Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Deciphering Tumour Microenvironment of Liver Cancer through Deconvolution of Bulk RNA-seq Data with Single-cell Atlas

Shaoshi Zhang [1,2], Wendi Bacon [3], Maikel P. Peppelenbosch [2], Folkert van Kemenade [3] and Andrew P. Stubbs [1,*]

[1] Department of Pathology and Clinical Bioinformatics, Erasmus University Medical Centre, Wytemaweg 80, 3015 CN, Rotterdam, the Netherlands; s.zhang@erasmusmc.nl (S.Z.); f.vankemenade@erasmusmc.nl (F.K.); a.stubbs@erasmusmc.nl (A.P.S.)

[2] Department of Gastroenterology and Hepatology, Erasmus MC-University Medical Center, Wytemaweg 80, 3015 CN, Rotterdam, The Netherlands; m.peppelenbosch@erasmusmc.nl (M.P.P.)

[3] School of Life, Health & Chemical Sciences, Faculty of Science, Technology, Engineering & Mathematics, The Open University, Milton Keynes, United Kingdom. (W.B)

* Correspondence: a.stubbs@erasmusmc.nl (A.P.S.).

Simple Summary: ScRNA-seq is a powerful tool to investigate the cancer microenvironment, but the cost of analysing every scientific scenario is prohibitive. Fortunately, deconvolution of bulk RNA-seq data with scRNA-seq cell atlas reference datasets provides a cheaper strategy. In this study, we validated the feasibility of deciphering the microenvironment of liver cancer through the estimation of cell fractions with Cibersortx and scRNA-seq atlas reference datasets. Five cell types are associated with patient outcomes, showing that deconvolution is a useful method to characterise the tumour microenvironment.

Abstract: Liver cancers give rise to a heavy burden of health care worldwide. Understanding the tumour microenvironment (TME) underpins the development of precision therapy. Single-cell RNA sequencing (scRNA-seq) technology has generated high-quality cell atlases of the TME, but its wider application faces enormous costs for various clinical circumstances. Fortunately, a variety of deconvolution algorithms can instead repurpose bulk RNA-seq data, alleviating the need for generating scRNA-seq datasets. In this study, we reviewed major public omics databases for relevance in this study and utilized 8 RNA-seq and 1 microarray datasets from clinical studies. To decipher the TME of liver cancer, we estimated the fractions of liver cell components by deconvoluting the samples with Cibersortx using three reference scRNA-seq atlases. We also confirmed that Cibersortx can accurately deconvolute cell types/subtypes of interest. Compared with non-tumorous liver, liver cancers showed multiple decreased cell types forming normal liver microarchitecture, as well as elevated cell types involved in fibrogenesis, abnormal angiogenesis and disturbed immune responses. Survival analysis shows that the fractions of five cell types/subtypes significantly correlated with patient outcomes, indicating potential therapeutic targets. Therefore, deconvolution of bulk RNA-seq data with scRNA-seq atlas references can be a useful tool to help understand the TME.

Keywords: liver cancer; tumour microenvironment; deconvolution

1. Introduction

Liver cancer is one of the leading causes of cancer-related mortality worldwide, making up 4.7% of newly diagnosed cases but 8.2% of deaths [1]. Hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA), which are frequently tallied together, constitute the major burden of liver cancer [2]. With a 5-year survival of 18%, liver cancer ranks as the second most lethal cancer [3]. The poor prognosis of liver cancer is partially

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

due to the insufficiency of effective therapies. Surgical interventions yielded the best outcomes but are limited by high recurrence rates or an easy loss of an operative window. Liver transplantation achieved better long-term survival but is hampered by an inadequate supply of donor organs [4,5]. Systematic non-specific chemotherapy resulted in disappointing results for both HCC and CCA [5,6]. Innovative agents targeting angiogenesis [7,8], fibrogenesis [9], and regulation of immune responses [10,11]. have shown the potential to improve outcomes. Such advances suggest that a shift from a cancer-centric to a tumour microenvironment (TME)-centric perspective is required in the future development of precision therapy [12,13].

Single-cell sequencing (scRNA-seq) technology delivers in-depth interrogation of the TME. The analysis of complex cancer tissues at a single cell level through scRNA-seq has brought insights into the heterogeneity and progression of cancer, as well as escape from immune surveillance, drug resistance and intercellular communication [14]. However, scRNA-seq technology is expensive, requires specific tissue collection methods, and can be difficult to implement. The cheaper and more common bulk RNA-seq studies make up the largest body of work in this area, filling public repositories, including flagship projects such as The Cancer Genome Atlas (TCGA) and its resulting resource the Pan-Cancer Atlas [15]. To make the most out of these available datasets, numerous algorithms have been proposed that enhance the informativeness of bulk transcriptome analysis using reference scRNA-seq data. Machine learning is the major approach of such methodologies. For example, stemness indices within the Pan-Cancer Atlas were estimated with the Progenitor Cell Biology Consortium datasets and one-class logistic regression [16]. Another group of deconvolution algorithms focus on the profiling of cell fractions of the bulk transcriptomic data, such as MuSiC [17] and Cibersortx [18]. Meanwhile, emerging scRNA-seq atlases (e.g., Human Cell Atlas) provide high-quality references, allowing accurate deconvolution of bulk RNA-seq data in increasingly wider contexts (e.g., profiling of TME for head and neck cancers [19]).

In this study, we reviewed major omics databases and selected ten studies that compared transcriptomes between HCC/CCA and normal liver. The cell fractions of tumour and non-tumour tissues were estimated with Cibersortx and three scRNA-seq reference atlases. The included studies contain two cohort studies of HCC, allowing us to investigate the clinical implications of TME abnormalities through survival analysis and gene set enrichment analysis (GSEA) [20]. We determined that the TMEs of liver cancer lack multiple cell types (e.g., hepatocytes) that normally form the liver microarchitecture, and instead are enriched with components involved in fibrogenesis, abnormal angiogenesis and irregular immune activities. Among all the cell types/subtypes in the HCC TME, hepatocytes and mature B cells are positively correlated with patient outcomes, while cholangiocytes, bi-potent stem cells, plasma B cells and regulatory T (T_{reg}) cells correlate with negative outcomes.

2. Materials and Methods

2.1. Data Obtainment

We searched three public omics databases – the Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), and the International Cancer Genome Consortium (ICGC) – for studies engaging on liver cancers. Inclusion criteria were: (1) histologically identified sample series from human tissues in clinical studies (including cohort studies and case review series); (2) Whole transcriptomes by microarray or RNA-seq; (3) with non-tumorous or normal liver tissues as controls (for comparison) or follow-up information (for interrogation of clinical outcomes); (4) studies covering two major liver cancers (HCC and CCA) were included and all other studies were excluded, i.e., metastatic liver cancers and non-malignant hyperplasia (e.g., hemangioma). Finally, ten curated datasets (nine RNA-seq and one microarray) were selected for this study (Table S1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

2.2. Preprocessing of Microarray Data

Microarray studies with raw data (.CEL files) in GEO database were obtained via R package SCAN.UPC [21] which provides one-step preprocessing through empirical correction of major bias (GC content-related). This normalization method for microarray datasets has been proven to be reliable for downstream analysis [22]. All gene names were translated into HGNC symbol with R package BioMart [23]. Duplication of features in expression matrices were collapsed by MaxMean strategy [24].

2.3. Preprocessing of RNA-seq Data

TCGA-LIHC was obtained via UCSCXenaTools [25] and datasets in ICGC were obtained via the official web portal [26]. Expression matrices of other studies from the GEO database were retrieved according to the authors' instructions. All datasets recruited in this study have been listed in Supporting Information. All gene names were translated to HGNC symbol with R package BioMart [23]. Duplication of features in expression matrices was removed with summation method [24].

2.4. Deconvolution of Cell Types with Cibersortx and Three Atlases

Three single-cell RNA-seq datasets were used in this study: (1) GSE115469 (Normal Liver), (2) GSE146409 (TME-Stroma), and (3) GSE156337 (TME-Immune). GSE115469 is a liver subset of Human Cell Atlas [26]. GSE146409 contains the TME of liver cancer (HCC and CCA), including malignant cells [27]. GSE156337 contains the HCC microenvironment, without malignant cells [28]. This dataset identified high-quality immune cells.

All expression matrices were normalized to 10,000 counts/cell and packed into an H5AD file with authors' annotation as metadata for subsequent estimations. Both the signature matrix of scRNA-seq datasets and the expression matrix of bulk RNA-seq datasets were transformed into tab-delimited tables. The signature matrices of reference scRNA-seq were built with the Cibersortx protocol for "scRNA-seq" ("S"). Deconvolution was performed with the "Impute Cell Fractions" module. In validation experiments, batch correction was disabled in within-study validation and activated in "S" mode (with single cell expression matrix as reference) in cross-study validation (two groups of validation experiments will be described below). In the estimation of real clinical data, batch correction was enabled in "S" mode for RNA-seq datasets and "B" mode (with single cell expression matrix collapsed into bulk matrix before used as reference) for microarray datasets. All other Cibersortx parameters were the default configurations [18].

Cibersortx fails when the cell type tree of the reference atlas is complicated (e.g., > 10 cell types). In this situation, collapsing some branches of the cell type tree would complete the calculation [18]. In our study, for example, when we used the Normal Liver atlas, the cell type tree was divided into two groups (immune and non-immune groups). Similarly, the TME-Stroma atlas was divided into three groups (mesenchymal, vasculature and immune groups), and the TME-Immune was divided into two groups (immune and non-immune groups) (All cell type trees are shown in **Figure S1**). To evaluate the influence of adjusting the cell type tree, we performed a group of validation experiments in which cell subtypes in the Normal Liver atlas were collapsed (such that macrophage = inflammatory macrophage + non-inflammatory macrophage, T cell = alpha-beta T cell + gamma-delta T cell, B cell = mature B cell + plasma B cell, and LSEC = periportal LSEC + central venous LSEC). To increase the matching in cross-study experiments (described below), the cell-type trees of the other two atlases were also adjusted (for TME-Stroma atlas, let macrophage = scar-associated macrophage + Kupffer cell + tissue macrophage, and for TME-Immune atlas, let T cell = $CD4^+$ T cell + $CD8^+$ T cell + T_{reg} cell).

2.5. Accuracy and Robustness of Cell Fraction Estimation

The accuracy and robustness of deconvolution with Cibersortx were tested in two groups of experiments: intra-study validation and cross-study validation. In intra-study

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

validation experiments, the signature matrix of scRNA-seq and the pseudobulk for testing were generated from the same scRNA-seq atlas. The advantage of this method is a full coverage of all cell types/subtypes. In cross-study validation experiments, the signature matrix of scRNA-seq and the pseudobulk dataset were generated from different atlases.

Pseudobulk datasets were generated using the random module of NumPy in the following procedure: the expression of the scRNA-seq atlas was separated into two groups, cell type of interest and all the remaining cell type/subtypes collapsed into a single group labeled as "others"; 10% of cells in each group were selected using Choice function of NumPy, then two representative expression vectors (V) were generated by calculating the mean value of each gene; a random number f (between 0 and 100) was generated by Uniform function of NumPy; finally, the expression vector of the pseudobulk sample was generated by $V_{celltype} \times f + V_{others} \times (100 - f)$. f was used as benchmarking target.

Two parameters were used to evaluate the accuracy of Cibersortx estimation, (1) Pearson correlation coefficient (PCC) between predefined proportions (f) and estimated fractions (f'); (2) mean absolute error (MAE =$1/n \sum_{i=1}^{n} |f'_i - f_i|, i = 1, \ldots, n$) and direction (overestimation or underestimation).

2.6. Survival analysis, Statistics and Data Visualization

To demonstrate the added value of our deconvolution analysis, we investigated the survival impact of estimated cell fractions on two HCC cohorts (TCGA-LIHC, GSE14520). The patient cohort was first ordered based on descending order of estimated level of each cell fraction and then separated into high- and low-level groups. All separation possibilities (from 1:n-1 to n-1:1) have been tested with log-rank tests. The one with the lowest P-value in log-rank test was selected as the optimized separation. If all the P-values were above 0.05, the cohort was equally separated into two groups (median-point separation).

Survival lengths were transformed into months and observed events (labeled as "1") were transformed into 'True' (Boolean value, according to the requirement of Scikit-Survival [29]). Kaplan-Meier (K-M) curves were then fitted with Scikit-Survival [29] and plotted with the Step function of Matplotlib. The log-rank test was calculated with Lifelines [30]. All boxplots of this study were generated with MatPlotLib.

GSEA was performed with GSEApy. The input gene list for GSEA was the marker genes selected by Cibersortx for the cell types. Our study shows a demonstration with "WikiPathway 2021 – Human" as the reference. GSEApy allows dozens of different libraries, and we provide scripts in our GitHub repository.

For better reproducibility, all the analysis scripts incluing preprocessing have been shared with GitHub (https://github.com/ErasmusMC-Bioinformatics/OIO_Shaoshi).

3. Results

3.1. Cibersortx Estimation of Cell Fraction

In this study, we aim to decipher the TME of liver cancer by estimating the cell fractions in transcriptomes. This requires an accurate and robust model, in addition to well-annotated single-cell atlases. We adopted a state-of-the-art deconvolution algorithm (Cibersortx) and three scRNA-seq atlases (Normal Liver, TME-Stroma and TME-Immune). These atlases describe more than 20 cell types or subtypes. **Figure 1A** outlines the workflow of this study, and the cell-type trees of the three atlases are outlined in **Figure S1**.

We first performed intra-study validation experiments to test whether all of the identified cell types from each atlas, can be accurately deconvoluted with Cibersortx for each pseudobulk dataset generated by the same atlas (**Figure 1**). In this validation mode, most cells achieved ideal PCCs (**Figure S2A,3-4**). The PCCs compare the relationship between predefined cell fractions in generated pseudobulk samples and estimated values by Cibersortx. However, MAEs vary between different cell types (**Table S2**, **Figure S5**). To evaluate the effect of merging some cell types, we did a group of tests by combining

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

the major subtypes in Normal Liver atlas (**Figure S2B**). Both panels in **Figure 1B** show the result of intra-study validation with Normal Liver atlas. Although both experiments show high levels of PCCs, subtle differences in accuracy exist. For example, after merging eight cell subtypes into four major cell types, the PCC of hepatocyte increases (0.979 vs. 0.992, **Table S2**, the same the following), cholangiocyte decreases (0.9915 vs. 0.9844), and hepatic stellate cell (HSC) decreases (0.9901 vs. 0.986).

Although Cibersortx allows "partial deconvolution", in which the samples may contain cell type/subtypes not present in the reference atlas, we performed a group of cross-study validation experiments to assess this impact. The results of these experiments demonstrate that the PCC between presets and prediction remain high (**Figure 1E**, **Figure S5**) whilst the MAEs vary significantly between cell types (**Table S2**). Our study suggests that Cibersortx normally guarantees high levels of PCCs but MAEs vary when using partial deconvolution method. Thus, we adopted a protocol in all subsequent analysis such that if a cell type (e.g., hepatocytes) is present in multiple atlases, the one with the best performances in both intra-study and cross-study validation experiments was chosen as the reference for real clinical data.

3.2. Difference of Cell Fraction between Tumour and Non-Tumor Liver Tissue

To determine the difference in cell fraction between tumour and non-tumour tissues, we compared seven RNA-seq datasets which provide paired tissues collected from HCC/CCA cohort studies or case review studies. LIRI-JP is an RNA-seq study which includes primary liver cancer, secondary liver cancer from stomach, colon, and prostate, etc., with adjacent non-tumorous liver tissues as controls. GSE119336 is an RNA-seq study comparing CCA and non-tumour liver. The other five RNA-seq studies compare HCC and non-tumour liver tissues. Three of these studies included cases with HBV infection as the risk factor.

The three atlases provide more than 20 cell types, allowing a panoramic view of TMEs. All of these cell types can be largely divided into three groups: (1) cell types underpinning the fundamental physiology of liver, e.g., hepatocytes, cholangiocytes; (2) cell types participating in the pathological evolution of cancer formation, e.g., HSCs, LVECts; (3) immune cells. Results of related cell types are described in the following groups.

3.2.1. Hepatocytes and Cholangiocytes

These two cell types constitute the major functional units of liver – liver lobule and bile ducts [26]. As both the Normal Liver and TME-Immune atlases have hepatocytes, we made a comparison of results by deconvolution with two different atlases for the same dataset. **Figure 2A-B** show the results of deconvolution for the same datasets with different atlases. Compared with the results determined with the TME-immune atlas, deconvolution with the Normal atlas returned higher levels of hepatocytes for some datasets (such as GSE119336) and lower levels for others (such as GSE94660). However, both experiments arrived at the same result whereby the decrease of hepatocytes and elevation of cholangiocytes are common in liver cancer. Similarly, **Figure S6D** generated by deconvolution with TME-Stroma resulted in the same conclusion. Compared with HCC, the decrease of hepatocytes in CCA is more significant (**Figure 2A-B**).

3.2.2. Fibrogenesis

HSCs are tissue-specific equivalents of pericytes, and pericytes can be rarely detected in normal liver [27,31]. Fibroblasts proliferate following the activation of HSCs. Compared with non-tumorous tissues, a decrease of HSCs and an elevation of pericytes can be seen in liver cancers. CAFs are rarely detected in normal liver and are more often seen in liver cancers. The opposite alterations between HSCs and pericytes/CAFs suggest active fibrogenesis in tumours (**Figure 2D-F**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

3.2.3. Vasculature

Liver sinusoidal endothelial cells (LSECs) form the wall of hepatic sinusoids and participate in immune responses. Vascular smooth muscle cells (vSMCs) are also key components of blood vessels. Tumour liver vascular endothelial cells (LVECts) are variants of normal LVECs [27]. Decreased LSECs, vSMCs, and LVECs can be seen in cancerous tissues, while the elevation of LVECt can be observed in liver cancers (**Figure 3A-D**). These results suggest the emergence of abnormal angiogenesis.

3.2.4. Immune Cells

T cells. In the three atlases, five subsets of T cells were identified: T cells with alpha-beta (αβ) TCR chains or gamma-delta (γδ) chains (Normal Liver atlas) [26]; $CD4^+$ (helper), $CD8^+$ (cytotoxic) T cells, and T_{reg} cells (TME-Immune atlas) [28]. We observed obvious elevations of αβ T cells but no clear alterations of γδ T cells in liver cancers. $CD4^+$ T cells rise sharply in CCA while $CD8^+$ T cells elevate moderately in HCC. Finally, T_{reg} cells are rarely detected in normal livers whereas elevations are common in liver cancers. T_{reg} cells have been recognized as a suppressor of tumour immune responses. Liver cancers also show high levels of overall T cells ($CD4^+$+$CD8^+$+T_{reg} cells). (**Figure 4A-F**).

B cells. The Normal Liver atlas identified two subtypes of B cells, mature B cells (antigen inexperienced) and plasma B cells (antigen secreting) [26]. Variation of mature B cells doesn't show a direct association with liver cancers. Plasma B cells were rarely detected in normal liver, but were widely detected in liver cancers (**Figure 5A-B**).

Macrophages. The Normal Liver atlas contained inflammatory and non-inflammatory macrophages [26]. The TME-Stroma atlas separated Kupffer cells, tissue monocytes (TMs), and scar-associated macrophages (SAMs) [27]. SAMs are often recruited in the process of liver fibrosis [27,32]. We observed elevations of inflammatory macrophages and SAMs in liver cancers. No obvious differences in Kupffer cells, TMs and non-inflammatory macrophages were seen (**Figure 5C-F**).

Dendritic cells. As the quintessential antigen-presenting cell, the dendritic cell (DC) is another hot research interest for the potential of immunotherapy. The TME-Stroma atlas was used to estimate conventional DC1 and DC2 (cDC1 and cDC2) cell types [27,33]. Deconvolution suggests that cDC1 cells are elevated in liver cancers while cDC2 cells are not (**Figure 6A-B**).

In summary, liver cancers show higher levels of overall immune cells, involving both the innate (monocyte-macrophages) and adaptive branches (T, B cells), as well as auxiliary components (dendritic cells). Meanwhile, suppressive components such as T_{reg} cells can also be observed, suggesting the disordered responses in tumours.

3.2.5. Bi-potent Stem Cells and Proliferative Cells

This group involves two cell types which can proliferate and differentiate. Bi-potent stem cells (from TME-Immune atlas) were named for their potential to differentiate into both hepatocytes and cholangiocytes [28]. Proliferating cells were identified in TME-Stroma atlas and elevation of these two cell types (HCC and CCA) was common in tumours [26]. Bi-potent cells were rare in normal livers and their elevation in CCA is prominent (**Figure 6C-D**).

3.2.6. Other Cell Types

These three atlases also identified some immune cell types which exist in the liver with a relatively low abundance. TME-Immune identified a cluster of natural killer (NK) cells (a key component of the innate immune branch) and a cluster of myeloid cells (the liver-resident precursors of monocytes-macrophages) [28]. The Normal Liver atlas isolated a cluster of NK-like cells, which may be an ambiguous mixture of natural killer T (NKT) cells and NK cells [26]. Different atlases may have some cell types/subtypes with the same labels. However, calculated signature matrices suggest that they have different

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

scopes, e.g., HSCs in Normal Liver atlas, **Figure 2D**, *vs*. HSCs in TME-Stroma atlas, **Figure S6A**).

3.3. Cell Fraction of HCC TME Correlates with Clinical Outcome

Finally, we investigated whether cellular alteration affects clinical outcome of HCC through survival analysis. In public repositories, TCGA-LIHC is the highest-cited cohort of a liver cancer study, with well-annotated follow-up information and substantial sample size (370 HCC patients). TCGA-LIHC is a pooled study of five cohorts with mixed risk factors. Available survival analyses include overall survival (OS) and disease-free survival (DFS) [34].

The distributions of estimated cell fractions show two typical shapes, "Sigmoid" or "Exponential" (**Figure 7C**, **F**). Using an optimization strategy (lowest log-rank test p-value), the patient cohort was typically separated at inflection points, although this separation may fail in cases of negative results or meaningless grouping (e.g., separating 1 case into a group). In these circumstances, we used the median-point strategy to finish complete K-M curves (**Figure 7**, **Figure S7-12**).

Among all the estimated cell types, hepatocytes and bi-potent cells show a prominent impact on patients' outcomes. The estimated fractions of hepatocytes show an S-shaped distribution. The optimization strategy (lowest log-rank test p-value) separates the cohort at a close-to-median point in OS analysis, and at an inflection point for DFS analysis (**Figure 7C**). High fractions of hepatocytes are associated with longer OS and DFS (**Figure 7A-B**). Estimated fractions of Bi-potent cells show exponential distribution. The optimization strategy isolated a subset with the cell fractions close to zero (**Figure 7F**). Those with high fractions of bi-potent cells show both lower OS and DFS (**Figure 7D-E**).

GSE14520 is a study which recruited more than 200 HBV-related HCC cases and provides both OS and DFS information [35,36]. Its transcriptomic tests are based on microarray platforms, which may provide less accuracy than RNA-seq. In our study, rare cell types were not often detected in the deconvolution of microarray data. However, given the subtle difference of study protocol, it still provided alternative evidence about the impact of cell fractions on patients' outcomes. A compilation of the survival analyses for GSE14520 is available in **Figure S10-12**, with a summary in **Table S3**. Consistent and significant results for both OS and DFS in the two cohorts include: hepatocyte (positive), cholangiocyte (negative), bi-potent stem cell (negative), Mature B cell (positive), Plasma B cell (negative), and T_{reg} cell (negative).

The mathematical essence of Cibersortx is the quantification of signature gene expression of a cell type in a tissue context [18]. The generated signature matrices by Cibersortx may provide useful information for further analysis, e.g., pathway activities. These analyses may reveal clues for development of precision therapy. GSEA varies depending on the selection of the reference library. **Figure S13** is an example using the signatures from the three atlases generated by Cibersortx with the library "WikiPathway 2021 – Human" as the source of pathway definitions. All Signature matrices and scripts have been shared online (see GitHub address).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

Fig. 1

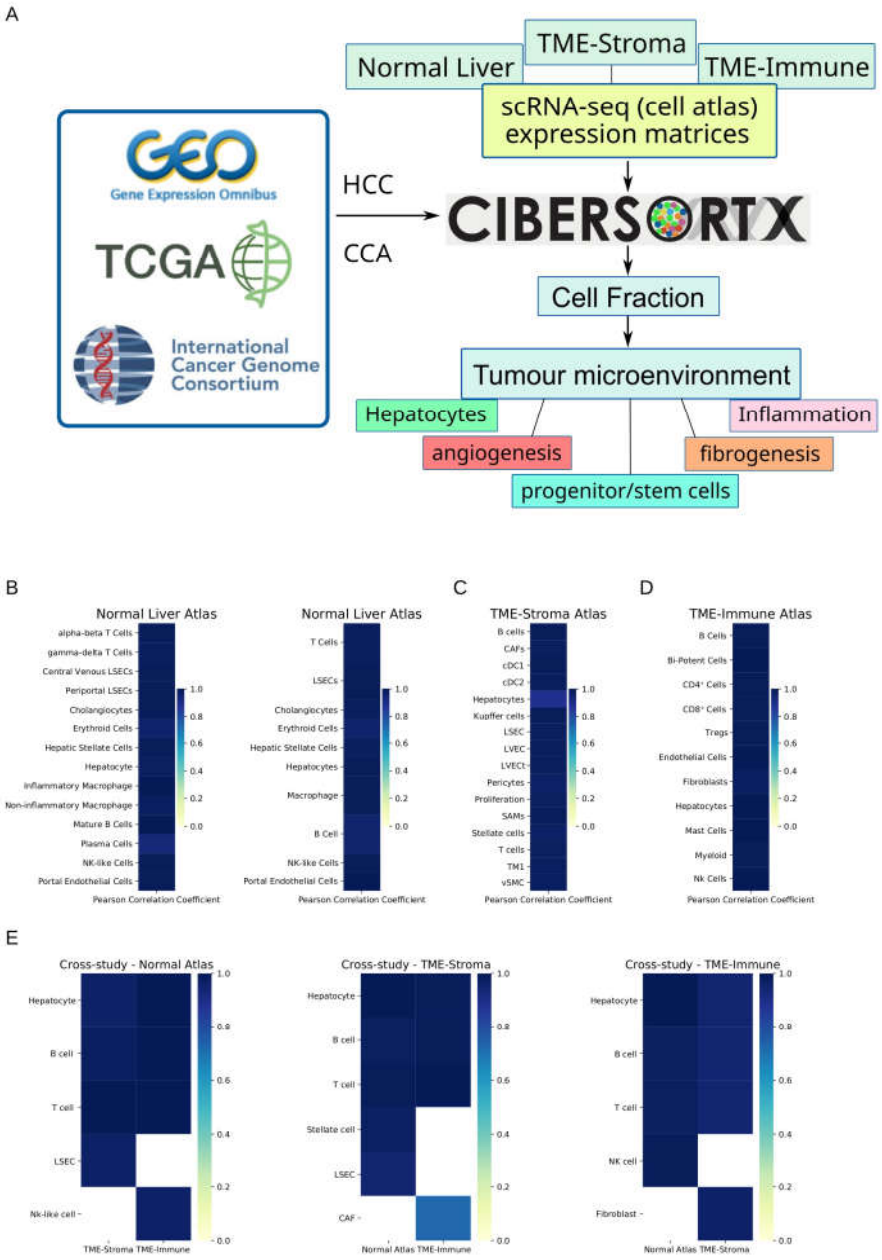

Figure 1. **A**. A general workflow of deciphering tumour microenvironment. Cibersortx estimates the cell fraction with scRNA-seq atlas and bulk RNA-seq/microarray data. Three expression matrices of scRNA-seq study were used as reference atlases. Through estimation of the fractions of more 20 cell types/subtypes, biological events could be inferred. **B-D**. Performances of Cibersortx deconvolution in intra-study validation experiments. **E**. Performances of Cibersortx deconvolution in cross-study validation experiments. The title of each panel indicates which reference atlas has been used in the deconvolution, two ticks on the bottom indicate which dataset has been used to generate the pseudobulk samples.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

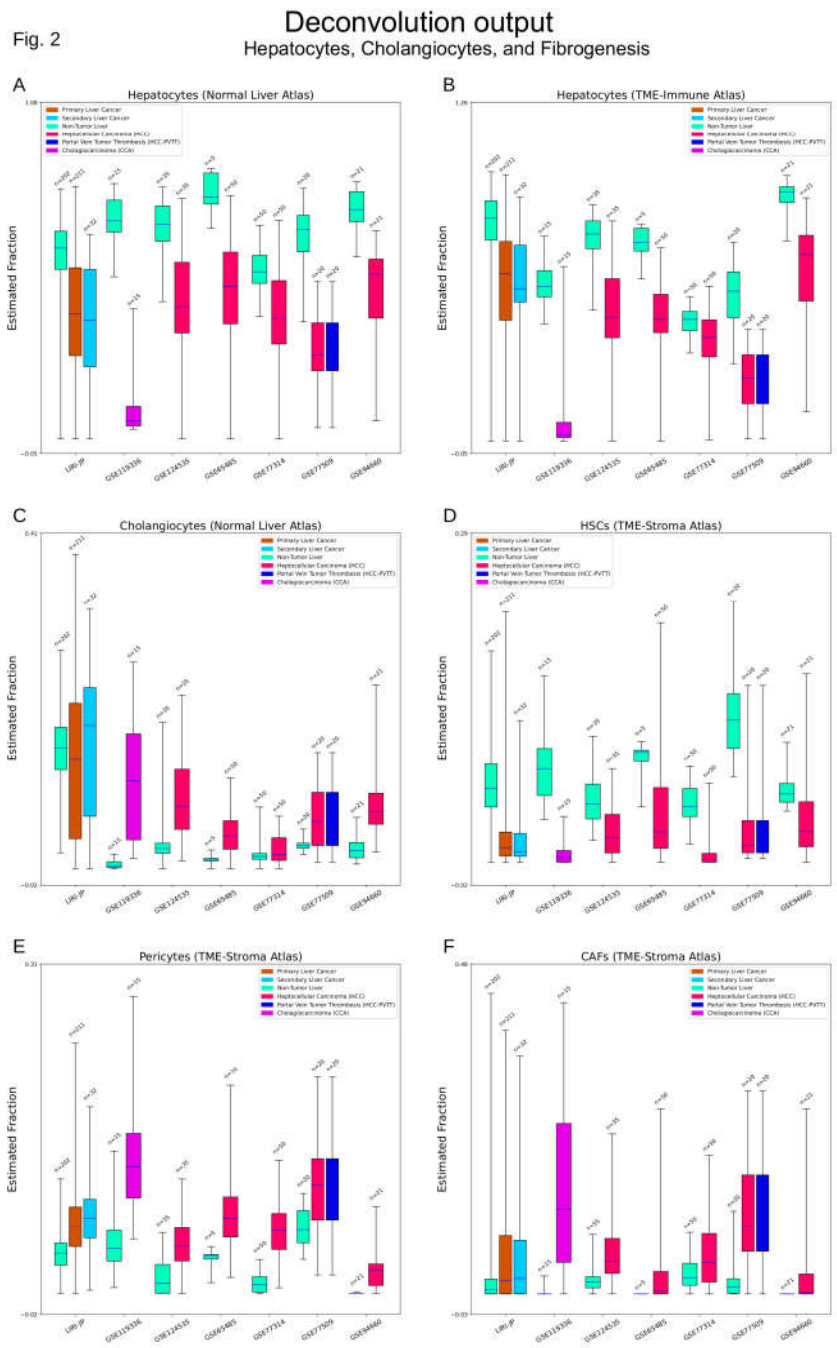

Figure 2. **A-B**. A comparison of hepatocytes estimation by Cibersortx with two different atlases. Subtle differences of cell fractions can be seen between two experiments, but they arrived at the same conclusion that hepatocytes decrease in cancerous tissue. In CCAs, this phenomenon is more prominent. **C**. Estimation of cholangiocytes. Elevation of cholangiocytes can be widely seen in cancerous tissues. **D-F**. Estimation of hepatic stellate cells (HSCs), pericytes and cancer-associated fibroblasts (CAFs). These three cell types are key components participating the fibrogenesis of liver cancer. HSCs are liver-specific pericytes, and often activated in liver cancer. Pericytes (broad sense) can be hardly detected in normal liver but widely present in liver cancer. CAFs are uniquely characterized fibroblasts in cancer. They can be hardly detected in normal liver but common in cancerous tissues.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

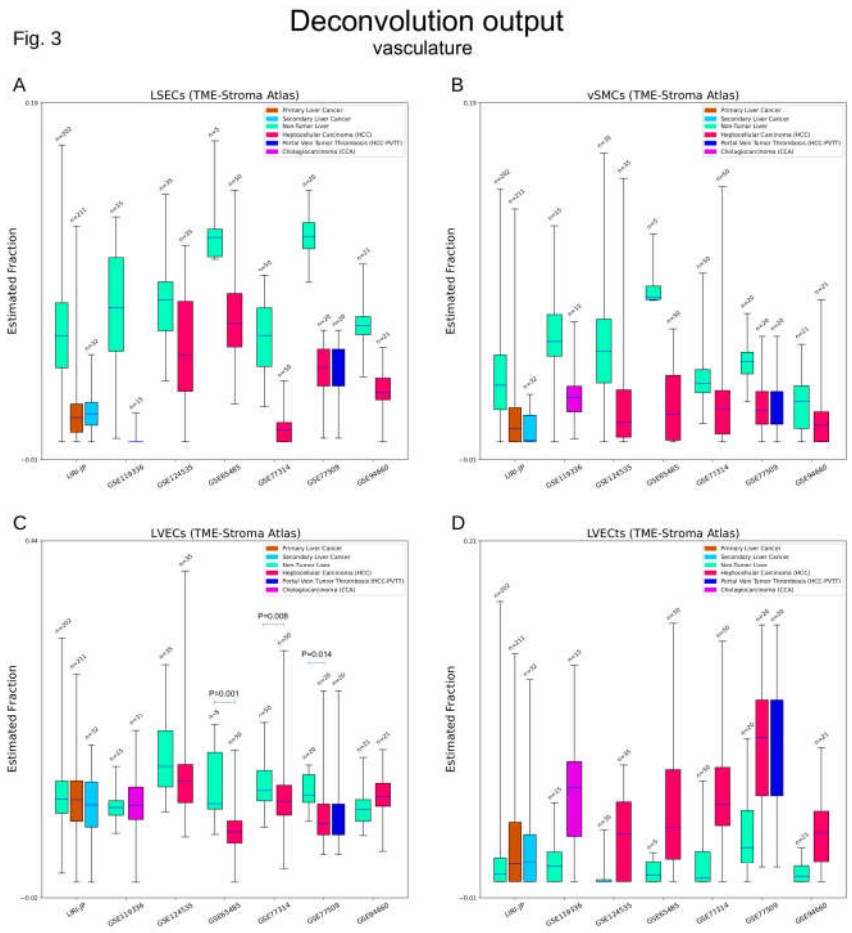

Figure 3. Estimation of liver sinusoidal endothelial cells (LSECs), vascular smooth muscle cells (vSMCs), liver vascular endothelial cells (LVECs) and tumour LVECs (LVECts). These four cell types/subtypes are key components of blood vessels in both normal and cancerous livers. Decreases of normal structural blocks (LSECs, vSMCs) and abnormal angiogenesis (LVECts) can be seen in liver cancers. Not all the differences are statistically significant. Those pairs with P<0.05 have been labeled.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

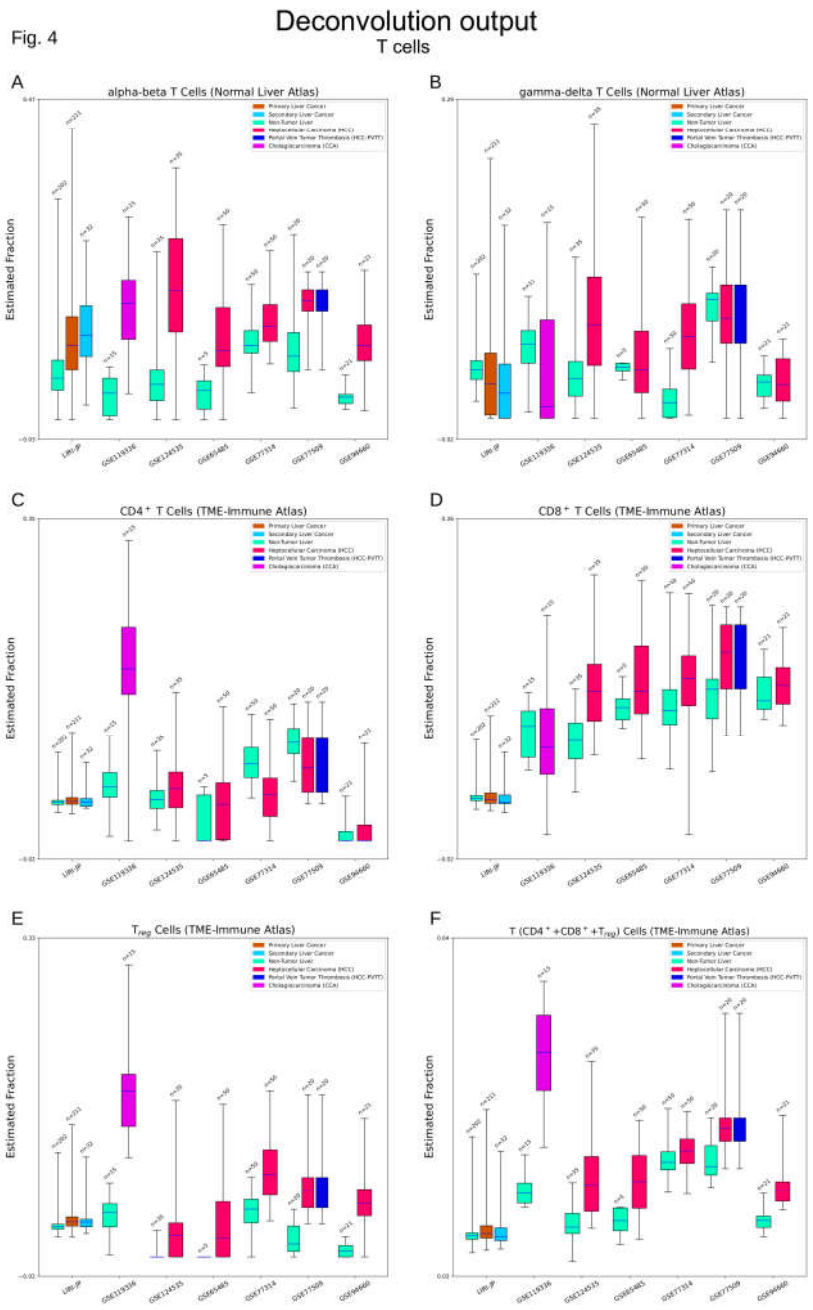

Figure 4. **A-E**. Estimation of T cells. Alpha-beta ($\alpha\beta$) T cells and gamma-delta ($\gamma\delta$) T cells are from Normal Liver atlas. $CD4^+$/$CD8^+$ T cells and regulatory T (T_{reg}) cells are from TME-Immune atlas. Overlap of cell type trees may exist between the two atlases. Obvious elevations of $\alpha\beta$ T cells and obscure changes of $\gamma\delta$ T cells can be seen in liver cancers. $CD4^+$ T cells rise prominently in CCAs and $CD8^+$ T cells moderately in HCCs. **F**. Estimation of overall T cells. Higher activities of T cells can be seen in liver cancers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

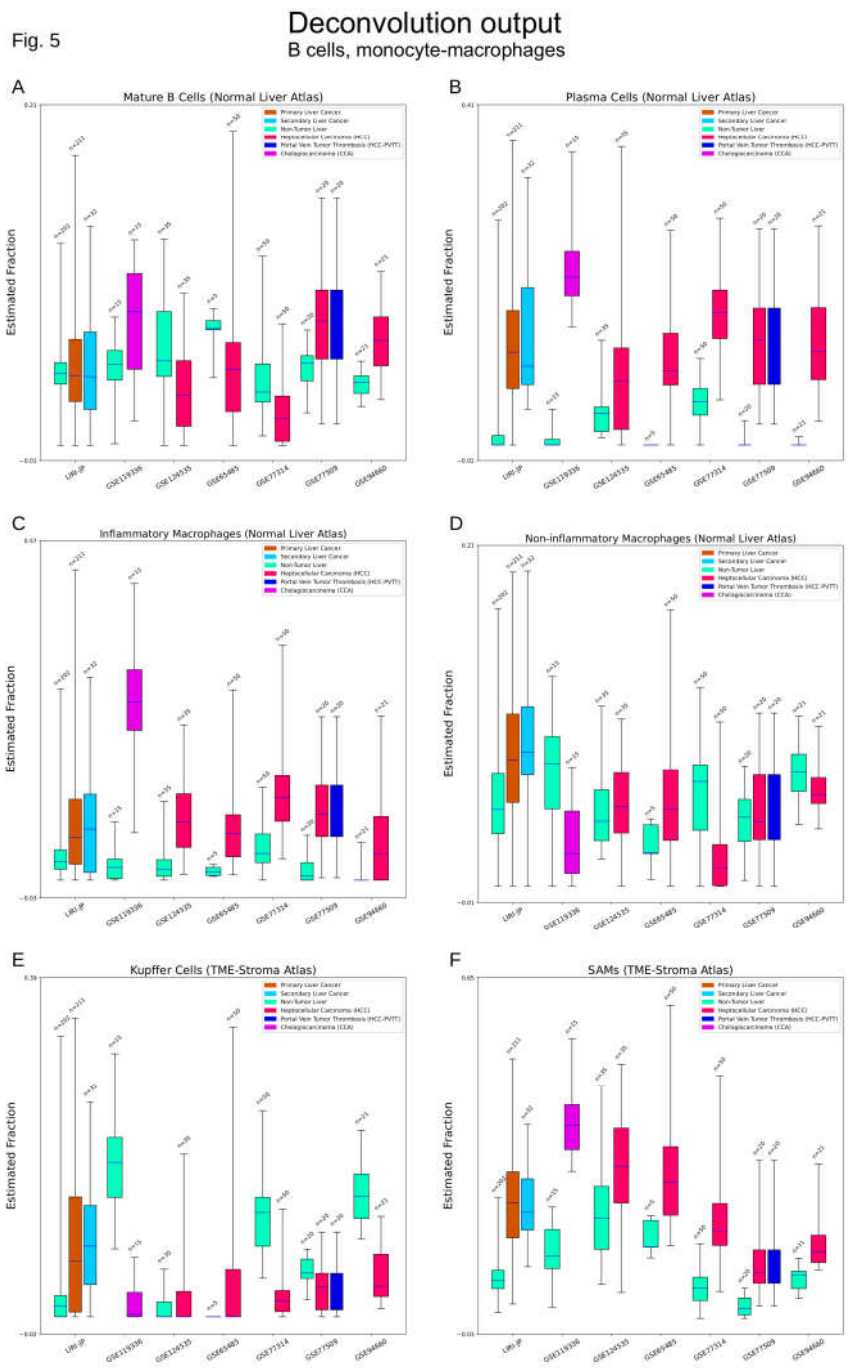

Figure 5. **A-B**. Estimation of B cells. Plasma B cells can be rarely detected in normal livers but common in liver cancers. **C-F**. Estimation of macrophages. The Normal Liver atlas separates macrophages into the inflammatory and non-inflammatory subtypes. The TME-Stroma atlas identified Kupffer cells, scar-associated macrophages (SAMs) and tissue macropahges (TMs, **Figure S6E**). Overlap of cell type trees may exist between the two atlases. Increase of inflammatory macrophages can be seen in liver cancers. SAMs are a pathological variant of macrophages and participate in the process of liver fibrosis. Their elevation can be seen in liver cancers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

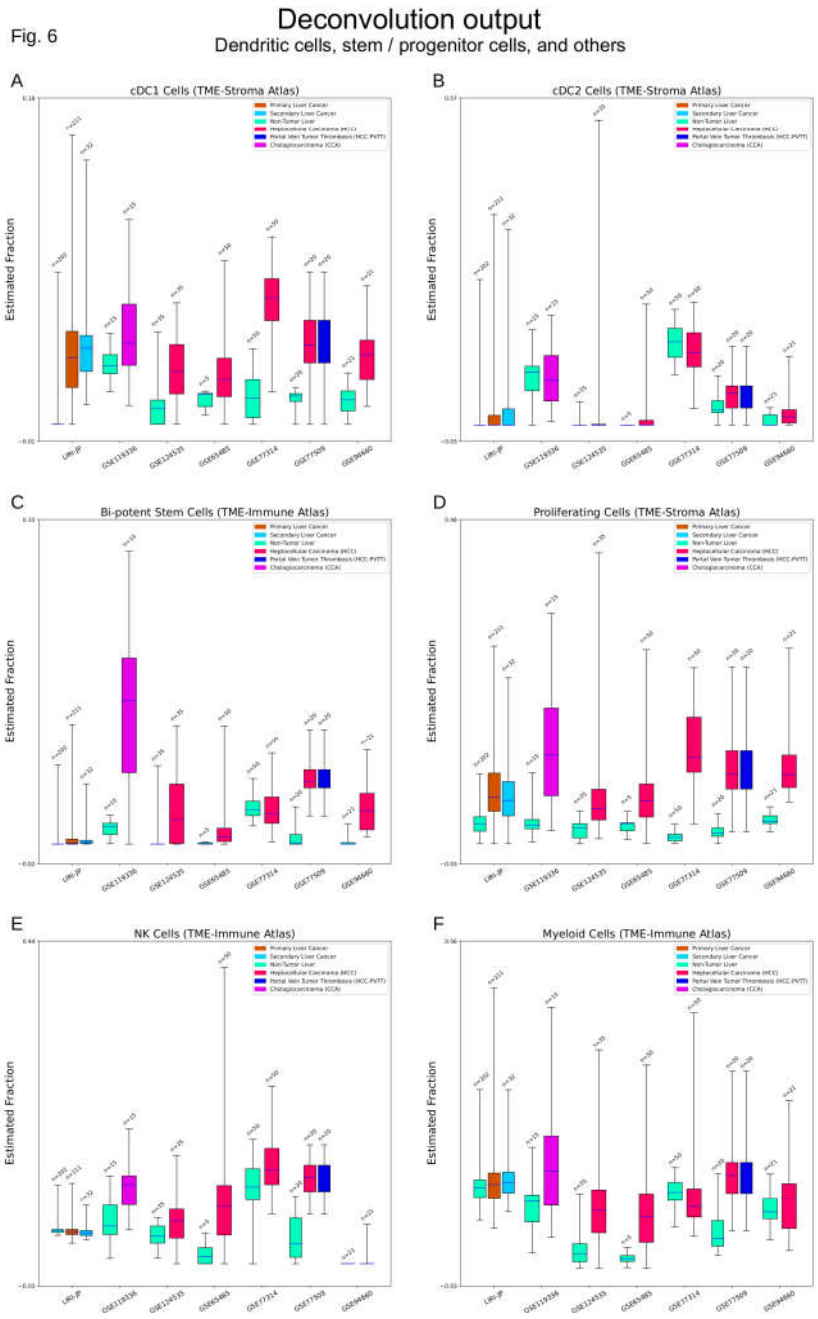

Figure 6. **A-B**. Estimation of conventional dendritic cells (cDCs). Elevations of cDC1s can be seen in liver cancers. **C-D**. Estimation of two types of proliferative cells. Bi-potent stem cells are a group of late-stage pluripotent cells with potential to differentiate into hepatocytes and cholangiocytes. The TME-Stroma atlas did not clarify the exact characteristics of proliferating cells. Its signature genes suggest its pluripotent origin. Elevations of these two cell types can be seen in liver cancers. **E-F**. Estimation of natural killer (NK) cells and myeloid cells. NK cells belong to innate immune branch and myeloid cells are the hematopoiesis-originated immune branch. Deconvolution of these two cell types shows altered activities in liver cancers but no direct association can be drawn.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

Fig. 7

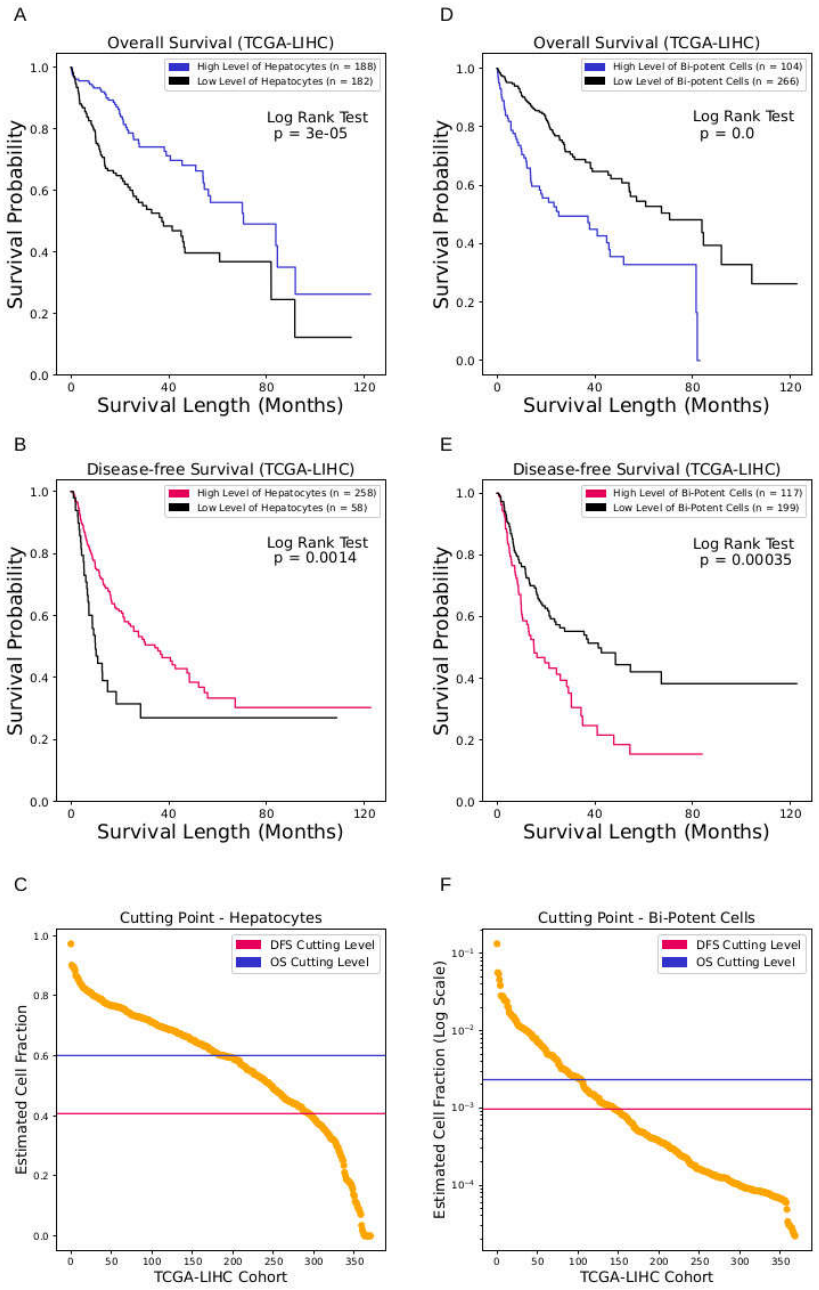

Figure 7. Impact of cell fractions on patients' survivals. **A-B**. Impact of hepatocyte fraction. High-level hepatocytes show both longer OS and DFS lengths. **D-E**. Impact of bi-potent cell fraction. High-level bi-potent cells show lower OS and DFS lengths. **C**, **F**. Cutting thresholds of overall survival (OS) and disease-free survival (DFS) analysis after optimization. The fraction of hepatocytes and bi-potent cells show two typical types of distribution. Hepatocyte fractions have a sigmoid-shaped curve and the cutting points are at an inflection point of the curve. Bi-potent cell fractions show an exponential distribution (Y-axis has been log-transformed for better visualization) and cutting points extract one group with estimated fractions close to zero.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

4. Discussion

Deconvolution algorithm attracts high interest for it allows the estimation of cell fractions with scRNA-seq atlases. Cibersortx and MuSiC are two state-of-the-art algorithms in benchmarking studies. Cibersortx was adopted in this study for its easy reproducibility while implementation with MuSiC involves R scripting and manual selection of markers. Cibersortx first constructs a representative profile and signature matrix for a given scRNA-seq atlas. In practice, the atlases with too many cell types often overdrive Cibersortx, leading to the failure of calculation. According to the official guidance of Cibersortx, an adjustment of cell labelling of the atlas is then warranted [18]. We performed a set of experiments with the Normal Liver in intra-study validation group, to determine the influence of this adjustment of cell type labelling. We found that collapsing the branches of LSEC, T/B cells and macrophages resulted in slightly different PCCs but did not substantially impact the accuracy of deconvolution. In our practice, we split the atlases into several groups to ensure that the deconvolution with Cibersortx ran to completion. In the estimation of cell types in one group, all cells of the other group were combined and labelled as "other".

Another salient advantage of Cibersortx is the partial deconvolution feature which is applied when the tissue contexts of atlases and bulk RNA-seq data are not necessarily identical [18]. This situation is widely present in applications with clinical data. Two main factors influence the differences between atlases: emerging new cell types/subtypes and transforming cell states. Diseased livers often contain newly generated cell types, such as fibroblasts in cirrhosis, infiltrated lymphocytes in viral hepatitis, or malignant cells in liver cancer, etc. Moreover, diseases can also shift the cell states, such as in the activation of HSCs or T/B cells. To validate the reliability of partial deconvolution by Cibersortx, we performed a set of cross-study validation experiments. Due to the limitation of available data, our validation did not cover the full spectrum of cell types found in liver. In these limited *in silico* experiments, we determined that partial deconvolution guarantees high-level PCCs but that MAEs vary between different cell types.

Cibersortx estimates cell fractions through quantifying the abundance of signature genes, which warrants careful consideration when interpreting the results. Cell fractions can be defined in diverse ways. In conventional histological studies, cell proportion has been calculated by volume, cell number, or mass, etc [37]. The prediction of Cibersortx seems to be close to the definition of "fraction by cell number". However, the expression of signature genes varies between cells, tissues, individuals, and the different disease states, leading to an ambiguity of the concept "fraction". This gap becomes prominent in situations where the biomedical conditions of the reference scRNA-seq atlases and the bulk RNA-seq samples are different (necessitating "partial deconvolution"). In addition, cell clusters between atlases with the same label may not be identical. In our study, we preserved all the signature matrices generated by Cibersortx for better comparison. Therefore, we recommend taking into account all of these factors when interpreting the biomedical implications of the deconvoluted results. [18].

Liver cancer is the second most lethal cancer due to the difficulty of early diagnosis and the lack of efficient therapies. Potent therapeutic strategies are often hampered by the accompanying components in the TME [4], most notably, the CAFs and T_{reg} cells. CAFs are believed to interfere with immune responses, providing local stiff niches that shelter tumour-initiating cells and low permeability barriers that impede the delivery of tumour-eradicating chemicals [9]. T_{reg} cells are a suppressive subpopulation of immune cells, leading to tolerance of malignancy. T_{reg} cells function through PD-1/PD-L1 pathway and a blockade of this communication results in the resurrection of immune responses in a minor group of patients. Further studies show that the therapeutic efficiency of a PD1 inhibitor depends on the interaction between the TME and other immune components (e.g., $CD8^+$ T cells) [38,39]. Our study shows elevations of both CAFs and T_{reg} cells in cancerous tissues, and a negative impact of T_{reg} cells on patient outcomes. These findings

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

are consistent with present literatures, suggesting the feasibility of deconvolution with scRNA-seq atlases as a data mining tool to reveal clues towards precision therapy.

The exact mechanism of liver cancer formation remains partially understood. $EpCAM^+$ cells were proposed as a tumour-initiating component in HCC development [40]. EpCAM expression has been found in fetal liver, hepatic progenitor cells, and carcinoma cells, etc., but not in mature hepatocytes [41]. In our study, EpCAM was selected by Cibersortx as a signature gene for cholangiocytes (Normal Liver atlas), proliferating cells (TME-Stroma), and bi-potent stem cells (TME-Immune). Survival analysis suggests a negative impact of bi-potent stem cells on patient outcomes, providing alternative clinical evidence for the tumour-initiating hypothesis of $EpCAM^+$ cells. Although the interpretation of these findings warrants careful consideration, our study demonstrates that deconvolution can also help understand the mechanism of cancer formation.

5. Conclusions

In this study, we decipher the TME of liver cancer by estimating the cell fractions of a sample given a transcriptome. By estimating more than 20 cell types/subtypes within bulk RNA-seq data using three atlases and Cibersortx, we found disruption of normal liver architecture, abnormal fibrogenesis and angiogenesis, as well as disturbed immune responses in HCC and CCA. Survival analysis demonstrated that five cell types/subtypes highly correlated with patient outcomes.

Deconvolution algorithm and emerging scRNA-seq atlases allow the decomposition of bulk RNA-seq data into cell-type fractions. By linking the cell fractions of samples and clinical follow-up information, we provide an innovative approach for the discovery of potential therapeutic targets. In the future, with the advent of more high-quality scRNA-seq atlases, deconvolution could be a powerful data mining tool to uncover the intricate nature of the TME of liver cancer, and reveal valuable information in the vast amount of available transcriptomic data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

Supplementary Materials: Figure S1: Cell type trees of Normal Liver atlas, TME-Stroma atlas and TME-Immune atlas; **Figure S2**: In silico experiments of Cibersortx estimation (Normal Liver atlas, intra-study validation mode). **A1-13**, cell type tree was defined by the authors; **B1-10**, cell subtypes were combined into major cell types; **Figure S3**: In silico experiments of Cibersortx estimation (TME-Stroma atlas, intra-study validation mode); **Figure S4**: In silico experiments of Cibersortx estimation (TME-Immune atlas, intra-study validation mode); **Figure S5**: In silico experiments of Cibersortx estimation (cross-study validation experiments); **Figure S6**: Deconvolution output of all the remaining cell type/subtypes not shown in the maintext; **Figure S7**: Impact of cell fraction on patient outcomes (TCGA-LIHC, Normal Liver atlas); **Figure S8**: Impact of cell fraction on patient outcomes (TCGA-LIHC, TME-Stroma atlas); **Figure S9**: Impact of cell fraction on patient outcomes (TCGA-LIHC, TME-Immune atlas); **Figure S10**: Impact of cell fraction on patient outcomes (GSE14520, Normal Liver atlas); **Figure S11**: Impact of cell fraction on patient outcomes (GSE14520, TME-Stroma atlas); **Figure S12**: Impact of cell fraction on patient outcomes (GSE14520, TME-Immune atlas); **Figure S13**: Gene set enrichment analysis of Cibersortx identified signature genes for the three atlases. **Table S1**: Summary of recruited studies; **Table S2**: Summary of *in silico* validation experiments; **Table S3**: Summary of survival analysis.

Author Contributions: Conceptualization, S.Z. and A.P.S.; methodology, S.Z. and A.P.S.; software, S.Z.; validation, S.Z.; formal analysis, S.Z.; investigation, S.Z. and A.P.S ; resources, S.Z. and A.P.S.; data curation, S.Z. and A.P.S.; writing—original draft preparation, S.Z. and A.P.S.; writing—review and editing, S.Z., W.B., F.K., and A.P.S.; visualization, S.Z.; supervision, M.P.P., F.K., and A.P.S.; project administration, A.P.S.; funding acquisition, M.P.P., F.K., and A.P.S.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was covered by the Human Subjects Protection and Data Access Policies. These are set of policies developed by NCI and NHGRI to protect the privacy of participants donating specimens to TCGA. Included are TCGA's informed consent policy, data access policy and information about HIPAA Privacy Rule compliance. January, 2014.

Data Availability Statement:

All data is publically available. The code to reproduce this work is available from our GitHub (https://github.com/ErasmusMC-Bioinformatics/OIO_Shaoshi).

Acknowledgement: We would like to thank Ju Jie, Willem de Koning, Teodora-Elena Trandafir, as well as all the Bioinformatics team, Erasmus MC.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **2021**, *71*, 209–249, doi:10.3322/CAAC.21660.
2. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2020. *CA. Cancer J. Clin.* **2020**, *70*, 7–30, doi:10.3322/caac.21590.
3. Jemal, A.; Ward, E.M.; Johnson, C.J.; Cronin, K.A.; Ma, J.; Ryerson, A.B.; Mariotto, A.; Lake, A.J.; Wilson, R.; Sherman, R.L.; et al. Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival. *JNCI J. Natl. Cancer Inst.* **2017**, *109*, doi:10.1093/jnci/djx030.
4. Llovet, J.M.; Kelley, R.K.; Villanueva, A.; Singal, A.G.; Pikarsky, E.; Roayaie, S.; Lencioni, R.; Koike, K.; Zucman-Rossi, J.; Finn, R.S. Hepatocellular Carcinoma. *Nat. Rev. Dis. Prim. 2021 71* **2021**, *7*, 1–28, doi:10.1038/s41572-020-00240-3.
5. Brindley, P.J.; Bachini, M.; Ilyas, S.I.; Khan, S.A.; Loukas, A.; Sirica, A.E.; Teh, B.T.; Wongkham, S.; Gores, G.J. Cholangiocarcinoma. *Nat. Rev. Dis. Prim. 2021 71* **2021**, *7*, 1–17, doi:10.1038/s41572-021-00300-2.
6. Deng, G.L.; Zeng, S.; Shen, H. Chemotherapy and Target Therapy for Hepatocellular Carcinoma: New Advances and Challenges. *World J. Hepatol.* **2015**, *7*, 787–798, doi:10.4254/WJH.V7.I5.787.
7. Morse, M.A.; Sun, W.; Kim, R.; He, A.R.; Abada, P.B.; Mynderse, M.; Finn, R.S. The Role of Angiogenesis in Hepatocellular Carcinoma. *Clin. Cancer Res.* **2019**, *25*, 912–920, doi:10.1158/1078-0432.CCR-18-1254/73039/AM/THE-ROLE-OF-ANGIOGENESIS-IN-HEPATOCELLULAR.
8. Simone, V.; Brunetti, O.; Lupo, L.; Testini, M.; Maiorano, E.; Simone, M.; Longo, V.; Rolfo, C.; Peeters, M.; Scarpa, A.; et al. Targeting Angiogenesis in Biliary Tract Cancers: An Open Option. *Int. J. Mol. Sci.* **2017**, *18*, doi:10.3390/IJMS18020418.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

9. Kaps, L.; Schuppan, D. Targeting Cancer Associated Fibroblasts in Liver Fibrosis and Liver Cancer Using Nanocarriers. *Cells* **2020**, *9*, doi:10.3390/CELLS9092027.
10. Sangro, B.; Sarobe, P.; Hervás-Stubbs, S.; Melero, I. Advances in Immunotherapy for Hepatocellular Carcinoma. *Nat. Rev. Gastroenterol. Hepatol. 2021 188* **2021**, *18*, 525–543, doi:10.1038/s41575-021-00438-0.
11. Fiste, O.; Ntanasis-Stathopoulos, I.; Gavriatopoulou, M.; Liontos, M.; Koutsoukos, K.; Dimopoulos, M.A.; Zagouri, F. The Emerging Role of Immunotherapy in Intrahepatic Cholangiocarcinoma. *Vaccines* **2021**, *9*, doi:10.3390/VACCINES9050422.
12. Jin, M.Z.; Jin, W.L. The Updated Landscape of Tumor Microenvironment and Drug Repurposing. *Signal Transduct. Target. Ther.* **2020**, *5*, doi:10.1038/S41392-020-00280-X.
13. Binnewies, M.; Roberts, E.W.; Kersten, K.; Chan, V.; Fearon, D.F.; Merad, M.; Coussens, L.M.; Gabrilovich, D.I.; Ostrand-Rosenberg, S.; Hedrick, C.C.; et al. Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy. *Nat. Med.* **2018**, *24*, 541–550, doi:10.1038/S41591-018-0014-X.
14. Zhang, Y.; Wang, D.; Peng, M.; Tang, L.; Ouyang, J.; Xiong, F.; Guo, C.; Tang, Y.; Zhou, Y.; Liao, Q.; et al. Single-Cell RNA Sequencing in Cancer Research. *J. Exp. Clin. Cancer Res.* **2021**, *40*, doi:10.1186/S13046-021-01874-1.
15. Cooper, L.A.D.; Demicco, E.G.; Saltz, J.H.; Powell, R.T.; Rao, A.; Lazar, A.J. PanCancer Insights from The Cancer Genome Atlas: The Pathologist's Perspective. *J. Pathol.* **2018**, *244*, 512–524, doi:10.1002/PATH.5028.
16. Malta, T.M.; Sokolov, A.; Gentles, A.J.; Burzykowski, T.; Poisson, L.; Weinstein, J.N.; Kamińska, B.; Huelsken, J.; Omberg, L.; Gevaert, O.; et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **2018**, *173*, 338-354.e15, doi:10.1016/j.cell.2018.03.034.
17. Wang, X.; Park, J.; Susztak, K.; Zhang, N.R.; Li, M. Bulk Tissue Cell Type Deconvolution with Multi-Subject Single-Cell Expression Reference. *Nat. Commun.* **2019**, *10*, doi:10.1038/s41467-018-08023-x.
18. Newman, A.M.; Steen, C.B.; Liu, C.L.; Gentles, A.J.; Chaudhuri, A.A.; Scherer, F.; Khodadoust, M.S.; Esfahani, M.S.; Luca, B.A.; Steiner, D.; et al. Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry. *Nat. Biotechnol.* **2019**, *37*, 773–782, doi:10.1038/S41587-019-0114-2.
19. Qi, Z.; Liu, Y.; Mints, M.; Mullins, R.; Sample, R.; Law, T.; Barrett, T.; Mazul, A.L.; Jackson, R.S.; Kang, S.Y.; et al. Single-Cell Deconvolution of Head and Neck Squamous Cell Carcinoma. *Cancers (Basel).* **2021**, *13*, 1–20, doi:10.3390/CANCERS13061230.
20. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15545–15550, doi:10.1073/PNAS.0506580102.
21. Piccolo, S.R.; Sun, Y.; Campbell, J.D.; Lenburg, M.E.; Bild, A.H.; Johnson, W.E. A Single-Sample Microarray Normalization Method to Facilitate Personalized-Medicine Workflows. *Genomics* **2012**, *100*, 337–344, doi:10.1016/J.YGENO.2012.08.003.
22. Piccolo, S.R.; Withers, M.R.; Francis, O.E.; Bild, A.H.; Johnson, W.E. Multiplatform Single-Sample Estimates of Transcriptional Activation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17778–17783, doi:10.1073/PNAS.1305823110.
23. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor Package BiomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191, doi:10.1038/nprot.2009.97.
24. Miller, J.A.; Cai, C.; Langfelder, P.; Geschwind, D.H.; Kurian, S.M.; Salomon, D.R.; Horvath, S. Strategies for Aggregating Gene Expression Data: The CollapseRows R Function. *BMC Bioinformatics* **2011**, *12*, doi:10.1186/1471-2105-12-322.
25. Wang, S.; Liu, X. The UCSCXenaTools R Package: A Toolkit for Accessing Genomics Data from UCSC Xena Platform, from Cancer Multi-Omics to Single-Cell RNA-Seq. *J. Open Source Softw.* **2019**, *4*, 1627, doi:10.21105/joss.01627.
26. MacParland, S.A.; Liu, J.C.; Ma, X.Z.; Innes, B.T.; Bartczak, A.M.; Gage, B.K.; Manuel, J.; Khuu, N.; Echeverri, J.; Linares, I.; et al. Single Cell RNA Sequencing of Human Liver Reveals Distinct Intrahepatic Macrophage Populations. *Nat. Commun.* **2018**, *9*, doi:10.1038/S41467-018-06318-7.
27. Massalha, H.; Bahar Halpern, K.; Abu-Gazala, S.; Jana, T.; Massasa, E.E.; Moor, A.E.; Buchauer, L.; Rozenberg, M.; Pikarsky, E.; Amit, I.; et al. A Single Cell Atlas of the Human Liver Tumor Microenvironment. *Mol. Syst. Biol.* **2020**, *16*, doi:10.15252/MSB.20209682.
28. Sharma, A.; Seow, J.J.W.; Dutertre, C.A.; Pai, R.; Blériot, C.; Mishra, A.; Wong, R.M.M.; Singh, G.S.N.; Sudhagar, S.; Khalilnezhad, S.; et al. Onco-Fetal Reprogramming of Endothelial Cells Drives Immunosuppressive Macrophages in Hepatocellular Carcinoma. *Cell* **2020**, *183*, 377-394.e21, doi:10.1016/J.CELL.2020.08.040.
29. Pölsterl, S. Scikit-Survival: A Library for Time-to-Event Analysis Built on Top of Scikit-Learn. *J. Mach. Learn. Res.* **2020**, *21*, 1–6.
30. Davidson-Pilon, C. Lifelines: Survival Analysis in Python. *J. Open Source Softw.* **2019**, *4*, 1317, doi:10.21105/joss.01317.
31. June, S.L.; Semela, D.; Iredale, J.; Shah, V.H. Sinusoidal Remodeling and Angiogenesis: A New Function for the Liver-Specific Pericyte? *Hepatology* **2007**, *45*, 817–825, doi:10.1002/HEP.21564.
32. Tacke, F.; Zimmermann, H.W. Macrophage Heterogeneity in Liver Injury and Fibrosis. *J. Hepatol.* **2014**, *60*, 1090–1096, doi:10.1016/J.JHEP.2013.12.025.
33. Lurje, I.; Hammerich, L.; Tacke, F. Dendritic Cell and T Cell Crosstalk in Liver Fibrogenesis and Hepatocarcinogenesis: Implications for Prevention and Therapy of Liver Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 1–25, doi:10.3390/IJMS21197378.
34. Ally, A.; Balasundaram, M.; Carlsen, R.; Chuah, E.; Clarke, A.; Dhalla, N.; Holt, R.A.; Jones, S.J.M.; Lee, D.; Ma, Y.; et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **2017**, *169*, 1327-1341.e23, doi:10.1016/J.CELL.2017.05.046.
35. Roessler, S.; Jia, H.L.; Budhu, A.; Forgues, M.; Ye, Q.H.; Lee, J.S.; Thorgeirsson, S.S.; Sun, Z.; Tang, Z.Y.; Qin, L.X.; et al. A Unique Metastasis Gene Signature Enables Prediction of Tumor Relapse in Early-Stage Hepatocellular Carcinoma Patients. *Cancer Res.* **2010**, *70*, 10202–10212, doi:10.1158/0008-5472.CAN-10-2607.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 November 2022 doi:10.20944/preprints202211.0344.v1

36. Roessler, S.; Long, E.L.; Budhu, A.; Chen, Y.; Zhao, X.; Ji, J.; Walker, R.; Jia, H.L.; Ye, Q.H.; Qin, L.X.; et al. Integrative Genomic Identification of Genes on 8p Associated with Hepatocellular Carcinoma Progression and Patient Survival. *Gastroenterology* **2012**, *142*, doi:10.1053/J.GASTRO.2011.12.039.
37. Blouin, A.; Bolender, R.P.; Weibel, E.R. Distribution of Organelles and Membranes between Hepatocytes and Nonhepatocytes in the Rat Liver Parenchyma. A Stereological Study. *J. Cell Biol.* **1977**, *72*, 441–455, doi:10.1083/jcb.72.2.441.
38. Sangro, B.; Sarobe, P.; Hervás-Stubbs, S.; Melero, I. Advances in Immunotherapy for Hepatocellular Carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **2021**, *18*, 525–543, doi:10.1038/S41575-021-00438-0.
39. Plitas, G.; Rudensky, A.Y. Regulatory T Cells in Cancer. *https://doi.org/10.1146/annurev-cancerbio-030419-033428* **2020**, *4*, 459–477, doi:10.1146/ANNUREV-CANCERBIO-030419-033428.
40. Yamashita, T.; Ji, J.; Budhu, A.; Forgues, M.; Yang, W.; Wang, H.Y.; Jia, H.; Ye, Q.; Qin, L.X.; Wauthier, E.; et al. EpCAM-Positive Hepatocellular Carcinoma Cells Are Tumor-Initiating Cells With Stem/Progenitor Cell Features. *Gastroenterology* **2009**, *136*, doi:10.1053/j.gastro.2008.12.004.
41. Yoon, S.M.; Gerasimidou, D.; Kuwahara, R.; Hytiroglou, P.; Yoo, J.E.; Park, Y.N.; Theise, N.D. Epithelial Cell Adhesion Molecule (EpCAM) Marks Hepatocytes Newly Derived from Stem/Progenitor Cells in Humans. *Hepatology* **2011**, *53*, 964–973, doi:10.1002/HEP.24122.