Intra-tumoral epigenetic heterogeneity and aberrant molecular clocks in hepatocellular carcinoma

Paula Restrepo[1,2+], Adrian Bubie[1+], Amanda J. Craig[2,3], Ismail Labgaa[4], Myron Schwartz[5], Swan Thung[6], Gustavo Stolovitzky[7], Bojan Losic*[1,2], Augusto Villanueva*[2,3]

[+] *These authors contributed equally to this study.*

[1]*Departments of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029*
[2]*Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029*
[3]*Division of Liver Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, 10029*
[4]*Department of Visceral Surgery, Lausanne University Hospital CHUV (Switzerland).*
[5]*Department of Surgery, Icahn School of Medicine at Mount Sinai, New York (US).*
[6]*Department of Pathology, Icahn School of Medicine at Mount Sinai, New York (US).*
[7]*IBM T. J. Watson Research Center, Yorktown Heights, New York, NY, USA*

***Address for correspondence:**

*Bojan Losic, PhD
Department of Genetics and Genomic Sciences
Tisch Cancer Institute, Cancer Immunology
Diabetes, Obesity and Metabolism Institute
Icahn Institute for Data Science and Genomic Technology
Icahn School of Medicine at Mount Sinai, New York, NY 10029
bojan.losic@mssm.edu

*Augusto Villanueva, MD, PhD
Division of Liver Diseases, Department of Medicine, Liver Cancer Program, Tisch Cancer Institute, Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, 1425 Madison Ave, Box 1123, Room 11-70E , New York NY 10029
augusto.villanueva@mssm.edu

ABSTRACT

There is limited understanding of the epigenetic drivers of tumor evolution in hepatocellular carcinoma (HCC). We quantify epigenetic intra-tumoral heterogeneity (ITH) using regional enhanced reduced-representation bisulfite sequencing (eRRBS) DNA methylation data from 47 early stage, treatment-naive HCC biopsies across 9 patients. Integrating these data with matching RNAseq, targeted DNA sequencing, tumor-infiltrating lymphocyte (TIL) and hepatitis-B viral (HBV) expression, we computed regional differential methylation (DM) ITH signatures across 19,327 promoter regions, and 654,133 CpG islands, while overlapping with known methylation age marker genes (240/354). We found substantial ITH signatures in promoter and enhancer sites across 4/9 patients highlighting novel molecular pathways of tumor progression not otherwise detectable from RNA analysis alone. Additionally, we identify an epigenetic tumoral aging measure that reflects a complex tumor fitness phenotype as a potential proxy for tumor evolution. In order to compute clinical associations with epigenetic tumoral age, we use 450k array data from 377 HCC patients in the TCGA-LIHC single-biopsy cohort to calculate tumoral age and find evidence implying that epigenetically old tumors have lower fitness yet higher TIL burden. Our data reveal a novel, unique epigenetic ITH axis in HCC tumors that furthers our understanding of tumor evolution and may serve as a potential avenue for enhancing patient stratification and treatment.

INTRODUCTION

Hepatocellular carcinoma (HCC) is a common yet lethal form of liver cancer that has one of the fastest growing incidence rates globally (1,2). HCC typically arises in patients with a background of chronic liver disease, such as hepatitis B or C (HBV or HCV) infection, alcohol use disorder, and nonalcoholic steatohepatitis (NASH). Despite recent improvements in systemic therapies, drug resistance remains a challenge and median survival in advanced stages is still less than two years. We and others have shown significant molecular intra-tumoral heterogeneity (ITH) in 30-40% of HCC (3–5). This includes complex mixtures of subclonal cancer populations evolving over time and interacting with non-tumoral stromal and immune cells in the microenvironment (6). This process of tumor clonal evolution in the context of the microenvironment is a key molecular feature underlying drug resistance in HCC.

To date, there have been relatively few studies that examine HCC ITH leveraging experimental designs with multi-regional sampling from the same tumor (3–7). These studies have found that increased ITH correlates with worse prognosis and clinical outcomes in patients with HCC (3). Additional studies have generally found that driver mutations in genes such as *TERT* promoter, *TP53*, and *CTNNB1* were trunk events, present in all regions of the tumor (4,7). We have recently carried out a study leveraging bulk and single-cell RNA-sequencing, SNP array, and TCR-seq profiling across multi-regional samples of 14 HCC patients, finding distinct ITH-dependent patterns of spatio-temporal interactions between HCC and immune cells along with strong evidence of spatially heterogeneous immunoediting (3). However, recent studies have also found evidence for profound immune modulation of clonal evolution and evidence of ITH arising from gene expression, indicating that HCC ITH cannot be fully characterized through mutation data alone (3,4).

It is crucial to better understand the molecular drivers of HCC ITH in order to identify new prognostic and treatment prediction biomarkers, develop therapies, and counter drug resistance. One notable yet relatively underexplored aspect of ITH in HCC is the role of epigenetics and, in particular, DNA methylation. Previous work examining the epigenetic drivers of HCC have found evidence of promoter hypermethylation of tumor suppressor genes as an important mechanism for tumorigenesis and disease progression in HCC and some of its background etiologies (8,9). We detected aberrant methylation patterns in HCC that can be distilled into signatures predictive of patient prognosis and survival, indicating a potentially useful avenue for biomarker discovery (8). A previous study examined DNA methylation using multi-regional sampling in a small cohort of HCCs and reported significant epigenetic heterogeneity among tumors, in contrast to strong interindividual homogeneity of nonmalignant liver tissue. Numerous studies in single-biopsy and ITH contexts have tied aberrant DNA methylation to cancer-related processes involved with HCC tumor oncogenesis, proliferation, and invasiveness (4,8–10).

However, the majority of methylation work in HCC and HCC ITH uses array-based Illumina 450k data, therefore limiting the discovery of epigenetic ITH arising from methylation in novel genomic loci such as enhancers and CpG open shelf, shore, and open sea positions (11–13). A key aspect of methylation data is its ability to predict molecular epigenetic age, which measures proxies of

biological age of a tissue as opposed to its chronological age (14). These so-called epigenetic clocks have been extensively validated in normal tissue and used to examine the biological aging process and its relation to human diseases, including cancer (14–16). However, the role of methylation age in HCC and its evolutionary process has yet to be established.

In order to build on our prior work and further investigate the unique contributions of DNA methylation to HCC evolution and ITH, here we use enhanced reduced representation bisulfite sequencing (eRRBS) to obtain methylation profiles from multiregional biopsies of 9 patients from our previous ITH dataset (3). By adding another layer of methylation data to existing extensive molecular ITH profiles, we explicitly compare ITH signatures and molecular pathways derived from RNA expression and DNA methylation data and find key differences. In particular, we leverage well-known associations between conserved methylation signatures and chronological patient age to derive estimates of differentially aging tumoral regions that reflect novel aspects of tumor evolution and heterogeneity not captured by expression profiling alone (14–16). Integrating the TCGA-LIHC (HCC) cohort with our data, we find a differential tumoral aging profile that strongly associates epigenetically younger TCGA-LIHC tumors with greater DNA clonal diversity, tumor mutation burden (TMB), tumor fitness (poorer patient prognosis), and lower tumor infiltrating lymphocyte (TIL) burden. The differential regional aging of tumors is a novel axis of epigenetic ITH (eITH) which may reflect a complex tumor fitness phenotype, serving as a useful proxy for tumoral evolution and ultimately treatment response stratification of patients.

RESULTS

Spatial tumor methylation profiling via multi-regional sampling

We performed enhanced reduced representation bisulfite sequencing (eRRBS) (17) on 47 samples from 9 HCC patients, including tumor (35 samples) and adjacent non-tumoral regions (12 samples, **Fig 1A**). On average, we analyzed a median of 5 samples per patient. All patients were treatment-naive at the time of the surgical resection. Most patients were male (67%) with a median age of 61 years, median tumor size of 6.5 cm, and about 55% (5/9) had hepatitis B infection as the most common underlying liver disease (**Fig 1B**). Drawing on our previous characterization of these samples (3), tumor purity was determined using SNP-array data via ASCAT (version 2.4) and histological examination.

Since eRRBS data offers a granular view of methylation status at specific CpG sites within and even outside promoter regions, we designed our regional methylation analysis to take advantage of both promoter-specific and site-specific sensitivity. First, we performed a site-specific analysis considering only well-covered CpG sites regardless of proximity to a transcriptional start site (TSS). Second, we aggregated sites within some proximal region of a TSS and conducted a gene-wise promoter-based analysis that facilitates a reasonable comparison to both 450k methylation array data and its derived methylation age, as well as bulk RNA-seq transcriptional profiling (**Fig. 1C**).

Regional epigenetic ITH (eITH) reveals novel differential methylation associations

Using the normalized site-specific methylation data, we performed a principal component analysis (PCA) analysis in order to quantify the leading axes of variation driving patient and regional specific DNA methylation signatures (**Fig. 2A**). While the first two principal components collectively explain only 26% of the total variance, the leading principal component separates tumor from adjacent normal samples and broadly stratifies tumors by degree of tumor-infiltrating lymphocytes (TIL), defined by VDJ read burden normalized for library size from RNA-seq. While there is strong evidence of patient-specific clustering, there are also outlier regions which demonstrate that the relative scale of eITH can overwhelm patient-specific variance. The second principal component also reveals that patient 8 is a strong outlier. All of these findings also hold using the promoter-specific methylation data (**Fig 2A, Supp Fig 1A**).

In order to directly derive multi-regional signatures reflecting epigenetic ITH (eITH) across patients' tumors, we carried out within-patient differential methylation analyses. Patients 2, 3, and 10 displayed significant eITH (**Fig 2B, Fig 3A**), while patient 4 revealed the highest ITH with region H4.c having the highest relative number of differentially methylated promoters (1993 DM promoters, FDR < 0.05). Patient 6, previously described as being heavily immune infiltrated (3), displayed only comparatively minor eITH DM (126 DM promoter regions, FDR < 0.05, **Fig 3A**) despite extensive immune heterogeneity across the different tumor regions found through TCR sequencing (3). Although increased sampling rate (number of biopsies per tumor) increases power to detect eITH, we noticed that the most densely sampled patient (P9, $n_{biopsies} = 6$), had very few eITH DM loci (**Fig 2B, Fig 3A**). This suggests that our average sampling rate of 5 biopsies per tumor was sufficient to capture the leading effect of eITH. Altogether, these data

suggest significant eITH in 4 out of 9 patients, and multiple, separate axes of eITH that are in part described by DM profiling and principal component analysis.

In order to compare regional eITH signatures with those associated with HBV viral expression, tumor-infiltrating-lymphocyte (TIL) burden, and HCC vs adjacent normal, we leveraged our previously reported transciptome profiling data. We identified site and promoter level eITH DM signatures driven by TIL, tissue type (normal vs HCC tumor), and HBV viral expression (**Fig 2B, Supp Fig 1B**). As expected, the HCC tumor-specific signature had the most DM loci (184,504 sites and 5,458 promoters, FDR < 0.05; **Fig 2B**, **Supp Fig 1B**), followed by TIL-related sites and promoters, with the HBV-driven signatures having the fewest sites or promoters (**Fig 2B, Supp Fig 1B**). In all global DM signatures, there is a bias towards hypomethylation (**Fig 2B, Supp Fig 1B**). We identified overlapping significant DM sites among global and within patient regional eITH DM signatures in order to examine the contribution of general cancer or immune specific methylation patterns to regional methylation gradients. Overall, we observed unexpectedly little overlap in DM sites among global and regional comparisons (**Fig 2C**) and among eITH DM profiles between patients. This suggests that eITH DM profiles are not only unique on a per-patient basis, but may also be uniquely independent of general tumor-or immune-driven transcriptional regulation programs.

<u>eITH signature is not a simple recapitulation of transcriptomic profiling</u>

Since ITH is a complex manifestation of cancer evolution, we combined DM promoters with differentially expressed (DE) genes from RNAseq data already available from these samples to gain an integrative view of the data (3). DM promoters and DE genes had zero overlap for 32 regional comparisons (i.e, the majority of the intra-tumor regional comparisons) across 9 patients. A handful of regional comparisons, which belong to patients 2, 3, 4, and 5, had low overlap relative to the DNA methylation (max = 34.06%, median = 17.07%) and to the RNA (max = 10%, median = 1.590%) (**Fig 3B**). Sample sizes from gene overlap only allowed us to correlate DNA methylation and gene expression log fold changes in three contrasts, with no significant correlation in two (H4.c and H4.a&c) and a moderate negative correlation in the third (H3.a; Spearman rho = -0.409, P = 4.5 e-03). Taken together, these observations suggest that the eITH described by DM profiling offers a unique signal that may not be captured by gene expression information alone.

<u>Tumoral hyper-aging and inter-regional heterogeneity predicted by methylation clock</u>

To better quantify how eITH relates to tumor clonal evolution, we calculated the predicted methylation age of the tissue across tumor samples and adjacent normal tissue using the methylation clock described by Horvath et al (14–16). Of the 353 CpG sites used to inform Horvath's model only 12 sites were covered with reasonable depth across sites in the eRRBS sequencing data (**Fig 1C**), making it difficult to compute age based on these data. However, the promoter-aggregated site analysis provided coverage overlap at the gene level of 240 of the 353 genes from the methylation clock mode (**Fig 1C**), allowing us to predict tissue methylation age across samples at the gene level.

To visualize how eITH might be driving differences in methylation age and in turn reflect a facet of clonal evolution, we compared the relative tissue methylation age across tissue types for patients 2,4,8,9 and 10, each of which had three or more regionally sampled tumor sites and sampled adjacent normal tissue (**Fig 4A**). While for all but patient 8 we found wide variance in the predicted methylation ages across regional samples within tumor, average relative tumoral methylation age was more advanced than the predicted methylation age of the adjacent normal liver tissue for all patients. Tumoral relative hyper-aging was greatest in patient 8, where average tumoral methylation age was 1.35 times greater than that of the average predicted age of the adjacent liver tissue (**Fig 4B**), with a median tumoral hyper-aging ratio of 0.17 across patients. Tumor-normal blood relative methylation age ratios calculated for two of three additional multi-regionally samples HCC patients (4) validated this trend, though the remaining patient had a significant hyper-aging signature of the normal tissue compared to the tumor (**Supp Fig. 3**) (4).

The ratio of predicted sample methylation age to true patient age was used to determine a relative tissue DNA methylation aging factor. This ratio was significantly positively correlated with the second principal component of a promoter-aggregated site methylation PCA across regional samples (**Fig 4C;** Spearman rho = 0.5, P = 0.002), with the tumoral samples from patient 8 demonstrating the largest hyper-aging effect compared to the true patient age. Methylation clock age was also significantly correlated with PC2 (**Fig 4C;** Spearman rho=0.31, P = 0.001), but to a lesser degree. Neither measure correlated with other sample clinical covariates. We also observed a negative correlation between methylation clock age and VDJ read burden (**Fig 4C;** Spearman rho = -0.35, P = 2.52e-04). These data indicate that tumoral hyper-aging is a potentially useful quantitative proxy for eITH in terms of a direct facet of tumor evolution, and that it helps uncover a key axis of variation (PC2) in our multi-regional dataset.

<u>Tumoral hyper-aging associates with reduced tumor clonality, decreased tumor fitness, and "hotter", less clonal tumor microenvironment.</u>

We applied our methylation-age analysis to 450k methylation array data for 366 single biopsy HCC samples from the TCGA-LIHC cohort in an attempt to not only validate our observation of tumor hyper-aging, but also establish the association with patient survival and known clinical biomarkers for immune activity such as TIL and tumor mutational burden (TMB). We observed advanced relative tumor methylation-age in 260/366 (71%) HCC patients in the cohort (ratio of predicted tumoral methylation age to patient diagnosis age; mean = 1.22, median =1.15; **Fig 5A**). Kaplan-Meier survival estimates were calculated for patients organized by tumor methylation clock age into "old" and "young" groups, defined by the median tumor aging factor. Patients with "old" methylation-aged tumors saw significant survival benefit over patients with "young" tumors (HR=0.667, p=0.037; **Fig 5B**).

To investigate whether the relative tumor methylation age remains a significant predictor of patient survival after regressing out known clinical covariates, we constructed a multivariable Cox proportional hazard model using patient TMB, sex, tumor stage, and relative tumor age factor (**Supp. Table 6**). A first order cubic spline was applied to the relative age factor to test for non-linear effects on survival. We found relative tumor methylation age to have a significant positive effect on patient survival (p=0.04), with no evidence for a significant non-linear effect. To

demonstrate the improvement in survival prediction error using relative tumor methylation age, we compared the time dependent integrated Brier scores for cph models using the clinical covariates with and without the relative methylation age factor (**Fig 5C**). Comparing the resulting prediction error curves under bootstrap resampling (see Methods) we found that inclusion of relative methylation age factor reduces survival prediction error compared to the model that does not include the measure (**Fig 5C, inset table**). Computing Brier scores under cross-validation indicates these results are robust and suggests that these results will favorably generalize to other datasets.

As we found advanced relative tumor DNA methylation age associated with improved patient survival (decreased tumor fitness) in the TCGA, we investigated its association with other key tumoral and immune molecular features. Performing VDJ-deconvolution RNA sequencing data from TCGA HCC samples matched with 450k array data (18), we observed TIL burden and immune clonality to be significantly higher amongst “old” relative tumor methylation age HCCs compared to “young” tumors ($p=0.016$, $p=0.046$; **Fig 5D-E**), and PD-1 receptor expression, typically associated with T cell influx, was similarly significantly upregulated in “old” tumors ($p<0.01$; **Fig 5G**). Interestingly, using the TCGA HCC DNA-sequencing data to determine and cluster somatic mutations into clones, “old” HCCs were also found to have significantly fewer tumor clones compared to “young” tumors ($p<0.01$; **Fig 5H**), and tended to have lower TMBs ($p=0.024$, **Fig 5F**). Together, these results bring further context to our earlier observations of eITH by directly relating tumoral hyper-aging to reduced clonal complexity, hotter and more clonal TIL burden, and decreased tumor fitness, in a much larger cohort of single-biopsy HCC samples from the TCGA.

DISCUSSION

DNA methylation alterations in cancer are often difficult to systematically interpret outside of well-studied examples such as promoter hypermethylation of tumor suppressor genes and hypo-methylation of repeat-rich regions. These exceptions define the rule of poor overall correlation of gene expression and promoter methylation in cancer, which we also dramatically confirm in the context of ITH (Figure 3). Indeed, there are few studies that have evaluated eITH in cancer. A recent comprehensive characterization of eITH in lung cancer found that ITH mapping to tumor suppressors was lower than that of oncogenes. The authors suggested a greater selection pressure in these regions with lower methylation ITH (19). To date, we have found only one other study using multiregional sampling and DNA methylation to assess eITH in HCC (4). We confirm the authors' prior findings on the prognostic value of eITH, and expand upon it in this analysis. By leveraging multiregional eRRBS methylation profiling on 47 samples from 9 patients, we provide a characterization of spatial differential methylation gradients as a novel, relatively unexplored axis of ITH. We characterized these eITH gradients and found they are unique, not fully recapitulated by RNA expression, and are distinct from global cancer or immune expression programs. Further, we report a novel property of eITH driven by tumoral hyper-aging that may capture a complex tumor fitness phenotype with clinical prognostic value. To our knowledge, this work provides the most comprehensive analysis of epigenetic ITH in HCC reported to date.

Regional differential methylation underlying hyper-aging tends to implicate promoters belonging to genes which are not differentially expressed in those same regional comparisons. Moreover, eITH patterns are both very specific to each patient and have no significant overlap between each other (intra-patient FDR adjusted over-enrichment Fisher p-value < 0.05). This indicates highly specific epigenetic reprogramming can characterize tumor evolution. Even so, we observed that patients with marked eITH, such as patients 4 and 5, had strong over-enrichment of their eITH signatures across global DM signatures associated with TIL burden and HCC vs. adjacent normal liver. This contrasts with patients exhibiting minimal eITH such as patients 7 and 10, who were significantly under-enriched in these contrasts. This underscores that eITH is unraveling an orthogonal axis of tumoral evolution that is normally not specifically captured in a single-biopsy context. It also provides a rich feature space from which to mine potential biomarkers of tumoral evolution and treatment response in HCC. To this end, we note that given the relative granularity of eRRBS data compared to array-based DNA methylation data, our spatial multi-regional eITH data will add a number of previously uncharacterized eITH-specific loci for potential downstream prioritization in other applications such as liquid biopsies.

Using well-known epigenetic clocks, we show that tumoral hyper-aging, which in principle may be biologically related to a higher mitotic rate in tumor cells, is a useful, novel metric for describing tumor evolution and eITH in both the multi-regional and single-biopsy context. The chronological alteration of epigenetic profiles via progressive DNA methylation has been extensively established in a number of normal tissues (16,20), however its utility as a proxy of eITH for malignancies has yet to be fully explored. From the point of view of multi-regional sampling, we show that eITH effectively describes differential epigenetic aging for spatially distal regions in HCC relative to the adjacent normal aging rate. In other words, different parts of the same tumor are epigenetically aging at different rates, and on average these rates are all higher than that of adjacent non-HCC liver tissue. Even from the under-sampled point of view of single-biopsy data, as in the TCGA

LIHC cohort, we show that on average even a single HCC biopsy is older than the diagnosis age of the patient. In our previous work, we established that a key facet of ITH was immune-editing resulting from tumor-immune interactions which can effectively constrain tumoral evolution (3). Here, we report an association of tumoral hyper-aging with a “hotter” immune microenvironment and decreased clonality. This finding not only corroborates recent evidence showing that DNA methylation loss supports immune evasion (21), it demonstrates this concept in the context of tumor clonality and may suggest this mechanism could play a larger role in ITH than previously thought.

Despite being among the largest HCC cohorts analyzed for eITH to our knowledge, our study is limited by its relatively small sample size, a lack of sufficiently powered publicly available multi-regional omics datasets for validation, and the absence of a robustly trained epigenetic clock specific for HCC. Our study has characterized an additional layer of molecular intra-tumoral heterogeneity, finding that significant regional differential DNA methylation patterns in patient HCC tumors are unique and poorly recapitulated in standard pooled or single-biopsy profiles. We also report on accelerated epigenetic aging on a regional basis within tumors which serves as a new proxy for tumor evolution. We envision that these data will provide new insights into characterizing epigenetically driven tumor evolution and provide a framework of analysis for uncovering potential new biomarkers of treatment resistance.

METHODS

Sample collection

Patient recruitment and sample collection was performed as previously described (3). Briefly, patients were enrolled in the study at Icahn School of Medicine at Mount Sinai (ISMMS) and provided informed consent for tissue biobanking. The study was approved by the Mount Sinai IRB (IRB# HS-14-01011) and samples were provided by the ISMMS Tissue Biorepository (IRB# HS-10-00135). All patients had early stage hepatocellular carcinoma (HCC) as per EASL guidelines, and were treatment-naïve prior to surgical resection (17). Frozen tissue samples from the same tumor nodule were collected allowing for at least 1 cm of distance between each other. Samples were selected from areas without macroscopic evidence of necrosis or hemorrhage.

eRRBS sequencing and processing

DNA was extracted using the DNeasy blood and tissue kit (Qiagen) from 30mg tissue following manufacturers protocol. RRBS sequencing was performed by the Epigenomics Core facility at Weill Cornell Medical College using an in-house developed protocol consisting of restriction enzyme digestion for enrichment of CpG sites, NGS library construction and bisulfite conversion of cytosines (17). 50bp single-read libraries were sequenced using the Illumina HiSeq 2500 platform in high output mode.

Raw eRRBS reads were trimmed using trim galore (version 0.5.0) with the '-rrbs' flag enabled, and quality control metrics were compiled using fastqc (v 0.5.0). Reads were then aligned to the hg38 UCSC reference genome using bismark (v 0.22.3), bowtie2 (v 2.4.1), and samtools (v 1.11) with the options to retain unmapped and ambiguously mapped reads enabled. A count matrix with methylated and unmethylated counts for each sample was created using bismark's methylation extractor tool (v 0.22.3).

Differential methylation analysis

Differential methylation signatures were computed with edgeR, generally following the procedures outlined in Chen et al (22). We used the generalized linear model (GLM) framework in edgeR because it allows for great flexibility and robustness in analyzing our complex experimental design while being able to account for technical and biological covariates.

The raw counts matrix for the eRRBS data was filtered for two analysis pipelines: one that looked at individual CpG island sites (site-based), and one that aggregated the averaged counts over promoter regions (promoter-based).

In the site-based analysis, CpG islands with a minimum coverage of 10 reads across all samples were retained. In the promoter-based analysis, counts for sites that fall within the same promoter region (within 2kb upstream or 1kb downstream of the nearest transcription start site (TSS) annotation per edgeR's nearestTSS function) were summed (22,23). Promoters with at least 10 aggregated reads across all samples were retained for further analysis.

Site-based and promoter-based DM signatures were computed on a within-patient, regional-relative basis, and on a global basis. In all analyses, sites or promoters from mitochondrial, sex, and unannotated chromosomes were removed from analysis. Total library size for all samples was set as the sum of methylated and unmethylated counts. All dispersion estimates assumed no mean-variance trend.

Regional-relative DM signatures

Within-patient regionally-relative DM signatures were computed using pooled sample comparisons in an effort to directly sample epigenetic ITH while compensating for lack of statistical power. Briefly, a generalized linear model using edgeR's eRRBS pipeline was created for each of the 39 within-patient comparisons across all 9 patients (22). HBV viral expression (measured in RPKM) and normalized VDJ read burden determined via RNA sequencing as previously described were included as covariates in the model (3,22,23). For each comparison, a separate, paired design matrix that accounts for methylation status in the counts matrix was used. Differential methylation was assessed using a likelihood ratio test via edgeR.

Global DM signatures

Differential methylation signatures using all tumor and adjacent samples over all patients were computed with generalized linear models using edgeR's eRRBS pipeline. The design matrix accounted for HBV viral expression (measured in RPKM) and normalized VDJ read burden determined as previously described, and multiple samples from the same patient. Differential methylation along three major axes was assessed with a likelihood ratio test: normal vs tumor, VDJ read burden (as a proxy measure for TIL), and HBV viral expression.

RNA-seq differential expression analysis and integration with DM promoters

In order to compare expression-based ITH with DM promoter-based eITH, matched RNA sequencing for the MSSM cohort was obtained and processed as previously described (3). In order to ensure a fair comparison between the RNAseq and the eRRBS, generalized linear models for DE analysis were generated using the edgeR pipeline and the same covariates as that used in the DM analysis where reasonably applicable (23). DE testing was performed using a likelihood ratio test in edgeR for the same regional-relative and global comparisons used in DM promoter testing as described above.

Methylation age prediction

Methylation based tumor age was calculated for the HCC regional samples using the CpG-site based age clock described by Horvath et al (14–16). As site-level coverage of the 354 CpG-sites used to determine methylation age was low in the eRRBS data, encompassing genes for the Horvath CpG-sites were identified and beta methylation normalized values for overlapping genes from the promoter-based analysis were used as proxy. Methylation age was only computed for patients with regional sampling of at least one or more normal sites and three or more tumor sites (patients 2, 4, 8, 9, and 10), such that age predictions could be compared intratumorally and between tissue types. Patient age was transformed via logit function as described previously (14).

Median computed tumor and normal regional sampling ages for each patient were contrasted to calculate the patient normalized relative aging factor ratio (**Fig 4A-C; Supp Fig 3**).

Survival Analysis

Kaplan-Meier curves and risk tables grouping patients along the median tumor methylation age factor into "old" and "young" cohorts were created using the survminer R package (24). Patients with missing survival times or model feature data were removed before modeling. The log-rank test was used to evaluate the difference in survival outcomes between groups, with the significance and confidence intervals reported for each comparison.

Multivariate cox proportional hazard (cph) models using tumor methylation age, tumor clinical stage, sex, and patient tumor mutation burden (TMB), were constructed using the 'cph' function from the rms package (25). The relative predictive power amongst survival models and to the reference model was assessed by plotting prediction error curves with the pec package in R (26). Integrated Breier scores for each model were computed with .632+ bootstrap resampling across 100 iterations (B=100) and compared, with the weights of each score corresponding to the inverse likelihood of being censored, with censoring times for each iteration estimated using the Kaplan-Meier estimator. Brier scores for each model were compared at 75 months with significance by KS-test calculated and included in the inset figure tables, corresponding to 90% of patient events.

Validation of clock-associated tumoral hyper-aging and outcome analysis in external datasets

Two additional Illumina 450k methylation array HCC datasets were used for validation of methylation-based tumor aging and patient survival findings.

To demonstrate the tumor hypermethylation and regional methylation patterns seen in the main dataset, we retrieved beta-normalized illumina 450k methylation array data from the GEO database project GSE83691 for 17 samples from multi-regionally sampled HCC liver biopsies for three patients (*HCC8010, HCC6952*, and *HCC8257*) was retrieved (4). These three patients were selected due to the availability of matched normal data from circulating blood. Methylation based age predictions were made for each sample and relative aging factor ratios between median tumor ages and normal age were calculated for each patient.

To demonstrate the effects of tumor methylation age on patient survival, Illumina 450k methylation array data for 376 tumor samples from the TCGA liver cancer cohort (LIHC) was retrieved from the GDC portal (https://www.cancer.gov/tcga). Clinical phenotype and survival data for these patients was also retrieved. TMB information was available for 362 samples, and tumor clonality was assessed using SciClone in 192 samples where appropriate matched mutation data was available (27). For 188 samples where matched RNA-seq data was available, the estimated VDJ read burden and TIL clonality was assessed using MiXCR, and logCPM expression values for PDCD1 were calculated (28). Normalized methylation beta values were calculated for each site, and tumor methylation-based ages were predicted using Horvath's clock. Relative tumoral age factors were calculated for each patient using the ratio of the predicted tumor methylation age to the patient's age at diagnosis. Patients were separated into 'old' and 'young' tumor groups based on median age factor, and survival analyses were carried out as described above. Tumor

clonality, VDJ burden, TMB, and PDCD1 expression was statistically compared between tumor groups using a wilcoxon test in R.

SUPPLEMENT

Enhancer overlap with differential methylation signatures

We overlapped DM sites with known enhancer regions reported in the HACER database (29) and identified sites residing in putative promoter regions defined as within 2Kbp downstream or 1Kbp upstream of a transcription start site (TSS). While most DM sites fall into promoter regions they do also overlap known enhancers and there are also a number of DM sites whose genomic coordinates overlap with both a known enhancer region and a putative gene promoter (i.e., dual overlap). Furthermore, we find an aberrant, global TIL-associated bias towards hypermethylation in enhancers, and a less pronounced hypermethylation in dual-overlap sites (**Supp Fig 2A)**. In within-patient, regional-relative eITH DM sites, we find a significant hypermethylation bias in enhancers or dual-overlap sites in 7 comparisons across 4 patients (**Supp Fig 2B**).

DATA AND CODE AVAILABILITY

RNA sequencing data is deposited in ArrayExpress under the accession E-MTAB-5905. Reduced-representation sequencing data is deposited in SRA under the accession PRJNA662974. Sequencing data from patient 5 are not deposited due to lack of patient-specific deposition consent. TCGA-LIHC data were generated and made available by the TCGA Research Network (https://www.cancer.gov/tcga) and were downloaded from the National Cancer Institute's GDC Data Portal (https://portal.gdc.cancer.gov/). Sequencing data from the Lin et al 2017 study was obtained from GEO under the accession GSE83691.

ETHICS STATEMENT

All tumor samples used in this study were collected from consented patients enrolled in the Icahn School of Medicine at Mount Sinai institutional IRB-approved cancer biorepository tissue procurement protocol.

FUNDING

This work was funded by the Genetics and Genomics department, Icahn School of Medicine at Mount Sinai, as well as the Icahn Institute for Data Science and Genomic Technology. AV is supported by the U.S. Department of Defense (CA150272P3).

AUTHOR CONTRIBUTIONS

B.L. and A.V. conceived the study concept and experimental design. A.J.C was responsible for sample collection and preparation. GS provided funding and scientific oversight. BL conceived the hyper-aging description of eITH and carried out the initial data analyses. PR and AB analyzed the data under supervision from BL and AV. PR, AB, BL, and AV wrote the manuscript.

ACKNOWLEDGEMENTS

The authors thank the office of Scientific Computing and the Genomics Core Facility at the Icahn School of Medicine at Mount Sinai (ISMMS) for providing computational resources and staff expertise, as well as the ISMMS Tissue Biorepository for providing the samples.

DISCLOSURES

AV receives consulting fees from Guidepoint, Fujifilm, Boehringer Ingelheim, FirstWord, and MHLife Sciences; advisory board fees from Exact Sciences, Nucleix, Gilead and NGM Pharmaceuticals; and research support from Eisai.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68:394–424.

2. Villanueva A. Hepatocellular Carcinoma. New England Journal of Medicine. Massachusetts Medical Society; 2019;380:1450–62.

3. Losic B, Craig AJ, Villacorta-Martin C, Martins-Filho SN, Akers N, Chen X, et al. Intratumoral heterogeneity and clonal evolution in liver cancer. Nature Communications. Nature Publishing Group; 2020;11:291.

4. Lin D-C, Mayakonda A, Dinh HQ, Huang P, Lin L, Liu X, et al. Genomic and Epigenomic Heterogeneity of Hepatocellular Carcinoma. Cancer Res. American Association for Cancer Research; 2017;77:2255–65.

5. Zhai W, Lim TK-H, Zhang T, Phang S-T, Tiang Z, Guan P, et al. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. Nature Communications. Nature Publishing Group; 2017;8:4565.

6. Craig AJ, Felden J von, Garcia-Lezana T, Sarcognato S, Villanueva A. Tumour evolution in hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol. 2019;1–14.

7. Friemel J, Rechsteiner M, Frick L, Böhm F, Struckmann K, Egger M, et al. Intratumor Heterogeneity in Hepatocellular Carcinoma. Clin Cancer Res. American Association for Cancer Research; 2015;21:1951–61.

8. 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;61:1945–56.

9. Tischoff I, Tannapfel A. DNA methylation in hepatocellular carcinoma. World J Gastroenterol. 2008;14:1741–8.

10. Song M-A, Tiirikainen M, Kwee S, Okimoto G, Yu H, Wong LL. Elucidating the Landscape of Aberrant DNA Methylation in Hepatocellular Carcinoma. Corvalan AH, editor. PLoS ONE. 2013;8:e55761.

11. Xiong L, Wu F, Wu Q, Xu L, Cheung OK, Kang W, et al. Aberrant enhancer hypomethylation contributes to hepatic carcinogenesis through global transcriptional reprogramming. Nature Communications. Nature Publishing Group; 2019;10:335.

12. Hamdan FH, Johnsen SA. Perturbing Enhancer Activity in Cancer Therapy. Cancers. Multidisciplinary Digital Publishing Institute; 2019;11:634.

13. Tsang FH-C, Law C-T, Tang T-CC, Cheng CL-H, Chin DW-C, Tam W-SV, et al. Aberrant Super-Enhancer Landscape in Human Hepatocellular Carcinoma. Hepatology. 2019;69:2502–17.

14. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:1–20.

15. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics. Nature Publishing Group; 2018;19:371–84.

16. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biology. 2019;20:249.

17. Garrett-Bakelman FE, Sheridan CK, Kacmarczyk TJ, Ishii J, Betel D, Alonso A, et al. Enhanced Reduced Representation Bisulfite Sequencing for Assessment of DNA Methylation at Base Pair Resolution. JoVE. 2015;52246.

18. Wheeler DA, Roberts LR. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327-1341.e23.

19. Hua X, Zhao W, Pesatori AC, Consonni D, Caporaso NE, Zhang T, et al. Genetic and epigenetic intratumor heterogeneity impacts prognosis of lung adenocarcinoma. Nat Commun [Internet]. 2020 [cited 2021 Feb 1];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7235245/

20. Yu M, Hazelton WD, Luebeck GE, Grady WM. Epigenetic Aging: More Than Just a Clock When It Comes to Cancer. Cancer Res. American Association for Cancer Research; 2020;80:367–74.

21. Jung H, Kim HS, Kim JY, Sun J-M, Ahn JS, Ahn M-J, et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nat Commun. 2019;10:4278.

22. Chen Y, Pal B, Visvader JE, Smyth GK. Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. F1000Res. 2017;6:2055.

23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

24. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using “ggplot2” [Internet]. 2020. Available from: https://CRAN.R-project.org/package=survminer

25. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis [Internet]. Cham: Springer International Publishing; 2015 [cited 2020 Aug 4]. Available from: http://link.springer.com/10.1007/978-3-319-19425-7

26. Mogensen U, Ishwaran H, Gerds T. Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. Journal of Statistical Software, Articles. 2012;50:1–23.

27. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution. PLOS Computational Biology. 2014;10:e1003665.

28. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12:380–1.

29. Wang J, Dai X, Berry LD, Cogan JD, Liu Q, Shyr Y. HACER: an atlas of human active enhancers to interpret regulatory variants. Nucleic Acids Res. 2019;47:D106–12.

FIGURE LEGENDS

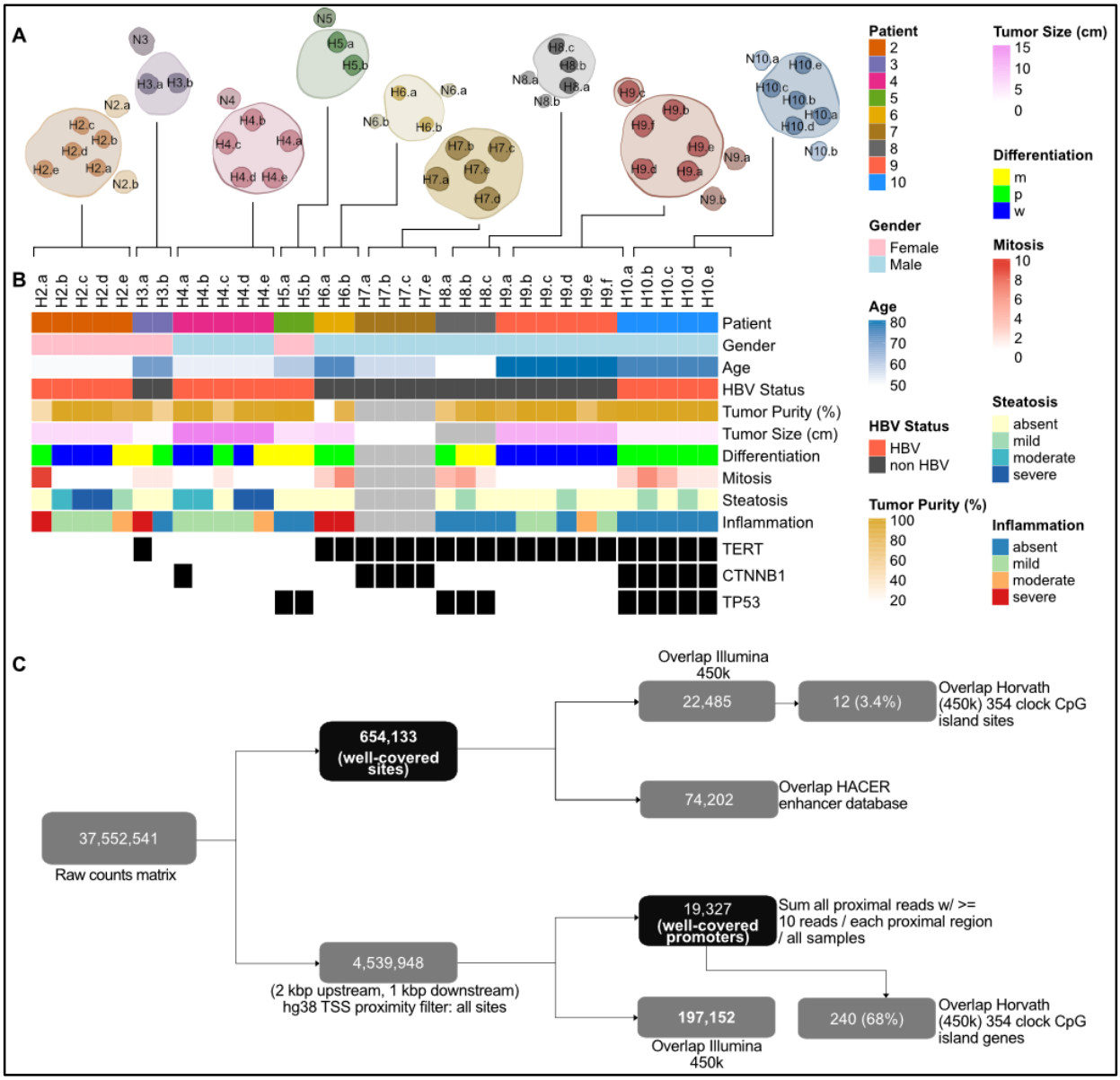


Figure 1. A) Representation of multi-regional sampling for tumors used in this study, and available data types used. B) Heatmap showing clinical features of samples, including TERT, CTNNB1, and TP53 mutation status. C) Flow chart indicating eRRBS processing and analysis framework.

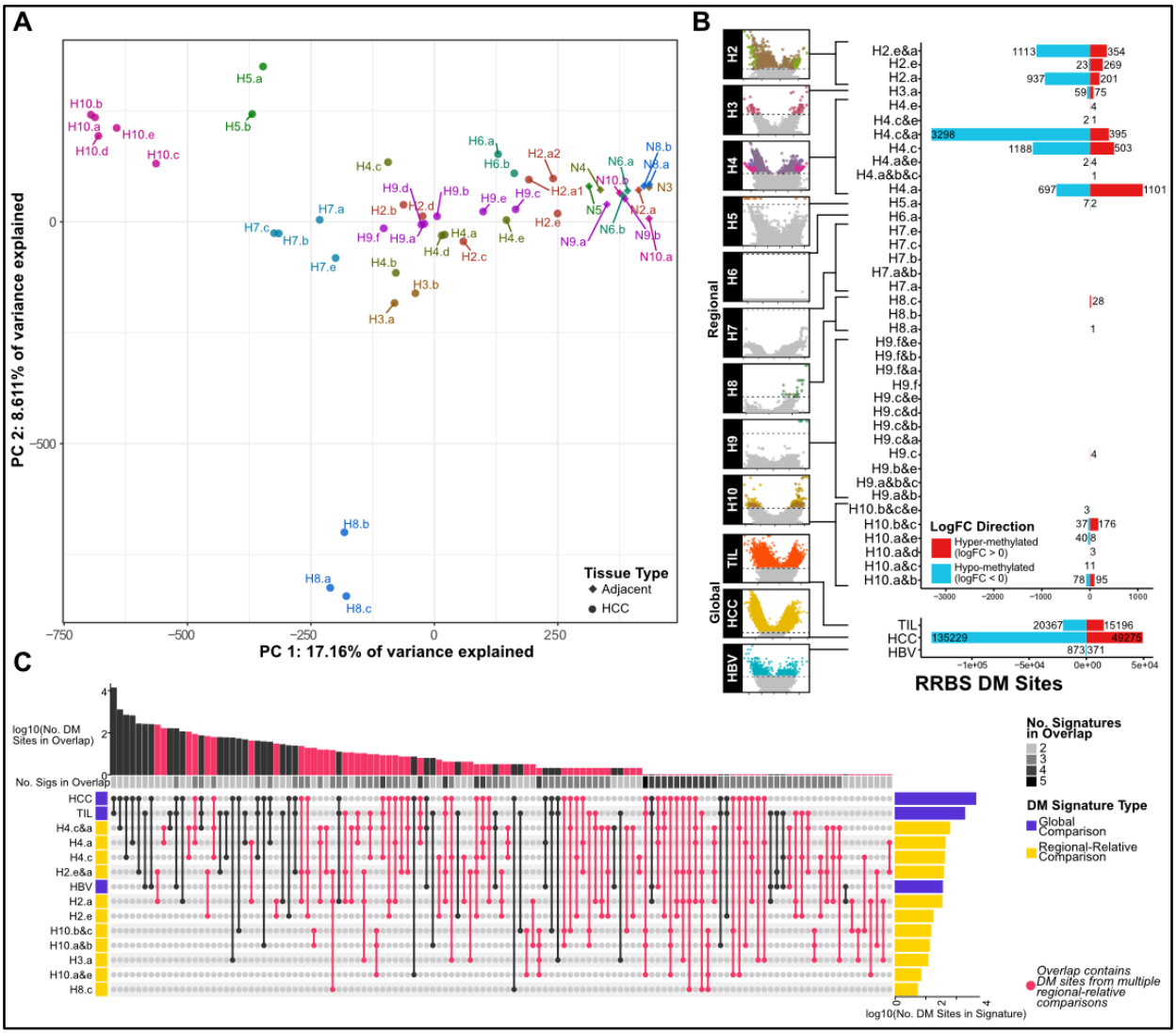


Figure 2. Differentially methylated sites. A) PCA of beta values. B) DM Sites in regional-relative and global comparisons. C) UpSet plot shows little overlap in significant DM sites across comparisons.

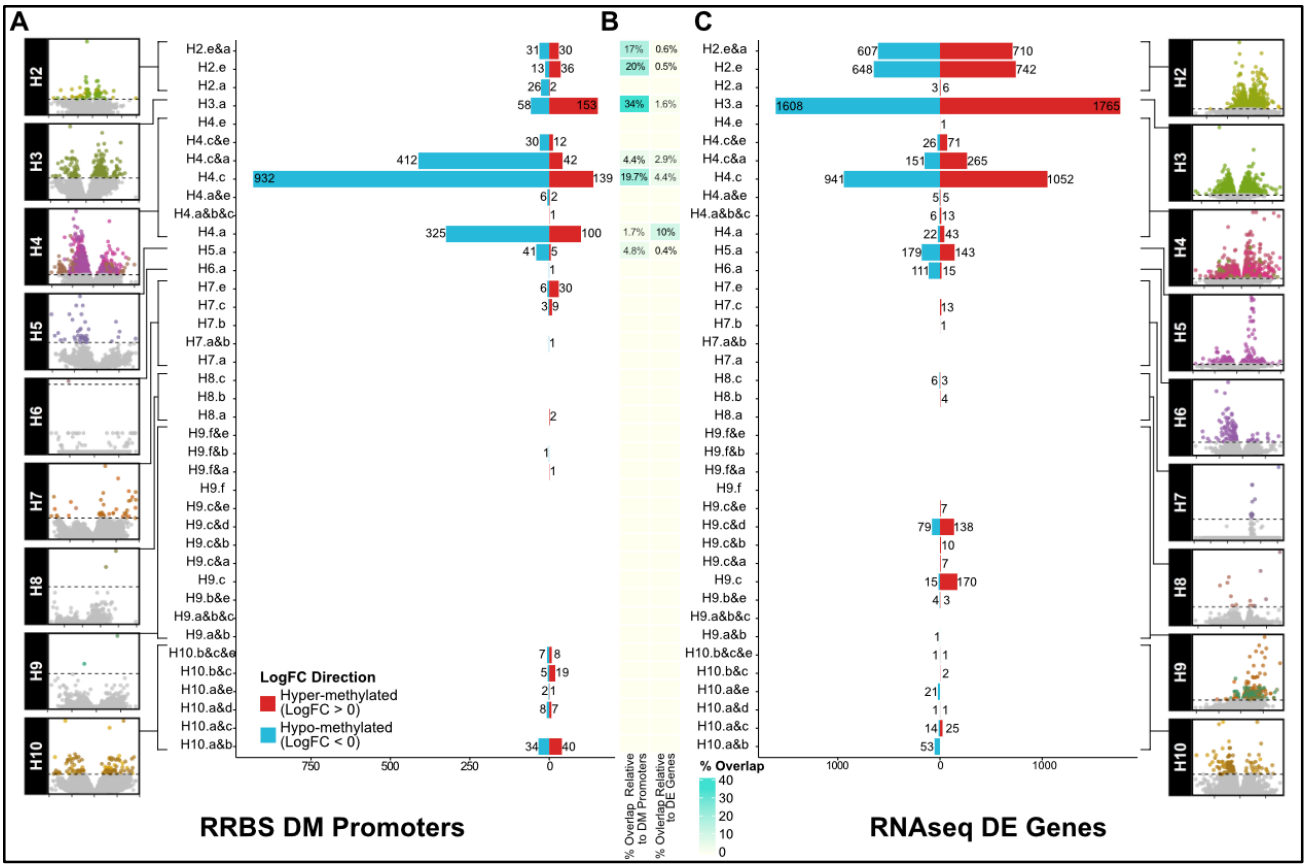


Figure 3. A) Regional-relative DM promoter profiling. B) Low overlap between significant DM promoters from eRRBS and significant DE genes in matched RNA seq. C) Regional DE gene profiling via RNA-seq.

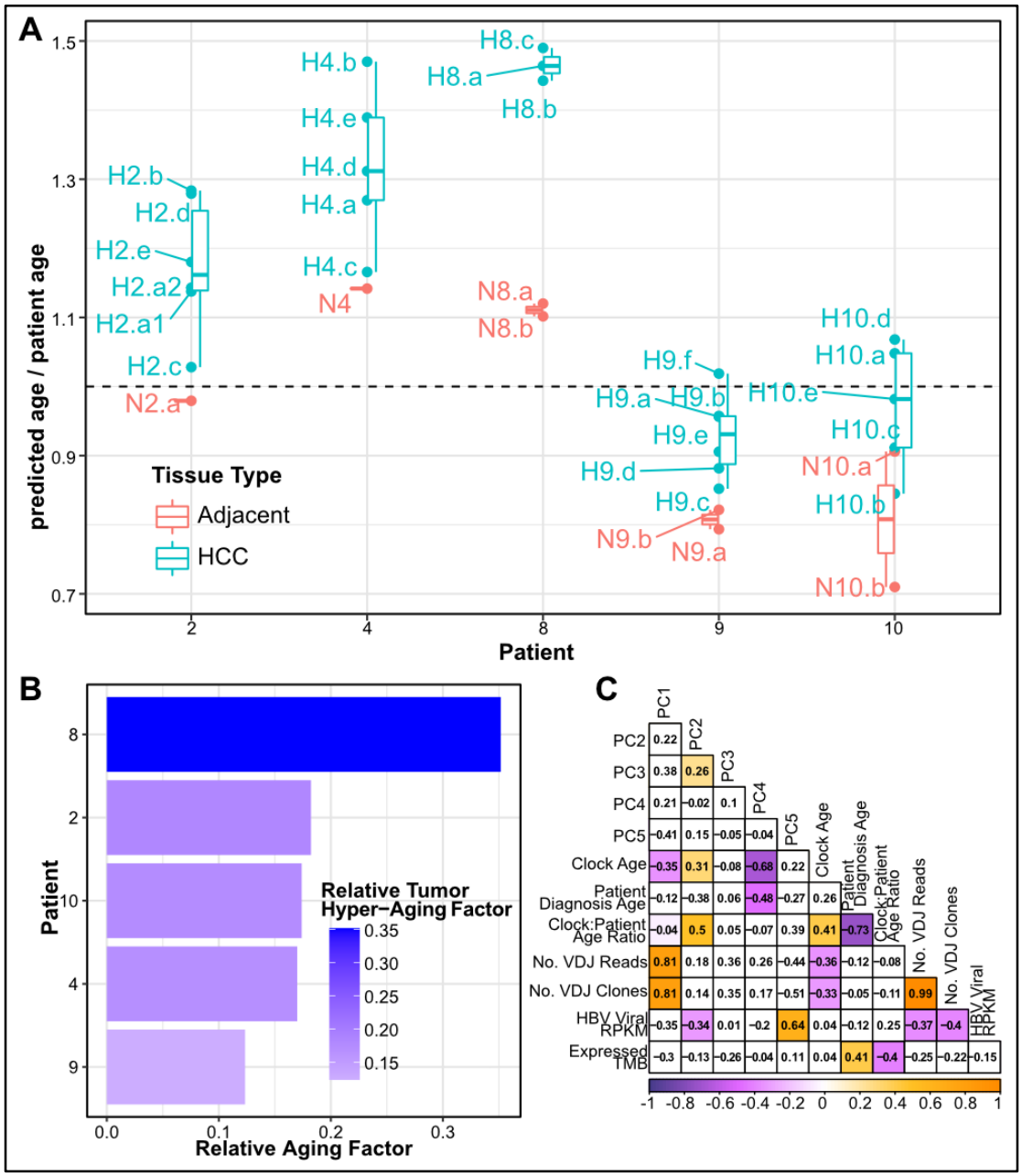


Figure 4. A) Ratio of predicted methylAge to patient age across regionally sampled patients . B) Hyper-aging relative to methylAge in adjacent normals explains patient 8 outlier status through accelerated tumoral hyper-aging. C) Correlation heatmap of MSSM cohort sample clinical, phenotype, and PCA with methylation clock age and relative age factors.

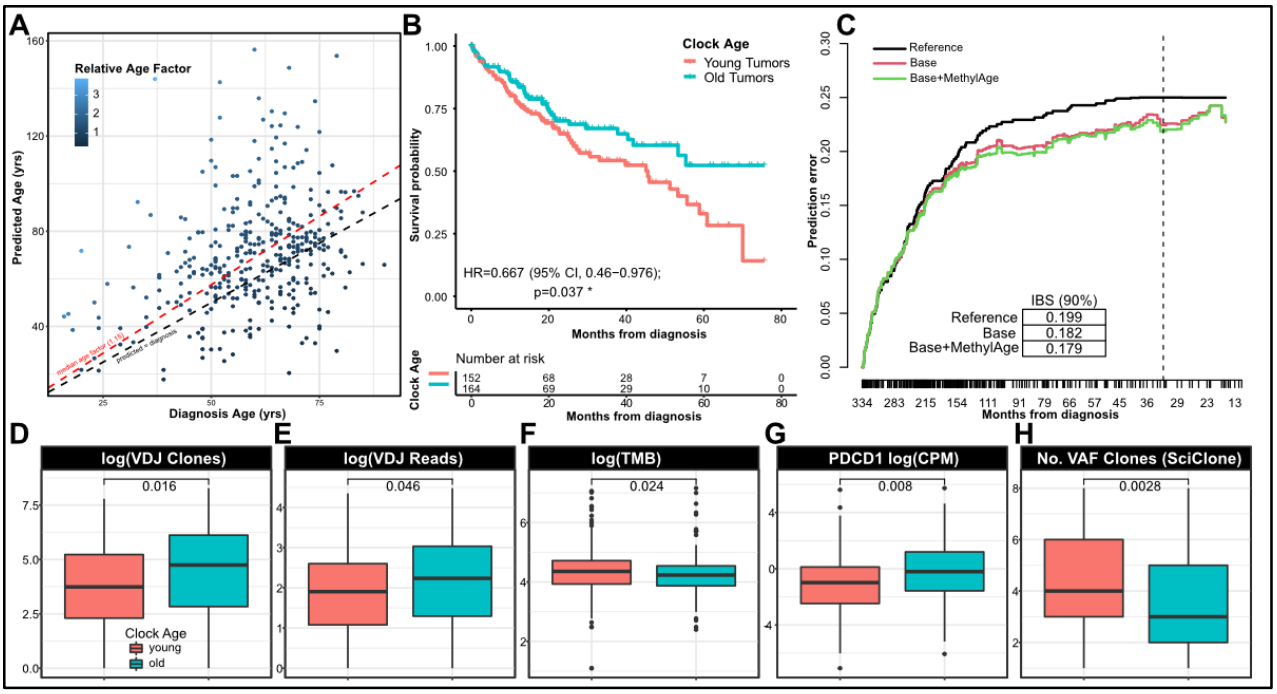
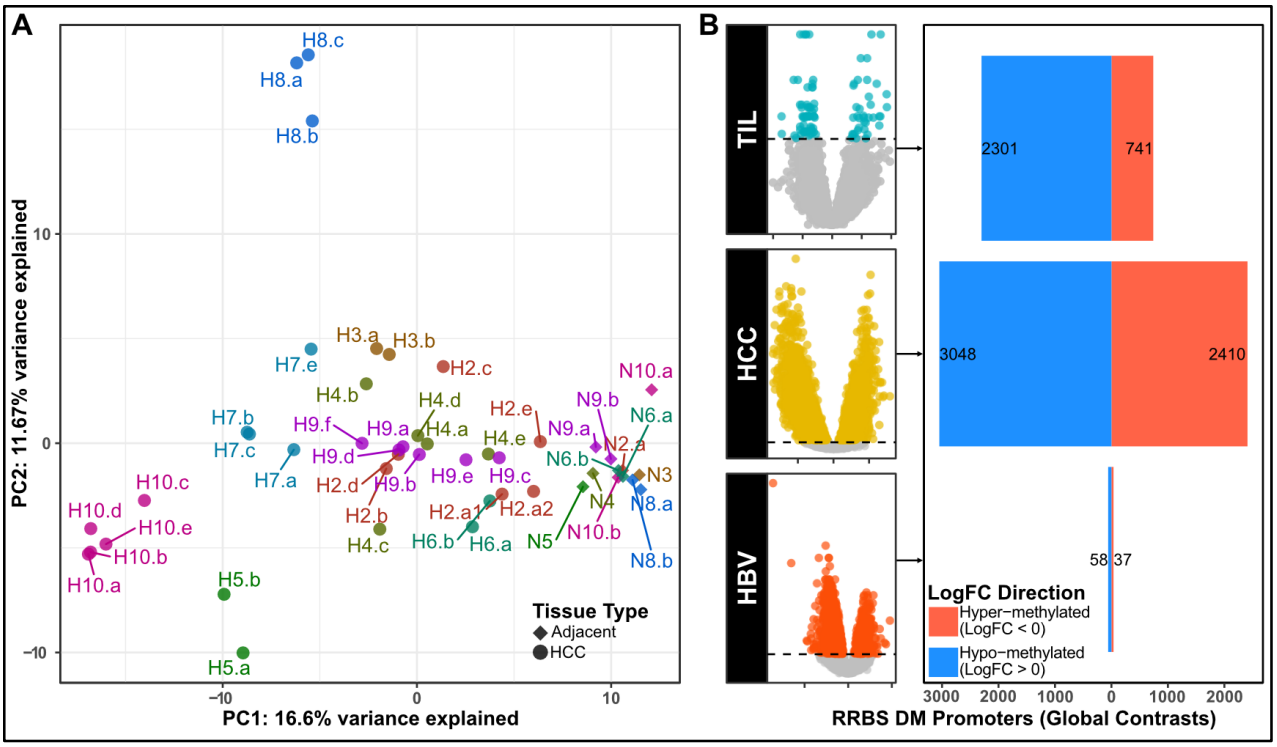
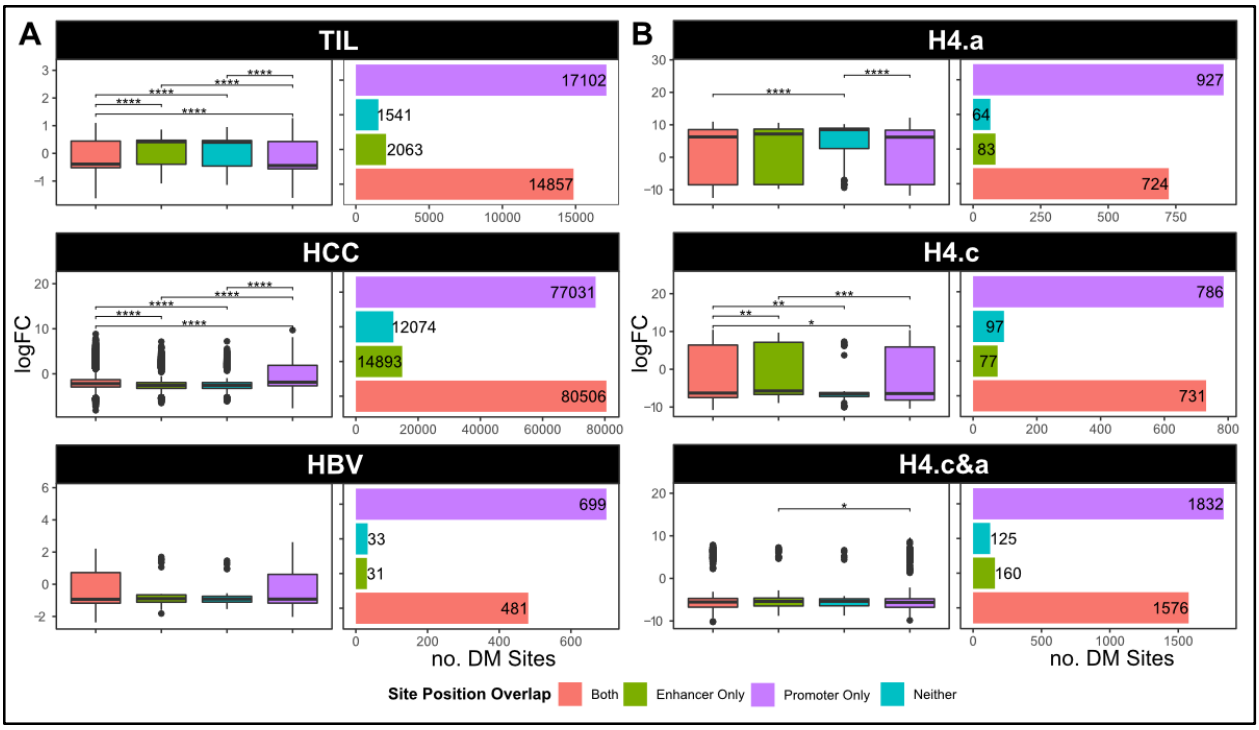


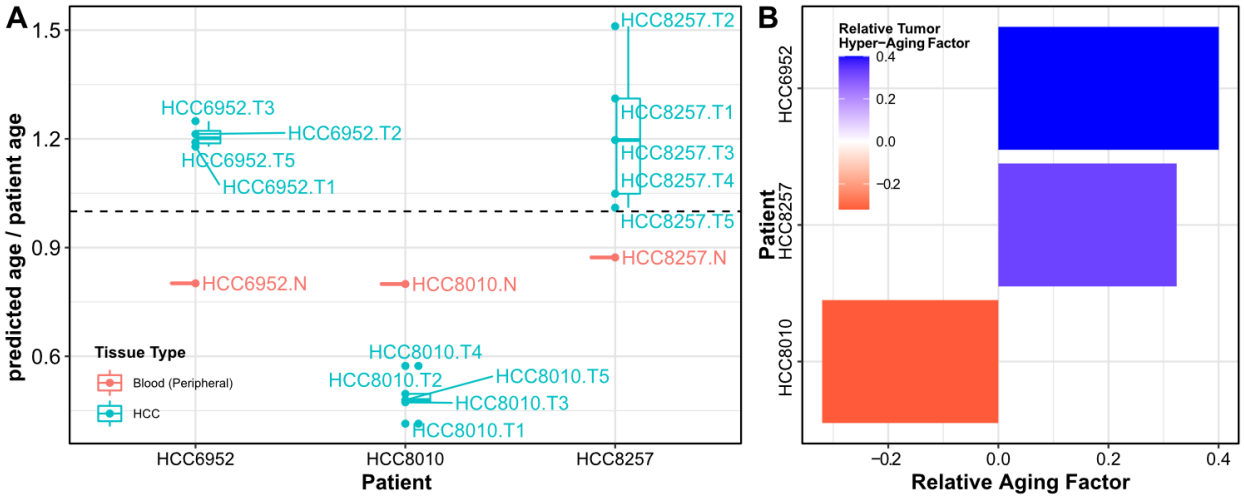
Figure 5. A) Age factor landscape of HCC tumors from TCGA LIHC. B) Kaplan-Meier analysis illustrates the survival benefit of “old” tumors with relative hyper-aging signature. C) Comparison of prediction error across cph models. A survival model including the predicted methylation age factor outperforms the model using only covariates of clinical tumor stage, sex, and TMB, across 90% of patient events. D) LIHC HCC relative methylation age factor tracks with known clinical markers including VDJ clone count, E) VDJ read burden, F) tumor mutation burden (total number of somatic mutations), G) PD-1 expression from RNA-seq, and H) number of tumor subclones estimated via SciClone.



Supplementary Figure 1. A) PCA of promoter aggregated methylation beta matrix. B) eRRBS DM promoters for global contrasts of (top to bottom) TIL, Tumor vs Adjacent, and HBV.



Supplementary Figure 2. Methylation change comparisons across DM regions covering promoter, enhancer, both, or neither feature loci for A) global contrasts (top to bottom) TIL, Tumor vs Adjacent, and HBV; B) regional contrasts across HCC in patient 4.



Supplementary Figure 3. Tumoral hyper-aging validation via GEO (Lin et al 2017 Cancer Discover Study). A) Site-by-site methylation predicted age to patient age ratios across three multi-regionally sampled HCCs. B) Tumoral methylation hyper-aging observed in two tumors, and hypo-aging in one tumor.