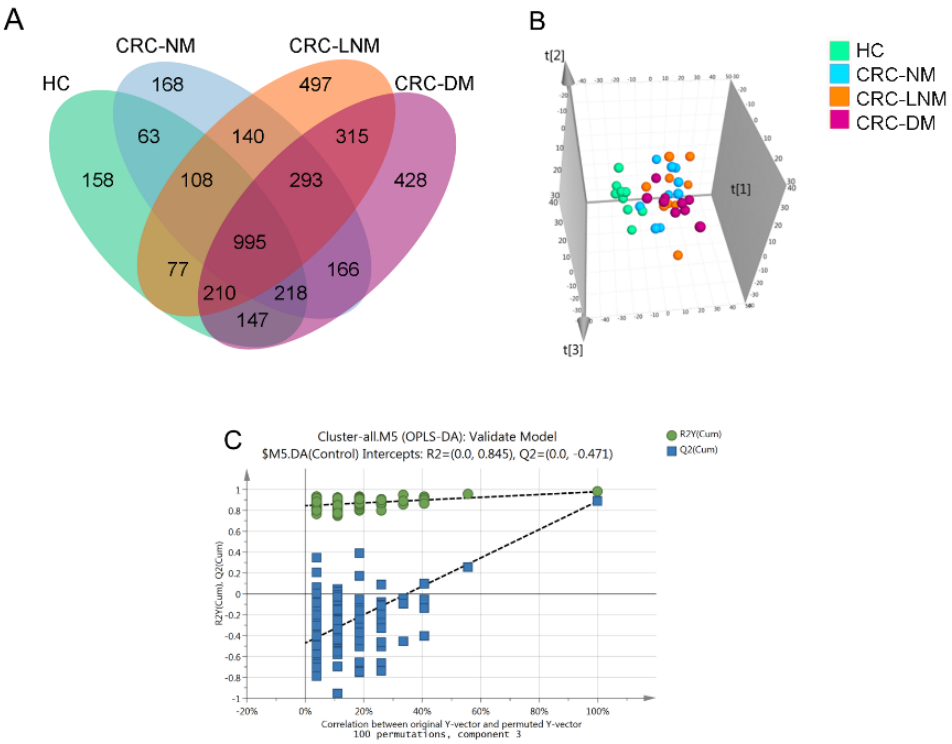
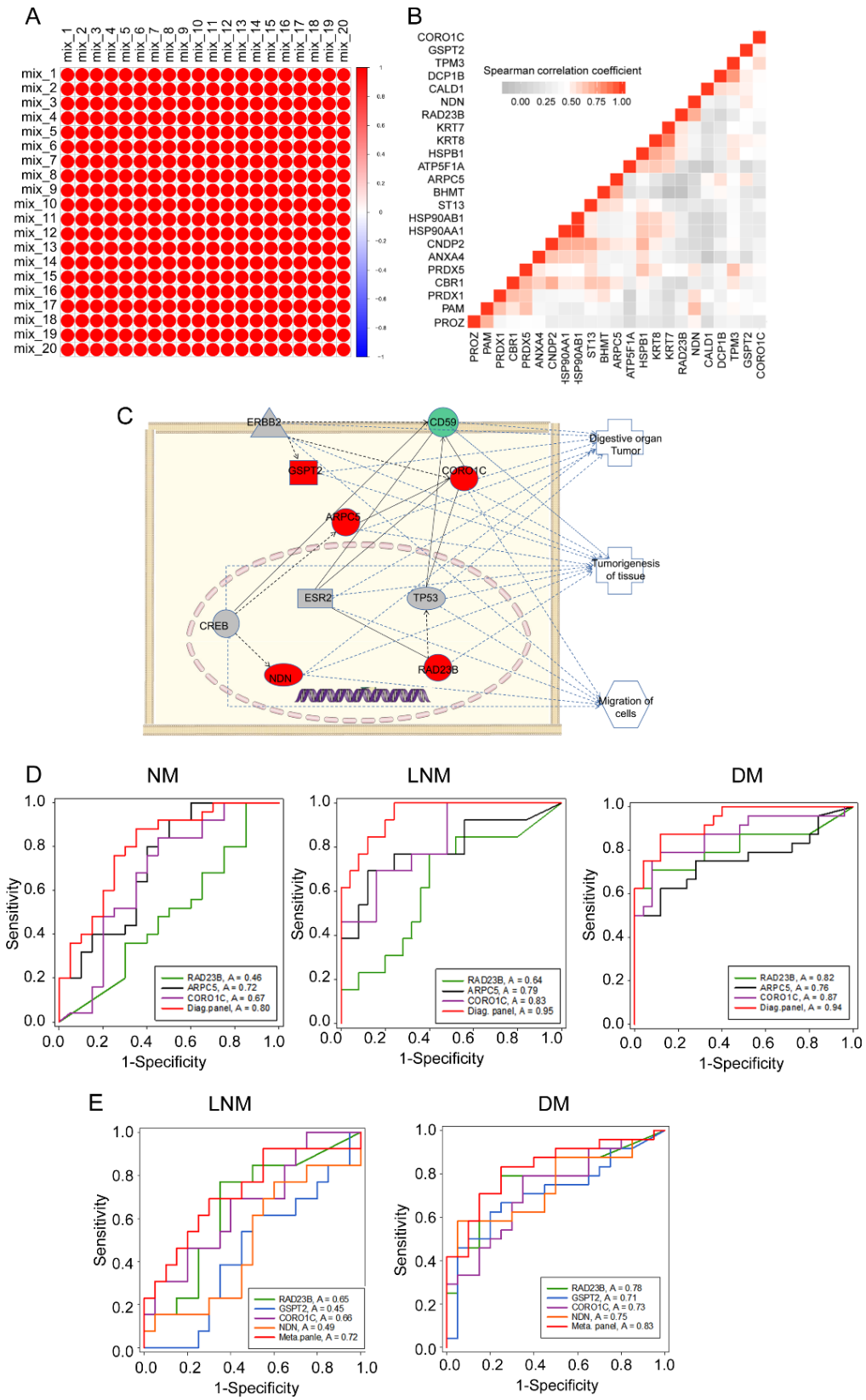
Supplementary Information

Noninvasive urinary protein signatures associated with colorectal cancer diagnosis and metastasis

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]*

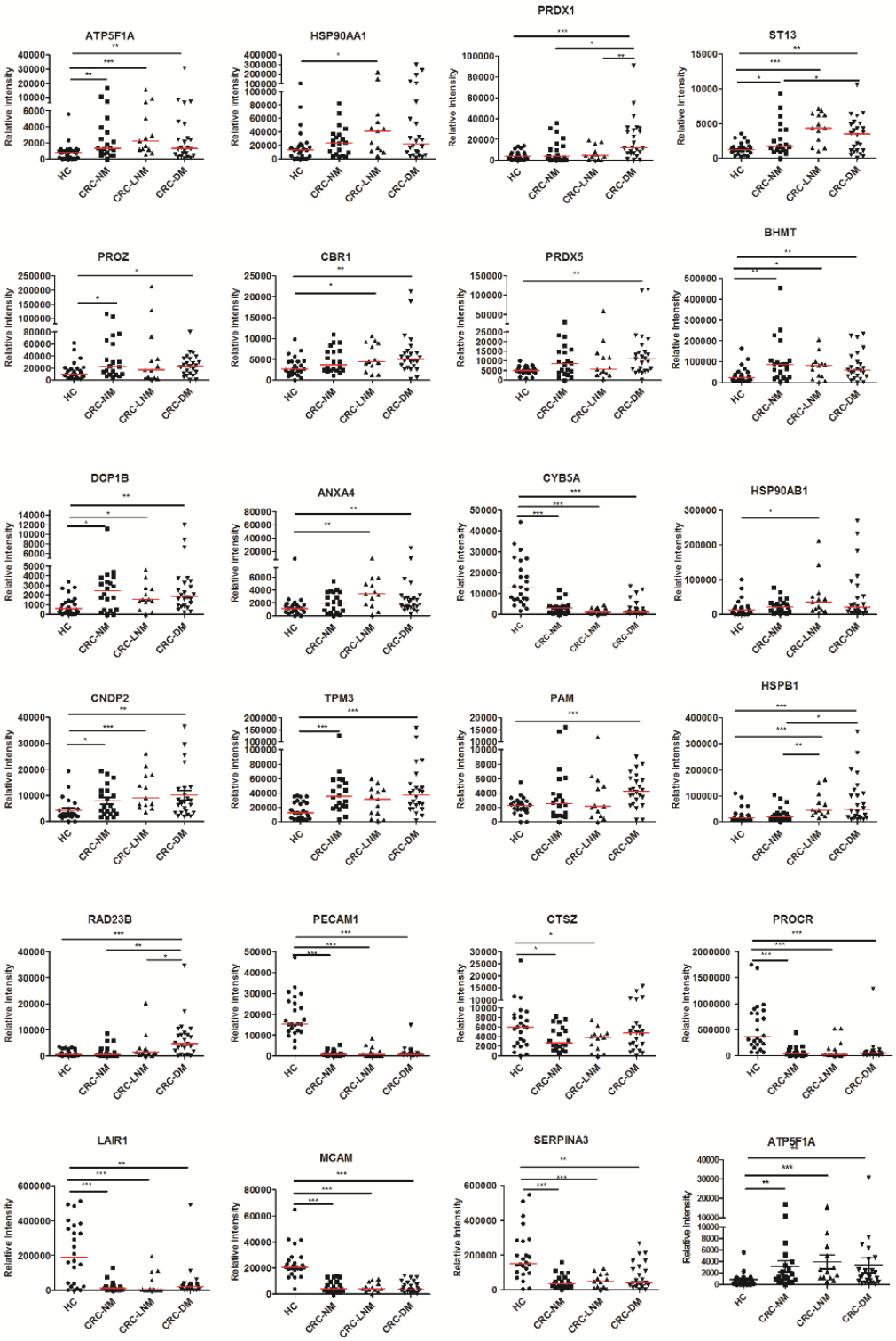


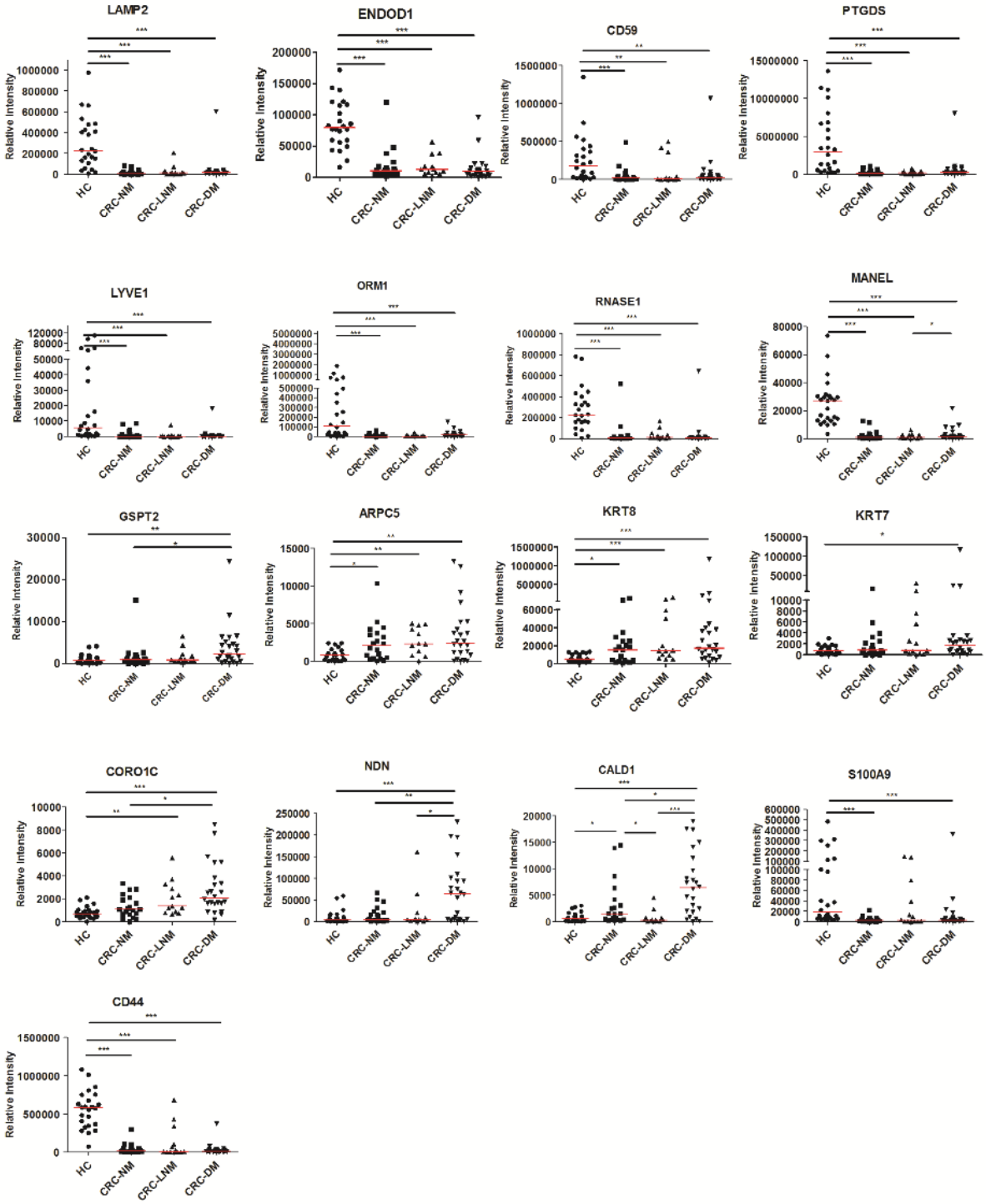
Supplementary Fig. 1. Quantitative urinary proteomics analysis in CRC at the discovery stage. (**A**) Venn diagram of the quantified proteins in the HC, NM, LNM, and DM groups in TMT quantification analysis. (**B**) Score plot of unsupervised principal component analysis (PCA) overview of urinary proteomics among the HC, NM, LNM, and DM groups. (C) One hundred permutation validations of the OPLS-DA model based on the proteome of the four groups.



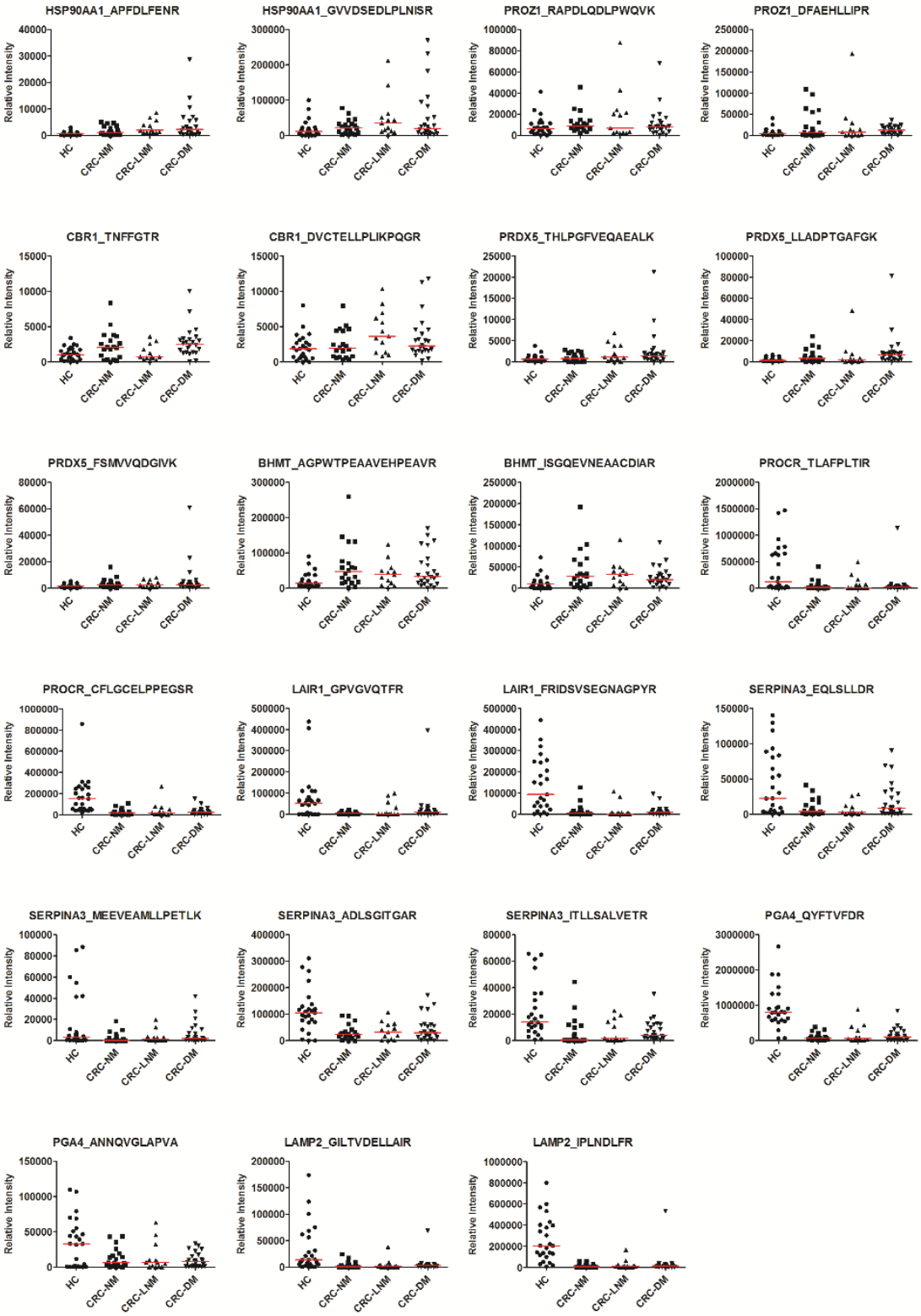
Supplementary Fig. 2. The generation of the CRC urinary protein biomarker signature. (**A**) Assessment of QC samples. The correlation map of QC samples. (**B**) The correlation matrix plot of 23 upregulated proteins in CRC patients based on PRM data. (**C**) Networks of the candidate CRC biomarkers. Red: upregulated proteins; blue: downregulated proteins based

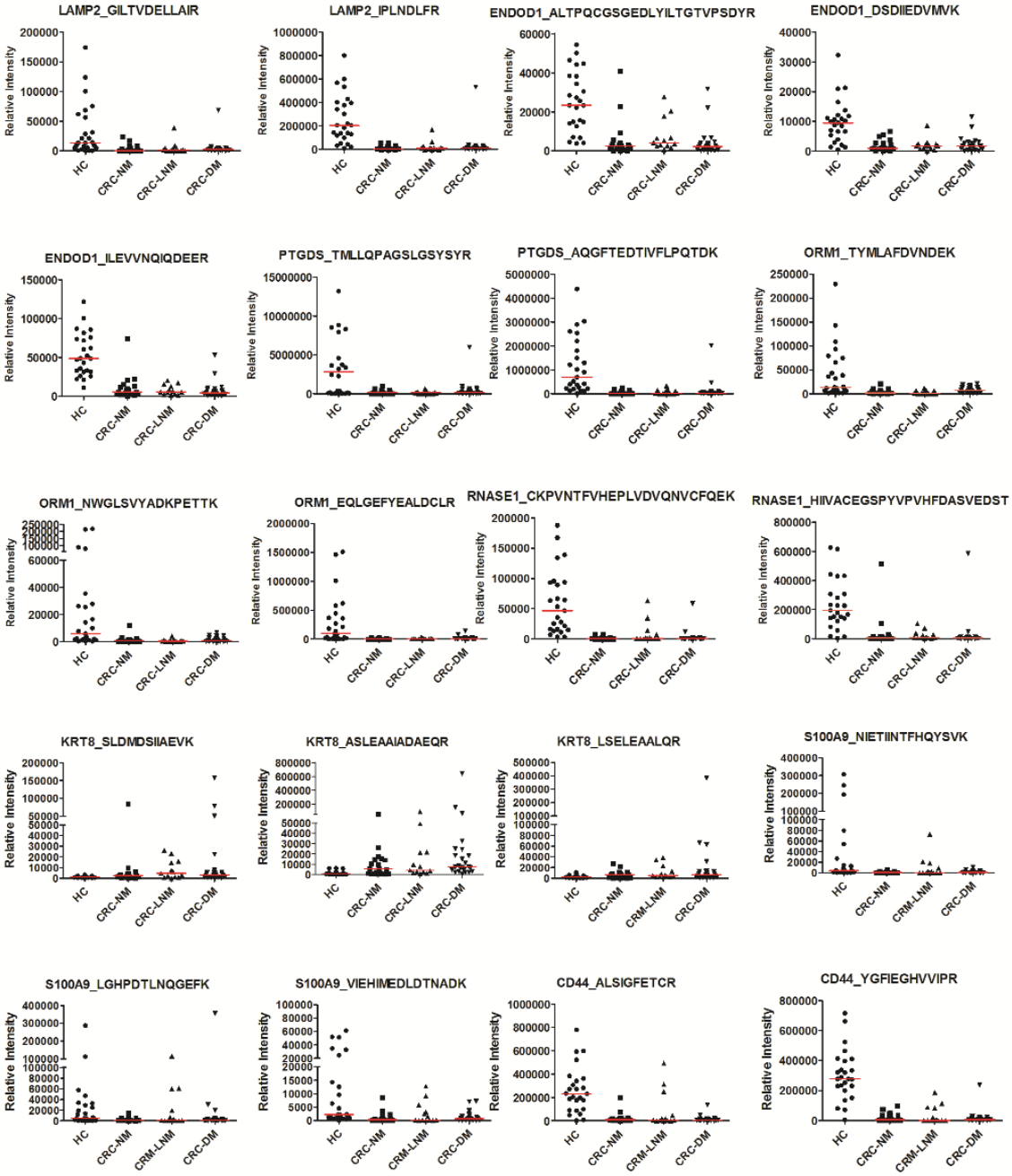
on PRM data. (**D**) Receiver operating characteristic (ROC) curve analysis was conducted by the diagnostic biomarkers to discriminate among the HC and NM, LNM, or DM groups. (**E**) Receiver operating characteristic (ROC) curve analysis was conducted by the metastatic biomarkers to discriminate among the LNM, DM, and NM groups.



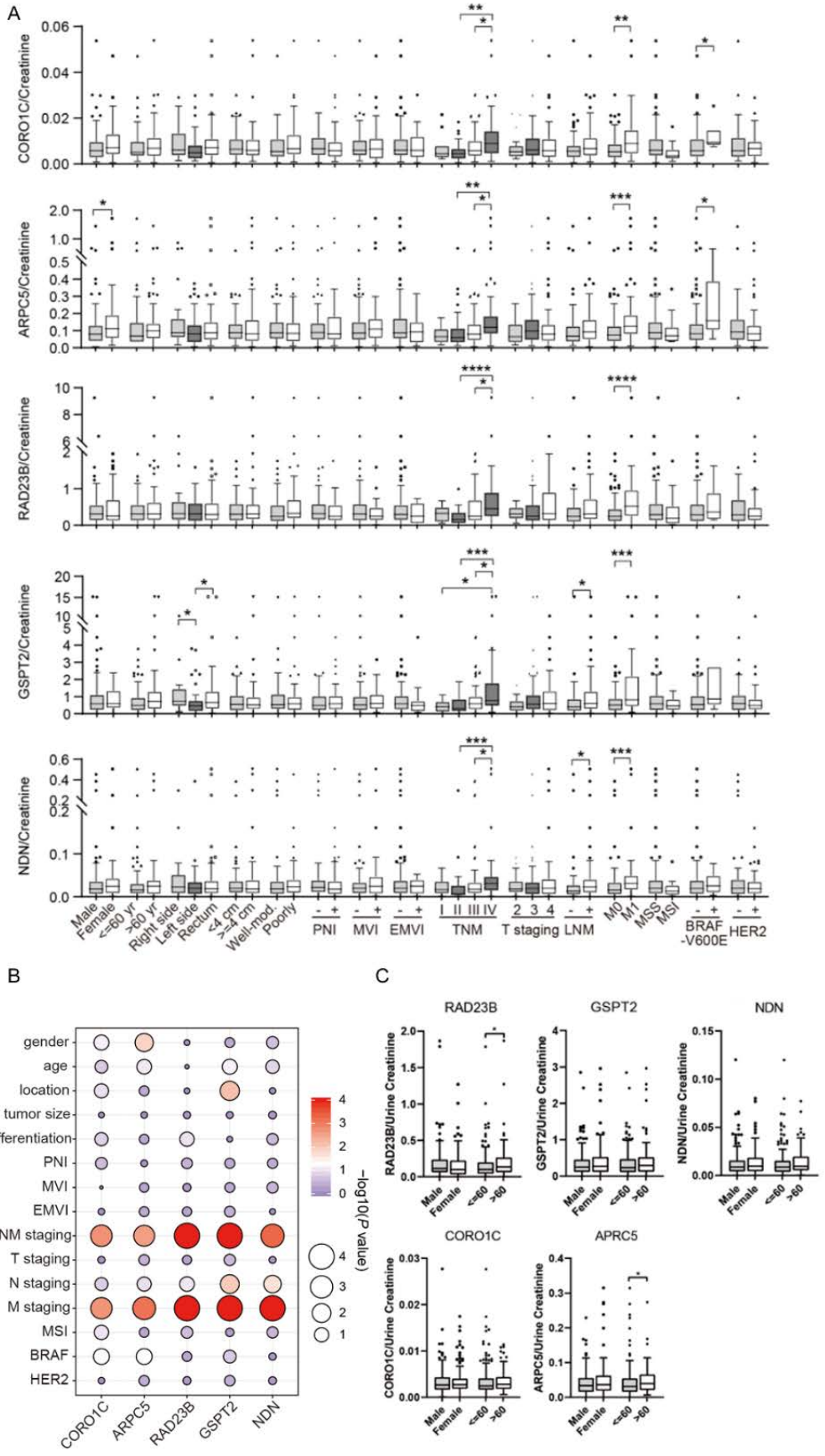


Supplementary Fig. 3. Scatter plots of the 41 proteins in the four groups.

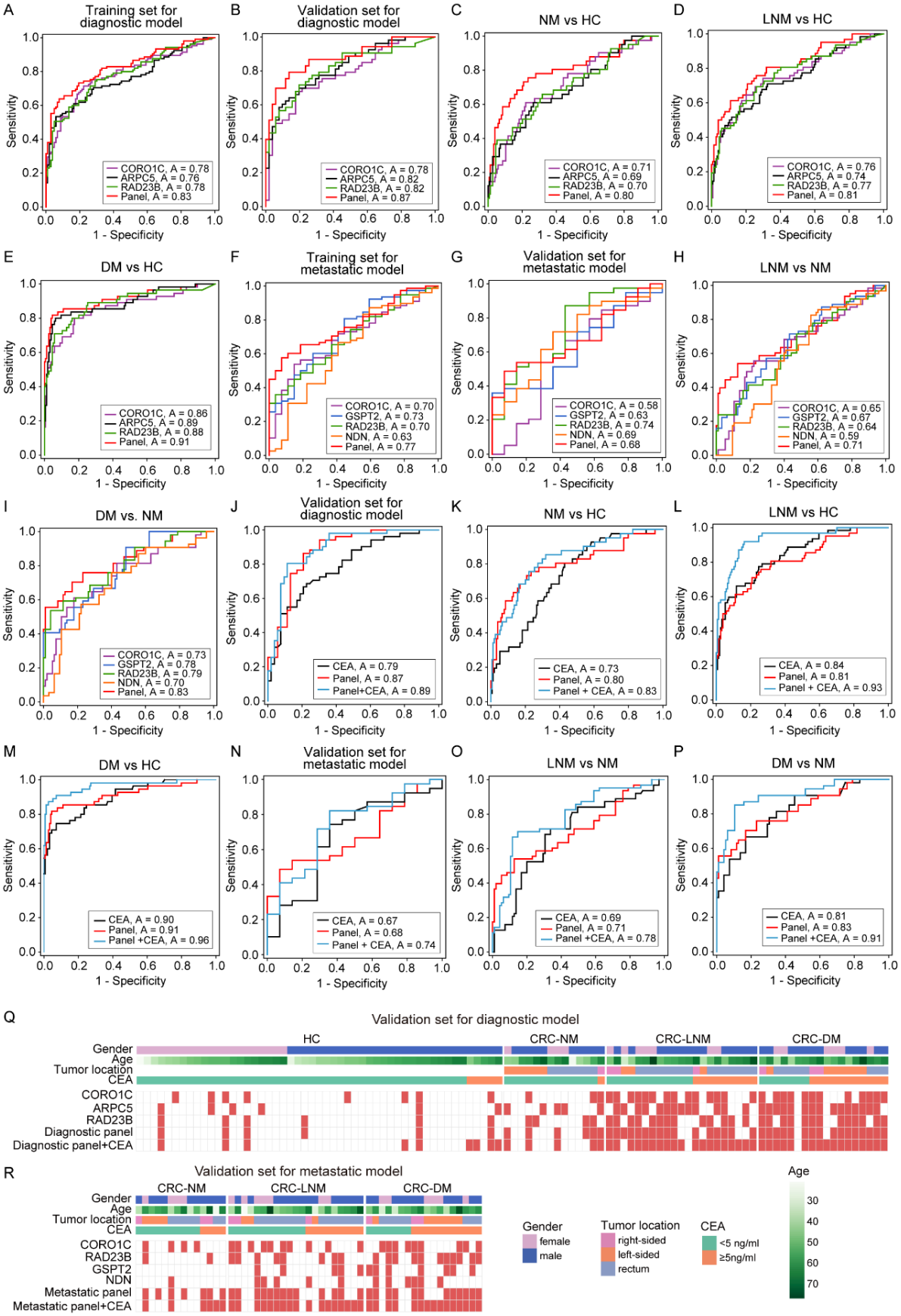

HSP90AA1_APFDLFENR
HSP90AA1_GVVDSEDLPLNISR
PROZ1_RAPDLQDLPWQVK
PROZ1_DFAEHLLIPR
CBR1_TNFFGTR
CBR1_DVCTELLPLIKPQGR
PRDX5_THLPGFVEQAEALK
PRDX5_LLADPTGAFGK
PRDX5_FSMVVQDGIVK
BHMT_AGPWTPEAAVEHPEAVR
BHMT_ISGQEVNEAACDIAR
PROCR_TLAFPLTIR
PROCR_CFLGCELPPEGSR
LAIR1_GPVGVQTFR
LAIR1_FRIDSVSEGNAGPYR
SERPINA3_EQLSLLDR
SERPINA3_MEEVEAMLLPETLK
SERPINA3_ADLSGITGAR
SERPINA3_ITLLSALVETR
PGA4_QYFTVFDR
PGA4_ANNQVGLAPVA
LAMP2_GILTVDELLAIR
LAMP2_IPLNDLFR
Relative Intensity
HC
CRC-NM
CRC-LNM
CRC-DM



Supplementary Fig. 4. Scatter plots of the different peptides within the same proteins in the four groups.

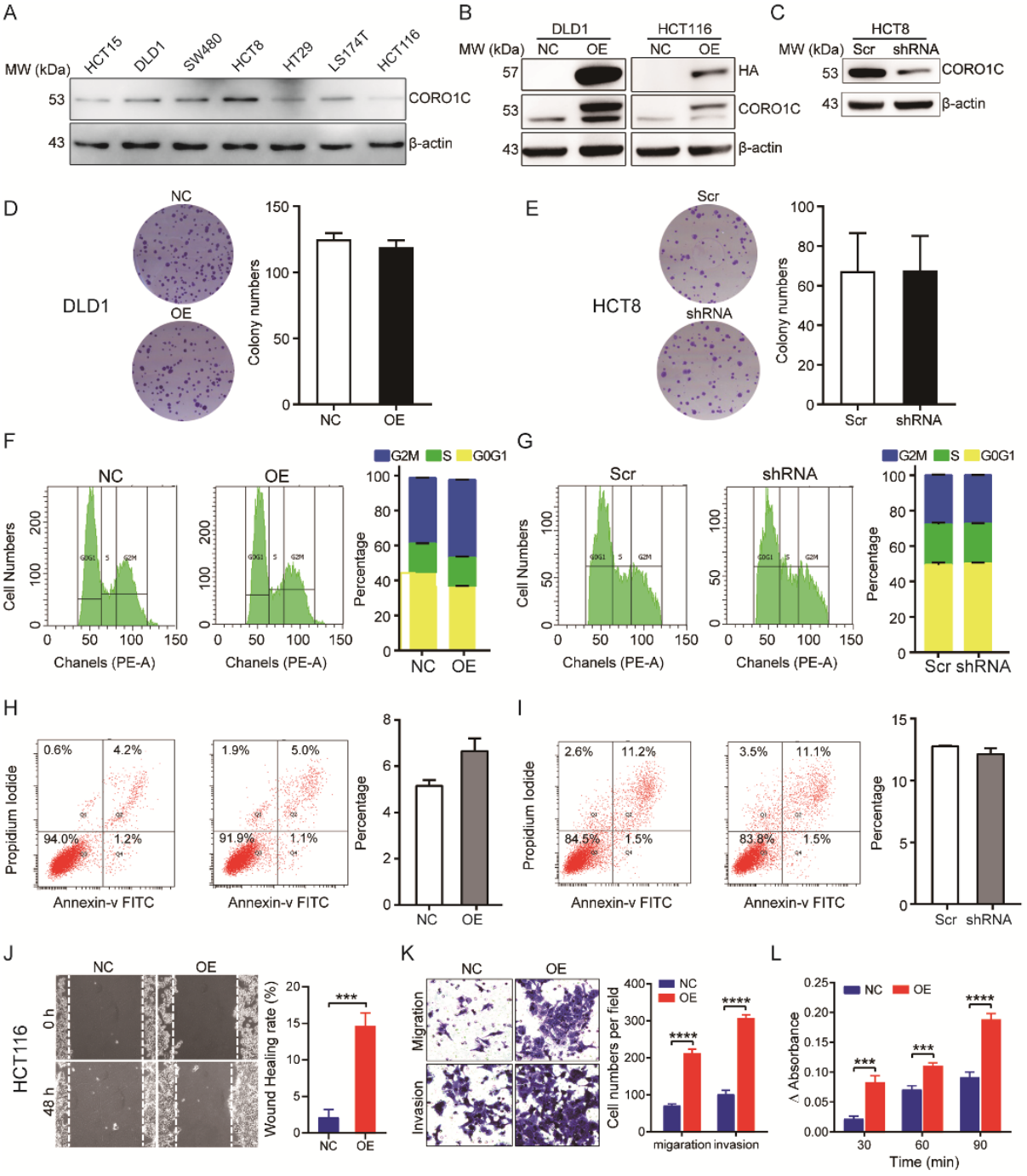

A
CORO1C/Creatinine
ARPC5/Creatinine
RAD23B/Creatinine
GSPT2/Creatinine
NDN/Creatinine
Male
Female
<=60 yr
>60 yr
Right side
Left side
Rectum
<4 cm
>=4 cm
Well-mod.
Poorly
PNI
MVI
EMVI
TNM
I II III IV
T staging
2 3 4
LNM
M0
M1
MSS
MSI
BRAF -V600E
HER2
B
gender
age
location
tumor size
differentiation
PNI
MVI
EMVI
TNM staging
T staging
N staging
M staging
MSI
BRAF
HER2
CORO1C
ARPC5
RAD23B
GSPT2
NDN
−log10(p value)
C
RAD23B
GSPT2
NDN
CORO1C
APRC5
RAD23B/Urine Creatinine
GSPT2/Urine Creatinine
NDN/Urine Creatinine
CORO1C/Urine Creatinine
ARPC5/Urine Creatinine
Male
Female
<=60
>60

Supplementary Fig. 5. The clinical significance of CORO1C, ARPC5, RAD23B, GSPT2 and NDN in CRC patients. The clinical significance of CORO1C, ARPC5, RAD23B, GSPT2 and NDN in CRC patients. (**A**) Data are represented in a box plot. (**B**) Data are represented in a balloon plot. The size and color of each circle represent the -log10 (*P* value), with red and large size correlating to a large *P* value. When the *P* value is equal to 0.05, the color of the circle is white, and the size is 1.3. (**C**) The distribution of urinary levels of CORO1C, APRC5, RAD23B, GSPT2 and NDN among healthy controls. Well-mod, well and moderately differentiated; Poorly, poorly differentiated; PNI, perineural invasion; MVI, microvascular invasion; EMVI, extramural vascular invasion; TNM, tumor, node, metastasis staging; T staging, tumor invasive depth; LNM, lymph node metastasis; M0, without distant metastasis; M1, with distant metastasis; MSS, microsatellite stable; MSI, microsatellite instable; BRAF-V600E, harboring the BRAF-V600E gene mutation; HER2, harboring the HER2 gene amplification. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$, ns, not significant.



Supplementary Fig. 6. Independent validation of the urinary protein signature using dot blot analysis in clinical samples. (**A** and **B**) ROC curves of CORO1C, APRC5, RAD23B and

their combination (diagnostic signature) for the diagnostic model in the training (**A**) and validation (**B**) sets. (**C**-**E**) ROC curve of CORO1C, APRC5, RAD23B and their combination (diagnostic signature) for the diagnostic model to discriminate between the HC and NM (**C**), LNM (**D**), or DM (**E**) groups. (**F** and **G**) ROC curves of CORO1C, GSPT2, RAD23B, NDN and their combination (metastatic signature) for the metastatic model in the training (**F**) and validation (**G**) sets. (H-I) ROC curve of CORO1C, GSPT2, RAD23B, NDN and their combination (metastatic signature) for the metastatic model to discriminate between the NM and LNM (**H**) or LNM (**I**) groups. (**J**) ROC curve of serum CEA, the diagnostic panel and the combination of the diagnostic panel and CEA for the diagnostic model in the validation set. (K-M) ROC curve of serum CEA, the diagnostic panel and the combination of the diagnostic panel and CEA for the diagnostic model to discriminate between the HC and NM (**K**), LNM (**L**), or DM (**M**) groups. (**N**) ROC curve of serum CEA, the metastatic panel and the combination of the metastatic panel and CEA for the metastatic model in the validation set. (**O** and **P**) (**G**) ROC curve of serum CEA, the metastatic panel and the combination of the metastatic panel and CEA for the metastatic model to discriminate between the NM and LNM (**O**) or LNM (**P**) groups. (**Q**) Heatmap of dot plot data for CORO1C, ARPC5, RAD23B, the diagnostic panel and the combination of the diagnostic panel and CEA for diagnostic model in all the samples of the validation group. (**K**) Heatmap of dot plot data for CORO1C, GSPT2, RAD23B, NDN, the metastatic panel and the combination of the metastatic panel and CEA for the metastatic model in all the samples of the validation group. Red: positive using the cutoff value indicated in Supplementary Table 6. The serum CEA, tumor location, sex and age are indicated by color coding (right side).



Supplementary Fig. 7. CORO1C promotes the migration and invasion but not colony formation, cell cycle progression or apoptosis in CRC cells *in vitro*. (A) Western blot detection of CORO1C protein levels in seven CRC cell lines. β-actin was used as a loading control. (B) Western blot analysis showing ectopic expression of CORO1C in the CORO1C (OE) or negative control (NC) vector-transfected DLD1 and HCT116 cells. (C) Western blot analysis showing silencing of CORO1C in HCT8 cells transfected with a specific shRNA (shRNA) or scramble negative control (Scr) vector. (D-E) CORO1C had no effect on colony formation in DLD1 (D) and HCT8 (E) cells. (F-G) Cell cycle analysis showed that CORO1C overexpression or depletion did not affect cell cycle progression in DLD1 (F) or HCT8 (G) cells. (H-I) Ectopic expression or knockdown of CORO1C had no significant effect on cell apoptosis in DLD1 (H) or HCT8 (I) cells. (J) Wound healing assay was used to measure cell front migration at 48 h in CORO1C-transfected HCT116 as well as in negative control cells.

The healing rate of the wound area was compared with that of controls. (K) The migration and invasion capacities were measured by Transwell assays in CORO1C-transfected HCT116 and negative control cells at 48 h after seeding. Representative images are shown on the left, and cell numbers per field are shown on the right panel. (L) Cell adhesion assay of CORO1C-transfected HCT116 as well as negative control cells to fibronectin coated wells. The absorbance at 450 nm is represented for the indicated progressive culture times.

Supplementary Table 1. Quantitative urinary proteomics results in CRC at the discovery stage. (A-D) The TMT quantification data for the four groups. (A) HC group; (B) NM group; (C) LNM group and (D) DM group. The mixed sample was labeled with 131 TMT reagents. Nine individual samples were respectively labeled with the 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, and 130C TMT reagents in each group. (E) The CRC diagnosis-related differential proteins. The differential proteins in the NM, LNM, and DM groups compared to the HC group were annotated. (F) The CRC metastasis-related differential proteins. The differential proteins between the NM and LNM, DM and NM, and LNM and DM groups were annotated. A criterion of ratio-fold change greater than 1.5 was used. (G-H) IPA of differential proteins in NM, LNM, and DM groups. (G) Canonical pathway analysis of differential proteins. -Log (*P*-value) was shown. (H) Disease and Biofunction analysis of differential proteins. The Z-scores are shown. (I-K) Diagnosis and metastasis biomarkers in previous studies. (I) urine; (J) serum/plasma and (K) tumor tissues.

Supplementary Table 2. PRM verification results of the CRC urinary candidate biomarkers. (A and B) PRM validation results of the 41 proteins. The normalized relative abundance of the proteins in each sample was shown. (A) Peptide results and (B) Protein results. (C and D) Summary of the performance of the biomarkers for CRC group discrimination, CRC-diagnosis and CRC-metastasis in the PPM validation. (C) CRC-diagnosis and (D) CRC-metastasis.

Supplementary Table 3. Summary of the performance of the biomarkers for CRC group discrimination, CRC diagnosis and CRC metastasis in the dot blot validation. (A) CRC diagnosis; (B) CRC metastasis; (C) CRC diagnosis performance with 90% sensitivity in the discovery group and (D) CRC metastasis diagnosis performance with 90% sensitivity in the discovery group.

A

	Training AUC	Validation AUC	NM vs HC AUC	LNM vs HC AUC	DM vs HC AUC	Training Sensitivity	Specificity	Validation Sensitivity	Specificity
CORO1C	0.779	0.780	0.713	0.755	0.857	0.714	0.777	0.660	0.824
ARPC5	0.760	0.821	0.695	0.741	0.889	0.533	0.942	0.585	0.922
RAD23B	0.779	0.816	0.702	0.767	0.883	0.505	0.951	0.717	0.804
CEA (≥5ng/ml)	0.850	0.790	0.730	0.845	0.900	0.533	0.981	0.400	0.902
Diagnostic panel	0.828	0.875	0.796	0.814	0.913	0.638	0.922	0.792	0.863
Diagnostic panel + CEA	0.922	0.890	0.825	0.929	0.965	0.810	0.893	0.887	0.804

B

	Training AUC	Validation AUC	LNM vs NM AUC	DM vs NM AUC	Training Sensitivity	Specificity	Validation Sensitivity	Specificity
CORO1C	0.696	0.584	0.648	0.733	0.538	0.852	0.538	0.714
RAD23B	0.695	0.738	0.639	0.792	0.359	0.963	0.872	0.571
GSPT2	0.729	0.632	0.667	0.780	0.808	0.556	0.359	1.000
NDN	0.630	0.692	0.589	0.699	0.872	0.407	0.718	0.643
CEA (≥5ng/ml)	0.766	0.669	0.690	0.812	0.577	0.815	0.513	0.714
Metastatic panel	0.773	0.679	0.710	0.826	0.577	0.926	0.487	0.929
Metastatic panel + CEA	0.854	0.736	0.784	0.906	0.731	0.926	0.821	0.643

C

	Training		Validation	
	Sensitivity	Specificity	Sensitivity	Specificity
CORO1C	0.533	0.903	0.547	0.843
ARPC5	0.552	0.903	0.623	0.882
RAD23B	0.514	0.903	0.887	0.882
Diagnostic panel	0.638	0.903	0.717	0.882
Diagnostic panel + CEA	0.771	0.903	0.774	0.823

D

	Training		Validation	
	Sensitivity	Specificity	Sensitivity	Specificity
CORO1C	0.449	0.889	0.487	0.714
RAD23B	0.410	0.889	0.487	0.857
GSPT2	0.333	0.889	0.308	1.000
NDN	0.308	0.889	0.282	0.928
Metastatic panel	0.603	0.889	0.641	0.500
Metastatic panel + CEA	0.756	0.889	0.821	0.571