Supplementary Materials

SUPPLEMENTARY METHODS (p2)

SUPPORTING FIGURES (p9)

Supporting FIG. S1. Reproducibility of methylation components across the three data sets.

Supporting FIG. S2. Definition of CGI-based, gene-based features and chromatin states.

Supporting FIG. S3. Correlation matrix of Roadmap chromatin domains between normal liver and the cancerous HepG2 cell line.

Supporting FIG. S4. Transcriptional impact of each methylation component.

Supporting FIG. S5. Association of methylation components with CGI- and gene-based features, chromatin states and replication timing in the LICA-FR, TCGA-LIHC and HEPTROMIC cohorts.

Supporting FIG. S6. Association of methylation components with CpG context in the LICA-FR, TCGA-LIHC and HEPTROMIC cohorts.

Supporting FIG. S7. Correlation of MC1 with age across diverse normal and tumor tissues.

Supporting FIG. S8. t-SNE plots showing the methylation-based classification of HCC with associated clinico-molecular features and methylation components.

Supporting FIG. S9. Pre-neoplastic DNA methylation changes in cirrhotic liver.

SUPPORTING TABLES (provided as Excel files)

Supporting Table S1. Clinical and molecular alterations for the LICA-FR series.

Supporting Table S2. Univariate analysis of clinical and molecular features associated with methylation components in LICA-FR and TCGA series.

Supporting Table S3. Multivariate analysis of clinical and molecular features associated with methylation components in LICA-FR and TCGA series.

SUPPLEMENTARY METHODS

Public data used in this study

TCGA-LIHC cohort

The TCGA-LIHC cohort comprises 275 HCC and 50 non-tumor liver tissues analyzed using Illumina Infinium HumanMethylation450 arrays, whole exome and RNA sequencing. Clinical annotations, DNA methylation (QC metrics and methylation beta values) and RNA-seq data (raw read counts per gene) were obtained from the TCGA data portal (https://tcga-data.nci.nih.gov). Cholangiocarcinomas and mixed forms of HCC were discarded to keep only pure HCC and non-tumor samples. Single somatic mutations and TERT promoter mutation were retrieved from the original article (1). HCC cases in this cohort are predominantly of American, Canadian or Vietnamese origin, mostly males (64%), with a median of 63 years and related to diverse risk factors: alcohol (34%), HBV (14%), HCV (17%). Detailed clinical characteristics and sequencing details for each sample are available at the TCGA website.

HEPTROMIC cohort

The HEPTROMIC cohort comprises 221 surgically resected HCC and 19 non-tumor liver tissues analyzed using Illumina Infinium HumanMethylation450 arrays (2). HCC cases in this cohort come from two institutions of the HCC Genomic Consortium: IRCCS Istituto Nazionale Tumori (Milan, Italy) and Hospital Clínic (Barcelona, Spain). HEPTROMIC cases are mostly males (78%), with a median of 66 years and a predominance of viral-related etiologies (HCV, 47%; HBV, 20%). No additional molecular data was available for this cohort

Description of (epi)genomic features analyzed for correlation with methylation components

For each component, we examined the association of the following (epi)genomic features were considered:

- CpG island-based features: CpG islands retrieved from UCSC database (release 19, GRCh37), shores (2 kb on each side of the islands) and shelves (2 kb outside shores)
- gene-based features: promoter (defined as transcription start site (TSS) +/- 500 bp) and gene body for each gene in GENCODE database (release 19 - GRCh37.p13)
- chromatin states in normal liver as defined by the ROADMAP consortium (3). Eighteen chromatin states were defined by the consortium based on the genome-wide analysis of 6 histone marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3 and H3K27ac) using

a multivariate Hidden Markov model (ChromHMM tool (4)). Each chromatin state corresponds to a particular combination of histone marks and is associated with a specific type of functional element (e.g. active TSS, genic enhancers, heterochromatin…). We downloaded the bed file of chromatin states in normal liver through the ROADMAP epigenomics website (http://www.roadmapepigenomics.org).

- DNA methylation domains derived from whole genome bisulfite sequencing of normal hepatocytes (5). Genome-wide CpG methylation analyses have shown that the epigenome is organized in megabase-scale partially methylated domains (PMD, methylation between 50% and 80%) and highly methylated domains (HMD, methylation $>$ 80%), as well as short (regulatory) lowly methylated (LMR, methylation between 10% and 50%) and unmethylated regions (UMR, methylation < 10%) (6,7). We retrieved these domains in normal liver defined by Salhab *et al.* (5) using a Hidden Markov Model-based detection method called methylSeekR (7).
- replication timing in the liver cancer cell line HepG2. We used Repli-seq data generated by the ENCODE project (8) to characterize the replication timing of each CpG site. To do so, we downloaded the wavelet-smoothed Repli-seq signals for HepG2 cell line through the UCSC genome browser, and we segmented this signal into 10 deciles from the earliest (decile 1) to the latest (decile 10) replicated regions.

Description of clinico-molecular annotations analyzed for correlation with methylation components

Clinical features for the LICA-FR cohort are detailed in Supporting Table S1 and included patient information (gender, geographic origin, age), risk factors (alcohol intake, HBV or HCV infection, metabolic syndrome), underlying liver disease (METAVIR fibrosis stage; F0-F1: no fibrosis, F2-F3: moderate fibrosis, F4: cirrhotic liver) and various tumor characteristics like the number of nodules and size of the largest nodule, vascular invasion, Barcelona Clinic Liver Cancer stage (BCLC 0, A, B, C and D from the earliest to the terminal stage) and Edmonson grade (I-II = well differentiated, III-IV = poorly differentiated). In the TCGA cohort, the same features were analyzed except the following that were not available: metabolic syndrome, number of nodules and largest nodule size, vascular invasion, BCLC stage and Edmonson grade. The Ishak fibrosis score was converted to METAVIR for comparison with the LICA-FR series as follows: "0,1,2 - No Fibrosis or Portal Fibrosis" = F0-F1; "3,4 - Fibrous Speta" = F2-F3; "5,6 - Nodular Formation, Incomplete Cirrhosis" and "Established Cirrhosis" = F4).

Molecular features analyzed in both cohorts included:

- driver alterations of 27 HCC driver genes defined by Schulze *et al.* (9) or characterized recently in the lab (10,11): *TERT*, *CTNNB1*, *TP53*, *ARID1A*, *AXIN1*, *CDKN2A*, *ARID2*, *RPS6KA3*, *NFE2L2*, *KEAP1*, *PTEN*, *HNF1A*, *ALB*, *ACVR2A*, *RPL22*, *CDKN1A*, *RB1*, *TSC2*, *ATP10B*, *FGA*, *MEF2C*, *ZNRF3*, *EPHA4*, *TSC1*, *CCNA2*, *CCNE1*, *BAP1*). Mutational status for these 27 genes was derived from whole exome or whole genome sequencing, completed by *TERT* promoter screening by Sanger sequencing for both the TCGA-LIHC (1) and LICA-FR (12) cohorts.
- molecular subgroups of HCC, G1 to G6, defined by Boyault *et al.* from gene expression data (13). In the LICA-FR cohort, G1-G6 groups were predicted using a combination of 16 marker genes analyzed in qRT-PCR, as previously described (14). We used the *MS.liverK* package (15) to predict the G1-G6 groups based on RNA-seq expression data in the TCGA LIHC cohort. selected transcriptional signatures related to hepatocellular carcinoma phenotypes were analyzed, including differentiation (*ALB*, *CDH1*, *APOF*, *CYP1A1*, *CYP2A6*, *UGT2B7*, *HNF1A*, *HNF4A*) and proliferation (*CDC20*, *GMNN*, *MKI67*, *RRM2*, *CCNA2*, *CCND1*, *CCNE1*, *AURKA*, *BUB1*, *PCNA*, *RAN*, *BIRC5*, *SPP1*) signatures defined by Nault *et al.* (14), as well as liver progenitor (*PROX1*, *AFP*, *EPCAM*, *IGF2*, *SALL4*, *PROM1*, *LGR5*, *GPC3*, *LIN28B*), stem cell (*CD47*, *CD44*, *KDR*, *IL6*, *NCAM2*, *THY1*, *KIT*) and epithelial-mesenchymal transition/metastasis (*SNAI2*, *ITGB3*, *TWIST1*, *ZEB2*, *PLAUR*, *VIM*) signatures defined by Caruso *et al.* (16). For each signature, a score was computed in each tumor as the mean expression of marker genes.
- immune infiltrate estimated from RNA-seq data using the MCPcounter tool (17). The overall immune infiltrate was obtained by summing MCPcounter scores for all immune cell populations.

Linking CpG methylation with transcriptional networks

We used ELMER tool (18) to identify CpG-gene pairs, i.e. correlations between the methylation level of a CpG site and the expression of one or more nearby genes, leveraging samples with matched methylation array and RNA-seq data. The *get.pair* function of ELMER v2 package (19) was used in unsupervised mode to compare the expression of the 10 genes closest to each CpG site between the

40% samples with the highest/lowest methylation level for that CpG. We used a permutation size of 10,000 and selected CpG-gene pairs with an empirical p-value *Pe* < 0.001. We used an in-house adaptation of the GSEA (Gene Set Enrichment Analysis) method (20), modified to take as input a ranked gene list instead of an expression matrix, to identify gene sets associated with each methylation component (MC). For each MC, genes were ranked according to the contribution of their paired CpG. Genes paired with several CpGs were assigned to the CpG with the strongest contribution to the component (in absolute value). GSEA was then used to identify gene sets from the MSigDB v6 database overrepresented among genes paired with the most contributing CpGs. We used the *get.enriched.motif* of ELMER v2 package to identify transcription factor binding motifs enriched around the most contributing CpGs of each MC.

DNA methylation-based classification of hepatocellular carcinomas and non-tumor liver tissues

We used consensus clustering (21) to identify HCC subgroups on the basis of their DNA methylation profiles. A same set of CpGs was used for the LICA-FR and TCGA LIHC cohorts, corresponding to the union of the 10,000 most variant probes (based on standard deviation) in each series. We then established consensus partitions of the data set in K clusters (for $K = 2, 3, ..., 8$), based on 1,000 resampling iterations of hierarchical clustering, with Pearson's dissimilarity as the distance metric and Ward's method for linkage analysis. We used the cumulative distribution functions (CDF) of the consensus matrices to determine the optimal number of clusters, considering both the shape of the functions and the area under the CDF curves. The Bioconductor ConsensusClusterPlus package was used for consensus clustering analysis. T-stochastic neighbor embedding (tSNE) was used to project the data set in two dimensions using the *Rtsne* package (https://github.com/jkrijthe/Rtsne). t-SNE was applied to a Pearson correlation matrix of CpGs with standard deviation > 0.25, with a theta value of zero over 2,000 iterations and perplexity of 9 for TCGA-LIHC and 6 for LICA-FR.

We also performed an unsupervised classification of non-tumor liver tissues from the LICA-FR cohort. Hierarchical clustering was done on the 15 000 most variant probes (based on standard deviation) using R function *hclust* with Pearson's dissimilarity as distance metric and Ward.D2 linkage method.

Supplementary Methods References

- 1. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu, Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327-1341.e23.
- 2. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology. 2015;61:1945–1956.
- 3. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–330.
- 4. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods. 2012;9:215–216.
- 5. Salhab A, Nordström K, Gasparoni G, Kattler K, Ebert P, Ramirez F, et al. A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. Genome Biology. 2018;19:150.
- 6. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462:315– 322.
- 7. Burger L, Gaidatzis D, Schübeler D, Stadler MB. Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res. 2013;41:e155.
- 8. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
- 9. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat. Genet. 2015;47:505–511.
- 10. Bayard Q, Meunier L, Peneau C, Renault V, Shinde J, Nault J-C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. Nat Commun. 2018;9:5235.
- 11. Hirsch TZ, Negulescu A, Gupta B, Caruso S, Noblet B, Couchy G, et al. BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. J. Hepatol. 2020;72:924–936.
- 12. Nault JC, Mallet M, Pilati C, Calderaro J, Bioulac-Sage P, Laurent C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. Nat Commun. 2013;4:2218.
- 13. Boyault S, Rickman DS, de Reyniès A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology. 2007;45:42–52.
- 14. Nault J-C, Martin Y, Caruso S, Hirsch TZ, Bayard Q, Calderaro J, et al. Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. Hepatology. 2020;71:164–182.
- 15. Petitprez F, Meunier L, Letouzé E, Hoshida Y, Villanueva A, Llovet J, et al. MS.liverK: an R package for transcriptome-based computation of molecular subtypes and functional signatures in liver cancer. bioRxiv. 2019;540005.
- 16. Caruso S, Calatayud A-L, Pilet J, La Bella T, Rekik S, Imbeaud S, et al. Analysis of Liver Cancer Cell Lines Identifies Agents With Likely Efficacy Against Hepatocellular Carcinoma and Markers of Response. Gastroenterology. 2019;157:760–776.
- 17. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218.
- 18. Yao L, Shen H, Laird PW, Farnham PJ, Berman BP. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. Genome Biol. 2015;16:105.
- 19. Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. Bioinformatics. 2019;35:1974–1977.
- 20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. 2005;102:15545–15550.
- 21. Monti S. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. :28.

SUPPORTING FIGURES

Supporting FIG. S1. Reproducibility of methylation components across the three data sets. Pearson correlation coefficients were used to link each methylation component (MC) extracted in one cohort with its closest equivalent in the other two cohorts. The figure displays the Pearson correlation scores between each MC pair. MCs identified in one series without a match in the second are indicated in red. Abbreviations: LICA-FR, Liver Cancer (France); MC, methylation component; TCGA-LIHC, The

Supporting FIG. S2. Definition of CGI-based, gene-based features and chromatin states. Each CpG site was annotated relative to its position with respect to genes, CGIs and chromatin states. (A) Gene-based features comprise transcription start sites (TSS) +/- 500 bp and gene bodies for each gene in GENCODE database (release 19 - GRCh37.p13). CGI-based features comprise CpG islands (UCSC database release 19, GRCh37), shores (2 kb on each side of the islands) and shelves (2 kb outside shores). (B) Chromatin states were defined in various cell types by the Roadmap consortium based on the genome-wide analysis of 6 histone marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3 and H3K27ac). This panel, adapted from Kundaje *et al.*, indicates the combination of histone marks associated with each chromatin state. (C) Integrative Genomics Viewer (IGV) visualization of histone marks and chromatin states across the *CTNNB1* locus in liver tissue (Roadmap data).

Supporting FIG. S3. Correlation matrix of Roadmap chromatin domains between normal liver and the cancerous HepG2 cell line. Chromatin states (numbered as in Figure 2A) display a highly significant overlap between normal liver and HepG2 (*P* < 2.2e-16, Chi-square test). When considering the 18 chromatin state categories, 37% of CpG sites analyzed with the Illumina Infinium 450k Beadchip have the same state in normal liver and HepG2. However, most discrepancies involve very close chromatin states, e.g. "TssA" in normal liver and "TssFlnk" in HepG2. When grouping chromatin states into more general categories (namely TSS, enhancer, transcription, and inactive chromatin), 71% of Contract the same state in normal liver and HepG2. However, the same state in normal liver and "B. Significant overlap between normal liver and HepG2. (Pet the 18 chromatin states, e.g. "TssA" in normal liver and "TssFlnk

Supporting FIG. S4. Transcriptional impact of each methylation component. Genes "paired" with each CpG (whose expression is correlated with CpG methylation) were identified using the ELMER tool. The bar plots represent the proportion of the most contributing CpG sites of each component with a "paired" gene. The color of each bar indicates the average Pearson correlation score between CpG methylation and the expression of their paired gene. Abbreviation: ELMER, Enhancer Linking by Methylation/Expression Relationships.

Supporting FIG. S5. Association of MCs with CGI- and gene-based features, chromatin states, and replication timing in the LICA-FR, TCGA-LIHC and HEPTROMIC cohorts. The most contributing CpG sites of each component were extracted in each cohort, and the enrichment of these CpG sites across CGI features, gene-based features, chromatin states, and replication deciles are shown below. Abbreviations: CGI, CpG island; TSS, transcription start site; ZNF, zinc finger.

Supporting FIG. S6. Association of MCs with CpG context in the LICA-FR, TCGA-LIHC and HEPTROMIC cohorts. The most contributing CpG sites of each component were extracted in each cohort, and the enrichment of these CpG sites across 48 CpG categories are shown below. Methylation categories are defined based on the methylation domain in normal liver (HMD: Highly Methylated Domain; PMD: Partially Methylated Domain; LMR: Lowly Methylated Region; UMR: UnMethylated Region), local CpG density (number of flanking CpGs within 35 base pairs on each side of the dyad), and sequence context (SCGS, SCGW, or WCGW, with S denoting C or G and W denoting A or T). Abbreviations: HMD, highly methylated domain; LMR, lowly methylated region; PMD, partially methylated domain; UMR, unmethylated region.

Supporting FIG. S7. Correlation of MC1 with age across diverse normal and tumor tissues. (A) The average methylation across the most contributing CpG sites of MC1 was calculated in hepatocellular carcinomas and normal liver tissues from the LICA-FR (left) and TCGA-LIHC (right) series. Linear regression was used to estimate the *P* value and slope, indicating the beta-value increase per year. (B) Similar analysis in various cancer types and matched normal tissues from the TCGA project. Linear regression for normal samples was only calculated when 10 or more samples were available. Abbreviations: bval, beta value; N, normal tissue; T, tumors.

Supporting FIG. S8. t-SNE plots showing the methylation-based classification of HCC with associated clinico-molecular features and MCs. t-SNE plots depict the classification of HCC from the (A) LICA-FR and (B) TCGA cohorts based on their DNA methylation profiles. Associated clinical and

molecular features and the intensity of each MC are represented by color codes. Abbreviations: ARID1A, AT-rich interactive domain-containing protein 1A; CCNA2, cyclin A2; CCNE1, cyclin E1; CTNNB1, catenin beta 1; M, mutated ; NM, non-mutated; t-SNE, t-distributed stochastic neighbor embedding; transcr., transcriptomic; WT, wild-type.

Supporting FIG. S9. Pre-neoplastic DNA methylation changes in cirrhotic liver. (A) Hierarchical clustering of non-tumor liver tissues reveals four homogenous subgroups strongly associated with fibrosis stage. (B) Activity of components MC6 and MC7 in HCC and adjacent non-tumor liver tissues with different levels of fibrosis (METAVIR stages) in the LICA-FR series. (C) Activity of components MC6 and MC7 in HCC and adjacent non-tumor liver tissues with different levels of fibrosis (METAVIR stages) in the TCGA series. (D) MC6 activity is strongly correlated to the immune-mediated cancer field (ICF) signature evaluated from RNA-seq data. Sample type (HCC or non-tumor liver with diverse fibrosis stage) is indicated with a color code as in panel (B). (E) Heatmap representing the methylation of the

MC6 MRCpGs and the expression of paired genes identified with ELMER. Samples are ordered according to MC6 activity (color code for sample type as in (B)), and the ICF gene expression signature is represented below. (F) Left: Transcription factor binding motif enrichment around CpG sites hypermethylated in samples with the highest activity of MC6. Right: Gene set enrichment analysis of genes paired with CpG sites hypermethylated in samples with the highest activity of MC6. (G) Same as (F) for hypomethylated CpG sites. Activity of components MC6 and MC7 in HCC and adjacent nontumor liver tissues with different levels of fibrosis (METAVIR stages) in TCGA-LIHC series. Abbreviations: ETV6, ETS variant TCF 6; HNF1A/B, hepatocyte nuclear factor 1 alpha/beta; HNF4A/G, HNF 4 alpha/gamma; ICF, immune-mediated cancer field; MRCpGs, most representative CpG sites;; RNA-seq, RNA sequencing; RUNX3, RUNX family transcription factor 3; TF, transcription factor.

Sample	Sample type	Gender	Age at sampling	Geographical origin
CHC018T	HCC	F	35	Africa
CHC229T	HCC	F	65	Europe
CHC231T	HCC	M	66	Europe
CHC013T	HCC	M	63	Europe
CHC441T	HCC	M	77	Europe
CHC333T	HCC	M	73	Europe
CHC239T	HCC	F	21	Africa
CHC399T	HCC	M	67	Europe
CHC014T	HCC	M	30	Africa
CHC043T	HCC	M	56	Asia
CHC037T	HCC	M	51	Africa
CHC339T	HCC	F	26	Africa
CHC245T	HCC	M	64	Europe
CHC253T	HCC	M	67	Europe
CHC158T	HCC	M	65	Europe
CHC445T	HCC	M	55	Europe
CHC080T	HCC	M	43	Europe
CHC335T	HCC	M	68	Europe
CHC230T	HCC	M	70	Europe
CHC228T	HCC	M	48	Europe
CHC010T	HCC	F	18	Europe
CHC137T	HCC	M	71	Europe
CHC205T	HCC	M	46	Europe
CHC218T	HCC	M	69	Europe
CHC081T	HCC	F	76	Asia
CHC031T	HCC	M	67	Europe
CHC242T	HCC	M	70	Europe
CHC059T	HCC	M	40	Europe
CHC220T	HCC	M	73	Europe
CHC206T	HCC	M	64	Europe
CHC152T	HCC	M	64	Europe
CHC046T	HCC	M	61	Europe
CHC211T	HCC	M	69	Europe
CHC437T	HCC	M	59	Europe
CHC725T	HCC	M	60	Europe
CHC317T	HCC	F	69	Europe
CHC789T	HCC	M	54	Europe

Table S1: Clinical and molecular annotations for the 274 samples of the LICA-FR series

Table S2a: Association between methylation components and clinico-molecular features in LICA-FR series (univariate) *P-values obtained with univariate linear models are indicated for each methylation component (MC) d*

Sacries (univariate)

P-values obtained with univariate linear models are indicated for each methylation component (MC) and each annotation.

 $\overline{}$

Table S2b: Association between methylation components and clinico-molecular features

P-values obtained with univariate linear models are indicated for each methylation component (MC) and analy each methylation component and annotation.

$\boldsymbol{\mu}$ in TCGA LIHC series (univariate)

nent (MC) and each annotation.

Table S3a: Association between methylation components and clinico-molecular features in LICA-F *P-values obtained with multivariate linear models are indicated for each methylation component (NC) Only features significant in univariate analyses of both LICA-FR and TCGA LIHC series were included i*

R series (multivariate)

P-values ob annotation.

Table S3b: Association between methylation components and clinico-molecular features in TCGA

P-values obtained with multivariate linear models are indicated for each methylation component (NC) Only features significant in univariate analyses of both LICA-FR and TCGA LIHC series were included i

TIAC series (multivariate)

PC) and each annotation.

n multivariate models. For other features the p-value is NA (not applicable).

