

An RNA-based signature enables high specificity detection of circulating tumor cells in hepatocellular carcinoma

Mark Kalinich^{a,1}, Irun Bhan^{a,b,1}, Tanya T. Kwan^a, David T. Miyamoto^{a,c}, Sarah Javaid^{a,2}, Joseph A. LiCausi^a, John D. Milner^a, Xin Hong^a, Lipika Goyal^{a,d}, Srinjoy Sil^a, Melissa Choz^a, Uyen Ho^a, Ravi Kapur^e, Alona Muzikansky^{a,f}, Huidan Zhang^g, David A. Weitz^g, Lecia V. Sequist^{a,d}, David P. Ryan^{a,d}, Raymond T. Chung^b, Andrew X. Zhu^{a,d}, Kurt J. Isselbacher^{a,b,3}, David T. Ting^{a,d}, Mehmet Toner^{e,h}, Shyamala Maheswaran^{a,h,3}, and Daniel A. Haber^{a,d,i,3}

^aMassachusetts General Hospital Cancer Center, Harvard Medical School, Charlestown, MA 02129; ^bDivision of Gastroenterology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ^cDepartment of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ^dDivision of Hematology Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ^eCenter for Bioengineering in Medicine, Massachusetts General Hospital and Harvard Medical School and Shriners Hospital for Children, Boston, MA 02114; ^fDivision of Biostatistics, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; ^gDepartment of Physics, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; ^hDepartment of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; and ⁱHoward Hughes Medical Institute, Chevy Chase, MD 20815

Contributed by Kurt J. Isselbacher, December 11, 2016 (sent for review October 14, 2016; reviewed by Anil K. Rustgi and Timothy Wang)

Circulating tumor cells (CTCs) are shed into the bloodstream by invasive cancers, but the difficulty inherent in identifying these rare cells by microscopy has precluded their routine use in monitoring or screening for cancer. We recently described a high-throughput microfluidic CTC-iChip, which efficiently depletes hematopoietic cells from blood specimens and enriches for CTCs with well-preserved RNA. Application of RNA-based digital PCR to detect CTC-derived signatures may thus enable highly accurate tissue lineage-based cancer detection in blood specimens. As proof of principle, we examined hepatocellular carcinoma (HCC), a cancer that is derived from liver cells bearing a unique gene expression profile. After identifying a digital signature of 10 liver-specific transcripts, we used a cross-validated logistic regression model to identify the presence of HCC-derived CTCs in nine of 16 (56%) untreated patients with HCC versus one of 31 (3%) patients with nonmalignant liver disease at risk for developing HCC ($P < 0.0001$). Positive CTC scores declined in treated patients: Nine of 32 (28%) patients receiving therapy and only one of 15 (7%) patients who had undergone curative-intent ablation, surgery, or liver transplantation were positive. RNA-based digital CTC scoring was not correlated with the standard HCC serum protein marker alpha fetoprotein ($P = 0.57$). Modeling the sequential use of these two orthogonal markers for liver cancer screening in patients with high-risk cirrhosis generates positive and negative predictive values of 80% and 86%, respectively. Thus, digital RNA quantitation constitutes a sensitive and specific CTC readout, enabling high-throughput clinical applications, such as noninvasive screening for HCC in populations where viral hepatitis and cirrhosis are prevalent.

circulating tumor cells | early cancer detection | hepatocellular carcinoma | blood biopsy | predictive modeling

The shedding by epithelial cancers of circulating tumor cells (CTCs) into the bloodstream underlies the blood-borne dissemination of cancer, although only a small fraction of CTCs gives rise to metastases (1). Enumeration and analysis of CTCs may thus enable noninvasive monitoring of advanced cancers, as well as early detection of invasive but localized tumors before they give rise to viable metastases. Recent advances in CTC isolation provide sensitive and high-throughput platforms to enrich for these rare tumor cells within blood specimens, but antibody staining and microscopic imaging of captured cancer cells remain a critical bottleneck limiting broad application of the technology (2). Classical CTC staining criteria include the presence of cell surface epithelial cell adhesion molecule (EpCAM) and cytoplasmic epithelial cytokeratins and the absence of the hematopoietic CD45 marker (3), but epithelial marker expression is highly variable and extensive imaging criteria must be applied to score immunofluorescent signals reliably from

rare cancer cells surrounded by contaminating leukocytes (4). Emerging microfluidic CTC isolation technologies that effectively deplete leukocytes without manipulating tumor cells (5) preserve cell viability and ensure high-quality RNA content, as demonstrated by single-cell RNA sequencing studies (6–8). These CTC isolation platforms now enable the application of powerful RNA-based digital PCR (dPCR) technologies to score molecular signatures of cancer cells, thus providing a potentially robust and high-throughput readout for the presence of CTCs within blood specimens. To test the feasibility of RNA-derived digital scoring of CTC-enriched cell populations, we applied this strategy to hepatocellular carcinoma (HCC), a cancer that lacks defining gene mutations but originates in liver cells with unique tissue-specific expression profiles.

Liver cancer is the second highest cause of cancer mortality worldwide, leading to 765,000 deaths in 2015 (9). In the developing world, the high prevalence of hepatitis B virus (HBV) infection drives the incidence of HCC; worldwide, it is estimated that 248 million individuals are infected with HBV (10). The risk of developing HCC is calculated as 0.5–1% per individual per year in HBV carriers without cirrhosis and as high as 8% per individual per year in patients with cirrhosis (11). Developed countries are also witnessing a rise in HCC incidence, linked with cirrhosis due to chronic hepatitis C virus (HCV) infection, alcohol abuse, obesity-associated nonalcoholic fatty liver disease

Significance

The early detection of hepatocellular carcinoma (HCC) is of paramount importance for improving patient outcomes, yet an accurate, high-throughput screening methodology has yet to be developed. By combining microfluidic depletion of hematopoietic cells from blood specimens with absolute quantification of lineage-derived transcripts, we demonstrate the highly specific detection of circulating tumor cells, enabling noninvasive detection and clinical monitoring of HCC.

Author contributions: M.K., I.B., T.T.K., D.T.M., S.J., R.K., H.Z., D.A.W., D.P.R., K.J.I., D.T.T., M.T., S.M., and D.A.H. designed research; M.K., I.B., T.T.K., D.T.M., S.J., J.A.L., J.D.M., X.H., L.G., S.S., M.C., U.H., L.V.S., R.T.C., and A.X.Z. performed research; X.H. contributed new reagents/analytic tools; M.K., I.B., T.T.K., D.T.M., S.J., A.M., D.T.T., M.T., S.M., and D.A.H. analyzed data; and M.K., I.B., K.J.I., D.T.T., M.T., S.M., and D.A.H. wrote the paper.

Reviewers: A.K.R., University of Pennsylvania; and T.W., Columbia University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹M.K. and I.B. contributed equally to this work.

²Present address: Merck Research Laboratories, Boston, MA 02114.

³To whom correspondence may be addressed. Email: dhaber@mgh.harvard.edu, maheswaran@helix.mgh.harvard.edu, or kisselbacher@mgh.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617032114/-DCSupplemental.

(NAFLD), and nonalcoholic steatohepatitis (NASH) (9). Early-stage HCC is potentially curable by thermal ablation, surgical resection, or liver transplantation, with a 5-y survival of 50–80% following these therapies (9). Once the tumor disseminates within or outside the liver, however, therapeutic options are limited and 5-y survival declines to below 15% (11).

Although early detection of HCC in high-risk individuals offers a strategy for successful curative treatment, screening in patient populations with liver cirrhosis has been limited by the poor test characteristics of the primary biomarker, serum levels of alpha fetoprotein (AFP) (11). CTCs have been reported in patients with HCC, but the characteristically low EpCAM cell surface expression in this tumor type has limited the utility of standard CTC measurements (12, 13). To establish a high-throughput, blood-based assay for HCC that would have broad applicability, we therefore adapted the microfluidic CTC-iChip isolation platform with a digital RNA-PCR readout combining liver-specific transcripts whose expression is retained in HCC.

We applied this molecular CTC assay as proof of principle in a pilot cohort of patients with HCC and high-risk patients suffering from liver disease.

Results

Establishment of the HCC-Specific RNA dPCR Assay. Fig. 1A outlines the RNA-based dPCR CTC-scoring assay. CTCs were isolated using the CTC-iChip microfluidic device, which depletes hematopoietic cells from blood by size-based exclusion of red blood cells, platelets, and plasma, followed by magnetic deflection of white blood cells (WBCs) tagged with magnetic bead-conjugated CD45, CD16, and CD66b antibodies (5, 14). The high efficiency depletion of WBCs (4- to 5-log purification) enriches CTCs, which are admixed with some contaminating WBCs (<500 WBCs per milliliter of processed whole blood) (5, 14). To replace CTC imaging with high-throughput detection of CTC-derived transcripts, we coupled whole-transcriptome amplification (WTA) of CTC-derived RNA with dPCR amplification, in which cDNA

molecules are encapsulated within individual aqueous droplets and multiple transcripts of interest are quantified by in-droplet PCR amplification (*Materials and Methods*).

We curated liver-specific transcripts whose expression is preserved in HCC cells, but virtually absent in blood components. Of the 20,000 genes measured from publicly available microarray datasets (15), the 100 most highly expressed genes in HCC were screened against expression profiles of hematopoietic cells and other normal tissues (16). These genes were then validated against a separate RNA-sequencing dataset, comparing 10 primary HCCs (17) with WBCs from eight healthy donor blood samples processed through the CTC-iChip (Fig. 1B). Quantitative RT-PCR (qRT-PCR) of cDNA from purified healthy donor WBCs ($n = 3$) was used to eliminate genes with low but detectable background signal (Fig. S1). Based on these results, 10 genes [AFP, alpha 2-HS glycoprotein (*AHSG*), albumin (*ALB*), apolipoprotein H (*APOH*), fatty acid binding protein 1 (*FABP1*), fibrinogen beta chain (*FGB*), fibrinogen gamma chain (*FGG*), glypican 3 (*GPC3*), retinol binding protein 4 (*RBP4*), and transferrin (*TF*)] were selected for developing the dPCR assay.

To technically validate this strategy, we determined the number of transcripts measured from the introduction of individually micromanipulated cells from the liver cancer cell line HepG2 known to express albumin into 4 mL of blood from healthy donors, followed by CTC-iChip processing and dPCR for the liver-specific transcript *ALB* (Fig. 1C). No *ALB* RNA-positive droplets were observed in unspiked blood processed through the CTC-iChip, whereas a range from 2 to 100 spiked HepG2 cells generated from 240 to 28,800 *ALB* transcripts. Given the heterogeneity of HCC cells within clinical specimens, we applied low-template WTA to maximize starting material and optimized the 10-gene liver-specific panel to ensure that this dramatic signal amplification preserved the relative distribution among multiple markers. We ensured the absence of amplification bias in three independent experimental replicates (Fig. 1D and Fig. S2). Compared with unamplified cDNA derived from 1 ng of

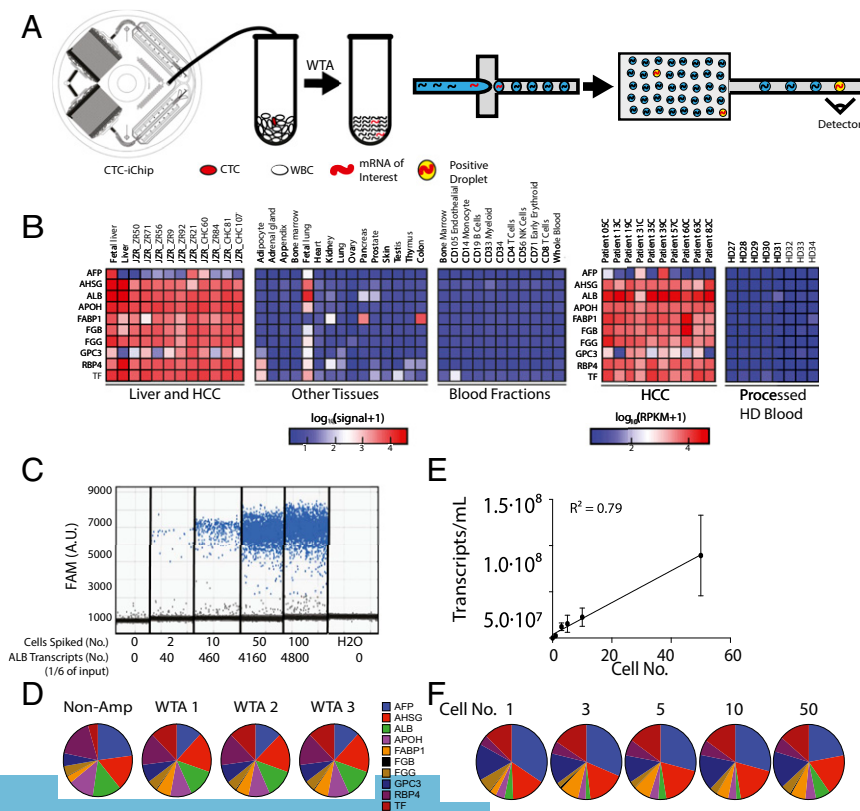


Fig. 1. dPCR quantitation of HCC cells after microfluidic enrichment from blood. (A) Schematic representation of the integrated platform for digital RNA-PCR scoring of CTCs. Hematopoietic components are depleted from whole blood through CTC-iChip processing as previously described (5). RNA from the CTC-enriched product is subjected to WTA, encapsulation of cDNA molecules within lipid droplets, and PCR amplification for transcripts of interest. (B) Heat maps derived from microarray (Left) and RNA-sequencing (RNAseq) (Right) datasets comparing expression of 10 liver-specific transcripts in HCC versus other tissues. The microarray dataset compares fetal liver, adult liver, and 10 cases of HCC (JZR samples) versus normal tissues (15 samples shown of 79 tissues tested) and blood components (15, 16). RNAseq compares 10 cases of HCC (17), with WBCs collected from eight independent healthy donor (HD) blood samples processed through the CTC-iChip. (C) dPCR quantitation of *ALB* transcripts from micromanipulated HepG2 spiked into whole blood and processed through the CTC-iChip. Each data point represents one-sixth of the CTC-iChip product. (D) Pie charts representing the distribution of transcripts for each of the 10 selected liver-specific genes, following dPCR analysis of 1 ng of HepG2-cell RNA. Samples were nonamplified or subjected to WTA (three independent reactions). (E) Total number of transcripts of interest after spiking increasing numbers of HepG2 cells into blood ($n = 3$), CTC-iChip processing, and dPCR. (F) Pie charts depicting the relative fraction of droplets for each of the 10 target transcripts after spiking increasing numbers of HepG2 cells into blood and CTC-iChip processing, as noted in E ($n = 3$).

HepG2 RNA, the increased overall signal resulting from WTA (25,000- to 100,000-fold amplification per gene) preserved the relative proportion of each transcript among the WTA replicates (Fig. S2 A and B).

To test the sensitivity of the 10-gene digital assay in rare tumor cells admixed with blood cells, we again micromanipulated either 1, 3, 5, 10, or 50 single HepG2 cells into 4 mL of healthy donor blood, which was then processed through the CTC-iChip, followed by WTA and dPCR. In two of three replicates, a single spiked HepG2 cell was detected, with an average 5,000-fold increase in signal over unspiked blood controls. All liver-specific genes were detected with progressively increasing signal and preservation of marker distribution as the numbers of input cells increased from 1 to 50 cells ($R^2 = 0.79$; Fig. 1 E and F). The high sensitivity of dPCR raised the possibility that CTC enrichment might not be required to detect tumor-derived transcripts in nucleated blood cell fractions. Past reports have suggested that standard RT-PCR amplification might identify PCR products comigrating with the expected *ALB* transcript from unpurified blood cells of patients with HCC (18); however, we were unable to reproduce this finding using the more sensitive and specific dPCR technology ($n = 9$ HCC samples; Fig. S3). We therefore applied the combination of microfluidic CTC enrichment followed by dPCR detection of HCC-derived transcripts to clinical specimens from HCC patients.

Digital CTC Detection in Patients with HCC. We evaluated the performance of the optimized digital CTC-scoring assay in blood samples (5–15 mL) from six patient cohorts, per Institutional Review Board (IRB)-approved protocols at Massachusetts General Hospital. The cohorts studied included (i) healthy blood donors ($n = 26$, median age = 55 y); (ii) patients with high-risk nonmalignant chronic liver disease (CLD), including hepatitis virus-associated cirrhosis, who were being routinely monitored

for HCC development ($n = 31$, median age = 58 y); (iii) newly diagnosed, untreated patients with HCC ($n = 16$, median age = 66 y); (iv) patients with HCC actively receiving therapy with radiographically evident disease ($n = 32$, median age = 67 y); (v) patients with HCC who had undergone curative-intent interventions, such as ablation, resection, or liver transplantation, and clinically had no evidence of disease (NED) any longer ($n = 15$, median age = 66 y); and (vi) patients with primary malignancies other than HCC, with or without liver metastases ($n = 44$, median age = 62 y). Patients in categories ii through vi were categorized by a clinician blinded to the dPCR data. Clinical characteristics of these cohorts are provided in Tables S1–S6.

dPCR analysis of CTC-iChip-processed blood specimens from patients with HCC generated high signal for individual transcripts, compared with healthy donors, patients with CLD, or patients with other malignancies (Fig. 2A). As expected, intra-cohort variability was observed, with some HCC patient samples exhibiting high signal from multiple liver-specific transcripts and others containing few transcripts of interest. HCC cases in which no signal was detected may reflect the absence of CTCs within the single 5- to 15-mL blood sample or expression of transcripts that are not captured by the 10-gene panel.

To integrate the results of multiple genes into a statistically robust scoring model, we screened each transcript to determine if it served as a statistically significant single-gene predictor of HCC status (Fig. S4). Nine of the 10 genes met this selection criterion (all excluding *GPC3*). These genes were then combined into a single metric CTC score, using a leave-one-out cross-validated (LOOCV) multivariate logistic regression model (Fig. S5). The LOOCV allowed us to build and test the model using a single HCC patient cohort, although the high stringency associated with this approach may underestimate the predictive value of the model; Fig. 2C demonstrates the change in model performance with and

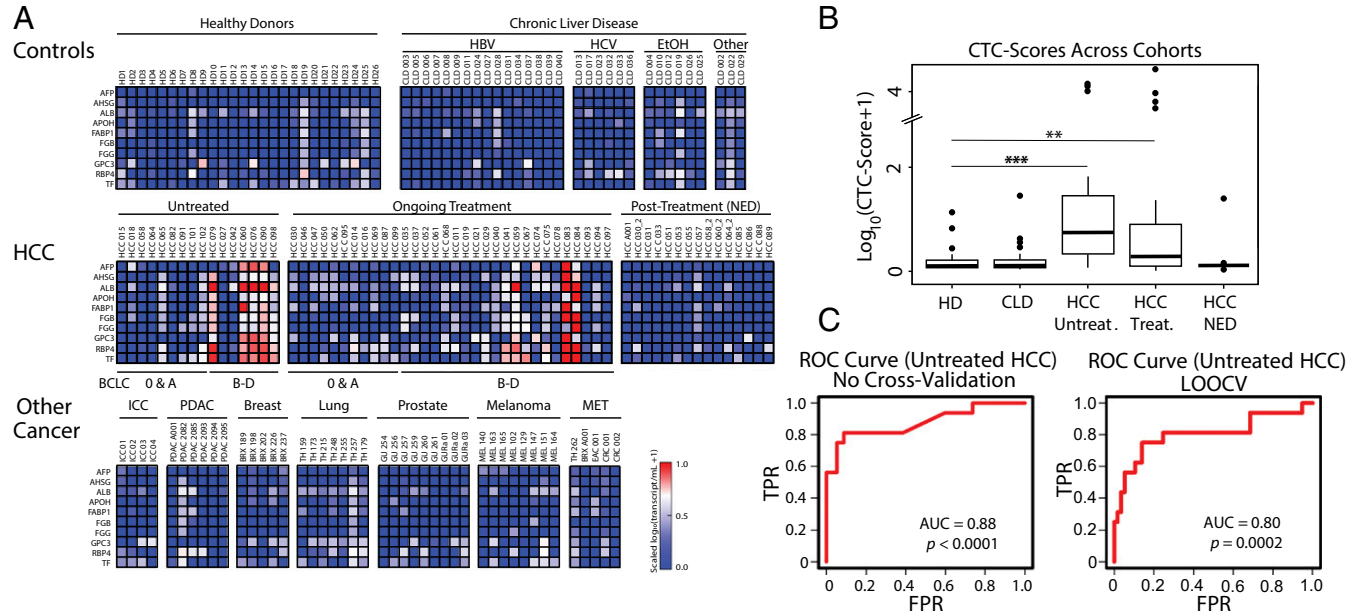


Fig. 2. CTC score from patients with HCC compared with at-risk patients. (A) Heat maps depicting relative signal intensities for each of the 10 liver-specific transcripts across different patient cohorts. Primary droplet numbers are log₁₀-transformed and scaled to the highest value for each transcript. (Upper) Healthy donors (blood bank, $n = 26$) and high-risk patients with CLD under active clinical surveillance for HCC ($n = 31$). Etiologies of CLD include HBV infection ($n = 16$), hepatitis C virus (HCV) infection ($n = 6$), alcohol (EtOH) ($n = 6$), or other causes ($n = 3$). (Middle) Patients with HCC, classified as untreated (newly diagnosed, $n = 16$) or receiving ongoing treatment (currently undergoing various therapies, $n = 32$). Patients are grouped according to Barcelona Clinic Liver Cancer (BCLC) criteria from early clinical stages (0 and A) to advanced clinical stages (B–D). Patients who have completed treatment and have NED are shown ($n = 15$). Four of these cases represent repeated analysis of patients initially tested before or during treatment (HCC-030_2, HCC-058_2, HCC-060_2, and HCC-064_2). (Lower) Patients with cancers other than HCC ($n = 43$): intrahepatic cholangiocarcinoma (ICC); pancreatic ductal adenocarcinoma (PDAC); breast, lung, and prostate cancers; melanoma; and cancers of nonhepatic origin metastatic to the liver (MET). Clinical data are listed in Table S6. (B) Box plots representing the integrated CTC score for the patient cohorts above. ** $P < 0.01$, *** $P < 0.0001$ (χ^2 , degrees of freedom = 5). (C) Receiver operator characteristic (ROC) curves for untreated HCC both without (Left) and with (Right) LOOCV. AUC, area under the curve; FPR, false-positive rate; TPR, true-positive rate.

without cross-validation. We tested the multigene model with 48 patients with HCC versus 57 patients without cancer (Fig. 2C).

Nine of 16 (56%) untreated patients with HCC were classified as positive by CTC score, compared with only one of 31 (3%) patients with at-risk nonmalignant CLD and two of 26 (7.6%) age-matched healthy blood donors ($P < 0.0001$, χ^2 ; Fig. 2B). Patients with HCC undergoing therapy but with radiographically detectable disease had a lower fraction of cases with positive CTC scores [nine of 32 (28%)], but this fraction was still significantly higher than the control population ($P = 0.004$, χ^2). Patients with NED after curative-intent treatment were only positive in one of 15 (7%) cases, an incidence comparable to the incidence of the control population ($P = 0.56$, not significant). Together, these results demonstrate that the CTC score can identify patients with active disease while maintaining a high degree of specificity. Among all patients with HCC, the CTC score was not associated with specific underlying risk factors ($P = 0.73$), but it was highly correlated with Barcelona Clinic Liver Cancer staging ($P = 0.011$) and trended toward significance when stratifying by vascular invasion ($P = 0.06$; Fig. S6).

We next determined the feasibility of differentiating between patients with HCC and patients with malignancies other than HCC using a separate logistic regression (Fig. S7). In this comparison, six of the 10 genes were statistically significant predictors of HCC vs. non-HCC status (*AFP*, *AHSG*, *APOH*, *FABP1*, *FGB*, and *FGG*), yielding a multipredictor model with different features than the previous model (Fig. S7 A and B). Among patients with pancreatic, prostate, breast, and non-small-cell lung cancers; cholangiocarcinoma; and melanoma, 39 of 44 (88%) cancers/carcinomas were correctly distinguished from HCC at a sensitivity of 50% for patients with HCC. Optimal differentiation of tissue of origin among CTCs will likely benefit from the inclusion of additional markers that are specific for other tumors, in addition to exclusion of HCC-associated transcripts.

Longitudinal Monitoring of Patients with HCC. The higher incidence of positive CTC scores in newly diagnosed, untreated patients with HCC (56%) compared with those patients with HCC who were undergoing active treatment (28%), and those patients with HCC who completed curative-intent therapy (7%) suggests a potential role for CTCs in longitudinal monitoring of tumor response. Furthermore, increasing tumor burden, as defined by the Barcelona Clinic Liver Cancer staging criteria, is associated with increased CTC score values (Fig. 2A and Fig. S6). A subset of patients with HCC in our study was monitored longitudinally for tumor response. Supporting the robustness of the assay, the CTC score remained high in two patients (HCC-041 and HCC-075), with no therapeutic intervention or change in clinical status between blood draws (Fig. 3A and Fig. S7E). Two other patients (HCC-058 and HCC-060) underwent surgical tumor resection and demonstrated a decrease in CTC score postoperatively (Fig. 3B). The CTC score for patient HCC-042 decreased impressively following an immune checkpoint inhibitor (nivolumab) treatment, and then showed a further reduction 3 wk after subsequent radioembolization of the tumor (Fig. 3C). Coincident computed tomography scans demonstrated a significant tumor response to radioembolization. Of note, for three of these five patients (HCC-042, HCC-058, and HCC-075), serum protein AFP measurements were below clinically informative values (<20 ng/mL) at all draw points. Although serum AFP protein measurements are often used for monitoring tumor response in patients with HCC, they are below detection in a significant fraction of cases. In such cases, CTC score monitoring may serve as a complementary marker to assess disease status.

Early Detection of HCC in High-Risk Populations. Although early detection of localized HCC in individuals with liver cirrhosis provides the only hope for curative therapy, serum AFP alone does not provide sufficient accuracy to enable screening of at-risk populations. Using a cutoff of 20 ng/mL, AFP has an estimated sensitivity of 53% with a specificity of 87%, leading to a positive

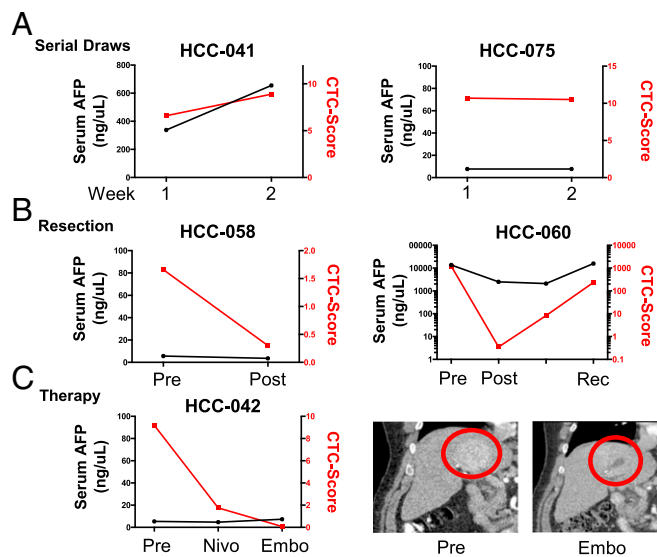


Fig. 3. Longitudinal monitoring of patients treated for HCC. (A) Serial blood measurements performed at 1-wk intervals in two patients (HCC-041 and HCC-075) in the absence of therapeutic intervention. Concurrent CTC score (red) and serum AFP (black) measurements are shown. (B) Longitudinal monitoring of two patients (HCC-058 and HCC-060), before (Pre) and after (Post) resection of localized HCC. HCC-060 had NED 1 mo after resection but then developed a recurrence of HCC (Rec). (C) Serial monitoring of a patient (HCC-042) initially treated with the immune checkpoint inhibitor nivolumab (Nivo), followed by radioembolization (Embo) of the residual mass. The tumor mass and postembolization changes are shown by computed tomography scan. Concurrent CTC score (red) and serum AFP (black) measurements are shown.

predictive value (PPV) of 6% in populations where the expected prevalence of HCC is 1% (19). Raising the AFP threshold to 100 ng/mL improves specificity to 99% but reduces the test sensitivity to 31% (20). Given these poor test characteristics, the American Association for the Study of Liver Diseases does not recommend using AFP as a screening tool for HCC (11).

To model the potential combination of AFP and CTC score in screening for HCC, we determined the correlation between these two biomarkers in all of the patients with HCC for whom concomitant assay results were available. CTC score and serum AFP levels were not significantly correlated ($R^2 = 0.0007$, $P = 0.74$), with concordance restricted to cases with high serum AFP levels (Fig. 4A). The discordance between AFP protein levels and the CTC score is consistent with the underlying basis for these assays: The multigene CTC score is a digital signature to quantify HCC cells that have invaded into blood, whereas serum AFP measures a single protein produced by HCC cells and released into blood from tumor deposits. As such, the two assays may be orthogonal and have added value as blood-based biomarkers for HCC.

Although our pilot study was not powered to test the accuracy of the CTC score directly in early detection of HCC, we modeled two strategies using either a high or low cutoff for AFP measurements. First, we tested the additive value of CTC score positivity and high-threshold AFP (100 ng/mL). Of the 15 patients with newly diagnosed HCC for whom both AFP and CTC scores were available, four (27%) were detected by CTC score alone, one (7%) by AFP alone, and five (33%) by both assays (Fig. 4B). Together, either AFP or CTC score was positive in 67% of patients, leaving only one-third undetected by either method. Importantly, among all 16 patients with newly diagnosed HCC, six (38%) patients met the Milan criteria for liver transplantation, a clinical indication that the disease is sufficiently localized to enable curative therapy. CTC scores were available for all six patients who met the Milan criteria, of whom two (33%) were positive (Fig. 4C). AFP levels were available for five of six patients, but none were above the 100-ng/mL threshold. Thus, a

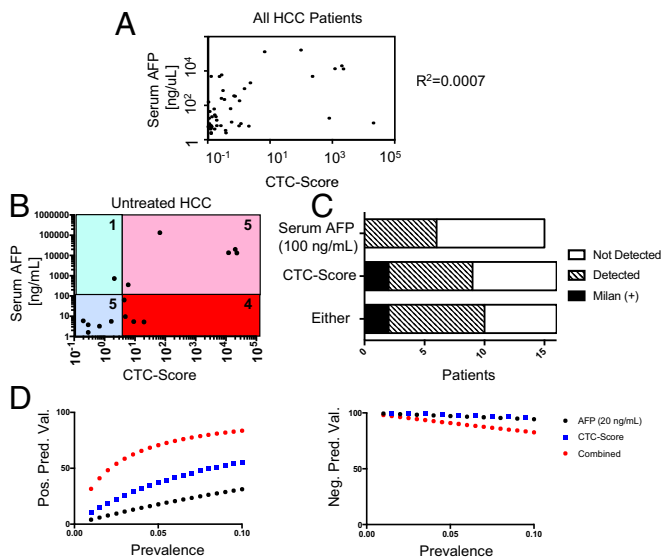


Fig. 4. Modeling early detection of HCC using CTC score and AFP measurements. (A) Absence of correlation between CTC score and serum AFP levels in all patients with HCC with concomitant measurements. (B) Proportion of 15 newly diagnosed, untreated patients with HCC identified by CTC score alone, AFP (>100 ng/mL) alone, or both CTC score and AFP. No AFP measurement was available for one patient with untreated HCC. (C) Bar graphs representing all newly diagnosed HCC patients, showing those patients identified by serum AFP (>100 ng/mL), CTC score, or the combination of the two tests (Either). Six of these newly diagnosed patients met the Milan criteria for localized disease amenable to curative liver transplantation [Milan (+)]. Two of six Milan(+) patients were identified by CTC score, but none of five had an AFP level >100 ng/mL [one Milan (+) patient did not have AFP measurement]. (D) PPV and NPV calculations for CTC score alone, AFP (>20 ng/mL) alone, or both, as a function of HCC prevalence. The CTC score model assumes 56% test sensitivity and 95% specificity, as observed in untreated patients with HCC; for AFP (>20 ng/mL), the 53% test sensitivity and 87% specificity are established from a population-based study (19).

subset of patients identified as having HCC by digital CTC assay may have curable HCC.

As an alternative HCC detection strategy, we modeled the initial screening of patients with cirrhosis using the higher sensitivity 20-ng/mL AFP cutoff, followed by CTC score analysis as a confirmatory test to provide the required specificity. Such a sequential strategy has been instrumental in population screening for infectious disease, such as HIV (21); it has the benefit of capturing most patients at risk with a rapid primary test, thereby increasing the disease prevalence among the group tested with the higher specificity confirmatory test. In patients with HBV hepatitis without cirrhosis (associated with a 0.5–1% annual incidence of HCC), such sequential AFP/CTC screening leads to a calculated PPV of 36%, with a negative predictive value (NPV) of 98% (Fig. 4D). In higher risk patients with HBV-induced cirrhosis (8% annual incidence of HCC), the PPV rises to 86%, with an NPV of 83%. Although these calculations require confirmation in large population studies, these predicted values are within the range that would warrant cancer surveillance initiatives within appropriate clinical settings (22).

Discussion

We have described a sensitive and specific RNA-based readout for detection of CTCs following their microfluidic enrichment from blood specimens. Our approach combines the high efficiency depletion of hematopoietic cells, enabling isolation of CTCs with intact RNA and without bias for expression of tumor- or epithelial-specific cell surface epitopes, together with CTC quantitation using a high-throughput, tissue lineage-specific dPCR assay. Together, these approaches overcome the rate-limiting hurdle in CTC detection, namely, antibody staining and microscopic scoring

of heterogeneous CTCs among an excess of contaminating WBCs. Moreover, the dramatic signal amplification and the molecular specificity derived from dPCR provide an effective way to detect the rare but highly biologically significant occurrence of intact tumor cells in the blood circulation. As an initial proof of concept, we applied this digital CTC measurement strategy to HCC, a highly lethal malignancy with worldwide impact, for which early detection strategies are currently inadequate (11).

The concept of liquid biopsies for noninvasive monitoring of cancer has emerged as one of the most promising approaches in cancer diagnostics, with applications ranging from early detection to treatment selection and monitoring response (23, 24). Three complementary technologies each interrogate different biological specimens and rely upon distinct technological assays. Circulating tumor DNA (ctDNA) is derived from small fragments of genomic DNA shed into the vasculature by dying tumor cells, amid the background of DNA released from normal tissues, and it provides powerful opportunities for targeted DNA-based genotyping (25). However, an inherent limitation of ctDNA-based genotyping is that it does not provide information as to the tissue of origin for mutations detected in the blood, in contrast to CTCs, which provide a source of intact RNA for lineage-based analysis. Genotype-based cancer detection is also of limited utility in tumors such as HCC, where highly prevalent mutations have not been identified. An alternative blood-based cancer detection technology takes advantage of exosomes, small membrane-bound cellular fragments encapsulating cytoplasmic RNA and other cellular components that are released by both tumor and normal cells. Although strategies for enrichment of tumor-derived exosomes are still being optimized, the analysis of pooled exosomes has allowed RNA-based detection of cancer-associated mutations (26). However, the fact that normal tissues also abundantly shed exosomes precludes the use of lineage and tissue markers to identify tumor-derived expression signatures in blood specimens. In contrast to exosomes, whole cells derived from normal tissues are extraordinarily rare in the blood circulation. Hence, the initial isolation of intact CTCs in the bloodstream, followed by their molecular quantitation, may provide a highly specific diagnostic assay that is amenable to large-scale clinical applications. HCC, which arises within well-defined at-risk populations, and in which early detection may be curative, is particularly appropriate to serve as a “proof of principle” for CTC-based screening. Of note, the sensitivity and specificity of this molecular CTC assay hinge on the successful identification of HCC-derived transcripts: Additional targets that capture the full heterogeneity of HCC may be identified by single-cell sequencing of HCC CTCs, and analysis of larger blood volumes may also improve the likelihood of CTC detection, especially in patients with small lesions.

The poor performance of AFP alone in HCC screening stems from the fact that only a minority of patients with liver cancer have very high elevations in this marker (>100 ng/mL), whereas low levels (>20 ng/mL) are common in conditions that predispose to HCC, including viral hepatitis. However, the combination of AFP screening with a second more specific assay has been shown to be effective in reducing disease mortality. In a large randomized trial, a 37% reduction in HCC-related mortality was reported among high-risk patients who underwent surveillance with AFP and liver ultrasound (27). Ultrasound-based surveillance is now practiced in many centers, but image quality is operator-dependent and degraded in the setting of obesity or cirrhosis. Furthermore, access to high-quality ultrasound is limited in developing countries, which bear the greatest burden of HCC. The technology that we describe here, together with the analysis of pilot clinical cohorts and clinical modeling studies, raises the possibility that digital CTC scoring may provide an important new tool for HCC detection. The scalability of digital CTC monitoring may be particularly useful in underserved populations that lack access to MRI and ultrasound screening. Although large population-based studies are now required to test and validate this technology’s performance against existing screening standards, we favor the combination of AFP

and CTC-scoring assays as orthogonal markers that, together, may provide both the sensitivity and specificity required for noninvasive and cost-effective HCC screening in high-risk populations that serum AFP alone is unable to provide. Alternative CTC-scoring algorithms may also enhance the differentiation of HCC from other hepatic lesions, an application of particular importance in the United States, where the use of imaging tests currently drives HCC detection and monitoring. Thus, quantitative analysis of multiple tissue-specific transcripts derived from CTCs may be optimized for distinct applications in the diagnosis and treatment of patients with HCC.

Finally, we note that digital scoring of CTCs for cancer monitoring is broadly applicable to other cancer types. Indeed, many cancers originate in tissues that express specialized gene transcripts that are absent in normal blood cells, and the curation and testing of these markers may enable high-sensitivity detection and monitoring of rare cancer cells in the blood. Such blood-based molecular monitoring for CTCs holds considerable promise for the early detection of cancer.

Materials and Methods

Gene Target Identification and Validation. Publicly available RNA-sequencing and microarray datasets were used to identify the top 100 genes highly expressed in HCC with very low to no expression in other tissues and blood components (15–17). The low expression of candidate genes within WBCs persisting in the CTC-iChip output was confirmed by RNA-sequencing of processed blood from healthy donors, and qRT-PCR of WBCs purified from whole blood was used as an additional exclusion criterion. Ten genes were selected to establish a signature of HCC-derived CTCs, enriched within a background of normal blood cells.

Cell Culture and RNA Processing. HepG2 cells were cultured following American Type Culture Collection-recommended culturing conditions. RNA was isolated using the RNeasy Plus Micro kit (Qiagen), and cDNA was generated using the SuperScript III Reverse Transcriptase kit (Thermo Fisher). Buffy coat WBC preparations were generated using standard procedures, followed by TRIzol RNA extraction (Ambion). For qRT-PCR assays, 5 ng of cDNA generated from HepG2 cells or from WBCs (buffy coats) of three independent blood donors was amplified and compared with GAPDH using the Applied Biosystems 7500 RT-qPCR assay (40 cycles). Primer and probe combinations are provided upon request. For single-cell manipulation and spike-in studies, individual cells were micropipetted using an

Eppendorf TransferMan NK2 micromanipulator and introduced into whole blood samples from healthy donors, before processing through the CTC-iChip.

Patient Cohorts and Blood Processing Through the Microfluidic CTC-iChip. Patient cohorts and clinical characteristics are provided in Tables S1–S6. Dana-Farber/Harvard Cancer Center IRB-approved protocols were used to obtain consent from all patients and healthy donors. Blood was processed using the CTC-iChip as previously described (5). A detailed description of each cohort and blood processing conditions can be found in *SI Materials and Methods*.

WTA and Droplet dPCR. WTA was performed on CTC-iChip-derived RNA using the SMARTer Ultra Low-input RNA kit, version 3 (Clontech Laboratory). One-third of the RNA extracted from the CTC-iChip product was loaded into each reaction. The dPCR experiments were performed using an AutoDG automated droplet generator, C1000 Touch Deep-well thermocycler, and QX200 plate reader (Biorad) following manufacturer recommendations, with 1% of the WTA product. Thermocycling conditions and primers are provided upon request.

Statistical Calculations and Logistic Regression. Genes that served as statistically significant predictors of disease status were used to build a multi-predictor logistic regression whose performance parameters were determined using LOOCV (Fig. 2C and Figs. S4, S5, and S7). For longitudinally collected samples or treated patients with HCC with NED, a single logistic regression was fit using the entirety of the aforementioned training set (Fig. S5). To calculate the PPV and NPV of the diagnostic assays at varying disease prevalences, we used published sensitivities and specificities for serum AFP at 20 ng/mL (19) and chose the CTC score sensitivity that yielded 95% specificity.

ACKNOWLEDGMENTS. We thank all of the patients who participated in this study. We also thank the Massachusetts General Hospital (MGH) nurses and clinical research coordinators for their assistance with the study; A. Bardia, R. Sullivan, P. Saylor, and R. Lee for graciously gathering clinical information and patient samples; and L. Libby for invaluable technical support. This work was supported by NIH Grant 2R01CA129933, the Howard Hughes Medical Institute, the National Foundation for Cancer Research (D.A.H.), NIH Quantum Grant 2U01EB012493 (to M.T. and D.A.H.), MD/PhD Training Grant T32GM007753 (to M.K.), NIH Grant DK078772 (to R.T.C.), NIH Grant T32DK007191 (to I.B.), the Department of Defense and Prostate Cancer Foundation (D.T.M.), National Science Foundation Grants DMR-1310266 and ECS-0335765, and Harvard Materials Research Science and Engineering Center Grant DMR-1420570 (to D.A.W.). The MGH has filed for patent protection for the digital CTC detection approach.

- Poste G, Fidler IJ (1980) The pathogenesis of cancer metastasis. *Nature* 283(5743):139–146.
- Yu M, Stott S, Toner M, Maheswaran S, Haber DA (2011) Circulating tumor cells: Approaches to isolation and characterization. *J Cell Biol* 192(3):373–382.
- Allard WJ, et al. (2004) Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* 10(20):6897–6904.
- Stott SL, et al. (2010) Isolation and characterization of circulating tumor cells from patients with localized and metastatic prostate cancer. *Sci Transl Med* 2(25):25ra23.
- Ozkumur E, et al. (2013) Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* 5(179):179ra47.
- Miyamoto DT, et al. (2015) RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* 349(6254):1351–1356.
- Aceto N, et al. (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158(5):1110–1122.
- Ting DT, et al. (2014) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Reports* 8(6):1905–1918.
- Llovet JM, et al. (2016) Hepatocellular carcinoma. *Nat Rev Dis Primers* 2:16018–16023.
- Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ (2015) Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *Lancet* 386(10003):1546–1555.
- Bruix J, Sherman M; Practice Guidelines Committee, American Association for the Study of Liver Diseases (2005) Management of hepatocellular carcinoma. *Hepatology* 42(5):1208–1236.
- Fan JL, Yang YF, Yuan CH, Chen H, Wang FB (2015) Circulating tumor cells for predicting the prognostic of patients with hepatocellular carcinoma: A meta analysis. *Cell Physiol Biochem* 37(2):629–640.
- Went PT, et al. (2004) Frequent EpCam protein expression in human carcinomas. *Hum Pathol* 35(1):122–128.
- Karabacak NM, et al. (2014) Microfluidic, marker-free isolation of circulating tumor cells from blood samples. *Nat Protoc* 9(3):694–710.
- Boyault S, et al. (2007) Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 45(1):42–52.
- Su AI, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101(16):6062–6067.
- Huang Q, et al. (2011) RNA-Seq analyses generate comprehensive transcriptomic landscape and reveal complex transcript patterns in hepatocellular carcinoma. *PLoS One* 6(10):e26168.
- Kar S, Carr BI (1995) Detection of liver cells in peripheral blood of patients with advanced-stage hepatocellular carcinoma. *Hepatology* 21(2):403–407.
- Gupta S, Bent S, Kohlwes J (2003) Test characteristics of α -fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann Intern Med* 139(1):46–50.
- Trevisani F, et al. (2001) Serum α -fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: Influence of HBsAg and anti-HCV status. *J Hepatol* 34(4):570–575.
- Mylonakis E, Paliou M, Lally M, Flanigan TP, Rich JD (2000) Laboratory testing for infection with the human immunodeficiency virus: Established and novel approaches. *Am J Med* 109(7):568–576.
- Skaane P, Hovind S, Skjennald A (2007) Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. *Radiology* 244(3):708–717.
- Haber DA, Velculescu VE (2014) Blood-based analyses of cancer: Circulating tumor cells and circulating tumor DNA. *Cancer Discov* 4(6):650–661.
- Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A (2013) Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* 10(8):472–484.
- Bettgowda C, et al. (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* 6(224):224ra24.
- Skog J, et al. (2008) Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* 10(12):1470–1476.
- Zhang B-H, Yang B-H, Tang Z-Y (2004) Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* 130(7):417–422.