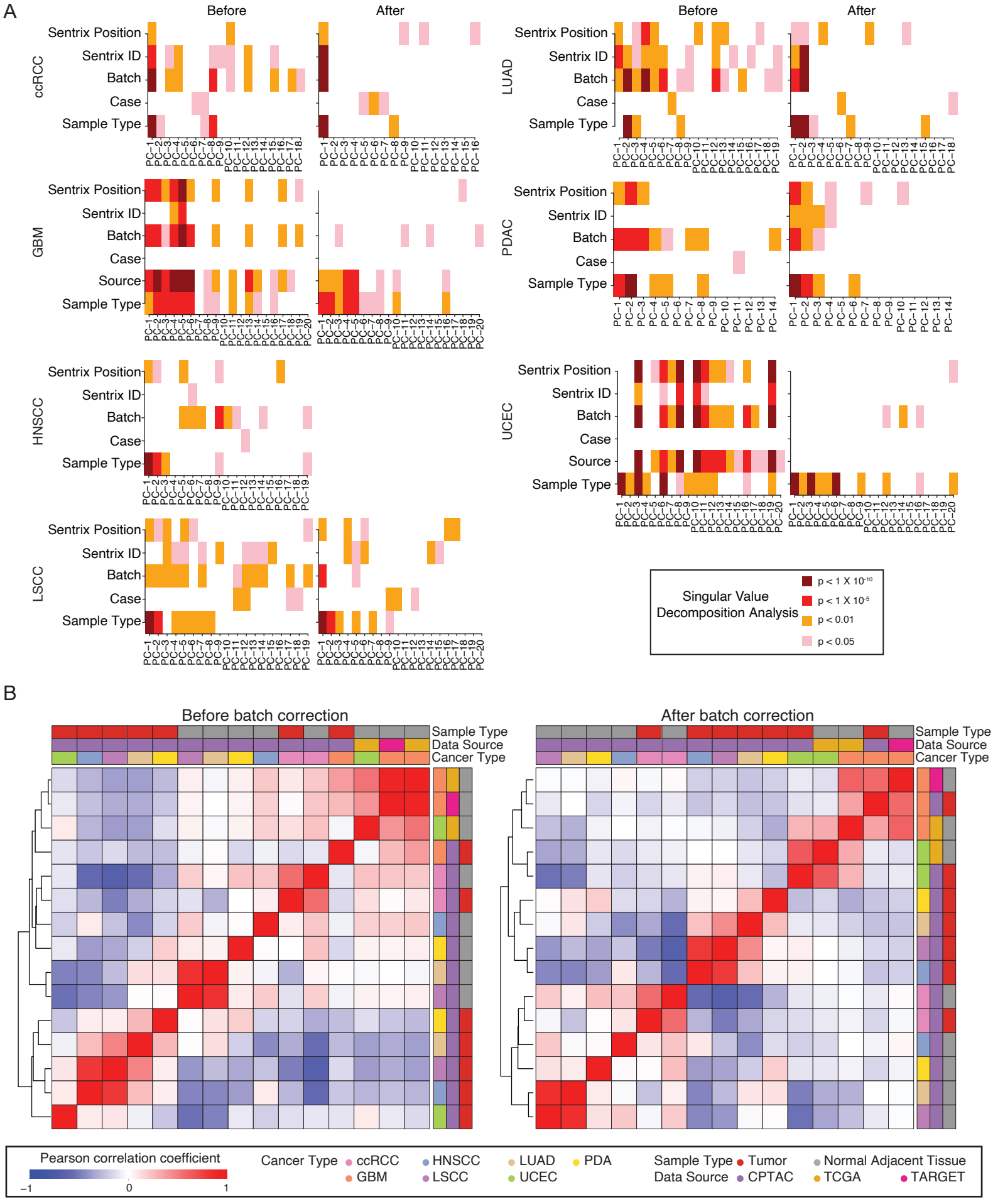
A
Before
After
ccRKCC
LUAD
GBM
PDAC
HNSCC
UCEC
LSCC
Sentrix Position
Sentrix ID
Batch
Case
Source
Sample Type
Singular Value Decomposition Analysis
p < 1 X 10-10
p < 1 X 10-5
p < 0.01
p < 0.05
B
Before batch correction
After batch correction
Sample Type
Data Source
Cancer Type
Pearson correlation coefficient
−1
0
1
Cancer Type ccRCC HNSCC LUAD PDA GBM LSCC UCEC
Sample Type Tumor Normal Adjacent Tissue
Data Source CPTAC TCGA TARGET

Figure S1

Figure S1. Batch correction of DNA methylome across seven cancer types. Related to Figure 1.

(A) Heatmap of P-values of association between the top 20 singular vectors (principal components) from the SVD of the raw beta value matrix and phenotypic as well as experimental factors. Phenotypic factors included sample type (solid tumor and adjacent normal tissue). Experimental factors included sentrix position, sentrix ID, data source (TCGA and CPTAC), and batch number. The left panels present SVD analysis before any adjustment, and the right panels present the result of SVD analysis after adjustment for batch (ccRCC, LUAD, LSCC, and PDA) or source (GBM and UCEC) effects. No adjustment has been made for HNSCC. Box colors indicate significance of association based on P-values from ANOVA F-statistic: P<10e-10 (dark red), 10e-10<P<10e-5 (red), 10e-5<P<0.01 (orange), 0.01<P<0.05 (pink), not significant (white).

(B) Heatmap showing the Pearson correlation coefficient between DNA methylation data obtained by hierarchical clustering using the Euclidean distance measure and the complete linkage method before (left) and after (right) ComBat batch correction.

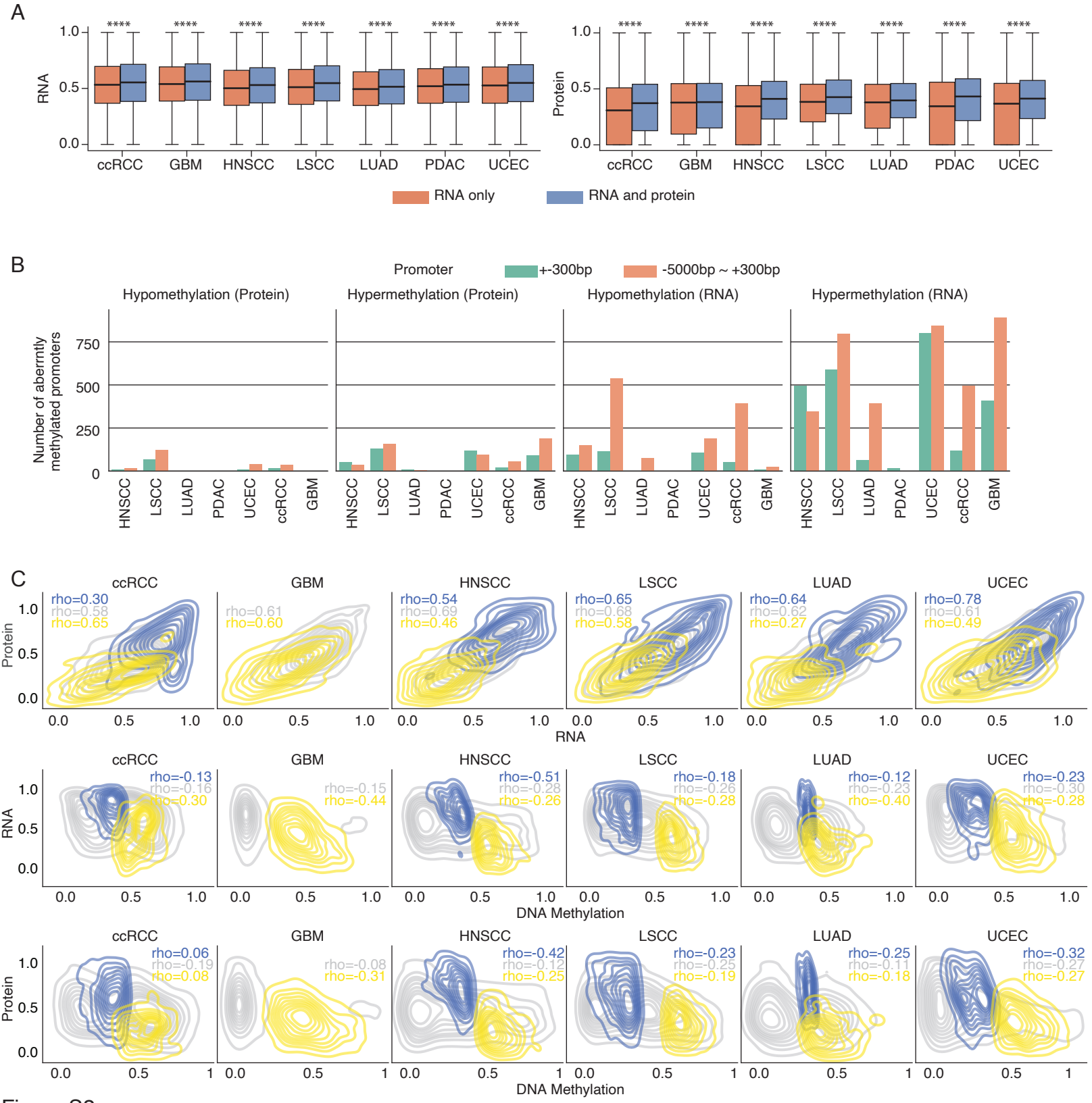
A
RNA
Protein
ccRCC
GBM
HNSCC
LSCC
LUAD
PDAC
UCEC
RNA only
RNA and protein
B
Promoter
+-300bp
-5000bp ~ +300bp
Hypomethylation (Protein)
Hypermethylation (Protein)
Hypomethylation (RNA)
Hypermethylation (RNA)
Number of aberrntly methylated promoters
C
RNA
DNA Methylation
Protein

Figure S2

Figure S2. Integrative analysis of DNA methylome, transcriptome, and proteome uncovers aberrant DNA methylation in tumors. Related to Figure 1 and 2.

(A) Box plots display the distribution of scaled RNA expression (left) and protein abundance values (right) for genes with promoter methylation correlated only with RNA expression (orange), compared to genes where promoter methylation correlates with both RNA and protein (blue). The median value is represented by the horizontal black line, while the first and third quartiles are indicated by the top and bottom lines, respectively. Statistical significance between groups was determined using a Wilcoxon rank sum test, with **** denoting $P < 2.2e\text{-}16$.

(B) Number of gene promoters with aberrant DNA methylation observed in RESET pipeline runs with different definitions of promoter regions.

(C) Correlation of RNA expression and protein abundance (first row), promoter methylation and RNA expression (second row), and promoter methylation and protein abundance (third row). The plot is based on the significant CpG sites in both RNA expression and protein abundance measurements, and colored based on methylation status: yellow, hypermethylation; blue, hypomethylation; gray: normal methylation. No significant correlation was found in the PDAC cohort.

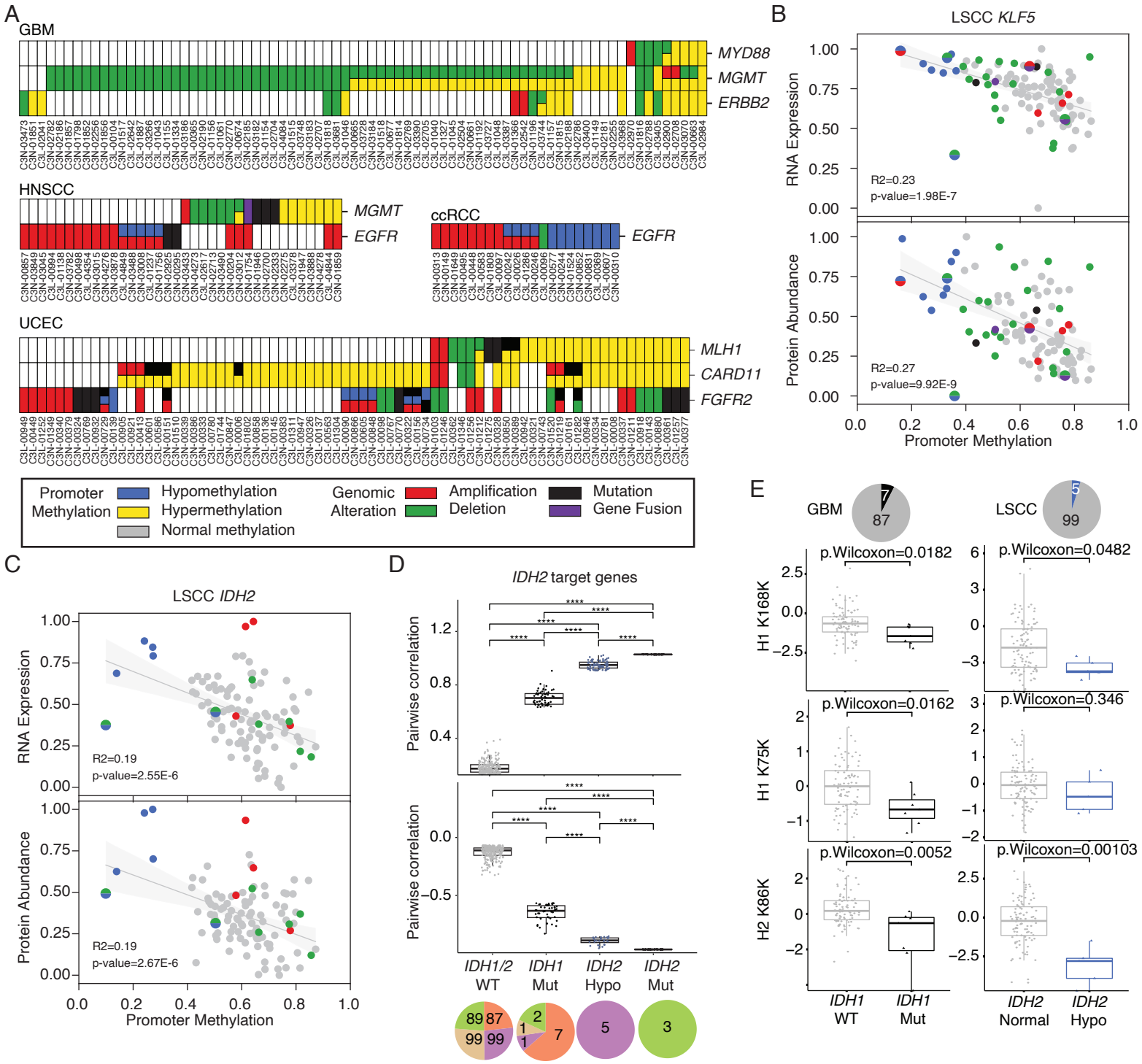
A
GBM
MYD88
MGMT
ERBB2
HNSCC
MGMT
EGFR
ccRCC
EGFR
UCEC
MLH1
CARD11
FGFR2
Promoter Methylation
Hypomethylation
Hypermethylation
Normal methylation
Genomic Alteration
Amplification
Deletion
Mutation
Gene Fusion
B
LSCC KLF5
RNA Expression
R2=0.23
p-value=1.98E-7
Protein Abundance
R2=0.27
p-value=9.92E-9
Promoter Methylation
C
LSCC IDH2
RNA Expression
R2=0.19
p-value=2.55E-6
Protein Abundance
R2=0.19
p-value=2.67E-6
Promoter Methylation
D
IDH2 target genes
Pairwise correlation
IDH1/2 WT
IDH1 Mut
IDH2 Hypo
IDH2 Mut
E
GBM
87
LSCC
99
H1 K168K
p.Wilcoxon=0.0182
p.Wilcoxon=0.0482
H1 K75K
p.Wilcoxon=0.0162
p.Wilcoxon=0.346
H2 K86K
p.Wilcoxon=0.0052
p.Wilcoxon=0.00103
IDH1 WT
IDH1 Mut
IDH2 Normal
IDH2 Hypo

Figure S3

Figure S3. Aberrant DNA methylation in the promoter of cancer-associated genes. Related to Figure 3.

(A) Mutual exclusivity and co-occurrence of genomic and epigenomic alterations in driver genes across cancer types. Each column represents a different tumor sample.

(B) Correlation of promoter methylation at significant CpG sites with gene expression (upper) and protein abundance (lower) of *KLF5* in LSCC. Tumors harboring multiple alterations at *KLF5* are highlighted by large dot size.

(C) Scatter plots demonstrating the correlations between promoter methylation and gene expression (upper) and protein abundance (lower) in *IDH2*. Each dot represents a LSCC tumor sample, and the color of the dots indicates the presence of genetic and/or epigenetic alterations of *IDH2*. Sample with multiple alterations are highlighted by large dot size.

(D) Correlation coefficients of histone acetylation levels and methylation levels at *IDH2* target genes were examined among *IDH1/IDH2* wild-type, *IDH1* mutant, *IDH2* hypomethylated samples, and *IDH2* mutant. The breakdown of each group was shown in the pie chart below. The box plot illustrates positive and negative correlations separately, with the y-axis representing the significant Pearson correlation coefficients (Pearson's $r > 0.2$, $P < 0.05$, or Pearson's $r < -0.2$, $P < 0.05$) of acetylation-methylation pairs. The x-axis represents the four groups categorized based on *IDH* alteration status. The boxes represent the interquartile range (IQR), with the horizontal line inside indicating the median correlation value. The whiskers extending from the boxes indicate the range of the data, excluding outliers, which are represented as individual data points beyond the whiskers. Statistically significant differences between groups were determined using FDR corrected P-values, with **** indicating $P < 2.2e\text{-}16$.

(E) Box plots comparing acetylation levels of H1 K168K, H1 K75K, and H2 K86K are shown for *IDH1* mutants and wild-type samples in GBM (left) and IDH2 normal methylated and hypomethylated samples in LSCC (right). The pie chart above provides the breakdown of each group. Boxes represent the interquartile range (IQR), with the median acetylation level indicated by the horizontal line inside. Whiskers extend from the boxes, representing the data range. Statistically significant differences between groups were assessed using the Wilcoxon rank-sum test.

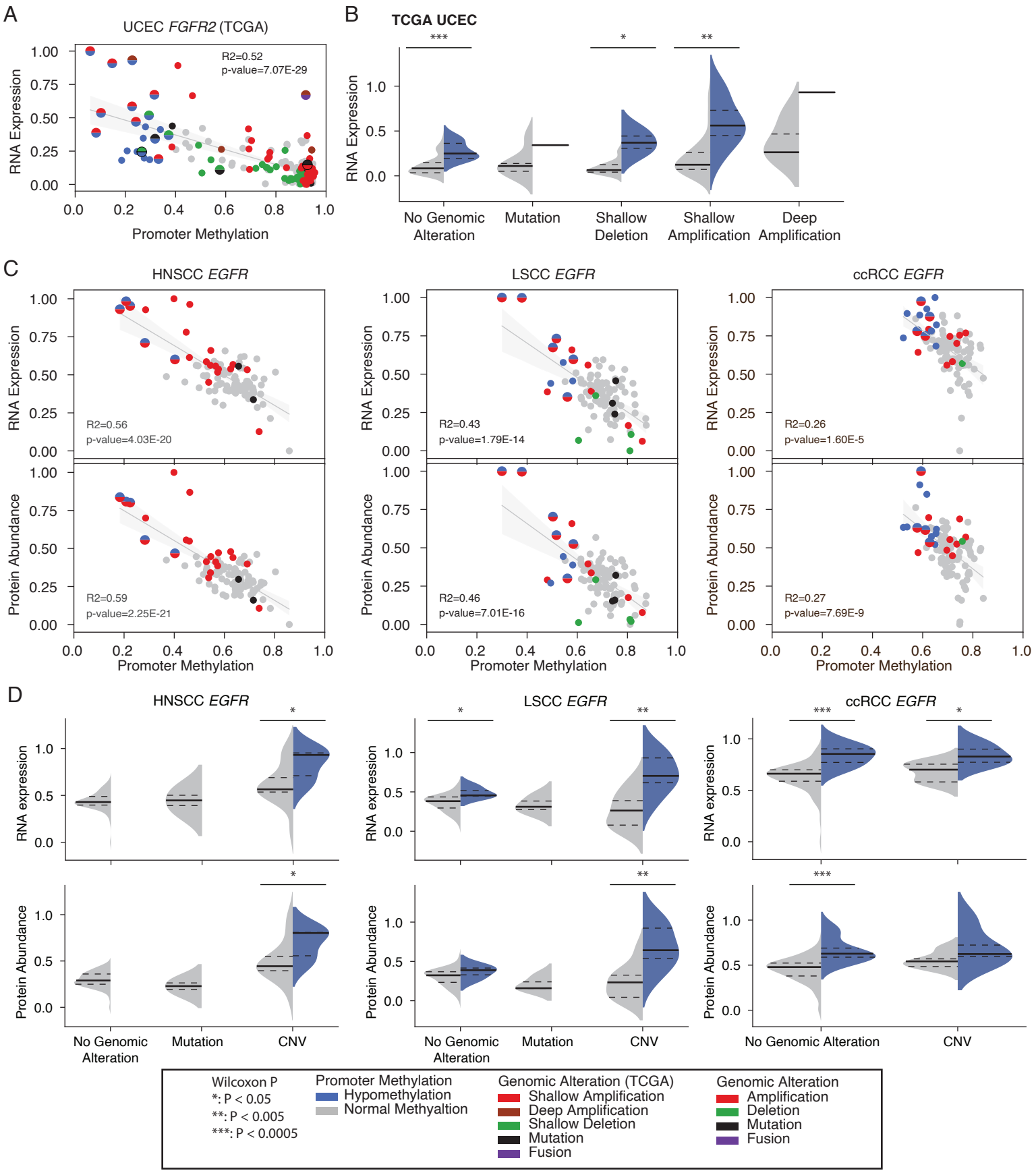
A
UCEC FGFR2 (TCGA)
R2=0.52
p-value=7.07E-29
RNA Expression
Promoter Methylation
B
TCGA UCEC
RNA Expression
No Genomic Alteration
Mutation
Shallow Deletion
Shallow Amplification
Deep Amplification
C
HNSCC EGFR
LSCC EGFR
ccRCC EGFR
RNA Expression
Protein Abundance
R2=0.56
p-value=4.03E-20
R2=0.59
p-value=2.25E-21
R2=0.43
p-value=1.79E-14
R2=0.46
p-value=7.01E-16
R2=0.26
p-value=1.60E-5
R2=0.27
p-value=7.69E-9
Promoter Methylation
D
HNSCC EGFR
LSCC EGFR
ccRCC EGFR
RNA expression
Protein Abundance
No Genomic Alteration
Mutation
CNV
Wilcoxon P
*: P < 0.05
**: P < 0.005
***: P < 0.0005
Promoter Methylation
Hypomethylation
Normal Methyaltion
Genomic Alteration (TCGA)
Shallow Amplification
Deep Amplification
Shallow Deletion
Mutation
Fusion
Genomic Alteration
Amplification
Deletion
Mutation
Fusion

Figure S4

Figure S4. Validation of *FGFR2* hypomethylation in TCGA UCEC cohort and hypomethylated *EGFR* is upregulated across cancer types. Related to Figure 4.

(A) Correlation of methylation at significant CpG site (cg10314760) with gene expression of *FGFR2* in TCGA UCEC cohort.

(B) RNA expression level in TCGA UCEC tumors stratified by *FGFR2* genomic alterations and *FGFR2* hypomethylation (blue) versus *FGFR2* normal methylation (gray). The median value is shown as a solid black line, and the first and third quartiles are represented by the top and bottom dashed lines, respectively. Statistical significance was assessed using a Wilcoxon rank-sum test (* $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$).

(C) Correlation of methylation at significant CpG sites with gene expression (upper) and protein abundance (lower) of *EGFR* in HNSCC (left), LSCC (middle), and ccRCC (right).

(D) RNA (upper) and protein (lower) levels comparison between *EGFR* hypomethylation (blue) and *EGFR* normal methylation (gray), stratified by *EGFR* genomic alterations in HNSCC (left), LSCC (middle), and ccRCC (right). Median values are represented by solid black lines, and first and third quartiles are indicated by dashed lines. Statistical significance was determined using a Wilcoxon rank-sum test (* $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$).

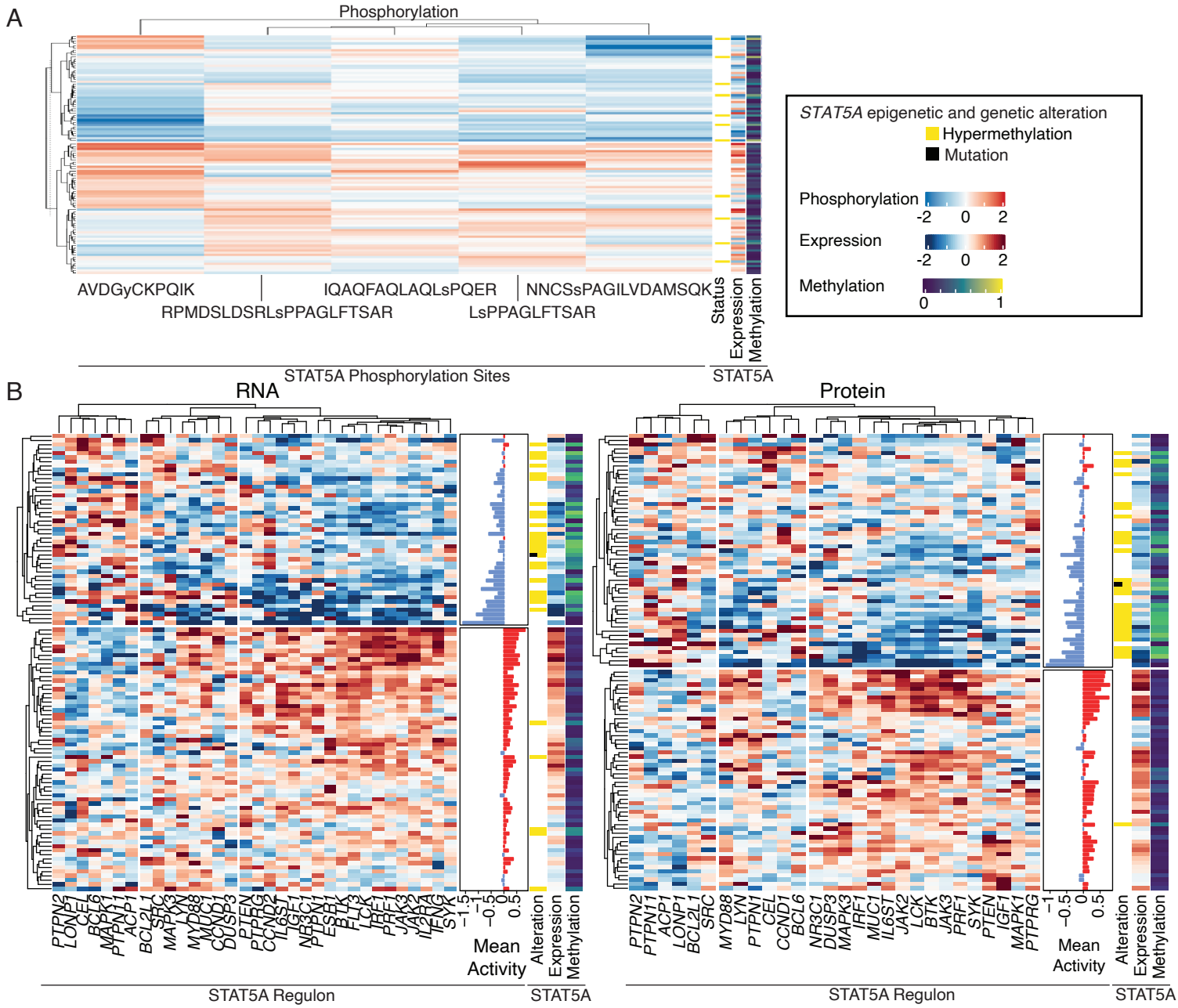
A
Phosphorylation
AVDGyCKPQIK
IQAQFAQLAQLsPQER
NNCSsPAGILVDAMSQK
RPMDSLDSRLsPPAGLFTSAR
LsPPAGLFTSAR
Status
Expression
Methylation
STAT5A Phosphorylation Sites
STAT5A
STAT5A epigenetic and genetic alteration
Hypermethylation
Mutation
Phosphorylation
-2 0 2
Expression
-2 0 2
Methylation
0 1
B
RNA
Protein
PTPN2 LONP1 CEL BCL6 MAPK1 PTPN11 ACP1 BCL2L1 SRC MAPK3 LYN MYD88 MUC1 CCND1 DUSP3 PTEN PTPRG CCND2 IL6ST IGF1 NR3C1 PTPN1 ESR1 BTK FLT3 LCK IRF1 PRF1 JAK3 JAK2 IL2RA IFNG SYK
−1.5 −1 −0.5 0 0.5
Mean Activity
Alteration
Expression
Methylation
STAT5A Regulon
STAT5A
PTPN2 PTPN11 ACP1 LONP1 BCL2L1 SRC MYD88 LYN PTPN1 CEL CCND1 BCL6 NR3C1 DUSP3 MAPK3 IRF1 MUC1 IL6ST JAK2 LCK BTK JAK3 PRF1 SYK PTEN IGF1 MAPK1 PTPRG
−1 −0.5 0 0.5
Mean Activity
Alteration
Expression
Methylation
STAT5A Regulon
STAT5A

Figure S5

Figure S5. Hypermethylated *STAT5A* and regulon activity. Related to Figure 5.

(A) Unsupervised clustering of *STAT5A* phosphorylation site using Pearson correlation of scaled RNA sequencing data. Annotations denote *STAT5A* expression and methylation levels. Color scale is proportional to phosphorylation activity (red: phosphorylated; blue: unphosphorylated).

(B) Unsupervised clustering of STAT5A regulon genes using Pearson correlation of scaled RNA sequencing data in LSCC. Annotations denote *STAT5A* expression and methylation levels. Mean activity indicates the overall sum of regulon activity. The color scale is proportional to expression activity (red: upregulation; blue: downregulation).

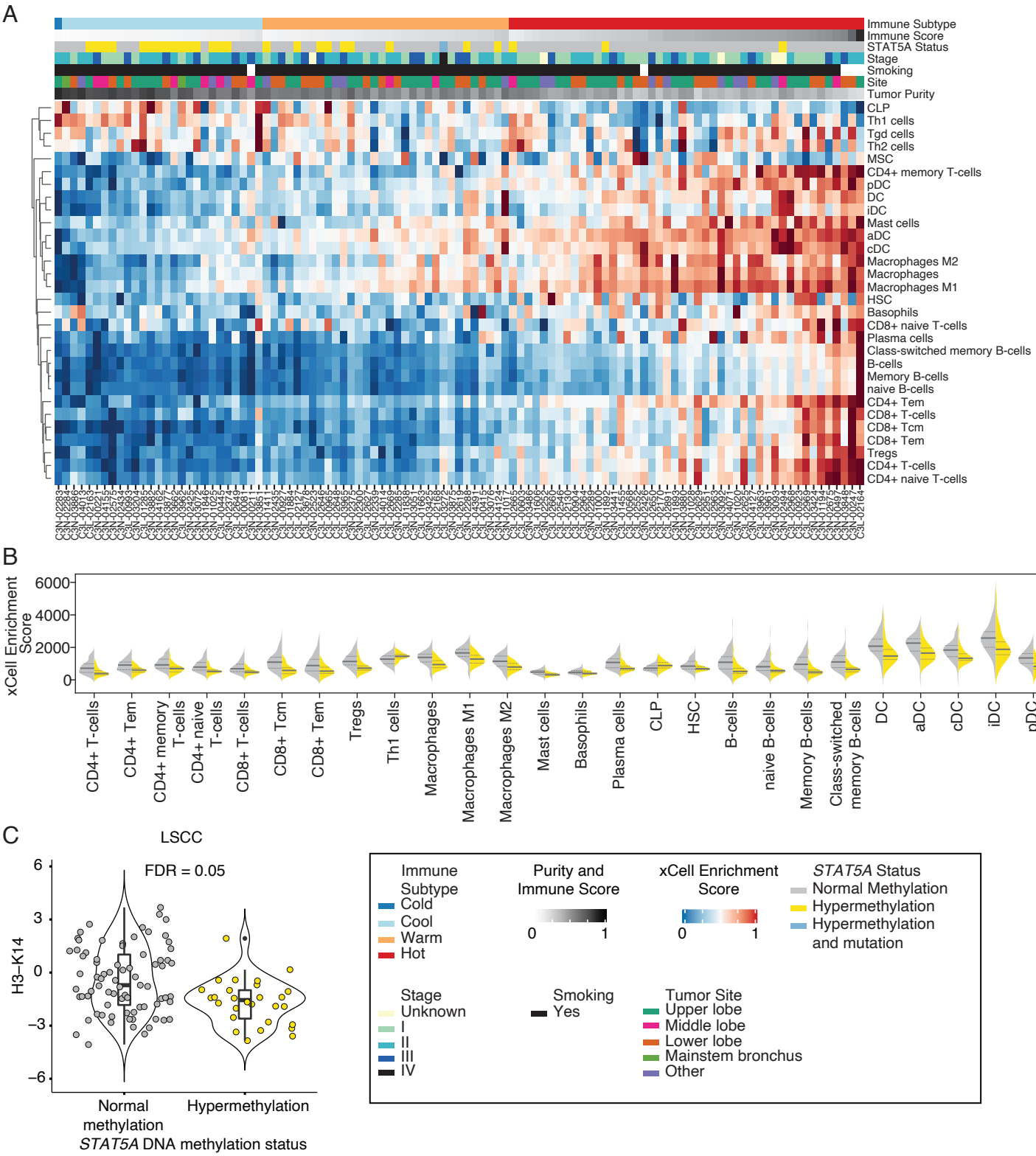
A
Immune Subtype
Immune Score
STAT5A Status
Stage
Smoking
Site
Tumor Purity
CLP
Th1 cells
Tgd cells
Th2 cells
MSC
CD4+ memory T-cells
pDC
DC
iDC
Mast cells
aDC
cDC
Macrophages M2
Macrophages
Macrophages M1
HSC
Basophils
CD8+ naive T-cells
Plasma cells
Class-switched memory B-cells
B-cells
Memory B-cells
naive B-cells
CD4+ Tem
CD8+ T-cells
CD8+ Tcm
CD8+ Tem
Tregs
CD4+ T-cells
CD4+ naive T-cells
B
xCell Enrichment Score
6000
4000
2000
0
CD4+ T-cells
CD4+ Tem
CD4+ memory T-cells
CD4+ naive T-cells
CD8+ T-cells
CD8+ Tcm
CD8+ Tem
Tregs
Th1 cells
Macrophages
Macrophages M1
Macrophages M2
Mast cells
Basophils
Plasma cells
CLP
HSC
B-cells
naive B-cells
Memory B-cells
Class-switched memory B-cells
DC
aDC
cDC
iDC
pDC
C
LSCC
FDR = 0.05
H3–K14
6
3
0
−3
−6
Normal methylation
Hypermethylation
STAT5A DNA methylation status
Immune Subtype
Cold
Cool
Warm
Hot
Purity and Immune Score
0
1
xCell Enrichment Score
0
1
STAT5A Status
Normal Methylation
Hypermethylation
Hypermethylation and mutation
Stage
Unknown
I
II
III
IV
Smoking
Yes
Tumor Site
Upper lobe
Middle lobe
Lower lobe
Mainstem bronchus
Other

Figure S6

Figure S6. Functional impact of *STAT5A* hypermethylation on immune cell depletion in LSCC. Related to Figure 6.

(A) Distinct immune subtypes of LSCC tumors identified by consensus clustering of LSCC tumors using xCell enrichment scores. The top panel shows the immune score, DNA methylation status of *STAT5A*, immune subtype, and tumor stage. The heatmap shows xCell enrichment scores deconvoluted from RNA-seq data.

(B) Violin plots comparing xCell enrichment scores of immune effectors of *STAT5A* hypermethylation (yellow) versus of *STAT5A* normal methylation (gray) in LSCC tumors. Median values are shown as solid black lines, and first and third quartiles are represented by dashed lines. Statistical significance was determined using a Wilcoxon rank-sum test.

(C) Violin plot comparing acetylation levels of H3 K14 between *STAT5A* hypermethylated samples (yellow) and normal methylated samples (gray) in LSCC. Boxes represent the IQR, with the median acetylation level shown as a horizontal line. Whiskers extend from the boxes to indicate the data range. Statistically significant differences between groups were determined using FDR-corrected P-values.

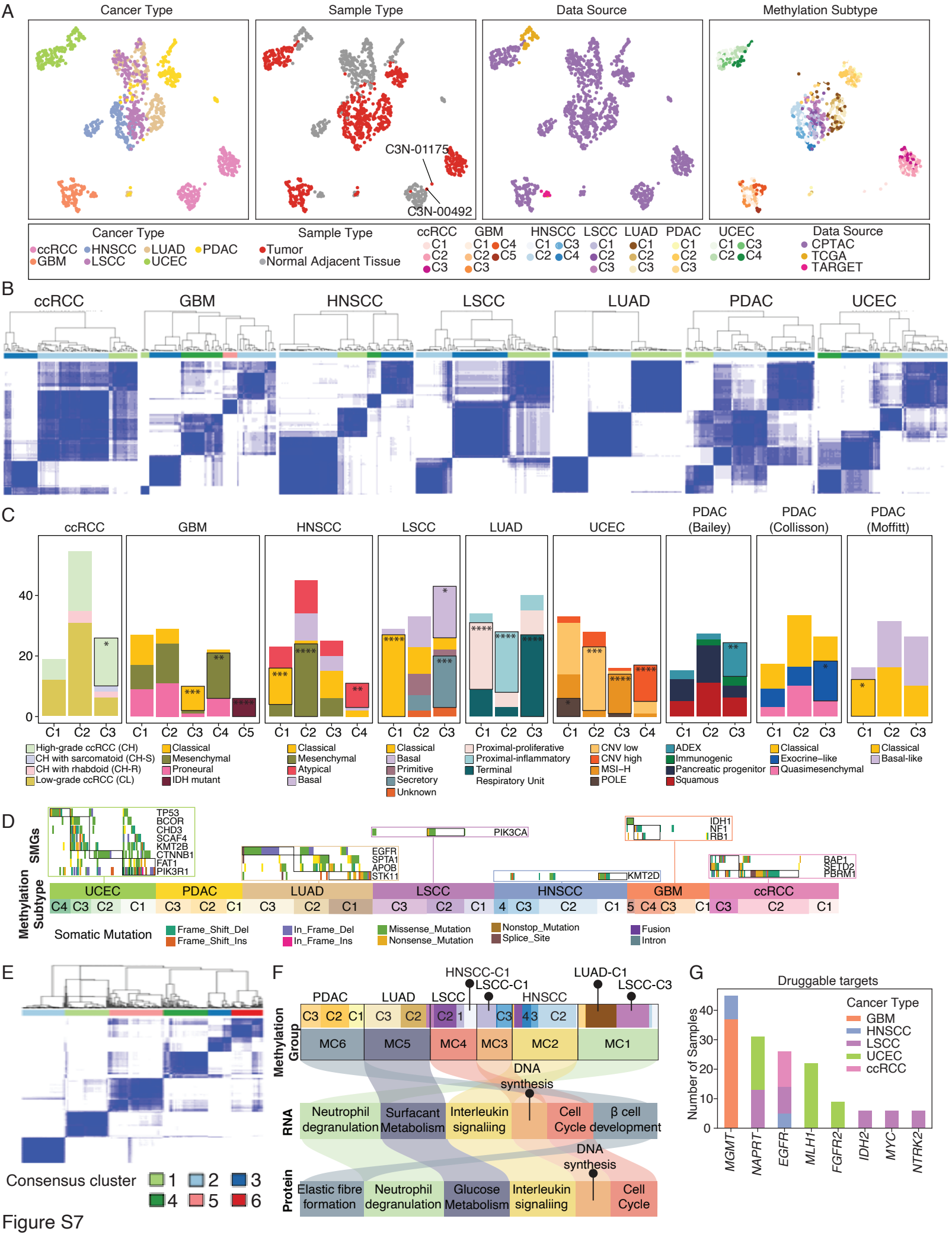
A
Cancer Type
Sample Type
Data Source
Methylation Subtype
C3N-01175
C3N-00492
Cancer Type
ccRCC HNSCC LUAD PDAC
GBM LSCC UCEC
Sample Type
Tumor
Normal Adjacent Tissue
ccRCC C1 C2 C3
GBM C1 C4 C2 C5 C3
HNSCC C1 C3 C2 C4
LSCC C1 C2 C3
LUAD C1 C2 C3
PDAC C1 C2 C3
UCEC C1 C3 C2 C4
Data Source
CPTAC
TCGA
TARGET
B
ccRCC
GBM
HNSCC
LSCC
LUAD
PDAC
UCEC
C
ccRCC
GBM
HNSCC
LSCC
LUAD
UCEC
PDAC (Bailey)
PDAC (Collisson)
PDAC (Moffitt)
High-grade ccRCC (CH)
CH with sarcomatoid (CH-S)
CH with rhabdoid (CH-R)
Low-grade ccRCC (CL)
Classical
Mesenchymal
Proneural
IDH mutant
Classical
Mesenchymal
Atypical
Basal
Classical
Basal
Primitive
Secretory
Unknown
Proximal-proliferative
Proximal-inflammatory
Terminal
Respiratory Unit
CNV low
CNV high
MSI-H
POLE
ADEX
Immunogenic
Pancreatic progenitor
Squamous
Classical
Exocrine-like
Quasimesenchymal
Classical
Basal-like
D
SMGs
TP53
BCOR
CHD3
SCAF4
KMT2B
CTNNB1
FAT1
PIK3R1
EGFR
SPTA1
APOB
STK11
PIK3CA
KMT2D
IDH1
NF1
RB1
BAP1
SETD2
PBRM1
Methylation Subtype
UCEC
PDAC
LUAD
LSCC
HNSCC
GBM
ccRCC
Somatic Mutation
Frame_Shift_Del
Frame_Shift_Ins
In_Frame_Del
In_Frame_Ins
Missense_Mutation
Nonsense_Mutation
Nonstop_Mutation
Splice_Site
Fusion
Intron
E
Consensus cluster
1
2
3
4
5
6
F
PDAC
LUAD
LSCC
HNSCC
HNSCC-C1
LSCC-C1
LUAD-C1
LSCC-C3
Methylation Group
MC6
MC5
MC4
MC3
MC2
MC1
DNA synthesis
RNA
Neutrophil degranulation
Surfacant Metabolism
Interleukin signaliing
Cell Cycle
β cell development
DNA synthesis
Protein
Elastic fibre formation
Neutrophil degranulation
Glucose Metabolism
Interleukin signaliing
Cell Cycle
G
Druggable targets
Number of Samples
Cancer Type
GBM
HNSCC
LSCC
UCEC
ccRCC
MGMT
NAPRT
EGFR
MLH1
FGFR2
IDH2
MYC
NTRK2

Figure S7

Figure S7. Cancer methylome and methylation-driven subtyping. Related to Figure 7.

(A) Projection of the cancer methylomes using UMAP. Each point is a sample and is colored based on the cancer type (first column), sample type (second column), data source (third column), or methylation subtypes (fourth column). The two chromophobe RCC were highlighted.

(B) Consensus matrix heatmap of the chosen optimal cluster number for all the per-cancer methylation-driven subtyping. The rows and columns represent DNA methylation values of tumor samples, and consensus matrix values range from 0 in white (meaning that tumors are never clustered together) to 1 in dark blue (meaning that tumors are always clustered together).

(C) Bar plot showing the enrichment of molecularly and clinically relevant subtypes in methylation-driven subtypes across cancer types. Fisher's exact test * < 0.05, ** < 0.005, *** < 0.0005, **** < 0.0001.

(D) Oncoplot showing the enrichment of mutations at driver genes across methylation-driven subtypes.

(E) Consensus matrix heatmap of the chosen optimal cluster number for the multi-cancer methylation groups. The rows and columns represent DNA methylation values of tumor samples, and consensus matrix values range from 0 in white (meaning that tumors are never clustered together) to 1 in dark blue (meaning that tumors are always clustered together).

(F) Alluvial plot showing the multi-cancer methylation groups (first row), their enriched RNA expression signature (second row) and enriched protein signature (third row). The curved lines across panels correspond to different methylation groups.

(G) Breakdown of potentially druggable genes, for which expression is altered by tumorigenic DNA methylation.