and RAC1 and induces the phosphorylation of FAK, SRC, AKT, p70S6K, 4EBP1, JNK and c-Jun, thereby activating integrin/FAK/SRC signaling and its downstream effectors and leading to actin reorganization-based membrane ruffling and lamellipodia formation, protein synthesis and altered gene expression for some proliferation-, migration- and invasion-related genes.

Table 1. Clinical information of all samples used in this study.

	HC				NM				LNM				DM				*P* value
	TMT	PRM	Dot blot	*P* value	TMT	PRM	Dot blot	*P* value	TMT	PRM	Dot blot	*P* value	TMT	PRM	Dot blot	*P* value	
Sex				0.6248				0.7946				0.1129				0.178	0.1329
Male	4	14	155		6	14	34		3	10	51		4	18	40		
Female	5	11	104		3	7	12		6	3	24		5	5	18		
Age (years old)				0.1333				0.2579				0.2044				0.3523	0.0951
<55	1	12	113		5	10	14		5	6	23		4	8	14		
≥55	8	13	146		4	11	32		4	7	52		5	15	44		
Histological grade								0.0671				0.1320				0.0298	0.0854
Well differentiated					2	1	1		0	0	0		1	0	1		
Moderately differentiated					7	16	30		5	9	40		5	16	31		
Poorly differentiated					0	4	14		4	1	30		0	3	18		
Unknown					0	0	1		0	3	5		3	4	8		
CEA (ng/mL)								0.6859				0.2594				0.9696	<0.0001
<5	-	-	156		4	6	32		3	4	34		2	5	16		
≥5	-	-	7		1	0	9		0	2	32		6	13	39		
Unknown	-	-	96		4	15	5		6	7	9		1	5	3		
CA19-9 (U/mL)								0.3816				1.0000				0.1564	<0.0001
<37	-	-	161		5	5	40		3	6	56		3	11	35		
≥37	-	-	2		0	1	1		0	0	7		5	3	20		
Unknown	-	-	96		4	15	5		6	7	12		1	9	3		
Location								0.0607				0.8839				0.3852	0.0013
Right side					2	4	6		1	2	8		2	4	11		
Left side					5	3	20		2	3	13		5	11	18		
Rectum					2	14	20		6	8	54		2	8	29		
TNM stage								0.0028				-				-	<0.0001
I					0	11	11		0	0	0		0	0	0		
II					9	10	35		0	0	0		0	0	0		
III					0	0	0		9	13	75		0	0	0		
IV					0	0	0		0	0	0		9	23	58		

Figures

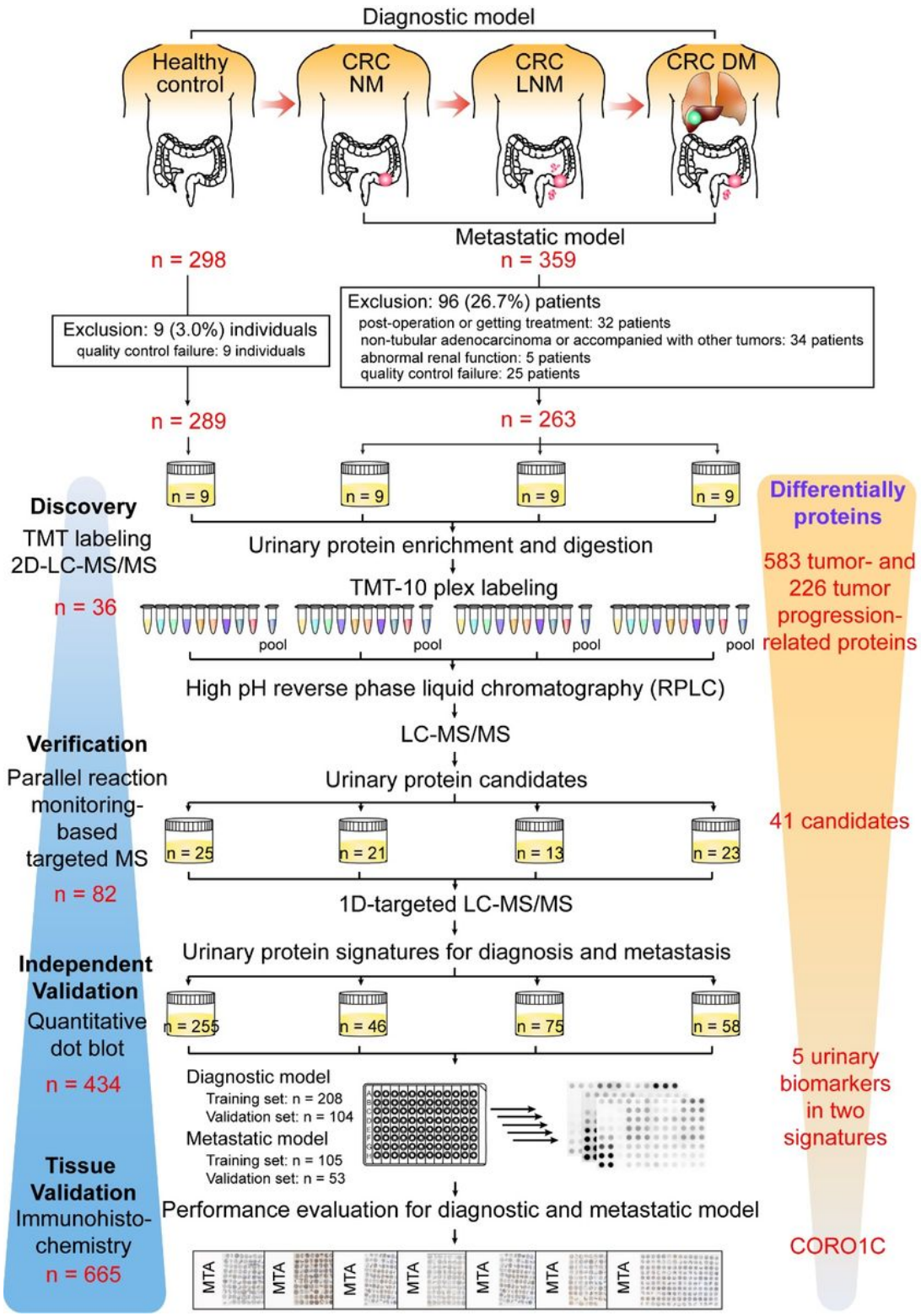


Figure 1

The overall workflow of study sample inclusion and exclusion criteria as well as the discovery, verification, independent validation and tissue validation for CRC urine biomarkers. The detailed inclusion and exclusion criteria of the samples are shown. The CRC patients were divided into three groups by

metastatic status, including patients without metastases (NM), with lymph node metastasis (LNM) and with distant metastasis (DM) groups. The four-stage workflow consisted of a series of mass spectrometry (MS) and immunoassay-based approaches, including TMT labeling-2D-LC-MS/MS quantitative proteomic strategy, parallel reaction monitoring-based targeted proteomic method, quantitative dot blot analysis and tissue immunohistochemistry, to construct a coherent and high-throughput cancer biomarker method in urine.

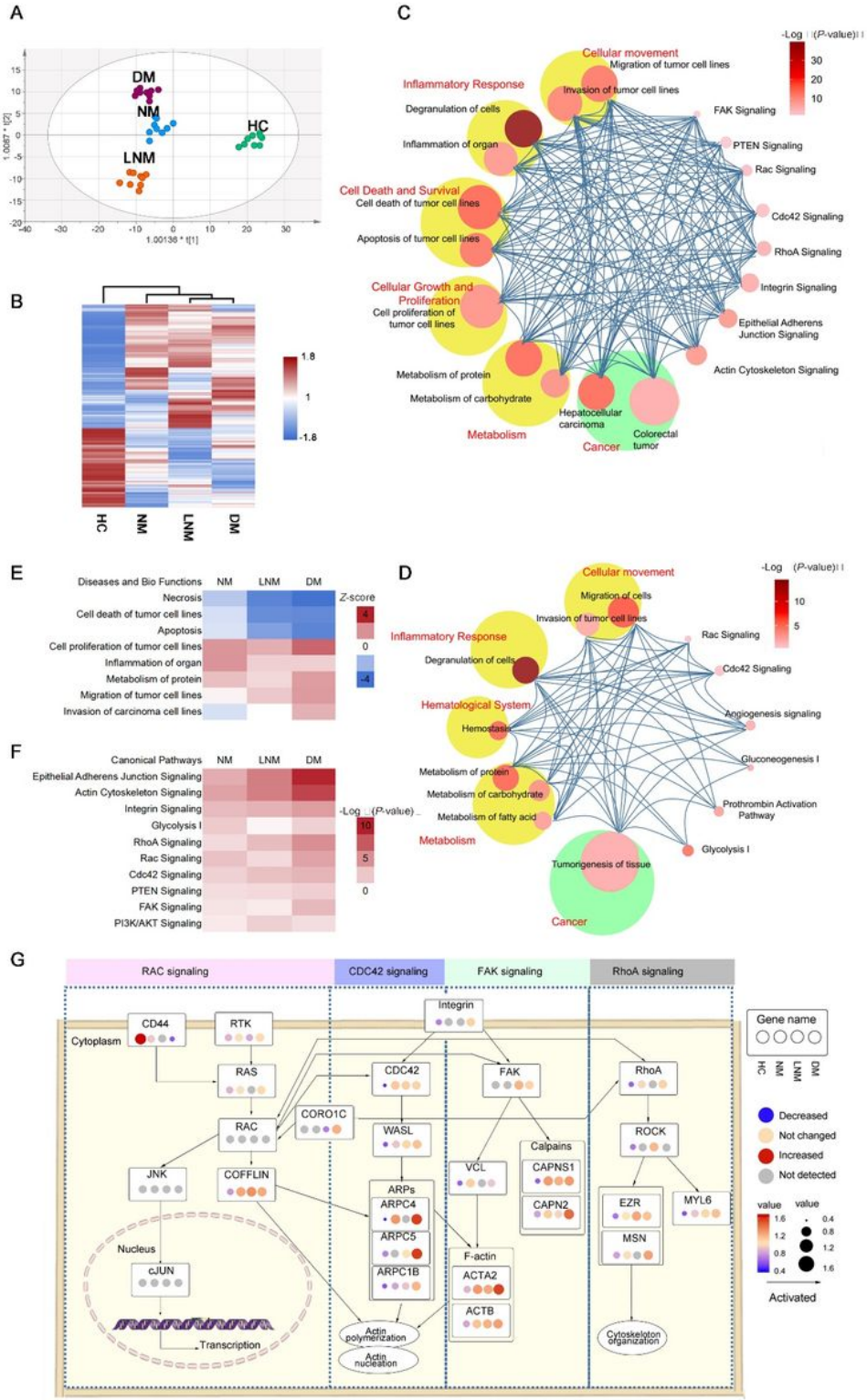


Figure 2

Quantitative urinary proteomics analysis in CRC at the discovery stage. (A) Score plot oforthogonal partial least squares discriminant analysis (OPLS-DA) model among the four groups. (B) Relative abundance of differential proteins among the four groups by unsupervised clustering. (C and D) CRC tumor-related (C) and tumor progression-related (D) pathway networks. Pathways are grouped vertically into three classes: disease, function and canonical pathways. The color of each node represents the -log10 (P value) of that pathway. The size of each node represents the number of differential proteins in that pathway/disease/function. Interactions between pairs of pathways are indicated by curves. (E) Heatmap of the deregulated biofunctions in the three CRC patient groups depicted by IPA. Red: Z-score>0, activated; Blue: Z-score<0, inhibited. (F) Heatmap of the deregulated canonical pathways in the three CRC groups depicted by IPA. The color represents the -log10 (P value) of that pathway. (G) Schematic diagram of tumor progression-related pathways, including RAC, CDC42, FAK and RhoA signaling pathways. The protein levels in the HC, NM, LNM, and DM groups are shown. The color and the size of the circle within each gene represent the expression levels of different stages of CRC of each gene.

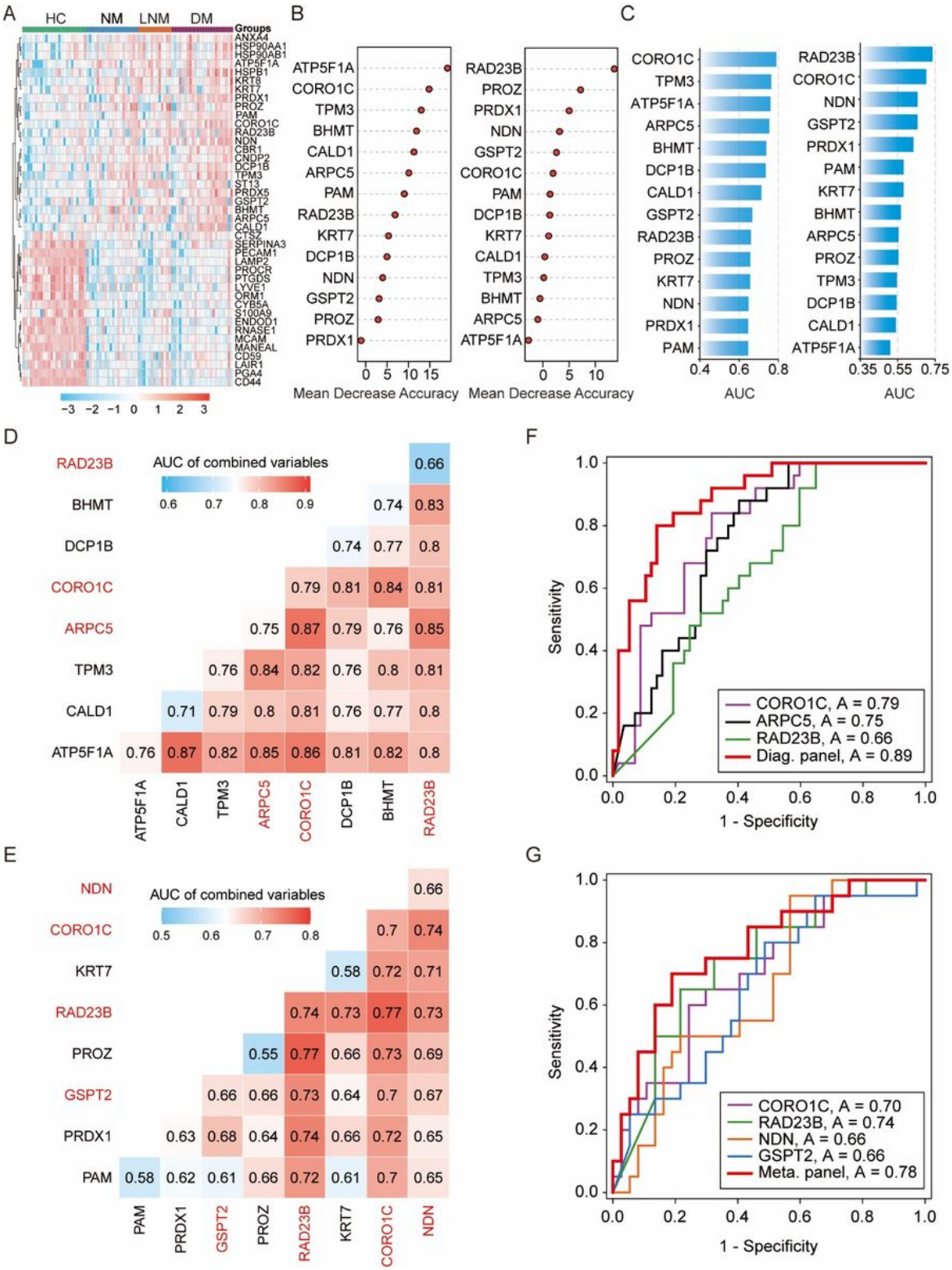


Figure 3

The generation of the CRC urinary protein biomarker signature. (A) The unsupervised clustering analysis of 41 deregulated proteins in the four groups (healthy controls and CRC patients) based on PRM data. (B) Variable importance plots produced by the random forestalgorithm measured as each variable's mean decrease in accuracy. The most important predictors have the highest mean decrease accuracy values. Left panel, for the class of CRC patients vs. healthy controls (diagnostic model); right panel, for the class

of patients with metastasis (lymph node metastasis and distant metastasis) vs. those without metastasis (metastatic model). (C) The AUC was used to evaluate the ability of individual proteins to distinguish between CRC patients and healthy controls (left) as well as between patients with metastasis and those without metastasis (right). (D and E) The AUC of combining any two variables was calculated and shown as matrix plots for the diagnostic model (D) and metastatic model (E). The proteins that show superior discrimination and complementarity are marked in red. (F and G) ROC curves for the diagnostic model (F) and metastatic model (G). The performance for the selected protein signature and individual proteins are compared.

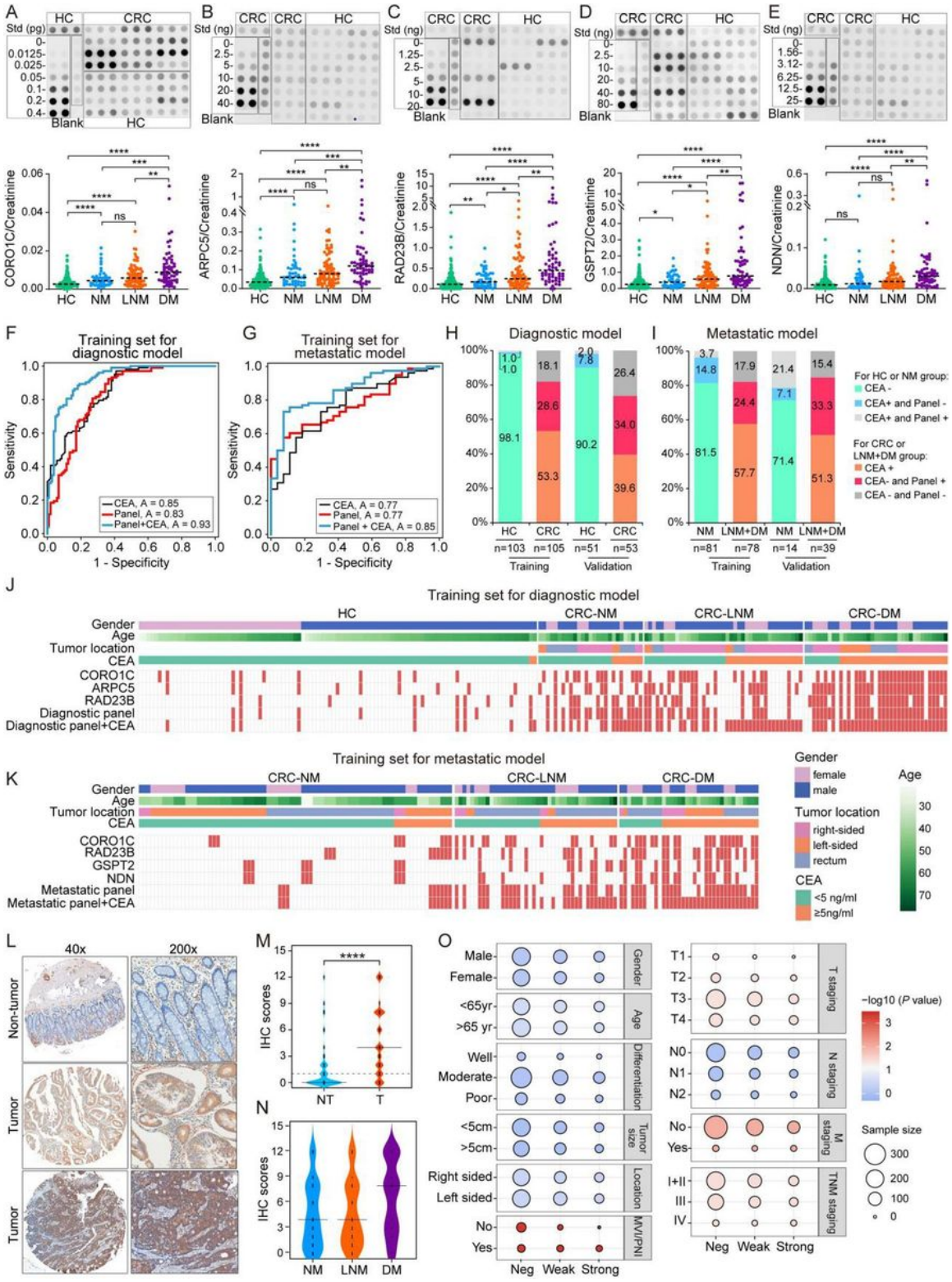


Figure 4

Independent urine and tissue validation of the urinary protein signature using dot blot analysis and immunohistochemical staining assay. (A-E) Representative dot blot and scatter plot for CORO1C (A), ARPC5 (B), RAD23B (C), GSPT2 (D) and NDN (E) in 255 healthy controls (HCs) and 179 CRC patients, including NM (n=46), LNM (n=75) and DM (n=58). Std, standards. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001, ns, not significant. (F) ROC 982 curve of serum CEA, diagnostic panel and the

combination of the diagnostic panel and CEA for the diagnostic model in the training set. (G) ROC curve of serum CEA, metastatic panel and the combination of the metastatic panel and CEA for metastatic model in the training set. (H) Diagnostic power of the diagnostic signature in the individuals who were misdiagnosed by serum CEA in the training and validation groups. (I) Metastatic predictive power of the metastatic signature in the individuals who were misdiagnosed by serum CEA in the training and validation groups. (J) Heatmap of dot plot data for CORO1C, APRC5, RAD23B, the diagnostic panel, and the diagnostic panel combined with serum CEA (Diagnostic panel+CEA) for the diagnostic model in the training samples. (K) Heatmap of dot plot data for CORO1C, RAD23B, GSPT2, NDN, the metastatic panel, and the metastatic panel combined with CEA (Metastatic panel+CEA) for the metastatic model in the training samples. Red: positive using the cutoff value indicated in Table S6. The serum CEA, tumor location, sex and age are indicated by color coding (right side). (L) Representative immunohistochemical staining of CORO1C in nontumor (upper panel) and CRC specimens (middle panel: weak positive; lower panel: strong positive). (M) Bean plot of immunohistochemical staining scores of CORO1C in the nontumor (NT) and tumor (T) tissues of colon adenocarcinoma (n=665). The median in each group of individuals is shown as long black lines. (N) Bean plot of immunohistochemical scores of CORO1C in samples of NM, LNM and DM patients with colon adenocarcinoma. (O) The balloon plot for the clinical significance of CORO1C in colon adenocarcinoma patients with distinct staining intensities.

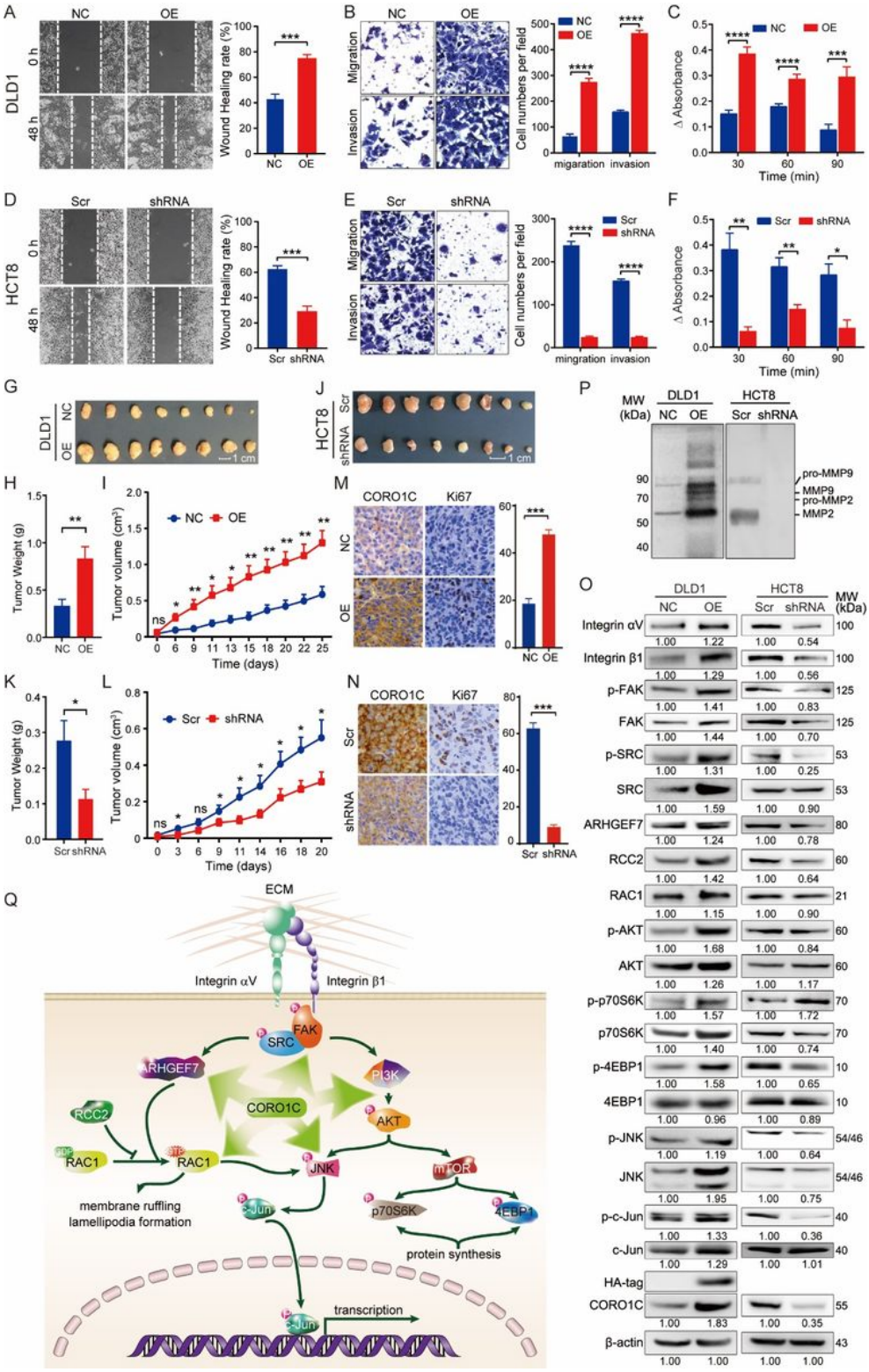


Figure 5

CORO1C promotes the migration and invasion of CRC cells via integrin/FAK/SRC downstream signaling pathways. (A and D) Wound healing assay was used to measure cell front migration at 48 h in CORO1C-transfected (OE) DLD1 (A) and CORO1C shRNA transfected (shRNA) HCT8 cells (D) compared with that in the relative negative control cells (NC or Scr). (Band E) The migration and invasion capacities were measured by Transwell assays in the cells as demonstrated above. (C and F) Cell adhesion assay of the

cells as demonstrated above the fibronectin-coated wells. The absorbance at 450 nm is represented for the indicated progressive culture times. Macroscopic images of xenograft tumors (G and J), tumor weights of extirpated tumor xenografts (H and K) and growth curve of subcutaneous implantation tumor volume (I and L) of CORO1C-transfected (OE) DLD1 and negative control (NC) cells (G-I) or CORO1C shRNA transfected (shRNA) HCT8 and scramble negative control (Scr) cells (J-L). (M and N) Immunohistochemical staining of CORO1C and Ki67 in xenograft tumor tissues with DLD1 (M) and HCT8 (N) cells. For A-F, the data shown are done in triplicate and displayed as the mean ± SEM. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001; ns, not significant. (O) Western blot analysis of integrin, FAK and SRC activation and their downstream signaling molecules in DLD1 and HCT8 cells with the overexpression or silencing of CORO1C. β-actin was used as a loading control. (P) Gelatin zymography analysis of pro and active forms of MMP9 and MMP2 in the conditional media of DLD1 and HCT8 cells transfected with the CORO1C or shRNA vector. (Q) Schematic diagram of CORO1C-regulated pathways in CRC cells. CORO1C overexpression upregulates the expression of integrin αV, integrin β1, ARHGEF7, RCC2 and RAC1 and induces the phosphorylation of FAK, SRC, AKT, p70S6K, 4EBP1, JNK and c-Jun, thereby activating integrin/FAK/SRC signaling and its downstream effectors and leading to actin reorganization-based membrane ruffling and lamellipodia formation, protein synthesis and altered gene expression for some proliferation-, migration- and invasion-related genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationv3120820.pdf
- SupplementaryTable23100620.xls
- SupplementaryTable14gzg111920.xlsx
- NCOMMS2048752reportingsummaryv6121820002.pdf