Single Cell and Plasma RNA Sequencing for RNA Liquid Biopsy for Hepatocellular Carcinoma

Joaquim S.L. Vong, Lu Ji, Macy M.S. Heung, Suk Hang Cheng, John Wong, Paul B.S. Lai, Vincent W.S. Wong, Stephen L. Chan, Henry L.Y. Chan, Peiyong Jiang, K.C. Allen Chan, Rossa W.K. Chiu and Y.M. Dennis Lo

BACKGROUND: Human plasma contains RNA transcripts released by multiple cell types within the body. Single-cell transcriptomic analysis allows the cellular origin of circulating RNA molecules to be elucidated at high resolution and has been successfully utilized in the pregnancy context. We explored the application of a similar approach to develop plasma RNA markers for cancer detection.

METHODS: Single-cell RNA sequencing was performed to decipher transcriptomic profiles of single cells from hepatocellular carcinoma (HCC) samples. Cell-type-specific transcripts were identified and used for deducing the cell-type-specific gene signature (CELSIG) scores of plasma RNA from patients with and without HCC.

RESULTS: Six major cell clusters were identified, including hepatocyte-like, cholangiocyte-like, myofibroblast, endothelial, lymphoid, and myeloid cell clusters based on HCC tumor tissues as well as their paired adjacent nontumoral tissues. The CELSIG score of hepatocyte-like cells was significantly increased in preoperative plasma RNA samples of patients with HCC ($n = 14$) compared with non-HCC participants ($n = 49$). The CELSIG score of hepatocyte-like cells declined in plasma RNA samples of patients with HCC within 3 days after tumor resection. Compared with the discriminating power between patients with and without HCC using the abundance of ALB transcript in plasma [area under curve (AUC) 0.72], an improved performance (AUC: 0.84) was observed using the CELSIG score. The hepatocyte-specific transcript markers in plasma RNA were further validated by ddPCR assays.

The CELSIG scores of hepatocyte-like cell and cholangiocyte trended with patients’ survival.

CONCLUSIONS: The combination of single-cell transcriptomic analysis and plasma RNA sequencing represents an approach for the development of new noninvasive cancer markers.

Introduction

The genetics and epigenetics of circulating cell-free DNA are being actively studied, leading to numerous novel diagnostic approaches for noninvasive prenatal testing and cancer detection (1–9). This field can be generally referred to as “liquid biopsies.” On the other hand, there is much less interest being paid to the diagnostic and research uses of plasma transcriptome, with early work directed to the development of noninvasive prenatal markers (10, 11). Tsui et al. demonstrated that plasma RNA sequencing could provide a holistic view of the maternal plasma transcriptomic repertoire, revealing a set of pregnancy-associated transcripts as well as the relative contributions from the placenta and maternal sources (12). Another study further confirmed that the use of plasma transcriptome would allow the obtaining of the relative transcriptional contributions to plasma RNA pool from different tissues (13). The plasma transcriptome thus possesses a wealth of information related to the tissues of origin.

A number of challenges, however, still exist in exploiting plasma RNA for diagnostic purposes. First,...
the quantities of tissue-specific transcripts in plasma are highly variable (12), which could be attributed by the variable cellular compositions of different cell types in that tissue, hampering the discovery of a set of robust transcript biomarkers. Second, many organ-specific transcripts derived from a bulk tissue might not be specifically associated with certain diseases. For example, the liver-specific ALB messenger RNA (mRNA) molecules in the plasma were reported to be increased both in patients with hepatocellular carcinoma (HCC) and those with chronic hepatitis B virus (HBV) infection, compared with healthy control participants (14).

A bulk tissue consists of a number of cell types. If one could dissect the cell-type-specific transcripts and determine the correlations between those transcripts and diseases, it might improve the signal-to-noise ratio regarding the targeted RNA molecules in plasma. In a previous study, we combined single-cell RNA sequencing and cell-free RNA sequencing for tracing the placental cellular dynamics during pregnancy in a noninvasive manner (15). More importantly, the expression of transcripts associated with extravillous trophoblasts was significantly higher in pregnant women affected by preeclampsia than unaffected pregnant women. On the other hand, the expression of transcripts associated with other cell types in the placenta such as syncytiotrophoblasts, decidual cells, and endothelial cells appeared to exhibit no significant difference between these 2 groups (15). Thus, based on single-cell RNA sequencing technology, the dissection of cell-type-specific transcripts in the plasma would enhance the noninvasive detection of diseases.

In this study, we explored the feasibility of using cell-type-specific transcript signals in the plasma RNA pool to inform the presence of cancer. We used HCC, the most common type of liver cancer, as a model. Cell-type-specific transcripts in plasma were determined by single-cell sequencing of HCC tumor and paired normal tissues. On the basis of plasma RNA sequencing of patients with HCC, the plasma RNA molecules contributed by such cell-type-specific transcripts were investigated.

### Materials and Methods

#### PARTICIPANTS, CLINICAL INFORMATION COLLECTION, AND SAMPLE COLLECTION

This study was approved by the Joint Chinese University of Hong Kong—Hospital Authority New Territories East Clinical Research Ethics Committee. Patients with HCC (n = 14) were recruited, 20 mL of peripheral blood was drawn into EDTA-anticoagulated blood tubes before (preoperative samples) and after (postoperative samples) surgical resections. Seven of the 14 patients with HCC had consented to the collection of postsurgical samples. Postoperative blood samples (n = 7) were taken within 72 h after tumor resection. For healthy controls (n = 8), participants with chronic HBV infection with cirrhosis (n = 23) and without cirrhosis (n = 18), 20 mL of peripheral blood from each participant was drawn into EDTA-anticoagulated tubes. For survival data analysis, day 0 was defined as the day of surgical resection. Details about the participants are shown in Table 1 in the online Data Supplement.

We obtained the independent tumoral tissue samples (n = 4) and the paired adjacent nontumoral tissue samples (n = 4) for single-cell RNA sequencing on the basis of the frozen-section from the Department of Surgery of The Chinese University of Hong Kong at the

### Table 1. Seven representative genes associated with the hepatocyte-like cell gene signatures.

| Gene Symbol | Hepatocyte-like (UMI/cell) | Cholangiocyte-like (UMI/cell) | Myofibroblast-like (UMI/cell) | Endothelial (UMI/cell) | Lymphoid (UMI/cell) | Myeloid (UMI/cell) | Full gene name | Biological process |
|-------------|-----------------------------|--------------------------------|-------------------------------|-----------------------|-------------------|-------------------|----------------|-------------------|
| TF          | 6.51                        | 0.17                           | 0.05                          | 0.07                  | 0.02              | 0.05              | Transferrin     | Blood plasma iron level control |
| GSTA1       | 6.03                        | 0.05                           | 0.04                          | 0.13                  | 0.02              | 0.05              | Glutathione S-transferase alpha 1 | Detoxification |
| ORM2        | 5.9                         | 0.10                           | 0.05                          | 0.07                  | 0.02              | 0.04              | Orosomucoid 2   | Acute phase glycoprotein |
| FGL1        | 5.65                        | 0.19                           | 0.07                          | 0.06                  | 0.02              | 0.05              | Fibrinogen-like 1 | Blood clotting |
| KNG1        | 5.51                        | 0.14                           | 0.05                          | 0.07                  | 0.01              | 0.03              | Kininogen 1     | Blood coagulation |
| ALOC18      | 5.34                        | 0.09                           | 0.10                          | 0.10                  | 0.03              | 0.05              | Aldolase fructose-bisphosphate B | Glycolysis, gluconeogenesis |
| HPD         | 1.03                        | 0.15                           | 0.02                          | 0.03                  | 0.01              | 0.02              | 4-Hydroxyphenylpyruvate dioxygenase | Blood tyrosine level regulation |
HCC TUMOR AND PAIRED ADJACENT NORMAL TISSUES

We used high-throughput droplet-based single-cell RNA sequencing (10x Genomics) to profile a total of 17 176 cells from 4 patients with HCC, consisting of 5782 cells originating from HCC tumor samples and 11 394 cells from paired normal liver samples, based on the analytical framework described previously (15). Briefly, the dissociated single cells were individually compartmentalized into each droplet. The transcripts served as templates for generating complementary DNA by barcoded primers comprising an oligo (dT) sequence, a 10-nucleotide (nt) molecular index for uniquely determining each mRNA strand, 16-nt barcode unique to each cell, and a universal sequence. All droplets were pooled together and broken to collect barcoded complementary DNA molecules for sequencing. The sequencing reads were traced back to the original cells accordingly to the barcode sequences, allowing for determining cell types and identifying cell-type-specific gene signatures (Fig. 1, A).

The median number of analyzable cells per sample was 1903 (range: 391–5400). The median number of detectable transcripts per cell was 6340 (range: 2122–16 812), corresponding to a median of 1718 genes per cell (range: 750–2898) (Supplemental Table 1).
Fig. 1. Cell types and cell-type-specific gene signatures identified in the human hepatocellular cancer (HCC) tumors and paired normal tissues based on single-cell RNA sequencing. (A), Schematic diagram of workflow for single-cell RNA sequencing. (B), Visualization of cell types for 17 176 HCC and adjacent nontumor liver cells using t-distributed stochastic neighbor embedding (t-SNE). (C), Expression patterns of those representative cell-type-specific gene signatures. (D), Heatmap plot for the expression of the cell-type-specific gene signatures.
We performed clustering analysis by $t$-stochastic neighborhood embedding ($t$-SNE) (15). Cells were clustered into different groups according to the similarities of transcriptomic profile between cells. As shown in the $t$-SNE plot (Fig. 1, B), each dot indicated one cell. The $t$-SNE analysis displayed 6 major cell clusters (Fig. 1, B). Cluster C1 was transcriptionally defined as hepatocyte-like cells with the expression of liver-specific genes such as $ALB$, $F2$, $AP0A2$, and $HP$; cluster C2 was defined as cholangiocyte-like cells, with the expression of cytokeratins (KRT7, KRT19), claudins (CLDN4, CLDN10) and MMP7; cluster C3 was deemed myofibroblast-like cells with the expression of genes specific to hepatic stellate cells such as $ACTA2$, and collagen-related genes (COL1A2, COL3A1); cluster C4 corresponded to endothelial cells with the expression of $TEK$, $DNASE1L3$, $CLEC4G$, and $FCN3$; cluster C5 represented lymphoid cells expressing $CD3D$, $CST7$, $GZMA$, and $NKG7$; and cluster C6 would be myeloid cells expressing $CD163$, $CD86$, $AIFI$, $S100A8$, $S100A9$, and $MARCO$ (Fig. 1, B). Figure 1, C shows the expression level of a representative marker of each cluster in the $t$-SNE map. The full list of cell-type-specific genes is shown in Supplemental Table 3.

**CELL-TYPE-SPECIFIC GENE SIGNATURES**

We attempted to identify the cell-type-specific gene signatures that were preferentially expressed in a particular cell cluster. The gene-expression z-score was used to measure the expression preference of a transcript in a particular cell type. The z-score quantified a transcript expression level in a cluster, equivalent to the number of standard deviations from the mean of the remaining clusters. A z-score cutoff of 3 was used for defining the cell-type-specific transcript. In total, we identified 39, 15, 21, 20, 19, and 20 cell-type-specific gene signatures from hepatocyte-like, cholangiocyte-like, myofibroblast-like, endothelial, lymphoid, and myeloid cells, respectively (Fig. 1, D). For example, a number of previously reported liver-specific transcripts such as $TF$, $ALDOB$, and $F2$ were overexpressed in the hepatocyte-like cells. In addition, we identified numerous transcripts specific to the hepatocyte-like cells, including $KNG1$, $FGL1$, and $HPD$. Among them, $FGL1$ (fibrinogen-like-1) gene expression was very recently reported to be higher in HCC tissues than in adjacent normal tissues (19), but $FGL1$ had not yet been well studied in the plasma RNA pool.

**PRESENTATION OF CELL-TYPE-SPECIFIC GENE SIGNATURES IN PLASMA RNA POOL**

We further studied whether the use of cell-free-specific gene signatures in plasma would be able to inform the presence of cancer. Based on plasma RNA sequencing results, CELSIG scores were calculated for healthy controls ($n = 8$), patients with HBV infection with ($n = 23$) cirrhosis and without ($n = 18$) cirrhosis (i.e., HBV carriers), and preoperative patients with HCC ($n = 14$), with 7 paired postoperative patients collected within 72 h after the tumor surgery. The CELSIG score represented the transcriptional contributions to the plasma RNA pool from different cell clusters.

A median of 185 million sequencing reads (range: 103–255 million) were obtained. Interestingly, the plasma CELSIG score related to hepatocyte-like cells was significantly increased in patients with HCC (median: 3.60; range 0.90–6.97) compared to the healthy controls and HBV carriers with and without cirrhosis (median: 1.54; range 0.28–7.24) ($P = 0.0002$; Mann–Whitney $U$-test) (Fig. 2, A). The plasma CELSIG score related to hepatocyte-like cells quickly declined within 72 h after surgical tumoral resection (median: 1.99; range 0.84–3.41) ($P = 0.03$; Mann–Whitney $U$-test) (Fig. 2, A). In contrast, no such changing patterns of the CELSIG score were observed in other cell clusters including cholangiocyte-like, myofibroblast-like, endothelial, lymphoid, and myeloid cell clusters (Supplemental Fig. 1, A). Furthermore, the CELSIG score of lymphoid cells was significantly decreased ($P = 0.006$; Mann–Whitney $U$-test) in patients with HCC (median: 7.47; range 5.28–9.69), compared with non-HCC groups (median: 8.46; range: 5.55–9.98) (Fig. 2, B). However, the CELSIG score of lymphoid cells paradoxically appeared to be further decreased in postoperative patients, perhaps suggesting the surgical removal of liver cancer would trigger the release of confounding transcripts from other cell types to the plasma RNA pool.

To further study whether the transcript signatures of hepatocyte-like cells were mainly derived from the platelets, we obtained the expression levels of 39 signatures of hepatocyte-like cells in platelets for healthy control participants from a published study (20) and the normal liver tissues from transcriptomic BodyMap (accession number: GSE30611; Gene Expression Omnibus database). We found that the expression levels of transcript signatures of hepatocyte-like cells were much higher in the liver tissues [median fragments per kilobase per million mapped fragments (FPKM): 375.2; IQR: 194.9–857.9], in comparison with platelets (median FPKM: 0.006; IQR: 0–0.07). These data indicated that there was a low probability for those signature signals of hepatocyte-like cells mainly originating from platelets.

**VALIDATION OF HEPATOCYTE-LIKE GENE SIGNATURES PRESENT IN PLASMA USING DDPCR**

Figure 3, A and Supplemental Fig. 1, B show the expression levels in plasma RNA for representative transcripts specific to hepatocyte-like cells. Such hepatocyte-like cell specific transcripts were found to be generally increased in the plasma RNA of patients with HCC.
compared with non-HCC participants, whose expression levels were decreased after the surgical removal of HCC tumors (Fig. 3, A and Supplemental Fig. 1, B). For example, the FGL1 gene displayed nearly no expression in plasma of HBV carriers with and without cirrhosis (mean expression level in FPKM: 0.90; range: 0–14.78) (Fig. 3, A), according to the plasma RNA sequencing results. The FGL1 expression level was sharply increased in plasma of preoperative patients with HCC (mean expression level: 4.92; range: 0–29.59), rapidly declining after the surgical resection of HCC tumors (mean expression level: 0.50; range: 0–3.32) (Fig. 3, A). In contrast, the pattern of the ALB gene expression appeared to be less specific to the preoperative patients with HCC (Fig. 3, B).

To orthogonally validate whether the cell-type-specific gene signatures would be robustly detectable in the plasma RNA pool, we designed ddPCR assays to quantify the expression levels for 7 representative genes (TF, GSTA1, ORM2, FGL1, KNG1, ALDOB, HPD) out of the 39 hepatocyte-like specific genes (Table 1). The gene expressions were normalized to the reference gene ACTB in plasma (Supplemental Fig. 2). The ACTB gene expression was chosen as the reference gene in this study because its expression level was shown to be relatively high (median FPKM: 1830; range: 933–3027), with the smallest variation across different participants [coefficient of variation (CV): 24%] when compared to 8 genes with a comparable expression level in plasma relative to the ACTB (median CV: 44%). These genes included RPS11, SH3BGR1L3, PPBP, RPS27, HBB, AC138811, AC006064, and AL138963. All 7 transcripts specific to hepatocyte-like cells showed generally consistent expression patterns in plasma RNA across groups measured by ddPCR, when compared with plasma RNA sequencing results (Supplemental Fig. 2). Figure 3, C and D shows that the patterns of the FGL1 and ALB gene expressions in plasma were confirmed by the corresponding ddPCR assays. The dot plots of FGL1 and ALB gene expressions for individual patients across groups are shown in Supplemental Fig. 3, A and B.

**CLINICAL IMPLICATIONS OF CELL-TYPE-SPECIFIC GENE SIGNATURES**

We further investigated the potential clinical applications of using those cell-type-specific gene signatures identified from this study. Because we observed that the CELSIG score of hepatocyte-like cells was increased in patients with HCC whereas the CELSIG score of
lymphoid cells was decreased, the ratio of CELSIG score of hepatocyte-like cells to lymphoid cells (i.e., H/L ratio) was used for studying the clinical potential of HCC detection. Receiver operating characteristic curve analysis revealed that H/L ratio achieved an area under the receiver operating characteristic curve (AUC) of 0.84 for differentiating between patients with and without HCC, suggesting feasibility in differentiating between patients with and without HCC. H/L ratio-based analysis appeared to be superior to that based on plasma ALB expression level (AUC: 0.72) \( (P = 0.03; \text{DeLong test}) \) (Fig. 4, A). More importantly, when differentiating between HCC patients and patients with chronic HBV infection, H/L ratio resulted in an AUC of 0.82 that was higher than the ALB expression-based analysis (AUC: 0.68) \( (P = 0.02; \text{DeLong test}) \) (Fig. 4, B). These results indicated that the combined analysis of single-cell and plasma RNA sequencing data might be superior to analysis based on a single plasma RNA biomarker.
As the CELSIG score reflected the transcriptional contribution of a cell type to the plasma RNA pool, it might be correlated with the turnover of cell death of particular tissues/organs that would contribute RNA transcripts to the blood circulation. We envisioned that the CELSIG score of a cell type might correlate with disease outcomes. Thus, we analyzed the correlation between the patient survival data and the CELSIG score of each cell type. We divided the preoperative patients with HCC ($n = 14$) into 2 groups based on a CELSIG score below or above the median value. The H/L ratio appeared to be correlated with survival data (Fig. 4, C) (i.e., the higher the H/L ratio, the shorter the survival time). However, H/L ratio analysis did not achieve statistical significance ($P = 0.15$; Gehan–Breslow–Wilcoxon tests) (Fig. 4, C), perhaps because of the small sample size available for this analysis. Of note, we observed a significantly longer survival time for patients with lower CELSIG scores of cholangiocyte-like cells than those with higher scores ($P = 0.03$; Gehan–Breslow–Wilcoxon tests) (Fig. 4, D).

**Fig. 4.** Clinical implication of CELSIG analysis for plasma RNA. (A), ROC analysis for CELSIG ratio of hepatocyte-like cells to lymphoid cells (H/L ratio) and the $\text{ALB}$ mRNA abundance between participants with and without HCC. (B), ROC analysis for the H/L ratio and the $\text{ALB}$ mRNA abundance between participants with HCC and participants with HBV infection. (C), Survival analysis based on the H/L ratio. (D), Survival analysis based CELSIG of cholangiocyte-like cells.
Discussion

Here we demonstrate that the plasma RNA transcriptomic profiling in combination with single-cell RNA sequencing can be used for the development of noninvasive tumor markers. Single-cell RNA sequencing enabled a high-resolution mapping of cell types, thus allowing the identification of those disease-associated cell-type-specific gene signatures that were preferentially present in the plasma RNA pool.

This proof-of-concept study illustrated that the analysis of multiple plasma transcript signals of hepatocyte-like cells (i.e., CELSIG score) led to a better power of differentiating patients with and without HCC, compared with a single plasma transcript marker. One possible reason may be that a single transcript marker may not be universally present in all HCC tumors, resulting in a poor clinical sensitivity, as HCC tumors were reported to be highly heterogeneous (21–23). Another possible reason may be that a single transcript marker is expressed in multiple cell types, resulting in a decrease of clinical specificity. In contrast, the use of multiple transcript markers specific to a cell type would offer a new possibility to not only increase the sensitivity with more transcript signals available, but also improve the clinical specificity because of the removal of confounding transcript signals from other cell types.

Among the 39 transcripts identified to be specific to hepatocyte-like cells, the FGL1 (fibrinogen-like 1) gene expression level quantified by plasma RNA sequencing was elevated in patients with HCC, showing a potential for further development as a biomarker. The FGL1 was recently reported to be transcriptionally upregulated in HCC tissues and the HCC cell line (HepG2) (24). Furthermore, there was a significant increase of FGL1-positive cells by multiplex immunofluorescence staining in HCC tumors compared with the adjacent nontumoral tissues (19), and FGL1 was suggested as a prognostic marker. Therefore, the circulating FGL1 mRNA concentration may be a useful marker for diagnosis and prognosis.

Because the sample size was relatively small in this proof-of-concept study, a validation study with a large sample size would be warranted. It would be of interest to design an assay for plasma RNA targeted sequencing based on those 39 hepatocyte-like signatures identified in this work, exploring a cost-effective approach for HCC detection.

The ratio of the CELSIG score of hepatocyte-like cells to lymphoid cells (i.e., the H/L ratio) demonstrated a trend of correlation with overall survival, although the results did not achieve statistical significance. On the other hand, we observed that patients with HCC who had lower CELSIG scores of cholangiocyte-like cells exhibited a longer overall survival time. It has been reported that HCC tumor tissues can acquire the characteristic of positivity of cholangiocyte markers during the process of tumor progression, showing a higher positivity of cholangiocyte markers in larger-sized and more poorly differentiated tumors (25, 26). We speculated that one possible mechanism may be that the release of mRNA from dying transdifferentiated cholangiocytes from hepatocytes during focal damage repair in HCC.

It has also been reported that RNA profiles of blood platelets derived from patients with cancer (i.e., tumor-educated platelets) are distinct from platelets of healthy individuals (20). However, in this study, there was a low probability of those signature signals of hepatocyte-like cells present in plasma that were contributed by platelets, as the expression levels of those hepatocyte-derived signatures were much lower in platelets in compared with liver tissues.

Those tissue-specific RNA transcripts that are readily detectable in plasma might be attributed to the potential protection of RNA species from degradation in blood circulation. For example, certain plasma RNAs would be contained in extracellular membrane vesicles (27), and thus might be protected from degradation. Yao et al. reported that many mRNA fragments in plasma were associated with protein-binding sites that could render such RNA species resistant to plasma nucleases (28). Whether such RNA protection might be associated with tissue specificity will be an interesting avenue for future research. Since a human genome could encode more than 80,000 mRNA transcripts (The GENCODE Project) (29), i.e., one order of magnitude more pervasive than mature micro RNAs (miRNAs) (approximately 2600) (30). Thus, theoretically, the information available for the analysis based on plasma mRNA transcripts would be 30 times more than that based on mature miRNAs. On the other hand, the miRNA species would be commonly bound by the argonaute2 protein. Such miRNA–argonaute2 complexes might specifically stabilize plasma miRNAs (31). The potential lengthening of the half-life of the miRNAs in circulation could be one possible advantage for biomarker development.

In summary, this proof-of-concept study has demonstrated an integrative analysis of single-cell and plasma RNA sequencing data, shedding light on an approach for the development of new cancer markers.

Nonstandard Abbreviations: HCC, hepatocellular carcinoma; CELSIG, cell-type-specific gene signature; HBV, hepatitis B virus; ddPCR, droplet digital PCR; t-SNE, t-stochastic neighborhood embedding

Human Genes: ALB, albumin; F2, coagulation factor II, thrombin; APOA2, apolipoprotein A2; HP, haptoglobin; KRT7, keratin 7; KRT19, keratin 19; CLDN4, claudin 4; CLDN10, claudin 10;
MMP7, matrix metalloproteinase 7; ACTA2, actin alpha 2, smooth muscle; COL1A2, collagen type I alpha 2 chain; COL3A1, collagen type III alpha 1 chain; TEK, TEK receptor tyrosine kinase; DNASE1L3, deoxyribonuclease 1 like 3; CLEC4G, C-type lectin domain family 4 member G; FCN3, ficollin 3; CD3D, CD3d molecule; CS77, cystatin F; GZMA, granzyme A; NKG7, natural killer cell granule protein 7; CD163, CD163 molecule; CD86, CD86 molecule; AIF1, allograft inflammatory factor 1; S100A8, S100 calcium binding protein A8; S100A9, S100 calcium binding protein A9; MARCO, macrophage receptor with collagenous structure; TF, transferrin; ALDOB, aldolase, fructose-bisphosphate B, F2, coagulation factor II, thrombin; KNG1, kininogen 1; FGL1, fibrinogen-like 1; HPD, 4-hydroxyphenylpyruvate dioxygenase; GATA1, glutathione S-transferase alpha 1; ORM2, orosomucoid 2; ACTB, actin beta; RPS11, ribosomal protein S11; SH3BGR1L3, SH3 domain binding glutamate rich protein like 3; PPRB, pro-platelet basic protein; RPS27, ribosomal protein S27; HBB, hemoglobin subunit beta; AC138811; AC006064; AL138963.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

J.S.L. Vong, statistical analysis; L. Ji, statistical analysis; M.M.S. Heung, administrative support; J. Wong, provision of study material or patients; P.B.S. Lai, provision of study material or patients; V.W.S. Wong, administrative support, provision of study material or patients; S.L. Chan, provision of study material or patients; H.L.Y. Chan, provision of study material or patients; P. Jiang, statistical analysis; K.C.A. Chan, statistical analysis.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest: Funding or sponsorship: This work was supported by the Research Grants Council of the Hong Kong SAR Government under the Theme-based research scheme (T12-401/16-W), a collaborative research agreement from Grail, the Innovation and Technology Fund under the ImmunoHK Initiative and the Vice Chancellor’s One-Off Discretionary Fund of The Chinese University of Hong Kong (VCF2014021). Y.M.D. Lo is supported by an endowed chair from the Li Ka Shing Foundation. K.C.A. Chan, R.W.K. Chu, and Y.M.D. Lo receive research funding from Grail.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: We thank Mr. Kam-Wing Chan, Ms. Irene Tse, and Mr. Pete Tse for their assistance and helpful discussions. We thank Dr. Jason C.H. Tsang and Dr. Gary J.W. Liao for generating some of the exploratory data during the initial stage of this project.

References

1. Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci USA 2008;105:20458-63.
2. Fan HC, Blumenfeld YJ, Chikira U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA 2008;105:16246-71.
3. Chan KCA, Jiang P, Chan CWM, Sun K, Wong J, Hui EP, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci USA 2013;110:18761-8.
4. Dawson SJ, Tsui DWK, Murtaza M, Biggs H, Rueda OM, Chen SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013;368:1199-209.
5. Yu SCY, Chan KCA, Zheng YW, Jiang P, Loo GJW, Sun H, et al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. Proc Natl Acad Sci USA 2014;111:8583-8.
6. Jiang P, Chan CWM, Chan KCA, Cheng SH, Wong J, Wong WKS, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci USA 2015;112:1317-25.
7. Moulere F, Chandrananda D, Piskorz AM, Moore KE, Morris J, Allibone LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med 2018;10:16266-71.
8. Liu MC, O’Neill GR, Klein EA, Swanton C, Smith M, Richards D, et al., CGGA Consortium. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 2020;31:745-59.
9. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, epigenomics, and topology of cell-free DNA in liquid biopsies. Science 2021;372:eaw23616.
10. Tsui NBY, Chiu SSC, Chiu RWK, Lau TK, Ng EKO, Leung TN, et al. Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. J Med Genet 2004;41:461-7.
11. Lo YMD, Tsui NBY, Chiu RWK, Lau TK, Leung TN, Heung MMS, et al. Plasma placental mRNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat Med 2007;13:218-23.
12. Tsui NBY, Jiang P, Wong YF, Leung TY, Chan KCA, Chiu RWK, et al. Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. Clin Chem 2014;60:954-62.
13. Koh K, Pan W, Gauw C, Fan HC, Kerchner GA, Wys-Cosar Y, et al. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. Proc Natl Acad Sci USA 2014;111:7361-6.
14. Chan RWK, Wong J, Chan HY, Mo KSK, Lo WW, Lee V, et al. Aberrant concentrations of liver-derived plasma albumin mRNA in liver pathologies. Clin Chem 2010;56:82-9.
15. Tsang JCH, Vong JSL, Li P, Poon ECY, Jiang P, Liu KO, et al. Integrative single-cell and cell-free plasma RNA transcription elucidates placental cellular dynamics. Proc Natl Acad Sci USA 2017;114:9786-95.
16. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Luu FMF, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010;2:61ra91.
17. European Genome-Phenome Archive, European Bioinformatics Institute. https://ega.ebi.ac.uk/ega/ (Accessed 2020 November).
18. The Comprehensive R Archive Network. https://cran.r-project.org/(Accessed 2020 November).

19. Guo M, Yuan F, Qi F, Sun J, Rao Q, Zhao Z, et al. Expression and clinical significance of LAG-3, FGL1, PD-L1 and CD8. J Transl Med 2020;18:306.

20. Best MG, Soi N, Kooi I, Tamourn J, Westerman BA, Rustenburg F, et al. RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. Cancer Cell 2015;28:666–76.

21. Yu CB, Zhu LY, Wang YG, Li F, Zhang XY, Dai WJ. Systemic transcriptome analysis of hepatocellular carcinoma. Tumour Biol 2016;37:13323-31.

22. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017;169:1327-41.

23. Ho DW, Tsui YM, Sze KM, Chan LK, Cheung TT, Lee E, et al. Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer. Cancer Lett 2019;459:176-85.

24. Sayeed A, Dalvano BE, Kaplan DE, Viswanathan U, Kulip J, Janneh AH, et al. Profiling the circulating mRNA transcriptome in human liver disease. Oncotarget 2020;11:2216-32.

25. Shibuya M, Kondo F, Sano K, Takada T, Asano T. Immunohistochemical study of hepatocyte, cholangiocyte and stem cell markers of hepatocellular carcinoma. J Hepatobiliary Pancreat Sci 2011;18:537-43.

26. Kumagai A, Kondo F, Sano K, Inoue M, Fujii T, Hashimoto M, et al. Immunohistochemical study of hepatocyte, cholangiocyte and stem cell markers of hepatocellular carcinoma: the second report: relationship with tumor size and cell differentiation. J Hepatobiliary Pancreat Sci 2016;23:414-21.

27. Gutiérrez García G, Galicia García G, Zalapa Soto J, Izquierdo Medina A, Rotangen-Rodríguez M, Casas Aguilar GA, et al. Analysis of RNA yield in extracellular vesicles isolated by membrane affinity column and differential ultracentrifugation. PLoS ONE 2020;15:e0238545.

28. Yao J, Wu DC, Nottingham RM, Lambowitz AM. Identification of protein-protected mRNA fragments and structured excised intron RNAs in human plasma by TGIRT-seq peak calling. eLife 2020;9:e60743.

29. The GENCODE Project. https://www.gencodegenes.org/ (Accessed 2020 November).

30. Plotnikova O, Baranova A, Skoblov M. Comprehensive analysis of human microRNA-mRNA interactome. Front Genet 2019;10:933.

31. Arroyo JD, Chevillet JR, Kosh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci USA 2011;108:5003-8.