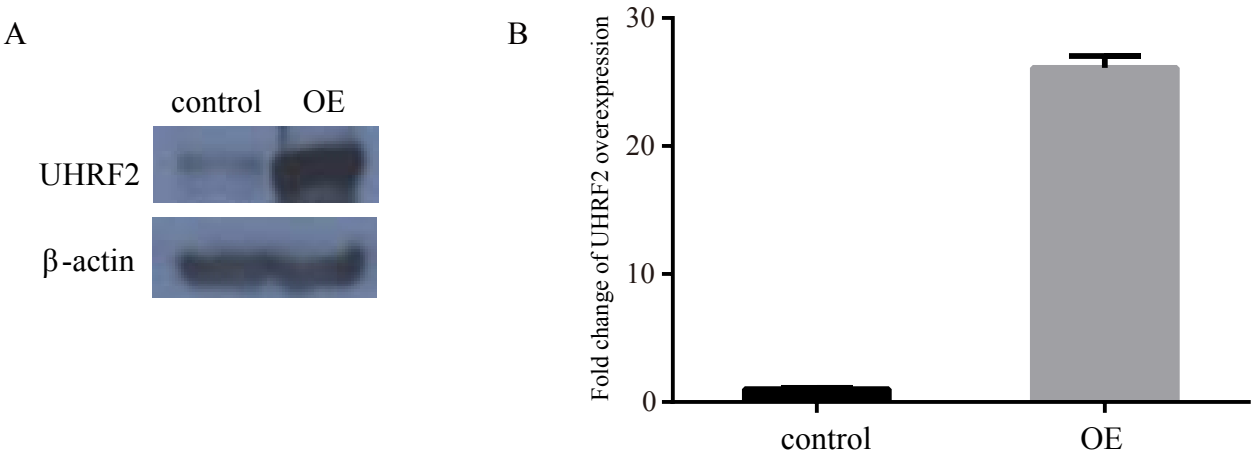
Supplemental information for Multi-dimensional Proteomics Reveals a Role of UHRF2 in the Regulation of Epithelial-Mesenchymal Transition (EMT)

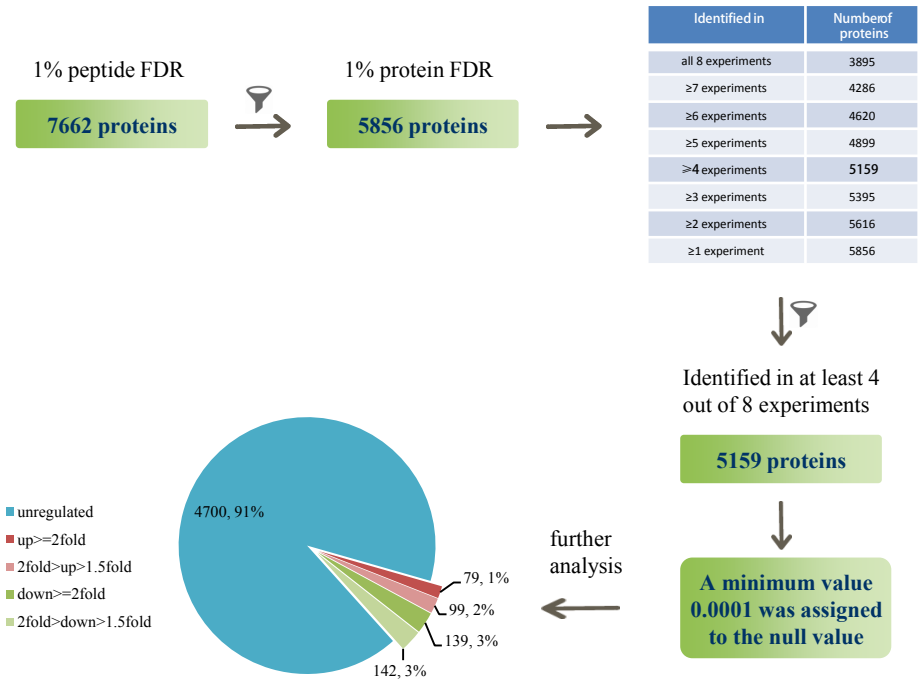
This file contains supplemental figures 1-7 and their legends; supplemental tables 3-4; the description of supplemental tables 1-7 and supplemental methods.

Supplemental figures

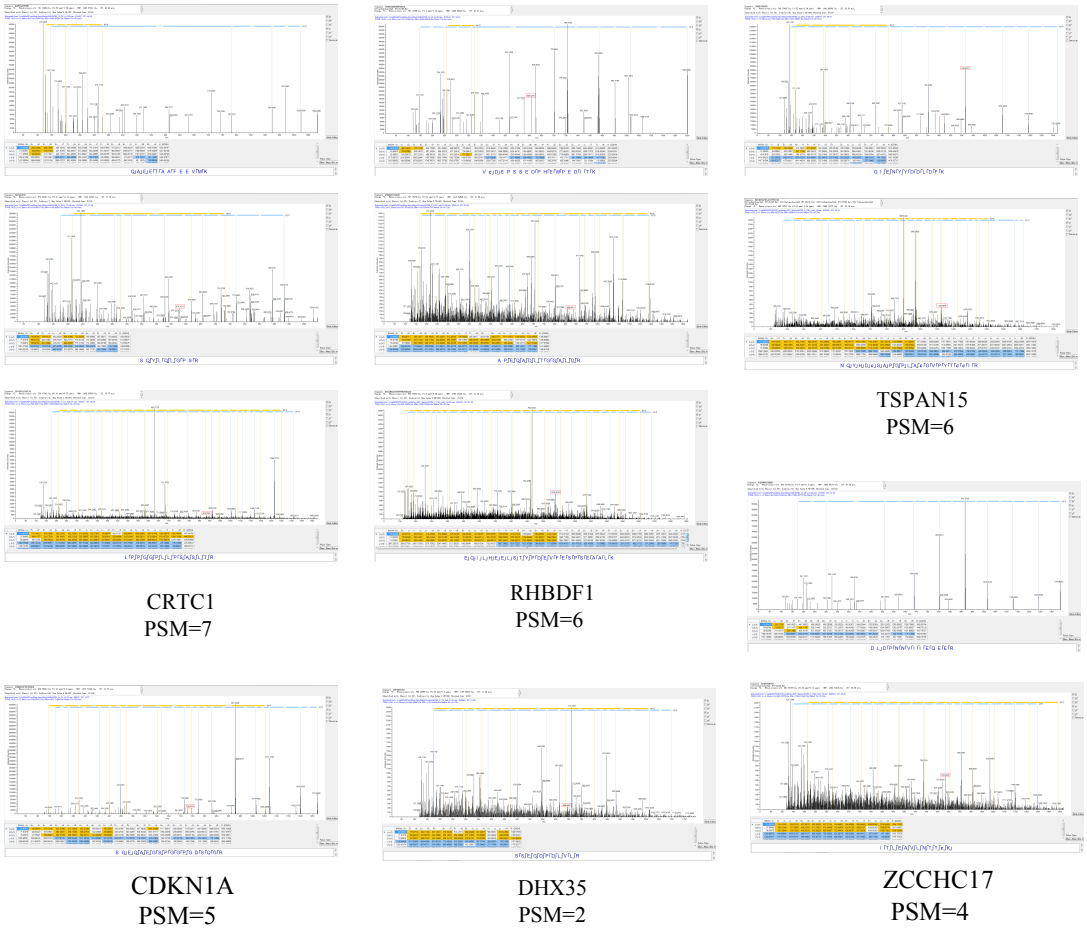


Supplementary Figure 1. UHRF2 ectopic overexpression level in gastric cancer cell line MKN74 was determined by WB and MS. (A) UHRF2 overexpression level was detected by WB. (B) The expression intensity of UHRF2 in control and OE cells was monitored by MS and the relative overexpression level was shown.

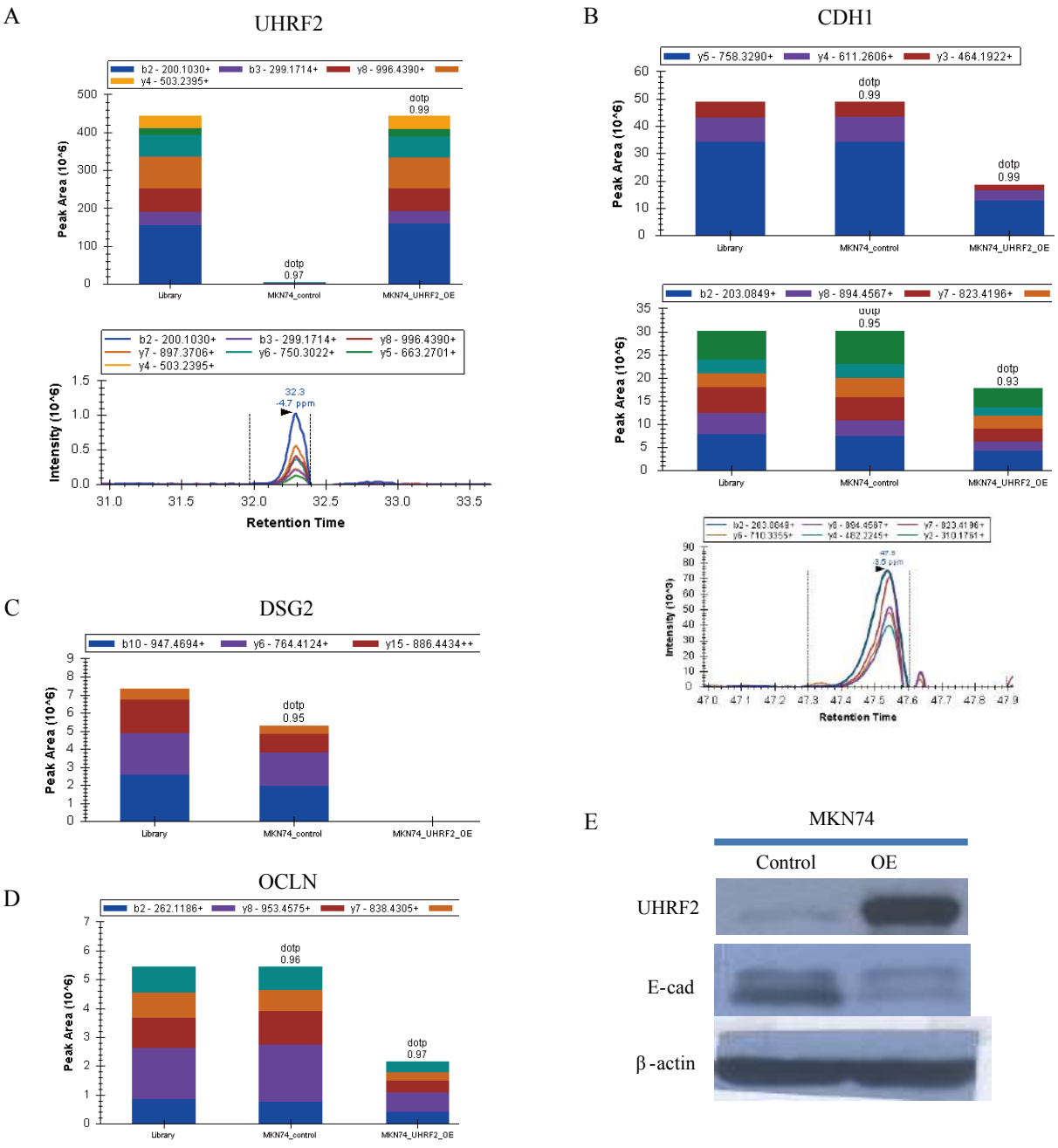
A



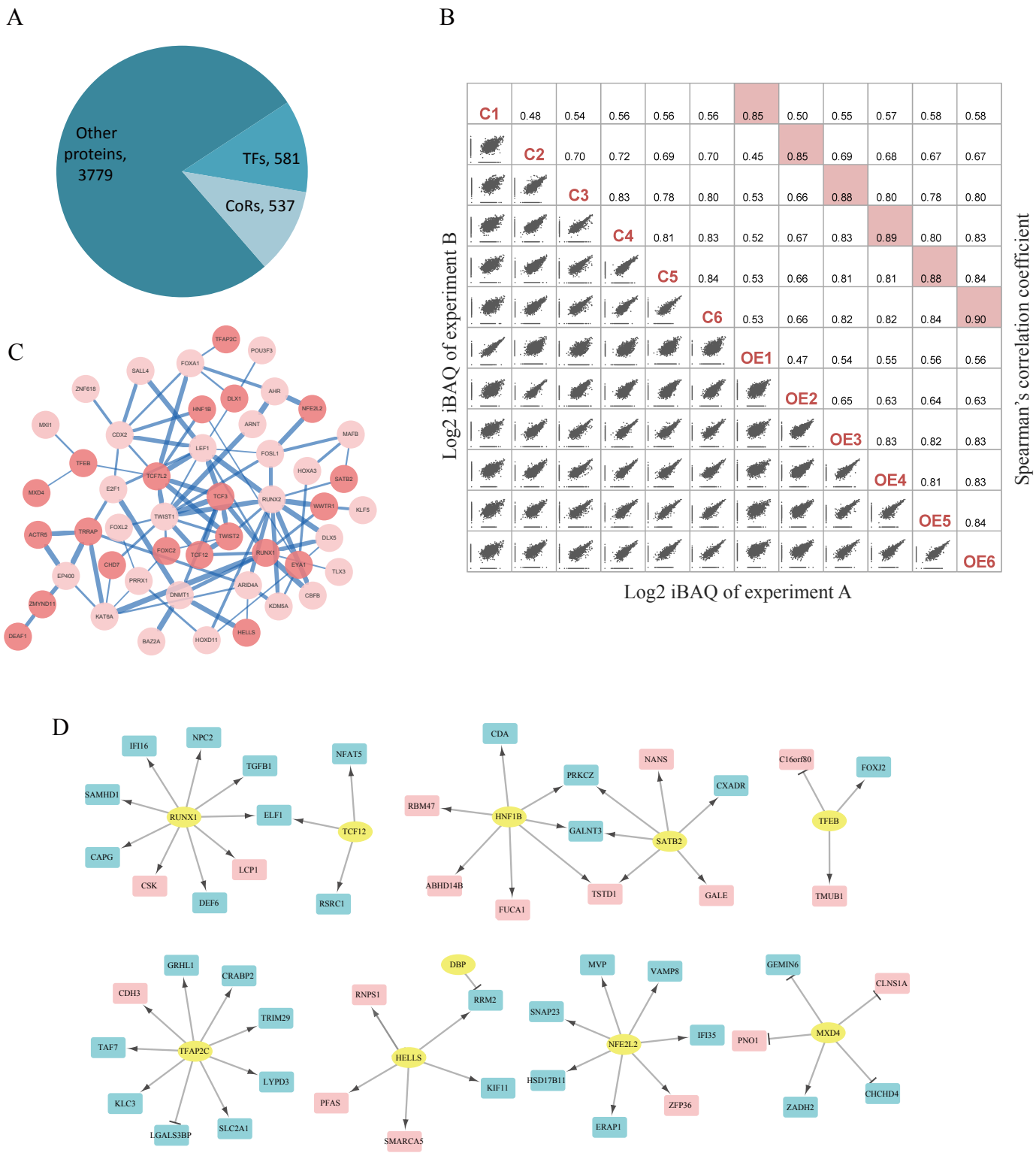
B



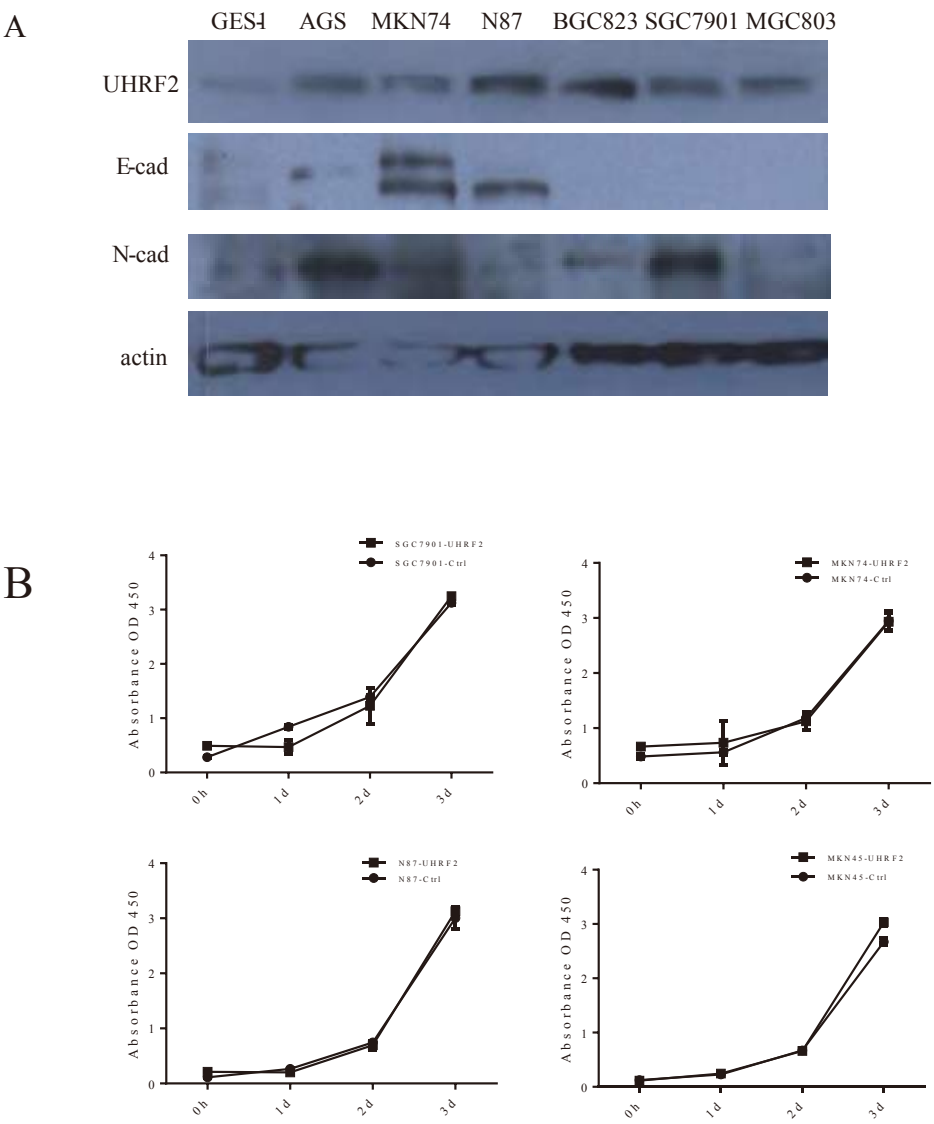
Supplementary Figure 2. The processing of whole lysate profiling data and manual validation of some changed proteins. (A) A total of 7662 proteins were identified at 1% peptide FDR. With the filter of 1% FDR on protein level as a cutoff, 5856 proteins were retained. The numbers of proteins shared in different experiments were listed and proteins identified in at least four out of eight experiments were used for further analysis. The distribution of protein expression changes upon UHRF2 overexpression was displayed. (B) Manual validation of 6 proteins identified by minimum number spectrum with 2-7 psms.



Supplementary Figure 3. The repression of three epithelial proteins CDH1, DSG2 and OCLN by UHRF2 were validated by PRM quantification. (A) UHRF2 overexpression was detected by PRM. Up panel shows intensities of selected peptide of UHRF2 in control and OE cells. Down panel exhibits the transition of the selected peptide. (B) The intensity decrease of CDH1 peptides was monitored in OE cells. (C) Expression change of DSG2 in control and OE cells. (D) Expression change of OCLN was quantified in control and OE cells. (E) The repression of CDH1 by UHRF2 was validated with WB.



Supplementary Figure 4. (A) Classification of proteins identified in TFRE data. (B) Pairwise scatter plots of the experiments. X and y axes represented log2 iBAQ intensity for experiments in corresponding columns and rows, respectively. (C) The network of altered TFs with interaction score>0.4. (D) Activated TFs and their regulated targets.

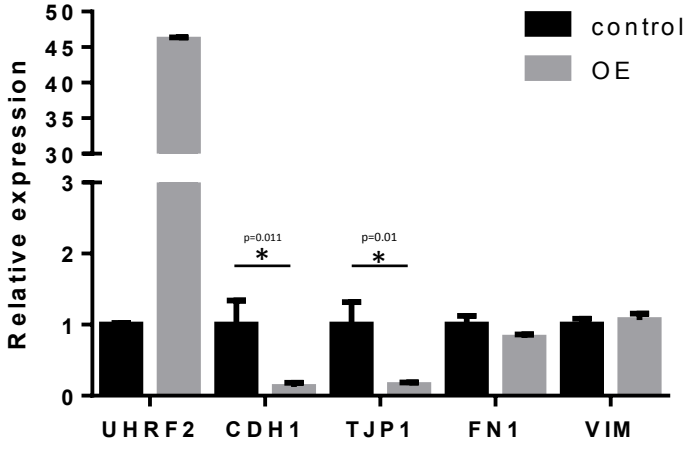


Supplementary Figure 5. (A) The expression levels of UHRF2, E-cad and N-cad in different gastric cancer cell lines. (B) UHRF2 did not affect the growth of four gastric cancer cell lines in our experiments.

A

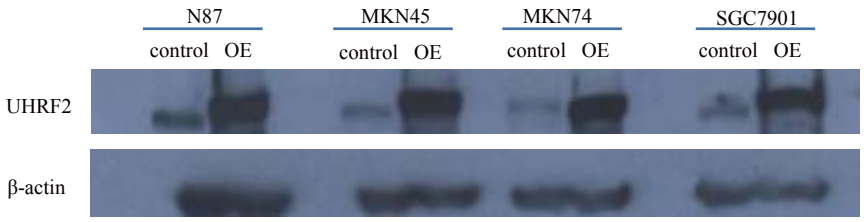
Rank	Motif	P-value	Log P-pvalue	% of Targets	% of Background	STD (Bg STD)
1		1e-73	-1.691e+02	23.62%	12.82%	54.7bp (59.0bp)
2		1e-30	-7.001e+01	34.95%	26.45%	54.4bp (63.1bp)
3		1e-27	-6.294e+01	35.47%	27.37%	55.9bp (59.3bp)
4		1e-26	-6.061e+01	15.07%	9.58%	54.2bp (63.5bp)
5		1e-21	-5.021e+01	19.93%	14.18%	54.8bp (58.9bp)
6		1e-20	-4.610e+01	26.94%	20.66%	54.4bp (64.8bp)
7		1e-19	-4.523e+01	31.89%	25.26%	55.2bp (59.7bp)
8		1e-19	-4.506e+01	9.39%	5.66%	52.3bp (59.6bp)
9		1e-18	-4.150e+01	0.31%	0.00%	53.4bp (23.0bp)
10		1e-17	-3.991e+01	0.26%	0.00%	53.3bp (18.5bp)
11		1e-15	-3.672e+01	0.31%	0.01%	58.9bp (23.5bp)
12		1e-15	-3.658e+01	35.73%	29.56%	56.4bp (58.8bp)
13		1e-14	-3.330e+01	1.54%	0.46%	55.0bp (63.5bp)
14		1e-14	-3.285e+01	0.29%	0.01%	51.8bp (50.5bp)
15		1e-13	-3.124e+01	10.91%	7.50%	56.2bp (60.7bp)
16		1e-13	-3.037e+01	1.44%	0.43%	48.6bp (58.0bp)

B



Supplementary Figure 6. UHRF2 binding motifs enriched in ChIP-seq experiment are listed as the rank of enrichment p-value(A). The RNA levels of several other EMT markers in addition to CDH1 were detected. Epithelial gene TJP1 was also repressed while mesenchymal makers VIM and FN1 were not significantly changed upon UHRF2 overexpression(B).

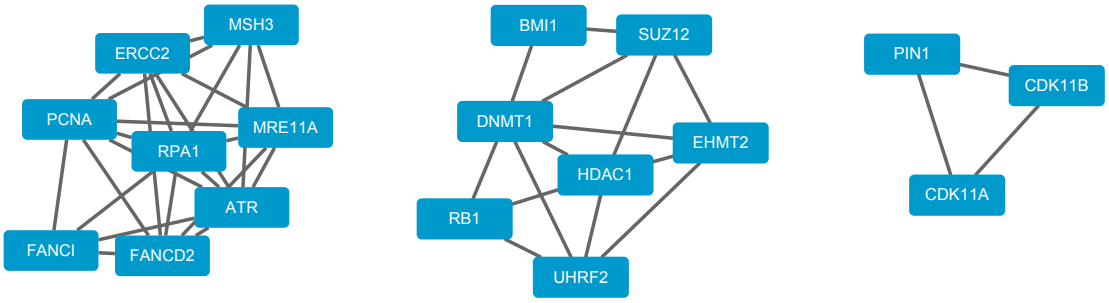
A



B

IP/MS with USP7 antibody		
giAccession	Gene symbol	Unique Peptides
gi150378533	USP7	75
gi23312364	UHRF2	2

C



Cluster 1, score 6.875 Cluster 2, score 4.667 Cluster 3,score 3

D

Components of NuRD Components of PRC1/2

Supplementary Figure 7. Analysis of the UHRF2 interactome. (A)UHRF2 ectopic overexpression level in 4 gastric cancer cell lines. (B) UHRF2 was detected in USP7-IP experiment. (C) Three subnetworks recreated by MCODE analysis are shown. (D) Some components of chromatin-modifying complexes (CMCs) (NuRD, PRC1, PRC2) are shown.

Supplemental tables

Table S1. Summary of proteins identified in profiling of UHRF2-OE and control cells at 1% protein FDR with at least 1 unique peptide. The table lists all proteins that were identified and quantified in 4 experiments at 1% protein FDR with at least 1 unique peptide. All proteins are listed with their corresponding FOT values in each experiment.

Table S2. Summary of TFs and CoRs identified in TFRE experiments with at least 1 unique peptide. The table lists all TFs and CoRs that were identified in 6 experiments with their corresponding iBAQ values. All TFs and CoRs were quantified in each experiment and the corresponding ratios of OE/Control are listed.

Table S3. Signature peptides and parameters of PRM assay.

Table S4. UHRF2 interactors identified with Mass Spectrometry analysis.

Table S5. Information of protein and peptide identification of MS profiling data. The accession number, number of unique peptides assigned for each protein, and % coverage of each protein assigned are contained in layer 1 of this table. The information of peptide identification including sequences, precursor charge, m/z, modifications, and identification scores is listed in layer 2.

Table S6. Information of protein and peptide identification of TFRE data.

Table S7. Information of protein and peptide identification of IP data.

Table S3. Signature peptides and parameters of PRM assay.

GI number	Gene Name	Peptide Sequence	Precursor Mz	Precursor Charge	Product Mz	Product Charge	Fragment Ion
gi|23312364	UHRF2	AQVFSCPACR	598.270975	2	200.102967	1	b2
		AQVFSCPACR	598.270975	2	299.171381	1	b3
		AQVFSCPACR	598.270975	2	996.438983	1	y8
		AQVFSCPACR	598.270975	2	897.370569	1	y7
		AQVFSCPACR	598.270975	2	750.302155	1	y6
		AQVFSCPACR	598.270975	2	663.270126	1	y5
		AQVFSCPACR	598.270975	2	503.239478	1	y4
gi|4757960	CDH1	TIFFCER	486.73402	2	102.054955	1	b1
		TIFFCER	486.73402	2	215.139019	1	b2
		TIFFCER	486.73402	2	871.413085	1	y6
		TIFFCER	486.73402	2	758.329021	1	y5
		TIFFCER	486.73402	2	611.260607	1	y4
		TIFFCER	486.73402	2	464.192193	1	y3
		TIFFCER	486.73402	2	304.161545	1	y2
		TIFFCER	486.73402	2	175.118952	1	y1
		TIFFCER	486.73402	2	379.668149	2	y5
		MALEVGDYK	513.252242	2	203.084874	1	b2
		MALEVGDYK	513.252242	2	894.456724	1	y8
		MALEVGDYK	513.252242	2	823.41961	1	y7
		MALEVGDYK	513.252242	2	710.335546	1	y6
		MALEVGDYK	513.252242	2	482.224539	1	y4
		MALEVGDYK	513.252242	2	310.176132	1	y2
gi|116534898	DSG2	VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	947.469354	1	b10
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	288.168642	2	b5
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	722.862597	2	b15
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	1106.566291	1	y9
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	764.412357	1	y6
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	886.443428	2	y15
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	821.922131	2	y14
		VIQPHGGGSNPLEGTQHLQDVPYVMVR	732.875079	4	382.709816	2	y6
gi|327478416	OCLN	NFDTGLQEYK	607.788035	2	262.118617	1	b2
		NFDTGLQEYK	607.788035	2	838.430509	1	y7
		NFDTGLQEYK	607.788035	2	737.382831	1	y6
		NFDTGLQEYK	607.788035	2	310.176132	1	y2

Table S4. UHRF2-interacting proteins as identified by mass spectrometry analysis

				PSMs							
GeneSymbol	cFAM	complex	Reported interactors	N87-OE	MKN45-OE	MKN74-OE	SGC7901-OE	N87-control	MKN45-control	MKN74-control	SGC7901-control
UHRF2	RING, UBI-E3, TC			764	253	194	186	39	25	11	6
USP7	UBI-DUB			148	49	51	19	4	2	2	2
GMPS				87	28	24	6	1	0	1	0
RPA1	REPAIR			63	19	11	5	1	0	0	0
DNMT1	TC		DNMT1	50	12	2	0	1	1	0	0
HDAC1	TC	NuRD	HDAC1	47	13	5	6	0	1	0	1
CDK11B	TC, KI			39	3	0	0	0	0	0	0
BCLAF1	TC			37	3	0	0	0	0	0	0
CDK11A	TC			36	3	0	0	0	0	0	0
KIF4A				36	3	1	0	0	0	0	0
CHD3	TC, UBI-E3	NuRD		34	9	6	2	0	0	0	0
CUL4B				28	5	5	7	1	1	1	1
CTBP2	TC			27	8	2	1	4	0	0	0
CDC73	TC			24	2	1	1	0	0	0	0
PCNA			PCNA	23	3	2	2	1	0	0	0
CHERP	UBI-UBD			22	2	0	0	0	0	0	0
PSMD5	UBI-PRO			22	7	0	2	2	0	0	0
PUF60				20	7	5	0	1	0	0	0
GATAD2B	DBTF ZNF-GATA	NuRD		18	9	1	1	0	0	0	0
LUC7L2				18	6	1	0	0	1	0	0
FANCI	REPAIR			17	3	0	2	0	0	0	0
MSH3	REPAIR			17	1	0	1	0	0	0	0
ERCC2	REPAIR, TC			15	1	1	0	0	0	0	0
NCAPG2	TC			15	1	0	0	0	0	0	0
DIDO1	CBTC			14	2	1	0	0	0	0	0
MRE11A	REPAIR			14	2	0	0	0	0	0	0
NCAPD3	TC			14	2	0	0	0	0	0	0
SUZ12	TC	PRC2		14	3	0	0	1	0	0	0
CHAF1B	TC			13	3	0	0	1	0	0	0
LUC7L3	PM			13	6	1	0	0	0	0	0
RCOR1	DBTF HOMEO			13	3	1	1	0	0	0	0
KDM2A	TC			12	4	0	1	0	0	0	0
ATR	KI			11	1	1	0	0	0	0	0
FANCD2				11	1	0	0	0	0	0	0
MBD3	DBTC	NuRD		11	1	0	0	0	0	0	0
NR2F2	DBTF NHR			11	1	3	0	0	0	0	0
CDC16				10	1	0	1	0	0	0	0
CHD8	TC			10	2	1	0	0	0	0	0
PBRM1	CBTC			10	1	1	0	0	0	0	0
EHMT2	TC		EHMT2	9	0	0	0	0	0	0	0
RB1	DBTF RB		RB1	9	1	0	0	0	0	0	0
EED		PRC2		8	1	0	0	0	0	0	0
FBXL18	UPS, UBI-FBX			8	2	3	0	0	0	0	0
KDM4B	TC			8	2	0	0	0	0	0	0
BRD3	TC, KI			7	1	0	0	0	0	0	0
CHD2	TC, E2			7	1	0	0	0	0	0	0
TLK1	KI			7	1	0	0	0	0	0	0
ZNF574	DBTF ZNF-C2H2			5	1	0	0	0	0	0	0
BTAF1	TC			4	2	2	0	0	0	0	0
TCF7L2	DBTF HMG			3	1	0	0	0	0	0	0
TFAP4	DBTF HLH			3	1	1	1	0	0	0	0
BMI1	RING, UBI-E3	PRC1		2	1	0	0	0	0	0	0
PCNP			PCNP	2	1	0	0	0	0	0	0
PIN1	TC			2	1	0	0	0	0	0	0
UBE2D3	UBI-E2			2	1	0	0	0	0	0	0

Supplemental methods

PRM

Peptides were prepared from control and UHRF2-OE cells as described above. PRM was performed on Q-Exactive Plus mass spectrometer (Thermo). DDA full scans were processed with samples from control and UHRF2-OE cells to establish peptides library. The unique peptides with high intensity and confidence of each target protein were chosen. PRM methods (collision energy charge state, and retention times) were established and optimized according to information in library of selected peptides. Targeted MS2 spectra were acquired and the raw data were analyzed using Skyline, in which signal intensities for individual peptide sequences of each target protein were quantified.

ChIP-seq data analysis

ChIP-Seq data were mapped to the reference genome hg19 using bowtie, only the uniquely mapped reads are used for the following analysis. Then the sam files were converted into bam files using samtools and 'rmdup' function from samtools was applied for removing potential PCR duplicates. Reads enrichment region (peak) was scanned from the whole genome using MACS14 with the default parameters and the appropriate control. Finally, known and predicted motifs were analyzed with HOMER limited in the default parameters.