SUPPLEMENTAL INFORMATION

Supplemental Methods

Chromatin-immunoprecipitation (ChIP)

ChIP was performed in a similar process as previously reported . Briefly, cells were fixed with 0.5% formaldehyde for 10 minutes at room temperature and subsequently treated with 125 mmol/L glycine for 10 minutes. For each ChIP assay 1x107 Cells were lysed in 750μL lysis buffer [10 mmol/L Tris-Cl (pH8.0), 10 mmol/L EDTA, 0.5 mmol/L EGTA, 0.25% Triton X-100, protease inhibitor cocktail]. Lysates were incubated for 15 minutes at room temperature with rotation and centrifuged. The pellet was resuspended in enrichment buffer [10 mmol/L Tris-Cl (pH 8.0), 200 mmol/L NaCl, 10 mmol/L EDTA, 0.5mmol/L EGTA, protease inhibitor cocktail]. The preparation was incubated for an additional 15 minutes at room temperature with rotation. Insoluble material was pelleted, resuspended in 1 mL immunoprecipitation buffer [20 mmol/L Tris-Cl (pH 8.0), 200 mmol/L NaCl, 0.5% Triton X-100, 0.05%sodium deoxycholate, 0.5% NP40, protease inhibitor cocktail], and sonicated to generate fragmented chromatin with approximately 500 bp in length. For each ChIP, 1μg antibody for OCT1, or 1μg nonspecific immunoglobulin G (Sigma) was added and incubated overnight at 4°C. Immune complexes were precipitated for 1 hour at room temperature with 1/20th volume protein-G beads. Beads were washed three times with immunoprecipitation buffer, once with high salt buffer (immunoprecipitation buffer with 800 mmol/L NaCl), once in LiCl buffer [10 mmol/L Tris-Cl (pH 8.0), 250 mmol/L LiCl, 1% Triton X-100, 0.5% NP40, 0.1% sodium deoxycholate, 5 mmol/L EDTA, protease inhibitor cocktail], and twice with Tris-EDTA. Immune complexes were eluted with 40 μL elution buffer [10 mmol/L Tris-Cl (pH8.0), 1% SDS, 5 mmol/L EDTA] at 65°C for 1 hour and cross-links were reversed by adding 200 mmol/L NaCl and incubating at 65°C overnight. Eluted material was treated with 30 μg proteinase K for 2 hours at 42°C and purified (QIAGEN). Amplification of DNA from OCT1 ChIP was carried out with PCR reaction containing 12.5μL SYBR green master mix (QIAGEN), 7μL water, 3μL of 5μM primer mix, and 2.5μL

DNA. An iCycler iQ real-time PCR detection system (Bio-Rad) was used to perform the quantitative PCR.

Cell viability assays

The MGC803 cells transfected with OCT1 vector in the presence or absence of synbindin siRNA and control vector were seeded onto 96-well plates at 2500cells/well. Cell proliferation was measured using the Cell Counting Kit-8 (CCK-8, Dojindo). At each time point, cells were incubated with 10mL CCK-8 reagent per well (100 ml medium/well) for 30min at 37°C, 5% CO_2. The absorbance was measured at 450 nm.

Flow cytometric analysis

Cell cycle progression was assayed by DNA content using propidium iodide and flow cytometry. MGC803 cells were transfected with OCT1 vector, OCT1 vector+synbindin siRNA and control vector. Approximately $1×106$ cells were trypsinized and washed twice with ice-cold PBS and then fixed overnight at -20°C in 70% ethanol. Immediately before flow cytometry, the cells were resuspended in PBS containing propidium iodide (50 μg/ml) and DNase-free RNase (10 μg/ml). Flow cytometry was performed using a FACScalibur (BD biosciences) system with CELLquest software.

Tumor cell invasion assays

Tumor cell invasion assays were performed using Boyden chambers with filter inserts (pore size, 8-μm) coated with Matrigel (40 μg; Collaborative Biomedical, Becton Dickinson Labware, Bedford, MA) in 24-well dishes (Nucleopore, Pleasanton, CA, USA) as described previously [1]. Briefly, $1×10^5$ cells after transfected with OCT1 vector in the presence or absence of synbindin siRNA were seeded in the upper chamber, while the same medium was placed in the lower chamber. The plates were incubated for 24 h. Then the cells were fixed in 4% formaldehyde and stained with 0.05% crystal violet in PBS for 20min at room temperature. Cells on the upper side of the filters were removed by cotton-tipped swabs, and the filters were washed with PBS. The cells on the lower side of the filters were defined as invasive cells and counted at x200 magnification in 10 different fields of each filter.

Annexin V analysis

The apoptotic status was analyzed by using a TACS annexin V-FITC kit (R&D Systems, Minneapolis, MN). Briefly, cells (1×10^6cells/mL) were transfected with OCT1 vector in the presence or absence of synbindin siRNA. After washing in ice-cold PBS, cells were collected by EDTA treatment and incubated with a mixture of annexin V-FITC and PI for 15 minutes at room temperature according to the manufacturer's instruction. Early apoptotic (only annexin V-positive) cells were distinguished from late apoptotic or necrotic (annexin V and PI double-positive) cells by a flow cytometric analysis.

Array-based Copy Number Variation (CNV) analysis

The altered chromosome segments in 293 stomach adenocarcinoma patients of The Cancer Genome Atlas (TCGA) cohort were obtained via the UCSC cancer genome browser [2], and the CNV data of Singapore cohort can be accessed from GEO database (GSE31168). Both datasets were based on the Affymetrix SNP6.0 microarray, and were analyzed by GISTIC 2.0 algorithm [3] using default parameters for 'threshold' and 'focal' modes, respectively. The VUMC data were based on customized array-CGH method, and was obtained from GEO database (GSE26389). The compatible CGHCall algorithm [4] was used calculate the CNV of all genes for in VUMC dataset. To identify significant relationships between the CNVs of OCT1 and other genes, a dimension reduction permutation (DRP) statistical algorithm was used to analyze the TCGA CNV dataset. The DRP algorithm has been described previously [5].

Gene Set Enrichment Analysis (GSEA)

GSEA is a method of analyzing and interpreting microarray and such data using biological knowledge [6]. The data in question is analyzed in terms of their differential enrichment in a predefined biological set of genes (representing pathways). These predefined biological sets can be published information about biochemical pathway or coexpression in a previous experiment. GSEA was performed using GSEA version 2.0 from the Broad Institute at MIT. Two datasets were analyzed by GSEA: a 'multi-cancer' dataset [7] including 64 solid tumors and 27 normal tissues (GSE28866) and The Cancer Genome Atlas (TCGA) dataset including 282 GC samples. Both datasets were determined by 3'-End Sequencing for Expression Quantification (3SEQ). In this study, GSEA firstly

generated an ordered list of all genes according to their correlation with OCT1 expression, and then a predefined gene set (signature of gene expression upon perturbation of certain cancer-related gene) receives an enrichment score (ES), which is a measure of statistical evidence rejecting the null hypothesis that its members are randomly distributed in the ordered list. Parameters used for the analysis were as follows. The "c6.all.v4.0.symbols.gmt" gene sets were used for running GSEA and 1000 permutations were used to calculate *P*-value and permutation type was set to gene_set. The maximum gene set size was fixed at 1500 genes, and the minimum size fixed at 15 genes. The expression level of OCT1 (POU2F1) was used as phenotype label, and "Metric for ranking genes" was set to Pearson Correlation. All other basic and advanced fields were set to default.

To overcome gene-set redundancy and help in the interpretation of altered pathways, we applied the "Enrichment Map" algorithm, a network-based method for interpreting gene-set enrichment results [8]. Gene-sets are organized in a network, where each set is a node and links represent gene overlap between sets. Automated network layout groups related gene-sets into network clusters, helping to identify the major enriched functional themes. The default parameters were used to generate enrichment networks: P-value cutoff=0.005; FDR Q-value cutoff=0.1; overlap coefficient cutoff=0.75.

Prediction of OCT1 binding sites by 3D structure-based energy calculations

An initial assignment using the TFsearch prediction tool indicated potential OCT1-binding sites within the synbindin promoter (ranged from -700 to TSS). Then, a more dedicated 3D structure-based method was used to identify OCT1-binding site with higher confidence.

Prediction of OCT1 binding sites based on ChIP-seq data

The features of OCT1- binding sequences can be interpreted by the enriched stretches in OCT1 ChIP-seq data. A previous study has identified 5-mer stretches that are specifically enriched in top 1% OCT1-binding sequences [9]. We analyzed the enrichment of OCT1-binding motifs within the scanning window (described above) as a measure to

search for potential OCT1 binding site in the synbindin promoter. This method was combined with 3DTF algorithm to identify highly confident candidate sequences.

Western blotting

Western blot assays were performed to examine OCT1 and synbindin proteins. Cell extracts were prepared using RIPA buffer (Thermo Fisher) from cells that were treated as indicated. After electrophoresis, proteins were electroeluted at 120 volts onto a polyvinylidene difluoride (PVDF) membrane (Invitrogen). Primary antibodies raised OCT1 were purchased from Abcam. The anti-synbindin antibody was purchased from Abnova, and the α-tubulin antibody was used as a control. The Western blotting analysis was repeated at least three times.

In vivo **experiments**

Briefly, male BALB/c athymic nude mice (4–6 weeks old) were obtained from the experimental animal center of Shanghai Institute for Biological Sciences (SIBS). All mice were injected subcutaneously into the right side of back with 1.0×107 stable MGC803 cells transfected with OCT1 vector and control vector to establish the GC xenograft model. 14 days after first injection,the two groups were sacrificed and the xenograft tissues were taken for Western blot and Immunofluorescence. Tumor diameters were measured at regular intervals with digital calipers, and tumor volume was calculated by the formula: tumor volume (mm3) = shorter diameter2 × longer diameter/2. The tumor volume data are presented as means ± SD (n = 6).
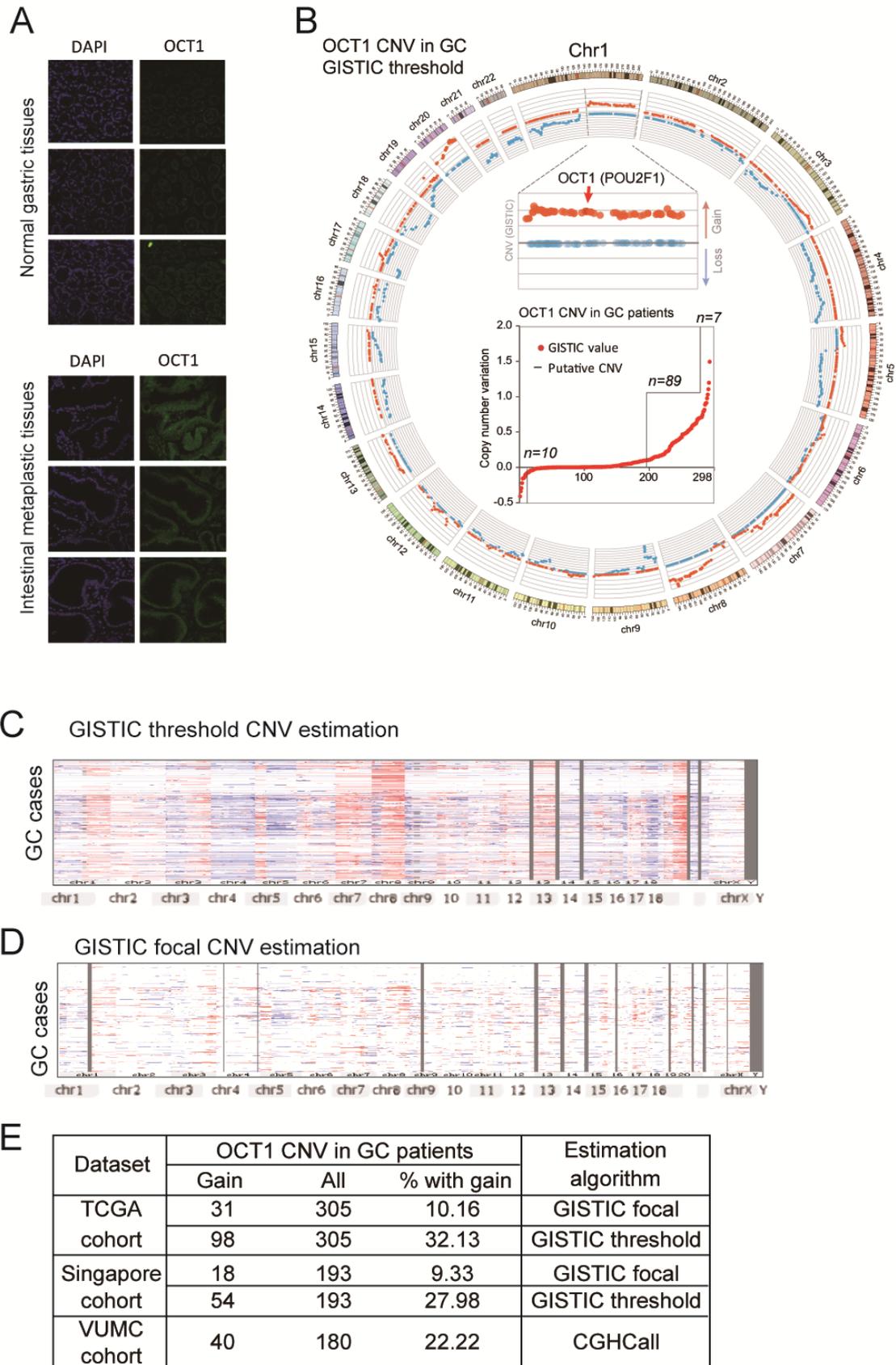
Supplementary Figures

Figure S1. Expression and amplification of OCT1 gene in gastric tissues.

(**A**) OCT1 is upregulated in intestinal metaplastic tissues. The upper panel shows OCT1 immunofluorescent staining (green) in normal gastric tissue, while the lower panel presents OCT1 expression in intestinal metaplastic tissues. Cell nucleus were stained with DAPI in blue.

(**B**) The Circos plot shows gene copy number variation (CNV) in the TCGA stomach adenocarcinoma (STAD) cohort. The level of CNV was based on the mean aggregation of log ratio signals of all probes within each gene, regardless of the level of alterations (focal or broad). OCT1 is located in a recurrently amplified region in chromosome 1 (magnified plot). The lower plot shows the CNV (log2 ratio and GISTIC-estimated events) of OCT1 in all patients by increasing order. While the copy number of OCT1 was increased in 96 cases, it was decreased in only 10 patients.

(**C**) Heat map showing CNVs of the TCGA gastric cancer cohort including both broad and focal events. Each column represents a segmented genomic region, and each row represents a GC patient.

(**D**) Focal CNV events in TCGA gastric cancer cohort, wherein the broad CNV events have been subtracted.

(**E**) Summary on the prevalence of OCT1 amplification in three indicated GC cohorts. The TCGA and Singapore datasets (based on SNP array platform) were analyzed by GISTIC algorithm for focal and broad CNV events. The frequencies for OCT1 amplification (including both focal and broad events) were respectively 32.1% (TCGA cohort) and 28.0% (Singapore cohort). Focal amplification of OCT1 was found in 10.2% cases in TCGA cohort and 9.3% patients in Singapore cohort. The VUMC dataset was based on ArrayCGH platform and analyzed CGHCall algorithm, which revealed 22.2% GC cases with amplified OCT1 gene.
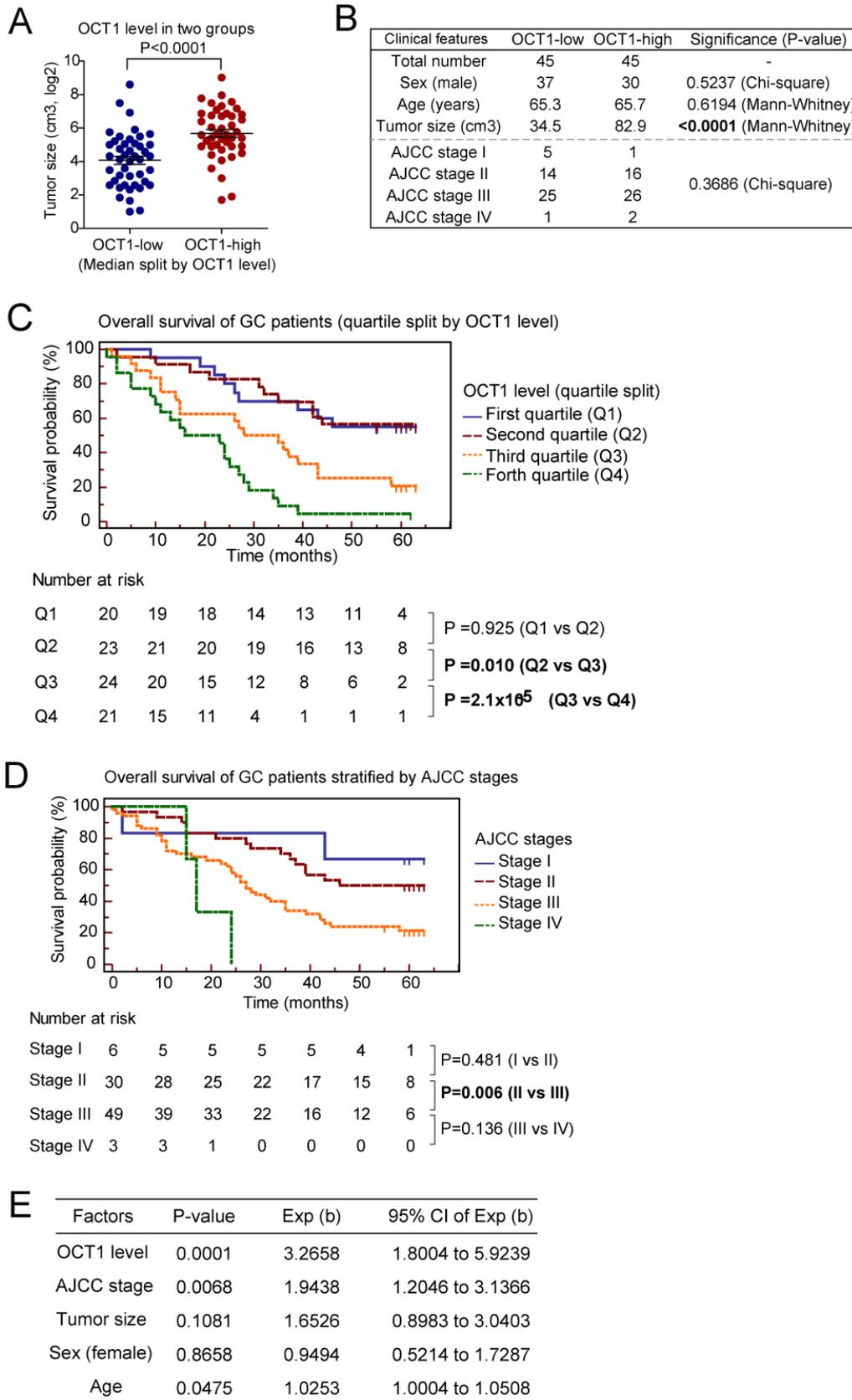
A OCT1 level in two groups
P<0.0001

B

Clinical features	OCT1-low	OCT1-high	Significance (P-value)
Total number	45	45	-
Sex (male)	37	30	0.5237 (Chi-square)
Age (years)	65.3	65.7	0.6194 (Mann-Whitney)
Tumor size (cm3)	34.5	82.9	**<0.0001** (Mann-Whitney)
AJCC stage I	5	1	
AJCC stage II	14	16	0.3686 (Chi-square)
AJCC stage III	25	26	
AJCC stage IV	1	2	

C Overall survival of GC patients (quartile split by OCT1 level)

OCT1 level (quartile split)
— First quartile (Q1)
— Second quartile (Q2)
— Third quartile (Q3)
— Forth quartile (Q4)

Number at risk

Q1	20	19	18	14	13	11	4
Q2	23	21	20	19	16	13	8
Q3	24	20	15	12	8	6	2
Q4	21	15	11	4	1	1	1

P =0.925 (Q1 vs Q2)
P =0.010 (Q2 vs Q3)
P =2.1x10^5 (Q3 vs Q4)

D Overall survival of GC patients stratified by AJCC stages

AJCC stages
— Stage I
— Stage II
— Stage III
— Stage IV

Number at risk

Stage I	6	5	5	5	5	4	1
Stage II	30	28	25	22	17	15	8
Stage III	49	39	33	22	16	12	6
Stage IV	3	3	1	0	0	0	0

P=0.481 (I vs II)
P=0.006 (II vs III)
P=0.136 (III vs IV)

E

Factors	P-value	Exp (b)	95% CI of Exp (b)
OCT1 level	0.0001	3.2658	1.8004 to 5.9239
AJCC stage	0.0068	1.9438	1.2046 to 3.1366
Tumor size	0.1081	1.6526	0.8983 to 3.0403
Sex (female)	0.8658	0.9494	0.5214 to 1.7287
Age	0.0475	1.0253	1.0004 to 1.0508

Figure S2. Association between OCT1 expression and clinicopathological patterns of GC.

(**A**) Statistical analysis on the staining intensity of OCT1 in patients stratified by tumor

sizes (median split). Tumors with larger sizes showed significantly higher OCT1 staining intensity than those with smaller tumor sizes (P<0.0001, Mann Whitney test).

(**B**) Association between OCT1 expression and clinicopathological features of gastric cancer patients. All 90 patients were stratified by OCT1 expression levelinto OCT1-low and OCT1-high groups (median split), and the clinicopathological features were compared between two groups using the statistical method indicated.

(**C**) Gastric cancer patients (n=90) were divided into 4 groups according to OCT1 expression levels (quartile split into Q1-Q4), and their survival were compared using Kaplan-Meier survival analysis. The numbers of cases at risk are shown in the lower panel. Significant differences were found between the pairs Q1 v.s. Q3 (P=0.017), Q2 v.s. Q3 (P=0.010), and Q3 v.s. Q4 (P=2.1×10^{-5}).

(**D**) Survival of patients with different AJCC grades were compared by Kaplan-Meier survival analysis. Significant different was only found between the groups with AJCC state II and stage III (P=0.006).

(**E**) Multivariate Cox regression survival analysis suggested independent prognostic value of OCT1. Different factors (including OCT1, AJCC stage, tumor size, sex and age) were analyzed for their association with patient survival using Cox regression model. The *P*-value, Exp (b) (hazard ratio) and its 95% confident interval (CI) are shown for each factor.
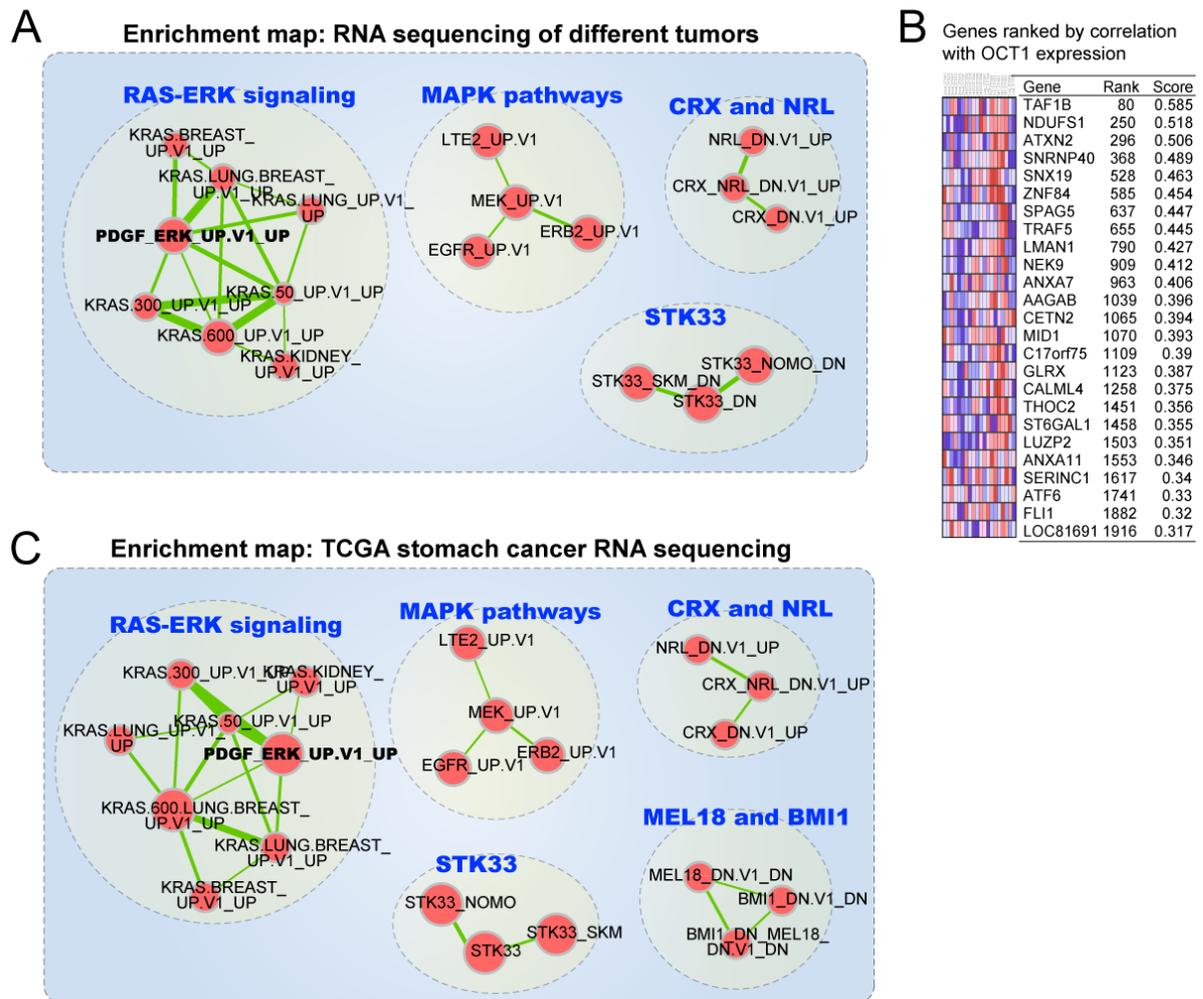
Figure S3. Enrichment map summary of GSEA on OCT1-associated pathways .

(**A**) OCT1-associated gene sets were identified by GSEA on gene expression profiles in multiple tumors. The multi-cancer dataset (GSE28866) included 64 solid tumors and 27 normal tissues, and a ranked gene list was calculated according to the correlation with OCT1 expression. Detailed analysis method for can be found in Methods section. The Enrichment Map analysis was used to overcome gene-set redundancy and improve the interpretation of large gene lists. Multiple gene sets related to RAS-ERK signaling were found to be associated with OCT1 expression (dashed circle), with PDGF/ERK signature as the most correlated gene set. Other clusters of gene sets include MAPK pathways, CRX and NRL, STK33, etc.

(**B**) The expression of ranked genes in the PDGF/ERK signature is shown in the heat map.

(**C**) The OCT1-associated gene sets were determined using TCGA stomach adenocarcinoma (STAD) gene expression profiles. The expression of all genes were determined by 3' RNA sequencing method, and the GSEA/Enrichment Map analyses were performed using the same approaches as in (**A**). The TCGA gastric cancer data also identified the strong correlation between OCT1 expression and RAS-ERK signaling, with PDGF/ERK signature being the most correlated gene set.
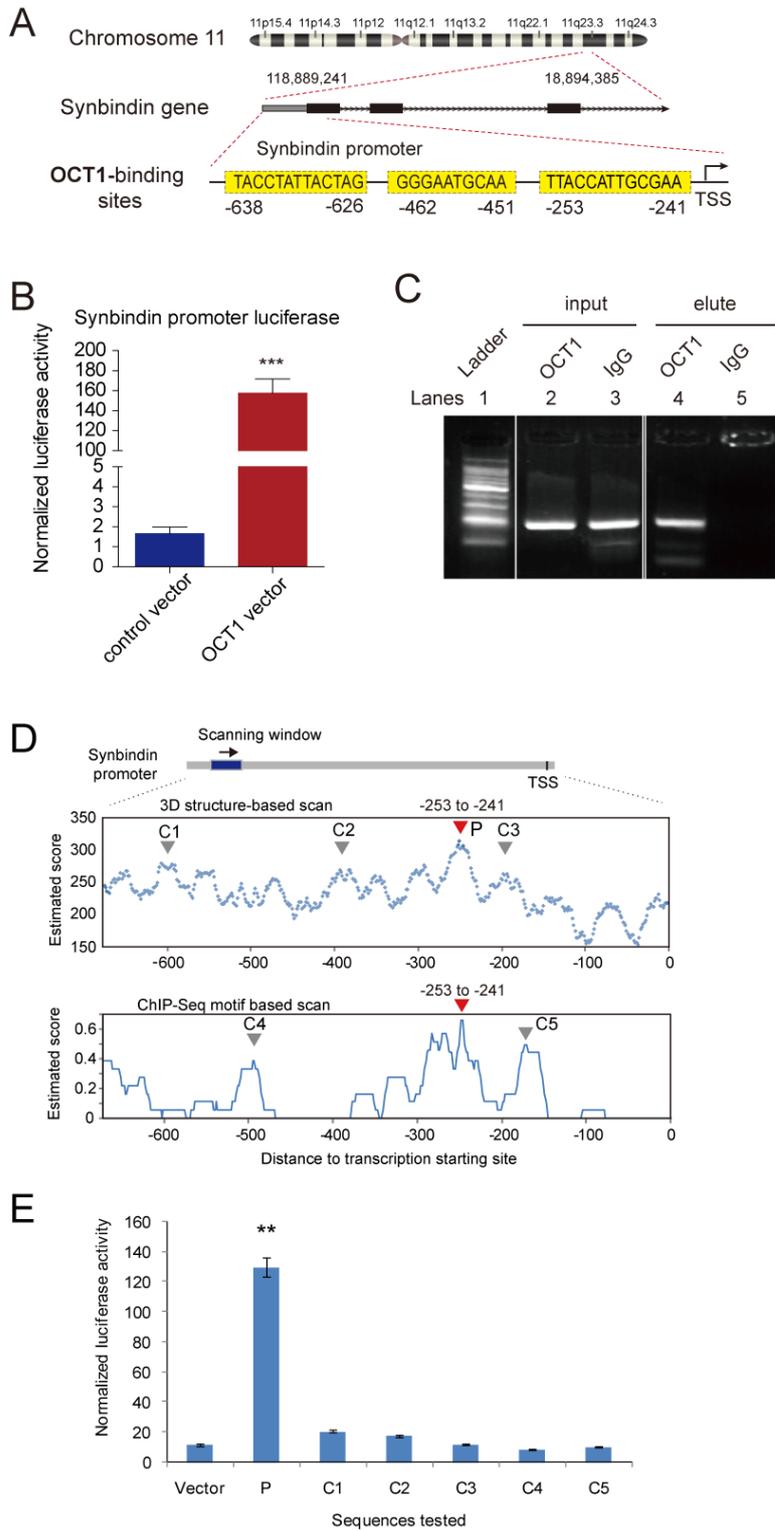
Figure S4. OCT1 binds to the promoter region of synbindin.

(**A**) An initial assignment on OCT1-binding sites on the synbindin promoter using the TFsearch prediction tool. The putative OCT1-binding sites were found upstream (-700) of the transcription start site (TSS) of synbindin gene.

(**B**) Luciferase reporter assay showing the transactivation of synbindin promoter by

OCT1. The synbindin promoter sequence (-654 to -241bp upstream of TSS) was inserted to a reporter vector, and then co-transfected with OCT1 in MGC803 cells. Ectopic expression of OCT1 strongly increased luciferase activity (*** $P<0.001$, two-sided student t-test).

(**C**) Chromatin immunoprecipitation (ChIP) assay showing the binding of OCT to synbindin promoter *in vivo*. The promoter region of synbindin was amplified from the DNA recovered from the immunoprecipitation complex using a specific antibody for OCT1. The input DNA and ChIP yield using non-specific IgG are included as controls.

(**D**) The sequences tested in luciferase assays were picked up from the prediction results by 3DTF (upper panel) and ChIP seq-based (lower panel) methods. The red arrow heads show the position of the true binding site, while grey arrow heads indicate the other tested sequences.

(**E**) Luciferase assay revealed that OCT1 transactivates the predicted binding site (marked as 'P') but not other control sites (C1-C5). (** $P<0.01$, t-test)
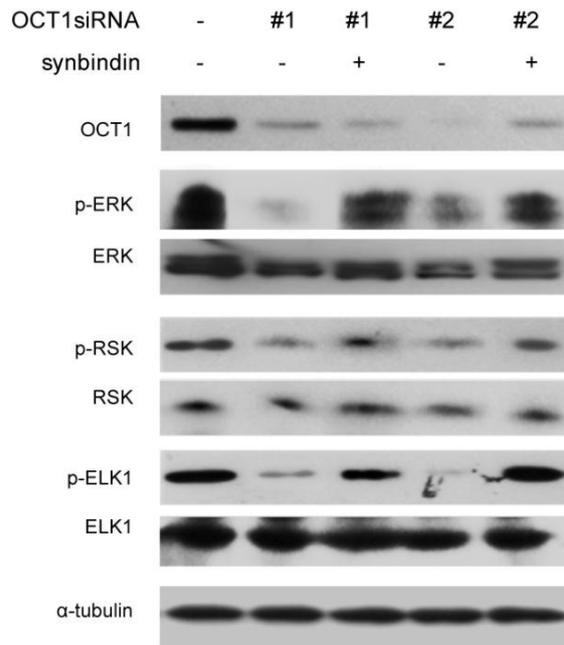
Figure S5. Effects of OCT1 silencing and synbindin ectopic expression on the phosphorylation of ERK and substrates in MKN45 cells.

The MKN45 cells were transfected with siRNAs for OCT1 (#1 and #2, respectively) in the absence or presence of synbindin expression vector, and Western Blot was used to determined the levels of the indicated proteins/phosphor-proteins. While OCT1 silencing suppressed phosphorylation of ERK/RSK/ELK1, synbindin overexpression blocked the effect of OCT1 silencing.
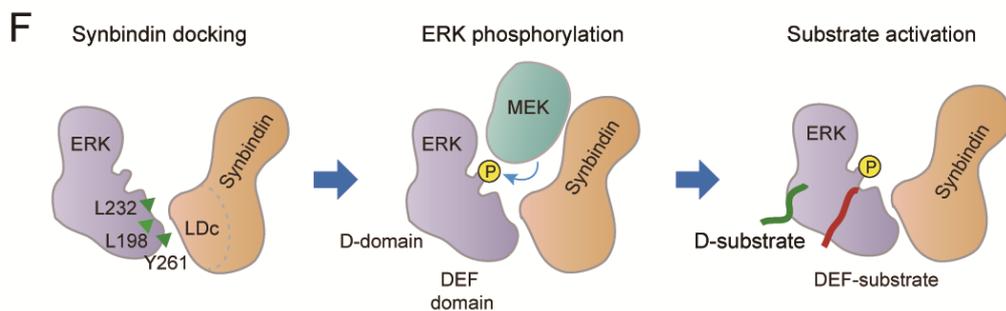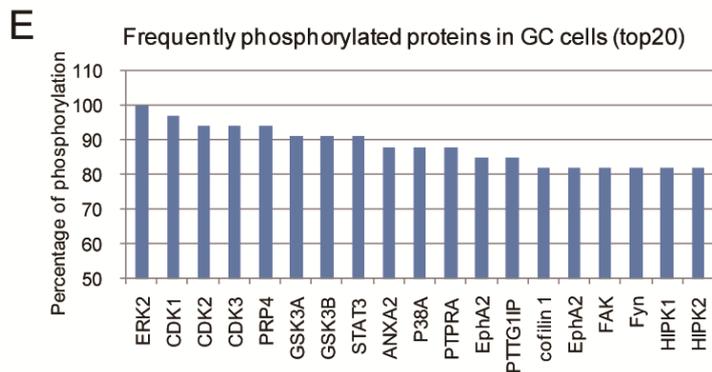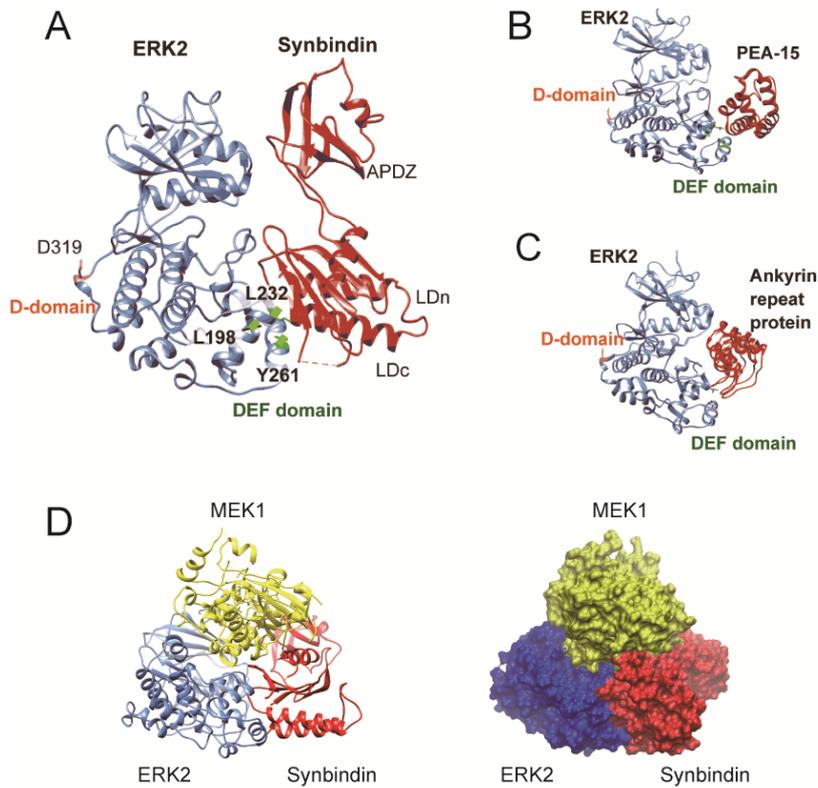
Figure S6. Synbindin binds to DEF-domain of ERK2 protein

(**A**) The binding mode of ERK-synbindin complex as predicted by the Hex spherical polar Fourier protein docking algorithm. In this model, the LDc domain of synbindin interacts with the DEF-domain of ERK2, in contact with its residues L191, L232 and Y261.

(**B**) Crystallography-determined structure of ERK/PEA15 binding complex (PDB accession 4IZA), wherein PEA15 binds to the DEF-domain of ERK but not D-domain.

(**C**) The structure of ERK binding to a designed Ankyrin Repeat protein, based on crystallography data (PDB accession 3ZUV). The DEF-domain of ERK is also involved in the interaction with Ankyrin Repeat protein.

(**D**) The ERK2/synbindin heteroduplex structure (validated by GST pull-down) created a binding surface to MEK1, and the picture shows a predicted structure of MEK1 binding to the ERK2/synbindin duplex. These results suggest that synbindin may facilitate MEK1/ERK2 interaction.

(**E**) The histogram shows the most frequently phosphorylated proteins found in gastric cancer cell lines, based on the PhosphoSite database (combined data from high-throughput proteomic experiments). The phosphorylation of ERK2 protein was identified in all the 32 analyzed gastric cancer cell lines, which represents the most frequent phosphorylation event in the whole proteome. Altogether 6496 proteins were found phosphorylated by different frequencies, ranging from 1/32 to 32/32. Synbindin was not found phosphorylated in any analyzed cell line, suggesting its phosphorylation is not required for ERK phosphorylation in GC cells.

(**F**) Proposed model for synbindin-involved ERK activation. The left panel shows the interaction between synbindin LDc domain and ERK DEF-domain (experimentally validated), which creates a binding surface for MEK1. The binding with MEK1 facilitates ERK phosphorylation (middle panel), which enables ERK to activate its substrates docked to DEF and D-domains (right panel).

Supplemental References

1   Kong X, Qian J, Chen LS, Wang YC, Wang JL, Chen H, *et al.* Synbindin in extracellular signal-regulated protein kinase spatial regulation and gastric cancer aggressiveness. Journal of the National Cancer Institute 2013;**105**:1738-49.

2   Goldman M, Craft B, Swatloski T, Ellrott K, Cline M, Diekhans M, *et al.* The UCSC Cancer Genomics Browser: update 2013. Nucleic acids research 2013;**41**:D949-54.

3   Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome biology 2011;**12**:R41.

4   van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics 2007;**23**:892-4.

5   Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. Nature 2008;**455**:1069-75.

6   Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics 2007;**23**:3251-3.

7   Brunner AL, Beck AH, Edris B, Sweeney RT, Zhu SX, Li R, *et al.* Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. Genome biology 2012;**13**:R75.

8   Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS one 2010;**5**:e13984.

9   Ferraris L, Stewart AP, Kang J, DeSimone AM, Gemberling M, Tantin D, *et al.* Combinatorial binding of transcription factors in the pluripotency control regions of the genome. Genome research 2011;**21**:1055-64.