Published in final edited form as:

Hepatology. 2018 July ; 68(1): 127–140. doi:10.1002/hep.29778.

Single cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma

Hongping Zheng^{1,*}, Yotsawat Pomyen^{1,5,*}, Maria Olga Hernandez¹, Caiyi Li², Ferenc Livak², Wei Tang¹, Hien Dang¹, Tim F. Greten³, Jeremy L. Davis³, Yongmei Zhao⁴, Monika Mehta⁴, Yelena Levin⁴, Jyoti Shetty⁴, Bao Tran⁴, Anuradha Budhu¹, and Xin Wei Wang^{1,**}

¹Laboratory of Human Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892 ²Flow Cytometry Core Facility, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892 ³Thoracic and Gastrointestinal Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892 ⁴Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland 21701 ⁵Translational Research Unit, Chulabhorn Research Institute, Bangkok 10210, Thailand

Abstract

Intratumor molecular heterogeneity of hepatocellular carcinoma (HCC) is partly attributed to the presence of hepatic cancer stem cells (CSCs). Different CSC populations defined by various cell surface markers may contain different oncogenic drivers, posing a challenge in defining molecular-targeted therapeutics. We combined transcriptomic and functional analyses of HCC cells at the single cell level to assess the degree of CSC heterogeneity. We provide evidence that hepatic CSCs at the single-cell level are phenotypically, functionally and transcriptionally heterogeneous. We found that different CSC subpopulations contain distinct molecular signatures. Interestingly, distinct genes within different CSC subpopulations are independently associated with HCC prognosis, suggesting that a diverse hepatic CSC transcriptome affects intratumor heterogeneity and tumor progression.

Conclusion—Our work provides unique perspectives into the previously unappreciated diversity of CSC subpopulations, whose molecular heterogeneity further highlights their role in tumor heterogeneity, prognosis and hepatic CSC therapy.

Keywords

intratumor heterogeneity; single cell genome; tumor cell community; cancer stem cells; cancer stem cell heterogeneity

Hepatocellular carcinoma (HCC) has had the fastest rising incidence and mortality among cancers in the United States over many decades. HCC is clinically, molecularly and

**Correspondence: Xin Wei Wang, National Cancer Institute, 37 Convent Drive, Building 37, Room 3044A, Bethesda, Maryland 20892; xw3u@nih.gov; Phone: 240-760-6858; Fax: 240-541-4496.

*These authors contributed equally to this work.

Potential Conflicting Interest: Nothing to report.

biologically heterogeneous, and highly resistant to treatment (1, 2). A major barrier to improving HCC patient outcome is the incomplete understanding of HCC heterogeneity and its impact on therapeutic intervention. Tumor heterogeneity consists of intertumor (tumor by tumor) and intratumor (within each tumor) heterogeneity. Strong evidence indicates the presence of intertumor heterogeneity, as several stable molecular subtypes are found in HCC (3–7). However, knowledge on intratumor heterogeneity, especially at the single cell level in HCC, is limited (8). The current view is that each primary tumor lesion consists of cells that may differ genetically, and epigenetically, which may result in phenotypic heterogeneity unique to each tumor type (9). Some of these tumor cells, referred to as cancer stem cells (CSCs), are thought to be responsible for generating a heterogeneous tumor lesion and contributing to treatment resistance, tumor relapse and metastasis. Thus, there is an urgent need to understand intratumor heterogeneity by characterizing tumor cell communities, cellular hierarchy and their diversity specific to a particular tumor type with unique clinical features and therapeutic responses.

Intratumor genomic heterogeneity has been recently documented in many tumor types including HCC (10, 11). Evidence linking intratumor genomic heterogeneity to cancer prognosis has also been noted (12). One hypothesis is that a tumor lesion is hierarchically organized, with each cell having different roles, which collectively ensure the survival of a given tumor cell community. However, current genomic analyses rely on bulk tissue with mixed tumor cells. Such methods are not robust enough to capture true the tumor evolution and cell communities. Currently, it is unclear how various heterogeneous tumor cells cooperate with each other and whether their collective behavior and regulation exist as an efficient community unique to each tumor subtype. Notably, tumor communities are poorly characterized.

The recent development of single-cell genome sequencing technologies has generated many new insights into complex biological systems, including human cancers (13). Single tumor cell analysis can provide the level of sensitivity and specificity to understand tumor biology regarding collective behavior and regulation of a given tumor cell community (14, 15). In this study, we hypothesized that CSC heterogeneity may contribute to a molecular and biological diversity of a HCC cell community and consequently, patient prognosis. Thus, we conducted a feasibility study by performing single-cell transcriptome analysis to characterize CSC heterogeneity in HCC.

Materials and Methods

Clinical Specimens

The Liver Cancer Institute (LCI) and Laboratory of Experimental Carcinogenesis (LEC) cohorts were previously described (4, 16). The study was approved by the Institutional Review Board of the LCI and the National Institutes of Health. For the Cancer Genome Atlas (TCGA) cohort, clinical and RNA-seq data related to 250 liver HCC samples with available survival data were collected from TCGA (<https://tcga-data.nci.nih.gov/tcga/>). A resected HCC sample was obtained with informed consent from a patient who had undergone resection at the NIH Clinical Center, and tissue acquisition procedures were approved by the Institutional Review Board of NIH.

Cell Lines

Human liver cancer cell lines (HuH1 and HuH7) were obtained from Health Science Research Resources Bank (JCRB0199 and JCRB0403, respectively). Cell line authentication was performed by the NCI Genome Core Laboratory using STR (short tandem repeats) analysis. For 2D and 3D culture, the cell lines were cultured with the specific media that has been established for each cell line as previously described (16).

Immunofluorescence Staining

For monolayer cell staining, cells growing on 12 mm coverslips were fixed in formaldehyde 4% (VWR) for 20 min at room temperature, washed, and blocked in IFF [1% bovine serum albumin, 2% FBS in PBS] followed by incubation with conjugated primary antibodies against EpCAM (FITC, #60136FI, Stemcell Technologies), CD133 (APC, #130-090-854, Miltenyi Biotec) and CD24 (PE, #555428, BD Pharmingen) for 2 hr at room temperature. For spheroid cells, single-cell suspensions of 1000 cells were seeded in 6-well Ultra-Low Attachment Microplates (Corning, Corning, NY) and cultured for 2 weeks. Collected spheres were seeded to 8-well chamber slides (Thermo Fisher Scientific, Waltham, MA) and fixed with 4% paraformaldehyde for downstream staining. Images were taken with a LSM 780 Confocal Microscope (Carl Zeiss, Jena, Germany).

Flow Cytometry and Sorting

Cell pellets were resuspended in staining buffer (PBS/ph 7.2, 0.5% bovine serum albumin and 2 mM EDTA) supplemented with penicillin and streptomycin. The cell suspension was further filtered through a .45 µm filter to remove cell aggregates. Antibody staining was performed on ice for 16 min. Stained cells were washed of excess unbound antibodies and resuspended in staining buffer. Flow analysis was done on a 5-laser BD LSRFortessa (Becton Dickinson) cell analyzer and single-cell sorting was done using a 5-laser BD FACSaria Fusion cell sorter (Becton Dickinson). Forward-scatter height versus forward-scatter width (FSC-H versus FSC-W) and side-scatter height versus side-scatter width (SSC-H versus SSC-W) were used to eliminate cell doublets and other aggregates and ensure single cell sorting. Dead cells were eliminated by excluding Sytox positive (SYTOX Blue dead cell stain, Thermo Fisher Scientific) cells, which increased the efficiency of sorting robust, live cells for single-cell experiments.

Single-cell Clonogenicity Assay and Flow analysis of Single-cell derived cell populations

Single cells were directly sorted based on cell surface marker status into 96-well plates with culture medium. After 2-hour incubation, the wells were checked under the microscope to confirm that there was one single-cell per well and the positive wells were recorded and counted. For the clonogenicity assay, single cells were grown for 2 weeks followed by counting of the number of clones and the cell number found within each single clone. For flow analysis, the clones were cultured for one month to obtain appropriate cell number for the analysis.

DEPArray procedure

Fresh primary HCC tissue was mechanically dissociated and digested in collagenase/dispase/DNaseI solution (2 mg/ml collagenase/dispase, 0.001% DNaseI) for 30 min at room temperature. Contaminated red blood cells were lysed with ammonium chloride solution (STEMCELL Technologies) according to the manufacturer's instructions. The resulting cell suspension was stained with Hoechst and antibodies against CD45 (PE, #555438, BD Biosciences), EpCAM (FITC, #60136FI, Stemcell Technologies), CD133 (APC, #130-090-854, Miltenyi Biotec) and CD24 (PerCP/Cy5.5, #311115, BioLegend). Single cells were isolated based on surface marker status using the DEPArray system (Silicon Biosystems) following the manufacturer's instructions. Leukocyte depletion for inflammatory infiltrate was performed based on CD45 staining.

Single-cell cDNA Synthesis, Library Construction, and Sequencing using SMART-seq platform

Cells were sorted directly into ice-cold lysis buffer. The SMART-seq protocol was performed on single or pooled sorted cells following the manufacturer's instructions with some modifications. Briefly, whole transcriptome was amplified with the SMART-Seq v4 Ultra Low Input RNA Kit (Clontech). Agencourt AMPure XP PCR purification kit (Beckman Coulter) was used for purification of amplified cDNA. Full length cDNA libraries were barcoded using the Nextera XT DNA Library Preparation Kit (Illumina). Libraries were pooled and sequenced on one lane of HiSeq2500 with Illumina TruSeq V4 chemistry (126 bp paired-end reads) at the Frederick National Laboratory for Cancer Research Sequencing Facility in Frederick, MD. In total, 129 single cells from HuH1 and HuH7 cell lines, and 34 single cells from a HCC biopsy sample were processed; 118 single cells passed the quality filter and were successfully sequenced, including 43 HuH1 cells, 55 HuH7 cells and 20 primary HCC cells.

Data processing of Single-cell RNA-seq from SMART-seq platform

Read alignment to the human reference genome GRCH38 and gencode version 24 was performed using STAR version 2.5.1 (17). RNA quantification was performed by using RSEM version 1.2.22 (18). Average mapped reads for the two cell lines are around 6 million reads, while average mapped reads for single cells from the patient biopsy sample are around 4 million reads. All downstream analyses were carried out using R version 3.3.3. Raw data from human embryonic stem cells from GSE69471 (19), mouse endothelial cells (ECs) and adult mouse hematopoietic stem cells (HSCs) stages T1 and T2 pre-HSCs from E11 AGM region and mature HSCs from E12 and E14 fetal liver from GSE66954 (20) were used as reference samples in principal component analysis (PCA). All data were median normalized before further analyses. To combine mouse and human gene expression, genes from mouse and human genomes were mapped to the gene list from human and mouse gene co-expression study (21). Human and mouse genes with the same gene symbols (11,493) were retained for further analyses. We also investigated the distribution of non-zero data points (transcripts that are "detectable" on the platforms) across all datasets and observed that each dataset shows a normal distribution independently within similar ranges as well as together, suggesting that while they are from different platforms, the data show concordant

variations (See Supplemental Figure S4A–B). We did not perform further data transformation due to the relatively small number of sequencing reads, the sparsity of data, and cell population heterogeneity present in single cell sequencing data, as data transformation would significantly decrease the dimension of the data, decrease intrinsic biological variability (especially in low abundant expressed genes) and cause less stable estimation (precise measurement) (22). To generate PCA, we clustered our single cell data with other groups, including hESC, mEC, and mHSC without transformation.

Single-Cell RNA-seq using GemCode platform

The 10XGenomics platform was used to generate gene expression data from single cells from HCC cell lines and the HCC biopsy sample. Briefly, frozen cells were thawed at 37°C, and washed with PBS+0.04% BSA and approximately 5000 cells were loaded onto the 10XGenomics Chromium Controller machine for Gel Beads-in-Emulsion (GEM) generation. During this step, cells were partitioned into the GEMs along with Gel Beads coated with oligos. These oligos provide poly-dT sequences to capture mRNAs released after cell lysis inside the droplets, as well as cell-specific and transcript-specific barcodes (14 bp 10X Barcode and 10 bp Randomer, respectively). mRNA from cell lines and the HCC biopsy sample were prepared using 10x Genomics Chromium Single Cell 3' reagent kit (V1 chemistry). Following RT, cDNA containing both barcodes were recovered, purified and amplified to generate sufficient quantities for library preparation. All three samples were pooled and sequenced in one run NextSeq500 with NextSeq V2 chemistry. The sequencing run was setup as asymmetric dual index run: read 1 had 98 bps for transcript read, index1 had 14bps for cell barcode index read, index 2 had 8bps for sample index read, and read2 had 10bps for Unique Molecular Identifier (UMI). Demultiplex was done allowing 1 mismatch in the barcodes.

Data processing of single-cell RNA-seq from GemCode platform

Sample demultiplex, read alignment and unique molecular identifier (UMI) counting and Graph-based clustering analysis with t-distributed stochastic neighbor embedding (tSNE) calculation were performed by using CellRanger version 1.3.1 (<https://support.10xgenomics.com>). All data processing was done on NIH HPC Biowulf cluster (<http://hpc.nih.gov>). All other downstream analyses were carried out using R version 3.3.3. In total, 3,847 cells were sequenced from the three cell types, which are not FACS-sorted cells. For HuH1, HuH7 and HCC patient cells, 899, 2088 and 860 cells were sequenced, respectively. After normalization, there are 75,826 mean reads per cell, with 2,703 median genes detected per cell. Finally, 3,649 cells with more than 1,660 UMI counts were retained for further analyses. tSNE analysis and graph-based clustering were performed using the first 10 principal components for projection. Normalized and z-scored expression of selected HCC-specific CSC marker genes were overlaid onto tSNE projection.

Additional methodologies can be found in supplemental materials.

Results

Phenotypic and biological heterogeneity of liver CSC

Compelling experimental data indicate the presence of HCC CSCs defined by various cell surface markers, e.g., CD13, CD24, CD44, CD90, CD133 and EpCAM, as the source of tumor initiating cells (23). These cells are capable of self-renewal, differentiation and production of heterogeneous tumors. Moreover, the discrete nature of CSCs, defined by CD90 and EpCAM, was noted on bulk tissue (24). To better define CSC heterogeneity, we performed single-cell analysis. First, we characterized HuH1 and HuH7, two independent and well-defined HCC cell lines, for the presence of CSC properties and heterogeneity (16, 25). We first characterized commercially available antibodies specific to CD13, CD24, CD44, CD90, CD133 and EpCAM for immunofluorescence (IF) and fluorescence-activated cell sorting (FACS) analyses for their abilities to detect different CSC subpopulations. We selected antibodies specific to CD24, CD133 and EpCAM for further studies due to their superior performance in sensitivity and specificity (data not shown). To determine biological and molecular heterogeneity of CSCs, we first compared different subpopulations of CSCs by examining CSC marker expression in HCC cells cultured either in monolayers or organoids at the single cell level using confocal imaging analysis (Fig. 1). Within each cell type, we observed vast heterogeneity of CSC subpopulations in both monolayers (Figure 1A) and organoids (Fig. 1B). Interestingly, some cells express all three markers, but other CSC subpopulations show discrete clusters and seem to not be randomly distributed among a colony of cells.

Next, we isolated distinct CSC subpopulations using a single marker via FACS, plated them onto 96-well plates and assessed their self-renewal potential. A total of 1,343 single HuH1 cells and 1,134 single HuH7 cells from three independent single cell sorts were assessed. We found that approximately 15–20% of marker-positive CSC subpopulations of HuH1 cells and approximately 8–12% of marker-positive subpopulations of HuH7 cells could divide (Fig 2A–B). Notably, only certain CSC subpopulations seem to divide better when cultured under hypoxia, a condition known to promote stemness (26). In contrast, approximately 0–5% of marker negative cells (Triple–) could divide, consistent with previous findings (16, 25, 27). We next monitored cell numbers in each colony derived from single marker expressing cells for a culture period of two weeks under normoxia or hypoxia. We found that there was a vast difference in self-renewal capacities among different CSC subpopulations in both HuH1 and HuH7 cells (Fig 2C–D). We further assessed several successfully-expanded cell populations derived from individual marker-expressing cells by FACS analysis. Compared to unsorted cells, a single marker-expressing cell could expand to a population of mixed CSCs with varying degrees in both HuH1 and HuH7 (Fig 2E). Interestingly, surviving triple– once expanded, regained a population of mixed CSCs, suggesting that either a few of these cells may not have been truly marker-negative or they underwent trans-differentiation from marker-negative cells to marker-positive cells. These results indicate that distinct CSC subpopulations identified using single-cell surface markers all had a higher self-renewal capacity compared to marker-negative cells, but displayed appreciable biological differences in cell division and responses to hypoxia.

Transcriptome landscapes of individual liver CSCs and heterogeneity

To determine if the biological difference among individual marker-expressing CSCs could be linked to different molecular features, we performed single cell transcriptome analysis using single-cell SMART-Seq. We isolated marker-expressing HuH1 and HuH7 cells by FACS. We also isolated marker-expressing cells (CD133+/CD24+/EpCAM+/CD45-) from a surgically resected tumor (P1) by DEPArray. We obtained RNA-seq data with an average of six million uniquely mapped reads per cell from a total of 118 single cells and pools of 10 or 100 cells. As a control, we first examined levels of marker expression among sorted single cells. We found that marker-positive cells defined by FACS generally had much higher levels of corresponding transcripts than marker negative cells (Suppl Fig S1), revealing excellent sensitivity of the single cell RNA transcriptome methodology.

To determine the global transcriptomic diversity of CSCs, we included publicly available single-cell RNA-seq data from human embryonic stem cells (hESC), murine endothelial cells (mEC) and murine hematopoietic stem cells (mHSC) for comparison (19, 20). Using principal component analysis (PCA), we found that each cell type has a distinct transcriptome since single cells are clustered largely based on their types, without or with batch effect removal (Fig 3A and Suppl Fig 4C). In addition, each single cell differs in their transcriptome within a cell community regardless of whether they are cultured HCC cell lines or freshly isolated tumor cells. However, such a difference was diminished when a pool of 10 or 100 cells were analyzed (Fig 3A), indicating that transcriptomic heterogeneity exists at the single cell level but decreases significantly in cell pools. We observed a greater transcriptomic diversity in HuH7 cells than HuH1 cells based on their marker expression. Moreover, when single HuH7 cells were clustered according to their transcriptome, they can be identified based largely on marker expression (Fig 3B). Similar results were obtained in HuH1 and P1 but the difference was smaller (Suppl Fig S2). Together, these results indicate that the cell transcriptome is heterogeneous at the single cell level.

To determine whether the difference in the single-cell transcriptome reflects its biological nature or is a result of stochastic gene expression at the single cell level due to low copy transcripts measured by single-cell RNA-seq, we assessed whether potential CSC gene signatures associated with marker-expressing cells are linked to HCC prognosis. We reasoned that if transcriptomic changes in single cells associated with marker expression are functionally important, these changes should be associated with HCC survival, analogous to previous reports using tumor bulk (16, 25, 27). We first searched for genes that are associated with marker expression by performing class comparison between triple marker-positive HuH7 (Triple+) and marker-negative HuH7 cells (TripleNeg). We identified a total of 286 genes associated with Triple+ HuH7 cells (Suppl Table S1). Using the 286-gene signature, we performed multivariable Cox regression survival risk prediction analysis with 10-fold cross validation and 1,000 random permutation of sample labels. We found that the triple+ gene signature could predict overall survival in 240 HCC tissues from the LCI cohort (log-rank $p < 0.0001$) and the cross-validated misclassification rates are significantly lower than expected by chance (permutation $p < 0.01$) (Fig 3C). This relationship was not observed in non-tumor tissues from the same LCI cohort (log-rank $p > 0.05$; permutation $p > 0.1$) (Fig 3D), indicating that the triple+ marker signature is tumor specific. Similar results were

observed in the TCGA cohort of 250 HCC cases with available survival data (log-rank $p < 0.01$; permutation $p < 0.05$) (Fig 3E).

Class comparison analysis of distinct CSC subpopulations based on individual markers in HuH7 cells revealed that gene signatures from EpCAM+, CD133+, CD24+ or Triple+ largely do not overlap (Fig 4A, Suppl Table S1). Furthermore, network analysis based on Ingenuity Pathway Analysis (IPA) revealed that the top networks associated with different marker gene signatures also largely do not overlap, e.g., AKT signaling for CD133+ cells, NF κ B signaling for CD24+ cells, p38 MAPK signaling for EpCAM+ cells and NF κ B for Triple+ cells, all of which are known canonical signaling pathways critical for maintaining cancer stemness (Fig 4C–F, Suppl Table S2). Interestingly, the above marker signatures can predict overall survival of the LCI cohort, TCGA cohort and LEC cohort using multivariable survival risk prediction algorithm described above (Fig 4B). Univariable and multivariable Cox Regression analysis revealed that gene signatures linked to CD133 and EpCAM, but not CD24, are independent predictors of HCC survival (Suppl Table S3). Similar results were obtained in HuH1 cells with borderline significance (Suppl Fig S2C, E). These results suggest that, while molecular signaling differs among different marker expressing cells, marker expression signatures are associated with HCC prognosis. Taken together, the above results indicate that single-cell transcriptome analysis provides sufficient sensitivity to capture biologically-relevant molecular signaling linked to distinct functional composition of CSCs and that significant molecular heterogeneity exists among individual tumor cells within a tumor cell community or among different tumor types.

Transcriptomic landscapes of HCC cell community, heterogeneity and HCC prognosis

We sought to determine global cell population diversity in HuH1, HuH7 and P1 by single-cell RNA-seq using the GemCode technology (10X Genomics), a droplet-based system enabling high cell throughput with ~50% cell capture efficiency (28). We generated transcriptome data from 3,847 single cells from HuH1, HuH7 and P1, which were visualized by tSNE projection (Fig 5). Consistent with transcriptome data from single cells with defined marker status, single cells are largely clustered based on their distinct cell types (Fig 5A). Furthermore, tSNE and hierarchical clustering analyses revealed that HuH1 had 2 main branches of cells and HuH7 had 4 main branches of cells, while P1 had 3 distinct branches of cells (Fig 5A, Suppl Fig S3A–B). It is noted that a cell population diversity in HuH1, HuH7 and P1 was similar between marker-expressing cells analyzed by single-cell SMART-seq and unsorted populations analyzed by 10X Genomics (Fig 3A, Fig 5A and Suppl Fig S2A), suggesting that single-cell transcriptomic diversity is an intrinsic nature of each cell type. Interestingly, expression patterns of known CSC markers, such as EpCAM, CD133 (PROM1), CK-19 (KRT19) and aldehyde dehydrogenase (ALDH1A1), varied greatly among individual cells (Fig 5B). Similarly, the expression patterns of several known HCC-specific genes, i.e., AFP (29), MDK (30), GPC3 (31), GOLM1 (32), YY1AP1 (33), PLK1, ECT2 and NELFE (7, 34), whose functions are associated with different aspects of hepatocarcinogenesis, were also heterogeneous within each cell community and among different cell types (Fig 5C). No discrete cell cluster with the above markers was noted. Since P1 gave rise to three distinct branches of cells, we examined lineage specific markers, i.e., Yamanaka factors (KLF4, POU5F1, MYC), stem cell marker (NANOG), leukocyte

marker (PTPRC), T-cell specific markers (CD3E, CD8A) and B-cell marker (CD79A) (Fig 5D) among these cells. We found that only a few cells express Yamanaka factors, except MYC, which is expressed relatively more abundantly in HuH1 and HuH7 than P1 with notable heterogeneity. One branch of P1 cells expressed T-cell markers, while another branch expressed a B-cell marker, indicating that these are infiltrating leukocytes in HCC specimens. The 3rd branch of P1 cells expressed various tumor markers, suggesting that these are tumor cells (Fig 5, Suppl Fig S3A–B).

CSC heterogeneity may be responsible for intratumor heterogeneity within each tumor cell population and may contribute to HCC prognosis. To test this hypothesis, we first searched for genes that are correlated with all known HCC CSC genes at the single-cell level. We selected ANPEP (35), CD44 (36), DLK1 (37), KRT19 (38), EPCAM (16), PROM1 (25), ICAM1 (39), CD47 (40), LGR5 (41) and SOX9 (42) whose functions have been demonstrated to be necessary for maintaining HCC stemness. Global correlation analysis revealed 45 genes for HuH1 and 1066 genes for HuH7 as CSC heterogeneity-surrogate gene signatures, referred as H1-HET and H7-HET, respectively. This observation is consistent with the above findings (Figure 2A), which suggests that HuH7 cells are more heterogeneous compared to HuH1 cells. We then tested whether H1-HET and H7-HET are associated with survival in the LCI, LEC and TCGA cohorts, a strategy analogous to the analysis of marker signatures described above. We reasoned that these surrogate-gene signatures should reflect CSC activities and thus predict HCC survival. Consistently, cross-validated survival risk prediction analyses revealed that both H1-HET and H7-HET could predict overall survival in three independent cohorts, i.e., LCI, LEC, TCGA cohorts (Suppl Fig S3C). The cross-validated misclassification rates were significantly lower than expected by chance (permutation $p < 0.05$) for the LEC cohort and had a borderline significance in the other two cohorts, suggesting that the CSC heterogeneity surrogate-gene signatures derived from HuH1 and HuH7 cells are associated with HCC prognosis. Again, a prognostic association of CSC surrogate gene signatures was not found in non-tumor tissues from LCI cohort (data not shown), indicating that these changes are linked to tumor cell activities.

Discussion

Intratumor genomic heterogeneity in various solid tumors has been well documented in recent years. However, it is unclear how such heterogeneity contributes to therapeutic failure and cancer progression. This is the first report to demonstrate the presence of biological and transcriptomic heterogeneity of CSCs at the single-cell level in HCC. We found evidence of CSC heterogeneity within each distinct CSC cell community, which was identified by different CSC markers. Interestingly, transcriptomic heterogeneity is diminished when a pool of cells was analyzed, suggesting that single-cell analysis is required to fully appreciate the heterogeneous nature of each tumor. Our findings are analogous to the results from other human malignancies, which demonstrate that tumors are transcriptomically heterogeneous at the single-cell level (43–46). Moreover, we show that CSC-marker expression-associated genes are linked to HCC prognosis, suggesting that the diverse transcriptome of a single cell may reflect its tumor biology. These results indicate that a meaningful understanding of intratumor heterogeneity, which reflects the true biological nature of the tumor should be

carried out at the single-cell level, rather than tumor bulk. Thus, our study may provide a feasible approach to study the impact of intratumor heterogeneity in HCC development.

Here, using a combined staining of three established CSC markers, i.e., CD24, CD133, EpCAM, we show that hepatic CSCs are spatially heterogeneous under different culture conditions. Accordingly, our data demonstrates that marker-expressing HCC cells are not randomly distributed, but further efforts are needed to investigate whether certain patterns of hepatic CSC distribution or organization exist and are linked to tumor biology. At a single-cell resolution, we found that different surface marker-defined CSCs are functionally heterogeneous in terms of self-renewal capacity and differentiation potential. Phenotypic heterogeneity of CSCs has been suggested (47) and recently, functional heterogeneity of skin and hematopoietic stem cells has been described (48). It is of interest to note that HCC CSCs exhibited an altered pattern of self-renewal heterogeneity when cultured under normoxia or hypoxia, suggesting a biological plasticity of these cells. One could speculate that such an intrinsic plasticity of HCC cells is an important feature for tumor progression, especially when facing different tumor microenvironments. Understanding the underlying mechanism that drives CSC plasticity may help us to understand how such events impact tumor microenvironmental heterogeneity and therapeutic response (49).

A limitation of this study is that the findings are largely based on two HCC cell lines and only one clinical specimen from a HCC patient. Thus, interpretation of CSC heterogeneity may not be generalized without data from additional clinical specimens. In principle, with single-cell resolution, coupled with cost-effective UMI-based approaches to increase throughput, one can classify patients into groups based on the diversity of CSC marker gene expression and predict the response of patients to certain chemo- or targeted therapies. Encouragingly, we have now established the logistic pipeline that allows us to acquire high quality single cells derived from HCC patients who have been enrolled in clinical trials at NCI for single-cell transcriptome analysis. Further studies will be focused on incorporating more patient samples to demonstrate the clinical values of this concept. We have now completed single cell transcriptome analysis in 20 additional HCC patients using the v2 version of 10x genomics to determine the degree of molecular distinctiveness among human HCC samples and their relationship with patient outcomes, which is the subject of a new manuscript. Another limitation is that there is no optimization of reads mapped to our samples before sequencing. As a proof-of-concept study, we utilized all sequencing data available on limited numbers of cells. Future studies will have to balance the cost of sequencing and the number of samples included for analysis. A careful calculation of average reads available for samples should be considered.

In summary, this study highlights the molecular composition and heterogeneity of CSCs in HCC, which may be used for CSC targeting strategies. Furthermore, evaluation of CSC heterogeneity and its prognostic value based on single-cell transcriptome data provide new insights into intratumor heterogeneity, tumor progression and their implications in the clinical setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Wang laboratory for critical discussions.

Financial support: This work was supported by grants (Z01 BC 010313 and Z01 BC 010877) from the Intramural Research Program of the Center for Cancer Research of the National Cancer Institute.

Abbreviations

HCC	Hepatocellular Carcinoma
CSC	Cancer Stem Cell
LCI	Liver Cancer Institute
LEC	Laboratory of Experimental Carcinogenesis
TCGA	The Cancer Genome Atlas
EC	Endothelial Cell
HSC	Haematopoietic Stem Cell
GEM	Gel Beads-in-Emulsion
UMI	Unique Molecular Identifier
IF	Immunofluorescence
FACS	Fluorescence-Activated Cell Sorting
hESC	human Embryonic Stem Cell
mEC	murine Endothelial Cell
PCA	principal component analysis
Triple+	CD133+/CD24+/EpCAM+
Triple–	CD133–/CD24–/EpCAM–
CD133+	CD133+/CD24–/EpCAM–
CD24+	CD133–/CD24+/EpCAM–
EpCAM+	CD133–/CD24–/EpCAM+
IPA	Ingenuity Pathway Analysis
tSNE	t-distributed Stochastic Neighbor Embedding
SR	Surface Rendering

References

1. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, Mariotto A, et al. Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival. *J Natl Cancer Inst.* 2017;109.
2. Worns MA, Galle PR. HCC therapies--lessons learned. *Nat Rev Gastroenterol Hepatol.* 2014; 11:447–452. [PubMed: 24492278]
3. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, Simon R, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med.* 2003; 9:416–423. [PubMed: 12640447]
4. Lee JS, Heo J, Libbrecht L, Chu IS, Kaposi-Novak P, Calvisi DF, Mikaelian A, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med.* 2006; 12:410–416. [PubMed: 16532004]
5. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* 2009; 69:7385–7392. [PubMed: 19723656]
6. The Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell.* 2017; 169:1327–1341. e1323. [PubMed: 28622513]
7. Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, Forgues M, et al. Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. *Cancer Cell.* 2017; 32:57–70. e53. [PubMed: 28648284]
8. Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, Wu X, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 2016; 26:304–319. [PubMed: 26902283]
9. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer.* 2012; 12:323–334. [PubMed: 22513401]
10. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012; 366:883–892. [PubMed: 22397650]
11. Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, Chen K, et al. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci U S A.* 2015; 112:E6496–6505. [PubMed: 26561581]
12. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med.* 2017; 376:2109–2121. [PubMed: 28445112]
13. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet.* 2016
14. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet.* 2015; 16:133–145. [PubMed: 25628217]
15. Chattopadhyay PK, Gierahn TM, Roederer M, Love JC. Single-cell technologies for monitoring immune systems. *Nat Immunol.* 2014; 15:128–135. [PubMed: 24448570]
16. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, Jia H, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. *Gastroenterology.* 2009; 136:1012–1024. [PubMed: 19150350]
17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29:15–21. [PubMed: 23104886]
18. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011; 12:323. [PubMed: 21816040]
19. Mertes F, Lichtner B, Kuhl H, Blattner M, Otte J, Wruck W, Timmermann B, et al. Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells. *BMC Genomics.* 2015; 16:925. [PubMed: 26564201]
20. Zhou F, Li X, Wang W, Zhu P, Zhou J, He W, Ding M, et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature.* 2016; 533:487–492. [PubMed: 27225119]

21. Monaco G, van Dam S, Casal Novo Ribeiro JL, Larbi A, de Magalhaes JP. A comparison of human and mouse gene co-expression networks reveals conservation and divergence at the tissue, pathway and disease levels. *BMC Evol Biol.* 2015; 15:259. [PubMed: 26589719]
22. Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat Methods.* 2017; 14:565–571. [PubMed: 28504683]
23. Yamashita T, Wang XW. Cancer stem cells in the development of liver cancer. *J Clin Invest.* 2013; 123:1911–1918. [PubMed: 23635789]
24. Yamashita T, Honda M, Nakamoto Y, Baba M, Nio K, Hara Y, Zeng SS, et al. Discrete nature of EpCAM+ and CD90+ cancer stem cells in human hepatocellular carcinoma. *Hepatology.* 2013; 57:1484–1497. [PubMed: 23174907]
25. Ma S, Chan KW, Hu L, Lee TK, Wo JY, Ng IO, Zheng BJ, et al. Identification and characterization of tumorigenic liver cancer stem/progenitor cells. *Gastroenterology.* 2007; 132:2542–2556. [PubMed: 17570225]
26. Ezashi T, Das P, Roberts RM. Low O₂ tensions and the prevention of differentiation of hES cells. *Proc Natl Acad Sci U S A.* 2005; 102:4783–4788. [PubMed: 15772165]
27. Lee TK, Castilho A, Cheung VC, Tang KH, Ma S, Ng IO. CD24(+) liver tumor-initiating cells drive self-renewal and tumor initiation through STAT3-mediated NANOG regulation. *Cell Stem Cell.* 2011; 9:50–63. [PubMed: 21726833]
28. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017; 8:14049. [PubMed: 28091601]
29. Johnson, PJ. Role of alpha-fetoprotein in the diagnosis and management of hepatocellular carcinoma. *Acta Med Okayama.* 1999; 14(Suppl):S32–S36.
30. Jia HL, Ye QH, Qin LX, Budhu A, Forgues M, Chen Y, Liu YK, et al. Gene expression profiling reveals potential biomarkers of human hepatocellular carcinoma. *Clin Cancer Res.* 2007; 13:1133–1139. [PubMed: 17317821]
31. Capurro M, Wanless IR, Sherman M, Deboer G, Shi W, Miyoshi E, Filmus J. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology.* 2003; 125:89–97. [PubMed: 12851874]
32. Ye QH, Zhu WW, Zhang JB, Qin Y, Lu M, Lin GL, Guo L, et al. GOLM1 Modulates EGFR/RTK Cell-Surface Recycling to Drive Hepatocellular Carcinoma Metastasis. *Cancer Cell.* 2016; 30:444–458. [PubMed: 27569582]
33. Zhao X, Parpart S, Takai A, Roessler S, Budhu A, Yu Z, Blank M, et al. Integrative genomics identifies YY1AP1 as an oncogenic driver in EpCAM(+) AFP(+) hepatocellular carcinoma. *Oncogene.* 2015; 34:5095–5104. [PubMed: 25597408]
34. Dang H, Takai A, Forgues M, Pomyen Y, Mou H, Xue W, Ray D, et al. Oncogenic Activation of the RNA Binding Protein NELFE and MYC Signaling in Hepatocellular Carcinoma. *Cancer Cell.* 2017; 32:101–114. e108. [PubMed: 28697339]
35. Haraguchi N, Ishii H, Mimori K, Tanaka F, Ohkuma M, Kim HM, Akita H, et al. CD13 is a therapeutic target in human liver cancer stem cells. *J Clin Invest.* 2010; 120:3326–3339. [PubMed: 20697159]
36. Zhu Z, Hao X, Yan M, Yao M, Ge C, Gu J, Li J. Cancer stem/progenitor cells are highly enriched in CD133+CD44+ population in hepatocellular carcinoma. *Int J Cancer.* 2010; 126:2067–2078. [PubMed: 19711346]
37. Xu X, Liu RF, Zhang X, Huang LY, Chen F, Fei QL, Han ZG. DLK1 as a potential target against cancer stem/progenitor cells of hepatocellular carcinoma. *Mol Cancer Ther.* 2012; 11:629–638. [PubMed: 22238367]
38. Kawai T, Yasuchika K, Ishii T, Katayama H, Yoshitoshi EY, Ogiso S, Kita S, et al. Keratin 19, a Cancer Stem Cell Marker in Human Hepatocellular Carcinoma. *Clin Cancer Res.* 2015; 21:3081–3091. [PubMed: 25820415]
39. Liu S, Li N, Yu X, Xiao X, Cheng K, Hu J, Wang J, et al. Expression of intercellular adhesion molecule 1 by hepatocellular carcinoma stem cells and circulating tumor cells. *Gastroenterology.* 2013; 144:1031–1041. e1010. [PubMed: 23376424]

40. Lee TK, Cheung VC, Lu P, Lau EY, Ma S, Tang KH, Tong M, et al. Blockade of CD47-mediated cathepsin S/protease-activated receptor 2 signaling provides a therapeutic target for hepatocellular carcinoma. *Hepatology*. 2014; 60:179–191. [PubMed: 24523067]
41. Lei ZJ, Wang J, Xiao HL, Guo Y, Wang T, Li Q, Liu L, et al. Lysine-specific demethylase 1 promotes the stemness and chemoresistance of Lgr5+ liver cancer initiating cells by suppressing negative regulators of beta-catenin signaling. *Oncogene*. 2015; 34:3214. [PubMed: 26062983]
42. Kawai T, Yasuchika K, Ishii T, Miyauchi Y, Kojima H, Yamaoka R, Katayama H, et al. SOX9 is a novel cancer stem cell marker surrogated by osteopontin in human hepatocellular carcinoma. *Sci Rep*. 2016; 6:30489. [PubMed: 27457505]
43. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016; 352:189–196. [PubMed: 27124452]
44. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJJ, Kong SL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet*. 2017; 49:708–718. [PubMed: 28319088]
45. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014; 344:1396–1401. [PubMed: 24925914]
46. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014; 512:155–160. [PubMed: 25079324]
47. Tang DG. Understanding cancer stem cell heterogeneity and plasticity. *Cell Res*. 2012; 22:457–472. [PubMed: 22357481]
48. Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, Sanchez Castillo M, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*. 2015; 16:712–724. [PubMed: 26004780]
49. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013; 501:346–354. [PubMed: 24048067]

Figure 1. Heterogeneous expression patterns of hepatic CSC surface markers CD133, CD24 and EpCAM, in spheroid and monolayer HCC cells shown by immunofluorescence

(A) HuH1 and HuH7 cells were grown as monolayers and immunostaining was performed for CD133, CD24 and EpCAM as indicated. Nuclear DNA was counterstained by DAPI. Shown are representative confocal images of each surface marker, a merged image and the corresponding image of surface rendering (SR). Scale bars are 50 μm. (B) HuH1 and HuH7 cells were grown as spheroids and immunostaining was performed for CD133, CD24 and EpCAM as indicated. Nuclear DNA was counterstained by DAPI. Shown are representative confocal images of each surface marker, a merged image and the corresponding image of SR. Scale bars are 50 μm.

Figure 2. Single-cell functional analysis reveals distinct self-renewal and differentiation capacities of HCC cells defined by hepatic CSC surface markers

(A and B) Barplots showing the percent of single cells with >2 cell divisions during 2-week culture after flow sorting. Single cells from HuH1 (A) or HuH7 (B) were stained and cells were sorted based on surface marker (CD133, CD24 and EpCAM) status, and further cultured individually in normoxia or hypoxia conditions for 2 weeks. Experiments were performed in triplicate and data are shown as mean $\pm$ SEM. (C and D) Violin plots showing the cell number (log2) of single cell-derived clones. The single cells from HuH1 (C) and HuH7 (D) were sorted and cultured as in (A and B), and the surface marker status is shown as indicated. (E) Barplots showing the percentage of CSC surface marker positive cells (CD133+, EpCAM+ and CD24+; bottom) within the single-cell derived cell populations. The single cells from HuH1 and HuH7 were defined and sorted as in (A and B) and cultured for one month. The derived cell population was analyzed by flow cytometry. The surface marker status of the original single cells is shown on the top panel as CD133+, EpCAM+

CD24+, Triple+ and Triple-. “n” indicates the number of the original cultured single cells. Unsorted cells were used as a control and in this instance, “n” indicates replicate analyses.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Single-Cell transcriptomic analysis reveals heterogeneity of marker-defined CSCs

(A) Multidimensional scaling analysis illustrates the relative similarity between all 118 single HCC cells, cell pools (10cell [n=8], 100cells [n=12]) and population controls. The small-size solid circles represent single cells, the medium-size shaded circles with a black outline represent 10 cells while the large-size open circles represent 100 cells. The distance between any two cells reflects the similarity of their transcriptomic profiles. Cells group mainly by their origin (color code) as indicated, i.e., HuH1, HuH7 cells, one primary HCC fresh tissue (P1 [n=20]), mouse hematopoietic stem cell (mHSC [n=7]), mouse embryonic cell (mEC [n=7]) and human embryonic stem cell (hESC [n=2]). All data were median normalized. (B) Multidimensional scaling illustrates the relative similarity between all 55 single cells from the HuH7 cell line with defined surface marker status. Cell cluster by marker expression status as indicated, but each subpopulation also contains outliers that are more similar to cells in other subpopulations. (C, D and E) Survival Risk Predictions of 240 tumor cases (C) and 231 non-tumor cases (D) in the LCI cohort, and 250 cases of tumor in TCGA with available survival data were performed using the Triple+ vs Triple- differential expressed gene set from HuH7 cells. The corresponding Kaplan-Meier survival curves are shown for high-risk and low-risk survival groups with the log-rank p value and the permuted p value.

(A) A Venn diagram is shown of genes with 5-fold alteration in the comparison of surface marker positive (Triple+, CD133+, EpCAM+ or CD24+) and Triple- HuH7 cells. (B) A Forest plot shows the hazard ratios and 95% confidence intervals (CI) for overall survival in various HCC cohorts (LCI, TCGA and LEC). The gene sets used for analyses are indicated in (A). (C-F) The top networks associated with CD133+, CD24+, EpCAM+ and triple+ HuH7 cells defined by single-cell transcriptome and IPA analyses are shown. Genes with shaded shapes are included in the signatures and are highlighted in colors, i.e., red (up-regulated), green (down-regulated). Open shapes represent genes that are not on the list of significant genes but are reported to be associated with the network. Arrows represent

positive regulation of gene expression, with solid arrows indicating direct regulation and broken arrows indirect regulation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5. Transcriptomic patterns of HuH1, HuH7 and P1 based on 10X genomics single cell analysis

(A) 2D visualization of (10X Genomics) data from 899 HuH1 cells, 2088 HuH7 cells and 860 freshly isolated primary HCC cells by t-SNE. (B) t-SNE plots showing expression and distribution of hepatic CSC related genes. (C) HCC associated genes. (D) stemness-associated genes or immune cell genes. For panels B–D, each cell was colored based on their normalized expression of indicated genes, i.e., a gradient of gray, yellow, and red indicating low to high expression.