Original Articles

Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer

Short title: Single-cell transcriptomics on liver cancer

Daniel Wai-Hung Ho[a, b, 1]

Yu-Man Tsui[a, b, 1]

Karen Man-Fong Sze[a, b]

Lo-Kong Chan[a, b]

Tan-To Cheung[c]

Eva Lee[a, b]

Pak-Chung Sham[d]

Stephen Kwok-Wing Tsui[e]

Terence Kin-Wah Lee[f]

Irene Oi-Lin Ng[a, b, *]

iolng@hku.hk

[a]Department of Pathology, The University of Hong Kong, Hong Kong

[b]State Key Laboratory ~~for~~ of Liver Research, The University of Hong Kong, Hong Kong

[c]Department of Surgery, The University of Hong Kong, Hong Kong

[d]Department of Psychiatry, The University of Hong Kong, Hong Kong

[e]School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong

[f]Department of Applied Biology and Chemical Technology, Hong Kong Polytechnic University, Hong Kong

[*]Corresponding author. Room 7-13, Block T, Queen Mary Hospital, Pokfulam, Hong Kong.

[1]Contributed equally.

Abstract

Hepatocellular carcinoma (HCC) is heterogeneous, rendering its current curative treatments ineffective. The emergence of single-cell genomics represents a powerful strategy in delineating the complex molecular landscapes of cancers. In this study, we demonstrated the feasibility and merit of using single-cell RNA sequencing to dissect the intra-tumoral heterogeneity and analyze the single-cell transcriptomic landscape to detect rare cell subpopulations of significance. Exploration of the inter-relationship among liver cancer stem cell markers showed two distinct major cell populations according to *EPCAM* expression, and the $EPCAM^+$ cells had upregulated expression of multiple oncogenes. We also identified a $CD24^+/CD44^+$-enriched cell subpopulation within the $EPCAM^+$ cells which had specific signature genes and might indicate a novel stemness-related cell subclone in HCC. Notably, knockdown of signature gene *CTSE* for $CD24^+/CD44^+$ cells significantly reduced self-renewal ability on HCC cells *in vitro* and the stemness-related role of *CTSE* was further confirmed by *in vivo* tumorigenicity assays in nude mice. In summary, single-cell genomics is a useful tool to delineate HCC intratumoral heterogeneity at better resolution. It can identify rare but important cell subpopulations, and may guide

better precision medicine in the long run.

Keywords: Single-cell sequencing; Tumor heterogeneity; HCC; Cancer stem cell; Cancer stemness

1 Introduction

Hepatocellular carcinoma (HCC) is a prevalent cancer and one of the leading causes of cancer death worldwide [1,2]. It has a poor prognosis and only few effective treatment options are available. In the recent decade, use of next generation sequencing (NGS) has successfully delineated the molecular landscapes of HCC. The identification of numerous genomic, transcriptomic and epigenomic alterations has revolutionized the understanding of HCC [3] and alerted us for the need of changing from the traditional "one-size-fits-all" treatment rationale to the strategy of stratifying patients with distinct biomarkers, e.g. *TSC1/2* mutations in dictating mTOR inhibitor hypersensitivity [4]. It is generally accepted that HCC tumors display high degree of both inter-tumoral and intra-tumoral heterogeneity, rendering existing treatment modalities ineffective [5]. With the inherent limitation of bulk-cell sequencing approach, which averages and masks signals from individual cells, cancer genomics should ideally be tackled on a single-cell basis.

In this study, we made use of the single-cell RNA sequencing (scRNA-seq) to determine the single-cell transcriptomic landscape of individual cells from HCC tumor. It serves as a proof-of-concept investigation, providing evidence to support the feasibility and advantages of single-cell genomics in delineating intra-tumoral heterogeneity and identifying rare cell subclones in HCC tumors. More importantly, we made use of the uniqueness of single-cell genomics to demonstrate the inter-relationship among different liver cancer stem cell (CSC) markers. We identified two major HCC cell populations characterized by differential *EPCAM* expression and a $CD24^+/CD44^+$-enriched stemness-related rare cell subclone with specific oncogenic gene expression signature, which distinguishes our study from the other recent reports [6,7].

2 Materials and methods

Please refer to the supplementary information for details on HCC specimen, sample preparation, experimental procedures of single-cell capture and sequencing, bioinformatics and statistical analyses, CSC marker-sorted transcriptome sequencing, *in vitro* sphere formation assays and *in vivo* tumorigenicity assays.

3 Results

3.1 HCC single cell isolation by fluidigm C1 and sequencing

After subjecting the integrated fluidic circuit (IFC)-captured HCC cells for further viability and multiplet checking, we isolated a total of 153 HCC single cells. NGS run was subsequently performed on these 153 HCC single-cell RNA samples. After removal of potential mouse cell contamination, we had data on a finalized set of 139 human HCC single cells. Based on the preliminary saturation analysis, sequencing depth of 1 M reads per cell reached saturation for gene expression detection using the FPKM threshold of 1 (Supplementary Fig. 1). The 139 HCC single cells were sequenced with a median depth of 5.5 M reads (range: 1M–16 M); hence the sequencing depth was adequate to cover the transcriptome.

3.2 Determination of HCC intra-tumoral heterogeneity using single-cell transcriptomic landscape

Based on the global protein-coding gene profiles of 139 HCC single cells, we performed unsupervised cell stratification analyses (h-clustering, PCA and t-SNE) (Supplementary Fig. 2). Interestingly, all the analyses consistently revealed the presence of 3 major cell clusters of HCC single cells (Supplementary Fig. 2A; Supplementary Figs. 2B and 2C, upper panel). In order to further determine the inter-relationship among the 3 analyses, we correlated the h-clustering result with both PCA and t-SNE plots (Supplementary Figs. 2B and 2C, lower panel). The HCC single cells tended to cluster together in the PCA and t-SNE plots according to the h-clustering grouping into 3 loosely separate cell clusters. Taken together, the various analyses of global protein-coding gene profiles consistently indicate that there existed no more than 3 major cell populations in the HCC single cells.

More importantly, to reduce the data dimensionality and account for the noisy data generated by scRNA-seq, we determined a subset of hypervariable genes that distinguished themselves among the genes of similar expression levels. They more likely represented subtle but genuine biological expression variations among the HCC single cells, after taking the technical noise into account. We identified 71 hypervariable genes (Supplementary Fig. 3A) and subjected their profiles to separate the same 139 HCC single cells. Based on cell stratification analyses, cells were then confined and substantiated into 2 distinct cell clusters (Supplementary Figs. 3B–3D). The data also indicate the presence of a cell sub-cluster H1A, which existed as a minor sub-branch in the h-clustering (dashed red line box in upper panel of Supplementary Fig. 3B) and its cells displayed enriched gene expression signature in a subset of hypervariable genes (*S100A6*, *VIM*, *CD44*, *CTSE* and *KRT20*) (dashed red line box in lower panel of Supplementary Fig. 3B), suggesting that there might be a rare cell subpopulation in the tumor. To further validate the existence of rare cell subpopulation, we performed GiniClust analysis, which identified high Gini genes ($n = 628$) that had significant higher Gini index values, indicating major variability among cells. Using gene expression profile of the high Gini genes, we

validated the existence of rare cell subpopulation (cell cluster GC1A) (dashed red line box in upper panel of Supplementary Fig. 4). Given that cell cluster H1A (Supplementary Fig. 3B) was identical to the cell cluster GC1A (Supplementary Fig. 4), the coherent findings essentially indicate the possible existence of a rare (<5%) cell subpopulation within the HCC tumor studied. Besides, the expression of stemness-related gene markers *EPCAM* was enriched in cell cluster H1, while *CD44* was enriched in both cell cluster H1A (which is equivalent to cell cluster GC1A) (indicated by red arrows in Supplementary Fig. 3B). We noted major (H1 or GC1) as well as rare (H1A or GC1A) cell clusters, which were differentially marked by expressional enrichment of stemness-related gene markers (*EPCAM* and *CD44*, respectively). It is likely that the HCC single cells and the represented cell subpopulations differed by variation in the expression of stemness-related genes.

3.3 Expression of stemness-related gene markers and their inter-relationship

As the data obtained implied subtle differences among the HCC single cells in terms of stemness-related gene expression, we further focused our scRNA-seq exploration to liver cancer stem cell (CSC) markers and their mutual expression pattern. Common liver CSC gene markers include *EPCAM*, *CD13*, *CD24*, *CD44*, *CD47*, *CD90* and *CD133* [8]. We found that *EPCAM* and *CD44* were enriched in restricted cell cluster/sub-cluster. In general, there was only a mild degree of correlation among the liver CSC markers. However, there was a positive correlation between *EPCAM* and *CD24* (Pearson $r = 0.26$), and a negative correlation between *EPCAM* and *CD13* (Pearson $r = -0.25$) (Supplementary Fig. 5). HCC single cells were clustered according to gene expression enrichment for multiple liver CSC markers (*EPCAM*, *CD13*, *CD24*, *CD44*, *CD47* and *CD133*) (*CD90* was discarded from our analysis due to its undetectable expression among the HCC cells examined) and stratified into liver CSC groups (Fig. 1A, upper panel). We used the term liver CSC group due to the cell stratification was based on expression pattern of liver CSC markers, but it did not necessarily imply all the cells studied were liver CSCs. HCC single cells were also classified into enrichment positive and negative cells based on *EPCAM*, *CD13*, *CD24*, *CD44* and *CD47* single-marker expression (Fig. 1A, upper panel) (There were too few CD133-expressing cells for derivation of the enrichment pattern). The majority of the cells were enriched in expression from a single liver CSC marker (group 1: *EPCAM*; group 6: *CD13*) to the maximum of 3 markers (group 3: *EPCAM*, *CD24* and *CD44*) (Fig. 1A). Such admixed nature of HCC single cells serves as another evidence illustrating the intra-tumoral heterogeneity of HCC tumor. In fact, most of the HCC single cells expressed more than one liver CSC marker (Fig. 1A, lower panel). Indeed, using traditional FACS strategy, only limited numbers of liver CSC markers can concurrently be used and it is technically challenging to distinguish or identify rare cell subpopulations using a panel of multiple markers. Despite the apparent differences in terms of *CD24*, *CD44* and *CD47* expressions, cells of liver CSC groups 1–4 would be regarded equally as $EPCAM^+$ cells using FACS, which would then undesirably mask the underlying important molecular and cellular characteristics. Hence, the use of single-cell genomics approach can provide comprehensive high-dimension profiling data on a large combination of liver CSC markers for more precise molecular sub-typing of HCC cells. The findings suggested that HCC single cells were heterogeneous and could be separated into different subpopulations based on intrinsic variations in liver CSC marker expression.

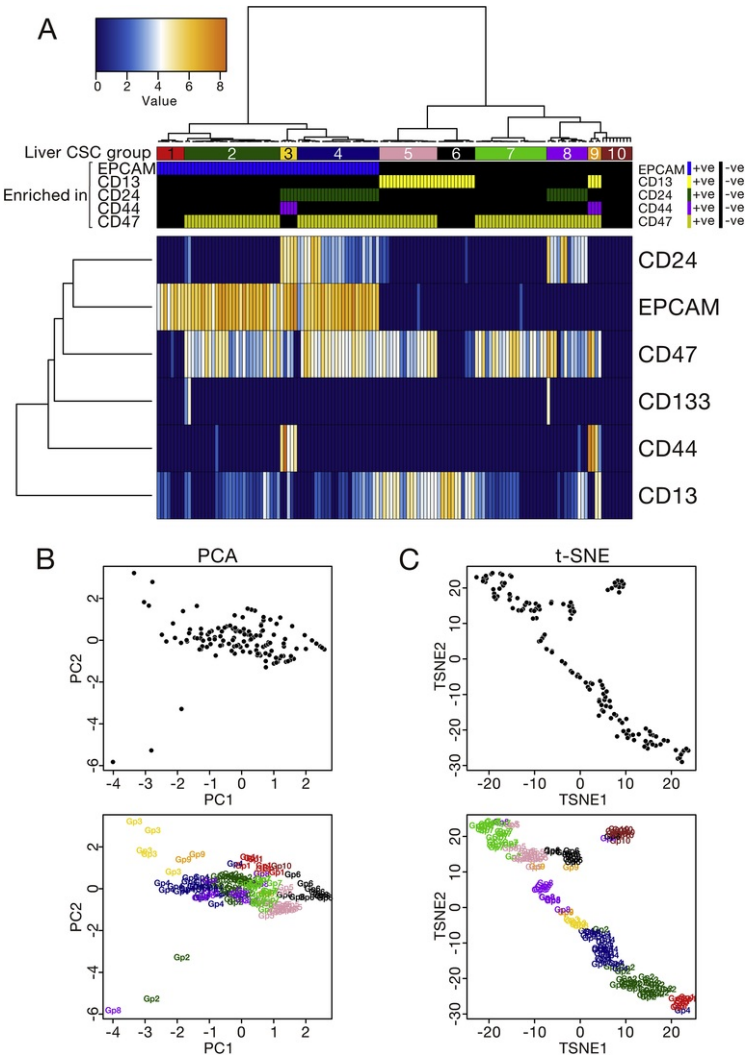


Fig. 1 Inter-relationship of HCC single cells in terms of liver CSC marker profiles and the enrichment pattern of individual liver CSC markers. (A) Based on the gene expression enrichment pattern of various liver CSC markers, HCC single cells were categorized into liver CSC group (10 groups; according to the h-clustering and expression pattern of liver CSC markers) and defined their enrichment of gene expression of liver CSC markers (*EPCAM*, *CD13*, *CD24*, *CD44* and *CD47*). Liver CSC groups differed in gene expression enrichment from a single to multiple liver CSC markers. (B and C) Similarly, PCA and t-SNE analyses suggested concordant separation of HCC single cells according to liver CSC groups, revealing intrinsic differences in liver CSC marker expression among potentially different tumor cell populations. Intriguingly, liver CSC group 3 cells relatively stood out among the HCC cells in the PCA plot (denoted by gold color). On the other hand, t-SNE revealed that liver CSC group 1–4 cells were distinguished from the remaining groups, with liver CSC group 8 cells being the intermediate. Liver CSC group 1–4 cells were essentially enriched with *EPCAM* expression, while liver CSC group 3 cells were further enriched with *CD24* and *CD44* expressions. Colors and liver CSC group naming were consistent among panel A–C. (For interpretation of the references to color in this figure legend the reader is referred to the Web version of this article.)

alt-text: Fig. 1

3.4 Correlation of HCC heterogeneity pattern and liver CSC status suggested the existence of both major and rare cell subpopulations

Based on the profiles of global protein-coding genes and hypervariable genes, we demonstrated the intra-tumoral heterogeneity pattern among HCC single cells. Here, when the previous analyses (Supplementary Figs. 2 and 3) were additionally augmented with the information of liver CSC group and the enrichment status of individual liver CSC markers, we identified a cell cluster that clearly contained the *EPCAM* expression-enriched cells, while the remaining clusters correlated with, but to a lesser extent, *CD13* enrichment. Moreover, the clustering of *EPCAM* and *CD13* expression-enriched cells was roughly non-overlapping, suggesting the potentially distinctive expressions of *EPCAM* and *CD13* in different subpopulations of HCC cells (Fig. 2; Supplementary Figs. 6–10). More importantly, in addition to the major cell clusters differed by *EPCAM* expression, we noted the cell cluster H1A (equivalent to GC1A) contained entirely the liver CSC group 3 cells with enriched *EPCAM*, *CD24* and *CD44* expressions (dashed red line box in upper panel of both Fig. 2 and Supplementary Fig. 11). Our data implied that the distinctive expressions of *EPCAM* and *CD1*3 might represent different major subpopulations of the HCC cells, while the concurrent *EPCAM*, *CD24* and *CD44* expressions possibly indicate a rare cell subpopulation.

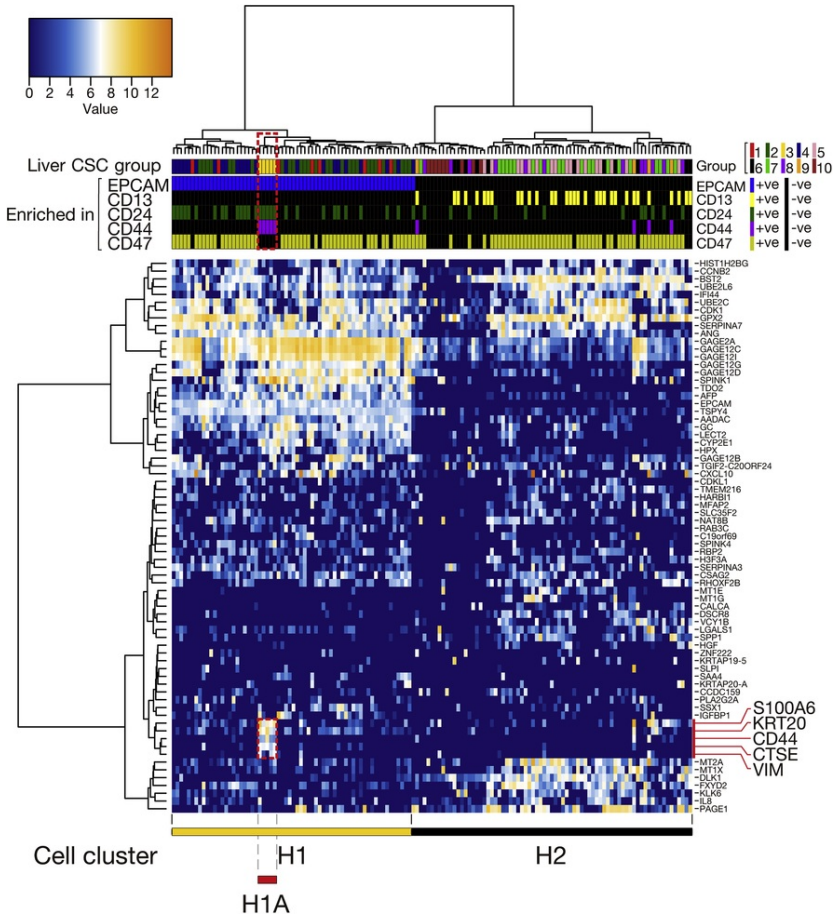


Fig. 2 Unsupervised h-clustering of hypervariable gene profiles of HCC single cells, correlated with liver CSC information. With the analysis using hypervariable genes, which could potentially better represent genuine biological expression variations, HCC single cells were further confined into 2 distinct cell clusters (H1 and H2). They were likely indicating 2 major cell subpopulations, which essentially differed in *EPCAM* expression, were contained in the HCC tumor. Besides, within the *EPCAM* expression-enriched major cell subpopulation (cell cluster H1), there existed a rare cell subpopulation (cell cluster H1A) with additional enrichment in *CD24* and *CD44* expressions and denoted as liver CSC group 3 (upper panel, dashed red line box). They displayed enriched gene expression signature in a subset of hypervariable genes (*S100A6, VIM, CD44, CTSE* and *KRT20*) (lower panel, dashed red line box). Taken together, both major and rare cell subpopulations could be identified in the HCC tumor, through scRNA-seq strategy. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

alt-text: Fig. 2

3.5 Mutational and expression landscapes between EPCAM⁺ and EPCAM⁻ cells

We examined the mutational landscape of HCC single cells and observed similar mutation profiles between EPCAM$^+$ and EPCAM$^-$ cells, suggesting that mutational acquisition may not be a major factor leading to their emergence. Similarly, EPCAM$^+$ and EPCAM$^-$ cells carried comparable levels of gene mutations (Supplementary Fig. 12). Furthermore, upon comparison of the gene expression profiles between the two groups of cells, we identified 180 and 126 genes that had significantly upregulated expression in EPCAM$^+$ and EPCAM$^-$ cells, respectively (Fig. 3). We further subjected these genes for gene set enrichment analysis. EPCAM+cells were enriched in the expression for genes regarding various metabolic processes, particularly lipid metabolism, while EPCAM$^-$ cells had upregulated genes related to the translation and RNA processing (Fig. 4A and B, left panel). Upon the protein-protein interaction analysis, elevated genes regarding EPCAM$^+$ and EPCAM$^-$ cells, as expected, displayed involvement of contrasting gene networks (Fig. 4A and B, right panel). To identify potentially important genes related to HCC development in the EPCAM$^+$ and EPCAM$^-$ cells, we examined our in-house and TCGA RNA-seq datasets of human HCCs and correlated the gene lists with elevated expression (as in Fig. 3) with the upregulated genes in human HCC samples. Thirty-one and 41 genes were consistently upregulated in both our in-house and TCGA HCC datasets and had significantly elevated expression levels in EPCAM$^+$ and EPCAM$^-$ cells, respectively (Supplementary Figs. 13 and 14). Five genes (EPCAM$^+$ cells: *NDC80*, *MEP1A* and *RRM2*; EPCAM$^-$ cells: *MDK* and *AKR1B10*) had more consistent and contrasting differential gene expression between human HCC and nontumorous liver tissues and concurrently showed elevated expression according to *EPCAM* enrichment status (Fig. 5A and B). To consolidate the scRNA-seq discovery, we also performed flow cytometry analysis on the same PDTX model. It confirmed the existence of 2 major subpopulations of HCC cells, namely EPCAM$^+$/CD13$^+$ (61.4%) and EPCAM$^-$/CD13$^+$ (34.3%), residing in the PDTX (Fig. 5C). Subsequent TCGA data analysis provided evidence supporting the possible oncogenic roles of the signature genes (*NDC80*, *MEP1A* and *RRM2*) for EPCAM$^+$ cells. Cases with upregulation of the signature genes for EPCAM$^+$ cells had significantly poorer overall survival (P = 8.54e-4) and disease-free survival (P = 0.004) (Supplementary Fig. 15). Overall, the data suggest that EPCAM$^+$ cells express specific oncogenic signature genes, which might confer oncogenic functions leading to a poorer prognosis.

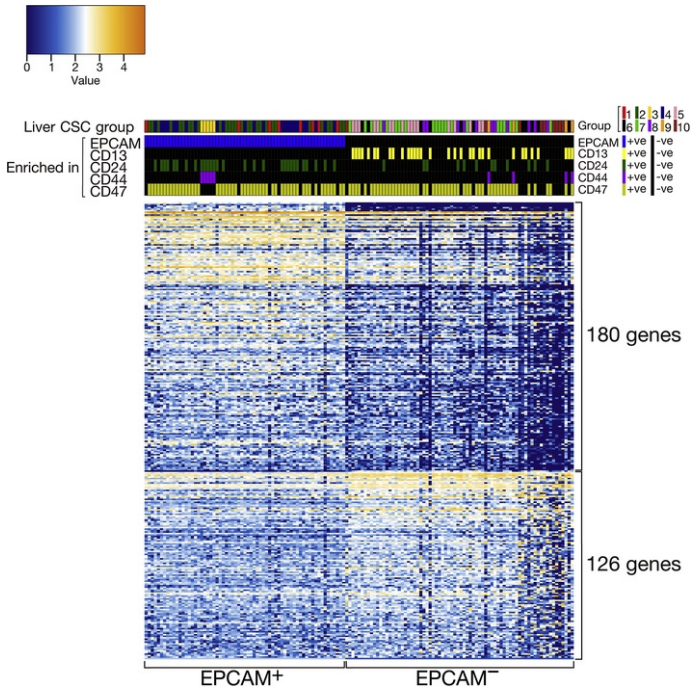


Fig. 3 Differential expression analysis between EPCAM$^+$ and EPCAM$^-$ cells. We stratified the 139 HCC single cells into EPCAM$^+$ and EPCAM$^-$ groups, according to EPCAM expression enrichment status. We identified 180 genes upregulated in EPCAM$^+$ cells and 126 genes upregulated in EPCAM$^-$ cells.

alt-text: Fig. 3

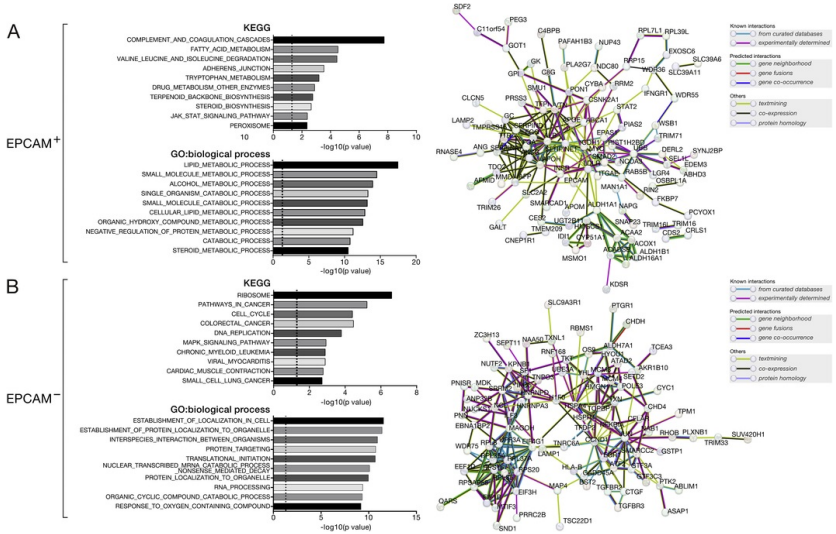


Fig. 4 Gene set enrichment and protein-protein interaction analyses for differentially expressed genes regarding EPCAM^{+} and EPCAM^{-} cells. (A and B) The 180 and 126 significantly upregulated genes in EPCAM^{+} and EPCAM^{-} cells respectively, were subjected to gene set enrichment analysis. Top 10 enriched gene sets based on KEGG canonical pathways and gene ontology: biological process (left panel). Protein-protein interaction network built by differentially expressed genes (right panel).

alt-text: Fig. 4

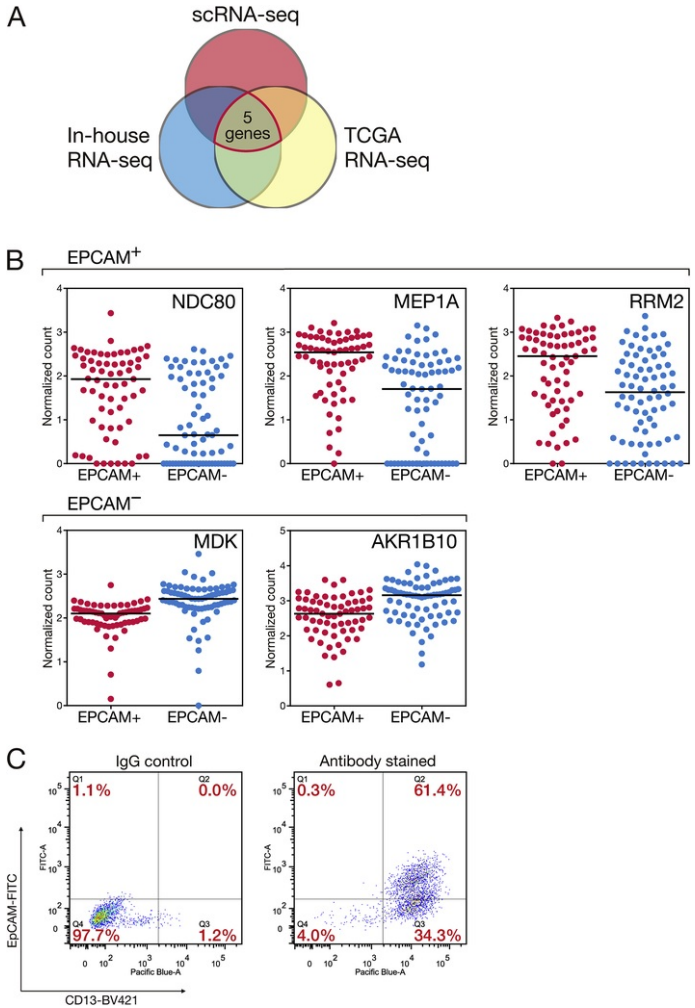


Fig. 5 Key differentially expressed genes for EPCAM$^+$ and EPCAM$^-$ cells and confirmation of major cell subpopulations stratified by EPCAM expression. (A) We integrated the differential expression analyses using scRNA-seq (EPCAM$^+$ vs EPCAM$^-$ cells), in-house RNA-seq (HCCs vs NTLs), TCGA RNA-seq (HCCs vs NTLs) and shortlisted 5 gene candidates. (B) There were 3 upregulated genes (*NDC80*, *MEP1A* and *RRM2*) for EPCAM$^+$ cells and 2 (*MDK* and *AKR1B10*) for EPCAM$^-$ cells. (C) Using the same HCC PDTX model used for scRNA-seq, FACS based on EPCAM and CD13 expressions confirmed the existence of EPCAM$^+$/CD13$^+$ and EPCAM$^-$/CD13$^+$ cells as major cell subpopulations.

alt-text: Fig. 5

3.6 Confirmation of gene expression from scRNA-seq using RNA-seq on CSC marker-sorted HCC cells

To substantiate the scRNA-seq results, we performed FACS on HCC tumors from the same passage of the PDTX model for liver CSC markers. By sorting out HCC cells of 4 different combinations of CD24 and CD44 CSC marker expression (CD24$^+$/CD44$^+$, CD24$^+$/CD44$^-$, CD24$^-$/CD44$^+$ and CD24$^-$/CD44$^-$) and subjecting them to RNA-seq, CD24$^+$/CD44$^+$-enriched HCC cells similarly had high expression of the signature genes (*CD44*, *KRT20*, *S100A6*, *VIM* and *CTSE*) as identified by scRNA-seq (Fig. 6A). We also performed immunohistochemistry (IHC) staining using antibodies for different signature genes in consecutive paraffin sections of the PDTX and were able to demonstrate overlap of IHC staining for CD44, KRT20, S100A6 and VIM in some HCC cells (Supplementary Fig. 16). Upon further exploration using TCGA HCC dataset, cases with upregulated expression levels of those signature genes for CD24$^+$/CD44$^+$ cells had significantly poorer disease-free survival ($P = 0.043$), while there was also a trend for poorer overall survival ($P = 0.091$) (Supplementary Fig. 17). Taken together, similar to the EPCAM$^+$ cells, CD24$^+$/CD44$^+$ cells also likely expressed

specific signature genes, which might potentially confer oncogenic functions leading to poorer prognosis of HCC.

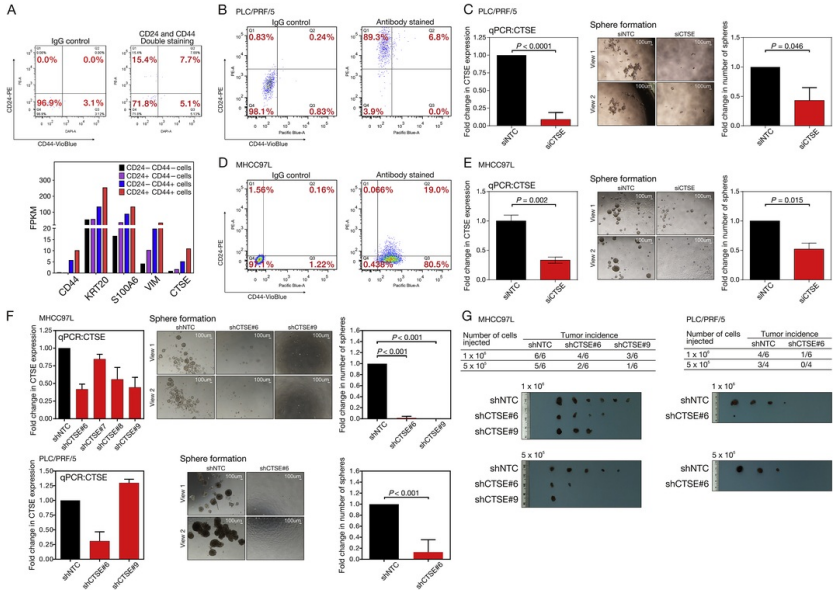


Fig. 6 Confirmation of single-cell transcriptome sequencing discovery on CD24$^+$/CD44$^+$ cell subpopulation and its signature gene *CTSE* in cancer stemness. (A) We extracted HCC cells according to various combinations of CD24 and CD44 expressions by FACS. RNA-seq data confirmed the gene expression signature for CD24$^+$/CD44$^+$ cells. (B and D) FACS confirmed the presence of CD24$^+$/CD44$^+$ cell subpopulation in PLC/PRF/5 and MHCC-97L HCC cells, which was a relevant cell model for subsequent experiment. (C and E) PLC/PRF/5 and MHCC-97L HCC cells displayed significant reduction in sphere formation ability upon siRNA knockdown on the signature gene *CTSE* specific to the enriched CD24$^+$/CD44$^+$ subpopulation. (F) Confirmation of reduced sphere formation ability upon stable knockdown of *CTSE* in MHCC-97L and PLC/PRF/5 HCC cells. (G) Stable knockdown of *CTSE* in MHCC-97L and PLC/PRF/5 HCC cells resulted in reduced tumor incidence.

alt-text: Fig. 6

3.7 Reduced sphere forming ability of HCC cells upon knockdown of CD24+/CD44 + signature genes

Among the few CD24$^+$/CD44$^+$ signature genes that potentially acted as downstream mediators of CD24$^+$/CD44$^+$ cells (*S100A6*, *VIM*, *CTSE* and *KRT20*), we examined their expressions using CSC marker-sorted RNA-seq. We selected *CTSE* (cathepsin E) for follow-up because it was the most upregulated (2.1-fold–13.2-fold) among the 4 signature genes in CD24$^+$/CD44$^+$-enriched cells, as compared to the others (cytoplasmic localization, Supplementary Fig. 18). Interestingly, we previously identified *CTSS* (cathepsin S), which also belongs to the cathepsin protease family, to be involved in the downstream mediation of CD47 signaling in HCC stemness [9]. We examined MHCC-97L and PLC/PRF/5 HCC cells and confirmed the presence of CD24$^+$/CD44$^+$ cell subpopulation (Fig. 6B and D). It served as a relevant cell model in our subsequent experiments. With transient si-knockdown of *CTSE* in HCC cells, we observed a significant reduction in the sphere formation ability, as compared to the siNTC control (Fig. 6C and E). We also sorted for CD24$^+$/CD44$^+$ cells using MHCC-97L HCC cells and consistently demonstrated a reduction in the sphere formation ability upon transient si-knockdown of *CTSE* (Supplementary Fig. 19). The findings preliminarily suggest a potential involvement of CD24$^+$/CD44$^+$-enriched cells in stemness-related functions, likely mediated by *CTSE*. As further substantiation of our findings, we verified the reduced sphere formation ability upon si-knockdown of *CTSE* in additional Huh7 HCC cells (Supplementary Fig. 20).

3.8 Suppressed sphere formation and tumor formation *in vivo* upon stable knockdown of CTSE

Among the four shCTSE sequences we created, shCTSE#6 and shCTSE#9 showed the best knockdown efficiency in MHCC-97L while shCTSE#6 showed the best in PLC/PRF/5 as compared to the respective shNTC (Fig. 6F). Therefore, we subjected these stable clones to sphere formation assays and *in vivo* tumorigenicity assays to assess the effects of *CTSE* on self-renewal ability of HCC cells. Stable knockdown of *CTSE* in MHCC-97L by shRNA #6 and #9 significantly suppressed sphere formation ($P < 0.001$) (Fig. 6F). Stable knockdown of *CTSE* by shRNA #6 in another HCC cell line PLC/PRF/5 also significantly suppressed sphere formation ($P < 0.001$) (Fig. 6F).

Furthermore, we found reduced tumor incidence upon knockdown of *CTSE* in MHCC-97L and PLC/PRF/5 cells. The knockdown of *CTSE* by shRNA #6 and #9 in MHCC-97L significantly reduced the number of tumors formed in both groups of nude mice injected with 1×10^6 and 5×10^5 cells as compared to the respective NTCs (Fig. 6G). Similar trend was also observed in PLC/PRF/5 cell line: the knockdown of *CTSE* by shRNA #6 markedly reduced the number

of tumors formed in both groups of nude mice injected with 1×10^6 and 5×10^5 cells as compared to the respective NTCs (Fig. 6G). Collectively, these functional assay results indicate an important role of *CTSE* in self-renewal ability of HCC cells.

4 Discussion

Intra-tumoral heterogeneity is a common phenomenon in cancers and it is a major hurdle to allow for disease cure. Previous studies have suggested that multiple subpopulations of cancer cells reside in HCC tumors and they display intrinsic differences in tumor behavior. CSCs are specific subpopulations of cells that are believed to be responsible for tumor relapse, metastasis and chemoresistance. Single-cell genomics offers an excellent modality to delineate the intra-tumoral heterogeneity of HCC, as compared to bulk-cell approach. In this proof-of-concept study, we successfully demonstrated the usefulness of scRNA-seq in complementing the conventional FACS-based approach in studying the inter-relationship between liver CSC markers. In fact, scRNA-seq offers unprecedented depth and breadth in pinpointing a variety of liver CSC markers. Not only is it capable of targeting many different markers at once, but it also provides better resolution and sensitivity. Through a simple power estimation taking subclonal frequency into account [10], our study achieved satisfactory statistical power (>95%) to detect a 2% cell subclone. Although more complicated issues e.g. false positives and negatives, as well as sequencing errors, had not been taken into consideration, our data clearly indicate the feasibility of scRNA-seq in detecting and studying rare cell subtypes. In contrast, the detection or examination of rare cell subclone (<5%) is usually not feasible with either bulk-cell or FACS-based strategies. In this study, it was not our primary focus to determine which liver CSC groups of HCC cells were genuine liver CSCs. Thus, the research question of which subpopulation of HCC cells is the most important in HCC, while interesting, is beyond the scope of our current investigation and requires a separate study.

Regarding the HCC PDTX model used in the current study, it has been utilized in studying various aspects of HCC, including CD47 upregulation and sorafenib resistance [11], combination treatment of anti-CD47 antibody and doxorubicin [12], and stearoyl-CoA desaturase (SCD1) overexpression and sorafenib resistance [13]. Previously, we have performed whole-transcriptome sequencing on 4 pairs of PDTXs and their corresponding primary tumors and data demonstrated PDTX models could serve as good representation of the corresponding primary tumors (data not shown). Importantly, by having the same PDTX model subjecting to cell sorting experiment and bulk-cell transcriptome sequencing, we provided the necessary matched validation to confirm our discovery observations using scRNA-seq. Further examination in additional clinical specimens is therefore justified and awaited.

In our current study, we demonstrated HCC single cells generally stratified according to *EPCAM* and *CD13* expressions and EPCAM+/CD13 + and EPCAM-/CD13 + cells were 2 major subpopulations residing in this HCC tumor. EPCAM and CD13 are individually recognized as liver CSC markers [14,15]. In accordance with Qin et al. [16], our scRNA-seq data also suggested EPCAM + cells were enriched in expression for multiple genes regarding metabolic processes, particularly lipid metabolism. Apart from involving in various metabolic processes as suggested by our gene set enrichment analysis, $EPCAM^+$ cells were particularly prominent in the expression of *NDC80*, *MEP1A* and *RRM2* genes. *NDC80* encodes a component of the kinetochore complex which links centromeres to mitotic spindle microtubules. It was overexpressed in various cancers, including HCC [17–20]. Similarly, *MEP1A* (encodes for meprin A subunit alpha) is a metalloprotease, which was recently found to contribute to tumor progression and predict poor clinical outcome in HCC [21]. It could induce HCC cell migration and invasion under the regulation of Reptin [22]. Moreover, *RRM2*, together with *RRM1*, encodes the two subunits for ribonucleotide reductase. This enzyme catalyzes the conversion of ribonucleotides into deoxyribonucleotides, which are important in DNA replication and repair [23]. Notably, RRM2 expression has been suggested to be a useful biomarker and treatment target in many cancers and its upregulation might contribute to chemoresistance and anti-apoptosis [24–28]. Altogether, $EPCAM^+$ cells seem to specifically have upregulated expression of multiple oncogenes that are likely to be involved in various aspects of HCC malignancy, including metabolism [29]. On the other hand, our data on EPCAM-cells suggested their involvement in cell cycle, which also in echo with the findings from a previous study on EPCAM-antibody induced cell cycle progression [30].

In addition, we have identified a rare cell subpopulation of $CD24^+/CD44^+$ cells. They have a specifically enriched gene expression signature. CD24 and CD44 are cell surface proteins and are liver CSC makers on their own [8]. CD24 and CD44 both are also common cell surface markers that expressed in many other cancer types. Independently, single-marker expression of them was frequently used to mark cancer stem cells and predict prognostic outcomes. For CD44, human anti-CD44 monoclonal antibody (RG7356) has already been implicated for the treatment of hematological malignant diseases [31,32] and there are several other potential drugs that targeting CD44 by conjugates of hyaluronic acid (ligand for CD44) [33]. Nevertheless, there were inconsistencies on the functional roles CD24 and CD44 in different cancers [34], suggesting single-marker expression of them may not be adequate in distinguish the genuine cancer stem cells [35]. Importantly, the combination of CD24 and CD44 expression e.g. CD24+/CD44+, CD24-/CD44+, has been used to define CSCs in various cancers, including breast, ovarian, prostate, pancreatic, colorectal, lung and renal [34], but this has not yet been reported in HCC so far [36]. Therefore, the identification of stemness-related CD24+/CD44 + cell subpopulation in HCC represents a likely unique finding that distinguish our current study from previous reports on single-marker basis. Interestingly, coincide with the recent report by Wei et al. studying ovarian cancer [37], our findings similarly suggesting that cancer cells concurrently expressing EPCAM, CD24 and CD44 are likely to possess stemness-related characteristics. $CD24^+/C44^+$-enriched cells have specific expression for signature genes of *KRT20*, *S100A6*, *VIM* and *CTSE*. Regarding *KRT20* (encoding keratin 20), previous studies have suggested elevated KRT20 expression in carbon tetrachloride-induced liver tumors [38] and exosomes derived from HCC cells [39]. On the other hand, S100A6 belongs to the S100 protein family. S100 family members are commonly overexpressed in various cancers [40]. Moreover, S100A6 was demonstrated to be upregulated and its silencing could inhibit HCC cell growth and motility, suggesting S100A6 may be involved in promotion and progression of HCC [41]. *VIM* (encoding vimentin) is a component of intermediate filament for cytoskeleton and is participated in epithelial-to-mesenchymal transition. It is functionally implicated in multiple aspects of cancer, particularly metastasis [42,43]. *CTSE* (encoding cathepsin E) is a major intracellular aspartic protease that is mainly present in immune cells, but is also detected in gastric epithelial cells and osteoclasts [44]. It is overexpressed in pancreatic

ductal adenocarcinoma and gastric cancer, and the family of cathepsins is highly expressed in various human cancers and associated with tumor metastasis [45,46]. Our *in vitro* sphere formation assays and *in vivo* tumorigenicity assays also support a contributory role of *CTSE* in HCC self-renewal ability. Taken together, our findings implicate the role of $CD24^+/CD44^+$ cells and their signature genes for cancer stemness in HCC.

Intriguingly, as demonstrated using TCGA data, the signature genes for both $EPCAM^+$ and $CD24^+/CD44^+$ cells were collectively upregulated in a substantial proportion of HCC cases. Moreover, their upregulated expressions were either associated or tended to be associated with a poorer prognosis (both overall and disease-free survival rates), indicating discovery obtained from single-cell genomics may likely reveal relevant clinical consequence. These findings further warrant the practical use of single-cell genomics to understand HCC tumor biology and guide future treatment.

In summary, our study has provided evidence highlighting the feasibility, merit and necessity of adopting single-cell genomics in studying HCC, with specific focus on tumor heterogeneity and cancer stemness. More importantly, the power of single-cell genomics has enabled the identification of rare cell subpopulations, and the determination of inter-relationship between different liver CSC markers and their unique gene expression signatures. The data will provide important clues for the targeting and eradication of liver CSCs. We believe single-cell genomics will provide necessary insight to complement this challenging exploitation.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgement

We thank Angela Wu for her advice on the data analysis. This study was supported by the Hong Kong Research Grants Council Theme-based Research Scheme (T12-704116-R), Innovation and Technology Commission grant for State Key Laboratory of Liver Research, and University Development Fund of The University of Hong Kong. IOL Ng is Loke Yew Professor in Pathology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.canlet.2019.06.002.

References

[1] H.B. El-Serag, Hepatocellular carcinoma, *N. Engl. J. Med.* **365**, 2011, 1118–1127.

[2] A. Villanueva and J.M. Llovet, Liver cancer in 2013: mutational landscape of HCC--the end of the beginning, *Nat. Rev. Clin. Oncol.* **11**, 2014, 73–74.

[3] D.W. Ho, R.C. Lo, L.K. Chan and I.O. Ng, Molecular pathogenesis of hepatocellular carcinoma, *Liver Cancer* **5**, 2016, 290–302.

[4] D.W.H. Ho, L.K. Chan, Y.T. Chiu, I.M.J. Xu, R.T.P. Poon, T.T. Cheung, C.N. Tang, V.W.L. Tang, I.L.O. Lo, P.W.Y. Lam, D.T.W. Yau, M.X. Li, C.M. Wong and I.O.L. Ng, TSC1/2 mutations define a molecular subset of HCC with aggressive behaviour and treatment implication, *Gut* **66**, 2017, 1496–1506.

[5] L.C. Lu, C.H. Hsu, C. Hsu and A.L. Cheng, Tumor heterogeneity in hepatocellular carcinoma: facing the challenges, *Liver Cancer* **5**, 2016, 128–138.

[6] H. Zheng, Y. Pomyen, M.O. Hernandez, C. Li, F. Livak, W. Tang, H. Dang, T.F. Greten, J.L. Davis, Y. Zhao, M. Mehta, Y. Levin, J. Shetty, B. Tran, A. Budhu and X.W. Wang, Single cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma, *Hepatology* 2018.

[7] M. Duan, J. Hao, S. Cui, D.L. Worthley, S. Zhang, Z. Wang, J. Shi, L. Liu, X. Wang, A. Ke, Y. Cao, R. Xi, X. Zhang, J. Zhou, J. Fan, C. Li and Q. Gao, Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing, *Cell Res.* **28**, 2018, 359.

[8] T. Yamashita and X.W. Wang, Cancer stem cells in the development of liver cancer, *J. Clin. Investig.* **123**, 2013, 1911–1918.

[9] T.K. Lee, V.C. Cheung, P. Lu, E.Y. Lau, S. Ma, K.H. Tang, M. Tong, J. Lo and I.O. Ng, Blockade of CD47-mediated cathepsin S/protease-activated receptor 2 signaling provides a therapeutic target for hepatocellular carcinoma, *Hepatology* **60**, 2014, 179–191.

[10] N.E. Navin, The first five years of single-cell cancer genomics and beyond, *Genome Res.* **25**, 2015, 1499–1507.

[11] J. Lo, E.Y. Lau, R.H. Ching, B.Y. Cheng, M.K. Ma, I.O. Ng and T.K. Lee, Nuclear factor kappa B-mediated CD47 up-regulation promotes sorafenib resistance and its blockade synergizes the effect of sorafenib in

hepatocellular carcinoma in mice, *Hepatology* **62**, 2015, 534–545.

[12] J. Lo, E.Y. Lau, F.T. So, P. Lu, V.S. Chan, V.C. Cheung, R.H. Ching, B.Y. Cheng, M.K. Ma, I.O. Ng and T.K. Lee, Anti-CD47 antibody suppresses tumour growth and augments the effect of chemotherapy treatment in hepatocellular carcinoma, *Liver Int. : official journal of the International Association for the Study of the Liver* **36**, 2016, 737–745.

[13] M.K.F. Ma, E.Y.T. Lau, D.H.W. Leung, J. Lo, N.P.Y. Ho, L.K.W. Cheng, S. Ma, C.H. Lin, J.A. Copland, J. Ding, R.C.L. Lo, I.O.L. Ng and T.K.W. Lee, Stearoyl-CoA desaturase regulates sorafenib resistance via modulation of ER stress-induced differentiation, *J. Hepatol.* **67**, 2017, 979–990.

[14] T. Yamashita, J. Ji, A. Budhu, M. Forgues, W. Yang, H.Y. Wang, H. Jia, Q. Ye, L.X. Qin, E. Wauthier, L.M. Reid, H. Minato, M. Honda, S. Kaneko, Z.Y. Tang and X.W. Wang, EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features, *Gastroenterology* **136**, 2009, 1012–1024.

[15] N. Haraguchi, H. Ishii, K. Mimori, F. Tanaka, M. Ohkuma, H.M. Kim, H. Akita, D. Takiuchi, H. Hatano, H. Nagano, G.F. Barnard, Y. Doki and M. Mori, CD13 is a therapeutic target in human liver cancer stem cells, *J. Clin. Investig.* **120**, 2010, 3326–3339.

[16] X.Y. Qin, N. Dohmae and S. Kojima, Reply to Yoshida: liver cancer stem cells: identification and lipid metabolic reprogramming, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2018, E6390–e6391.

[17] Q.-C. Meng, H.-C. Wang, Z.-L. Song, Z.-Z. Shan, Z. Yuan, Q. Zheng and X.-Y. Huang, Overexpression of NDC80 is correlated with prognosis of pancreatic cancer and regulates cell proliferation, *American Journal of Cancer Research* **5**, 2015, 1730–1740.

[18] E. Diaz-Rodriguez, R. Sotillo, J.M. Schvartzman and R. Benezra, Hec1 overexpression hyperactivates the mitotic checkpoint and induces tumor formation in vivo, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2008, 16719–16724.

[19] I. Bieche, S. Vacher, F. Lallemand, S. Tozlu-Kara, H. Bennani, M. Beuzelin, K. Driouch, E. Rouleau, F. Lerebours, H. Ripoche, G. Cizeron-Clairac, F. Spyratos and R. Lidereau, Expression analysis of mitotic spindle checkpoint genes in breast carcinoma: role of NDC80/HEC1 in early breast tumorigenicity, and a two-gene signature for aneuploidy, *Mol. Cancer* **10**, 2011, 23.

[20] L.L. Ju, L. Chen, J.H. Li, Y.F. Wang, R.J. Lu, Z.L. Bian and J.G. Shao, Effect of NDC80 in human hepatocellular carcinoma, *World J. Gastroenterol.* **23**, 2017, 3675–3683.

[21] H.Y. OuYang, J. Xu, J. Luo, R.H. Zou, K. Chen, Y. Le, Y.F. Zhang, W. Wei, R.P. Guo and M. Shi, MEP1A contributes to tumor progression and predicts poor clinical outcome in human hepatocellular carcinoma, *Hepatology* **63**, 2016, 1227–1239.

[22] O. Breig, M. Yates, V. Neaud, G. Couchy, A. Grigoletto, C. Lucchesi, J. Prox, J. Zucman-Rossi, C. Becker-Pauly and J. Rosenbaum, Metalloproteinase meprin alpha regulates migration and invasion of human hepatocarcinoma cells and is a mediator of the oncoprotein Reptin, *Oncotarget* **8**, 2017, 7839–7851.

[23] V. D'Angiolella, V. Donato, F.M. Forrester, Y.T. Jeong, C. Pellacani, Y. Kudo, A. Saraf, L. Florens, M.P. Washburn and M. Pagano, Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair, *Cell* **149**, 2012, 1023–1034.

[24] M.A. Rahman, A.R.M.R. Amin, D. Wang, L. Koenig, S. Nannapaneni, Z. Chen, Z. Wang, G. Sica, X. Deng, Z. Chen and D.M. Shin, RRM2 regulates bcl-2 in head and neck and lung cancers: a potential target for cancer therapy, *Clin. Cancer Res. : an official journal of the American Association for Cancer Research* **19**, 2013, 3416–3428.

[25] V.K. Grolmusz, K. Karászi, T. Micsik, E.A. Tóth, K. Mészáros, G. Karvaly, G. Barna, P.M. Szabó, K. Baghy, J. Matkó, I. Kovalszky, M. Tóth, K. Rácz, P. Igaz and A. Patócs, Cell cycle dependent RRM2 may serve as proliferation marker and pharmaceutical target in adrenocortical cancer, *American Journal of Cancer Research* **6**, 2016, 2041–2053.

[26] L. Wang, L. Meng, X.W. Wang, G.Y. Ma and J.H. Chen, Expression of RRM1 and RRM2 as a novel prognostic marker in advanced non-small cell lung cancer receiving chemotherapy, *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **35**, 2014, 1899–1906.

[27] N. Wang, T. Zhan, T. Ke, X. Huang, D. Ke, Q. Wang and H. Li, Increased expression of RRM2 by human papillomavirus E7 oncoprotein promotes angiogenesis in cervical cancer, *Br. J. Canc.* **110**, 2014, 1034–1044.

[28] Y. Yoshida, T. Tsunoda, K. Doi, Y. Tanaka, T. Fujimoto, T. Machida, T. Ota, M. Koyanagi, Y. Takashima, T. Sasazuki, M. Kuroki, A. Iwasaki and S. Shirasawa, KRAS-mediated up-regulation of RRM2 expression is essential for the proliferation of colorectal cancer cell lines, *Anticancer Res.* **31**, 2011, 2535–2539.

[29] M.A. Chaudry, K. Sales, P. Ruf, H. Lindhofer and M.C. Winslet, EpCAM an immunotherapeutic target for gastrointestinal malignancy: current experience and future challenges, *Br. J. Canc.* **96**, 2007, 1013–1019.

[30] K. Maaser and J. Borlak, A genome-wide expression analysis identifies a network of EpCAM-induced cell cycle regulators, *Br. J. Canc.* **99**, 2008, 1635–1643.

[31] G. D'Arena, G. Calapai and S. Deaglio, Anti-CD44 mAb for the treatment of B-cell chronic lymphocytic leukemia and other hematological malignancies: evaluation of WO2013063498, *Expert Opin. Ther. Pat.* **24**, 2014, 821–828.

[32] S.C. Ghosh, S. Neslihan Alpay and J. Klostergaard, CD44: a validated target for improved delivery of cancer therapeutics, *Expert Opin. Ther. Targets* **16**, 2012, 635–650.

[33] G. Mattheolabakis, L. Milane, A. Singh and M.M. Amiji, Hyaluronic acid targeting of CD44 for cancer therapy: from receptor biology to nanomedicine, *J. Drug Target.* **23**, 2015, 605–618.

[34] A. Jaggupilli and E. Elkord, Significance of CD44 and CD24 as cancer stem cell markers: an enduring ambiguity, *Clin. Dev. Immunol.* 2012, 708036, 2012.

[35] T.N. Flores-Téllez, S. Villa-Treviño and C. Piña-Vázquez, Road to stemness in hepatocellular carcinoma, *World J. Gastroenterol.* **23**, 2017, 6750–6776.

[36] N. Wang, S. Wang, M.-Y. Li, B.-G. Hu, L.-P. Liu, S.-L. Yang, S. Yang, Z. Gong, P.B.S. Lai and G.G. Chen, Cancer stem cells in hepatocellular carcinoma: an overview and promising therapeutic strategies, *Therapeutic advances in medical oncology* **10**, 2018, 1758835918816287-1758835918816287.

[37] X. Wei, D. Dombkowski, K. Meirelles, R. Pieretti-Vanmarcke, P.P. Szotek, H.L. Chang, F.I. Preffer, P.R. Mueller, J. Teixeira, D.T. MacLaughlin and P.K. Donahoe, Mullerian inhibiting substance preferentially inhibits stem/progenitors in human ovarian cancer cell lines compared with chemotherapeutics, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2010, 18874–18879.

[38] X. Chen, M. Yamamoto, K. Fujii, Y. Nagahama, T. Ooshio, B. Xin, Y. Okada, H. Furukawa and Y. Nishikawa, Differential reactivation of fetal/neonatal genes in mouse liver tumors induced in cirrhotic and non-cirrhotic conditions, *Cancer Sci.* **106**, 2015, 972–981.

[39] S. Wang, G. Chen, X. Lin, X. Xing, Z. Cai, X. Liu and J. Liu, Role of exosomes in hepatocellular carcinoma cell mobility alteration, *Oncol Lett* **14**, 2017, 8122–8131.

[40] A.R. Bresnick, D.J. Weber and D.B. Zimmer, S100 proteins in cancer, *Nat. Rev. Canc.* **15**, 2015, 96–109.

[41] Z. Li, M. Tang, B. Ling, S. Liu, Y. Zheng, C. Nie, Z. Yuan, L. Zhou, G. Guo, A. Tong and Y. Wei, Increased expression of S100A6 promotes cell proliferation and migration in human hepatocellular carcinoma, *J. Mol. Med. (Berl.)* **92**, 2014, 291–303.

[42] L. Hu, S.H. Lau, C.H. Tzang, J.M. Wen, W. Wang, D. Xie, M. Huang, Y. Wang, M.C. Wu, J.F. Huang, W.F. Zeng, J.S. Sham, M. Yang and X.Y. Guan, Association of Vimentin overexpression and hepatocellular carcinoma metastasis, *Oncogene* **23**, 2004, 298–302.

[43] A. Satelli and S. Li, Vimentin in cancer and its potential as a molecular target for cancer therapy, *Cell. Mol. Life Sci. : CM* **68**, 2011, 3033–3046.

[44] N. Zaidi and H. Kalbacher, Cathepsin E: a mini review, *Biochem. Biophys. Res. Commun.* **367**, 2008, 517–522.

[45] Z. Cruz-Monserrate, W.R. Abd-Elgaliel, T. Grote, D. Deng, B. Ji, T. Arumugam, H. Wang, C.H. Tung and C.D. Logsdon, Detection of pancreatic cancer tumours and precursor lesions by cathepsin E activity in mouse models, *Gut* **61**, 2012, 1315–1322.

[46] G.J. Tan, Z.K. Peng, J.P. Lu and F.Q. Tang, Cathepsins mediate tumor metastasis, *World J. Biol. Chem.* **4**, 2013, 91–101.

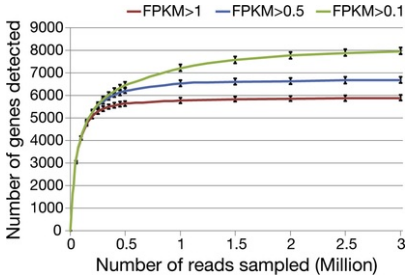
Appendix A. Supplementary data

The following are the supplementary data to this article:

Multimedia Component 1

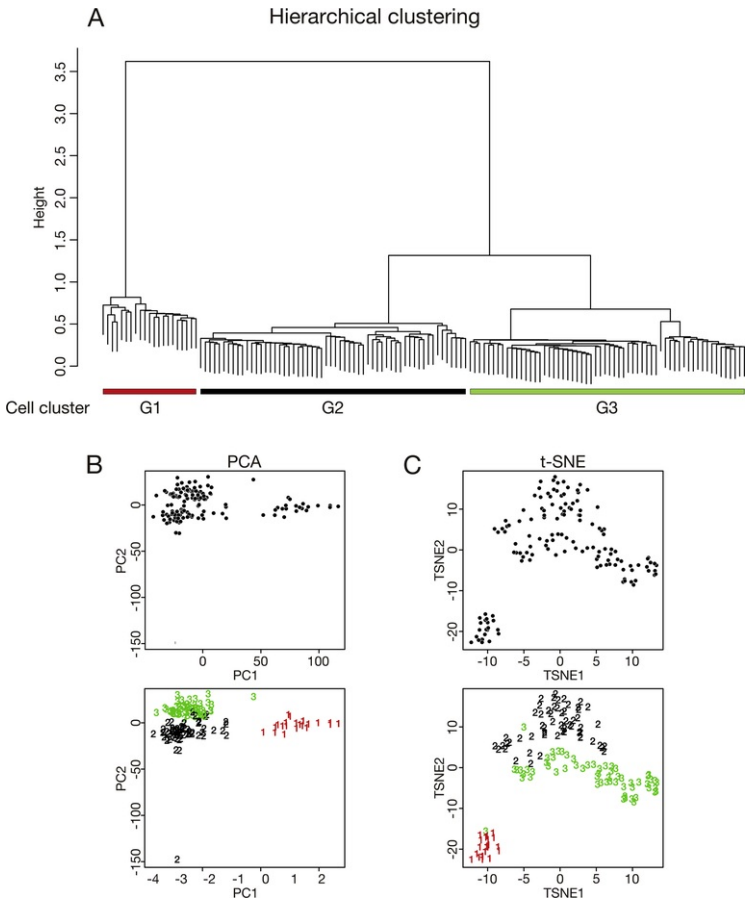
Multimedia component 1

alt-text: Multimedia component 1



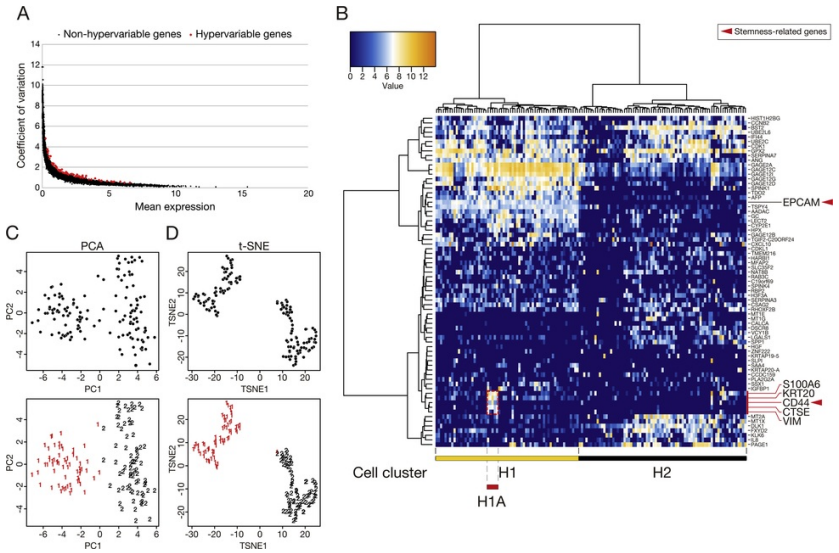
Supplementary Fig. 1 Saturation analysis of HCC single cells. Data were subsampled to estimate transcriptome coverage at various read depth, using different cutoffs for gene expression level.

alt-text: Supplementary Fig. 1



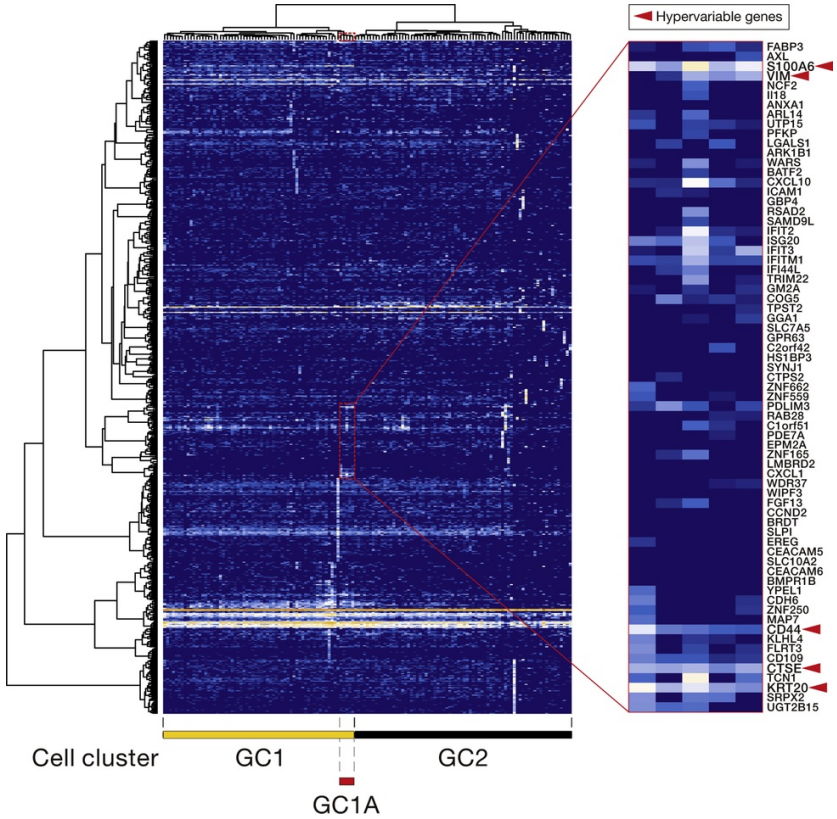
Supplementary Fig. 2 Global profiles of single-cell transcriptomic sequencing revealed intra-tumoral heterogeneity of HCC. (A) Unsupervised h-clustering suggested the existence of 3 major cell clusters, namely G1-G3. (B and C) PCA and t-SNE analyses similarly suggested 3 loosely distinguished cell clusters in the tumor sample (upper panel). Correlation between h-clustering grouping (G1-G3) with PCA and t-SNE results further suggested good consistency in classifying HCC single cells based on global transcriptomic landscape (lower panel).

alt-text: Supplementary Fig. 2



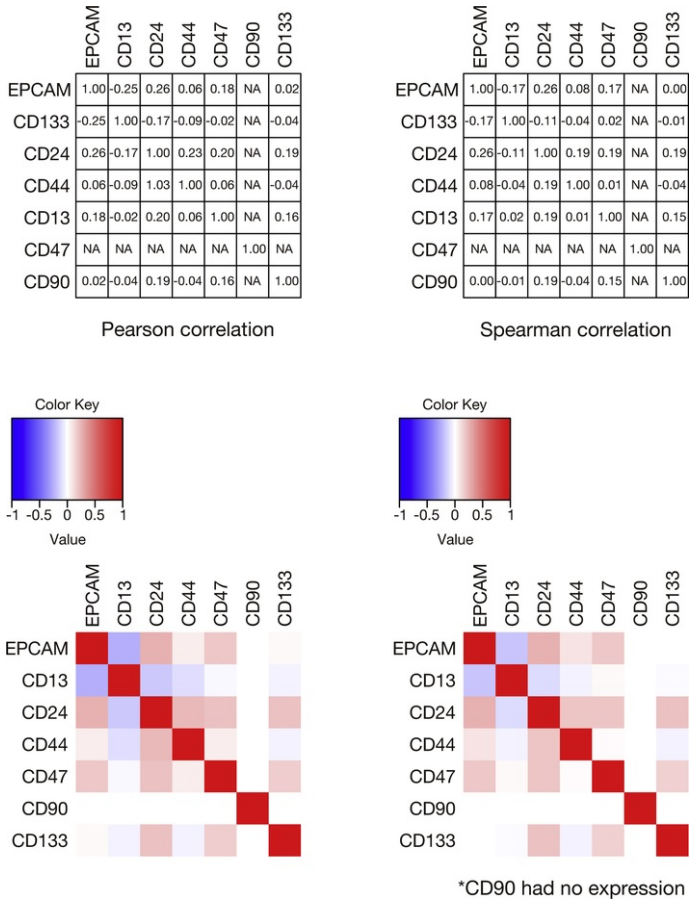
Supplementary Fig. 3 Profiles of hypervariable genes of HCC single cells further consolidated a heterogeneity pattern of HCC tumor and suggested the existence of rare cell subpopulation. (A) Identification of 71 hypervariable genes that displayed their CVs with 2.5 SD above the mean CV of corresponding reference gene sets. (B) Unsupervised h-clustering suggested the existence of 2 major cell clusters, namely H1 and H2. There was a minor cell sub-cluster of H1A within H1 (upper panel, dashed red line box), with corresponding enriched signature genes (*S100A6, VIM, CD44, CTSE* and *KRT20*) (lower panel, dashed red line box). Stemness-related genes *EPCAM* and *CD44* were indicated by red arrows. (C and D) PCA analysis and t-SNE analyses consistently suggested 2 major cell clusters in the tumor sample (upper panel). Correlation between h-clustering grouping (H1 and H2) with PCA and t-SNE results further suggested good consistency in classification of HCC single cells based on expression pattern of hypervariable genes (lower panel).

alt-text: Supplementary Fig. 3



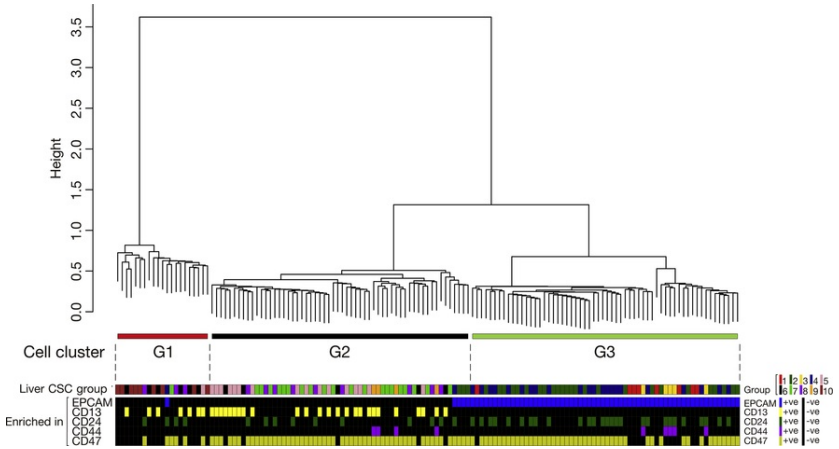
Supplementary Fig. 4 Profiles of high Gini genes of HCC single cells. Similar to the analysis of hypervariable genes, unsupervised h-clustering similarly suggested the existence of 2 major cell clusters (GC1 and GC2), with the existence of a minor cell cluster of GC1A (upper panel, dashed red line box) within GC1. GC1A contains the same set of HCC single cells as cell cluster H1A.

alt-text: Supplementary Fig. 4



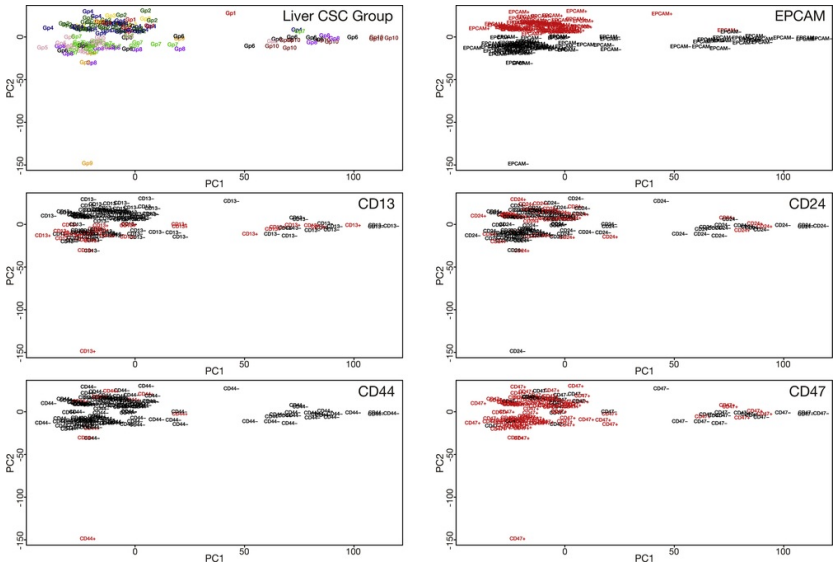
Supplementary Fig. 5 Expression correlation between liver CSC markers. There was consistently low degree of expression correlation between liver CSC markers based on both Pearson and Spearman correlation.

alt-text: Supplementary Fig. 5



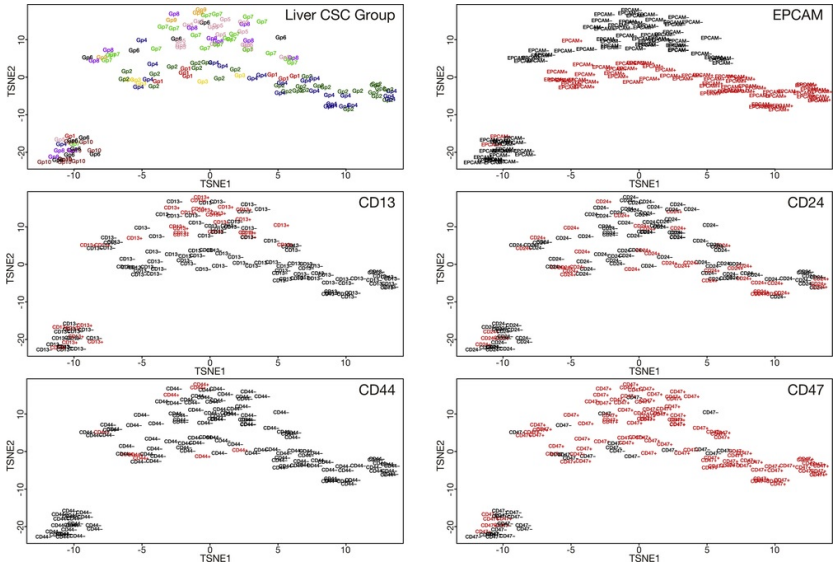
Supplementary Fig. 6 Unsupervised h-clustering of global gene profiles of HCC single cells, correlated with liver CSC information. Data suggested the existence of 3 major cell clusters, namely G1-G3. *EPCAM* expression-enriched cells predominantly contained in cell cluster G3, while the remaining cell clusters (G1 and G2) correlated with *CD13* enrichment but to a lesser extent. Moreover, their seemingly non-overlapping pattern suggests distinctive *EPCAM* and *CD13* expressions in different subpopulations of HCC cells in the tumor.

alt-text: Supplementary Fig. 6



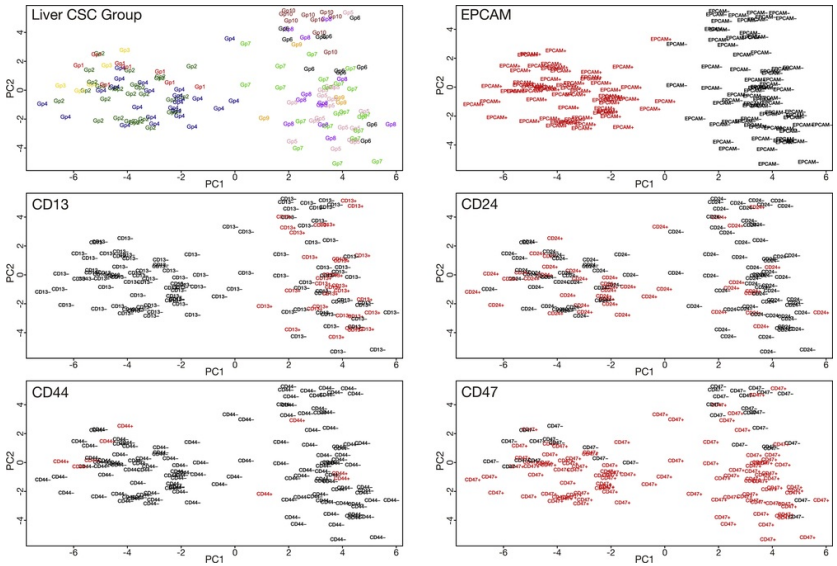
Supplementary Fig. 7 PCA analysis of global gene profiles of HCC single cells, correlated with liver CSC information. PCA plot based on global protein-coding gene profiles of HCC single cells was correlated with the status of liver CSC group (10 groups as same as in Fig. 1) and single-marker liver CSC expression enrichment (enrichment-positive cells marked as red). Upper left cell clusters consisted of mainly liver CSC group 1–4 cells and were EPCAM expression-enriched.

alt-text: Supplementary Fig. 7



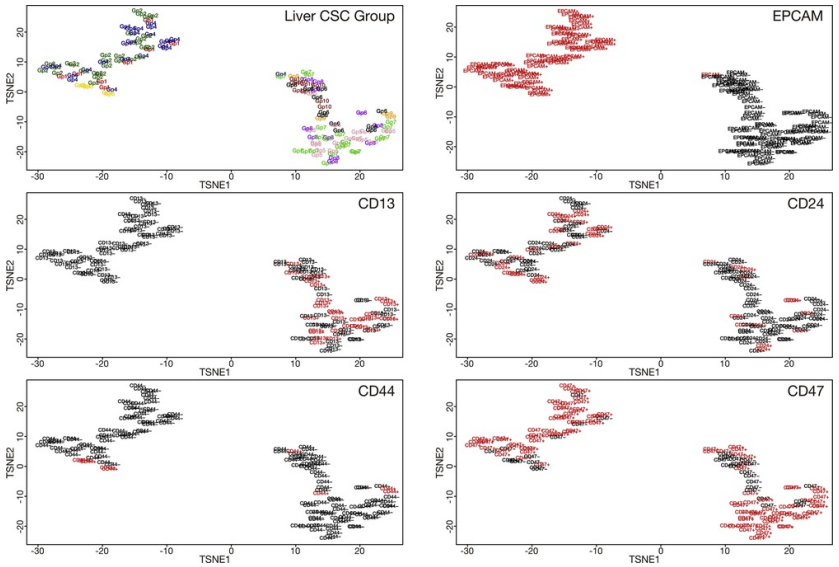
Supplementary Fig. 8 t-SNE analysis of global gene profiles of HCC single cells, correlated with liver CSC information. t-SNE plot based on global protein-coding gene profiles of HCC single cells was correlated with the status of liver CSC group (10 groups as same as in Fig. 1) and single-marker liver CSC expression enrichment (enrichment-positive cells marked as red). Middle cell clusters consisted of mainly liver CSC group 1–4 cells and were EPCAM expression-enriched.

alt-text: Supplementary Fig. 8



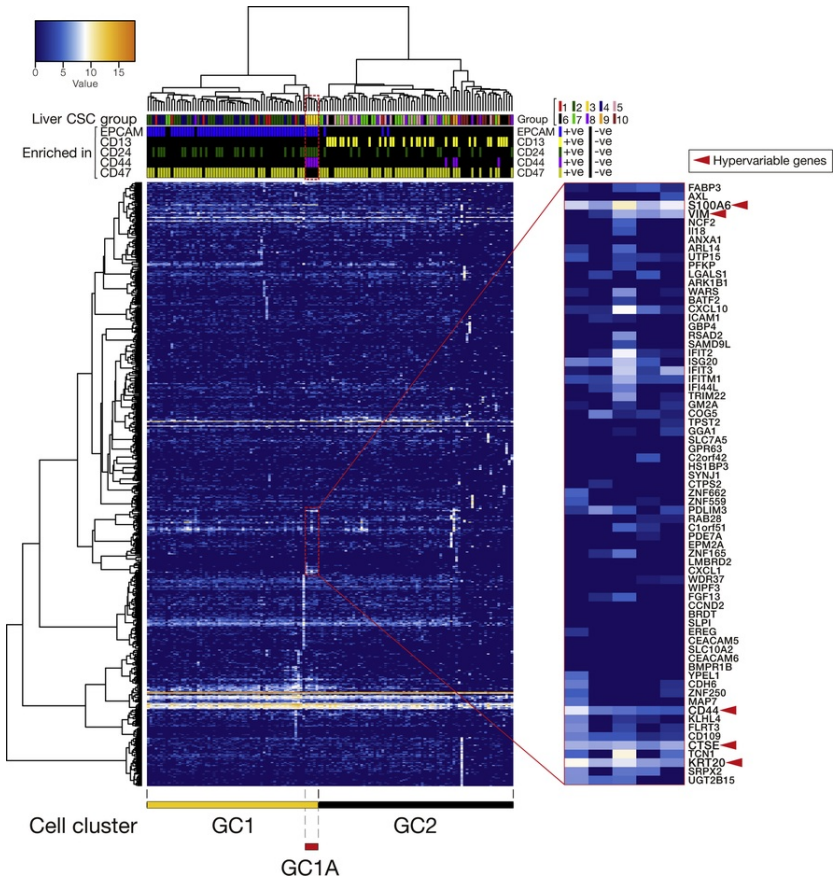
Supplementary Fig. 9 PCA analysis of hypervariable gene profiles of HCC single cells, correlated with liver CSC information. PCA plot based on hypervariable gene profiles of HCC single cells was correlated with the status of liver CSC group (10 groups as same as in Fig. 1) and single-marker liver CSC expression enrichment (enrichment-positive cells marked as red). Left cell clusters consisted of mainly liver CSC group 1–4 cells and were EPCAM expression-enriched.

alt-text: Supplementary Fig. 9



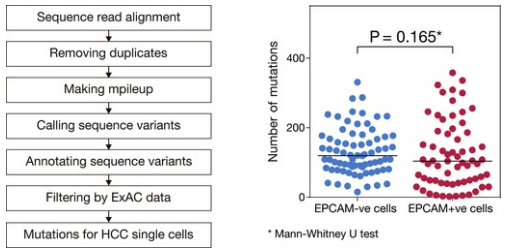
Supplementary Fig. 10 t-SNE analysis of hypervariable gene profiles of HCC single cells, correlated with liver CSC information. t-SNE plot based on hypervariable gene profiles of HCC single cells was correlated with the status of liver CSC group (10 groups as same as in Fig. 1) and single-marker liver CSC expression enrichment (enrichment-positive cells marked as red). Left cell clusters consisted of mainly liver CSC group 1–4 cells and were EPCAM expression-enriched.

alt-text: Supplementary Fig. 10



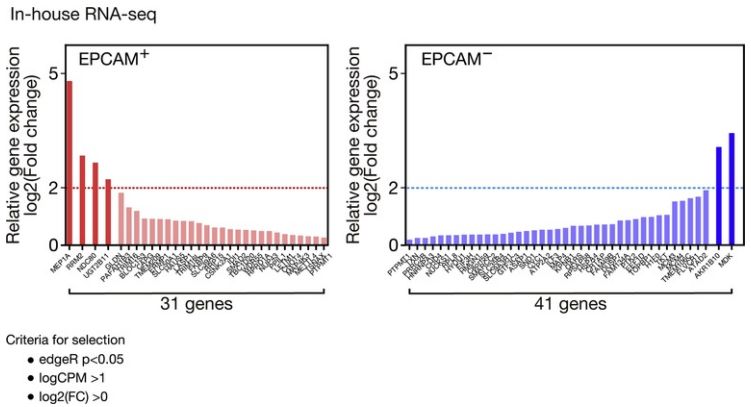
Supplementary Fig. 11 Unsupervised h-clustering of high Gini gene profiles of HCC single cells, correlated with liver CSC information. There was a minor cell cluster GC1A within GC1, having enriched *EPCAM, CD24* and *CD44* expressions (upper panel, dashed red line box).

alt-text: Supplementary Fig. 11



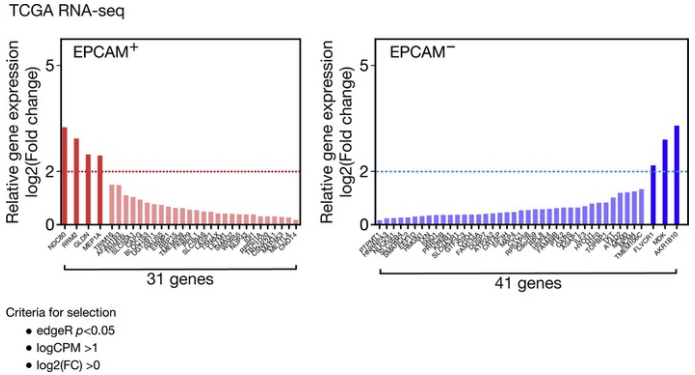
Supplementary Fig. 12 Mutation calling on EPCAM^{+} and EPCAM^{-} HCC single cells. EPCAM^{+} and EPCAM^{-} cells were having similar number of mutations.

alt-text: Supplementary Fig. 12



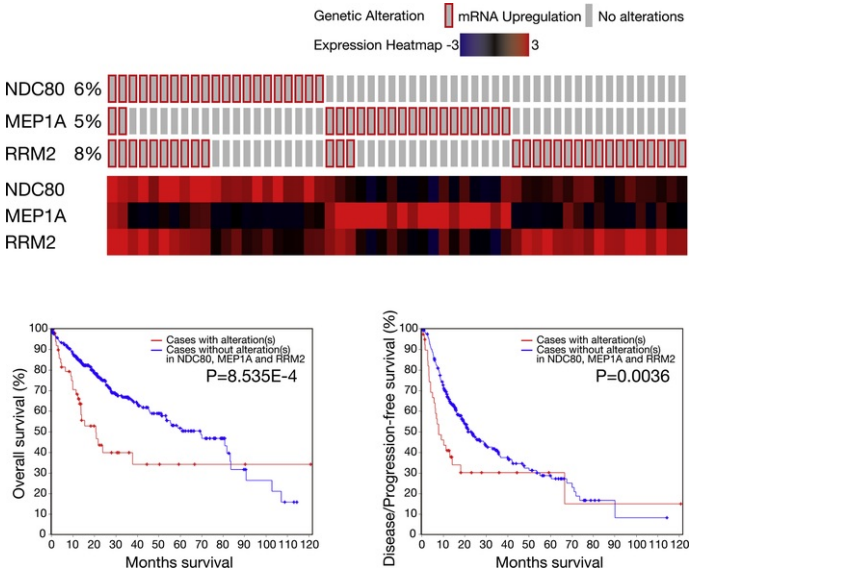
Supplementary Fig. 13 Candidate selection for EPCAM$^+$ and EPCAM$^-$ cells using in-house RNA-seq. Genes were shortlisted that were concurrently overexpressed in HCC and significantly upregulated genes in either EPCAM$^+$ and EPCAM$^-$ cells. There were 4 (MEP1A, RRM2, NDC80 & UGT2B11) and 2 (MDK & AKR1B10) genes identified in EPCAM$^+$ and EPCAM$^-$ cells respectively, that exceeded log2(fold change) threshold of 2.

alt-text: Supplementary Fig. 13



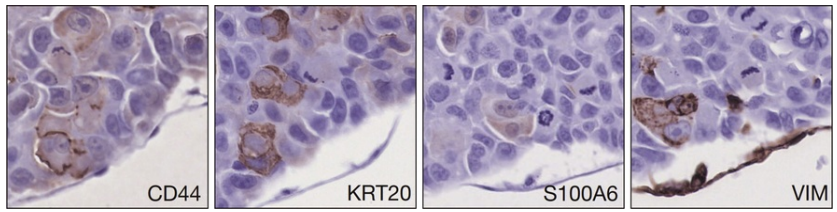
Supplementary Fig. 14 Candidate selection for EPCAM$^+$ and EPCAM$^-$ cells using TCGA RNA-seq. Genes were shortlisted that were concurrently overexpressed in HCC and significantly upregulated genes in either EPCAM$^+$ and EPCAM$^-$ cells. There were 4 (NDC80, RRM2, GLDN & MEP1A) and 3 (AKR1B10, MDK & FLVCR1) genes identified in EPCAM$^+$ and EPCAM$^-$ cells respectively, that exceeded log2(fold change) threshold of 2.

alt-text: Supplementary Fig. 14



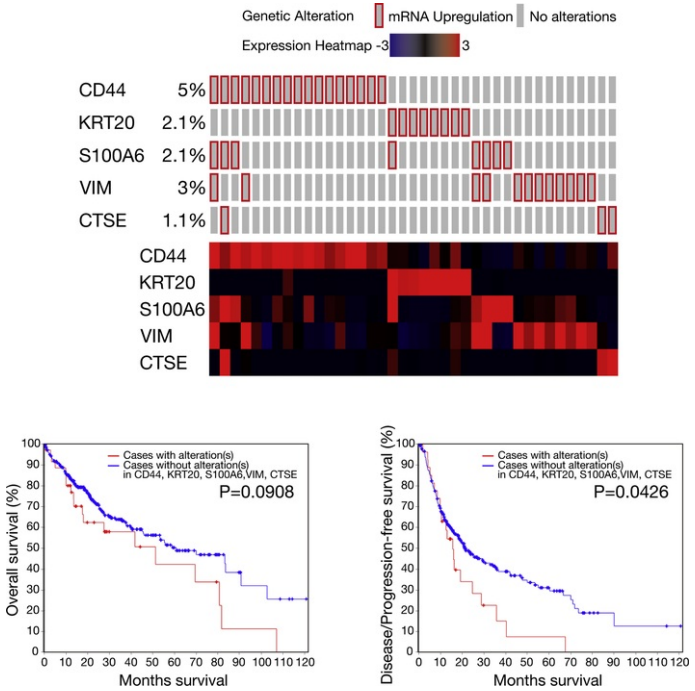
Supplementary Fig. 15 TCGA exploration on EPCAM^{+} signature genes. Cases with mRNA upregulation on signature genes (NDC80, MEP1A & RRM2) were compared to those without for overall ($P = 8.535E-4$) and disease-free survival ($P = 0.0036$).

alt-text: Supplementary Fig. 15



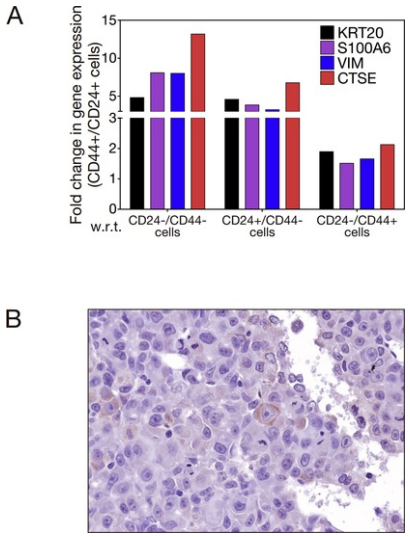
Supplementary Fig. 16 IHC suggested overlap staining for CD24^{+}/CD44^{+} signature genes in some HCC cells. IHC staining was performed using antibodies for different signature genes (CD44, KRT20, S100A6 and VIM) in consecutive paraffin sections of the PDTX.

alt-text: Supplementary Fig. 16



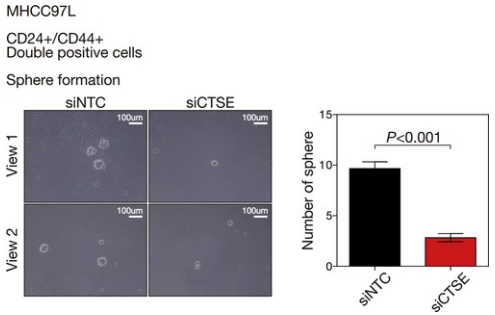
Supplementary Fig. 17 TCGA exploration on CD24^{+}/CD44^{+} signature genes. Cases with mRNA upregulation on signature genes (CD44, KRT20, S100A6, VIM & CTSE) were compared to those without for overall ($P = 0.0908$) and disease-free survival ($P = 0.0426$).

alt-text: Supplementary Fig. 17



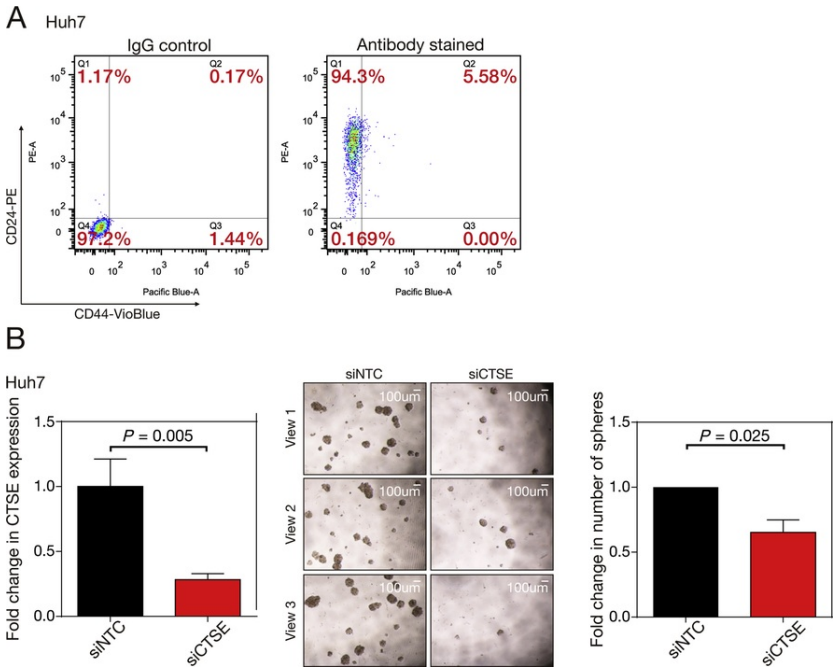
Supplementary Fig. 18 (A) Gene expression comparison in searching for key CD24^{+}/CD44^{+} downstream mediator. Fold changes in gene expression of gene candidates were displayed comparing CD24+/CD44 + cells, with respect to either the CD24-/CD44-, CD24+/CD44-, or CD24-/CD44 + counterparts. (B) IHC for CTSE showed mainly cytoplasmic localization in HCC cells.

alt-text: Supplementary Fig. 18



Supplementary Fig. 19 Stemness-related function of CD24$^+$/CD44$^+$ cell subpopulation mediated by *CTSE*. Transient siRNA knockdown of *CTSE* on CD24$^+$/CD44$^+$ cell subpopulation sorted from MHCC-97L HCC cells significantly reduced sphere formation ability, as compared to siNTC control.

alt-text: Supplementary Fig. 19



Supplementary Fig. 20 Confirmation of reduced sphere formation ability upon si-knockdown of *CTSE* in Huh7 HCC cells. (A) FACS confirmed the presence of CD24$^+$/CD44$^+$ cell subpopulation in Huh7 HCC cells, which was a relevant cell model for subsequent experiment. (B) Huh7 HCC cells displayed significant reduction in sphere formation ability upon siRNA knockdown on the signature gene *CTSE* specific to the enriched CD24$^+$/CD44$^+$ subpopulation.

alt-text: Supplementary Fig. 20

Multimedia Component 2

Supplementary Table 1 Demographic data of HCC patient.

alt-text: Supplementary Table 1

Highlights

- Single-cell transcriptomics dissected the intra-tumoral heterogeneity of hepatocellular carcinoma (HCC).
- HCC single cells showed two distinct major cell populations according to *EPCAM* expression.
- $CD24^+/CD44^+$-enriched cell subpopulation was identified within the $EPCAM^+$ cells.
- *CTSE* was the most upregulated signature gene in $CD24^+/CD44^+$-enriched cells.
- Knockdown of *CTSE* significantly reduced self-renewal ability *in vitro* and tumorigenicity *in vivo*.

Queries and Answers

Query: Please confirm that the provided email “iolng@hku.hk” is the correct address for official communication, else provide an alternate e-mail address to replace the existing one, because private e-mail addresses should not be used in articles as the address for communication.
Answer: Yes, this is correct.

Query: Have we correctly interpreted the following funding source(s) you cited in your article: Hong Kong Research Grants Council Theme-based Research Scheme?
Answer: Yes

Query: Please provide the volume number or issue number or page range or article number for the bibliography in Ref(s). [6].
Answer: *Hepatology* **68**, 2018, 127-140.

Query: Please confirm that given names and surnames have been identified correctly and are presented in the desired order and please carefully verify the spelling of all authors’ names.
Answer: Yes