Integrative multi-omic cancer profiling reveals DNA methylation patterns associated with therapeutic vulnerability and cell-of-origin

Wen-Wei Liang  
*Washington University School of Medicine in St. Louis*

Rita Jui-Hsien Lu  
*Washington University School of Medicine in St. Louis*

Reyka G Jayasinghe  
*Washington University School of Medicine in St. Louis*

Steven M Foltz  
*Washington University School of Medicine in St. Louis*

Michael C Wendl  
*Washington University School of Medicine in St. Louis*

See next page for additional authors

Follow this and additional works at: [https://digitalcommons.wustl.edu/oa_4](https://digitalcommons.wustl.edu/oa_4)

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.

---

**Recommended Citation**

Liang, Wen-Wei; Lu, Rita Jui-Hsien; Jayasinghe, Reyka G; Foltz, Steven M; Wendl, Michael C; Song, Yizhe; Govindan, Akshay; Li, Xiang; Li, Yize; Sethuraman, Sunantha; Ding, Li; and et al., "Integrative multi-omic cancer profiling reveals DNA methylation patterns associated with therapeutic vulnerability and cell-of-origin." *Cancer Cell.* 41, 9. 1567 - 1585.e7. (2023).  
[https://digitalcommons.wustl.edu/oa_4/3098](https://digitalcommons.wustl.edu/oa_4/3098)

---

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Integrative multi-omic cancer profiling reveals DNA methylation patterns associated with therapeutic vulnerability and cell-of-origin

Graphical abstract

Highlights

- Pan-cancer epigenetic aberrations and their transcriptional and translational changes
- *FGFR2* and *EGFR* hypomethylation are bona fide driver DNA methylation events
- STAT5A methylation is a potential switch for immunosuppression in squamous tumors
- Methylation subtypes illuminate cell origin, tumor heterogeneity, tumor phenotype, and links to therapeutic potential

Authors

Wen-Wei Liang, Rita Ju-i-Hsien Lu, Reyka G. Jayasinghe, ..., Ana I. Robles, Li Ding, Clinical Proteomic Tumor Analysis Consortium

Correspondence

lding@wustl.edu

In brief

Liang et al. catalog pan-cancer DNA methylation with concordant transcriptional and translational changes, revealing lineage-specific epigenetic driver *FGFR2* hypomethylation in uterine corpus endometrial carcinoma, and STAT5 hypermethylation as an immunosuppression switch in squamous tumors. They also identify methylation-driven subtypes associated with cell-of-origin, tumor heterogeneity, tumor phenotype, and links to therapeutic potential.
Integrative multi-omic cancer profiling reveals DNA methylation patterns associated with therapeutic vulnerability and cell-of-origin

Wen-Wei Liang,1,2,24 Rita Jui-Hsien Lu,1,2 Reyka G. Jayasinghe,1,2 Steven M. Foltz,1,2 Eduard Porta-Pardo,3,4 Yifat Geffen,5,6 Michael C. Wendl,2,7,8 Rossana Lazcano,9 Iga Kolodziejczak,10,11 Yizhe Song,1,2 Akshay Govindan,1,2 Elizabeth G. Demicco,12 Xiang Li,1,2 Yize Li,1,2 Sunantha Sethuraman,1,2 Samuel H. Payne,13 David Fenyö,14,15 Henry Rodriguez,16 Maciej Wiznerowicz,17,18,19 Hui Shen,20 Karin D. Rodland,21,22 Alexander J. Lazar,9 Ana I. Robles,10 Li Ding,1,2,23,25,* and Clinical Proteomic Tumor Analysis Consortium

1Department of Medicine, Washington University in St. Louis, St. Louis, MO 63110, USA
2McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108, USA
3Josep Carreras Leukaemia Research Institute (IJC), 08916 Badalona, Spain
4Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
5Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA
6Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA 02115, USA
7Department of Genetics, Washington University in St. Louis, St. Louis, MO 63130, USA
8Department of Mathematics, Washington University in St. Louis, St. Louis, MO 63130, USA
9Departments of Pathology & Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
10International Institute for Molecular Oncology, 60-203 Poznań, Poland
11Postgraduate School of Molecular Medicine, Medical University of Warsaw, 02-091 Warsaw, Poland
12Department of Pathology and Laboratory Medicine, Mount Sinai Hospital and Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
13Department of Biology, Brigham Young University, Provo, UT 84602, USA
14Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY 10016, USA
15Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY 10016, USA
16Office of Cancer Clinical Proteomics Research, National Cancer Institute, Rockville, MD 20850, USA
17International Institute for Molecular Oncology, 60-203 Poznań, Poland
18Heliodor Święcicki Clinical Hospital in Poznań, ul. Przybyszewskiego 49, 60-355 Poznań, Poland
19Poznań University of Medical Sciences, 61-701 Poznań, Poland
20Van Andel Research Institute, Grand Rapids, MI 49503, USA
21Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA
22Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR 97221, USA
23Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63130, USA
24Present address: New York Genome Center, New York, NY 10013, USA
25Lead contact
*Correspondence: lding@wustl.edu
https://doi.org/10.1016/j.ccell.2023.07.013

SUMMARY

DNA methylation plays a critical role in establishing and maintaining cellular identity. However, it is frequently dysregulated during tumor development and is closely intertwined with other genetic alterations. Here, we leveraged multi-omic profiling of 687 tumors and matched non-involved adjacent tissues from the kidney, brain, pancreas, lung, head and neck, and endometrium to identify aberrant methylation associated with RNA and protein abundance changes and build a Pan-Cancer catalog. We uncovered lineage-specific epigenetic drivers including hypomethylated FGFR2 in endometrial cancer. We showed that hypermethylated STAT5A is associated with pervasive regulon downregulation and immune cell depletion, suggesting that epigenetic regulation of STAT5A expression constitutes a molecular switch for immunosuppression in squamous tumors. We further demonstrated that methylation subtype-enrichment information can explain cell-of-origin, intra-tumor heterogeneity, and tumor phenotypes. Overall, we identified cis-acting DNA methylation events that drive transcriptional and translational changes, shedding light on the tumor’s epigenetic landscape and the role of its cell-of-origin.
INTRODUCTION

Cytosine methylation is an epigenetic modification that confers stability and flexibility in the spatiotemporal gene regulation of many biological processes, including establishment and maintenance of cell identity. Aberrant DNA methylation is a hallmark of human cancer development and progression. Aberrant DNA methylation has been observed in the global hypomethylation of repetitive sequences and the gene-specific hypermethylation of numerous CpG islands (CGI). Such changes within promoter regions can silence tumor suppressor genes or deregulate oncogenes. Moreover, the widespread changes in DNA methylation patterns usually arise in the early stages of tumorigenesis, suggesting a driving role of aberrant DNA methylation. Given the reversible and dynamic nature of DNA methylation, treating cells with DNA demethylating agents might reprogram cell identity. Aberrant DNA methylation is a hallmark of many biological processes, including establishment and maintenance and flexibility in the spatiotemporal gene regulation of numerous cancer types from the Clinical Proteomic Tumor Analysis Consortium (CPTAC), and systematically examined the impact of cis-acting aberrant DNA methylation events on information flow, features, and functional consequences. Our analysis identified common and tissue-specific epigenetic events, including significant alterations in cancer genes affecting hallmark pathways. Distinct cancer subtypes were characterized by unique methylation patterns, validated by RNA and protein signatures reflecting each subtype’s molecular characteristics. Moreover, we identified putative druggable genes tightly regulated by DNA methylation, offering potential targets for tailored therapeutic interventions. This comprehensive catalog advances our understanding of DNA methylation-mediated tumorigenesis and offers insights toward the development of epigenetic therapies.

RESULTS

Pan-Cancer landscape of DNA methylation and associated functional changes

To construct a landscape of cancer methylomes and associated functional changes, we collected 687 human tumors with available DNA methylation profiles (Infinium EPIC array), gene expression (RNA-seq), and protein abundance (mass spectrometry) across seven cancer types from CPTAC. This cohort was composed of 107 clear cell renal cell carcinomas (ccRCCs), 94 glioblastomas (GBMs), 104 head and neck squamous cell carcinomas (HNSCCs), 107 lung squamous cell carcinomas (LSCCs), 102 lung adenocarcinomas (LUADs), 79 pancreatic ductal ade- nocarcinomas (PDACs), and 94 uterine corpus endometrial carcinomas (UCECs). We also collected DNA methylation data from matched normal adjacent tissues (NATS) including the kidney, head and neck, pancreas, and lung from CPTAC. For GBM and UEC datasets that lacked DNA methylation data from NATs, we acquired the corresponding DNA methylene from The Cancer Genome Atlas (TCGA) and the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) datasets in the form of Infinium HumanMethylation BeadChip (HM450) (Table S1, STAR Methods). All DNA methylation profiles were processed from raw array data, with standardized processing, quality control, and batch correction procedures (STAR methods). This expanded dataset allowed us to estimate methylation among reference tissues and to interrogate aberrant DNA methylation in a tissue-specific context (Figure S1).

First, to estimate how methylene, transcriptome, and proteome are correlated, we quantified the pairwise correlation between datasets for the 12,943 genes where all measurements are available using linear models, taking into account cancer type and tumor purity as covariates (STAR Methods, Figure 1A; Table S1). Consistent with expectations, RNA expression positively correlated with protein abundance, and promoter DNA methylation negatively correlated with RNA expression or protein abundance, indicating that promoter hypermethylation leads to gene silencing, while hypomethylation enhances gene expression (Figure 1A). Focusing on the 10,844 genes showing significant mRNA-protein correlation, we categorized them into groups based on the correlation between promoter methylation and RNA/protein expression. The correlation coefficients varied in strength and direction (Figure 1B), with 64.2% of promoter methylation showing no correlation, 18.4% correlated with RNA expression, and 3.5% correlated with protein abundance (Figure 1C). Only 13.9% of the promoter methylation exhibited correlation with both RNA expression and protein abundance (Figure 1C). For example, MxRas and Mnda were correlated with RNA expression and protein abundance, respectively, while Card11 and Mgmt showed correlations with both RNA expression and protein abundance (Figure 1D). When parallel correlation tests were conducted within single cancer types without considering tumor purity, the fractions of promoters correlated with both RNA and protein abundance were even lower. This suggests that pan-cancer analysis with a larger sample size enhances discovery power (Table S1).

The limited impact of methylation on the proteome may be attributable to biological factors (e.g., translational regulation, tissue-specific expression, protein degradation, and post-translational modifications) or to technical factors resulting from the low detection sensitivity of low-abundance transcripts or proteins. Indeed, we observed a significant decrease in RNA expression and protein abundance in genes where promoter methylation correlated only with RNA expression, compared to those where methylation correlated with both RNA expression and protein abundance (Figure S2A, Wilcoxon p < 2.2e-16). Furthermore, among the 353 genes with promoter methylation correlated solely with protein abundance, we identified 31 genes, including the previously reported Mnda, with promoter hypermethylation and gene upregulation, indicative of a biological effect rather than a technical artifact (Figures 1D and 1E; Table S1). These genes exhibited enrichment in neutrophil...
degranulation and glycogen metabolism pathways (Figure 1F; Table S1). The significant correlation between promoter methylation and protein abundance suggests the presence of an additional regulatory layer likely detectable through proteomic data, emphasizing the intricate and context-specific nature of DNA methylation’s impact on gene regulation.

To characterize the prevalence of aberrant DNA methylation across seven cancer types, we next detected recurrent and deregulated DNA methylation in tumors compared to NATs. We then determined the association of DNA methylation with changes in both mRNA expression and protein abundance using the published pipeline RESET.9 To minimize batch effects or confounding bias across cancer types, each cohort was analyzed individually. Our analysis aimed to profile aberrant DNA methylation for regions actively contributing to gene regulation.15 Therefore, probe sets were limited to those within...
promoter regions, or 300 bp upstream and downstream of the transcription start sites for the RESET pipeline (Figure S2B; Table S2). Overall, we detected 5,570 hypermethylated CpG sites associated with mRNA downregulation of 2,549 genes, 889 hypermethylated CpG sites associated with protein downregulation of 425 genes, 537 hypomethylated CpG sites associated with mRNA upregulation of 442 genes, and 166 hypomethylated CpG sites with protein upregulation of 124 genes (FDR <10%, Figure 2A). No aberrant DNA methylation has been identified in the PDAC cohort, which could be explained by low tumor purity.16 Our results confirmed that several well-known tumor-associated genes are regulated epigenetically, including MLH1 in UCEC,18 and MGMT in GBM and HNSCC.19 However, most identified genes have not been previously implicated in DNA methylation-mediated regulation (Table S2).

Next, for individual aberrant DNA methylation events, we examined the concordance between mRNA and protein changes. Out of 964 total hypermethylation and 217 total hypomethylation events, we identified 365 hypermethylation and 74 hypomethylation events associated with both RNA expression and protein abundance changes (Figures 2B and S2C), with a 70.5% validation rate for the genes having available DNA methylation (HM450) and RNA-seq data in TCGA cohorts (FDR <10%, Table S2). Among them, we observed 98.9% of the aberrant methylation had the same direction of effect across the seven cancer types. About 78.8% of the aberrant methylation have been recovered from the same analysis with tumor purity-adjusted methylation values (FDR <10%, Table S2), suggesting that our reported events are not being driven by underlying differences in tumor purity. Having constructed a comprehensive map of the cis-acting cancer methylome, we were able to delineate functional impacts of deregulated DNA methylation for a set of genes directly related to tumorigenesis (Table S2). Figure 2C illustrates the median methylation, RNA, and protein differences for cancer-associated genes, presented as the difference between NATs and tumor samples. In line with previous studies, we observed 41 of the aberrant methylation events are cancer type-specific.17,20,21 Tumor necrosis receptor was expressed in cancer cells and immune cells, indicating the possibility of aberrant DNA methylation affecting different populations of cells in the tumor microenvironment. In total, 190 out of 436 events (43.6%) were annotated for their specific expression (Table S2). While our integrative multi-omic analysis based on bulk tissue provides a reliable estimate across a large number of tumors, the complementary scRNA-seq annotations deepen these analyses by evaluating gene expression patterns of the reported events at the cellular level for future study.

cis-Acting aberrant DNA methylation as a possible driver event
DNA methylation-mediated modulation might be an important mechanism affecting the regulation of oncogenes and tumor suppressor genes. To test this hypothesis, we characterized cis-acting aberrant DNA methylation on 299 driver genes.29 Since DNA methylation can affect the binding of transcription factors (TFs),30 we examined the number and enrichment of TF binding sites (transcription factor bind sites [TFBS]) in loci associated with expression changes. All loci were associated with at least one TFBS. Both hypomethylated and hypermethylated CGI sites were characterized by a similar number of TFBS (mean 8.14 v.s. 7.49 TFBS) (Figure 3A). PREP1, a master regulator that functions as a tumor suppressor in maintaining genome stability,31 was the most enriched TFBS in hypermethylated loci. This result supports the notion that TFs might serve as both the readers and effectors of aberrant DNA methylation in tumors,30 leading to altered expression as revealed by transcriptomic and proteomic data.

Next, we explored the relationship among genetic alterations, DNA methylation, or histone acetylation. For the driver genes with cis-acting DNA methylations, most methylation events were mutually exclusive with genomic alterations, as exemplified in LSCC (Figures 3B and S3A). For example, the correlation between KLF3 expression and promoter DNA methylation suggests that methylation is the main factor regulating its expression (Figure S3B). Investigating the interplay between DNA methylation and histone acetylation revealed that tumors exhibiting a high frequency of hypermethylation are significantly linked to decreased levels of H3 acetylation (Figure 3C, Wilcoxon p < 0.05), indicating an overall repression of genes. Furthermore, a distinct relationship was observed among immune subtypes, hypermethylation frequency, and histone acetylation profiles, where the immune cool group showed an association with tumors displaying high hypermethylation frequency (Figure 3D). The pattern of mutual exclusivity and the interaction between DNA methylation and histone acetylation suggests that DNA methylation contributes to positive selection and histone changes, respectively, potentially playing a driving role in tumorigenesis.

Comparison to NATs showed DNA hypomethylation is the main perturbation occurring in the IDH2 gene, found in 6 of 107 LSCC tumors (Figures 3E and S3C). Since overexpression of IDH2 contributes to altered energy metabolism,32,33 we examined the metabolic activity of those cancers in samples with or without IDH2 hypomethylation. The results showed preferential upregulation of genes involved in cancer metabolism, including KDMs, ALKBHs, TETs, and MTOR, with the specific genes and extent of their expression changes varying slightly across different pathways (Figure 3F). To detect the downstream metabolic remodeling effect of IDH2 hypomethylation, we investigated the relationship between IDHs alterations and histone
acetylation profiles. A total of 393 patients with available acetylation and methylation data were categorized into four groups: IDH1/IDH2 wild-type (n = 374), IDH1 mutants (n = 11), IDH2 hypomethylated samples (n = 5), and IDH2 mutants (n = 3). We found that IDH2 hypomethylation, similar to IDH1 and IDH2 mutants, can significantly impact the correlations between histone
acetylation and DNA methylation levels at 2-hydroxyglutarate (α-KG) target genes (Figure 3G; Table S2) and IDH2 target genes (Figure S3D; Table S2). For example, significantly lower levels of H1 K168K, H1 K75K, and H2 K86K were observed in IDH1/2 mutant and IDH2 hypomethylated samples compared to IDH1/2 wild types and IDH2 normal methylated samples, respectively (Figure S3E). Despite the distinct nature of IDH2 hypomethylation compared to IDH1 and IDH2 mutations, the similarity in correlation patterns suggests a potential metabolic convergence of these alterations across cancer types. The combined methylation and acetylation profiling revealed driver mutation-independent IDH2 activation.

Hypomethylated RTKs are newly identified driver events
We found that several receptor tyrosine kinases (RTKs), including FGFR2 and EGFR, are frequently hypomethylated in UCEC and across cancer types, respectively (Figure 2C). To dissect the contribution of hypomethylated RTKs to oncogene activation, we examined the relationship between promoter methylation and genetic alteration. Specifically, we identified mutations, fusions, and copy number variations (CNVs), and their effects on RTK RNA and protein levels.

About 63.2% (12 of 19) of FGFR2 missense and indel mutations were activating mutations that enabled high-grade inflammation and cell proliferation without hypomethylation34-39 (Figure 4A). However, we identified 9 UCEC tumors carrying cis-acting hypomethylated FGFR2, 8 of which had co-occurring genomic alterations. Furthermore, unsupervised clustering of DNA methylation data across 94 CPTAC UCEC tumors and 43 TCGA normal samples from the same organ type revealed that FGFR2 hypomethylated cases formed a distinct cluster with lower DNA methylation than normal tissues (Figure 4B). Specifically, one CGL (cg10314760) within the FGFR2 promoter displayed a strong correlation between promoter hypomethylation and active gene expression both at the RNA and protein levels (Figure 4C). Our results suggest that promoter methylation is a major factor modulating expression of FGFR2, and that FGFR2 hypomethylation represents another mechanism of RTK activation potentially commensurate with activating mutations.

To distinguish the oncogenic effects of FGFR2 hypomethylation from co-occurring aberrations, we stratified UCEC tumors by the type of FGFR2 genomic alteration and examined FGFR2 expression accordingly. We found that the median of FGFR2 RNA expression or protein abundance are consistently higher in tumors with hypomethylated FGFR2 than in tumors with normal methylated FGFR2. Higher expression of FGFR2 is significant in the group with CNVs at the cognate locus (Wilcoxon p < 0.005) (Figure 4D). Validation in the TCGA UCEC cohort confirmed our findings. FGFR2 exhibited hypomethylation at CGL site (cg10314760) in 21 of 174 tumor samples, strongly correlating with active gene expression (Figure S4A). Consistent correlation between FGFR2 hypomethylation and upregulation was observed in samples with wild-type FGFR2, shallow deletion of FGFR2, and shallow amplification of FGFR2 (Figure S4B). This result suggests that, whereas activating mutations, amplifications, and promoter hypomethylation enable FGFR2 upregulation to differing extents, co-occurring FGFR2 hypomethylated sites result in even more profound expression changes than amplification alone. These results emphasize the important role of promoter hypomethylation in contributing to oncogenic gain-of-function. Similarly, hypomethylated EGFR was associated with EGFR upregulation in HNSCC, LSCC, and ccRCC (Figures S4C and S4D). Our results are consistent with the Pan-Cancer analysis suggesting that tumors harbor multiple aberrations within individual oncogenes such as FGFR2 or EGFR, likely conferring enhanced oncogenicity in combination.40 Overall, although recurrent gain-of-function genomic alterations in RTKs have long been known to promote a variety of cancers,41 our results reveal that RTK hypomethylation is also a bona fide epigenetic driver across several cancer types.

Hypermethylation of STAT5A is associated with pervasive changes in STAT5A regulon activity
Altered expression of TF can disrupt the activity of its regulon, a group of genes that are regulated by a common regulatory element. To estimate the impact of aberrant DNA methylation on regulon activity, we first identified 14 cis-acting hypermethylation events at TFs, and then tested their association with corresponding regulons, such as receptors, activator, repressor, and target genes involved in the same pathway (Fisher’s exact test, FDR p < 0.1, Table S3). We identified significant associations between hypermethylated TFs and low regulon activity at both RNA and DNA levels. For example, significantly lower levels of IDH1 mutant, IDH2 hypomethylated samples, and IDH2 mutant. The breakdown of each group was shown in the pie chart below. Boxes represent the IQR, with the median correlation value indicated by the horizontal line. Whiskers extend from the boxes to show the data range. Statistically significant differences between groups were determined using FDR-corrected p-values, with **** indicating p < 2.2e-16.

See also Figure S3 and Table S2.
Figure 4. Collaborative effects of FGFR2 mutations and hypomethylation on FGFR2 upregulation
(A) Lollipop showing missense mutations of FGFR2 in UCEC samples. The amino acids and types of mutations are labeled. Positions that are recurrently mutated are highlighted with the number of occurrences. The FGFR2 functional domains are colored.
(B) Unsupervised clustering of UCEC tumors (upper) and normal adjacent tissues (lower) based on DNA methylation of the FGFR2 promoter.
(C) Correlation of methylation with gene expression (upper) and protein abundance (lower). Samples are colored based on genetic and/or epigenetic alterations of FGFR2. Tumors harboring FGFR2 hypomethylation are highlighted by large dot size.
(D) RNA expression (upper) and protein abundance (lower) levels stratified by FGFR2 genomic alterations and FGFR2 hypomethylation (blue) versus FGFR2 normal methylation (gray), with ** denoting Wilcoxon p < 0.005. Median values are shown as solid black lines, and first and third quartiles are represented by dashed lines.
See also Figure S4.
expression and protein abundance levels. Since STAT5A regulon comprises the largest number of interacting genes, we therefore focus on STAT5A for the downstream analysis.

STAT5A controls cell identity, cytotoxicity, and cell survival; dysregulation of these processes can contribute to tumorigenesis.63 Unsupervised clustering of STAT5A-interacting proteins, using expression data from HNSCC tumors divided samples into two groups: those with high regulon activity and those with low regulon activity (Figure 5A). Notably, samples with hypermethylated STAT5A were significantly enriched in the regulon-low group (Fisher’s exact test, p = 7.8E-05). The same pattern was observed in protein abundance (Figure 5B, Fisher’s exact test, p = 0.025). Furthermore, exome sequencing of these tumors did not identify any distinct, recurrent coding sequence mutations in STAT5A-interacting genes (Table S3), suggesting that other genetic drivers were not involved. Additionally, STAT5A phosphorylation was not significantly associated with either STAT5A methylation status or STAT5A regulon activity (Figure S5A). Altogether, the enrichments we observed indicate that aberrant methylation of STAT5A leads to pervasive regulon changes in HNSCC. Similar to HNSCC, hypermethylated STAT5A was associated with low regulon activity in LSCC (Figure S5B).

Since samples with hypermethylated STAT5A were significantly associated with lower regulon activity at both RNA and protein levels (Figure 5C), we hypothesized that STAT5A-interacting components (i.e., receptors, kinases, repressors, co-activators, and target genes) would be downregulated in samples with hypermethylated STAT5A. Among the target genes, we observed significant downregulation of IRF1, PRF1, IFNG, IL2RA, and IL8ST (Wilcoxon p < 0.05) at either/both RNA and protein levels (Figure 5D). These observations link hypermethylated STAT5A to the regulation of cytokine production, cytotoxicity, cell proliferation, and interferon signaling.

Hypermethylated STAT5A is associated with immune cell depletion in squamous tumors

Recent study has shown that STAT5A-mediated interferon signaling regulates the expression of CD274 (encodes PD-L1) and PDCD1LG2 (encodes PD-L2), reflecting the clinical significance of STAT5A signaling in immunotherapy.43 STAT5A target genes are directly implicated in immune response (e.g., IL2RA, IRF1, and IFNG) and their low expression implies alteration of normal immune function and homeostasis.44 We therefore focused on characterizing the immune component of HNSCC and LSCC tumors to understand how hypermethylated STAT5A affects the tumor microenvironment.

To explore the microenvironment of HNSCC and LSCC tumors, we stratified transcriptome data from cell mixtures into multiple immune cell types using xCell.45 Consensus clustering of 64 different immune-related cell types identified four major immune clusters, including immune-cold, immune-cool, immune-warm, and immune-hot subtypes, which harbored general characteristics of immunosuppressive to inflammatory microenvironment (Table S3). STAT5A hypermethylated samples were significantly enriched in the immune-cool group both in HNSCC and LSCC (Figures 6A and S6A, Fisher’s exact test, p < 0.00001 and p = 0.0002, respectively). STAT5A hypermethylated samples displayed significantly decreased expression of genes associated with immune effectors and dendritic cells (Figures 6B and S6B, Wilcoxon p < 0.05). Interestingly, for squamous tumors with available histone acetylation data, we found samples with hypermethylated STAT5A have significantly lower acetylation of histone H3 lysine 14 (H3K14) (FDR = 0.05, Figure S6C), representing an overall gene repression.46 Our findings suggested that samples with hypermethylated STAT5A are significantly associated with immune cell depletion in squamous tumors.

Next, to further identify the expression specificity of STAT5A hypermethylation in HNSCC, we evaluated 29 representative HNSCC cases by immunohistochemistry (IHC) markers STAT5A (Table S3). We found prominent STAT5A protein expression in the tumor-infiltrating lymphocytes (TILs) and peritumoral lymphocytes, while STAT5A was minimally expressed or absent in tumor cells (Figure 6C). The overall level of STAT5A abundance in the stained slide was correlated with the level of TILs or peritumoral lymphocytes (Figure 6D). Interestingly, samples with hypermethylated STAT5A showed a distinct boundary between tumor cells and STAT5A-expressing immune cells (Figure 6E, left panels), while tumor samples with normal methylated STAT5A showed a mixture of tumor cells and STAT5A-expressing immune cells (Figure 6E, right panels). We speculated that immune cells with hypermethylated STAT5A might limit the migration of lymphocytes in the tumor microenvironment; however, the working model requires further investigation in the future.

Our finding of lower tumor microenvironment factors in STAT5A-hypermethylated samples is in agreement with previous studies showing that hematopoietic stem cell proliferation was severely impaired in Stats5A-deficient mice.46-61 It is also consistent with STAT5A being identified as a key tumor suppressor in lymphoma cell lines.62 In addition, studies have indicated that the development of HNSCC is closely related to immunosuppression and immune escape.63 These findings suggest that STAT5A hypermethylation may mediate the disease-dependent expression of STAT5A-targeted genes64 and contributes to altered tumor immunogenicity.

Aberrant methylation associated with therapeutic vulnerabilities

We next explored if epigenetic features can classify tumors into transcriptionally and translationally distinct subtypes. To identify subtypes based on methylation patterns, we generated methylation profiles from 687 tumors and NATs, and used Uniform Manifold Approximation and Projection (UMAP) to reduce methylation signals from 340,000 CGIs into two dimensions (Figures 7 and S7A). We found that tumors clustered by organ system, including the brain (GBM), kidney (ccRCC), lung (LUAD and LSCC), head and neck (HNSCC), pancreas (PDAC), and uterus (UCEC) (Figure 7A, first column). Additionally, squamous cell cancers, LSCC and HNSCC, formed a distinct cluster adjacent to LUAD. The distinct pattern between chromophobe renal cell carcinomas (C3N-00492 and C3N-01175) and ccRCC samples (Figure S7A) were consistent with their distinct origins.65,66 Direct comparisons between tumors and corresponding NATs revealed pronounced methylation differences (Figure 7A, second column). To investigate intrinsic lineage differences in DNA methylation between tumors possessing
different cells-of-origin, we examined differentially methylated CGIs in tumor samples compared to NAT samples and found 99 cancer-specific aberrant methylation promoters (Figure 7B; Table S4). Together, the results suggest that the DNA methylome of tumors faithfully reflects cell-of-origin and malignant transformation.

Next, using CGIs showing significant differences between tumors and NATs, we identified between 3 and 5 clusters from each cancer type (Figure 7A, third column; Figure S7B). We correlated those methylation subtypes with existing RNA expression-based subtypes (16,24,56,58–61) (Fisher’s exact test p < 0.05), and found that methylation subtypes captured several

Figure 5. STAT5A hypermethylation associated with pervasive STAT5A regulon changes
(A) Unsupervised clustering of STAT5A regulon genes using Pearson correlation of scaled RNA sequencing data. Annotations denote STAT5A expression and methylation levels. Mean activity indicates the overall sum of regulon activity. The color scale is proportional to expression (red: upregulation; blue: downregulation).
(B) Unsupervised clustering of STAT5A regulon genes using Pearson correlation of scaled global proteome data.
(C) Violin plot comparing regulon activity in hypermethylated STAT5A (yellow) and normally methylated STAT5A (gray) samples. 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.
(D) Pathway members and interactions in the STAT5A regulon. The mean expression differences between STAT5A hypermethylated samples and normally methylated samples are indicated by shading of the filled squares.
See also Figure S5 and Table S3.
A Himone Subtype Immune Score Score Immune Cell Status Stage Smoking Site Tumor Purity HSC Basophils CD8+ naive T-cells Mast cells Plasma cells Memory B-cells B-cells naïve B-cells Deep-seated memory B-cells pDCs CD8+ T-cells CD8+ Tem CD4+ Tem CD4+ memory T-cells Tregs CD4+ T-cells CD4+ naïve T-cells Macrophages M1 Macrophages M2 aDC DC tDC mDC MSC CLP T cell T cell Top cells TIL cells

B xCell Enrichment Score

C

D

E

(legend on next page)
important genomic features and clinical characteristics (Figure S7C; Table S4). For example, the UCEC C1 to C4 methylation subtypes are enriched with POLE, CNV-low, MSI-high, and CNV-high tumors, respectively. The LUAD C1 to C3 methylation subtypes are enriched with proximal-proliferative, proximal-inflammatory, and terminal respiratory unit tumors, respectively. GBM subtypes C4 and C5 feature mesenchymal phenotype and CpG island methylator phenotype associated with IDH1 mutation, respectively. Altogether, our results suggest that clustering cancer samples based on DNA methylation can help identify molecularly and clinically relevant subtypes.

To further explore the biological differences between methylation subtypes, we performed an over-representation pathway analysis using differentially expressed genes (DEGs) and proteins, revealing significantly enriched pathways related to tumorigenesis (Figure 7C; Table S5, p < 0.05). Some significant subtype-specific tumorogenic signatures were consistently observed at the transcriptomic and proteomic levels, such as enrichment of LSCC-C1 for NFE2L2 orchestrating the adaptive response to oxidative stress, and enrichment of ccRCC-C1 for tumoringenic transcriptional network coordinated by HSF1. Despite comprising distinct cancer types, LSCC-C3, LUAD-C2, PDAC-C2, HNSCC-C2, and GBM-C4 were characterized by immune-related signatures at RNA and/or protein levels. This correlation suggests signaling convergence among various cancers, in line with previous studies demonstrating that some methylation subtypes are significantly associated with immune signature. On the other hand, enrichment of pathway signatures was not detected in some methylation subtypes, which were thus only characterized by their distinct methylation pattern. For example, deficiency in DNA polymerase ε (POLE) proofreading generates an extensive number of somatic mutations and leads to a methylation profile, which is shown as the UCEC subtype C1 (Figures 7C and S7D). Overall, these results demonstrate heterogeneity within cancer types and how distinct methylation patterns may give rise to various cancer phenotypes.

To identify the interplay between driver mutation and methylation subtypes in tumorigenesis, we correlated methylation subtypes with 299 driver mutations. We found that 13 out of 25 methylation subtypes are significantly associated with cancer-specific driver mutations (Figures 7C and S7D). Consistently, enrichment of seven driver mutations in particular RNA-based subtypes have been identified previously, including UCEC-C4 (CNV-high) enriched with TP53 mutations, UCEC-C3 (MSI-high) with KMT2B mutations, UCEC-C2 (CNV-low) with CTNNB1 mutations, LUAD-C3 (terminal respiratory unit) with EGFR mutations, LUAD-C1 (proximal-proliferative) with STK11 mutations, and GBM-C4 (mesenchymal) with NF1 mutations. We also identified enrichment of mutations at epigenetic modifiers that could directly affect cancer methylome. For example, IDH1 deficiency generates high levels of 2-KG and leads to epigenetic reprogramming, which is shown as the GBM subtype C5. Mutations in PBRM1 (a chromatin remodeler), SETD2 (a histone methyltransferase), and BAP1 (a histone deubiquitinating enzyme) lead to distinct DNA methylation phenotypes in ccRCC. The significant correlation between driver mutations and methylation subtype suggested that deregulation of driver genes induces epigenetic reprogramming, rewireing regulatory networks together during tumorigenesis.

Beyond the methylation subtyping within each cohort, we also conducted methylation profiling and signature enrichment analysis across HNSCC, LSCC, LUAD, and PDAC, which form a distinct cluster in the UMAP projection (Figure 7A, fourth column; Figure S7E). We identified six methylation groups with various signatures, including groups enriched with immune-related signatures (MC1 and MC2), groups enriched with squamous tumors (MC3 and MC4), LUAD-dominant group (MC5), and PDAC-dominant group (MC6) (Figure S7F; Table S4, p < 0.05). MC5 and MC6 are enriched with cancer type-specific signatures such as β-cell development and surfactant metabolism, respectively. Following the observation of methylation subtypes, we again observed signaling convergence among various cancers in MC1 and MC2. Notably, cancers with squamous features (MC3 and MC4) were enriched with replication stress signatures, coinciding with their high degree of genomic instability. Our multi-omic integrative analysis enables the identification of common functionality arising from the same methylation profile across different cancer types.

Finally, we investigated druggable targets for sites with cis-acting dysregulated DNA methylation. We integrated cis-acting DNA methylation with the Clinical Interpretation of Variants in Cancer (CIVIC) and analyzed target genes with outlier expressions for which pharmacological intervention might be available (Table S6). Allowing for “off-label” drug treatment, we found that 19.2% of samples (132 of 687 tumors) would likely benefit from one or more treatments targeting genes altered by DNA methylation (Figure S7G). The most frequent druggable DNA methylation events across the seven cancer types are those on MGMT (n = 45 tumors), NAPRT (n = 31 tumors), and EGFR (n = 26 tumors). These findings may have important clinical implications. For example, tumor-specific loss of NAPRT, mediated...
by promoter hypermethylation, is synthetically lethal with NAMPT inhibitor treatment in multiple cancer types, resulting in inactivation of nicotinic acid salvage pathways.80 However, it is worth mentioning that we do not claim superiority over any other drug discovery method since it is difficult to evaluate the performance without gold standards. We claim that: (1) the high correlation among different data types provides validation to each other for the functional events in tumors; and (2) this study provides highly regulated aberrant DNA methylation events that are reflected in RNA expression and protein abundance readout, which is useful and complementary to other methods. Thus, collectively, these characterizations of cis-acting aberrant DNA methylation in cancer reveal potential new directions for treatment optimization.

DISCUSSION

The pan-cancer multi-omic analysis revealed driver gene regulation via DNA methylation, providing insights into methylation-based stratification of cancer patients. We identified and characterized methylation subtypes enriched with various RNA and protein signatures that have potential therapeutic and prognostic implications. Of interest, we observed that subsets of tumors in different organs may share a convergent immune-related signature. For these, the cancer methylome may offer opportunities for patient stratification to increase the efficacy of immune-based therapies. Moreover, we observed clinically relevant alterations with important therapeutic potential in 132 out of 687 tumors. Targeting those common aberrant methylation events could enhance the therapeutic reach of existing drugs by broadening the treatable patient and tumor populations. To maximize this benefit, future studies may look toward optimizing epigenetic therapies.

We uncovered several bona fide DNA methylation drivers showing functional consequences, including hypomethylated FGF2 and EGFR, which could be informative in expanding patient eligibility of conventional genotype-directed clinical trials. In addition, the apparent co-occurrence of FGF2 hypomethylation and amplification suggests that epigenetic enhancement of FGF2 expression may offer selective advantage for developing a second FGF2 alteration, perhaps from enhanced FGF2 signaling in these cells. The converse relationship is also possible: genomic alterations to FGF2 disrupt the reading, writing, or maintaining of DNA methylation machinery in tumor cells, subsequently leading to aberrantly reduced methylation within the FGF2 promoter. Regardless of the direction this relationship takes, since UCCE tumors harboring epigenetic and genetic alterations are significantly associated with FGF2 upregulation, it is likely that FGF2 hypomethylation works in concert with amplification at FGF2 to promote tumorigenesis.81

Studies have shown that HNSCC tumors evade the host immune system by manipulating their own immunogenicity.83 Our findings reinforce the critical nature of STAT5A as a signaling hub in modulating tumor immunogenicity across squamous cancers. Together, these findings suggest opportunities for therapeutic intervention by targeting epigenetic alterations within the STAT5A promoter. DNMT inhibitors, such as 5′-azacytidine, have been shown to reduce methylation of the STAT5A promoter in cell lines82 and are FDA-approved for treating myelodysplastic syndrome.83 Furthermore, activation of STAT5A signaling may transform an immunologically cold, inactive tumor into a hot, inflamed one and thus increase the anti-tumor immune response. Additional investigation is required to uncover the mechanisms mediating STAT5A hypermethylation and downstream immune-related signaling pathways, including the interaction between them, and the impact on therapeutic sensitivity.

There are several limitations to this study. First, recent investigations have shown that not only promoters, but also intragenic and intergenic regions, are widely modulated during disease progression.84 Here, we only focus on cis-acting DNA methylation at promoter regions, while trans-acting DNA methylation (i.e., DNA methylation acting upon other target genes) and other regulatory elements are not discussed. Second, previous studies have shown that DNA methylation and gene expression are not as frequently correlated as previously thought.85 Therefore, DNA methylation may have critical functions other than gene expression regulation. One possibility is that DNA methylation changes influence transcriptional potential rather than actual transcription status and could therefore be involved in the epigenetic plasticity of tumor cells.86,87 Another possibility is that aberrant intra- or intergenic DNA methylation in cancer cells may lead to increased non-synonymous mutation rate. Finally, although our integrative multi-omic analysis has limitation in differentiating DNA methylation changes specifically in cancer epithelial cells or TILs, we provide a reliable estimate of across a wide range of tumors, laying the groundwork for future single-cell analyses and spatial omics investigations. We emphasize the importance of isolating cancer epithelial cells and employing DNA methylation sequencing for future study. Altogether, future studies at single cell resolution may help reveal additional mechanistic details underlying the contribution of aberrant DNA methylation to tumor development.

Overall, our results help identify the contribution of DNA methylation in tumorigenesis and delineate its role in initiating and maintaining malignancies. This thorough account of cis-acting events and characterization of the cancer methylome will inform systematic explorations of aberrant DNA methylation and associated functional consequences, ultimately revealing potential new disease mechanisms and therapeutic opportunities.

Figure 7. Summary of the cancer methylome for cell-of-origin, tumor signatures, and therapy
(A) Projection of the cancer methylomes. Each point is a sample and is colored based on the cancer type (first column), sample type (second column), methylation subtype (third column), or multi-cancer methylation group (fourth column).
(B) Heatmap of differentially methylated CpG sites at promoter regions in seven cancer cohorts compared to normal adjacent tissue. Selected promoters annotated with the number of differentially methylated CpG sites are shown.
(C) Alluvial plot showing the per-cancer methylation subtypes (second row), their enriched significantly mutated genes (SMGs) (first row), enriched RNA expression signature (third row), and enriched protein signature (fourth row). The curved lines across panels correspond to different methylation subtypes. Signatures with FDR p-values <0.05 are highlighted with *. See also Figure S7 and Tables S4, S5, and S6.
STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Human participants
  - Clinical data annotation
- **METHODS DETAILS**
  - CPTAC datasets description
  - DNA methylation data preprocessing
  - Batch correction of DNA methylation value
  - Multi-omic data correlation analysis
  - Defining aberrant DNA methylation using RESET
  - Validation in TCGA cohorts
  - DEG analysis using scRNA-seq dataset
  - Defining cancer-associated genes
  - TFBS enrichment analysis
  - Profiling genetic alterations at driver genes
  - Histone acetylation analysis
  - Frequency of gene hyper- or hypomethylation
  - Regulon association analysis
  - Cell type enrichment deconvolution
  - Immunochemistry analysis
  - Methylation-driven subtyping
  - Pathway over-representation analysis
  - Identification of significantly mutated genes
  - RNA expression-based subtypes
  - Druggable genes with aberrant methylation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Statistical analysis
- **ADDITIONAL RESOURCES**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2023.07.013.

CONSORTIA


ACKNOWLEDGMENTS

We thank InPrint for the editing, Matthew A. Wyzalkowski for feedback on figures, and BioRender for diagrams. The Clinical Proteomic Tumor Analysis Consortium (CPTAC) is supported by the National Cancer Institute of the National Institutes of Health under award numbers U24CA210955, U24CA210986, U24CA210998, U24CA210654, U24CA210967, U24CA210972, U24CA210979, U24CA210993, U01CA214114, U01CA214116, and U01CA214125 as U24CA210972 to D.F., L.D., and S.P., and Contract GR0012005 to L.D. This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261201500003I, Task Order HHSN261000064. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the U.S. Government.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author(s) used ChatGPT to enhance its readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Received: July 12, 2022
Revised: May 30, 2023
Accepted: July 31, 2023
Published: August 14, 2023

REFERENCES


Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17, 98–110.


# STAR★METHODS

## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit polyclonal anti-STAT5A antibody</td>
<td>Atlas Antibodies</td>
<td>Catalog: HPA042128, RRID: AB_2677864</td>
</tr>
<tr>
<td><strong>Biological samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary tumor and normal adjacent tissue samples</td>
<td>Li et al.47</td>
<td><a href="https://pdc.cancer.gov/pdc/cptac-pancancer">https://pdc.cancer.gov/pdc/cptac-pancancer</a></td>
</tr>
<tr>
<td><strong>Chemicals, peptides, and recombinant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dako Protein Block, Serum-free blocking solution</td>
<td>Agilent Technologies Inc</td>
<td>Catalog: X090930-2</td>
</tr>
<tr>
<td>Dako Target Retrieval Solution, pH=6</td>
<td>Agilent Technologies Inc</td>
<td>Catalog: S236984-2</td>
</tr>
<tr>
<td>Dako Wash Buffer 10X</td>
<td>Agilent Technologies Inc</td>
<td>Catalog: S3006</td>
</tr>
<tr>
<td><strong>Critical commercial assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold</td>
<td>Illumina</td>
<td>Catalog: RS-122-2301</td>
</tr>
<tr>
<td>Infinium MethylationEPIC Kit</td>
<td>Illumina</td>
<td>Catalog: WG-317-1003</td>
</tr>
<tr>
<td>Nextera DNA Exosome Kit</td>
<td>Illumina</td>
<td>Catalog: 20020617</td>
</tr>
<tr>
<td>KAPA Hyper Prep Kit, PCR-free</td>
<td>Roche</td>
<td>Catalog: 07962371001</td>
</tr>
<tr>
<td>TMT-11 Reagent Kit</td>
<td>ThermoFisher Scientific</td>
<td>Catalog: A34808</td>
</tr>
<tr>
<td>BCA Protein Assay Kit</td>
<td>ThermoFisher Scientific</td>
<td>Catalog: 23225</td>
</tr>
<tr>
<td>PTMScan® Acetyl-Lysine Motif [Ac-K] Kit</td>
<td>Cell Signaling</td>
<td>Catalog: 13416</td>
</tr>
<tr>
<td>EnVision FLEX Visualizing Kit</td>
<td>Agilent Technologies Inc</td>
<td>Catalog: K800221-2</td>
</tr>
<tr>
<td><strong>Deposited data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIViC nightly, 062220</td>
<td>Griffith et al.79</td>
<td><a href="https://civicdb.org/home">https://civicdb.org/home</a></td>
</tr>
<tr>
<td>FANTOM5</td>
<td>FANTOM Consortium et al.87</td>
<td><a href="http://fantom.gsc.riken.jp/5/">http://fantom.gsc.riken.jp/5/</a></td>
</tr>
<tr>
<td>TARGET Methylation data</td>
<td>Pugh et al.78</td>
<td><a href="https://ccrg.cancer.gov/projects/target/">https://ccrg.cancer.gov/projects/target/</a></td>
</tr>
<tr>
<td>JASPAR</td>
<td>Khan et al.93</td>
<td><a href="https://jaspar.genereg.net/">https://jaspar.genereg.net/</a></td>
</tr>
<tr>
<td>InfiniumAnnotation</td>
<td>Zhou et al.94</td>
<td><a href="https://zwdzwd.github.io/">https://zwdzwd.github.io/</a></td>
</tr>
<tr>
<td>OmniPath</td>
<td>Turei et al.95</td>
<td><a href="http://omnipathdb.org/">http://omnipathdb.org/</a></td>
</tr>
<tr>
<td>ccRCC scRNA-seq data</td>
<td>Li et al.96</td>
<td><a href="https://portal.gdc.cancer.gov/projects/">https://portal.gdc.cancer.gov/projects/</a> CPTAC-3</td>
</tr>
<tr>
<td>PDAC scRNA-seq data</td>
<td>Cui Zhou et al.25</td>
<td><a href="https://data.humantumoratlas.org/">https://data.humantumoratlas.org/</a></td>
</tr>
<tr>
<td>Lungs scRNA-seq data</td>
<td>Travaglini et al.27</td>
<td><a href="https://www.synapse.org/#">https://www.synapse.org/#</a>!</td>
</tr>
<tr>
<td>CPTAC clinical and proteomic data</td>
<td>Li et al.47</td>
<td><a href="https://pdc.cancer.gov/pdc/cptac-pancancer">https://pdc.cancer.gov/pdc/cptac-pancancer</a></td>
</tr>
<tr>
<td>CPTAC genomic and transcriptomic data</td>
<td>Li et al.47</td>
<td><a href="https://pdc.cancer.gov/pdc/cptac-pancancer">https://pdc.cancer.gov/pdc/cptac-pancancer</a> and Cancer Data Service (CDS)</td>
</tr>
<tr>
<td>CPTAC DNA methylation data</td>
<td>This study</td>
<td><a href="https://pdc.cancer.gov/pdc/cptac-pancancer">https://pdc.cancer.gov/pdc/cptac-pancancer</a></td>
</tr>
<tr>
<td>CPTAC acetylation data</td>
<td>Li et al.47</td>
<td><a href="https://pdc.cancer.gov/pdc/cptac-pancancer">https://pdc.cancer.gov/pdc/cptac-pancancer</a></td>
</tr>
</tbody>
</table>

(Continued on next page)
## REAGENT or RESOURCE SOURCE IDENTIFIER

<table>
<thead>
<tr>
<th>Software and algorithms</th>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESET</td>
<td>Saghaeminia et al.⁹⁶</td>
<td><a href="http://ciriellolab.org/reset/reset.html">http://ciriellolab.org/reset/reset.html</a></td>
<td></td>
</tr>
<tr>
<td>ChAMP</td>
<td>Morris et al.⁹⁶</td>
<td><a href="https://www.bioconductor.org/packages/release/bioc/vignettes/ChAMP/inst/doc/ChAMP.html">https://www.bioconductor.org/packages/release/bioc/vignettes/ChAMP/inst/doc/ChAMP.html</a></td>
<td></td>
</tr>
<tr>
<td>Methylation array analysis pipeline for CPTAC</td>
<td>This study</td>
<td><a href="https://github.com/ding-lab/cptac_methylation/methylationArrayAnalysis">https://github.com/ding-lab/cptac_methylation/methylationArrayAnalysis</a> v3.9 Maksimovic et al.⁹⁷</td>
<td><a href="https://master.bioconductor.org/packages/release/workflows/html/methylationArrayAnalysis.html">https://master.bioconductor.org/packages/release/workflows/html/methylationArrayAnalysis.html</a></td>
</tr>
<tr>
<td>Illumina EPIC methylation array v0.6</td>
<td>See link</td>
<td><a href="https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylationEPICanno.ilm10b2.hg19.html">https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylationEPICanno.ilm10b2.hg19.html</a></td>
<td></td>
</tr>
<tr>
<td>ConsensusClusterPlus v1.48.0</td>
<td>Wilkerson et al.⁹⁸</td>
<td><a href="https://bioconductor.org/packages/ConsensusClusterPlus/">https://bioconductor.org/packages/ConsensusClusterPlus/</a></td>
<td></td>
</tr>
<tr>
<td>xCell v1.2</td>
<td>Aran et al.⁴⁵</td>
<td><a href="http://xcell.ucsf.edu/">http://xcell.ucsf.edu/</a></td>
<td></td>
</tr>
<tr>
<td>SomaticWrapper</td>
<td>Ding Lab</td>
<td><a href="https://github.com/ding-lab/somaticwrapper">https://github.com/ding-lab/somaticwrapper</a></td>
<td></td>
</tr>
<tr>
<td>Strelka2</td>
<td>Saunders et al.⁹⁹</td>
<td><a href="https://github.com/Illumina/strelka">https://github.com/Illumina/strelka</a></td>
<td></td>
</tr>
<tr>
<td>MUTECT v1.1.7</td>
<td>Cibulskis et al.¹⁰⁰</td>
<td><a href="https://software.broadinstitute.org/gatk/download/archive">https://software.broadinstitute.org/gatk/download/archive</a></td>
<td></td>
</tr>
<tr>
<td>VarScan v2.3.8</td>
<td>Koboldt et al.¹⁰¹</td>
<td><a href="http://varscan.sourceforge.net">http://varscan.sourceforge.net</a></td>
<td></td>
</tr>
<tr>
<td>Pindel v0.2.5</td>
<td>Ye et al.¹⁰²</td>
<td><a href="http://gmt.genome.wustl.edu/packages/pindel/">http://gmt.genome.wustl.edu/packages/pindel/</a></td>
<td></td>
</tr>
<tr>
<td>Fusion calling pipeline for CPTAC</td>
<td>Gao et al.¹⁰³</td>
<td><a href="https://github.com/cuidaniel/Fusion_hg38">https://github.com/cuidaniel/Fusion_hg38</a></td>
<td></td>
</tr>
<tr>
<td>STAR-Fusion v1.5.0</td>
<td>Haas et al.¹⁰⁴</td>
<td><a href="https://github.com/STAR-Fusion/STAR-Fusion/wiki">https://github.com/STAR-Fusion/STAR-Fusion/wiki</a></td>
<td></td>
</tr>
<tr>
<td>EricScript v0.5.5</td>
<td>Benelli et al.¹⁰⁵</td>
<td><a href="https://sites.google.com/site/bioericscript">https://sites.google.com/site/bioericscript</a></td>
<td></td>
</tr>
<tr>
<td>Integrate v0.2.6</td>
<td>Zhang et al.¹⁰⁶</td>
<td><a href="https://sourceforge.net/p/integrate-fusion/wiki/Home/">https://sourceforge.net/p/integrate-fusion/wiki/Home/</a></td>
<td></td>
</tr>
<tr>
<td>Copy Number Variant Calling</td>
<td>Ding Lab</td>
<td><a href="https://github.com/ding-lab/BICSEQ2">https://github.com/ding-lab/BICSEQ2</a></td>
<td></td>
</tr>
<tr>
<td>BIC-seq2</td>
<td>Xi et al.¹⁰⁷</td>
<td><a href="http://compbio.med.harvard.edu/BIC-seq/">http://compbio.med.harvard.edu/BIC-seq/</a></td>
<td></td>
</tr>
<tr>
<td>SomaticSV</td>
<td>Ding Lab</td>
<td><a href="https://github.com/ding-lab/SomaticSV">https://github.com/ding-lab/SomaticSV</a></td>
<td></td>
</tr>
<tr>
<td>Manta v1.6.0</td>
<td>Chen et al.¹⁰⁸</td>
<td><a href="https://github.com/Illumina/manta">https://github.com/Illumina/manta</a></td>
<td></td>
</tr>
<tr>
<td>MSFragger v3.4</td>
<td>Kong et al.¹⁰⁹</td>
<td><a href="https://msfragger.nesvilab.org/">https://msfragger.nesvilab.org/</a></td>
<td></td>
</tr>
<tr>
<td>Philosopher toolkit v4.0.1</td>
<td>da Veiga Leprevost et al.¹¹⁰</td>
<td><a href="https://philosopher.nesvilab.org/">https://philosopher.nesvilab.org/</a></td>
<td></td>
</tr>
<tr>
<td>TMT- Integrator</td>
<td>Djomehri et al.¹¹¹</td>
<td><a href="http://tmt-integrator.nesvilab.org/">http://tmt-integrator.nesvilab.org/</a></td>
<td></td>
</tr>
<tr>
<td>HTSeq v0.11.2</td>
<td>Anders et al.¹¹⁷</td>
<td><a href="https://htseq.readthedocs.io/en/master/">https://htseq.readthedocs.io/en/master/</a></td>
<td></td>
</tr>
<tr>
<td>DreamAI</td>
<td>Ma et al.¹¹²</td>
<td><a href="https://github.com/WangLab-MSSM/DreamAI">https://github.com/WangLab-MSSM/DreamAI</a></td>
<td></td>
</tr>
<tr>
<td>ProteintPaint Lollipop</td>
<td>Zhou Lab</td>
<td><a href="https://viz.stjude.cloud/zhou-lab/visualization/proteintpaint-lollipop-example-57">https://viz.stjude.cloud/zhou-lab/visualization/proteintpaint-lollipop-example-57</a></td>
<td></td>
</tr>
<tr>
<td>HOMER</td>
<td>Heinz et al.¹¹⁴</td>
<td><a href="http://homer.salk.edu/homer/">http://homer.salk.edu/homer/</a></td>
<td></td>
</tr>
<tr>
<td>QuPath v0.3.2</td>
<td>Bankhead et al.¹¹⁵</td>
<td><a href="https://qupath.github.io/">https://qupath.github.io/</a></td>
<td></td>
</tr>
<tr>
<td>Reactome</td>
<td>Fabregat et al.¹¹⁶</td>
<td><a href="https://reactome.org/">https://reactome.org/</a></td>
<td></td>
</tr>
<tr>
<td>Scannpy v1.7.0</td>
<td>Wolf et al.¹¹⁷</td>
<td><a href="https://github.com/scverse/scannpy">https://github.com/scverse/scannpy</a></td>
<td></td>
</tr>
<tr>
<td>Python v3.7</td>
<td>Python Software Foundation</td>
<td><a href="https://www.python.org/">https://www.python.org/</a></td>
<td></td>
</tr>
<tr>
<td>R v3.6</td>
<td>R Development Core Team</td>
<td><a href="https://www.R-project.org/">https://www.R-project.org/</a></td>
<td></td>
</tr>
<tr>
<td>Bioconda</td>
<td>Gruning et al.¹¹⁸</td>
<td><a href="https://bioconda.github.io/">https://bioconda.github.io/</a></td>
<td></td>
</tr>
<tr>
<td>Bioconductor v3.9</td>
<td>Huber et al.¹¹⁹</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
<td></td>
</tr>
</tbody>
</table>
RESOURCE AVAILABILITY

**Lead contact**
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Li Ding (lding@wustl.edu).

**Materials availability**
This study did not generate new unique reagents.

**Data and code availability**
- Raw and processed proteomics as well as open-access genomic data, can be obtained via Proteomic Data Commons (PDC) at https://pdc.cancer.gov/pdc/cptac-pancancer. Raw genomic and transcriptomic data files can be accessed via the Genomic Data Commons (GDC) Data Portal at https://portal.gdc.cancer.gov with dbGaP Study Accession: phs001287.v16.p6. Complete CPTAC Pan-Cancer controlled and processed data can be accessed via the Cancer Data Service (CDS, https://dataservice.datacommons.cancer.gov/). The CPTAC Pan-Cancer data hosted in CDS is controlled data and can be accessed through the NCI DAC approved, dbGaP compiled whitelists. Users can access the data for analysis through the Seven Bridges Cancer Genomics Cloud (SB-CGC) which is one of the NCI-funded Cloud Resource/platform for compute intensive analysis. Instructions to access data: 1. Create an account on CGC, Seven Bridges (https://cgc-accounts.sbgene.com/auth/register) 2. Get approval from dbGaP to access the controlled study (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001287.v16.p6) 3. Log into CGC to access Cancer Data Service (CDS) File Explore 4. Copy data into your own space and start analysis and exploration. 5. Visit the CDS page on CGC to see what studies are available and instructions and guides to use the resources. (https://docs.cancergenomicscloud.org/page/cds-data).
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

**Human participants**
A total of 687 participants were included in strict accordance with the CPTAC-3 protocol with informed consent from the patients. Prospective biospecimen collection (tumor and adjacent normal samples where feasible) followed a tumor type specific protocol and standard operating procedures, where sample collection, qualification and processing were optimized for both genomics and proteomics.16,24,56,58–61 CPTAC samples were collected by 30+ tissue source sites from both domestic and international locations and processed by a central biospecimen core resource. The samples were pathology qualified by a general pathologist and later reconfirmed by a disease-specific expert pathologist through histopathology image review and immunohistochemistry assays where applicable.

**Clinical data annotation**
Clinical data were obtained from TSS and aggregated by the Biospecimen Core Resource (BCR, Van Andel Research Institute (Grand Rapids, MI)). Data forms were stored as Microsoft Excel files (.xls). Clinical data can be accessed and downloaded from the CPTAC Data Portal and https://pdc.cancer.gov/pdc/cptac-pancancer as described in Li et al.47

METHODS DETAILS

**CPTAC datasets description**
We aggregated somatic variants, copy number variations, transcriptomic, proteomic, and clinical data generated by the National Cancer Institute CPTAC from CPTAC data portal, Genomic Data Commons (GDC), and published studies47 (See Data and Code Availability). The datasets include CPTAC Clear Cell Renal Cell Carcinoma (ccRCC) Discovery Study26, CPTAC Glioblastoma (GBM) Discovery Study24, CPTAC Lung Adenocarcinoma (LUAD) Discovery Study24, CPTAC Lung Squamous Cell Carcinoma (LSCC) Discovery Study58, CPTAC Head and Neck Cancer (HNSCC) Discovery Study58, CPTAC Pancreatic Ductal Adenocarcinoma (PDAC) Discovery Study16, and CPTAC Uterine Corpus Endometrial Carcinoma (UCEC) Discovery Study.57 Of note, previous CPTAC cohorts (Breast Invasive Carcinoma, Colon Adenocarcinoma, and Ovarian Serous Cystadenocarcinoma) did not obtain DNA methylation measurement, therefore we could not include those CPTAC cohorts in this study.

All the data were harmonized by CPTAC pipelines,47 which included alignment to the GDC hg38 human reference genome (GRCh38.d1.vd1), annotation with GENCODE v22 (RNA expression quantification) or v34 (the others), and thorough quality checks. Briefly, somatic mutations were called by the SomaticWrapper pipeline from Washington University in St Louis, which includes four different callers: Strelka2,99 MUTECT v1.1.7,100 VarScan v.2.3.8,101 and Pindel v.0.2.5.102 Copy number variation and structure variants were identified by BIC-seq2107 and Manta v.1.6.0,108 respectively (CPTAC pipeline from Washington University in St. Louis).
Gene fusions in RNA-Seq samples were identified using three callers: STAR-Fusion, EricScript, and Integrate, with fusions reported by at least 2 callers or reported by STAR-Fusion being retained (CPTAC pipeline from Washington University in St Louis). For transcriptomic data, gene-level stranded read counts were obtained using HTSeq v.11.2 and then converted to Fragments Per Kilobase of transcript per Million mapped reads (FPKM-UQ) values by following the GDC’s RNA-Seq pipeline, except running the quantification tools in the stranded mode (CPTAC pipeline from Washington University in St Louis). Tandem mass tags-based global proteomic and phosphoproteomic data were searched using the MSFragger search engine v3.4 against a GENCODE v34 protein FASTA database, processed by Philosopher toolkit v4.0.1 and quantified by TMT-Integrator. All the cohorts were uniformly processed and harmonized using the Philosopher toolkit. Missing values for the proteins or phosphosites that appeared in at least 50% of samples were imputed using DreamAI on each cohort separately. The acetylation data for six cancer types (BRCA, GBM, LUAD, LSCC and UCEC) were generated using MS/MS spectra. These spectra were analyzed with Spectrum Mill (SM) v7.08 to identify and quantify the shared acetylation sites among the cohorts.

**DNA methylation data preprocessing**

Raw methylation image files generated by Illumina Infinium EPIC BeadChip were downloaded from the CPTAC GDC (See Data and Code Availability). We calculated methylated (M) and unmethylated (U) intensities for tumor and normal adjacent tissue samples as described in the methylation processing pipeline on GitHub (See key resources table). Generally, we flagged a locus as NA if probes did not meet a detection P-value of 0.01. Probes with a minor allele frequency more than 0.1 were removed. Probes located on the sex chromosomes and samples with more than 85% NA values were removed from subsequent analysis. Infinium EPIC probes annotated as poor performing were filtered out, leaving 832,749 unique probes. To map EPIC arrays to GRCh38 assembly, all probes are reannotated by InfiniumAnnotation GENCODE v36, which was downloaded from the InfiniumAnnotation website (https://zwdzwd.github.io/InfiniumAnnotation).

The raw methylation image files of TARGET-NBL, TCGA-GBM, and TCGA-UCEC datasets generated by Illumina Infinium HumanMethylation450 BeadChip assays, used here for normal adjacent tissue approximation, were downloaded from TCGA FireHose or TARGET data matrix and processed as described above. We combined the data derived from the two platforms, but only used data on common probes (GBM: 364518 probes, UCEC: 364518 probes) (Table S1).

**Batch correction of DNA methylation value**

To account for batch effects introduced by different data sources, we performed batch-effect correction on the DNA methylation data using the ChAMP Bioconductor package with default parameters. The methylation data matrix was first imputed using the p.impute() function and then batch-corrected using champ.runCombat() function. Singular value decomposition (SVD) was conducted before and after batch correction for each cohort using champ.SVD() function. The choice of variable for batch correction was determined based on SVD diagnostic plots, where most cancer types were corrected using the “batch” variable, while for GBM and UCEC samples, the “source” variable was utilized. TCGA GBM and TCGA UCEC tumor samples were included solely for the purpose of batch correction and subsequently removed (Table S1). No batch correction was applied to HNSCC samples due to the lack of improvement in SVD diagnostic plots.

**Multi-omic data correlation analysis**

To investigate the correlation between mean promoter methylation, RNA expression, and protein abundance for 12,934 genes, we utilized linear regression models to fit three different relationships (methylation versus RNA expression, methylation versus protein expression, and RNA expression versus protein expression) and evaluated the significance of each model using adjusted R-squared values, regression coefficients, and P-values of coefficients. Specifically, we used the following model:

\[ Y_g = \beta_0 + \beta_1 M_g + \beta_2 P + \beta_3 C + \epsilon \]

where \( Y_g \) is a (n x 1) vector representing either protein or RNA abundance of a given gene, \( M \) is a vector indicating the methylation status for that particular gene \( g \) in a tumor sample, and the tumor purity \( P \) calculated by ESTIMATE was also included as a covariate. Cancer type was one-hot encoded and was included as an additional covariate \( C \). Lastly, the error \( \epsilon \) is assumed to be normally distributed with a constant variance \( \sigma \).

Additionally, we accounted for multiple comparisons and applied an FDR correction to determine the adjusted P-values and assess the significance of the associations with a threshold of P-value < 0.05. This comprehensive approach enabled us to identify which pairs were significantly associated.

**Defining aberrant DNA methylation using RESET**

We mapped CpG probes to canonical transcriptional start sites and unconventional exonic TSS as defined by FANTOM5 consortium, and then applied the RESET algorithm separately on each tumor type using the corresponding adjacent normal tissue samples. For integrating HM450 (adjacent normal tissue) and EPIC (tumor) datasets for GBM and UCEC, the arrays were intersected to a HM450 probeset to ensure comparability. To identify aberrant DNA methylation events associated with transcriptional or translational changes, aggregated mRNA expression and global proteomic data were used as an input, respectively. We considered the association as significant if the methylation event had FDR < 10%.
A parallel RESET run was conducted using purity-adjusted beta values. Briefly, a linear regression model is fitted to each tumor cohort using the original methylation estimate as the dependent and “1-purity” as the independent variable to obtain a linear fit, where the resultant intercept is the pure tumor methylation state. After normalizing values between 0 to 1 to satisfy the beta distribution of methylation values, the purity-adjusted beta value at individual CpG sites were obtained. We then repeated our association analysis with purity-adjusted beta values, and provided the mean FDR value in the “FDR_mean|RNA(Purity)” and “FDR_mean|Protein(Purity)” columns of Table S2 as an additional annotation to the reported events.

For Figure 2C, the methylation difference is derived from the mean beta values differences between samples with or without aberrant methylation. For RNA and protein differences, the values are derived from the mean differences of scaled RNA sequencing data or proteomic data between samples with or without aberrant methylation.

Validation in TCGA cohorts
For aberrant DNA methylation events with concurrent RNA and protein changes where DNA methylation (HM450 array) and RNA-seq data were available from TCGA, we downloaded and uniformly processed the raw DNA methylation data from 2,011 TCGA tumors across 7 cancer types, including ccRCC (N of tumors = 273), GBM (N = 58), HNSCC (N = 514), LSNC (N = 363), LUAD (N = 452), PDAC (N = 177), and UCEC (N = 174). The harmonized RNA expression, protein abundance data (reverse phase protein array, RPPA), mutation profiles, and copy number variation files were downloaded from GDAC. We then identified the aberrant methylation associated with RNA expression or protein abundance changes using the same RESET analysis aforementioned. Since the EPIC array used in this study represents a bigger genomic coverage when compared to the one from TCGA (HM450), we performed a direct comparison with RNA expression or protein abundance changes using the same RESET analysis aforementioned. Due to the sparsity of protein measurements included in the RPPA dataset from TCGA (the mean number of protein measurements in each cohort is 171.1), we calculated the validation rate at RNA level only. The validation results were shown in the “FDR_mean|RNA(TCGA)” and “FDR_mean|Protein(TCGA)” columns of Table S2.

DEG analysis using scRNA-seq dataset
We used published single-cell RNA-seq (scRNA-seq) datasets to annotate whether the reported aberrant DNA methylation event was likely to be identified in cancer cells, immune cells, or stromal cells. Briefly, we separated the cells into three main categories including tumor, immune, and stromal. Differentially expressed genes (DEGs) corresponding to each category were identified by the FindMarkers function in Seurat using the Wilcoxon rank sum test. DEGs were further filtered using the criteria of FDR < 0.05. DEGs identified in the same cancer type with a reported DNA methylation event at the corresponding promoters were annotated in the “scRNA-seq|cluster_matched_cancertype” column of Table S2. DEGs that were uniquely identified in cancer cells, immune cells, or stromal cells of the other cancer types were annotated in the “scRNA-seq|cluster_cross_cancertype” column of Table S2.

Defining cancer-associated genes
Cancer-associated genes were compiled from a 299 driver gene list defined by Bailey et al. and cancer-associated genes listed in Mertins et al. and adapted from Vogelstein et al.

TFBS enrichment analysis
The frequency distribution of transcription factor bind sites (TFBS) of hypermethylation and hypomethylation probes was conducted at locations of these probes and human hg38 TFBS locations in the JASPAR database. We further defined hyper and hypomethylation significant probes with driver genes and adopted the HOMER software to explore motifs enriched in these significant probes.

Profiling genetic alterations at driver genes
We obtained the somatic mutations in 299 driver genes harboring aberrant DNA methylation as described in “Identification of significantly mutated genes in methylation subtypes”. We further collected gene fusions involved in the driver genes. Aberrant DNA methylation was obtained as described in “Defining aberrant DNA methylation using RESET”. For the CPTAC samples, we defined the log2 copy ratio of each gene larger than 0.3 or smaller than -0.3 as amplification or deletion, respectively. For the TCGA samples, we defined the GISTIC value of -2, -1, 1, 2 as deep deletion, shallow deletion, shallow amplification, and deep amplification, respectively. We separated all the tumors into “No Genomic Alteration”, “Mutation”, and “CNV” categories based on the genomic alteration profiles of the locus being tested. For tumors harboring mutation, CNV, and gene fusions, we categorized them as “Mutation”. The missense mutations at FGFR2 locus were visualized using ProteinPaint Lollipop.

Histone acetylation analysis
To investigate the impact of IDH1/2 somatic alterations and aberrant methylations on the correlations between histone acetylation levels and methylation levels at IDH2 or α-KG target genes, we categorized a total of 393 patients with available acetylation and methylation data into four groups based on their IDH1/2 mutation and methylation status: IDH1/IDH2 wild-type (n=374), IDH1 mutants (n=11), IDH2 hypomethylated samples (n=5), and IDH2 mutants (n=3). To assess the levels of histone acetylation and DNA methylation at target genes in each group, we calculated Pearson correlation coefficients and determined the significance levels.
(P-values) of the correlations using the R function cor.test(). We focused on the acetylation-methylation pairs that showed significant correlations for further analysis (Pearson’s r > 0.2, P < 0.05, or Pearson’s r < -0.2, P < 0.05). To examine the differences in correlation coefficients among the groups, we utilized the R package rstatix (version 0.7.2). Specifically, we compared the group differences separately for positively and negatively correlated acetylation-methylation pairs.

We also compared the impact of IDH1 mutation in GBM and IDH2 hypomethylation in LSCC on histone acetylation levels. By examining 46 specific sites on histones H1, H2, H3, and H4, we compared the abundance of histone acetylation between IDH1 mutant (n=7) and wild-type (n=87) groups in GBM, as well as between IDH2 hypomethylation (n=5) and normal methylation (n=99) groups in LSCC using the Wilcoxon signed-rank test.

**Frequency of gene hyper- or hypomethylation**

To determine the frequency of hyper- or hypomethylation in each tumor, we calculated the proportion by aggregating the counts of CGI sites exhibiting aberrant methylation and dividing it by the total number of CGI sites in the tumor sample. This analysis was performed using the xCell web tool.

**Regulon association analysis**

We collected transcription factor-interacting genes from publicly available databases OmniPath, and grouped the samples into regulon-high (>0) and regulon-low (<0) based on the sum of transcription factor-interacting genes using RNA expression and protein abundance data. To test the association of the regulons with the DNA methylation status of transcription factors, we used Fisher’s exact test to test for overrepresentation of samples with aberrantly methylated transcription factors in the set of samples defining the regulon activity as described above. To account for multiple comparisons in Fisher’s exact test, we applied the Benjamini-Hochberg (BH) method to adjust the P-values for false discovery rate (FDR) control. We narrowed the list of transcription factors down to STAT5A and STAT5A-interacting genes based on FDR P-value <= 0.1 and the number of interacting genes >=5. We then performed unsupervised clustering of the STAT5A regulon using the Ward.D2 linkage method to generate heatmaps.

**Cell type enrichment deconvolution**

We inferred the abundance of each cell type using the xCell web tool using the FPKM-UQ expression matrix as input. xCell is a gene signature-based method learned from thousands of pure cell types from various sources, which performed the cell type enrichment analysis from gene expression data for 64 immune andstromal cell types (default xCell signature). xCell generates an immune score per sample that integrates the enrichment scores of immune cells (B cells, CD4+ T-cells, CD8+ T-cells, DC, eosinophils, macrophages, monocytes, mast cells, neutrophils, and NK cells), a stroma score, and a micro-environment score which is the sum of the immune score and stroma score.28,47

**Immunohistochemistry analysis**

Immunohistochemistry (IHC) staining was performed on 4-micron formalin-fixed, paraffin-embedded (FFPE) tissue sections. Prior to staining, antigen retrieval was performed using the heat-induced epitope retrieval method at pH 6. Staining employed the Dako Autostainer Link 48 with EnVision FLEX visualizing kit (K800221-2; Dako, Agilent Technologies Inc.) and rabbit polyclonal antibody against human STAT5A (HPA042128, Atlas Antibodies, 1:150 dilution). Appropriate known positive and negative control tissue were run in each assay batch.

A semi-quantitative product score for tumor-infiltrating lymphocytes (TILs) and peritumoral lymphocytes were scored by the study pathologists. Percentage of lymphocytes represents the positive lymphocytes relative to the total number of cells (tumors and lymphocytes) in the stained slide, while the intensity represents staining intensity of STAT5A (none, 0; weak, 1; moderate, 2; strong, 3). An independent H-score of STAT5A abundance in the stained slide is calculated using QuPath.115

**Methylation-driven subtyping**

To identify subtype based on cancer methylome, we took CGIs showing significant differences between tumors and NATs and performed unsupervised classification of tumors using consensus clustering on the most variable 8,000 CGIs by the R package ConsensusClusterPlus (parameters: reps = 2000, pItem = 0.9, pFeature = 0.9, clusterAlg = "kmdist", distance = "spearman"). Samples were assigned to the optimal number of clusters.

**Pathway over-representation analysis**

To designate the representative pathways of methylation subtypes from transcriptomic and proteomic data, we used the Wilcoxon rank sum test to select the top 250 differentially expressed features (RNA expression and protein abundance) for each subtype or methylation group. We then performed hierarchical clustering on these features. Each set of clustered features underwent pathway enrichment analysis using Reactome.116 Pathways with P-value smaller than 0.05 were manually reviewed and selected based on the following rules: (1) if similar signatures showed up repeatedly in the significant list, (2) if the significant pathway is supported by literature, and (3) if the signature had been supported by FDR test with P-value smaller than 0.05, (4) if the signatures were consistently observed at the transcriptomic and proteomic levels. Those manually reviewed pathways are highlighted in Figures 7C and S7F.
Identification of significantly mutated genes
A list of 299 driver genes was downloaded from Bailey et al.29 and results of somatic mutations were downloaded from the CPTAC discovery studies.16,24,56,58–61 We compiled a genomic alteration profile for each sample (file name: PanCan_Union_Maf_Broad_WashU_v1.1.maf), including frameshift deletion, frameshift insertion, inframe deletion, inframe insertion, missense mutation, nonsense mutation, nonstop mutation, splice site mutation, and intron mutation in 299 driver genes. We then categorized samples with driver mutations based on their methylation subtypes, and conducted Fisher’s exact test to test for overrepresentation of any key driver somatic alterations in the methylation subtypes.

RNA expression-based subtypes
The classifications of RNA expression-based subtypes of ccRCC, GBM, HNSCC, LSCC, LUAD, PDAC, and UCEC samples were downloaded from the CPTAC discovery studies.16,24,58–61 We then compared the association of RNA expression-based subtype and methylation subtypes using Fisher’s exact test.

Druggable genes with aberrant methylation
CIViC is a curated list of druggable variants describing their therapeutic, prognostic, diagnostic and predisposing relevance. We downloaded the list and intersected the list of genes with aberrant methylation. The potential druggability of each gene was manually reviewed to see if the altered expression of such gene is associated with therapeutic relevance supported by literature.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis
All statistical analyses were performed using R or Python unless explained otherwise. Multiple comparisons were adjusted by the Benjamini-Hochberg.127 Statistical parameters for each experiment are reported in the respective figure legend.

ADDITIONAL RESOURCES
The CPTAC program website, which includes details about program initiatives, investigators, and datasets, can be accessed at https://proteomics.cancer.gov/programs/cptac.