MR. CANER ERCAN (Orcid ID : 0000-0002-5611-2699)
DR. SALVATORE PISCUOGLIO (Orcid ID : 0000-0003-2686-2939)

Received Date : 20-Jul-2021
Revised Date : 18-Oct-2021
Accepted Date : 21-Nov-2021
Article type : Research Article

Epigenetic priming in chronic liver disease impacts the transcriptional and genetic landscapes of hepatocellular carcinoma

John Gallon[1], Mairene Coto-Llerena[1,2], Caner Ercan[2], Gaia Bianco[1], Viola Paradiso[2], Sandro Nuciforo[3], Stephanie Taha-Melitz[1,4], Marie-Anne Meier[3], Tujana Boldanova[3,4], Sofía Pérez-del-Pulgar[5], Sergio Rodríguez-Tajes[5], Markus von Flüe[4], Savas D Soysal[4], Otto Kollmar[4], Josep M. Llovet[6,7], Augusto Villanueva[7], Luigi M. Terracciano[8,9], Markus H. Heim[3,4], Charlotte K.Y. Ng[10,11#] and Salvatore Piscuoglio[1,2#]

[1]Visceral surgery and Precision Medicine research laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland;
[2]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland;
[3]Hepatology Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland;
[4]Clarunis, Department of Visceral Surgery, University Centre for Gastrointestinal and Liver Diseases, St. Clara Hospital and University Hospital Basel, Switzerland.
[5]Liver Unit, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERehd, Barcelona, Spain.
[6]Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS, Hospital Clínic, University of Barcelona, Barcelona, Catalonia, Spain;

[7]Liver Cancer Program, Divisions of Liver Diseases and Hematology/Medical Oncology, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA;
[8]Department of Pathology, Humanitas Clinical and Research Center, IRCCS, Milan, Italy;
[9]Humanitas University, Department of Biomedical Sciences, Milan, Italy;
[10]Department for BioMedical Research, University of Bern, Bern, Switzerland;
[11]SIB, Swiss Institute of Bioinformatics, Lausanne, Switzerland.

#Correspondence: Dr. Salvatore Piscuoglio, Visceral Surgery and Precision Medicine Research Laboratory, Department of Biomedicine, Basel, Hebelstrasse, 20, 4031, Switzerland Tel: +4161 328 68 74; Fax: +41612653194. E-mail: s.piscuoglio@unibas.ch or **Dr. Charlotte K Y Ng,** Department of BioMedical Research, University of Bern, Murtenstrasse 40, CH-3008 Bern, Switzerland; Tel: +41 31 632 87 79; Email: charlotte.ng@dbmr.unibe.ch.

Running title: Epigenetic priming in CLD impacts HCC

Keywords: Chronic liver disease, hepatocellular carcinoma, methylation, epigenetic priming.

Abbreviations: CLD, Chronic liver disease; CLDme, Chronic liver disease methylation; EMT, Epithelial-mesenchymal transition; HCC, Hepatocellular carcinoma; DE, Differentially expressed (genes); NAFLD, nonalcoholic fatty liver disease; PC, principal component; DM, differential methylation; DMR, differentially methylated regions.

Abstract

Hepatocellular carcinomas (HCCs) usually arise from chronic liver disease (CLD). Pre-cancerous cells in chronically inflamed environments may be 'epigenetically primed', sensitising them to oncogenic transformation. We investigated whether epigenetic priming in CLD may affect HCC outcomes by influencing the genomic and transcriptomic landscapes of HCC. Analysis of DNA methylation arrays in ten paired CLD-HCC identified 339 shared dysregulated CpG sites and 18 shared differentially methylated regions compared to healthy livers. These regions were associated with dysregulated expression of genes with relevance in HCC, including Ubiquitin D (*UBD*), Cytochrome P450 Family 2 Subfamily C Member 19 (*CYP2C19*) and O-6-Methylguanine-DNA Methyltransferase (*MGMT*). Methylation changes were recapitulated in an independent cohort of nine paired CLD-HCC. High CLD methylation score, defined using the 124 dysregulated CpGs in CLD and HCC in both cohorts, was associated with poor survival, increased somatic genetic alterations, and *TP53* mutations in two independent HCC cohorts. Oncogenic transcriptional and methylation dysregulation is evident in CLD and compounded in HCC. Epigenetic priming in CLD sculpts the transcriptional landscape of HCC and creates an environment favouring the acquisition of genetic alterations, suggesting that the extent of epigenetic priming in CLD could influence disease outcome.

Molecular Oncology (2020)

1. Introduction

Hepatocellular carcinoma (HCC) typically arises in the context of chronic inflammation and tissue necrosis.[1] Viral infections, excessive alcohol consumption, ingestion of aflatoxin B1 and nonalcoholic fatty liver disease (NAFLD) are all well-defined causes of chronic liver disease (CLD) and risk factors for HCC development.[2] Regardless of the aetiology, hepatocarcinogenesis usually occurs as a multistep progression from the healthy liver to fibrosis, cirrhosis, and ultimately HCC; a process which relies heavily on changes in the tissue microenvironment and the accumulation of epi/genetic alterations in the hepatocytes and stellate cells.[3–5]

The concept of epigenetic priming has been proposed in other cancers emerging from chronic health conditions or environmental factors, such as obesity in colon cancer or cigarette smoke in lung cancer.[6,7] In this model, pre-cancerous cells assume a new, epigenetically defined identity which sensitises them to oncogenic transformation. Similar to these cancers, HCC arises from a background of chronic disease. Indeed, epigenetic dysregulation was initially reported in CLD, with hypermethylation of the promoters of tumour suppressors such as *RASSF1A*, *APC* and *CDKN2A*.[8–10] These studies demonstrated that select epigenetic alterations that exist in HCC are also present in CLD, suggesting that they may contribute to disease initiation and/or progression. Subsequently, DNA methylation changes in non-alcoholic fatty liver disease have been associated with aberrant gene expression in non-tumoural tissue, while genome-wide analysis of methylation patterns have revealed the extent of epigenetic dysregulation in precancerous nodules.[8,11,12] The prognostic utility of DNA methylation patterns in HCC, following tumourigenesis, has also been demonstrated, and particular DNA methylation signatures have recently been linked to specific driver gene alterations.[13,14]

This literature points to critical roles for epigenetic changes acquired during CLD in the initial emergence of HCC, and for those acquired during HCC on disease progression. However, the impact on the transcriptional and genetic landscapes of HCC, and prognostic utility of genome-wide DNA methylation changes acquired specifically during CLD remains unexplored. Here we identify genome-wide DNA methylation changes acquired in non-tumoural CLD tissue, associated with distinct transcriptional and genetic landscapes in tumour samples. Using the results obtained from these analyses we developed a score that may have prognostic value in HCC.

2. Materials and Methods

2.1 Patients and samples

For the discovery cohort, ten patients with HCC were diagnosed at the University Hospital Basel and were prospectively recruited for this study after written informed consent. HCC biopsies, concomitant chronic liver disease (CLD) biopsies and peripheral blood leukocytes were collected from the HCC patients (**Fig. S1 and Table S1A**).

From each patient undergoing a diagnostic liver biopsy, two ultrasound-guided core needle biopsies of the primary tumor and two biopsies from the CLD tissue and whole blood were collected at diagnosis at the same time. Of the two biopsies taken from the primary tumor and from CLD tissue, one was processed and embedded in paraffin for clinical purposes and the other one was snap-frozen and stored at −80°C for research purposes. 10mL of whole blood was collected and processed immediately for the isolation of peripheral blood leukocytes ('buffy coat'). All biopsies were histologically characterized by two hepatopathologists (CE and LMT) to confirm the initial diagnosis of HCC.[15] the study was performed in accordance with the Declaration of Helsinki and the approval for the use of these samples has been granted by the ethics committee (Protocol Number EKNZ 2014-099).

For the validation cohort, 9 patients with HCC and concomitant CLD diagnosed at the Hospital Clinic, Barcelona or Mount Sinai, New York were prospectively recruited after written informed consent (Protocol Number 2010/5896 (IRB Hospital Clinic, Barcelona), **Fig. S1** and **Table S1A**).

As controls for methylation array profiling, healthy livers from two patients with colorectal cancer metastatic to the liver (University Hospital Basel, Protocol Number EKNZ 2014-099) and histologically normal tissues from ten patients undergoing hepatic resection due to non-cancer related diseases (Protocol Number 2010/5896 (IRB Hospital Clinic, Barcelona)) were used. As controls for transcriptomic analysis, liver biopsies with normal histology obtained from 15 patients without HCC and with normal liver values were used (University Hospital Basel, Protocol Number EKNZ 2014-099, **Fig. S1**).

For all patients in the discovery and validation cohorts, the clinical staging was determined according to the Barcelona Clinic Liver Cancer staging system.[16] Sex and age of the patients, clinical diagnosis, underlying liver disease (hepatitis B/C infection, alcoholic liver disease, non-alcoholic fatty liver disease) were retrieved from clinical files (**Table S1A**).

The samples encompassed the diverse backgrounds of HCC; of the 10 patients, 5 were diagnosed with alcohol-related HCC, 3 with HBV/HCV-related HCC, and 2 NAFLD-related HCC (**Table S1A**). Our discovery cohort of ten patients largely consisted of early-stage tumors (70% BCLC stages 0-A) with non-multinodular HCC (70% < 2 nodules)**.** Using the data generated from these samples we investigated how transcriptional changes might drive disease progression.

2.2 Nucleic Acid extraction

Genomic DNA and total RNA from biopsies from the discovery cohort were extracted using the ZR-Duet DNA and RNA MiniPrep Plus kit (Zymo Research) following the manufacturer's instructions. Prior to extraction, biopsies were crushed in liquid nitrogen to facilitate lysis. Extracted DNA was quantified using the Qubit Fluorometer (Invitrogen).[17] DNA from peripheral blood leukocytes ('buffy coat') was extracted using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. For the validation cohort DNA was extracted using Char-geSwitch genomic DNA Mini Tissue kit (Invitrogen) following the manufacturer's instructions.[8]

2.3 Exome sequencing and analysis

Whole-exome capture was performed using the SureSelectXT Clinical Research Exome (Agilent) platform according to the manufacturer's guidelines (**Fig. S1)**. Sequencing was performed on an Illumina HiSeq 2500 at the Genomics Facility Basel according to the manufacturer's guidelines. Paired-end 101-bp reads were generated. Reads obtained were aligned to the reference human genome GRCh38 using Burrows-Wheeler Aligner (BWA, v0.7.12).[18] Local realignment, duplicate removal, and base quality adjustment were performed using the Genome Analysis Toolkit (GATK, v3.6) and Picard (http://broadinstitute.github.io/picard/).[19] Somatic single nucleotide variants (SNVs) and small insertions and deletions (indels) were detected using MuTect (v1.1.4) and Strelka (v1.0.15), respectively.[20,21] We filtered out SNVs and indels outside of the target regions, those with a variant allelic fraction (VAF) of <1% and/or those supported by <3 reads. We also excluded variants for which the tumor VAF was <5 times that of the paired non-tumor VAF. We further excluded variants identified in at least two of a panel of 123 non-tumor samples, including the non-tumor samples included in the current study, captured and sequenced using the same protocols using the artifact detection mode of MuTect implemented in GATK. To account for the presence of somatic mutations that may be present below the limit of

sensitivity of somatic mutation callers, we used GATK Unified Genotyper to interrogate the positions of all unique mutations in all samples from a given patient to define the presence of additional mutations. Variants identified by this genotyping step supported by a minimum of 2 reads are annotated as "Genotyped". Hotspot missense mutations were annotated using the published resources.[22,23]

Allele-specific copy number alterations were identified using FACETS (v0.5.6), which performs joint segmentation of the total and allelic copy ratios and infers purity, ploidy, and allele-specific copy number states.[24] Copy number states were collapsed to the gene level using the median values to coding gene resolution based on all coding genes retrieved from the Ensembl (release GRCh37.p13). Genes with total copy number greater than gene-level median ploidy were considered gains; greater than ploidy + 4, amplifications; less than ploidy, losses; and total copy number of 0, homozygous deletions. Somatic mutations associated with the loss of the wild-type allele (i.e., loss of heterozygosity [LOH]) were identified as those for which the lesser (minor) copy number state at the locus was 0. All mutations on chromosome X in male patients were considered to be associated with LOH.[25]

2.4 RNA sequencing and analysis

200 ng total RNA was used for RNA-seq library prep with the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold (Illumina) according to manufacturer's specifications (**Fig. S1)**. Sequencing was performed on an Illumina HiSeq 2500 using v4 SBS chemistry at the Genomics Facility Basel according to the manufacturer's guidelines. Sequence reads were aligned to the human reference genome GRCh37 by STAR using the two-pass approach.[26] Transcript quantification was performed using RSEM.[27] Genes without >10 counts in at least 2 samples were discarded. Counts were normalized using the median of ratios method from the DESeq2 package in R version 3.6.1 (R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

Comparisons of the intra-group variation, defined as the within-group pairwise euclidean distance based on their principal components, were performed using Wilcoxon tests. Differential expression analysis was performed using the wald test in DESeq2.[28] Genes with |logFC| > 1.5 and FDR < 0.05 were considered differentially expressed. Gene set enrichment analysis was performed using the fgsea package using Hallmark gene sets, with genes ranked based on the *t* statistic from DESeq2.[28,29]

2.5 Methylation profiling and analysis

Methylation profiling was performed using Infinium® MethylationEPIC BeadChips and Infinium® HumanMethylation450 BeadChip (Illumina) on the discovery and validation cohorts, respectively (**Fig. S1**). After whole-genome amplification and enzymatic fragmentation, the samples were hybridized to the BeadChips and scanning was conducted with the Illumina iScan. Idat files were exported and analyzed using the minfi package in R.[30] All arrays were reduced to probes present on both the HumanMethylation 450 and MethylationEPIC BeadChips, as 10/12 normal samples, those from Barcelona, were analysed on the HumanMethylation 450 BeadChip. Probes associated with SNPs, on the sex chromosomes or with a detection *P* value > 0.01 in any sample were removed prior to analysis. Data were normalised using the Noob algorithm from the minfi package.[30] Probes were annotated using the IlluminaHumanMethylation450kanno package in Bioconductor.

Principal component analysis was performed using the top 500 most variable CpG sites. Comparisons of the intra-group variation, defined as the within-group pairwise euclidean distance based on their principal components, were performed using Wilcoxon tests. Comparisons of the inter-group variation, as measured by pairwise euclidean distance based on their principal components between samples of different groups, were performed using Wilcoxon tests. Probe-level differential methylation analysis was performed for 42,925 CpG sites using limma. Probes with |logFC| > 1.5 and FDR < 0.05 were considered differentially methylated. Differentially methylated regions (DMRs) were called using DMRcate using the parameters 'lambda=500, C=5'.[31–33] DMRs with mean change in *B* value > |15%| and FDR < 0.05 were considered differentially methylated. DMRs were annotated using the annotateTranscripts function from the bumphunter and the TxDb.Hsapiens.UCSC.hg19.knownGene packages from Bioconductor.[34]
To assess the relationship between DMRs and methyl-binding domain proteins and repressive histone modifications we downloaded ENCODE ChIP-seq data for ZBTB38, ZBTB4 and Histone 3 Lysine 27 trimethylation,[35] and intersected these with the DMRs using bedtools.[36]

2.6 Downloading and annotation of the TCGA cohort

DNA methylation, gene expression, mutation and survival data for 430 HCC samples were downloaded from TCGA using the TCGAbiolinks package in R on 28th July 2020.[37,38] Copy

number alteration data was downloaded from TCGA Firehose (Broad Institute TCGA Genome Data Analysis Center (2016): Firehose VERSION run. Broad Institute of MIT and Harvard. doi:10.7908/C11G0KM9). Assessment of the presence or absence of cholestasis, Mallory bodies, tumor-infiltrating lymphocytes, vessel infiltration, and necrotic areas was performed as previously described.[39] TCGA samples were reduced to the 368 for which complete DNA methylation data, survival data and histological annotation were available.

2.7 Development of CLD DNA methylation (CLDme) prognostic score

124 probes were DM in CLD and HCC in both the discovery and validation cohorts; after removing 15/124 probes with NA values in the TCGA dataset an elastic net Cox regression model was built using the remaining 109 probes and overall survival as the response variable. Elastic net regression is a regularization method that balances the trade-off between bias and variance using L1 and L2 regularisation parameters.[40] These are combined into a single parameter, lambda, in the implementation of elastic net regression in the glmnet R package.[41] The optimal value for lambda was selected using the training set and 10-fold cross validation using the cv.glmnet function from the glmnet package.[41] The model was built on a training set consisting of a randomly selected 70% (n=257) of the 368 TCGA HCC samples. A fixed seed was used in order to ensure reproducibility. The remaining 30% (n=111) samples were reserved for testing. Samples were classified as CLDme score high or low based on the median score of samples in the TCGA training set after defining the optimal value for lambda. Differences in survival between CLDme High/Low groups were compared using the log-rank test, adjusted for disease history and stage (the only factors significantly associated with survival).

2.8 Analysis of TCGA samples stratified by CLDme score

To compare the gene expression profiles between CLDme High and Low samples, differential gene expression analysis was performed using 362 TCGA samples for which DNA methylation, transcriptomic and clinical information were available. Differential gene expression analysis was performed using the wald test in DESeq2.[28] Comparisons of numbers of mutations, and copy number alterations between CLDme High and Low samples were carried out using Wilcoxon tests on the 306 and 364 TCGA samples for which clinical, DNA methylation, and mutation/copy number data were available, respectively. Comparison of lymphocyte invasion between CLDme High/Low groups was carried out using the histological annotation as described previously.

ImmuneScores for each TCGA sample were downloaded from https://xcell.ucsf.edu/ and compared between CLDme High/Low groups.[42]

2.9 Immunohistochemistry

Immunohistochemical staining was performed on a Benchmark immunohistochemistry staining system (Bond, Leica) with BOND polymer refine detection solution for DAB, using anti-MGMT (1:800, abcam ab39253) primary antibody as substrate as previously described.[43] Images were acquired using an Olympus BX46 microscope as previously described. MGMT immunoreactivity was scored semi-quantitatively by multiplying the proportion of MGMT positive cells (%) and the staining intensity (0 = none; 1 = weak; 2 = intermediate; 3 = strong). Statistical comparison was performed using paired Wilcoxon test.

3. Results

3.1 Transcriptional alterations present in HCC are detectable in CLD tissue

To identify transcriptional alterations in diseased liver tissues which progressed to HCC, we performed RNA-sequencing on needle biopsies from ten HCC tissue and matched adjacent chronic liver disease (CLD) tissue, along with 15 healthy liver samples against which CLD and HCC transcriptional profiles were compared (**Fig. S1**). Unsupervised analysis of gene expression data showed that normal and HCC samples form distinct clusters (**Fig. 1A**), with CLD tissues clustering closer to the normal tissues than HCCs. This was reflected in unsupervised consensus clustering which showed normal and HCC clustering separately, with CLD tissues split between these two clusters (**Fig. S2**). Differential gene expression analysis detected a significant overlap between transcriptional alterations in CLD and HCC when compared to normal samples. 978/1,269 (77.1%) and 697/996 (70.0%) genes down- and upregulated, respectively, in CLD were also differentially expressed in HCC (both $P < 0.0001$, hypergeometric tests, **Fig. 1B, Tables S1B-S1D, Fig. S3**). HCCs, however, acquired a further 1,562 and 1,818 genes down- and upregulated respectively. Furthermore, the change in expression of the 1,675 genes showing DE in both CLD and HCC was significantly amplified in HCC compared to CLD ($P < 2.22e^{-16}$, paired Wilcoxon test, **Fig. 1C**).

Pathway analysis of the dysregulated genes show upregulation of epithelial-to-mesenchymal transition (EMT)-related genes in CLD and HCC **(Fig. 1D)**, consistent with the tissue regeneration and fibrogenic processes occurring during CLD.[29] Interestingly, cancer-related pathways, such as cell cycle (MYC targets V1) and MTORC1 signaling, were also upregulated in both HCC and CLD samples, suggesting that these pathways may already be transcriptionally dysregulated in the pre-cancerous lesion. The magnitude of upregulation of these pathways was greater in the HCCs than in the CLDs, highlighting the progressive nature of these changes. By contrast, we also found upregulation of DNA repair and mitotic spindle pathways and downregulation of the xenobiotic and bile acid metabolism in HCC samples, but not the CLD samples (**Fig. 1D**). Conversely, we found significant alteration of the complement and interferon gamma response pathways in the CLD samples but not the HCC.

To determine whether the transcriptional alterations observed in CLD were driven by somatic genetic alterations we performed whole exome sequencing on the matched CLD and HCC samples (**Fig. S1**). We detected at least one somatic mutation in the most commonly mutated genes in HCC[44] and substantial copy number alterations in 9/10 HCCs (**Fig. S4, Tables S1E, S1F**). However, except for one low confidence mutation in *APOB* in the CLD from patient 6, we found no evidence for shared mutations in the commonly mutated genes or copy number alterations between CLD and HCC samples from the same patient.

Together these data demonstrate the significant accumulation of cancer-associated transcriptional changes in CLD, which are compounded in HCC, and suggest that the aberrant transcriptional landscape of HCC may start developing during CLD independent of genetic alterations.

3.2 DNA methylation alterations in HCC are detectable in CLD

Given that cancer-associated transcriptional changes in CLD do not appear to be underpinned by genomic changes frequently observed in HCC, we asked whether epigenetic alterations may help explain the transition towards HCC. In support of this hypothesis, we found progressive loss of expression of *MAT1A* (CLD q = 0.02, Log2FC = -0.80, HCC q = 1.43e^{-11}, Log2FC = -2.14, **Tables S1B, S1C**), which catalyses synthesis of the universal methyl donor S-adenosylmethionine, as previously reported in cirrhotic livers.[45] As the loss of S-adenosylmethionine availability

suggests the potential for epigenetic reprogramming, we subjected the same 10 pairs of CLD and HCC, and 12 normal liver samples to methylation profiling (**Fig. S1**).

Principal component (PC) analysis of the methylation profiles reflected the findings from the transcriptional analysis; CLD/normal livers were separated from HCCs on PC1 but CLDs were separated from normal livers by PC3 (**Fig. 2A, Fig. S5A**), reflecting a recent study showing a gradient of methylation changes spanning the progression from health liver to HCC.[46] We identified 54,888 differential methylated (DM, |log2FC| > 1.5, q < 0.05) CpG sites in the HCC samples compared to normal tissue, the majority of which (46,669, 85%) were hypomethylated (**Fig. 2B, Table S1G**), consistent with the phenomenon of genome-wide hypomethylation in cancer cells.[47,48] Differential methylation was observed at CpGs associated with *P14* and *RASSF1A,* previously shown to be aberrantly methylated in HCC (**Fig. S5B**).[49–51] In the CLD samples, we detected 586 DM CpGs compared to normal liver (**Fig. 2C, Table S1H**). Of these, 339 CpGs, associated with 222 genes, were also DM in the HCC samples, representing a highly significant overlap (P < 0.0001, hypergeometric test, **Fig. 2C, Fig. S6, Table S1I**). Importantly, as with the genes which were DE in both CLD and HCC, the 339 CpG sites DM in both CLD and HCC compared to normal showed significantly larger methylation changes in HCC than CLD (P < $2.22e^{-16}$, paired Wilcoxon test, **Fig. 2D**). Compared to HCC, a greater proportion of the methylation changes observed in CLD had the potential to regulate gene expression. In HCC samples, 53.5% DM CpG sites were hypomethylated and in Open Sea regions (> 4 kb from a CpG island), compared to 21.5% in CLD samples. On the other hand, DM CpGs in CLD samples were enriched in CpG island and shore regions (< 2 kb from a CpG island) compared to HCC (P = 0.016, OR = 1.36, Fisher's exact test, **Fig. 2E**), which could suggest that methylation alterations in CLD have proportionally greater effect on transcriptional regulation than those in HCC.

Given that DMRs have been shown to be more strongly linked to gene expression than methylation changes at single CpG sites,[52] we further identified differentially methylated regions (DMR), regions of adjacent CpG sites showing significantly altered methylation (mean change in B value > |0.15|, q < 0.05) in CLD and HCCs.[31] As with the probe-level analysis, we detected substantially more DMRs in the HCC samples than the CLD samples, compared to the normal (11,582 and 121 respectively). Intersecting these regions identified 67 DMRs, containing 262 CpGs, showing altered methylation in both CLD and HCC samples (**Fig. 2F**).

Our data demonstrate the extent of epigenetic changes in CLD, and that many of those changes are amplified in HCC. As genetic alterations typically observed in HCC were not detected in CLD,

while HCC-associated methylation changes were evident, this suggests the aberrant methylome of HCC may, in part, have emerged before tumourigenesis.

3.3 DNA methylation changes in CLD sculpt the transcriptional landscape of HCC

To determine how the DNA methylation changes observed in CLD and HCC shape their transcriptional profiles, we interrogated the 67 DMRs to search for those associated with DE genes (|log2FC| > 1.5, q < 0.05). We therefore removed candidate DMRs for which we did not have gene expression data, those which could not be associated with a gene promoter, i.e., annotated as 'downstream', and those whose change in methylation was not reflected in a significant change in gene expression, in the expected direction. This filtering left 18 regions DM in both CLDs and HCCs associated with DE genes **(Fig. 3A and Tables S1J, S1K**). The genes affected by the epigenetic priming occurring in CLD included hypermethylated regions associated with the cytochrome P450 family gene *CYP2C19* and tuberin sclerosis complex 2 (*TSC2*), both downregulated in the CLD and HCC samples and reported to be lost in HCC with implications for prognosis.[53,54] As an exploratory analysis to further demonstrate the relevance of these regions in the epigenetic regulation of gene expression, we found methyl binding domain protein (ZBTB4 and ZBTB38) and H3K27me^3 peaks from a previously published study overlapped with the DMRs associated with *HDAC11*, *SYT8*, and *TLDC2* [35] suggesting MBD proteins may interact with the identified DMRs (**Fig. S7**). We also identified a hypermethylated DMR within intron 3 of *MGMT*, containing the CpG site cg07554771 (CLD log2FC = 2.89, q = 0.02, HCC log2FC = 3.25, q = 0.0002; **Fig. 3B,** top), hypermethylation of which is correlated with *MGMT* repression in NAFLD and HCC.[11] Furthermore, an additional DMR was detected in the HCC samples, containing the CpG site cg00639517, hypermethylation of which is also correlated with loss of *MGMT* expression (**Fig. S8**).[11] The hypermethylation of *MGMT* was concomitant with a loss of *MGMT* expression in CLD and HCC (CLD log2FC = -0.99, q = 0.007, HCC log2FC = -1.80, q = 9.04e^{-8}; **Fig. 3B,** bottom). Corroborating the progressive loss of *MGMT* expression in HCC progression, MGMT immunohistochemical analysis of an independent set of 12 matched CLD and HCC samples showed significant reduction of MGMT expression in HCC compared to matched CLD samples (P = 0.03, Wilcoxon test, **Fig. 3C**).

While studies on DNA methylation in CLD progression have mainly focussed on hypermethylation and silencing of tumour suppressor genes, 13 of the 18 identified DMRs showed hypomethylation and upregulation in the CLD and HCC samples compared to normal liver (**Fig. 3A**). These

included the promoters of *UBD* (*FAT10*), a ubiquitin-like modifier, the calponin *TAGLN2,* both implicated in the progression of HCC, and *BAIAP2L2* coding for Insulin Receptor Tyrosine Kinase substrate, associated with actin remodelling and promoting HCC proliferation.[55–57]

With the changes in DMRs reflected in gene expression changes in CLD and HCC, our findings demonstrate the potential for epigenetic priming in CLD, not only to influence tumourigenesis as has been extensively reported, but also to sculpt the transcriptional landscape of the subsequent HCC.

3.4 CLD-associated DNA methylation changes distinguish CLD and HCC from normal livers across cohorts

To rule out the possibility that the DNA methylation changes we detected in CLD were cohort-specific, we analysed the DNA methylation data from an independent validation cohort of 9 pairs of CLD and HCC samples (**Fig. S1**). Principal component analysis of the validation cohort using the *B* values of the 51 CpG sites in the 18 DE-gene associated DMRs identified in both CLD and HCC in the discovery cohort (**Fig. 3A**) separated the normal samples from the CLD and HCC samples (**Fig. 4A**).

Differential methylation analysis of the samples in the validation cohort identified 2,970 and 86,473 DM CpG sites in the CLD and HCC respectively ($|\log2FC| > 1.5$, $q < 0.05$, compared to normal livers). As in the discovery cohort, the overlap of 1,268 DM CpG sites in both CLD and HCC in the validation cohort was highly significant ($P < 0.0001$, hypergeometric test; **Fig. 4B)**. These CpG sites included those identified in genes already reported in the discovery cohort such as in *MGMT* **(Fig. S9).** Importantly, the overlap between the set of shared DM CpG sites identified in both cohorts (124 CpG sites) was also highly significant ($P < 0.0001$, hypergeometric test, **Fig. 4B**). The consistency of the observed methylation changes was also conserved at the DMR level where 8 of 18 identified CLD-HCC DMRs, associated with DE genes in the discovery cohort, were shared with the validation cohort (**Fig. 4C**). **Fig. 4D** shows the change in methylation between normal, and CLD and HCC samples at a representative gene promoter; *UBD*, found to be upregulated in the discovery cohort, concomitant with loss of methylation at its promoter. This was also observed in the validation cohort.

Together these data suggest that specific epigenetic changes, with the potential to influence gene expression, occur consistently in CLD and are maintained in HCC.

3.5 Epigenetic priming in CLD creates a permissive environment for the accumulation of somatic mutations in HCC

Next, we assessed whether the methylation state of the 124 DM CpG sites in both CLD and HCC samples in both datasets was of clinical relevance, using methylation, clinical and survival data of 368 HCCs from The Cancer Genome Atlas (TCGA).[38] We randomly split the TCGA dataset 70:30 into training (n=257) and testing (n=111) set and, after removing 15 CpG sites with missing values, we trained an elastic net regression model using the remaining 109 CpG sites to define a "CLD Methylation (CLDme)" score for each sample (**Methods**, **Fig. 5A and Table S1L**). A multivariate Cox proportional hazards model, adjusted for disease history and stage (the only factors significantly associated with survival, **Table S1M**), showed a high CLDme score to be an independent predictor of poor survival in the test set of 111 TCGA samples (**Fig. 5B,** log-rank P = $5e^{-07}$, HR = 7.97). We confirmed our findings using an independent dataset of 241 patients[8] and showed that a high CLDme score was again significantly associated with survival independent of disease history and stage (log-rank P = 0.001, HR = 1.28; **Fig. 5C)**.

We next sought to determine whether the CLDme score was associated with genetic and transcriptomic alterations. Using the entire TCGA cohort, a differential expression analysis between CLDme High and Low samples found 8/18 genes associated with DMRs in the initial discovery samples were DE between CLDme High and Low samples (q < 0.05, DESeq2 Wald test; **Fig. S10A**). On the genetic level, we found that the CLDme High samples had significantly more mutations than the CLDme Low samples (P = 0.0015, Wilcoxon test; **Fig. 5D**). As mutations in *TP53* define a class of HCCs with poor prognosis,[58] we further asked whether CLDme was associated with *TP53* mutations. We found that *TP53* mutations were significantly enriched among CLDme High samples (44.5% vs 29% in CLDme Low, P = 0.0065, OR = 1.94, Fisher's exact test, **Fig. 5E**). Similarly, we observed that CLDme High samples showed significantly higher copy number changes than CLDme Low samples (P = 0.0001, Wilcoxon test; **Fig. 5F**).

Together these data suggest epigenetic priming in CLD may have roles in HCC that go beyond a role in tumourigenesis. By shaping the transcriptional landscape of HCC and creating a more

permissive environment for the acquisition of genetic alterations, aberrant methylation patterns in CLD may influence HCC outcome.

4. Discussion

In this proof-of-concept study we demonstrate the extent of epigenetic and associated transcriptomic changes occurring in the progression from normal tissue, to CLD and HCC. We show that methylation changes acquired during CLD may not only have a role in tumourigenesis, but also sculpt the transcriptional landscape of the subsequent HCC, with implications for disease outcomes. We detected significant hypermethylation affecting genes previously reported to be aberrantly methylated and silenced, and incorporated in HCC prognostic scores e.g. *RASSF1A, APC* and *P14*.[8,9,59–62] However, here, using two cohorts, we expand upon these by showing the extent and impact of DNA methylation changes in CLD is more far-reaching than has previously been reported, affecting genes for which aberrant methylation has not, to our knowledge, been reported in CLD e.g. *CYP2C19, TSC2* and *TAGLN2*. Firstly, we showed that genes reported to be upregulated and, in some cases, to promote HCC progression, such as *HDAC11*, *UBD* (*FAT10*) and *TAGLN2*,[55,63,64] are hypomethylated in CLD samples, suggesting these prognostically relevant epigenetic and transcriptional changes may arise before HCC has developed. Secondly, we showed that high CLDme score was associated with higher levels of *TP53* alterations, a poor prognostic indicator,[58] suggesting epigenetic changes acquired during CLD may be permissive for genetic alterations with the potential to influence HCC prognosis. While derived from DNA methylation changes initially detected in a small dataset, we validated the prognostic relevance of our model in two independent cohorts of HCC patients.

Our results reflect the recently proposed 'epigenetic priming' model, whereby epigenetic changes induced by chronic exposure to cigarette smoke were shown to sensitise cells to an oncogenic *KRAS* mutation by promoting EMT in lung cancer, or the epigenomic alterations driven by obesity were detectable in pre-cancerous colonic epithelium.[7,65] Importantly, while many of the genes affected by epigenetic priming are not necessarily cancer drivers, in the case of hypomethylated/upregulated genes such as *UBD* and *CREB5*, these genes have been linked to prognosis and disease outcome.[54,66,67] We therefore hypothesise that epigenetic priming during CLD may have implications for HCC prognosis through two possible mechanisms; by sculpting the transcriptional landscape of the emergent HCC, and by creating a permissive environment for the acquisition of genetic alterations affecting genes such as *TP53* that influences outcome.[68]

RNA-seq analysis revealed the nature of transcriptional reprogramming during the progression from CLD to HCC. First, we observed increased expression of immune gene sets in the CLD samples but not the HCC samples. CLD is characterised by the continued expression of cytokines and recruitment of immune cells to the liver.[69] However, during progression to HCC there is a shift towards a suppressive immune environment allowing the growth of cancer cells.[70] Secondly, in keeping with the tissue regeneration and fibrogenic processes occurring during CLD,[71] we found enrichment for genes associated with EMT in CLD samples. Beyond this, we also found gene sets, such as E2F and MYC targets are upregulated in CLD as well as HCC in a progressive manner. Indeed, the upregulation of E2F targets has been reported to define a subclass of HCC.[72,73] Thus tumorigenic transcriptional programmes may already be activated in CLD.

Our genome-wide evaluation of epigenetic dysregulation in matched CLD and HCC revealed that some of the epigenetic alterations in HCC are already detectable in CLD and are associated with transcriptional dysregulation. Of note, we found that DM CpG sites in CLD more frequently affected CpG islands and shores than those in HCC, suggesting that the methylation alterations in CLD may have a greater effect on transcriptional regulation than those in HCC. On the other hand, we also hypothesise that metabolic perturbations on the transcriptional level, such as *MAT1A* loss, may contribute to the epigenetic reprogramming. *MAT1A* loss results in reduced S-adenosylmethionine (SAM) synthesis which is a feature of both cirrhosis and HCC that leads to global hypomethylation in rat livers during hepatocarcinogenesis, and is associated with increased proliferation in human liver cancer cells.[74,75] Our results support a model of epigenetic priming occurring in CLD prior to the development of HCC and, more interestingly, that the influence of epigenetic priming in CLD may go beyond a role in tumorigenesis as it has the potential to create a transcriptional environment that influences disease outcomes.

Expanding upon previous work on epigenetic changes in CLD and HCC,[9,59] here we show that methylation changes acquired during CLD associate with outcome and genetic alterations in HCC. Notably, we detected hypermethylation of CpG sites within the O-6 methylguanine DNA repair gene *MGMT*, concomitant with its downregulation in CLD and HCC. Loss of *MGMT* permits liver cancer development *in vivo*, but recent studies have variably found links and no link between *MGMT* methylation and HCC risk.[76–79] As *MGMT* is the sole enzyme responsible for O-6 methylguanine repair, its hypermethylation-induced silencing, initiated during CLD, may result in increased rates of mutation. Indeed, loss of *MGMT* has been associated with *TP53* mutations in

HCC.[78] Loss of *MGMT* expression, associated with methylation of its promoter, defines a subset of HCCs [78] and has been reported in tumour-adjacent tissue from HCC patients, however this loss of expression was without associated promoter hypermethylation as measured using methylation-specific PCR.[80] In conjunction with our data showing the hypermethylation of non-promoter CpGs in *MGMT*, which have been shown to correlate with *MGMT* expression in NAFLD, this may imply the loss of expression of *MGMT* in HCC may be initiated by non-promoter methylation changes acquired during CLD, which become 'locked-in' by promoter methylation, as has been reported in HCC.[11,78]. Future work will focus on defining whether *MGMT* loss is more associated with tumour emergence, or progression.

The effect of epigenetic changes on the genetic landscapes of HCC is further illustrated by the association between CLDme score and the overall tumor mutational burden and *TP53* mutations in HCC, suggesting that the epigenetic state when a driver gene mutation occurs may influence outcome. Indeed, despite the small cohorts used to discover CLD-associated methylation changes, we showed that the prognostic relevance of the detected changes was consistent in two large-scale cohorts. We also noted that the prognostic relevance of the CLDme score is not purely a result of altered levels of immune infiltration given the lack of association between CLDme score and the presence of lymphocytes or the 'ImmuneScore' as defined by xCell in the TCGA cohort (**Fig. S10B-D**). Between the two cohorts, we also observed that the difference in survival between CLDme high and CLDme low patients was less pronounced than in the TCGA. We posit that this discrepancy may be due to differences in the ethnicity of the patients included in the two cohorts. The TCGA is composed of 43% Asian patients, while the validation cohort was collected in Spain, France and the United States so is likely to have a lower proportion of Asian patients. Secondly, the validation cohort had median AFP levels of 51, whereas the TCGA had a median value of 15. Elevated AFP is associated with the CpG island methylator phenotype in HCC so this may also impact the methylomes of patients in the validation cohort, affecting the accuracy of prediction.[68]

Other studies have shown the potential for methylation changes at specific gene promoters to predict hepatocarcinogenesis.[9] Therefore, an obvious extension of the work presented is to ask whether the CLD-associated methylation signature may have predictive as well as prognostic potential. To test the feasibility of this we performed the same array profiling on CLD tissue from six patients with decompensated liver disease, and had advanced CLD for > 10 years without HCC development. With this small cohort we were able to detect a trend (P = 0.059) towards lower CLDme score in non-progressing CLD, compared to HCC-associated CLD (**Fig. S11**).

While this remains to be validated in a larger cohort, these preliminary data indicate that a lack of this epigenetic dysregulation may be associated with a reduced risk of HCC emergence. While there was a strong inverse correlation between the methylation status of the identified DMRs and the expression of their associated genes, for genes such as *CYP2C19* and *TLDC2* the change in expression was disproportionate to the change in methylation. This observation may point to roles for other epigenetic mechanisms, such as altered patterns of histone modifications and chromatin organisation, in transcriptional regulation of these genes and in the progression of CLD to HCC. Indeed, ongoing research is focussed on the notion of the reversibility of changes to histone modifications occurring in the CLD-HCC transition, and other groups have shown the susceptibility of epigenetic reprogramming (H3K27ac in particular) to therapeutic intervention to prevent the onset of HCC in mice.[81]

5. Conclusions

In summary, we have shown that CLD and HCC samples from the same patient share broad transcriptional and epigenetic alterations which are compounded in HCC. Our results highlight how methylation changes in CLD may help not only to create a transcriptional landscape favourable for HCC emergence, but that the influence of these changes may extend to consequences for disease outcomes. The development of the CLDme score demonstrates that epigenetic changes occurring in CLD, and affecting both genes previously reported to be aberrantly methylated in CLD, as well as those we identify here, can be leveraged to predict HCC outcomes. Future studies will focus on identifying DNA methylation changes that may help identify CLD that would progress to HCC.

Data Availability: The data that support the findings of this study are available on request from the corresponding author.

Authors contributions: CKYN and SP conceived and supervised the study; JG performed the bioinformatic analysis; MCL, VP, SN, STM and GB performed nucleic acid extraction, immunohistochemistry and sequencing reactions; CE and LMT performed histopathologic review of the samples; AV, JML MAM, TB, SW, SPdP, SR-T, MvF, SDS, OK and MHH provided the samples and the clinical information included in the discovery cohort and critically discussed the results; AV and JML provided the samples included in the validation cohort and critically discussed the results; JG, MCL, CKYN and SP interpreted the results and wrote the manuscript.

Acknowledgements: None

Funding sources: L.M.T., C.K.Y.N. and S.P. were supported by the Swiss Cancer League (KLS-3639-02-2015, KFS-4543-08-2018, KFS-4988-02-2020-R, respectively; L.M.T., was supported by AIRC grant number IG 2019 Id.23615, S.P. is supported by The Professor Dr Max Cloëtta Foundation, from the University of Basel (Research Fund Junior Researchers), from the Krebsliga Beider Basel (KLbB-4473-03-2018), from the Theron Foundation, Vaduz (LI) and from the Surgery Department of the University Hospital Basel. M.H.H. was supported by ERC Synergy Grant 609883. J.M.L. is supported by grants from the European Commission (EC) Horizon 2020 Program (HEPCAR, proposal number 667273-2), the US Department of Defense (CA150272P3), the National Cancer Institute (P30 CA196521), the Samuel Waxman Cancer Research Foundation, the Spanish National Health Institute (MICINN, SAF-2016-76390 and PID2019-105378RB-I00), through a partnership between Cancer Research UK, Fondazione AIRC and Fundación Científica de la Asociacion Española Contra el Cáncer (HUNTER, Ref. C9380/A26813), and by the Generalitat de Catalunya (AGAUR, SGR-1358). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

A.V. has received 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. J.M.L. has received consulting fees from Eli Lilly, Bayer HealthCare Pharmaceuticals, Bristol-Myers Squibb, Eisai Inc, Celsion Corporation, Merck, Ipsen, Genentech, Roche, Glycotest, Nucleix, Sirtex, Mina Alpha Ltd and AstraZeneca; and research support from Bayer HealthCare Pharmaceuticals, Eisai Inc, Bristol-Myers Squibb, Boehringer-Ingelheim and Ipsen.

Figure Legends

Fig. 1: Oncogenic transcriptional alterations in CLD are compounded in HCC. A) Principal component analysis 15 healthy liver and 10 paired CLD and HCC samples. **B**) Venn diagrams of down- and upregulated genes in CLD and HCC compared to normal ($|\log_2 FC| > 1.5$ and $q < 0.05$). **C**) T statistics (absolute values) of 1,675 DE genes in both CLD and HCC compared to normal livers. DE: |log2FC| > 1.5, $q < 0.05$. *P* computed from paired Wilcoxon tests. **D**) Hallmark pathways with significaxnt enrichment in CLD, HCC or both from gene set enrichment analysis are shown (GSEA; $P < 0.05$). CLD: chronic liver disease; HCC: Hepatocellular carcinoma; DE: differentially expressed.

Fig. 2: DNA methylation alterations in HCC are detectable in CLD. A) Principal component (PC) analysis of 12 healthy liver samples, and ten paired CLD and HCC samples, showing PC1 and PC3. **B**,**C)** Differential methylation analysis ($-\log_{10}(q)$ against $\log_2$ fold-change (M value)) comparing CLD (**B**) and HCC (**C**) to healthy livers. Significant CpG sites ($|\log_2$ fold change| > 1.5, $q < 0.05$) are coloured according to the legend. **D**) T statistics (absolute values) of 339 differentially methylated CpG sites in both CLD and HCC compared to normal livers. DM: |log2FC| > 1.5, $q < 0.05$. *P* computed from paired Wilcoxon tests. **E**) Distribution of differentially methylated CpG sites (DMPs) according to their genomic features, detected in 10 CLD, 10 HCC compared to 12 normal livers. **F**) Venn diagram showing intersection of DMRs called in CLD and HCC samples compared to normal livers. PC: Principal component; CLD: chronic liver disease; HCC: Hepatocellular carcinoma; DM: differentially methylated; DMP: differentially methylated probes; DMR: differentially methylated region.

Fig. 3: Gene expression and DNA methylation changes define the transition from CLD to HCC. A) Heatmaps of methylation and gene expression at differentially expressed genes associated with DMRs (|mean change in methylation *B* value| > 0.15 and FDR < 0.05). Dot plots showing mean change in *B* value (methylation) and $\log_2$ fold change (gene expression) between CLD (n=10) and normal (n=12) and HCC (n=10) and normal (n=12). **B**) cg07554771 methylation and *MGMT* expression from 10 paired CLD (n=10) and HCC samples (n=10). *P* values computed from limma (methylation) and DESeq2 (gene expression). **C**) Representative immunohistochemistry images from two paired CLD and HCC biopsies stained with anti-MGMT antibody and MGMT protein expression IHC scores from 12 paired CLD and HCC samples (paired Wilcoxon test). Scale bar 20μm and 100μm. DMR: differentially methylated region; CLD: chronic liver disease; HCC: Hepatocellular carcinoma.

Fig. 4: DNA methylation alterations in CLD and HCC are conserved across cohorts. A) Principal component analysis of 12 healthy liver samples, and nine paired CLD and HCC samples from validation cohort, based on *B* values from the 51 CpG sites in the 18 DMRs associated with DE genes in the discovery cohort. **B**) Venn diagrams showing overlap between DM CpG sites in CLD and HCC in the discovery and validation cohorts, and the overlap between CLD-HCC DM CpG sites across cohorts. **C**) Heatmap of methylation across normal, CLD and HCC samples at 8 differentially expressed genes associated with DMRs in HCC and CLD samples across both cohorts. Dot plot on right shows mean difference in *B* values between CLD and normal and HCC and normal. **D**) Track plots showing *B* values at two CLD-HCC overlapping DMRs in *UBD* in validation and discovery cohorts. DNAse hypersensitivity sites are denoted in the bottom tracks with ENSEMBL gene annotation. CLD: chronic liver disease; HCC: Hepatocellular carcinoma; DMR: differentially methylated region; DE: differentially expressed; DM: differentially methylated; DMR: differentially methylated region.

Fig. 5: Implications of epigenetic priming in CLD on HCC. **A**) The approach used to select CpG sites for elastic net regression using TCGA data. **B-C**) Kaplan Meier plots of survival probability in the TCGA test set (**B**, n=111) and in an external validation cohort (**C**, n=246), stratified into High/Low CLDme score. Logrank *P* value adjusted for disease stage and history. **D**) Number of somatic mutations in TCGA samples stratified by CLDme score (163 High, 142 Low). Violin plots show distributions of mutations per sample. Boxplot shows the mean and interquartile range. Whiskers show the range of the data up to 1.5 x IQR. Samples outside this range are plotted as points. *P* value computed from Wilcoxon test. **E**) Barplot showing percentage of samples with *TP53* mutation types, stratified by CLDme (178 CLDme High, 183 CLDme Low). **F**) Extent of copy number alterations in TCGA samples stratified by CLDme score (178 CLDme High, 183 CLDme Low). Violin plots, boxplots and statistics as for **D**. TCGA: The Cancer Genome Atlas; CLDme: chronic liver disease methylation; IQR: inter quartile range.

Supplementary Figure Legends

Fig. S1: Schematic summarising the origin of samples used for methylation, transcriptomic, WES, and IHC analysis. WES: Whole exome sequencing; IHC: immunohistochemistry.

Fig. S2: Heatmap showing gene expression of top 500 most variably expressed genes across normal livers, CLDs and HCCs. Most variable genes were determined based on standard deviation across all samples. Dendrogram determined by consensus clustering (k means). Each row is a gene, and heatmap colours show z scaled, normalised expression. CLD: chronic liver disease; HCC: Hepatocellular carcinoma.

Fig. S3: Heatmap showing expression of genes differentially expressed in CLD, HCC or both, compared to normal livers. Each row is a gene, and heatmap colours show z scaled, normalised expression. CLD: chronic liver disease; HCC: Hepatocellular carcinoma.

Fig. S4: Genetic alterations detected in HCC are not present in matched CLD samples. Genetic alterations are not detectable in tumour-adjacent cirrhotic tissue. **A**) Summary of coding mutations in CLD and HCC samples from ten patients, affecting 20 frequently altered genes in HCC. The effects of the somatic alterations are color-coded according to the legend. **B**) Summary of copy number alterations affecting 20 genes showing frequent CNA in HCC. Samples are annotated with fibrosis stage (CLD samples) and HCC samples with Edmondson and BCLC stage. **C**) Representative genome-wide copy number plot (Patient 9). CLD: chronic liver disease; HCC: Hepatocellular carcinoma; CNA: copy number alteration.

Fig. S5: DNA methylation in CLD, HCC, and normal liver**. A**) Normal liver and CLD samples are separated from HCC samples by PC1. **B**) Hypermethylation of CpG sites associated with *RASSF1A* and *P14* in HCC. B values for 12 normal livers and 10 paired CLD and HCC samples. *P* values from moderated t-test, limma. CLD: chronic liver disease; HCC: Hepatocellular carcinoma;

Fig. S6: Venn diagram of CpG sites showing differential methylation in CLD and HCCs compared to normals. Sites are split according to the direction of methylation change. CLD: chronic liver disease; HCC: Hepatocellular carcinoma;

Fig. S7: Overlap between methyl-binding domain protein ChIP-seq data and CLD-HCC DMRs. CLD: chronic liver disease; HCC: Hepatocellular carcinoma; DMR: Differentially methylated region.

Fig. S8: Differentially methylated regions in CLD and HCC samples, compared to normal livers. Track plot of DMRs in *MGMT* shared between CLD and HCC and gained in HCC. DMRs were called by DMRcate with an FDR < 0.05 and change in B > |0.15|. CpG sites correlated with *MGMT* expression in Murphy *et al.* are highlighted. CLD: chronic liver disease; HCC: Hepatocellular carcinoma; DMR: Differentially methylated region; FDR: False discovery rate.

Fig. S9: Methylation changes in MGMT in CLD and HCC are conserved across cohorts. MGMT is hypermethylated in CLD and HCC compared to healthy livers. Methylation of cg07554771 in 12 healthy livers and 10 CLD and HCC pairs (discovery cohort) and a further 9 CLD - HCC pairs in the validation cohort. Wilcoxon Test. CLD: chronic liver disease; HCC: Hepatocellular carcinoma.

Fig. S10: Characterisation of CLDme High and Low TCGA samples. A) Volcano plot showing results of differential gene expression analysis comparing CLDme High to CLDme low tumours. 18 genes showing DE and associated with DMRs in the initial discovery samples are labelled. Dashed line shows significance threshold P < 0.05, DESeq2 Wald test. B) High and Low CLDme score TCGA samples show no difference in ImmuneScore. ImmuneScore was calculated for each sample as the sum of 11 cell types ('B-cells', 'CD4+T-cells', 'CD8+ T-cells' , 'DC', 'Eosinophils', 'Macrophages', 'Monocytes', 'Mast cells', 'Neutrophils' and 'NK cells') calculated by xCell using the expression data for 360 TCGA samples. Wilcoxon test ns = non-significant. C) Percentage of TCGA samples, stratified by CLDme score, with and without lymphocyte invasion. D) CLDme score in TCGA samples with and without lymphocyte invasion. Wilcoxon test. TCGA: The Cancer Genome Atlas; CLDme: chronic liver disease methylation; DE: Differentially expressed; DMR: differentially methylated region; CLDme: chronic liver disease methylation.

Fig. S11: CLDme scores in HCC, CLD, and non-progressing CLD (NPC). CLDme scores were generated for an additional 6 CLD patients with CLD for > 10 years, without HCC development, and compared with the CLD and HCC samples used throughout the study. Beta values were median centred before score generation to minimise cross-array batch effects. HCC n = 19, CLD n =19, NPC = 6. Wilcoxon Test. CLD: chronic liver disease; HCC: Hepatocellular carcinoma; CLDme: chronic liver disease methylation; NPC; non-progressing CLD

References

1 Villanueva A (2019) Hepatocellular Carcinoma. *N Engl J Med* **380**, 1450–1462.

2 Fujiwara N, Friedman SL, Goossens N & Hoshida Y (2018) Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J Hepatol* **68**, 526–549.

3 Iwahashi S, Shimada M, Morine Y, Imura S, Ikemoto T, Saito Y, Yamada S & Rui F (2019) The effect of hepatic stellate cells on hepatocellular carcinoma progression. *Journal of Clinical Oncology* **37**, 265–265.

4 Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M & Gores G (2016) Hepatocellular carcinoma. *Nat Rev Dis Primers* **2**, 16018.

5 Zucman-Rossi J, Villanueva A, Nault J-C & Llovet JM (2015) Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* **149**, 1226–1239.e4.

6 Beyaz S, Mana MD, Roper J, Kedrin D, Saadatpour A, Hong S-J, Bauer-Rowe KE, Xifaras ME, Akkad A, Arias E, Pinello L, Katz Y, Shinagare S, Abu-Remaileh M, Mihaylova MM, Lamming DW, Dogum R, Guo G, Bell GW, Selig M, Nielsen GP, Gupta N, Ferrone CR, Deshpande V, Yuan G-C, Orkin SH, Sabatini DM & Yilmaz ÖH (2016) High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58.

7 Vaz M, Hwang SY, Kagiampakis I, Phallen J, Patil A, O'Hagan HM, Murphy L, Zahnow CA, Gabrielson E, Velculescu VE, Easwaran HP & Baylin SB (2017) Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations. *Cancer Cell* **32**, 360–376.e6.

8 Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, Imbeaud S, Letouzé E, Hernandez-Gea V, Cornella H, Pinyol R, Solé M, Fuster J, Zucman-Rossi J, Mazzaferro V, Esteller M, Llovet JM & HEPTROMIC Consortium (2015) DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* **61**, 1945–1956.

9 Nishida N, Kudo M, Nagasaka T, Ikai I & Goel A (2012) Characteristic patterns of altered DNA methylation predict emergence of human hepatocellular carcinoma. *Hepatology* **56**, 994–1003.

10 Kaneto H, Sasaki S, Yamamoto H, Itoh F, Toyota M, Suzuki H, Ozeki I, Iwata N, Ohmura T, Satoh T, Karino Y, Satoh T, Toyota J, Satoh M, Endo T, Omata M & Imai K (2001) Detection of hypermethylation of the p16(INK4A) gene promoter in chronic hepatitis and cirrhosis associated with hepatitis B or C virus. *Gut* **48**, 372–377.

11 Murphy SK, Yang H, Moylan CA, Pang H, Dellinger A, Abdelmalek MF, Garrett ME, Ashley-Koch A, Suzuki A, Tillmann HL, Hauser MA & Diehl AM (2013) Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease.

Gastroenterology **145**, 1076–1087.

12 Hlady RA, Zhou D, Puszyk W, Roberts LR, Liu C & Robertson KD (2017) Initiation of aberrant DNA methylation patterns and heterogeneity in precancerous lesions of human hepatocellular cancer. *Epigenetics* **12**, 215–225.

13 Hernandez-Vargas H, Lambert M-P, Le Calvez-Kelm F, Gouysse G, McKay-Chopin S, Tavtigian SV, Scoazec J-Y & Herceg Z (2010) Hepatocellular carcinoma displays distinct DNA methylation signatures with potential as clinical predictors. *PLoS One* **5**, e9749.

14 Meunier L, Hirsch TZ, Caruso S, Imbeaud S, Bayard Q, Roehrig A, Couchy G, Nault J-C, Llovet JM, Blanc J-F, Calderaro J, Zucman-Rossi J & Letouzé E (2021) DNA Methylation Signatures Reveal the Diversity of Processes Remodeling Hepatocellular Carcinoma Methylomes. *Hepatology*.

15 Ng CKY, Di Costanzo GG, Tosti N, Paradiso V, Coto-Llerena M, Roscigno G, Perrina V, Quintavalle C, Boldanova T, Wieland S, Marino-Marsilia G, Lanzafame M, Quagliata L, Condorelli G, Matter MS, Tortora R, Heim MH, Terracciano LM & Piscuoglio S (2018) Genetic profiling using plasma-derived cell-free DNA in therapy-naïve hepatocellular carcinoma patients: a pilot study. *Ann Oncol* **29**, 1286–1291.

16 Llovet J, Brú C & Bruix J (1999) Prognosis of Hepatocellular Carcinoma: The BCLC Staging Classification. *Seminars in Liver Disease* **19**, 329–338.

17 Nuciforo S, Fofana I, Matter MS, Blumer T, Calabrese D, Boldanova T, Piscuoglio S, Wieland S, Ringnalda F, Schwank G, Terracciano LM, Ng CKY & Heim MH (2018) Organoid Models of Human Liver Cancers Derived from Tumor Needle Biopsies. *Cell Rep* **24**, 1363–1376.

18 Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.

19 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M & DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303.

20 Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ & Cheetham RK (2012) Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817.

21 Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES & Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213–219.

22 Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P, Chakravarty D, Phillips S, Kandoth C, Penson A, Gorelick A, Shamu T, Patel S, Harris C, Gao J, Sumer

SO, Kundra R, Razavi P, Li BT, Reales DN, Socci ND, Jayakumaran G, Zehir A, Benayed R, Arcila ME, Chandarlapaty S, Ladanyi M, Schultz N, Baselga J, Berger MF, Rosen N, Solit DB, Hyman DM & Taylor BS (2018) Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discov* **8**, 174–183.

23 Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, Zhang H, Solit DB, Taylor BS, Schultz N & Sander C (2017) 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med* **9**, 4.

24 Shen R & Seshan VE (2016) FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **44**, e131.

25 Bertucci F, Ng CKY, Patsouris A, Droin N, Piscuoglio S, Carbuccia N, Soria JC, Dien AT, Adnani Y, Kamal M, Garnier S, Meurice G, Jimenez M, Dogan S, Verret B, Chaffanet M, Bachelot T, Campone M, Lefeuvre C, Bonnefoi H, Dalenc F, Jacquet A, De Filippo MR, Babbar N, Birnbaum D, Filleron T, Le Tourneau C & André F (2019) Genomic characterization of metastatic breast cancers. *Nature* **569**, 560–564.

26 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

27 Li B & Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.

28 Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.

29 Korotkevich G, Sukhov V & Sergushichev A Fast gene set enrichment analysis. .

30 Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD & Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369.

31 Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ & Molloy PL (2015) De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6.

32 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W & Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47.

33 Mallik S, Odom GJ, Gao Z, Gomez L, Chen X & Wang L (2019) An evaluation of supervised methods for identifying differentially methylated regions in Illumina methylation arrays. *Brief Bioinform* **20**, 2224–2235.

34 Jaffe AE, Murakami P, Lee H, Leek JT, Daniele Fallin M, Feinberg AP & Irizarry RA (2012) Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies.

International Journal of Epidemiology **41**, 200–209.

35 ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

36 Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics* **47**.

37 Mounir M, Silva TC, Lucchetta M, Olsen C, Bontempi G, Noushmehr H, Colaprico A & Papaleo E Analyses of cancer data in the Genomic Data Commons Data Portal with new functionalities in the TCGAbiolinks R/Bioconductor package. .

38 Consortium TIP-CA of WG & The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020) Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93.

39 Kancherla V, Abdullazade S, Matter MS, Lanzafame M, Quagliata L, Roma G, Hoshida Y, Terracciano LM, Ng CKY & Piscuoglio S (2018) Genomic Analysis Revealed New Oncogenic Signatures in TP53-Mutant Hepatocellular Carcinoma. *Frontiers in Genetics* **9**.

40 Zou H & Hastie T (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301–320.

41 Friedman J, Hastie T & Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1–22.

42 Aran D, Hu Z & Butte AJ (2017) xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220.

43 Coto-Llerena M, Ercan C, Kancherla V, Taha-Mehlitz S, Eppenberger-Castori S, Soysal SD, Ng CKY, Bolli M, von Flüe M, Nicolas GP, Terracciano LM, Fani M & Piscuoglio S (2020) High Expression of FAP in Colorectal Cancer Is Associated With Angiogenesis and Immunoregulation Processes. *Front Oncol* **10**, 979.

44 Paradiso V, Garofoli A, Tosti N, Lanzafame M, Perrina V & Quagliata L (2018) Diagnostic targeted sequencing panel for hepatocellular carcinoma genomic screening. *J Mol Diagn* **20**, 836–848.

45 Avila MA, Berasain C, Torres L, Martín-Duce A, Corrales FJ, Yang H, Prieto J, Lu SC, Caballería J, Rodés J & Mato JM (2000) Reduced mRNA abundance of the main enzymes involved in methionine metabolism in human liver cirrhosis and hepatocellular carcinoma. *J Hepatol* **33**, 907–914.

46 Hernandez-Meza G, von Felden J, Gonzalez-Kozlova EE, Garcia-Lezana T, Peix J, Portela A, Craig AJ, Sayols S, Schwartz M, Losic B, Mazzaferro V, Esteller M, Llovet JM & Villanueva A (2020) DNA methylation profiling of human hepatocarcinogenesis. *Hepatology*.

47 Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW & Ehrlich M (1983) The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Research*

11, 6883–6894.

48 Lin CH, Hsieh SY, Sheen IS, Lee WC, Chen TC, Shyu WC & Liaw YF (2001) Genome-wide hypomethylation in hepatocellular carcinogenesis. *Cancer Res* **61**, 4238–4243.

49 Zhang C, Li J, Huang T, Duan S, Dai D, Jiang D, Sui X, Li D, Chen Y, Ding F, Huang C, Chen G & Wang K (2016) Meta-analysis of DNA methylation biomarkers in hepatocellular carcinoma. *Oncotarget* **7**, 81255–81267.

50 Mekky MA, Salama RH, Abdel-Aal MF, Ghaliony MA & Zaky S (2018) Studying the frequency of aberrant DNA methylation of APC, P14, and E-cadherin genes in HCV-related hepatocarcinogenesis. *Cancer Biomark* **22**, 503–509.

51 Jain S, Xie L, Boldbaatar B, Lin SY, Hamilton JP, Meltzer SJ, Chen S-H, Hu C-T, Block TM, Song W & Su Y-H (2015) Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis. *Hepatol Res* **45**, 1110–1123.

52 Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K & Beck S (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* **38**, 1378–1385.

53 Ashida R, Okamura Y, Ohshima K, Kakuda Y, Uesaka K, Sugiura T, Ito T, Yamamoto Y, Sugino T, Urakami K, Kusuhara M & Yamaguchi K (2018) The down-regulation of the CYP2C19 gene is associated with aggressive tumor potential and the poorer recurrence-free survival of hepatocellular carcinoma. *Oncotarget* **9**, 22058–22068.

54 Huynh H, Hao H-X, Chan SL, Chen D, Ong R, Soo KC, Pochanard P, Yang D, Ruddy D, Liu M, Derti A, Balak MN, Palmer MR, Wang Y, Lee BH, Sellami D, Zhu AX, Schlegel R & Huang A (2015) Loss of Tuberous Sclerosis Complex 2 (TSC2) Is Frequent in Hepatocellular Carcinoma and Predicts Response to mTORC1 Inhibitor Everolimus. *Molecular Cancer Therapeutics* **14**, 1224–1235.

55 Lee CGL, Ren J, Cheong ISY, Ban KHK, Ooi LLPJ, Yong Tan S, Kan A, Nuchprayoon I, Jin R, Lee K-H, Choti M & Lee LA (2003) Expression of the FAT10 gene is highly upregulated in hepatocellular carcinoma and other gastrointestinal and gynecological cancers. *Oncogene* **22**, 2592–2603.

56 Han M-Z, Xu R, Xu Y-Y, Zhang X, Ni S-L, Huang B, Chen A-J, Wei Y-Z, Wang S, Li W-J, Zhang Q, Li G, Li X-G & Wang J (2017) TAGLN2 is a candidate prognostic biomarker promoting tumorigenesis in human gliomas. *J Exp Clin Cancer Res* **36**, 155.

57 Wang Y-P, Huang L-Y, Sun W-M, Zhang Z-Z, Fang J-Z, Wei B-F, Wu B-H & Han Z-G (2013) Insulin receptor tyrosine kinase substrate activates EGFR/ERK signalling pathway and

promotes cell proliferation of hepatocellular carcinoma. *Cancer Lett* **337**, 96–106.

58 Villanueva A & Hoshida Y (2011) Depicting the role of TP53 in hepatocellular carcinoma progression. *J Hepatol* **55**, 724–725.

59 Um T-H, Kim H, Oh B-K, Kim MS, Kim KS, Jung G & Park YN (2011) Aberrant CpG island hypermethylation in dysplastic nodules and early HCC of hepatitis B virus-related human multistep hepatocarcinogenesis. *Journal of Hepatology* **54**, 939–947.

60 Nishida N, Nagasaka T, Nishimura T, Ikai I, Boland CR & Goel A (2008) Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. *Hepatology* **47**, 908–918.

61 Yang B, Guo M, Herman JG & Clark DP (2003) Aberrant Promoter Methylation Profiles of Tumor Suppressor Genes in Hepatocellular Carcinoma. *The American Journal of Pathology* **163**, 1101–1107.

62 Zhang Y-J, Wu H-C, Shen J, Ahsan H, Tsai WY, Yang H-I, Wang L-Y, Chen S-Y, Chen C-J & Santella RM (2007) Predicting hepatocellular carcinoma by detection of aberrant promoter methylation in serum DNA. *Clin Cancer Res* **13**, 2378–2384.

63 Wurmbach E, Chen Y-B, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J, Bottinger E, Friedman S, Waxman S & Llovet JM (2007) Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* **45**, 938–947.

64 Shi J, Ren M, She X, Zhang Z, Zhao Y, Han Y, Lu D & Lyu L (2020) Transgelin-2 contributes to proliferation and progression of hepatocellular carcinoma via regulating Annexin A2. *Biochemical and Biophysical Research Communications* **523**, 632–638.

65 Li R, Grimm SA, Chrysovergis K, Kosak J, Wang X, Du Y, Burkholder A, Janardhan K, Mav D, Shah R, Eling TE & Wade PA (2014) Obesity, rather than diet, drives epigenomic alterations in colonic epithelium resembling cancer progression. *Cell Metab* **19**, 702–711.

66 Liu L, Dong Z, Liang J, Cao C, Sun J, Ding Y & Wu D (2014) As an independent prognostic factor, FAT10 promotes hepatitis B virus-related hepatocellular carcinoma progression via Akt/GSK3β pathway. *Oncogene* **33**, 909–920.

67 Wu J, Wang S-T, Zhang Z-J, Zhou Q & Peng B-G (2018) CREB5 promotes cell proliferation and correlates with poor prognosis in hepatocellular carcinoma. *Int J Clin Exp Pathol* **11**, 4908–4916.

68 Zhang C, Li Z, Cheng Y, Jia F, Li R, Wu M, Li K & Wei L (2007) CpG island methylator phenotype association with elevated serum alpha-fetoprotein level in hepatocellular carcinoma. *Clin Cancer Res* **13**, 944–952.

69 Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, Higuera M, Pfister D,

Ramadori P, Pinyol R, Solé M, Heikenwälder M, Friedman SL, Sia D & Llovet JM (2019) An Immune Gene Expression Signature Associated With Development of Human Hepatocellular Carcinoma Identifies Mice That Respond to Chemopreventive Agents. *Gastroenterology* **157**, 1383–1397.e11.

70 Albillos A, Lario M & Álvarez-Mon M (2014) Cirrhosis-associated immune dysfunction: distinctive features and clinical relevance. *J Hepatol* **61**, 1385–1396.

71 Choi SS & Diehl AM (2009) Epithelial-to-mesenchymal transitions in the liver. *Hepatology* **50**, 2007–2013.

72 Bayard Q, Meunier L, Peneau C, Renault V, Shinde J, Nault J-C, Mami I, Couchy G, Amaddeo G, Tubacher E, Bacq D, Meyer V, La Bella T, Debaillon-Vesque A, Bioulac-Sage P, Seror O, Blanc J-F, Calderaro J, Deleuze J-F, Imbeaud S, Zucman-Rossi J & Letouzé E (2018) Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nature Communications* **9**.

73 Xu W, Wang N-R, Wang H-F, Feng Q, Deng J, Gong Z-Q, Sun J, Lou X-L, Yu X-F, Zhou L, Hu J-P, Huang X-F, Qi X-Q, Deng Y-J, Gong R, Guo Y, Wang M-M, Xiao J-C & Deng H (2016) Analysis of epithelial-mesenchymal transition markers in the histogenesis of hepatic progenitor cell in HBV-related liver diseases. *Diagnostic Pathology* **11**.

74 Pascale RM, Simile MM, Satta G, Seddaiu MA, Daino L, Pinna G, Vinci MA, Gaspa L & Feo F (1991) Comparative effects of L-methionine, S-adenosyl-L-methionine and 5'-methylthioadenosine on the growth of preneoplastic lesions and DNA methylation in rat liver during the early stages of hepatocarcinogenesis. *Anticancer Res* **11**, 1617–1624.

75 Cai J, Mao Z, Hwang JJ & Lu SC (1998) Differential expression of methionine adenosyltransferase genes influences the rate of growth of human hepatocellular carcinoma cells. *Cancer Res* **58**, 1444–1450.

76 Iwakuma T, Sakumi K, Nakatsuru Y, Kawate H, Igarashi H, Shiraishi A, Tsuzuki T, Ishikawa T & Sekiguchi M (1997) High incidence of nitrosamine-induced tumorigenesis in mice lacking DNA repair methyltransferase. *Carcinogenesis* **18**, 1631–1635.

77 Li C-C, Yu Z, Cui L-H, Piao J-M & Liu M (2014) Role of P14 and MGMT gene methylation in hepatocellular carcinomas: a meta-analysis. *Asian Pac J Cancer Prev* **15**, 6591–6596.

78 Zhang Y-J, Chen Y, Ahsan H, Lunn RM, Lee P-H, Chen C-J & Santella RM (2003) Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation and its relationship to aflatoxin B1-DNA adducts and p53 mutation in hepatocellular carcinoma. *Int J Cancer* **103**, 440–444.

79 Matsukura S, Soejima H, Nakagawachi T, Yakushiji H, Ogawa A, Fukuhara M, Miyazaki K, Nakabeppu Y, Sekiguchi M & Mukai T (2003) CpG methylation of MGMT and hMLH1

promoter in hepatocellular carcinoma associated with hepatitis viral infection. *Br J Cancer* **88**, 521–529.

80 Herath NI, Walsh MD, Kew M, Smith JL, Jass JR, Young J, Leggett BA & Macdonald GA (2007) Silencing of O6-methylguanine DNA methyltransferase in the absence of promoter hypermethylation in hepatocellular carcinomas from Australia and South Africa. *Oncol Rep* **17**, 817–822.

81 Jühling F, Hamdane N, Crouchet E, Li S, El Saghire H, Mukherji A, Fujiwara N, Oudot MA, Thumann C, Saviano A, Roca Suarez AA, Goto K, Masia R, Sojoodi M, Arora G, Aikata H, Ono A, Tabrizian P, Schwartz M, Polyak SJ, Davidson I, Schmidl C, Bock C, Schuster C, Chayama K, Pessaux P, Tanabe KK, Hoshida Y, Zeisel MB, Duong FH, Fuchs BC & Baumert TF (2021) Targeting clinical epigenetic reprogramming for chemoprevention of metabolic and viral hepatocellular carcinoma. *Gut* **70**, 157–169.

Supplementary Information

Supplementary Table Legends

Table S1:

Sheet A: Cohort overview
Sheet B: Differential gene expression - CLD - Normal
Sheet C: Differential gene expression - HCC - Normal
Sheet D: Genes DE in CLD and HCC compared to Normal
Sheet E: Coding alternations in CLD and HCC pairs
Sheet F: Copy number alterations in CLD and HCC pairs
Sheet G: CpG sites DM in CLD compared to Normal
Sheet H: CpG sites DM in HCC compared to Normal
Sheet I: CpG sites DM in both CLD and HCC compared to Normal
Sheet J: DMRs in both CLD and HCC compared to Normal, associated with DE genes
Sheet K: CpG sites falling in DMRs from Table J
Sheet L: Annotation of 109 probes used in prognostic model
Sheet M: Univariate analysis of clinical features and survival in TCGA HCC cohort

Supplementary Figure Legends

Fig. S1: Schematic summarising the origin of samples used for methylation, transcriptomic, WES, and IHC analysis.

Fig. S2: Heatmap showing gene expression of top 500 most variably expressed genes across normal livers, CLDs and HCCs.

Fig. S3: Heatmap showing expression of genes differentially expressed in CLD, HCC or both, compared to normal livers.

Fig. S4: Genetic alterations detected in HCC are not present in matched CLD samples.

Fig. S5: DNA methylation in CLD, HCC, and normal liver**.**

Fig. S6: Venn diagram of CpG sites showing differential methylation in CLD and HCCs compared to normals.

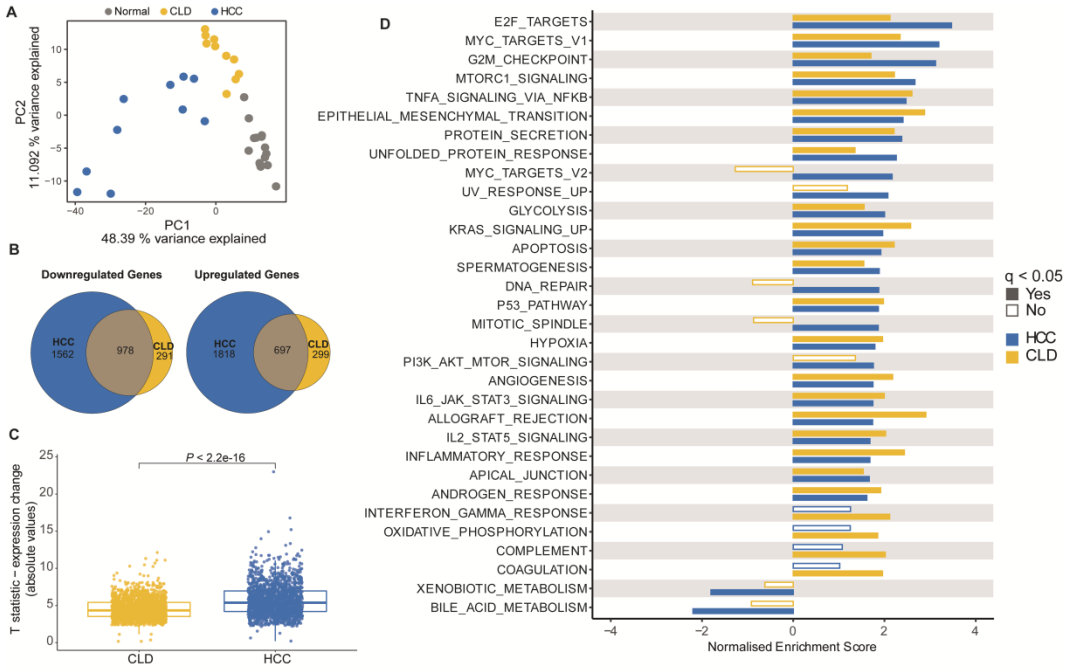
Fig. S7: Overlap between methyl-binding domain protein ChIP-seq data and CLD-HCC DMRs.

Fig. S8: Differentially methylated regions in CLD and HCC samples, compared to normal livers.

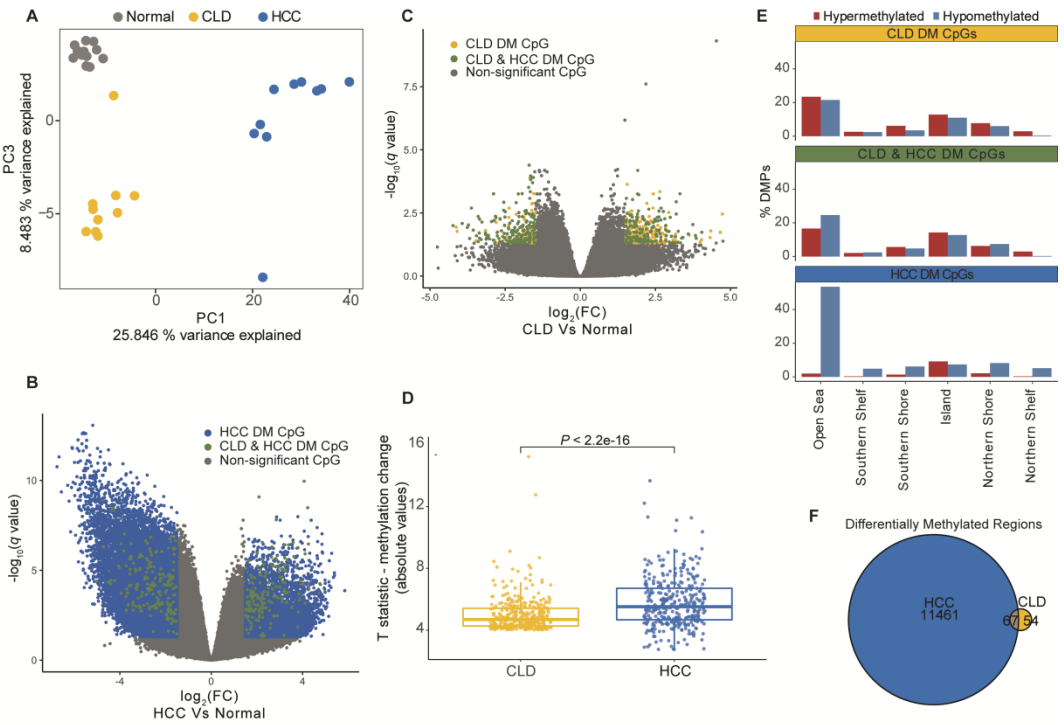
Fig. S9: Methylation changes in MGMT in CLD and HCC are conserved across cohorts.

Fig. S10: Characterisation of CLDme High and Low TCGA samples.

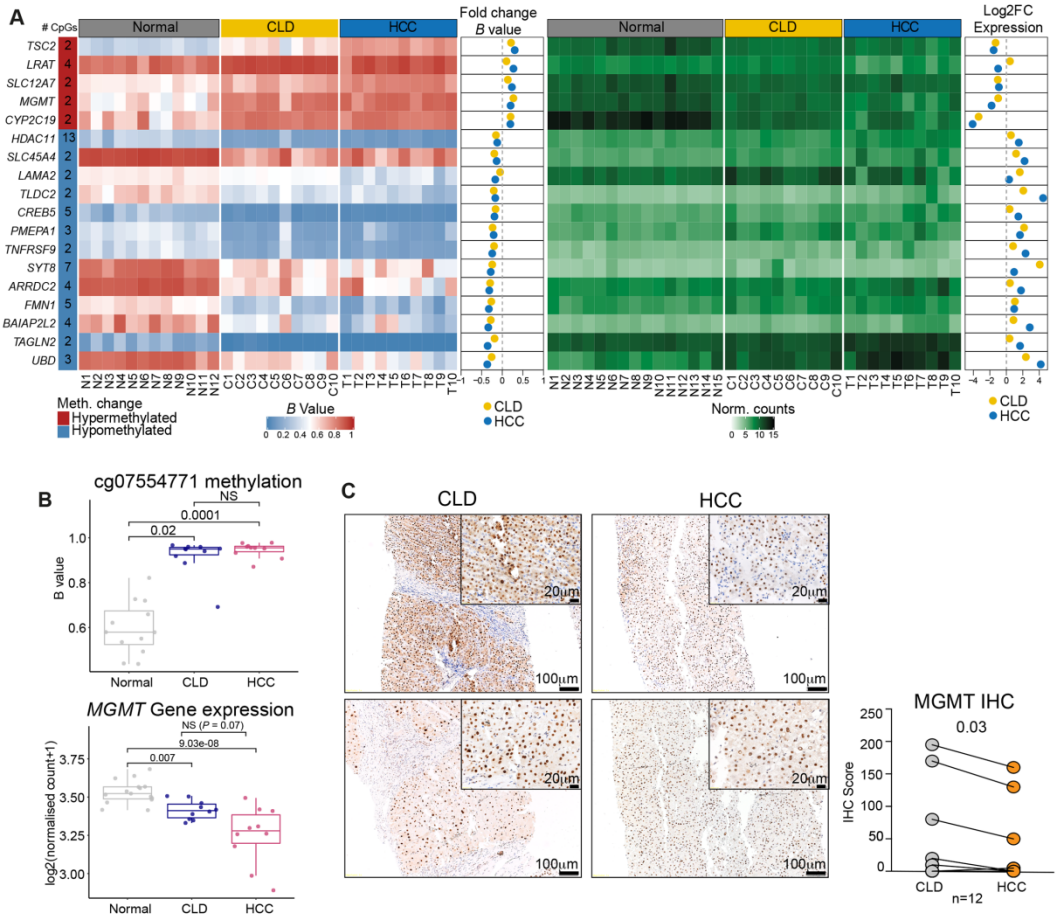
Fig. S11: CLDme scores in HCC, CLD, and non-progressing CLD (NPC).



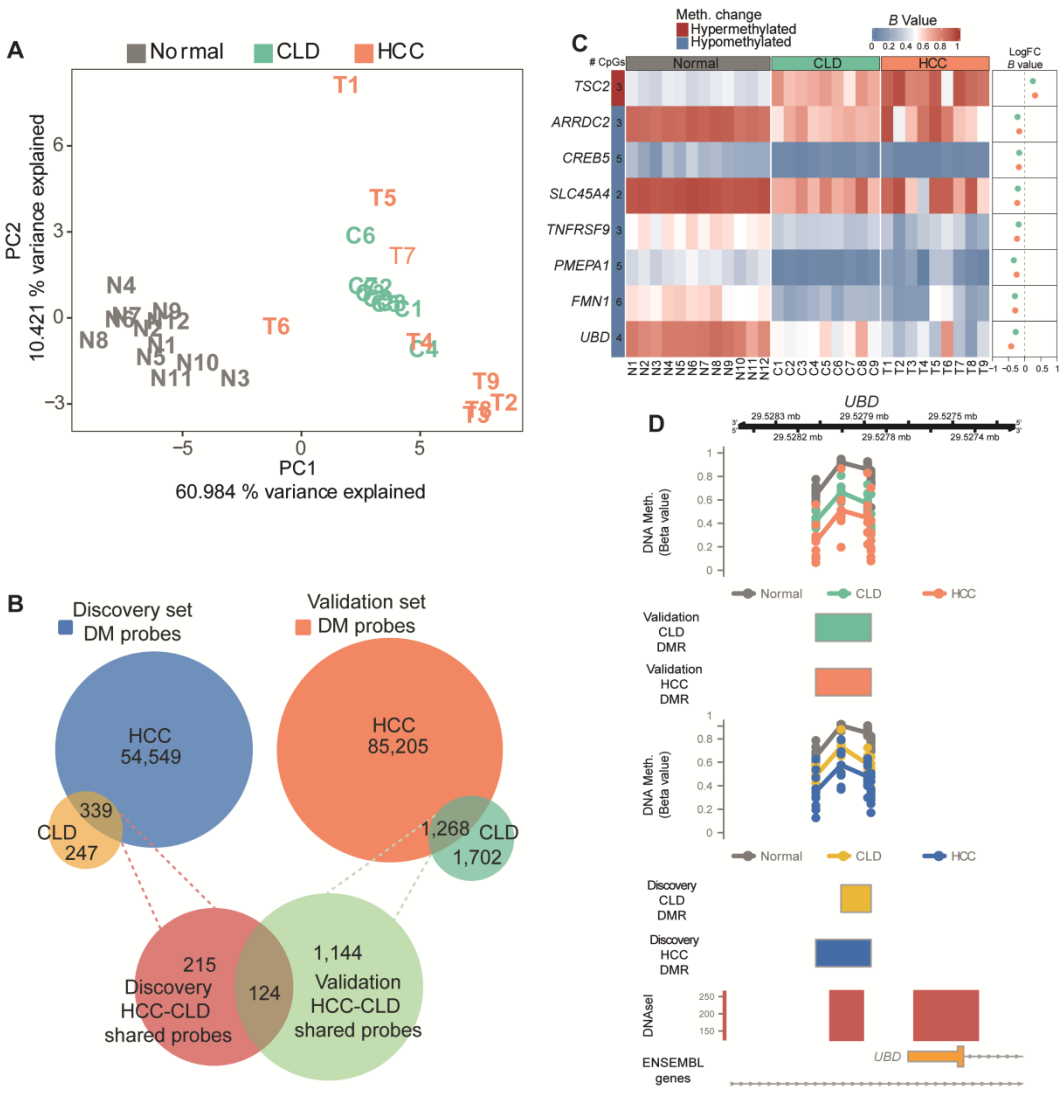
mol2_13154_f1.tif



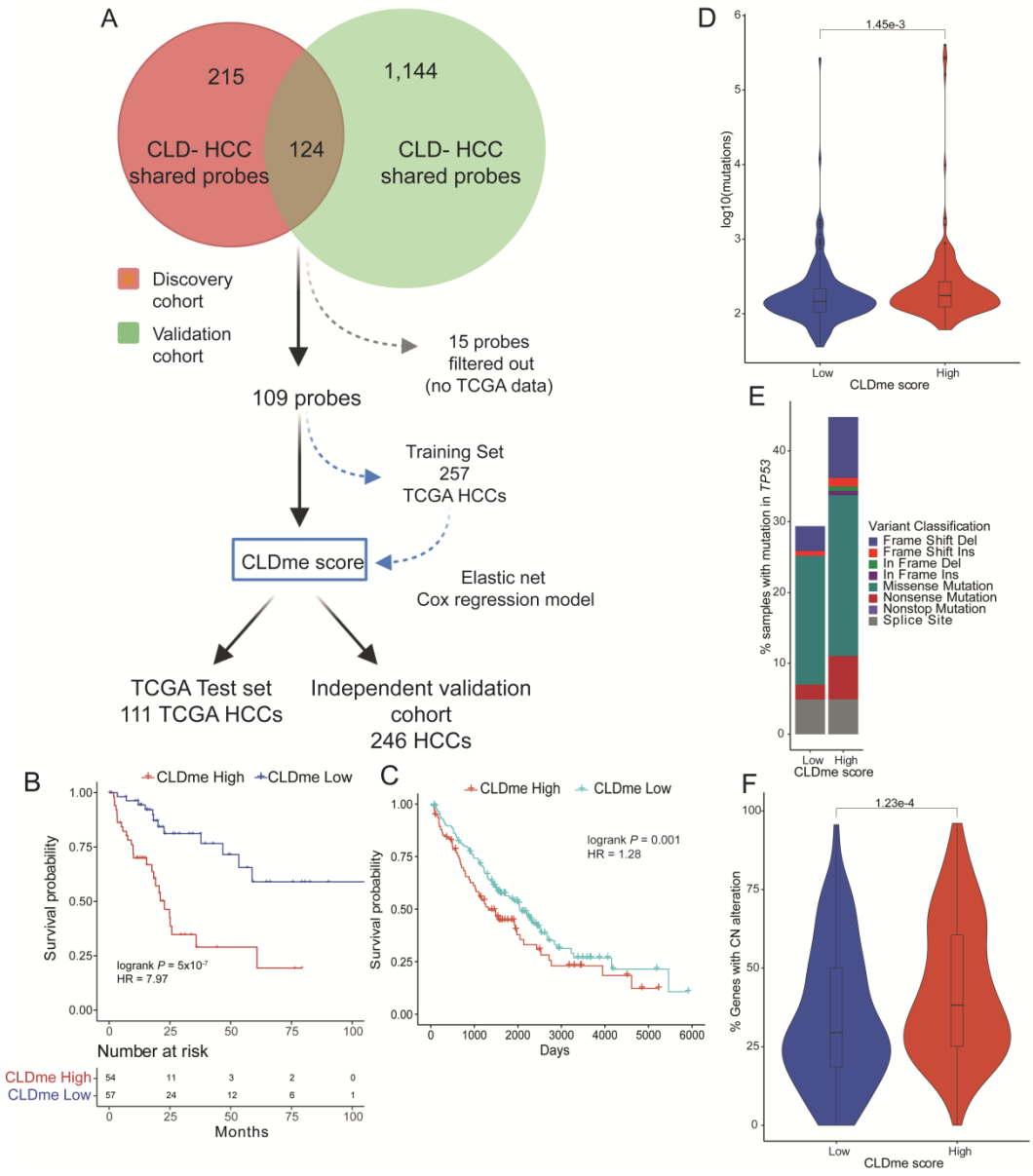
mol2_13154_f2.tif



mol2_13154_f3.tif



mol2_13154_f4.tif



mol2_13154_f5.tif