A genomic enhancer signature associates with hepatocellular carcinoma prognosis

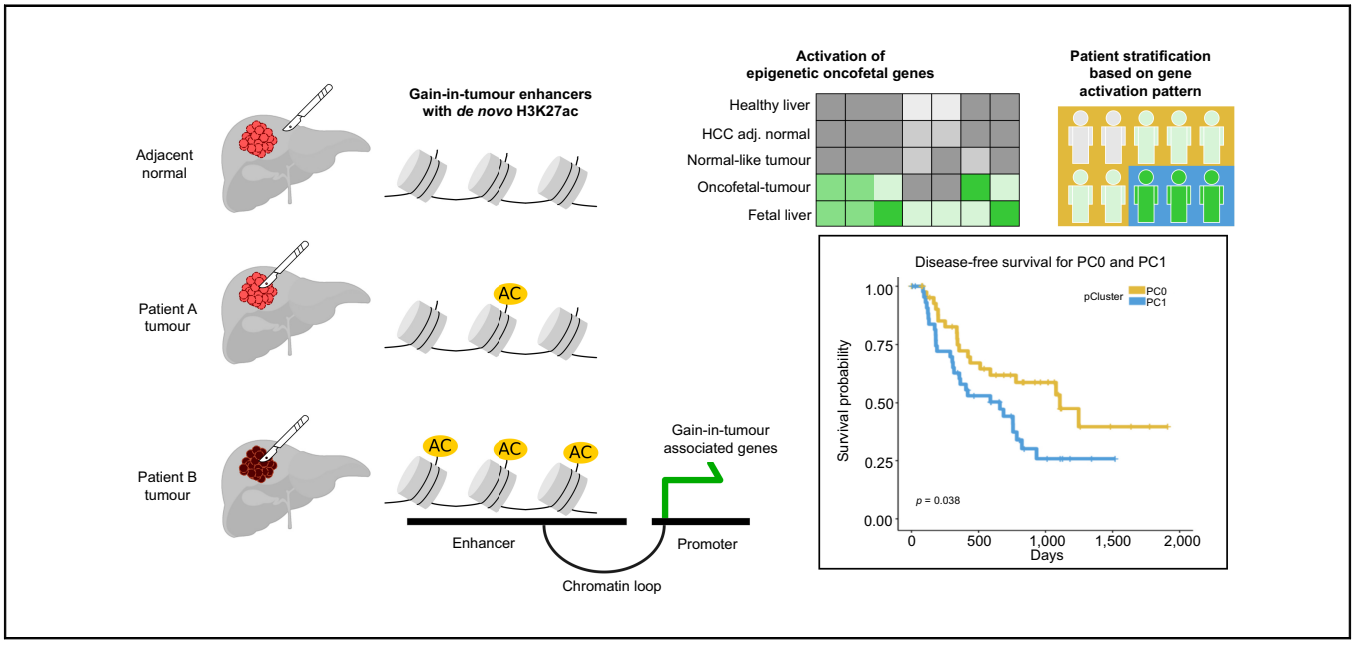
Authors

Ah-Jung Jeon, Chukwuemeka George Anene-Nzelu, Yue-Yang Teo, Shay Lee Chong, Karthik Sekar, Lingyan Wu, Sin-Chi Chew, Jianbin Chen, Raden Indah Kendarsari, Hannah Lai, Wen Huan Ling, Neslihan Arife Kaya, Jia Qi Lim, Alexander Yaw Fui Chung, Peng-Chung Cheow, Juinn Huar Kam, Krishnakumar Madhavan, Alfred Kow, Iyer Shridhar Ganpathi, Tony Kiat Hon Lim, Wei-Qiang Leow, Shihleone Loong, Tracy Jiezhen Loh, Wei Keat Wan, Gwyneth Shook Ting Soon, Yin Huei Pang, Boon Koon Yoong, Diana Bee-Lan Ong, Jasmine Lim, Vanessa H. de Villa, Rouchelle D. dela Cruz, Rawisak Chanwat, Jidapa Thammasiri, Glenn K. Bonney, Brian K.P. Goh, Roger Sik Yin Foo, Pierce Kah-Hoe Chow

Correspondence

pierce.chow@duke-nus.edu.sg (P.K.-H. Chow).

Graphical abstract



Highlights

- Epigenetic dysregulation is prevalent in HCC, accompanied by *de novo* enhancers.
- Differential enhancers and the associated gene expression changes are heterogeneous.
- Differential enhancer genes included cellular proliferation and foetal liver markers.
- Patients with tumour-activated gain-in-tumour enhancer genes showed worse prognosis.

Impact and Implications

Lifestyle and environmental-related exposures are the important risk factors of hepatocellular carcinoma (HCC), suggesting that tumour-associated epigenetic dysregulations may significantly underpin HCC. We profiled tumour tissues and their matched normal from 30 patients with early-stage HCC to study the dysregulated epigenetic changes associated with HCC. By also analysing the patients' RNA-seq and clinical data, we found the signature genes – with epigenetic and transcriptomic dysregulation – associated with worse prognosis. Our findings suggest that systemic approaches are needed to consider the surrounding cellular environmental and epigenetic changes in HCC tumours.

https://doi.org/10.1016/j.jhepr.2023.100715

Research article

JHEP Reports

A genomic enhancer signature associates with hepatocellular carcinoma prognosis

Ah-Jung Jeon,[1] Chukwuemeka George Anene-Nzelu,[2,3,4] Yue-Yang Teo,[1] Shay Lee Chong,[1] Karthik Sekar,[1] Lingyan Wu,[1] Sin-Chi Chew,[1] Jianbin Chen,[5] Raden Indah Kendarsari,[5] Hannah Lai,[5] Wen Huan Ling,[1] Neslihan Arife Kaya,[5] Jia Qi Lim,[5] Alexander Yaw Fui Chung,[6,7] Peng-Chung Cheow,[6,7] Juinn Huar Kam,[6,7] Krishnakumar Madhavan,[8] Alfred Kow,[8] Iyer Shridhar Ganpathi,[8] Tony Kiat Hon Lim,[9] Wei-Qiang Leow,[9] Shihleone Loong,[9] Tracy Jiezhen Loh,[9] Wei Keat Wan,[9] Gwyneth Shook Ting Soon,[10] Yin Huei Pang,[10] Boon Koon Yoong,[11] Diana Bee-Lan Ong,[11] Jasmine Lim,[11] Vanessa H. de Villa,[12] Rouchelle D. dela Cruz,[13] Rawisak Chanwat,[14] Jidapa Thammasiri,[15] Glenn K. Bonney,[8] Brian K.P. Goh,[6,7] Roger Sik Yin Foo,[2,5,16] Pierce Kah-Hoe Chow[1,6,7,*]

*[1]Program in Clinical and Translational Liver Cancer Research, Division of Medical Science, National Cancer Center Singapore, Singapore; [2]Cardiovascular Disease Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; [3]Montreal Heart Institute, Montreal, QC, Canada; [4]Department of Medicine, University of Montreal, QC, Canada; [5]Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore; [6]Department of Hepatopancreatobiliary and Transplant Surgery, National Cancer Centre Singapore and Singapore General Hospital, Singapore; [7]Academic Clinical Programme for Surgery, Duke-NUS Medical School, Singapore; [8]Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, University Surgical Cluster, National University Health System, Singapore; [9]Department of Anatomical Pathology, Singapore General Hospital, Singapore; [10]Department of Pathology, National University Hospital, Singapore; [11]Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; [12]Department of Surgery and Center for Liver Health and Transplantation, The Medical City, Pasig City, Philippines; [13]Department of Laboratory Medicine and Pathology, The Medical City, Pasig City, Philippines; [14]Hepato-Pancreato-Biliary Surgery Unit, Department of Surgery, National Cancer Institute, Bangkok, Thailand; [15]Division of Pathology, National Cancer Institute, Thailand; [16]Cardiovascular Research Institute, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*

JHEP Reports **2023.** https://doi.org/10.1016/j.jhepr.2023.100715

Background & Aims: Lifestyle and environmental-related exposures are important risk factors for hepatocellular carcinoma (HCC), suggesting that epigenetic dysregulation significantly underpins HCC. We profiled 30 surgically resected tumours and the matched adjacent normal tissues to understand the aberrant epigenetic events associated with HCC.
Methods: We identified tumour differential enhancers and the associated genes by analysing H3K27 acetylation (H3K27ac) chromatin immunoprecipitation sequencing (ChIP-seq) and Hi-C/HiChIP data from the resected tumour samples of 30 patients with early-stage HCC. This epigenome dataset was analysed with previously reported genome and transcriptome data of the overlapping group of patients from the same cohort. We performed patient-specific differential expression testing using multiregion sequencing data to identify genes that undergo both enhancer and gene expression changes. Based on the genes selected, we identified two patient groups and performed a recurrence-free survival analysis.
Results: We observed large-scale changes in the enhancer distribution between HCC tumours and the adjacent normal samples. Many of the gain-in-tumour enhancers showed corresponding upregulation of the associated genes and *vice versa*, but much of the enhancer and gene expression changes were patient-specific. A subset of the upregulated genes was activated in a subgroup of patients' tumours. Recurrence-free survival analysis revealed that the patients with a more robust upregulation of those genes showed a worse prognosis.
Conclusions: We report the genomic enhancer signature associated with differential prognosis in HCC. Findings that cohere with oncofoetal reprogramming in HCC were underpinned by genome-wide enhancer rewiring. Our results present the epigenetic changes in HCC that offer the rational selection of epigenetic-driven gene targets for therapeutic intervention or disease prognostication in HCC.
Impact and Implications: Lifestyle and environmental-related exposures are the important risk factors of hepatocellular carcinoma (HCC), suggesting that tumour-associated epigenetic dysregulations may significantly underpin HCC. We profiled tumour tissues and their matched normal from 30 patients with early-stage HCC to study the dysregulated epigenetic changes associated with HCC. By also analysing the patients' RNA-seq and clinical data, we found the signature genes – with epigenetic and transcriptomic dysregulation – associated with worse prognosis. Our findings suggest that systemic approaches are

Keywords: Epigenetics; Multi-omics; Hepatocellular carcinoma; Cancer prognosis; Personalised medicine.
Received 6 September 2022; received in revised form 20 January 2023; accepted 9 February 2023; available online 26 February 2023
* Corresponding author. Address: National Cancer Centre Singapore, 30 Hospital Boulevard, Singapore 168583, Singapore. Tel.: +65-63065424
E-mail address: pierce.chow@duke-nus.edu.sg (P.K.-H. Chow).

needed to consider the surrounding cellular environmental and epigenetic changes in HCC tumours.

Introduction

The cancer genome has revealed important insights, as evident from the widely accessed cancer genome databases, including The Cancer Genome Atlas (TCGA; http://tcga-data.nci.nih.gov/) and the International Cancer Genome Consortium (https://dcc.icgc.org/). The key driver mutations have thus far been identified for different types of cancer.[1] In addition to genomic disruptions, epigenetic dysregulation is also a key to tumorigenesis, with early efforts primarily focusing on DNA methylation.[2] Other genome-wide techniques of chromatin immunoprecipitation sequencing (ChIP-seq) and Hi-C now elaborate on enhancer dysregulation and reveal new insights on distinct cancer-type specific enhancer patterns.[3]

Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer and the third leading cause of cancer death worldwide in 2020.[4] Well-known risk factors include chronic infection with HBV or HCV, aflatoxin-contaminated foods, heavy alcohol consumption, and type 2 diabetes.[5] However, major risk factors for liver cancer appear to be shifting, given a declining prevalence of HBV or HCV and a corresponding increase in the prevalence of excess body weight and diabetes as risk factors for HCC in many countries.[6] The latter factors support the view that epigenetic dysregulation may significantly underpin HCC and, thus, the importance of studying the epigenome profiles.

Changes in histone marks and gene expression in liver cancer cells have been previously described *in vitro*.[7] Immunohistochemical analysis of H3K27 modifications – acetylation or methylation – in HCC showed variability and also a correlation to the degree of cellular de-differentiation and disease prognosis.[8] However, the enhancer landscape using H3K27 acetylation (H3K27ac) ChIP-seq and Hi-C has never been assessed using human samples from patients with HCC. Among the different regulatory elements and histone marks, active enhancers marked by H3K27ac are highly cell type and cell state dependent.[9,10] Therefore, to investigate the individuality and heterogeneity of HCC-related enhancers, we profiled H3K27ac using HCC liver tissues from 30 patients and mapped distal interacting genes by chromatin loop analysis from contemporaneous Hi-C profiles. By correlating the epigenomes of HCC samples to their corresponding genomes and transcriptomes, we analysed the functional and clinical implications of the dysregulated enhancers. A set of signature genes with correlated epigenetic changes proved useful in stratifying patient prognosis.

Patients and methods

Patient recruitment, sample preparation, and sequencing

Thirty patients were recruited from three local hospitals (National Cancer Centre Singapore, Singapore General Hospital, and National University Hospital) collaborating under the auspices of the Asia-Pacific Hepatocellular Carcinoma trial group.[11] The inclusion and exclusion criteria used during the patient recruitment are shown in Table S10. This included 12 patients enrolled in the Translational and Clinical Research (TCR) Flagship Programme: Precision Medicine in Liver Cancer across an Asia-pacific NETwork (PLANet; NCT03267641), funded by the Singapore National Medical Research Council. This study has been approved by Singhealth Centralised Institutional Review Board (2016/2626 and 2018/2112). Informed consent was taken from each patient before enrolment. These patients had undergone liver resection, and grid sampling with multiregion sampling was performed on the tumours as previously described.[11,12]

RNA-sequencing (RNA-seq) libraries were prepared for sequencing and processed using the method and pipeline described previously.[11] ChIP-seq was performed as previously described,[13] whereas Hi-C and HiChIP were performed using the Arima Hi-C Protocol described in the Arima Hi-C Kit (material part number: A410110; document part number: A160162 v00) (Arima Genomics, Carlsbad, United States). The details of the library preparation are provided in Supplementary methods 1.0.

Sources of data from previously published work

Of the 90 patients who contributed transcriptome data, the genome and transcriptome data of 44 patients have been previously reported by Zhai *et al.*[12]. The rest of the transcriptome and clinical data were reported by Jeon *et al.*[14]. The median time to follow-up was 839 days.

H3K27ac ChIP-seq data processing

The reads in ChIP-seq data were first trimmed off the adapter and index sequences using BBDuk from BBTools (DOE Joint Genome Institute, Berkeley, United States).[15] Read alignment and index generation were done using Bowtie2[16] and Samtools.[17] Duplicated reads and unmapped reads were removed using Sambamba.[18] DeepTools bamCoverage[19] was used to generate bigwig files for visualisation, and bedtools bamtobed was used to generate BED (Browser Extensible Data) files.

Hi-C/HiChIP data processing

Hi-C and H3K27ac HiChIP libraries were analysed using the Juicer pipeline.[20] Loops were called with HiCCUPS at 10-kB resolution. Contact domains were called at 50-kB resolution, and differential topologically associated domains (TADs) were analysed using TADCompare.[21]

Enhancer state calling and annotation

ChromHMM[22] was used to annotate the genome with a significant H3K27ac signal ('enhancer locus') for each sample (see Supplementary methods 2.0 for the details).

Differential enhancer testing

DESeq2[23] was used to identify enhancer loci with significant differences in H3K27ac ChIP-seq coverage between the tumour and adjacent normal (adj.normal) samples (Supplementary methods 3.0). An adjusted p value less than 0.05 and the absolute value of $\log_2$ fold change greater than 1 were used to identify differential enhancers.

Promoter–enhancer association

We filtered enhancer loci based on either (1) the enhancer locus overlaps with the promoter region ('promoter enhancer'), or (2) any of its interacting locus based on detected chromatin loops overlaps with the promoter region ('distal enhancer'). The promoter regions were inferred from the known transcription

start site loci (downloaded from University of California Santa Cruz (UCSC) genome browser Table Browser[24] for all known genes in hg38) to 500 bp upstream. For the first category, we used bedtools intersect to identify enhancer loci occurring directly at the promoters. For the second category, we merged all loops called from Hi-C and HiChIP data in BEDPE, and intersected first with the enhancer loci BED and then again with the promoter regions using bedtools pairToBed. The transcript ID assigned to each promoter locus was used to map the gene to the enhancer.

Visualisation of ChIP-seq and Hi-C data

Normalised ChIP-seq signal with the respective input control was obtained using deepTools bamCompare.[19] The sushi package[25] in R (R Foundation for Statistical Computing, Vienna, Austria) was used to plot coverage and chromatin loops. Computematrix and plotHeatmap from deepTools were used to generate ChIP-seq heatmaps. For visualisation, the heatmaps were limited to those enhancer loci with a width below 10 kbp. The midpoint of each enhancer locus was taken and extended by 5 kbp upstream and downstream for heatmap generation.

Pie charts in Fig. 1A were generated using the ChIPseeker package in R.[26]

Single-cell RNA-sequencing analysis

Processed single-cell RNA-sequencing (scRNA-seq) data were obtained from Sharma *et al.*[27]. We first filtered cells in clusters identified as hepatocytes and endothelial cells. We binarised the normalised expression counts by fitting a bimodal distribution. We followed the previously published work for the overall methodology for bimodal gene detection.[28] Diptest[29] was used to test for unimodality using Hartigan's dip test, and mixtools[30] was used to fit the multimodal distribution. For each gene, the binarising threshold was determined by the expression value where the minimum point of density occurs. For each patient, we then calculated the percentage of cells expressing the gene of interest based on the binarised gene expression counts.

DNA binding motif analysis at differential enhancer loci

We performed DNA binding motif analysis at differential enhancer loci using the findMotifsGenome function in HOMER,[31] with the consensus enhancer loci as the background loci and with the following parameters: '-size 800 -preparse -bg -noweight'. The hTFtarget database was used to infer the target genes corresponding to the motif.[32]

DNA mutation hotspot detection at differential enhancer loci

The non-coding module of MutEnricher[33] was used with default parameters.

Patient stratification using per-patient differential expression values

Log fold-change matrix is obtained from per-patient analysis and simplified into a categorical matrix with three values, namely -1, 0, or 1, representing the downregulation, no change, or upregulation of the gene in the tumour samples compared with the patient's normal samples, respectively. Co-clustering of this categorical matrix is performed using the blockcluster R package, which stratifies the patients into two patient clusters.

Analysis of TCGA-LIHC RNA-seq dataset

The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) data were downloaded from the cBioportal.[34] The data were filtered with patient criteria, and co-clustering was performed using the blockcluster package (Supplementary methods 4.0).

Recurrence-free survival analysis

Recurrence-free survival analysis is performed using the survival and survminer R packages.[35] All *p* values are obtained using the Tarone–Ware test. Disease-free survival analysis was done using recurrence-free survival days in our dataset, whereas progression-free interval days were used for the TCGA dataset, as recommended by TCGA guidelines.

Results

Multi-omics data collection from patients with early resectable HCC

We performed epigenome profiling for genomic enhancers using H3K27ac ChIP-seq and mapped gene-enhancer interactions using Hi-C or H3K27ac HiChIP on tumour and adj.normal tissues of 30 surgically resected HCC patients. The patients were part of a larger cohort whose genome and transcriptome data were previously published,[11,12,14] together with scRNA-seq data.[27] Overall, we used bulk RNA-seq and clinical data from 90 patients, of which 14 had scRNA-seq data and 30 had epigenome profiles. An overview of the multi-omics data collected and the overlap of patient groups between different profiling subsets are summarised in the schematic diagram in Fig. 2A and Table S1.

Differential enhancer loci are prevalent in HCC tumours

We first constructed a set of consensus enhancer loci for HCC tumours (Fig. 2B). The chromatin state with the highest emission probability of 0.935 for H3K27ac signal (state E5) was chosen as the enhancer chromatin state (Fig. 2C). H3K27ac ChIP-seq signals were then quantified at each locus. The quantified count matrix was then used for differential enhancer testing (see Patients and methods). We identified 1,650 gain-in-tumour enhancer loci that showed a stronger H3K27ac signal in the tumour, compared with adj.normal samples (adjusted p <0.05; $\log_2$ fold change >1). Similarly, 1,415 lost-in-tumour enhancer loci were identified ($\log_2$ fold change <-1). Fig. 2E shows H3K27ac coverage at differential enhancers for a subset of samples (combined profiles in Fig. 2D; a heatmap for the other enhancer landscape in Fig. S1). About 23.4% (717/3,065) of the differential enhancer loci overlapped with H3K27ac chromatin states of the normal liver by the ROADMAP epigenomics project.[36,37] The largest overlap was with the Weak Enhancer state (Table S2), indicating the emergence of HCC-specific and patient-specific enhancer loci present in our dataset. Loci for gain-in-tumour and lost-in-tumour enhancers are listed in Table S3 and S4, respectively.

Differential enhancers overlapped with various genomic elements

Overall, the consensus enhancer loci showed similar proportion of overlap to gene promoters (23.23%) and to distal intergenic regions (24.76%) (Fig. 1A). Gain-in-tumour enhancers, however, showed larger overlap to distal intergenic loci, whereas lost-in-tumour enhancers occurred more at the gene promoters. Hence, gained H3K27ac marks accumulated more at distal

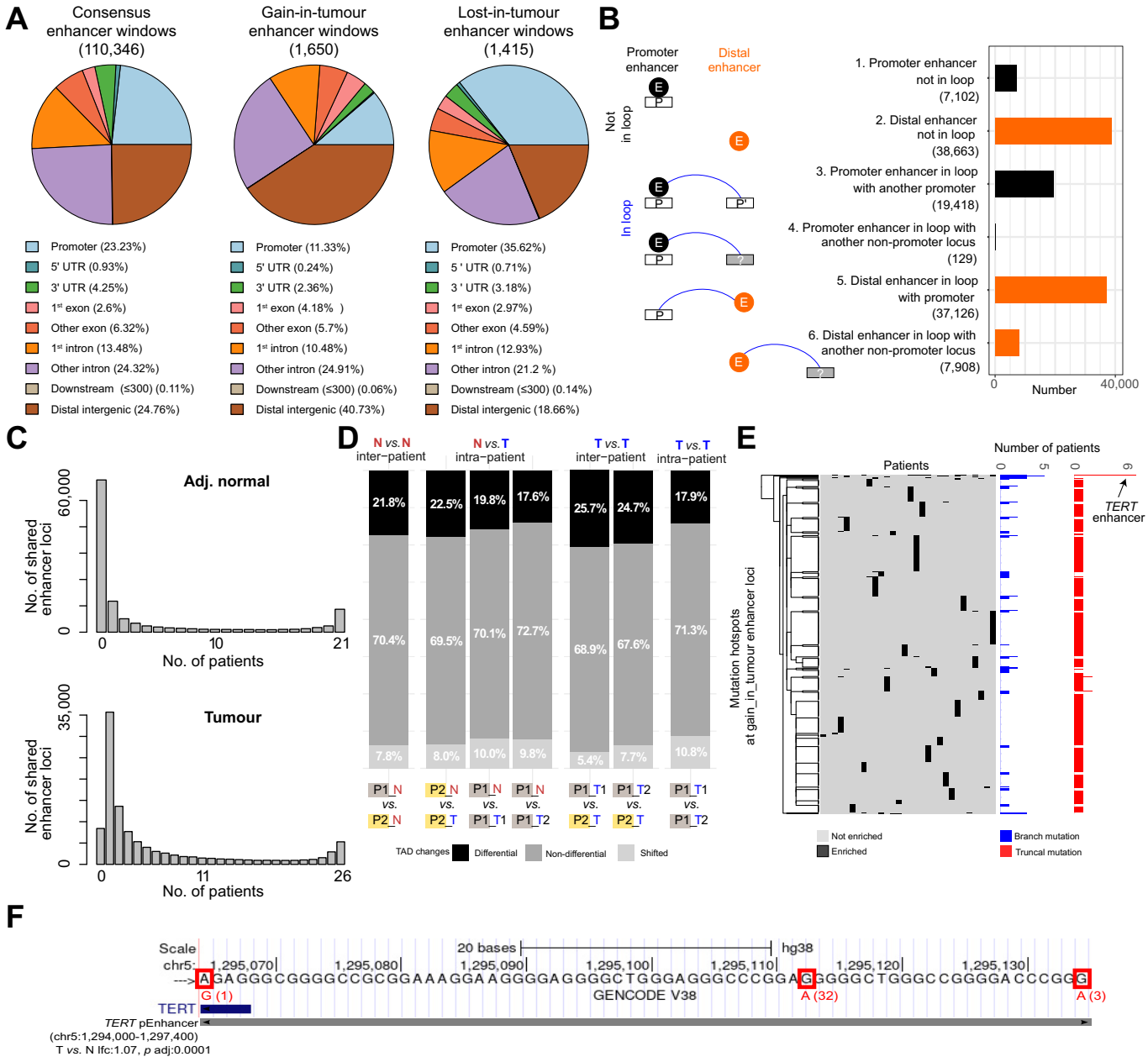


Fig. 1. Enhancer rewiring is partly attributable to genomic mutations. (A) Pie charts showing proportions of different genomic elements in all enhancers (left), gain-in-tumour enhancers (middle), and lost-in-tumour enhancers (right). (B) Types of enhancers and the total number detected. (C) Number of enhancer loci detected across the adj.normal (top) and tumour (bottom) tissues of different numbers of patients. (D) Stacked percentage bar chart of types of changes in the TAD boundaries between different pairs of samples, analysed using TADCompare.[21] N *vs.* N, adj.normal to adj.normal; T *vs.* T, tumour to tumour; inter-patient, between samples of different patients; intra-patient, between samples of the same patient. (E) Heatmap of the type of mutations at mutation hotspots identified from MutEnricher.[33] The left heatmap shows mutation hotspots at the gain-in-tumour enhancer regions, whereas the right heatmap shows mutation hotspots at the lost-in-tumour enhancer regions. Each row represents a mutation hotspot, corresponding to an enhancer locus. Barplots next to each heatmap represents the number of patients from which the mutation hotspot was detected. A mutation hotspot was considered truncal (red) if it was detected in all tumours of a patient, whereas mutations occurring in certain tumour sectors of a patient were considered branch mutations (blue). (F) Enriched somatic mutations at the most frequent mutation hotspot detected in (E). Numbers in parentheses represent the number of tumour sectors in which the mutation was detected from. The enhancer, 'chr5:1295700' (denoted in grey bar), overlaps with the *TERT* promoter (denoted in blue bar) and is a gain-in-tumour enhancer. adj.normal, adjacent normal; TAD, topologically associated domain; UTR, untranslated region.

enhancer loci, whereas lost H3K27ac were localised to promoter enhancer loci, indicating enhancer rearrangement in HCC.

Based on Hi-C and H3K27ac HiChIP data, gene–enhancer association was inferred by either (1) the direct occurrence of enhancers at a gene promoter or (2) distal enhancers in contact with a gene promoter, assuming that any changes at a distal enhancer would also affect the promoter in contact with the enhancer locus. There were more distal enhancers than promoter

enhancers, and strikingly, many promoter enhancers were in loop contact with other promoters (Fig. 1B).

Changes at the gain-in-tumour enhancers were more prominent than those at the lost-in-tumour enhancers

The H3K27ac coverage heatmap showed that adj.normal H3K27ac enrichment was faint for gain-in-tumour enhancers, whereas tumour H3K27ac enrichment for lost-in-tumour enhancers was clearly visible (Fig. 2E). The presence of *de novo* enhancer loci at gain-in-tumour enhancers was further supported by the frequency of unique enhancers observed in tumour tissues (tall bar at the zero column in Fig. 1C, top). By contrast, only a very small number of enhancers were uniquely detected in adj.normal tissues (Fig. 1C, bottom). Moreover, more chromatin loops were detected across tumour samples (97,255 pairs) than across 28,116 pairs in adj.normal. A higher number of detected loops in tumour samples also suggested the creation of *de novo* loops in HCC tumours. Altogether, these results indicate that H3K27ac mark changes are more prominent at gain-in-tumour enhancers with the creation of *de novo* enhancers and loops.

Few mutation hotspots existed under differential enhancers, and they were highly patient-specific

To investigate whether somatic mutations underlie changes in tumour H3K27ac enrichment, we looked for these at the differential enhancer loci. Whole genome sequencing data of the 30 patients in the epigenome cohort were used in this analysis. Among all gain-in-tumour enhancers, only ~20% (346 of 1,650) were enriched with 450 unique mutation hotspots. Similarly, for lost-in-tumour enhancers, ~18% (256 of 1,415) were enriched with 323 unique mutation hotspots (Fig. S3). However, even though the identified mutation hotspots presented a strong statistical base from MutEnricher,[33] most of them were detected uniquely in only one or two patients, and many branch mutations occurred uniquely in only subsets of tumour sectors for some patients (Fig. 1E and Fig. S2B). The high degree of individuality in the enhancers in HCC was also consistent with the high number of patient-specific tumour enhancers (Fig. 1C, bottom). One exception was the somatic mutation detected under a gain-in-tumour enhancer that overlapped with the TERT promoter enhancer (chr5:1,295,113, G>A) (Fig. 1F), which appeared to be the only true mutation hotspot under a differential enhancer. Even though the enhancer locus overlaps with the TERT promoter, the somatic mutation locus is upstream of the promoter. This enhancer also showed a significantly higher level of H3K27ac in tumour samples, and many of our patients did not harbour the known TERT activation mutation. Altogether, our observations suggest genetic mutations are not the main drivers of genome-wide enhancer changes in HCC.

Tumour samples showed variable H3K27ac signal at the differential enhancers

We found that K-means clustering (based on H3K27ac signals at gain-in-tumour enhancers) separated the samples into at least three groups (Fig. 3C). Any more than the cluster size of 3 resulted in group samples containing one or two patients only. The three 'enhancer clusters' (eClusters) A, B, and C showed increasing H3K27ac signals at gain-in-tumour enhancers and, conversely, decreasing H3K27ac signals at lost-in-tumour enhancers (Fig. 3D). The heatmap of H3K27ac signals at gain-in-tumour and lost-in-tumour enhancers, which grouped by eCluster, showed homogenous H3K27ac signals from adj.normal samples, whereas tumour samples showed highly variable H3K27ac signals (Fig. 3A and B). Adj.normal samples from all patients formed the majority of samples in eCluster A. A few tumour samples that showed highly similar H3K27ac pattern to adj.normal samples were also clustered to eCluster A. We also noticed that patients in eCluster C, the more divergent group with H3K27ac signal distinctly different from adj.normal, showed a higher degree of de-differentiation, a higher rate of disease recurrence (Fig. 3E), and shortened time to recurrence (Fig. 3F).

Enhancers in HCC showed high degree of interindividual variations

Among the detected chromatin loops in tumour samples, only 0.06% (60 of 97,255) were shared between all samples, whereas many of the loops were uniquely detected in only one sample (75.0%, 72,960 of 97,255). We also compared TAD boundaries between adj.normal and tumour samples from two patients. Interpatient comparisons showed high TAD boundary differences compared with intrapatient comparisons (Fig. 1D). Tumour samples from different patients showed greater TAD boundary differences than adj.normal samples, indicating a higher degree of divergence created by tumorigenesis.

Notably, a similar observation was made for enhancer heterogeneity. We have performed H3K27ac ChIP-seq on at least two tumour sectors for 12 of 30 patients. Three patients who had tumour sectors assigned to different eClusters showed intratumour heterogeneity (Fig. 3E). Among the 12 patients, six patients had three tumour sectors sequenced. For these six patients, we compared the enhancer loci detected in each sample to the merged consensus enhancer loci. This validated that many consensus enhancer loci were not detected in any of the six patients' tumour samples, again indicating a high degree of interindividual heterogeneity (Fig. S2).

Genes associated with differential enhancers indicated variable levels of de-differentiation and cellular proliferation

By associating enhancers to gene promoters with chromatin loops from Hi-C and HiChIP data, we identified 1,300 genes associated with gain-in-tumour enhancers and 1,360 genes associated with lost-in-tumour enhancers. There were many single gene promoter to multiple enhancer associations, and *vice versa*. Differential gene expression analysis was performed between tumour and adj.normal samples using the bulk RNA-seq data available from the same group of 29 patients. Although stronger H3K27ac signals in tumours did not always translate to a stronger expression of the associated gene, many gain-in-tumour enhancer-associated genes were indeed upregulated in tumour samples as compared with adj.normal samples (Fig. 3H). A similar observation was made for lost-in-tumour enhancer-associated genes. Enhancer changes therefore broadly correlated with transcriptional changes and motivated us to analyse the gene set more closely.

Many differential enhancer-associated genes encoded for transcription factors (TFs) (Fig. S5) may explain the genome-wide rearrangement of enhancers. Noticeably, more lost-in-tumour enhancer-associated genes overlapped with cell differentiation marker genes could indirectly support the process of de-differentiation in HCC. This observation was consistent with the enrichment of the HNF4A binding motif underlying gain-in-

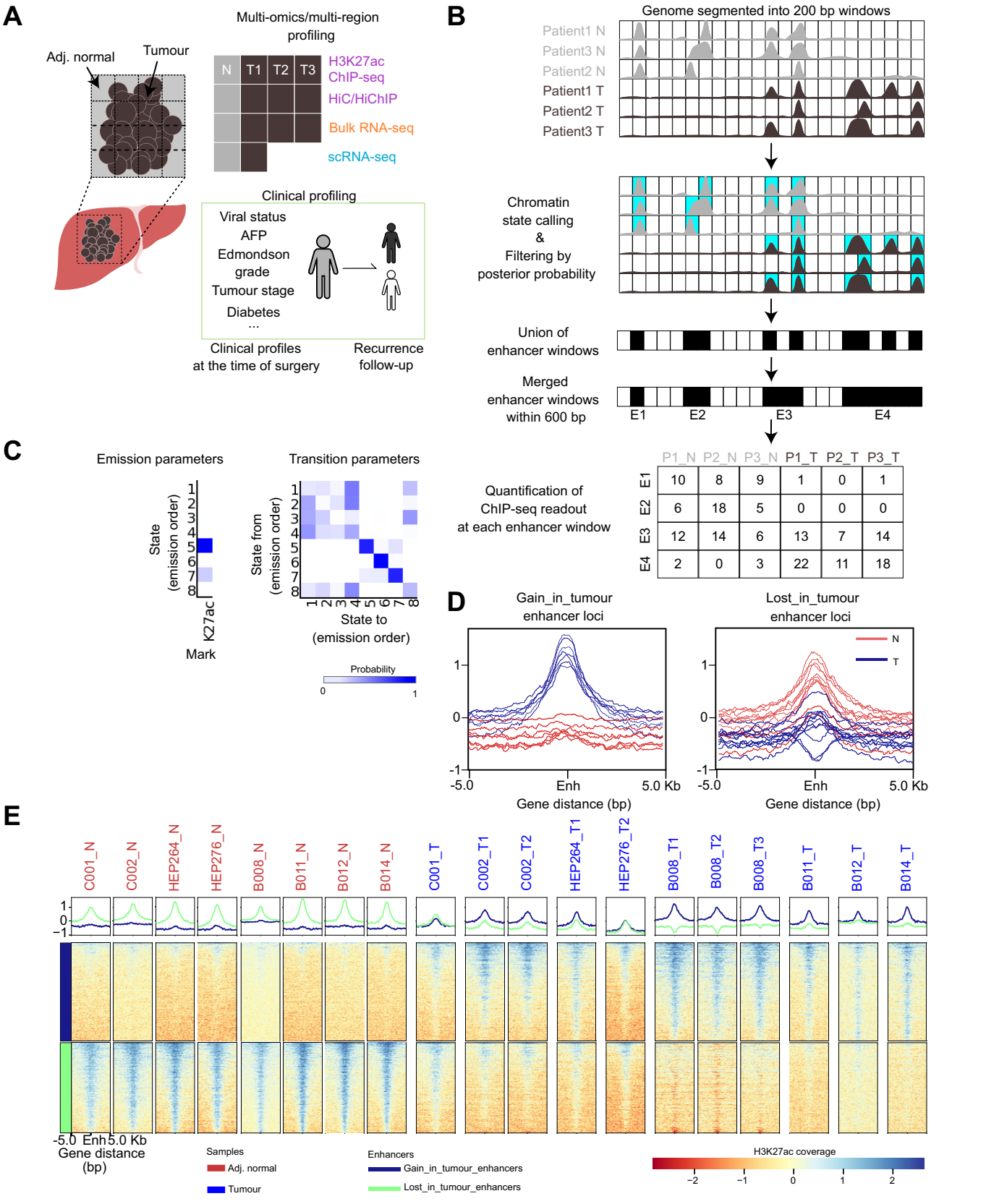


Fig. 2. Differential enhancers in HCC tumours. (A) Schematic diagram of the overall project design. (B) Schematic diagram of how differential enhancers were identified (see Patients and methods). (C) Emission and transition probability matrix from ChromHMM. (D) H3K27ac signal profiles at the gained and lost enhancers. Red, normal; blue, tumour. (E) Heatmap of H3K27ac signal from adj.normal and tumour samples at differential enhancers, normalised to the respective control samples. Heatmaps are centred at the midpoint of each enhancer and extended upstream and downstream by 5 kb. Gained-in-tumour (top) and lost-in-tumour (bottom) enhancer regions are shown. The same heatmaps for all samples can be found in Fig. S1. adj.normal, adjacent normal; AFP, alpha-foetoprotein; ChIP-seq, chromatin immunoprecipitation sequencing; H3K27ac, H3K27 acetylation; HCC, hepatocellular carcinoma; RNA-seq, RNA sequencing; scRNA-seq, single-cell RNA sequencing.

tumour enhancers (Fig. S4), as HNF4A is a known marker for hepatic progenitor cells.[38] Gain-in-tumour enhancer-associated genes showed strong enrichment for cell proliferation and regulation of metabolic and biosynthetic processes (Fig. 3G, top). Lost-in-tumour enhancer-associated genes were mainly those of signal transduction, biological adhesion, and epithelial–mesenchymal transition processes (Fig. 3G, middle and bottom). Notably, the patients whose tumour samples were assigned to eCluster C, which showed the most divergent enhancer pattern from adj.normal, had the poorly differentiated Edmondson grade 3 tumour, and all showed the shortest time to recurrence, compared with the patients whose tumours belonged to eCluster A or B (Fig. 3E and F).

Both enhancer patterns and expression of the associated genes were heterogeneous between patients

Many associated genes showed overall expression changes concordant with the enhancer changes (Fig. 3H). However, given the interindividual variability of H3K27ac signals at differential enhancer loci, we hypothesised that transcriptional changes of associated genes might also be patient-specific. We performed per-patient differential expression analysis by following the approach taken by Jeon *et al.*[14] In this approach, the multiregion tumour samples were treated as biological replicates to perform differential expression testing between each patient's tumour sectors and the matched adj.normal. Among the 30 patients with epigenome profiling, we selected 25 patients who had adj.normal tissue and at least two tumour sectors sequenced for RNA-seq. We assigned three categories, namely -1, 0, and +1, to denote downregulation, no change, and upregulation of the gene in each patient's tumour, respectively. Fig. 3I shows the categorised expression changes in differential enhancer-associated genes for all 25 patients. Overall, differential expression results from conventional analysis (Fig. 3H) and that of the patient-specific analysis (Fig. 3I) were broadly similar. The patient-specific analysis, however, provided granularity and information for each patient on whether a specific gene was dysregulated in the patient's tumour tissue. We then used the categorised differential expression values for patient stratification in the larger group of 90 patients.

Two examples of genes – *SOX4* and *GPC3* – in Fig. 4 exemplify the variability and patient specificity in enhancer patterns and gene expression changes. For *SOX4*, many enhancers linked by chromatin loops were unique to tumour tissues (Fig. 4A, bottom). Only one SOX4 distal enhancer showed a significant H3K27ac increase, and notably, the H3K27ac signal was highly variable between patients (Fig. 4C). For some patients (*e.g.* C001 and B011), H3K27ac was barely detectable, whereas in one patient (HEP262), there was markedly increased enrichment in tumour tissues. This is thus an example of a highly variable enhancer locus that exhibited a high degree of patient-specific heterogeneity. The variability was also seen in *SOX4* gene expression. Patient-specific differential expression analyses revealed *SOX4* to be upregulated in a subset of the 90 patients, where some of them even showed *SOX4* downregulation in their tumour tissues (Fig. 4E). By contrast, *GPC3* is an example of a gene with more consistent dysregulation across patients. *GPC3* promoter and distal enhancers showed significant and consistent H3K27ac in tumour samples (Fig. 4B) across all patients (Fig. 4D). *GPC3* was also upregulated consistently in a larger group of patients (Fig. 4F).

Genes associated with the gain-in-tumour enhancers show strong overlap with foetal liver expressed genes

Gene set enrichment analysis[39] showed a striking overlap between the gain-in-tumour enhancer-associated genes and marker genes for specific cell types (pancreas ductal cells, foetal liver hepatoblasts, and epithelial cell adhesion molecule (EpCAM)-positive bile duct cells) (Fig. 5A). Linear regression analysis showed that some genes, such as *GPC3, MSI1, PEG10*, and *ELOVL2*, also showed a strong positive correlation to serum alpha-foetoprotein (AFP) levels in corresponding patients (Fig. 5D and Fig. S7). By contrast, other genes, such as *UGT2B7* and *PROX1*, showed a negative correlation. The correlation was present only in tumour tissues, whereas gene expression across adj.normal tissues was largely unchanged. Variable gene expression in tumours may be partly associated with the extent of foetal-like transformation taking place in the liver.

Among all the gain-in-tumour enhancer-associated genes, we looked at 46 foetal liver hepatoblast markers (Table S5). Gain-in-tumour enhancers associated with these 46 genes indeed showed increased H3K27ac enrichment in tumour tissues, especially those in eCluster C (Fig. 5B). Because these genes with corresponding enhancer changes were the hallmark genes for foetal liver, and some showed significant correlation to serum AFP levels, we termed this gene set as 'epigenetic oncofoetal genes'. We hypothesise that when HCC cells are de-differentiated through the emergence of cancer stem cells or hepatic progenitor cells, 'epigenetic oncofoetal genes' are upregulated, whereas they are not normally expressed in a non-diseased liver. This hypothesis also concords with the preceding observation of HNF4 binding motif enrichment at gain-in-tumour enhancer loci. Moreover, gain-in-tumour enhancer-associated genes were related to processes of proliferation and de-differentiation.

We explored whether H3K27ac and corresponding gene expression variability in tumour tissues was caused by a differing proportion of poorly differentiated cells or hepatic progenitor cells. scRNA-seq of 14 patients were analysed for the expression of the epigenetic oncofoetal genes, as well as *SOX4* and *HNF4A*. From the scRNA-seq, we focused on cells identified as either endothelial cells or hepatocytes. We fitted a mixture model to select genes that showed bimodal expression and used gene-specific thresholds to binarise the gene expression values. The proportion of cells expressing a gene of interest was then calculated by taking the ratio of cells expressing the gene to the total number of cells. Using hierarchical clustering, we clustered the samples that comprised one adult non-diseased liver tissue, two foetal liver tissues, and adj.normal and tumour tissues from 14 patients with HCC based on the proportion of the cells expressing the epigenetic oncofoetal genes (Fig. 5C). Clustering revealed two groups of samples, one that was non-diseased liver-like and the other foetal liver-like. All but one adj.normal sample was grouped into the non-diseased liver-like group, so we termed the first group as 'adj.normal-like'. Many adj.normal-like liver tissues showed a low proportion of cells expressing epigenetic oncofoetal genes. The other group, by contrast, showed a distinctly higher proportion of cells expressing these epigenetic oncofoetal genes. Genes such as *GPC3* and *PEG10* were

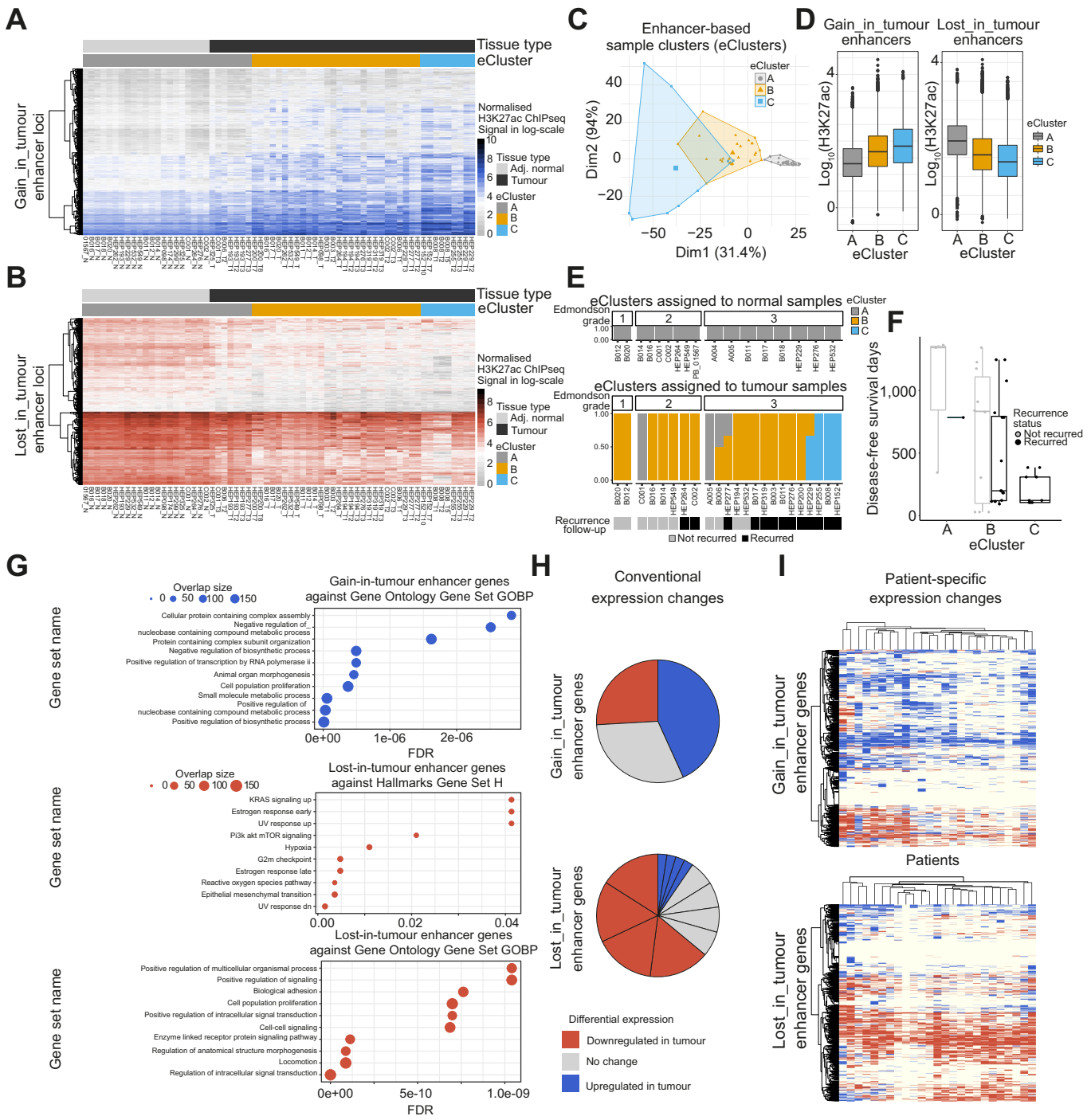


Fig. 3. Heterogeneity in the H3K27ac signal at differential enhancers and the associated gene expression changes. (A) Heatmap of H3K27ac ChIP-seq signal at the gain-in-tumour enhancers. Enhancers shown on the rows were ordered using hierarchical clustering. H3K27ac ChIP-seq samples are shown on the columns, ordered by the eClusters identified by K-means clustering. (B) Heatmap of H3K27ac ChIP-seq signal at the lost-in-tumour enhancers. (C) PCA plot of the samples shown in (A) and (B), grouped by the eClusters. (D) Total H3K27ac signal at the differential enhancers (left, gain-in-tumour; right, lost-in-tumour), grouped by the eClusters. (E) Proportion of eClusters assigned to adj.normal and tumour tissues of each patient (top and middle), and the recurrence status for the corresponding patients with tumour tissues sequenced (bottom). All bar plots are grouped by the Edmondson grade of the patients' tumour sample. (F) Boxplot of disease-free survival for patients shown in (E), grouped by the eClusters. (G) Gene set enrichment analysis results for the genes associated with the gain-in-tumour enhancers (top) and lost-in-tumour enhancers (middle and bottom). (H) Proportion of differentially expressed genes for genes associated with the gain-in-tumour enhancers (top) and with the lost-in-tumour enhancers (bottom). Results obtained from the conventional differential expression testing. (I) Heatmap of categorised expression changes matrix from patient-specific differential expression testing. Top shows the gain-in-tumour associated genes, whereas the bottom shows the lost-in-tumour associated genes. Colours of blue, light yellow, and red each represent the expression changes of upregulation, no change, and downregulation, respectively. adj.normal, adjacent normal; ChIP-seq, chromatin immunoprecipitation sequencing; eCluster, enhancer cluster; FDR, false discovery rate; GOBP, Gene Ontology Biological Process; H3K27ac, H3K27 acetylation; PCA, principal component analysis.

expressed in the majority of cells in both foetal-liver tissues, and also in some patient tumours. In Sharma *et al.*,[27] the authors in fact selected P8 and P15 patient tumour tissues for validation and showed experimentally that these tumours exhibited oncofoetal reprogramming of endothelial cells together with an immune escape from the surrounding T cells. In our analysis here, both P8 and P15 patient tumours (P8_Tumor and P15_Tumor, respectively) clustered with foetal liver tissues (Fig. 5C). We, therefore, termed the second group as 'oncofoetal-like'. This subgroup of tumours showed a high number of cells expressing epigenetic oncofoetal genes, with a similar expression pattern to that of foetal-liver tissues. The number of cells expressing *SOX4* was variable across all tissues, but *HNF4A* was markedly positive in many cells from the oncofoetal-like liver tissues. It is also noteworthy to mention that Domcke *et al.*[40] also reported that the genomic loci with chromatin accessibility specific to the foetal hepatoblast cells were enriched with the *HNF4A* motif. Altogether, our integrated analysis with scRNA-seq highlights that epigenetic oncofoetal genes are indeed associated with HCC tumours that are foetal liver-like, with characteristics of oncofoetal reprogramming.

Gain-in-tumour enhancer-associated genes that were patient subgroup-activated showed prognostic value

The patient-specific differential expression analyses revealed that tumour-related genes were dysregulated to varying degrees in different patients. Some genes, possibly related to common cancer pathways, were shared in a large group of patients, whereas other genes were dysregulated only in a small subgroup of patients. Our earlier work, which investigated this observation in more detail, had shown that such subgroup-activated genes were clinically relevant.[14] As the list of signature genes that we identified based on enhancer changes harboured many proliferation, de-differentiation, and oncofoetal reprogramming-

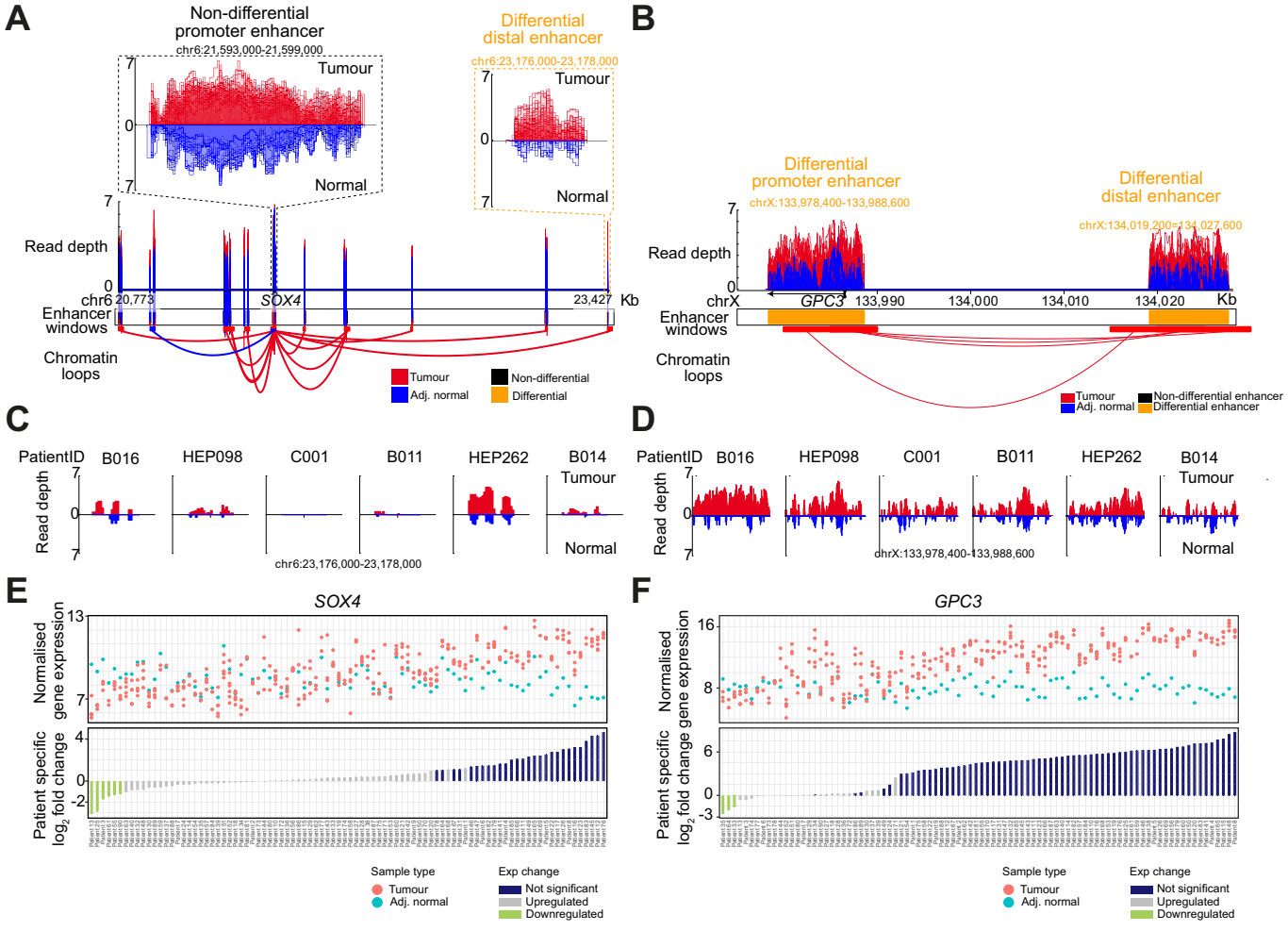


Fig. 4. Patient-dependent H3K27ac signal at the associated enhancers and expression changes of SOX4 (left) and GPC3 (right). (A) H3K27ac ChIP-seq signal at the enhancers associated with SOX4. Differential enhancer status of each enhancer is denoted in yellow. Chromatin loops were inferred from Hi-C/HiChIP libraries and are coloured based on the frequency of detection in Hi-C/HiChIP libraries. (B) Similar plot as A for GPC3. (C) Normalised H3K27ac coverage at the differential distal enhancer of SOX4, for the paired adj.normal (blue, bottom) and tumour (red, top) tissues of 6 patients. (D) Similar plot to (C) for the differential promoter enhancer of GPC3. (E) Per-patient differential expression testing results for SOX4 across 90 patients. Normalised gene counts for each adj.normal (blue) and tumour (pink) samples are shown on top, whereas the bottom plot shows the $\log_2$ fold change values for those detected as differentially expressed (blue bars, upregulated in tumour; green bars, downregulated in tumour; grey, not significantly different). In both plots, each column represents a patient. (F) Similar plot to (E) for GPC3. adj.normal, adjacent normal; ChIP-seq, chromatin immunoprecipitation sequencing; H3K27ac, H3K27 acetylation.

related genes, we hypothesised that subgroup-activated genes among them must present a strong correlation to recurrence. We hence applied these genes to patient stratification and disease-free survival analysis.

Given the earlier observations of the genes relating to de-differentiation and developing foetal-like characteristics, we hypothesised that the differential activation pattern of some of the gain-in-tumour enhancer-associated genes might be associated with different rates of HCC recurrence. We obtained the categorised expression changes matrix for 90 patients from their multisector bulk RNA-seq data and selected the differential enhancer-associated genes (Fig. S6A). These genes varied from being upregulated in just a few patients to more commonly across the patients (Fig. S6B). With the focus on gain-in-tumour enhancer-associated genes, we filtered out genes that were upregulated in only a small group of patients, as these may be

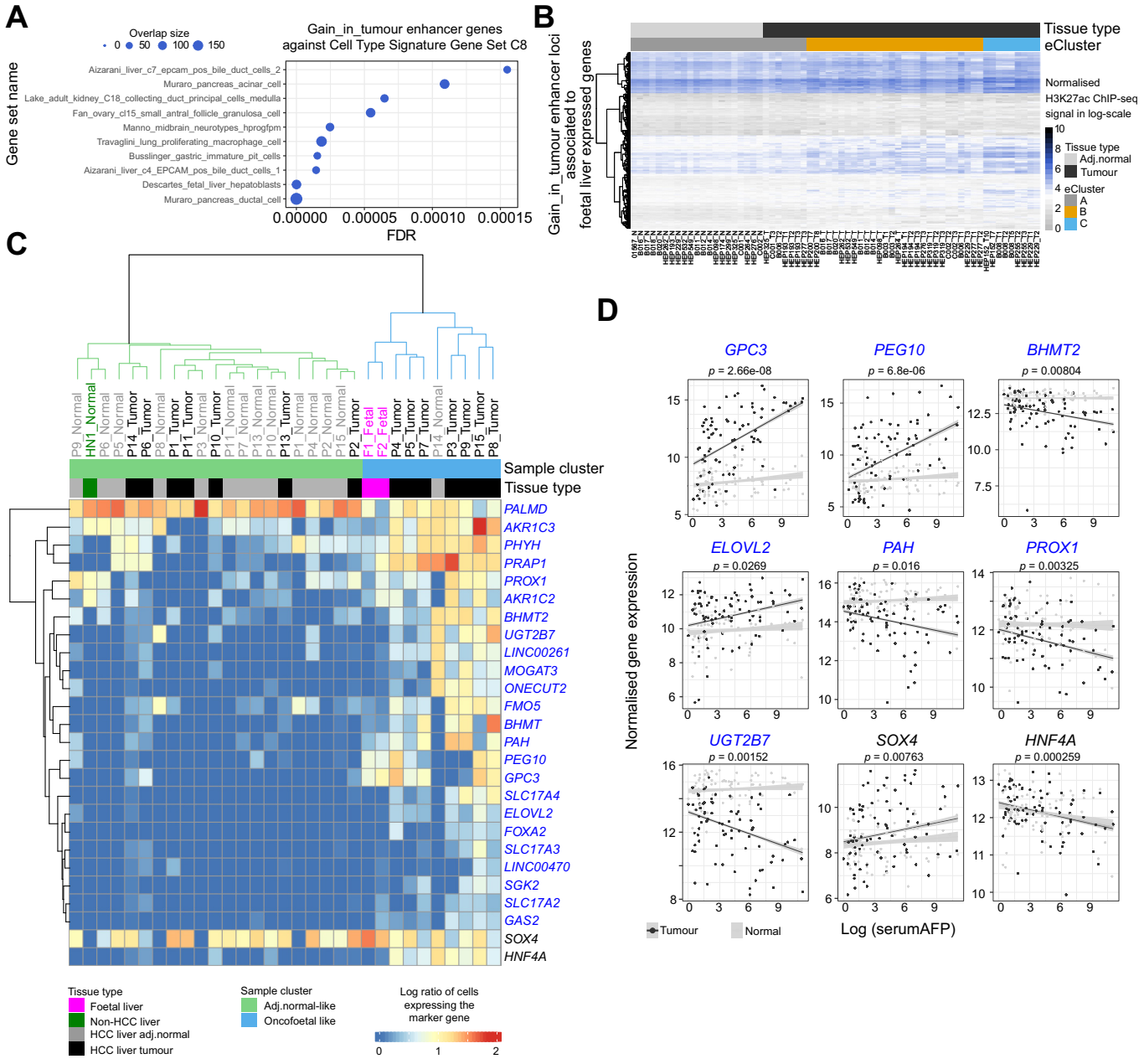


Fig. 5. Gain-in-tumour associated genes show enrichment for foetal liver genes. (A) Gene set enrichment analysis result of gain-in-tumour enhancer-associated genes against cell type signature gene set C8 from MSigDB.[48] (B) Heatmap of H3K27ac ChIP-seq signal for the gain-in-tumour enhancers associated with the foetal liver-expressed genes, ordered by the eCluster. (C) Hierarchical clustering of different liver tissues based on the amount of cells expressing foetal liver and gain-in-tumour enhancer-associated genes. Genes were further filtered based on the availability in the scRNA-seq data. (D) Selected pool of genes that showed significant correlation between its expression level and the serum AFP level. More genes are shown in Fig. S6. adj.normal, adjacent normal; AFP, alpha-foetoprotein; ChIP-seq, chromatin immunoprecipitation sequencing; eCluster, enhancer cluster; H3K27ac, H3K27 acetylation; HCC, hepatocellular carcinoma; scRNA-seq, single-cell RNA sequencing.

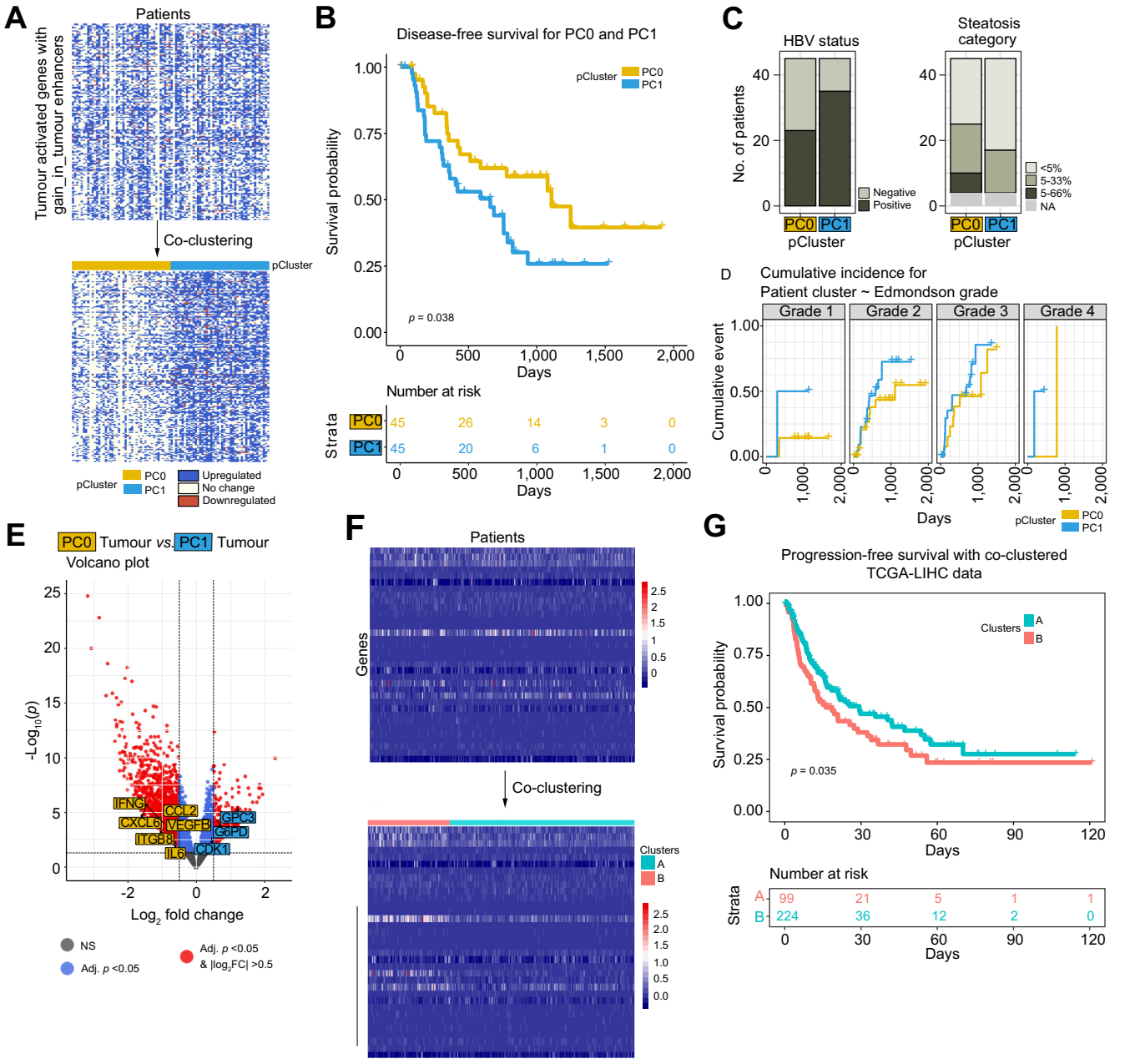


Fig. 6. The patient-specific upregulation pattern of gain-in-tumour enhancer-associated genes shows prognostic values. (A) Heatmap of patient-specific differential expression status for genes upregulated in at least 30% of the patients. The bottom heatmap shows the effect of co-clustering, and the columns are coloured by the assigned patient strata. (B) Disease-free survival plot and the risk table for the two patient clusters identified from co-clustering (Tarone–Ware test, $p = 0.038$). (C) Distribution of clinical parameters that were significantly different between the two patient clusters. The full breakdown of the clinical tables is shown in Table 1. (D) The cumulative event plots for patients with different Edmondson grades, grouped by the patient clusters. (E) Volcano plot of differential expression testing between PC0 and PC1 from RNA-seq data. (F) Heatmap of tumour-normal difference expression values of the epigenetic oncofoetal genes from TCGA-LIHC RNA-seq data, grouped by the patient clusters identified from co-clustering. (G) Progression-free survival plot and the risk table for the two patient clusters identified from TCGA-LIHC data (Tarone–Ware test, $p = 0.035$). PC, patient cluster; RNA-seq, RNA sequencing; TCGA-LIHC, The Cancer Genome Atlas Liver Hepatocellular Carcinoma.

related to individual patient characteristics, rather than a shared tumour profile. We finally narrowed the gain-in-tumour enhancer-associated genes down to 185 genes that were upregulated in at least 30% of patients. The categorised expression changes matrix showed that even though these genes were upregulated in many patient tumours, some genes still showed downregulation in some patients (Fig. 6A, top, and Fig. S6A). We used co-clustering and identified two groups of patients ('pClusters') – PC0 and PC1 – based on the categorised expression changes matrix of the 197 genes (Fig. 6A and Table S7). PC1 patients showed more gene upregulation in their tumour tissues than in their adj.normal tissues. The recurrence-free survival analysis revealed that PC1 patients showed a significantly worse prognosis with a shorter time to recurrence (Fig. 6B). Median

Table 1. Clinicopathological table for patient clusters PC0 and PC1.

	PC0	PC1	*p* value
N	45	45	
Sex			
Female	6	10	0.396
Male	39	35	
Race			
Chinese	28	28	0.319
Filipino	5	1	
Indian	1	2	
Indonesian	0	2	
Malay	3	2	
Thai	3	7	
Others	5	3	
Drinker			
No	22	22	0.498
Yes	15	11	
Unknown	8	12	
Child–Pugh			
A	44	45	1.000
B	1	0	
Diabetes			
N	26	29	0.662
Y	19	16	
Tumour multiplicity			
N	37	35	0.802
Y	8	10	
Fibrosis stage			
0	13	10	0.048*
1	1	6	
2	2	9	
3	11	6	
4	13	13	
Microvascular invasion			
No	29	25	0.508
Yes	16	20	
Edmondson grade			
1	7	2	0.266
2	24	22	
3	13	19	
4	1	2	
Steatosis category			
0	20	28	0.021*
1	15	13	
2	6	0	
Vital status			
Alive	35	34	1.000
Dead	10	11	
Hepatitis B status			
0	22	10	0.017*
B	23	35	
Hepatitis C status			
0	41	42	1.000
C	4	3	
Tumour stage			
1	23	22	0.493
2	17	14	
3	5	9	
Tumour diameter (cm)	6.51 ± 4.12	6.46 ± 4.55	0.971
Albumin (g/L)	41.39 ± 4.54	40.91 ± 3.38	0.383
Bilirubin (μmol/L)	13.6 ± 4.96	13.23 ± 5.43	0.674
Aspartate aminotransferase (U/L)	47.26 ± 33.15	53.64 ± 49.89	0.264
Alanine aminotransferase (U/L)	47.14 ± 51.02	37.31 ± 25.11	0.692
Alkaline phosphatase (U/L)	107.65 ± 62.38	123.09 ± 111.07	0.383
Prothrombin time (s)	11.1 ± 1.09	11.41 ± 1.27	0.240
Platelets ($\times 10^9$)	223.59 ± 75.83	239.22 ± 79.59	0.208
alpha-foetoprotein (ng/ml)	516.7 ± 1,554.74	4,658.25 ± 13,205.04	0.096

Full clinical variables are included in Table S6.
*Level of significance: p <0.05 (Chi-square test and Wilcoxon test).

time to recurrence was 1,107 days for PC0 patients and 659 days for PC1 patients. Comparing clinical parameters of PC0 and PC1 patients showed significant differences in the degrees of fibrosis and steatosis and hepatitis B status (Table 1). The PC1 group consisted of more patients with chronic hepatitis B infection, whereas the PC0 group showed more patients with high levels of steatosis (Fig. 6C). Our earlier observation showed that patients with high Edmondson grade have high H3K27ac intensity at gain-in-tumour enhancers in their tumour tissues. Even though the Edmondson grade was not significantly different between PC0 and PC1 (Table 1), a cumulative incidence plot between patient cluster and Edmondson grades showed that PC1 patients had a higher rate of recurrence, regardless of the Edmondson grade (Fig. 6D). This suggests that detectable molecular characteristics stratify those with higher recurrence rate, regardless of their histological assessment of cellular de-differentiation. This observation may be especially valuable for patients with Edmondson grades 1 and 2, even though they are usually considered to have low recurrence risk in current clinical practice.

To identify differences in patient tumour tissue profiles, we compared tumour tissues between PC0 and PC1 patients through differential expression testing of their RNA-seq data. Many upregulated genes included hallmark genes of glycolysis, *MYC* targets, and cell cycle-related genes (Fig. S8A). PC1 patient tumours also highly expressed genes related to liver cancer metastasis and subgroups of HCC with stem cell characteristics. Downregulated genes in PC1 patient tumour tissues included those related to the regulation of the immune system process, defence response, interferon-gamma response, and biological adhesion (Fig. S8B). Genes related to glucose metabolisms and cell cycles, such as *G6PD* and *CDK1* were upregulated in PC1 patients, whereas genes related to immune response (*IFNG* and *CXCL6*) were downregulated (Fig. 6E). The results indicate that our signature genes were indeed able to identify patient subgroups with more proliferative tumour tissues and greater potential to metastasize, possibly with suppressed immune response and metabolic processes.

Further studies are required to investigate the molecular mechanism behind epigenetic dysregulation and HCC. One possible mechanism, however, could involve the differential recruitment of TFs. For the TFs with motifs enriched in gain-in-tumour enhancers (Fig. S4), we studied the expression level changes of the predicted target genes in each pClusters (PC0 and PC1) (Table S9). More patients in PC1 showed upregulated target genes than those in PC0, indicating that the TFs may be the connection between the enhancer dysregulation and the downstream transcriptional changes, with clinical relevance in prognosis.

The epigenetic oncofoetal genes identify patients with a more aggressive HCC subtype

We compared our 185-gene panel used in patient stratification to previously reported genes related to the HCC subtypes of different prognoses from Hoshida *et al.*[41] None of the positive survival genes from Hoshida *et al.*,[41] and only one poor survival gene, *ELOVL2*, were found in our 185 genes. Despite the minimal overlap between the gene lists, the expression changes in the pattern of survival genes from Hoshida *et al.*[41] were in broad agreement for our two identified patient clusters, PC0 and PC1 (Fig. S9). Many positive survival genes were downregulated across PC1 tumour tissues. The expression changes of poor survival genes were less distinguished between the two patient clusters. Importantly, we again noted that expression changes were variable and patient-specific.

To compare our tumour subtypes to previously reported subtypes by Boyault *et al.*[42] and Hoshida *et al.*,[43] we tabulated the somatic mutation and gene expression features of the tumours in each patient group (Table S8). Upregulated genes in PC1 tumours, which showed a worse prognosis, showed enrichment of the BOYAULT_LIVER_CANCER_SUBCLASS_G3_UP gene set (Fig. S8A). PC1 tumours also showed a higher percentage of TP53 mutations (Table S8), which was one of the mutational features of the G2 and G3 subclasses by Boyault *et al.*[42]. Boyault *et al.*[42] had indeed reported that the G3 subclass showed the worst prognosis, even though the difference did not reach statistical significance. It shows that the subtypes we discovered may have a more selective prognostic value. About the Hoshida's subtypes,[43] we believe PC1 tumours correspond to the S2 subtype described by Hoshida *et al.*,[43] which was highlighted with a high AFP level, poor differentiation, and poor survival. Similarly, PC1 tumours showed association to EPCAM-positive cells (Fig. 5A), a high level of proliferation (Fig. 3G, top), a high level of serum AFP (Fig. 5D), and upregulation of MYC, PI3K-AKT, and E2F1 targets (Fig. S8A). All of them correspond to the features of the S2 subtype described by Hoshida *et al.*[43] Altogether, we believe our 185-gene panel has a higher specificity in identifying more aggressive HCC subtypes, and our findings show that the dysregulation of these genes is associated with the epigenetic changes that happen in the liver.

Finally, we validated our 185-gene panel using the TCGA-LIHC dataset by stratifying patients with HCC into clinically relevant groups. As the TCGA-LIHC dataset does not have similar multi-region sampling data that were used in our stratification, we stratified TCGA patients based on tumour-normal differences calculated using the average expression values among the normal samples. Our gene panel separated TCGA-LIHC patients into two clusters, cluster A and cluster B, consisting of 100 and 224 patients, respectively (Fig. 6F). From the normalised gene expression heatmap (Fig. 6G), a set of genes showed noticeably higher expression in cluster A patients than in cluster B patients. The top six differentially expressed genes are GPC3, ACSL4, PEG10, PRAP1, AKR1C2, and AKR1C3 (Fig. S10). The progression-free survival analysis between the two patient clusters showed that indeed the patients in cluster A (higher upregulation) showed worse prognosis than the patients in cluster B (Fig. 6G), similar to our earlier results (Fig. 6B). The results show that our gene panel has the selection power for stratifying patients into differential prognosis groups, as proven in both our dataset and the TCGA-LIHC dataset.

Discussion

Our systematic epigenome analysis of HCC reveals a highly variable enhancer distribution in HCC that is important and, to a large degree, patient-specific. Functional analysis of genes associated with differential enhancers and the clinical profile of patients with different enhancer signals showed that the different extent of cellular proliferation and de-differentiation of tumour cells appear to be key factors contributing to the heterogeneity.

There was minimal overlap between somatic mutation hotspots and differential enhancer loci, indicating that genetic perturbation appears not to be the main driver of HCC enhancer dysregulation. The potential cause for H3K27ac landscape

changes may instead be metabolic changes in the tumour microenvironment. Our results showed that some of the most striking differences between PC0 and PC1 patients were the hallmarks of glycolysis and PI3K–AKT–MTOR signalling genes. These may reflect the Warburg effect in HCC tissues to varying degrees among patients. A shift in glucose metabolism has been reported in HCC,[44] and it is expected to influence the acetyl-CoA abundance in the surrounding environment, in turn affecting histone acetylation.[45] Further studies are required to assess this possibility.

Oncofoetal reprogramming of endothelial cells in HCC and the emergence of foetal-liver-like characteristics in the tumour microenvironment of HCC tissues have been described in detail by Sharma *et al.*[27]. We confirmed a set of epigenetic-driven oncofoetal genes, which were expressed by a higher proportion of cells in oncofoetal-like tissues, indicating that the development of foetal-liver-like features in some HCC liver tissues is likely accompanied by the *de novo* (gain-in-tumour) enhancers. As the epigenetic oncofoetal genes are gain-in-tumour enhancer-associated genes known to foetal liver hepatoblast markers, our findings are concordant with the re-emergence of foetal liver marker genes in some patients with HCC together with underpinning enhancer rewiring. Furthermore, we have shown that some of the epigenetic oncofoetal genes were informative in identifying patient groups with differential prognosis.

Altogether, we present an integrative overview of the epigenomic and transcriptomic dysregulation in HCC, with an emphasis on patient-dependent heterogeneity with direct clinical relevance. Before developing HCC, patients often show progressive stages of various chronic liver diseases. Recent studies suggest that the liver undergoes dynamic and aberrant epigenetic changes accompanying metabolic changes.[46] Given the progressive nature of liver diseases before tumorigenesis, studies such as Jühling *et al.*[47] suggest the potential effectiveness of epigenetic drugs such as bromodomain inhibitors in chemoprevention. Our findings add to an ongoing approach that presents a paradigm shift in cancer research, from the convention of identifying specific targets for intervention to more systemic approaches that consider the tumour and the surrounding cellular environment in developing treatment strategies. For widespread and systemic genome-wide dysregulation, epigenetic therapies that can target genome-wide loci could be beneficial. Indeed, a panel of genes appears to undergo coherent epigenetic and transcriptional dysregulation in a subgroup of patients with HCC. Understanding the variability and developing methods to identify patient groups with different epigenomes will be important for the potential application of personalised cancer therapeutics.

Abbreviations

adj.normal, adjacent normal; AFP, alpha-foetoprotein; BED, Browser Extensible Data; ChIP-seq, chromatin immunoprecipitation sequencing; eCluster, enhancer cluster; EpCAM, epithelial cell adhesion molecule; H3K27ac, H3K27 acetylation; HCC, hepatocellular carcinoma; PC, patient cluster; PLANet, Precision Medicine in Liver Cancer across an Asia-pacific NETwork; RNA-seq, RNA sequencing; scRNA-seq, single-cell RNA sequencing; TAD, topologically associated domain; TCGA-LIHC, The Cancer Genome Atlas Liver Hepatocellular Carcinoma; TCR, Translational and Clinical Research; TF, transcription factor; UCSC, University of California Santa Cruz.

Financial support

This work was supported by Singapore National Medical Research Council Grants (TCR/015-NCC/2016, CSA-SI/0018/2017, and CIRG18may-0057).

Conflicts of interest

The authors declare no potential conflicts of interest.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Conceptualisation: PKC, AJ, RSYF. Formal analysis: AJ, YYT. Writing –original draft: AJ. Writing – review and editing: PKC, AJ, SLC, KS, CGA, RSYF. Project administration: SC. Resources: CGA, LW, JC, RIK, HL, WHL, NAK, JQL, AYFC, PC, JHK, KM, AK, ISG, TKHL, WL, SL, TJL, WKW, GSTS, YHP, BKY, DBO, JL, VHV, RDC, RC, JT, GKB, BKPG, PKC. Supervision: PKC, RSYF.

Data availability statement

The Hi-C/HiChIP, ChIP-seq, and scRNA-seq data analysed or generated during the current study are available in the Gene Expression Omnibus (GEO) with accession numbers GSE212055, GSE212342, and GSE156337, respectively. The DNA and RNA sequencing data are deposited in the European Genome-Phenome Archive (EGA) under accession EGAD00001009041. All data supporting the findings of this study are available within the article and its Supplementary information, or from the corresponding author upon request.

Acknowledgements

We thank all patients involved in this study. We thank members of NCCS, SGH, NUH, GIS, and the Liver TCR group for the all the work and effort involved.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhepr.2023.100715.

References

Author names in bold designate shared co-first authorship

[1] Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer 2020;20:555–572.
[2] Muntean AG, Hess JL. Epigenetic dysregulation in cancer. Am J Pathol 2009;175:1353–1361.
[3] Gopi LK, Kidder BL. Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. Nat Commun 2021;12:1419.
[4] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–249.
[5] London WT, McGlynn KA. Liver cancer. Cancer epidemiology and prevention. 3rd ed. New York: Oxford University Press; 2006.
[6] Petrick JL, McGlynn KA. The changing epidemiology of primary liver cancer. Curr Epidemiol Rep 2019;6:104–111.
[7] Liu Y-X, Li Q-Z, Cao Y-N, Zhang L-Q. Identification of key genes and important histone modifications in hepatocellular carcinoma. Comput Struct Biotechnol J 2020;18:2657–2669.
[8] Hayashi A, Yamauchi N, Shibahara J, Kimura H, Morikawa T, Ishikawa S, et al. Concurrent activation of acetylation and tri-methylation of H3K27 in a subset of hepatocellular carcinoma with aggressive behavior. PLoS One 2014;9:e91330.
[9] Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature 2021;590:300–307.
[10] Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol 2015;16:144–154.

[11] Zhai W, Lai H, Kaya NA, Chen J, Yang H, Lu B, et al. Dynamic phenotypic heterogeneity and the evolution of multiple RNA subtypes in hepatocellular carcinoma: the PLANET study. Natl Sci Rev 2022;9:nwab192.

[12] 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. Nat Commun 2017;8:4565.

[13] Anene-Nzelu CG, Tan WLW, Lee CJM, Wenhao Z, Perrin A, Dashi A, et al. Assigning distal genomic enhancers to cardiac disease–causing genes. Circulation 2020;142:910–912.

[14] **Jeon A-J**, **Teo Y-Y**, Sekar K, Chong SL, Wu L, Chew S-C, et al. Multi-region sampling with paired sample sequencing analyses reveals sub-groups of patients with novel patient-specific dysregulation in hepatocellular carcinoma. BMC Cancer 2023;23:118.

[15] Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Berkley: Lawrence Berkeley National Lab (LBNL); 2014.

[16] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.

[17] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–2079.

[18] Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics 2015;31:2032–2034.

[19] Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187–W191.

[20] Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 2016;3:95–98.

[21] Cresswell KG, Dozmorov MG. TADCompare: an R package for differential and temporal analysis of topologically associated domains. Front Genet 2020;11:158.

[22] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 2012;9:215–216.

[23] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

[24] Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res 2004;32:D493–D496.

[25] Phanstiel DH, Boyle AP, Araya CL, Snyder MP. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics 2014;30:2808–2810.

[26] Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 2015;31:2382–2383.

[27] Sharma A, Seow JJW, Dutertre C-A, Pai R, Blériot C, Mishra A, et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell 2020;183:377–394.e21.

[28] Jeon A-J, Tucker-Kellogg G. Bivalent genes that undergo transcriptional switching identify networks of key regulators of embryonic stem cell differentiation. BMC Genomics 2020;21:614.

[29] Maechler M. Diptest: Hartigan's dip test statistic for unimodality – corrected. 2021. https://github.com/mmaechler/diptest

[30] Young D, Benaglia T, Chauveau D, Hunter D, Elmore R, Hettmansperger T, et al. Mixtools: tools for analyzing finite mixture models.. https://github.com/dsy109/mixtools.

[31] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576–589.

[32] Zhang Q, Liu W, Zhang H-M, Xie G-Y, Miao Y-R, Xia M, et al. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. Genomics Proteomics Bioinformatics 2020;18:120–128.

[33] Soltis AR, Dalgard CL, Pollard HB, Wilkerson MD. MutEnricher: a flexible toolset for somatic mutation enrichment analysis of tumor whole genomes. BMC Bioinformatics 2020;21:338.

[34] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401–404.

[35] Kassambara A, Kosinski M, Biecek P, Fabian S. Survminer: drawing survival curves using 'ggplot2'. 2021. https://rpkgs.datanovia.com/survminer/index.html.

[36] Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. Nat Biotechnol 2010;28:1045–1048.

[37] Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–330.

[38] DeLaForest A, Di Furio F, Jing R, Ludwig-Kubinski A, Twaroski K, Urick A, et al. HNF4A regulates the formation of hepatic progenitor cells from human iPSC-derived endoderm by facilitating efficient recruitment of RNA pol II. Genes 2019;10:21.

[39] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci 2005;102:15545–15550.

[40] Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, et al. A human cell atlas of fetal chromatin accessibility. Science 2020;370:eaba7612.

[41] Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 2008;359:1995–2004.

[42] Boyault S, Rickman DS, De Reyniès A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology 2007;45:42–52.

[43] Hoshida Y, Nijman S, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–7392.

[44] De Matteis S, Ragusa A, Marisi G, De Domenico S, Casadei Gardini A, Bonafè M, et al. Aberrant metabolism in hepatocellular carcinoma provides diagnostic and therapeutic opportunities. Oxid Med Cell Longev 2018;2018:e7512159.

[45] Martínez-Reyes I, Chandel NS. Acetyl-CoA-directed gene transcription in cancer cells. Genes Dev 2018;32:463–465.

[46] Suzuki H, Kohjima M, Tanaka M, Goya T, Itoh S, Yoshizumi T, et al. Metabolic alteration in hepatocellular carcinoma: mechanism of lipid accumulation in well-differentiated hepatocellular carcinoma. Can J Gastroenterol Hepatol 2021;2021:e8813410.

[47] Jühling F, Hamdane N, Crouchet E, Li S, Saghire HE, Mukherji A, et al. Targeting clinical epigenetic reprogramming for chemoprevention of metabolic and viral hepatocellular carcinoma. Gut 2021;70:157–169.

[48] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739–1740.