UCLA
UCLA Electronic Theses and Dissertations

Title
Development of Circulating Tumor Cell Sequencing as a Biomarker in Gastrointestinal Cancers

Permalink
https://escholarship.org/uc/item/57q167j5

Author
Court, Colin

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Development of Circulating Tumor Cell Sequencing as a Biomarker in Gastrointestinal Cancers

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Molecular, Cellular, & Integrative Physiology

by

Colin MacDonald Court

2019

© Copyright by

Colin MacDonald Court

2019

ABSTRACT OF THE DISSERTATION

Development of Circulating Tumor Cell Sequencing as a Biomarker in Gastrointestinal Cancers

by

Colin MacDonald Court

Doctor of Philosophy in

Molecular, Cellular & Integrative Physiology

University of California, Los Angeles, 2019

Professor Thomas G. Graeber, Chair

For precision oncology to become a reality, clinicians need access to tumor tissue in a real-time, repeatable, and cost-effective manner. Circulating tumor cells (CTCs), cells of tumor origin that circulate in the blood, are one possible solution, with the potential to serve as a "liquid biopsy" for gastrointestinal (GI) tumors.

In this thesis, we demonstrate the development and validation of an assay for performing molecular analysis of CTCs as a biomarker for GI cancers. We used a microfluidic platform to develop GI tumor-specific CTC assays for pancreatic and liver cancer through incorporation of tissue-specific markers for CTC capture and identification. We then conducted prospective studies testing the utility of CTC enumeration as a biomarker in pancreatic (n=100) and liver (n=61) cancer. We found that the use of a tissue-specific assay resulted in higher CTC counts than in previous studies, that CTC counts correlated with increasing stage and that CTC enumeration may have diagnostic and prognostic utility as a biomarker in GI cancers.

Using a modified assay that allows for the complete isolation of the CTCs identified, we confirmed the tumor origin of the CTCs we identified by comparison of mutational status with the primary tumor using both Sanger and next generation sequencing methods.

After confirming their tumor origin, we developed a quality control assay that allowed us to determine how many CTCs we would need to isolate to reliably perform CTC molecular analysis. Using this metric, we modified our CTC sequencing protocol that resulted in reliable sequencing from as few as 3-5 CTCs as opposed to the 10-20 needed using the traditional CTC sequencing protocol. We then used our optimized CTC sequencing protocol in a prospective study to test the feasibility and potential utility of CTC molecular analysis as a biomarker in liver cancer.

The research described in this thesis serves two primary goals. The first is as a validated method that allows for the use of CTCs in liquid biopsy-type applications for both HCC and PDAC patients. The second is as a framework for validating CTC assays and definitions to ensure the reliability and reproducibility of CTC sequencing results.

The dissertation of Colin MacDonald Court is approved.

Joseph R. Pisegna

Amy Catherine Rowat

James S Tomlinson

Xinshu Xiao

Thomas G. Graeber, Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

Table of Contents

List of Figures

List of Tables

Acknowledgements

This work was supported in part by a UCLA Department of Surgery Levey Award as well as a Jonsson Comprehensive Cancer Impact Grant. The NanoVelcro Chips used in this research were supported by research grants (R33CA174562 and U01 CA198900) and an SBIR grant (R44 CA180482) from National Institute of Health. I would like to thank Dr. James Tomlinson for his mentorship throughout this process. Dr. Thomas Graeber provided an enormous amount of guidance and help and attending his lab meetings was what made me realize I could learn the computational biology necessary to complete this project. Thank you for always being available and willing to work around my haphazard schedule. Dr. Vatche Agopian has been tremendously helpful and his expertise and direction were pivotal in making this project a success. Dr. Jonathan Yahalom provided essential qualitative analysis as well. None of this work would have been possible without Dr. HR Tseng's development of the NanoVelcro CTC assay as well as his countless hours of support. I would especially like to thank all the members of my committee for their guidance and support.

I would like to sincerely thank the UCLA STAR program for their support and for establishing a program that allowed me to pursue my goals. I would especially like to thank the STAR Directors Drs. Linda Demer and Mitchell Wong for allowing physicians outside of the Department of Medicine to participate. I am deeply indebted to Dr. Joe Hines and Timothy Donahue for believing in me and allowing me the extra time in my surgery residency to finish this research.

I would like to thank my family and friends for their continued support throughout the long process of my research and residency. They have been unwavering in their support and are truly the reason I have been able to pursue my dreams. I would like to thank my children Theo and Millie for being excellent babies and learning to sleep through the night at such an early age. Most importantly, I would like to thank my wife Kelly who has been a constant source of love and support every step of the way and without whom I never would have considered pursuing this project. Thank you for always supporting my dreams.

Biographical Sketch

EDUCATION/TRAINING

Graduate School

University of Texas, San Antonio
San Antonio, Texas, 2008-2012
MD, May 2012

University of Texas, Houston
San Antonio, Texas, 2008-2012
MPH, May 2012

Undergraduate

Claremont McKenna College
Claremont, California, 2003-2007
Bachelor of Science in Neuroscience, June 2007

PUBLICATIONS

Sho, S., **Court, C. M.**, Winograd, P., Toste, P. A., Pisegna, J. R., Lewis, M., . . . Tomlinson, J. S. (2018). A Prognostic Scoring System for the Prediction of Metastatic Recurrence Following Curative Resection of Pancreatic Neuroendocrine Tumors. *J Gastrointest Surg.* doi:10.1007/s11605-018-4011-7

DiPardo, B. J., Winograd, P., **Court, C. M.**, & Tomlinson, J. S. (2018). Pancreatic cancer circulating tumor cells: applications for personalized oncology. *Expert Rev Mol Diagn*, 18(9), 809-820. doi:10.1080/14737159.2018.1511429

Court, C. M., & Hines, O. J. (2018). The New American Joint Committee on Cancer TNM Staging System for Pancreatic Cancer-Balancing Usefulness With Prognostication. *JAMA Surg,* e183629. doi:10.1001/jamasurg.2018.3629

Court, C. M., Hou, S., Winograd, P., Segel, N. H., Li, Q. W., Zhu, Y., . . . Agopian, V. G. (2018). A Novel Multimarker Assay for the Phenotypic Profiling of Circulating Tumor Cells in Hepatocellular Carcinoma. *Liver Transpl*. doi:10.1002/lt.25062

Court, C. M., Ankeny, J. S., Sho, S., Winograd, P., Hou, S., Song, M., . . . Tomlinson, J. S. (2018). Circulating Tumor Cells Predict Occult Metastatic Disease and Prognosis in Pancreatic Cancer. *Ann Surg Oncol*, 25(4), 1000-1008. doi:10.1245/s10434-017-6290-8

Lunsford, K. E., **Court, C.M.**, Lee, Y. S., Lu, D. S., Naini, B. V., Harlander-Locke, M. P., . . . Agopian, V. G. (2018). Propensity Matched Analysis of Patients with Mixed Hepatocellular-Cholangiocarcinoma and Hepatocellular Carcinoma Undergoing Liver Transplantation. *Liver Transpl*. doi:10.1002/lt.25058

Sho, S., **Court, C. M.**, Winograd, P., Russell, M. M., & Tomlinson, J. S. (2017). A prognostic mutation panel for predicting cancer recurrence in stages II and III colorectal cancer. *J Surg Oncol, 116*(8), 996-1004. doi:10.1002/jso.24781

Sho, S., **Court, C. M.**, Winograd, P., Lee, S., Hou, S., Graeber, T. G., . . . Tomlinson, J. S. (2017). Precision oncology using a limited number of cells: optimization of whole genome amplification products for sequencing applications. *BMC Cancer, 17*(1), 457. doi:10.1186/s12885-017-3447-6

Court, C. M., & Harari, A. (2017). Biomarkers used in the Screening, Diagnosis, Management, and Surveillance of Endocrine Cancers. In N. Lavin (Ed.), Manual of Endocrinology and Metabolism (5th ed.). Baltimore, MD: Lippincott Williams & Wilkins.

Sho, S., **Court, C. M.**, Kim, S., Braxton, D. R., Hou, S., Muthusamy, V. R., . . . Tomlinson, J. S. (2017). Digital PCR Improves Mutation Analysis in Pancreas Fine Needle Aspiration Biopsy Specimens. *PloS one*, 12(1), e0170897. doi:10.1371/journal.pone.0170897

Court, C. M., Ankeny, J. S., Sho, S., Hou, S., Li, Q., Hsieh, C., . . . Tomlinson, J. S. (2016). Reality of Single Circulating Tumor Cell Sequencing for Molecular Diagnostics in Pancreatic Cancer. *J Mol Diagn*, 18(5), 688-696. doi:10.1016/j.jmoldx.2016.03.006

Court, C. M., Ankeny, J. S., Sho, S., & Tomlinson, J. S. (2016). Circulating Tumor Cells in Gastrointestinal Cancer: Current Practices and Future Directions Gastrointestinal Malignancies (pp. 345-376): Springer.

Ankeny, J. S., **Court, C. M.**, Hou, S., Li, Q., Song, M., Wu, D., . . . Tomlinson, J. S. (2016). Circulating tumour cells as a biomarker for diagnosis and staging in pancreatic cancer. *Br J Cancer*, 114(12), 1367-1375. doi:10.1038/bjc.2016.121

Court, C. M., Ankeny, J. S., Hou, S., Tseng, H. R., & Tomlinson, J. S. (2015). Improving pancreatic cancer diagnosis using circulating tumor cells: prospects for staging and single-cell analysis. *Expert Rev Mol Diagn*, 15(11), 1491-1504. doi:10.1586/14737159.2015.1091311

Court, C. M., Cooper, S, Peterson, R.M. (2012). Should Texas provide public funding for bariatric surgery? (Master's thesis). Available from ProQuest Dissertations and Theses database (UMI: 1515585).

Chapter 1 Introduction

1.1 Overview

The work presented in this thesis focuses on developing molecular analysis of circulating tumor cells (CTCs) as a biomarker in gastrointestinal (GI) cancers. The work is divided into 4 main projects as follows:

Project 1 Using a combination of techniques, we developed and optimized GI tumor specific CTC assays for pancreatic and liver cancer and utilized our assays in prospective studies to confirm the accuracy of our method and study the potential utility of CTC enumeration as a biomarker (described in **Chapter 2**).

Project 2 We used modified versions of our CTC assay that allowed for CTC isolation and sequencing to compare the sequencing results of CTCs and primary tumors as a means of confirming the tumor origin of CTCs captured using our assay (**Chapter 3**).

Project 3 We studied the results of our single CTC sequencing in depth to determine the cellular limit of reliable CTC sequencing as well as optimized our single cell sequencing protocol for CTCs (**Chapter 4**).

Project 4 We used our optimized CTC sequencing protocol in a prospective study analyzing the feasibility and potential utility of CTC copy number alteration profiling in liver cancer (**Chapter 5**).

A schematic overview of the projects detailing the workflow and development of the projects is shown in **Figure 1**. The remainder of this chapter will provide an overview of gastrointestinal malignancies, circulating tumor cells, and the rationale behind our study.

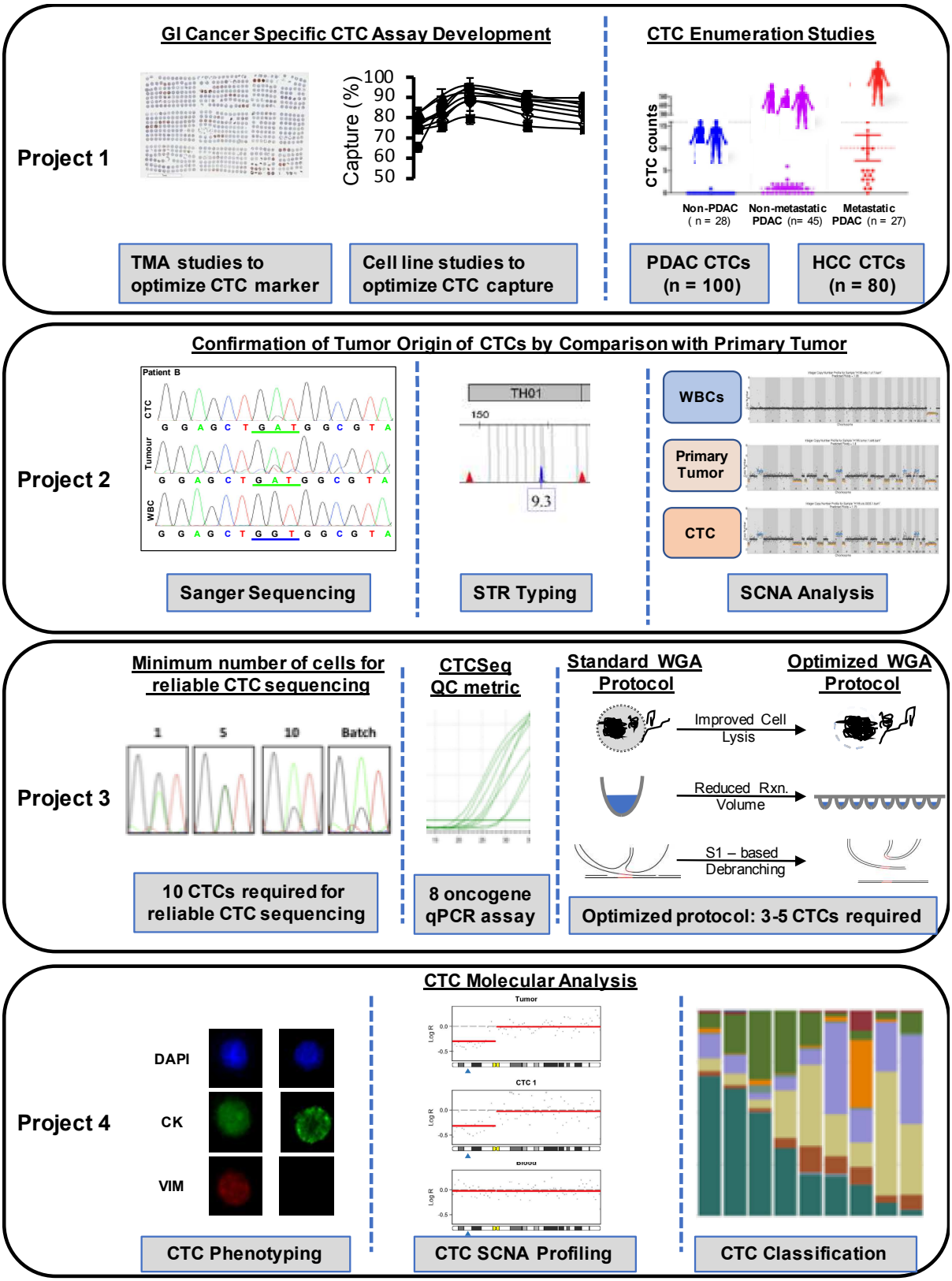


Figure 1. Graphical overview of the research described in the dissertation.

1.2 Gastrointestinal cancers

GI cancers are among the most prevalent cancers, and are the leading cause of cancer-related death, worldwide.[1] Due to their late presentation and aggressive nature, progress in GI malignancies has proven to be slower than for other major cancers. To date, the majority of the progress made in GI cancers has been due to the detection of lesions at an earlier stage through better screening programs or by treatment of cancer-related infectious disease.[2] Unfortunately, screening programs for GI cancers either don't exist, in the case of pancreatic and gastric cancer, or suffer from poor compliance, in the case of colon and liver cancer. Historically, the chemo-resistant nature of many GI cancers has hampered advances in life expectancy and treatment. However, the successes of new multidrug chemotherapeutic regimens and targeted therapies have the potential to make an impact in the treatment of advanced GI malignancies. Unfortunately, treatment response rates vary considerably between patients secondary to differences in tumor biology that most likely are reflected in the genetic profile of each individual tumor.[3] Furthermore, as patients live longer the relevance of the genetic makeup of their primary tumor has decreasing value, and additional biopsies are needed to adjust to the changing genomic landscape of the metastatic tumor deposits. Therefore, with the age of personalized and precision medicine seemingly approaching, there is a pressing need to evaluate the genetic makeup of tumors both at the time of diagnosis and longitudinally in response to therapy. CTCs theoretically have the biomarker attributes that are required to satisfy this need.

1.3 Circulating tumor cells

Circulating tumors cells (CTCs), which are thought to originate from primary tumors or metastatic sites, invade into a blood vessel or "intravasate", and then circulate in the blood (Figure 2).[4] Although this concept of circulating tumor cells was first introduced by Thomas Ashworth while investigating the blood of a widely metastatic breast cancer patient in 1869, it has only been over the past two decades that technological advances have allowed for their reliable capture and identification of these rare cells in the blood.[5] These advances have led to an explosion of research: A PubMed search for "circulating tumor cell" returned 16,070 publications, and a search of ClinicalTrials.gov resulted in 615 clinical trials, as of April, 2015. The first CTC platform to gain FDA approval was the CellSearch platform, which looks at CTC enumeration as a prognostic biomarker in metastatic lung, breast, and colon cancer.[4] Beyond simple capture and enumeration, second generation CTC platforms offer the ability to isolate CTCs for further testing in a manner similar to fine needle aspiration (FNA). Therefore, CTCs hold particular promise for GI cancers due to the difficulties involved in obtaining tumor tissue from visceral organs for the diagnosis, staging, prognosis, and management of these tumors. As a potential "liquid biopsy," CTCs may offer clinicians access to tumor tissue in a safe, convenient, and repeatable fashion from a simple peripheral blood draw.

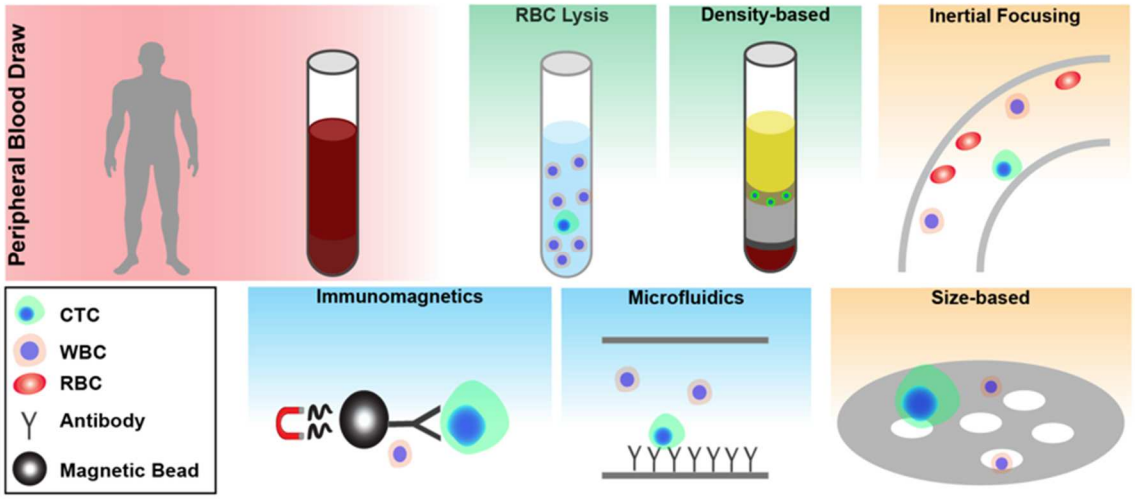


Figure 2. Methods for the enrichment and isolation of CTCs. **Peripheral Blood Draw:** Peripheral blood is obtained from patients, usually 2-20 mL. **RBC lysis:** involves incubation of whole blood with a mixture of ammonium chloride, potassium carbonate, and EDTA resulting in lysis of RBCs but not WBCs or CTCs. **Density-based:** density gradient centrifugation uses a density medium to separate mononuclear cells from RBCs and granulocytes based on cell density. **Inertial focusing:** Using a spiral device, cells are separated by size based on different flow patterns due to inertial microfluidics. **Immunomagnetics:** Antibodies are bound to magnetic beads allowing for the capture of CTCs as well as their subsequent manipulation. **Microfluidics:** Antibody-coated nanofabricated microfluidic channels use various methods to ensure cell-antibody interactions, allowing for the capture and manipulation of CTCs. **Size-based:** CTCs are generally larger than RBCs and WBCs and can be trapped by filtration on a micropore membrane. CTC = circulating tumor cell, WBC = white blood cell, RBC = red blood cell, EDTA = Ethylenediaminetetraacetic acid.

1.3.1 Enrichment and detection of CTCs

The difficulties involved in isolating rare CTCs have led to a vast array of different platforms for the enrichment and detection of CTCs. One recent review found over 50 unique techniques, more than 10 of which were first described in the last 5 years. [6] The rarity of CTCs, representing just a few cells in the millions of white blood cells and billions of red blood cells per milliliter of blood, is the primary difficulty that CTC detection methods must address. Most platforms involve a combination of different techniques to first enrich the CTC population against that of the RBC and WBC background followed by identification and isolation of CTCs. Methods used to both enrich and identify CTCs can broadly be broken down into those that utilize physical properties and those that utilize biological ones (**Figure 2**). Physical properties

such as size, density, and stability are commonly used. Additional properties including electrical charge, optical characteristics, photoacoustics, and deformability have also been studied. In general, enrichment of CTCs by physical properties alone is fast and relatively inexpensive, but does not allow for the level of enrichment seen with biological properties. Thus, many technologies use an initial physical enrichment before more specific biologic enrichment and detection is performed. Biologic properties, including protein expression, as well as DNA and RNA signatures, offer highly specific enrichment and isolation of CTCs. In particular, methods utilizing cell surface markers for enrichment and capture of CTCs are currently the most common. Of the cell surface label-based techniques, the majority use antibodies against the epithelial marker EpCAM to positively select the epithelial cells. While highly specific, some researchers have shown that the use of these anti-EpCAM antibodies results in the loss of CTCs that do not express EpCAM, especially CTCs that have undergone epithelial to mesenchymal transition (EMT).[7] Therefore, newer technologies use tumor specific or mesenchymal cell surface markers to allow for the enrichment and capture of these additional CTC subtypes.

Following enrichment, the identification of CTCs must be confirmed. Techniques for doing so can broadly be broken down into visual and molecular categories. The visual identification of CTCs by immunocytochemistry (ICC), which combines cytometry and immunofluorescence, is the most common technique used today. The most common visual definition of a CTC in use today is based on 3-channel immunofluorescence. It utilizes a nuclear stain (usually DAPI), cytokeratins (CK) as an epithelial marker, and CD45 as a hematopoietic marker (**Figure 3**). Thus, a Nuclear+/CK+/CD45- cell is a CTC whereas a Nuclear+/CK-/CD45+ cell is a WBC. While these categories are widely accepted, the discovery of Nuclear+/CK+/CD45+ cells, with both epithelial and hematopoietic markers, by many different

CTC platforms has led to some uncertainty in the ICC definition of a CTC. Studies to date seem to indicate that these cells may represent nonspecifically stained macrophages or polymorphonucleocytes, or occur due to technical antibody processing errors.[8]

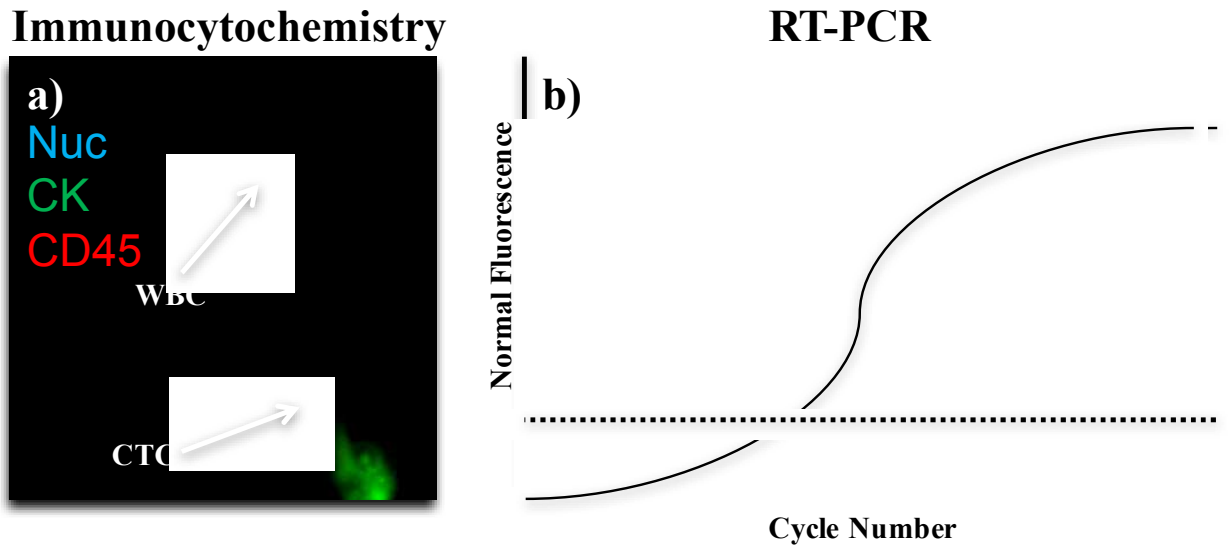


Figure 3. Traditional CTC identification and enumeration methods. **a) ICC:** Representative image of immunocytochemical staining obtained from the NanoVelcro platform from a patient with pancreatic cancer. CTCs are distinguished from other mononuclear cells based on differential immunostaining and cytologic characteristics. The blue nuclear stain (DAPI) identifies all nucleated cells. The epithelial marker cytokeratin is identified by green fluorescence (AlexaFluor 488) on CTCs while the hematopoietic marker CD45 is identified by red fluorescence (AlexaFluor 555) on WBCs. **b) RT-PCR:** The figure shows a representation of a readout from a real-time quantitative RT-PCR study. The PCR reaction is specific for a CTC associated mRNA transcript (e.g. CEA or hTERT). The graph plots the PCR cycle number versus the normalized relative fluorescent signal emitted at a given cycle count. In these types of studies, thresholds are calculated based on the baseline variance, and CTC positivity is defined as fluorescence greater than that threshold at a predetermined cycle count. If the fluorescent signal is greater than the calculated threshold at a cycle count lower than the cutoff, then the sample is positive for a CTC-associated transcript. ICC = Immunocytochemistry, CTC = circulating tumor cell, WBC = white blood cell, Nuc = nuclear stain, CK = cytokeratin, DAPI = 4',6-diamidino-2-phenylindole, RT-PCR = reverse transcription polymerase chain reaction.

Advances in molecular biology have made molecular detection techniques very sensitive, allowing for the detection of a single CTC in a background of thousands to hundreds of thousands of WBCs. The most common methods in use today are reverse transcriptase polymerase chain reaction (RT-PCR) based technologies that identify CTCs based on the expression of tumor-associated mRNA transcripts. Unfortunately, RT-PCR based techniques are

not as specific as ICC due to illegitimate transcription by WBCs that can result in tumor-associated transcripts being found in the blood of normal patients.[9] Therefore, the only truly tumor specific molecular biomarkers are those looking at gene fusion products or ubiquitous driver mutations. While initial studies were limited to RT-PCR of a single transcript, advances in multiplexing and single cell sequencing are increasingly being explored. In fact, with the cost of sequencing technology rapidly decreasing, the possibility of routine sequencing of CTCs is quickly becoming a reality.[10]

1.3.2 Clinical implications of CTCs

Despite the large body of research on CTCs, basic questions about the detection, isolation, enumeration, meaning, and clinical utility of CTCs remain to be answered. Additionally, important fundamental questions about the source of CTCs and their importance in tumor biology, especially the metastatic cascade, are yet to be definitively answered. CTCs are very heterogeneous, both between patients and among CTCs of a single patient, expressing markers of both epithelial and mesenchymal origin.[11] While studies have proven the viability of some CTCs through culture and functional assays, many CTCs appear to be apoptotic.[12,13] Additionally, their half-life appears to vary considerably between studies, from hours to weeks, in studies of patients with resected tumors. Perhaps the most important unanswered question is why CTCs do not correlate directly to the presence of metastasis. For example, patients with peritoneo-venous shunts do not immediately develop widespread metastasis despite large numbers of circulating tumor cells.[14] One possibility is that CTCs demonstrate considerable metastatic inefficiency, likely due to the issues outlines above. In fact, studies have demonstrated that only 2.5% of CTCs can form micrometastases and only 0.01% of them can form

macrometastases.[15,16,17] Thus, CTCs as a whole represent tumor cells that have accomplished the metastatic step of intravasation; however, whether these cells have the ability to extravasate from the vascular system and seed a metastatic distant metastatic colony is not clear. The isolation of the subset of CTCs with metastatic potential remains an active area of research.[9] One relevant area of research has been the study of intraoperative CTC counts to determine if surgical technique may affect the risk of metastasis following surgery. Studies have found that pressure, biopsies, and other invasive manipulations during surgery can transiently increase CTC counts up to 66-fold.[18] However, despite the development of "touchless" techniques that decrease the shedding of CTCs into the blood stream, studies have not consistently demonstrated an associated decrease in survival.

1.4 CTCs as a biomarker in gastrointestinal cancers

Studies to date have found several promising clinical applications for CTCs in GI cancers. As a biomarker in GI cancers, CTCs have a very high specificity coupled with a lower sensitivity. Their high specificity, equivalent to that of an FNA, may eventually allow CTCs to function as a liquid biopsy. However, several important limitations have been found. While false positive CTCs are rarely found in healthy controls, circulating epithelial cells have been detected in patients with benign diagnoses such as cystic pancreatic lesions, inflammatory bowel disease, and cirrhosis.[19,20] While newer molecular techniques may allow for the confirmation of tumor origin of circulating epithelial cells, studies on patients with inflammatory and cystic diseases are yet to be performed. Additionally, GI cancers generally have lower CTC counts than other malignancies. The filtering or "first pass" effect of the liver is thought to be the culprit, and this theory has been demonstrated in a mouse model.[21] Additional support for this theory comes from

several studies that found higher CTC counts in portal blood versus systemic blood in patients during surgery.[22] Currently, the only FDA-approved CTC application in GI cancers is the enumeration of CTCs as a prognostic biomarker in patients with colorectal cancer. However, the success of recent clinical studies makes the approval of CTCs as a biomarker in other GI cancers likely.

1.4.1 CTC molecular analysis

The most useful potential application of CTCs in GI cancers would be as a "liquid biopsy." If realized, it would allow clinicians access to tumor tissue in a real time, repeatable, safe, and cost-effective manner. While true for all cancers, such a liquid biopsy holds particular promise in GI cancer due to the difficulty and cost involved in obtaining tumor tissue by current endoscopic and percutaneous techniques. For pancreatic cancer in particular, the combination of anatomic location and the relative paucity of tumor cells within a desmoplastic stromal background makes traditional biopsies difficult, often requiring multiple endoscopies to obtain useful tumor tissue.[23] Furthermore, the separation of the few tumor cells from the dense stromal background for further molecular characterization is difficult and not routinely performed.[24] Therefore, with regards to molecular analysis in PDAC, CTCs may actually perform better than EUS-FNA due to the enrichment of tumor cells during CTC isolation. Furthermore, the difficulties of performing molecular analysis on the low number of tumor cells obtained by CTC assays have mostly been solved by recent developments in the field of single cell sequencing.[25] The development of single cell sequencing has realized the goal of having CTCs function as a liquid biopsy due to the increased specificity that molecular analysis can provide. A single cell contains approximately 4 picograms of DNA while the input requirements for most next

generation sequencing platforms is on the order of nanograms to micrograms.[26] Techniques for bridging the gap between these amounts are now available allowing for the whole genome amplification (WGA) of single cells to levels required for next generation sequencing (NGS). A number of studies have demonstrated that not only can CTCs act as a liquid biopsy, providing diagnostic accuracy comparable to that of an FNA, but they also have the potential to offer additional biologically relevant information about tumor heterogeneity and metastatic potential.[25,27-29] Furthermore, in some cases CTCs may actually provide more accurate molecular information than a traditional biopsy due to the enrichment of an important subset of tumor cells against stromal contaminants and less aggressive tumor subclones.[30] Both single cell RNA sequencing (RNAseq) and DNA sequencing (DNAseq) have advanced rapidly over the past 5 years, and targeted cancer gene panel sequencing is now routinely used in clinical practice.[31,32] Larger exome and genome sequencing studies that were once prohibitively expensive for even small numbers of patients, let alone the numbers required for a clinical trial, are now possible due to the exponential decrease in the cost of sequencing. At the same time, projects like The Cancer Genome Atlas (TCGA) and the Catalog of Somatic Mutations in Cancer (COSMIC) database provide researchers with the information they need to parse the large datasets that result from next generation sequencing. Combined, these developments have introduced a vast array of new molecular targets with prognostic and predictive importance.

1.5 Conclusion

CTC sequencing has the potential to be a liquid biopsy for GI cancers, acting as both a prognostic biomarker at the time of diagnosis as well as a longitudinal assay to track progression and the development of actionable genetic alterations. In this thesis, we will demonstrate our

work on developing and implementing such an assay. In *Project 1*, we develop and validate tissue-specific CTC assays for PDAC and HCC as well as demonstrate the utility of CTC enumeration as a diagnostic, staging, and prognostic biomarker. In *Project 2,* we validate our ICC definition of a CTC using molecular analysis and confirm the tumor origin of the CTCs we isolate using the CTC assays developed in *Project 1*. In *Project 3*, we investigate the number of CTCs required for reproducible and reliable CTC sequencing as well as develop an optimized CTC sequencing protocol that allows for reliable results from as few as 3-5 CTCs. Finally, in *Project 4*, we conduct a pilot study of CTC sequencing in HCC and demonstrate the potential utility of such an assay.

Chapter 2 Development and prospective trial of tissue-specific CTC assays for GI cancers.

2.1 Overview

Historically, the yield of CTC assays for GI cancers has been among the lowest for all cancer types.[4] While the precise mechanism is unknown, it is thought that the filtering effect of the liver may be partially responsible.[21,33] Additionally, the expression of epithelial cell-surface adhesion molecule (EpCAM), the most common antibody used for CTC isolation, is variably expressed in GI cancers, from nearly universal in the case of PDAC to less than 30% for HCC and neuroendocrine tumors.[34,35] Newer microfluidic techniques for CTC isolation have consistently resulted in higher CTC yields, and the use of cell surface markers other than EpCAM have also shown promise in EpCAM-low tumors.[33] Similarly, while cytokeratin (CK) has traditionally been the CTC identification marker of choice, the use of additional markers has been shown to aid in reducing false positives and provide additional prognostic information.[30,36]

Given the heterogeneity of prior results using general CTC assays, we sought to develop CTC assays tailored to each tissue-type individually. Our hope was that using such a tissue-specific approach would aid in both increasing the accuracy and performance of our assay as well as providing additional phenotypic information that could be used for prognostication. Using a combination of tissue microarray (TMA) staining studies, cell line experimentation, and literature review, cell surface and cytoplasmic markers of potential utility were identified and incorporated into our tissue-specific CTC assays based on the microfluidic NanoVelcro CTC platform.[37] We used this methodology for 2 tissue types: PDAC (**Project 2.2 and 2.3**) and HCC (**Project 2.4**). For PDAC we incorporated CEA, expressed by the majority of PDAC tumors,[38] as

an additional marker for CTC identification. For HCC, we used a multimarker CTC capture cocktail consisting of antibodies against the HCC-specific cell surface markers asialoglycoprotein receptor (ASGPR) and glypican-3 (GPC-3) in addition to EpCAM. Additionally, we utilized vimentin as a phenotypic marker of mesenchymal transformation that proved to be independently prognostic of both survival and recurrence for HCC patients.

2.2 Circulating tumor cells as a diagnostic and staging biomarker in pancreatic cancer

2.2.1 Abstract

Background: Current diagnosis and staging of pancreatic ductal adenocarcinoma (PDAC) has important limitations, and better biomarkers are needed to guide initial therapy. We investigated the performance of circulating tumor cells (CTCs) as an adjunctive biomarker at the time of disease presentation. **Methods:** Venous blood (VB) was collected prospectively from 100 consecutive, pre-treatment patients with PDAC. Utilizing the microfluidic NanoVelcro CTC chip, samples were evaluated for the presence and number of CTCs. KRAS mutation analysis was used to compare the CTCs with primary tumor tissue. CTC enumeration data was then evaluated as a diagnostic and staging biomarker in the setting of PDAC. **Results:** We found 100% concordance for KRAS mutation subtype between primary tumor and CTCs in all 5 patients tested. Evaluation of CTCs as a diagnostic revealed the presence of CTCs in 54/72 patients with confirmed PDAC (sensitivity=75.0%, specificity=96.4%, area under the curve (AUROC) = 0.867, 95%CI = 0.798-0.935, $p<0.001$). Furthermore, a cut-off of $\geq$3 CTCs in 4mL VB was able to discriminate between local/regional and metastatic disease (AUROC = 0.885; 95%CI = 0.800–0.969; $p<0.001$). **Conclusion:** CTCs appear to function well as a biomarker for diagnosis and staging in PDAC.

2.2.2 Introduction

Pancreatic ductal adenocarcinoma (PDAC) is currently the 4th most common cause of cancer death in the United States, and is projected to be the second most common cause by

2030[39,40]. Its incidence approximates its mortality secondary to a 5-year survival rate of less than 6%.[39] These poor outcomes are undoubtedly related to a late presentation coupled with an aggressive biology.[41]

Initial diagnosis of PDAC typically utilizes tissue acquisition via image-guided fine needle aspiration (FNA), most commonly by endoscopic ultrasound (EUS).[42] Unfortunately, these methods are expensive, inconvenient, require local expertise, and are not without risk to the patient.[43] Additionally, for PDAC tumors, the abundance of stromal elements relative to cancer cells can lead to non-diagnostic biopsies. Thus, multiple needle passes and sometimes even repeated EUS/FNA is necessary in order to obtain adequate tissue for diagnosis.[44] Clinical staging is based on cross-sectional imaging (CT or MRI) which does not have the necessary sensitivity to detect small volume metastatic disease leading to routine under-staging. As evidence, 10-25% of patients who are initially believed to be resectable based on imaging are found to have metastases at surgical exploration.[45] Moreover, approximately 80% of patients who undergo successful surgery will experience distant cancer recurrence, reflecting the presence of metastatic disease at the time of surgery.[46] Retrospectively, surgery may not be the appropriate initial therapy in these occult metastatic patients as it often delays administration of systemic therapy.[47] These data highlight the need for a biomarker that could improve the diagnosis and accuracy of staging at the time of disease presentation to better inform first line therapy.

Circulating tumor cells (CTCs) are one promising biomarker that may be useful for these purposes in PDAC. While CTCs have been studied in PDAC, they are not as well established as a biomarker in PDAC as compared with other solid cancers.[48,49] One reason may be the low sensitivity of current technology in detecting CTCs from peripheral blood in PDAC. Initial

studies using the Food and Drug Administration (FDA) approved CellSearch™ assay have confirmed the presence of CTCs in patients with PDAC, but in significantly lower overall numbers[49] when compared to other epithelial cancers.[50] Additionally, it is unclear if CTCs identified across various studies are truly cells of tumor origin.[30] In PDAC, mutations in *KRAS* are among the most common found in any cancer and occur with an incidence of greater than 95%.[51] Moreover, 98% of the KRAS mutations found in PDAC are located in codon 12. Given the nearly uniform occurrence of KRAS mutations, tumor origin of captured CTCs can be inferred for most patients by sequencing a single gene. Unfortunately, the CellSearch™ assay does not conveniently enable comprehensive molecular analysis of captured CTCs to confirm tumor origin or further inform the biology of metastasis.

Sensitivity has been improved with the development of new CTC detection and enumeration platforms.[30] The NanoVelcro platform utilizes anti-EpCAM-coated nanosubstrates in conjunction with microfluidic chaotic mixers to improve CTC capture and identification.[52,53] One advantage of the NanoVelcro platform is that it allows for the addition of tumor identification markers, such as CEA for PDAC.[38] Additionally, NanoVelcro allows for seamless integration with Laser Capture Micro-dissection (LCM) for single CTC isolation, referred to as NanoVelcro/LCM.[31] The individually isolated CTCs can then be subjected to downstream molecular analyses, such as Sanger sequencing or next generation sequencing (NGS).

In the present study, our goal was to evaluate CTCs as a biomarker for diagnosis and staging of PDAC at the time of disease presentation. To do this, we first developed a sensitive and specific method for PDAC CTC detection using the NanoVelcro platform in conjunction with high-resolution fluorescent microscopy and a multi-color immunocytochemistry (ICC) approach. We validated our CTC definition using NanoVelcro/LCM by mutational analysis of

KRAS codon 12, demonstrating concordance of *KRAS* mutation subtype between CTCs and primary tumor tissue (See **Chapter 4**). We then conducted CTC enumeration on 100 consecutive patients with suspicious pancreatic lesions or recent PDAC diagnosis and correlated these data with eventual diagnostic, pathologic, and staging information. Together, our results highlight the utility of CTCs as a liquid biopsy to better inform diagnosis and staging of PDAC, importantly, at the time of disease presentation.

Table 1. Characteristics of PDAC cell lines used for NanoVelcro Chip calibration, efficiency assessment, and feasibility of single cell mutational analysis.

PDAC Cell Line	Source	Differentiation	EpCAM Expression	*KRAS* Mutation	Zygosity
CFPAC-1	Metastasis	Well	High	G12V	Heterozygous
BxPC-3	Primary	Moderate	Moderate	WT*	WT*
PANC-1	Primary	Poor	Low	G12D	Heterozygous
AsPC-1	Metastasis	Poor	Moderate	G12D	Homozygous

*WT = Wild-type

2.2.3 Materials and methods

PDAC cell lines and cell culture

All PDAC cell lines were purchased from American Type Culture Collection (ATCC). Phenotypic information regarding the 4 PDAC cell lines (CFPAC-1, BxPC-3, PANC-1, and AsPC-1) can be found in (Table 1). BxPC-3 and AsPC-1 were maintained with RPMI medium (ATCC) supplemented with 10% heat-inactivated FBS (Omega Scientific), 2 mmol L^{-1} glutamine, and 1% penicillin-streptomycin (Gibco). PANC-1 was maintained in Dulbecco's Modified Eagle Medium (DMEM) containing 10% heat-inactivated FBS, 2 mmol L^{-1} glutamine, and 1% penicillin-streptomycin. CFPAC-1 was maintained in Iscove's Modified Dulbecco's Medium (IMDM) also containing 10% heat-inactivated FBS, 2 mmol L^{-1} glutamine, and 1%

penicillin-streptomycin. All cells were cultured in a humidified, 5% CO_2 incubator at 37°C and were proliferated according to reported literature and ATCC recommended procedures. Cells were detached from plates using 0.2 μm filtered 1X sodium citrate solution.

NanoVelcro calibration and capture efficiency assessment

For calibration and capture efficiency assessment, 4 different PDAC cell lines (CFPAC-1, BxPC-3, PANC-1, AsPC-1) with various cellular phenotypes were utilized (**Table 1**). Cells were lifted from culture plates using a 1X sodium citrate solution. After washing with the corresponding culture medium, each cell line was diluted to various densities (~20, 100, 500, and 1000 per 100 μL) that were verified by counting 100 μL volumes of each cell dilution on 3 48-well plates and taking the average number as the cell density. The various cell mixtures were treated with fluorescent DiO (Life. Tech.), and then fixed using 4% paraformaldehyde (PFA) for 10 minutes at room temperature. For initial calibration, 200 cells from each cell line were spiked into 1 mL healthy donor blood and were processed on the NanoVelcro platform at various flow rates (0.1, 0.5, 1.0, 2.0, and 3.0 mL/hour) in triplicate to determine the optimum flow rate for capture efficiency (**Figure 4**). A flow rate of 1.0 mL/hour was found to be the optimum flow rate. We then created several artificial samples from each cell line containing approximately 20, 100, 500, and 1000 PDAC cells spiked into 1 mL healthy donor blood. After processing on NanoVelcro Chips at the optimum flow rate of 1.0 mL/hour, captured cell were fixed by 4% PFA, stained with a DAPI mounting solution (Molecular Probes, Life Technologies), and enumerated under an automatic imaging system. The imaging system was composed of an upright fluorescent microscope (Eclipse 90i, Nikon) with the NIS-Element 4.1 imaging software (Nikon), a precision motorized stage (ProScan II system, Prior Scientific), and a fluorescent light

source (SPECTRAX, Lumencor). After automatically scanning through the entire substrate under 40x magnification, all DAPI+/DiO+ events were automatically selected, and a manual confirmation followed to determine the exact number of cancer cells on the substrate. The capture efficiencies were calculated as the number of cancer cells captured on the substrate divided by the total number of cells loaded into the chip. Results of the capture efficiency studies are shown (**Figure 4**).

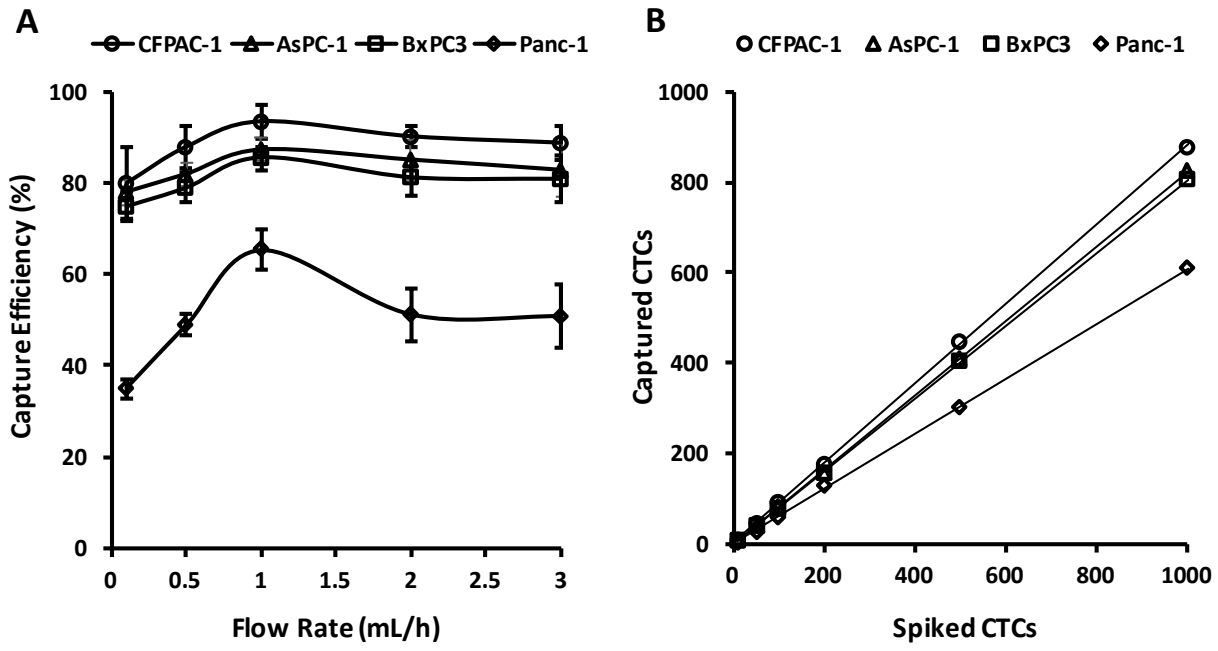


Figure 4. Calibration curves and capture efficiency data for PDAC cell lines. **(A)** Diagram of capture efficiency for CFPAC-1, AsPC-1, BxPC-3, and PANC-1 as a function of NanoVelcro device flow rates. **(B)** Capture efficiency of CFPAC-1, AsPC-1, BxPC-3, and PANC-1 as a function of the number of captured cells versus the number spiked into the artificial sample.

Materials for construction of silicon nanowire substrates (SiNS)

Oriented Silicon wafers (p-type, resistivity of ca. 10-20 Ωcm) were obtained from Silicon Quest International. 4-well Lab-Tek™ Chamber Slides were purchased from Thermo Fisher Scientific. Sulfuric acid (98%), Hydrogen peroxide (30%), Silver Nitrate (>99.8%), 4-Maleimidobutyric acid N-hydroxysuccinimide (≥98% HPLC), Hydrofluoric acid (48%), ethanol

(>99.5%), and 3-mercaptopropyl trimethoxysilane (95%) were purchased from Sigma-Aldrich Co.

Construction of SiNS NanoVelcro chips

The components of a NanoVelcro chip include a chaotic mixing chip made of polydimethylsiloxane (PDMS), and a patterned silicon nanowire substrate (SiNS) SiNS). The chaotic mixing chip was fabricated using a soft lithography method. The patterned silicon mold was produced by a standard two-step photolithographic process. In the first step, a thin layer (100 μm) of negative photoresist (SU8-2100, MicroChem Corp., Newton, MA, USA.) was spin-coated onto a 3-inch silicon wafer (Silicon Quest, San Jose, CA, USA). After UV exposure and development, a serpentine fluidic channel with a rectangular shape in the cross section was obtained (length 800 mm and width 1.0 mm). In the second step, another layer (35 μm) of negative photoresist (SU8-2025, MicroChem Corp.) was spin-coated onto the same wafer. Prior to UV exposure, the second mask was aligned (Karl Suss America Inc., Waterbury, VT, USA.) to ensure a good alignment between the previous pattern and the pattern to be fabricated. The new pattern created the ceiling "ridges" that induce chaotic mixing within the fluid channel. These steps produced the positive mold for fabrication of the PDMS chaotic mixer. The mold was then exposed to trimethylchlorosilane (TMSC) vapor for 2-3 minutes and transferred into a Petri dish. A well-mixed PDMS pre-polymer (GE Silicones, Waterford, NY, USA.; RTV 615 A and B in 10 to 1 ratio) was poured onto the surface of the mold to replicate the pattern, producing a chip of approximately 6 mm thick. After oven curing at 80^{o}C for 48 hours, the PDMS chip was ready to use. After the PDMS chip was peeled off from the mold, two through-holes were punched at both ends of the fluidic channel for connection of the tubing.

Patterned SiNS was fabricated in parallel by standard photolithography and a chemical wet etching process. First, a thin film photoresist (AZ 5214, AZ Electronic Materials USA Corp., Branchburg, NY, USA.) was spin-coated onto a silicon wafer (Silicon Quest, San Jose, CA, USA). After UV exposure and development, the silicon wafer covered by patterned photoresist was ready for the chemical etching process. This etching process has been reported previously (Peng et. al., *Adv. Mater.* 2002, *14*, 1164.) Briefly, an etching solution composed of deionized water, HF, and silver nitrate was prepared in a Teflon vessel. The concentrations of HF and silver nitrate were 4.6 and 0.02 M, respectively. The length of the nanowires was mainly determined by the duration of the wet etching process. In this manuscript, the etching conditions were optimized to produce approximately 25 μm nanowire structures on the silicon substrate. Then, the substrate was immersed in boiling aqua regia (3:1 (v/v) HCl/HNO_3) for 15 min to remove the silver film. The patterned photoresist on the silicon substrate was removed with acetone and ethanol. After washing with DI water and drying by nitrogen, the silicon substrate with patterned nanowire structures became ready for subsequent surface modification. The surface of the silicon substrate was modified with 4 % (v/v) 3-mercaptopropyl trimethoxysilane in ethanol at room temperature for 45 minutes, and then by a treatment of the coupling agent N-maleimidobutyryloxy succinimide ester (GMBS, 0.25 mM in DMSO) for 30 minutes. The substrate surface was then coated with streptavidin (SA, 10 μg/ml in 1×PBS) at room temperature for 30 minutes. Excess chemicals from each step were washed away using appropriate solvents or buffer solutions.

Patient recruitment and sample processing

Consecutive pre-treatment patients with either suspicion for, or recently diagnosed, PDAC were approached regarding participation in our study between December 2012 and

December 2014 and enrolled in the study under University of California, Los Angeles IRB#11-002112. Inclusion criteria were suspicious cystic or solid pancreatic lesions based on cross-sectional imaging or recent diagnosis of PDAC via biopsy, as well as willingness to give informed consent. Exclusion criteria included previous PDAC-directed treatment of any kind (surgical or non-surgical), active inflammatory bowel disease, synchronous malignancies, or other malignancy in the past 5 years. After a 5 mL waste to prevent epithelial cell contamination, 10 mL venous blood (VB) was taken from a peripheral vein. 4 mL was processed in parallel 2 mL samples for enumeration studies and 6 mL was cryopreserved. CA19-9 values were recorded when available. A database with demographic, pathologic, and relevant clinical outcome/survival variables was maintained in a prospective manner. All PDAC diagnoses were confirmed via biopsy or surgical pathology. Non-adenocarcinoma diseases were confirmed via biopsy or surgical pathology when possible. In some cases, intraductal papillary mucinous neoplasm (IPMN) was diagnosed clinically using a combination of cross-sectional imaging appearance, EUS characterization, lack of malignant cells on FNA, and cyst fluid analysis.

Patient sample/NanoVelcro chip preparation

In each 4 mL VB sample, red blood cells (RBCs) were lysed using a 0.15 M Tris-ammonium chloride solution. Cold phosphate buffered saline (PBS) (Gibco) was then added for termination of lysis and samples were subsequently centrifuged at 300 x g for 10 minutes at 4^{o}C. After removal of supernatant, cell pellets were re-suspended in cold RPMI-1640 (Cellgro) supplemented with 5% cell-free foetal bovine serum (FBS) (Gibco). Following a second centrifugation and removal of supernatant, cell pellets were re-suspended in PBS+2% donkey serum (DS) (Jackson Immunoresearch). Biotinylated goat anti-EpCAM (R&D Systems) was

added to the samples and allowed to incubate at room temperature for 30 minutes with gentle agitation to prevent cell clumping. Following antibody incubation, samples were washed in PBS and re-suspended in 400-µL PBS for chip loading. All experiments were carried out on parallel duplicate chips each running 200-µL of the re-suspended sample. NanoVelcro was assembled and operated as outlined in the supplemental materials and methods and as previously described (**Figure 5**) [31]. The prepared samples were injected into the device at a flow rate of 1.0 mL/h, followed by 4% paraformaldehyde (PFA) at the same rate for fixation.

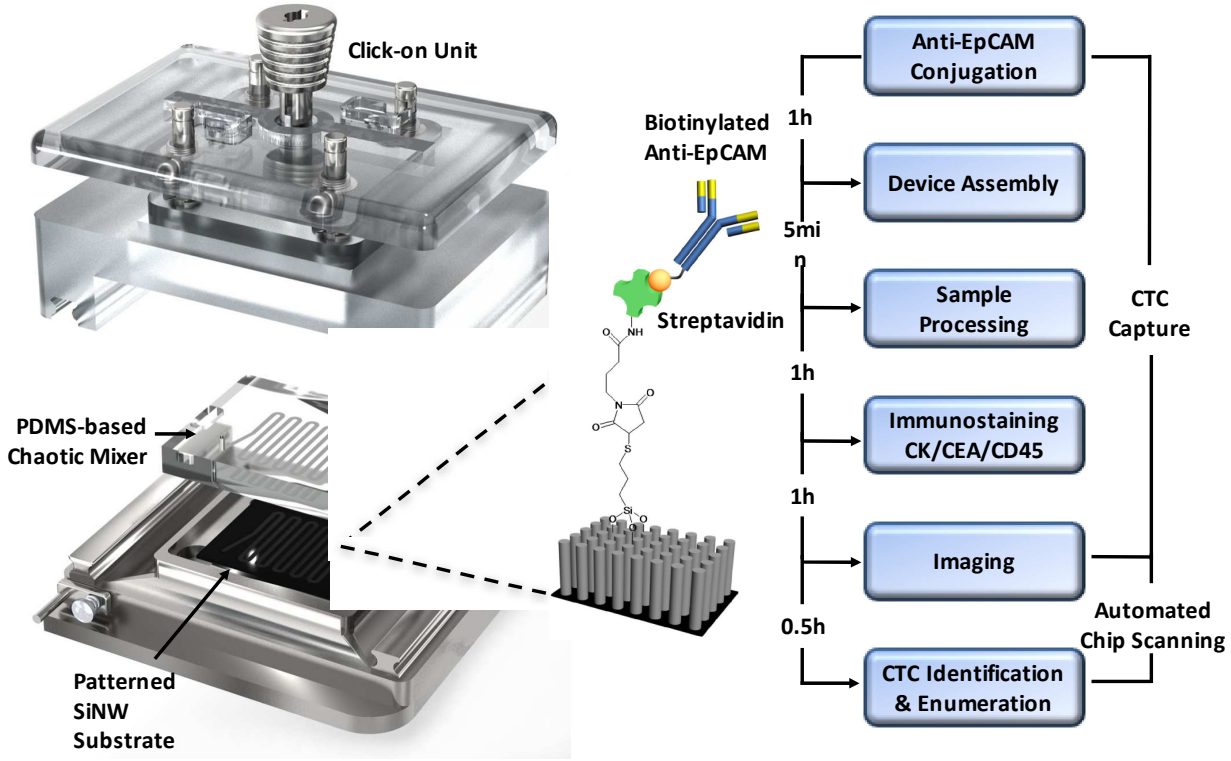


Figure 5. General workflow diagram for CTC capture, identification, and enumeration on SiNS NanoVelcro Chips. The schematic shows the individual components of SiNS NanoVelcro Chips, chip holding device, and elements/duration of the workflow for processing a typical blood sample.

Immunocytochemistry (ICC) and chip mounting

Following disassembly of the device, processed chips were washed in PBS for 15 minutes. Chips were blocked and cells permeabilized using PBS+2%DS+0.1% Triton X-100

(Sigma) for 15 minutes. Chips were then incubated with a cocktail of primary antibodies containing two rabbit anti-pancytokeratin (CK) antibodies (Invitrogen, Abcam), one chicken anti-CEA antibody (Abcam), and two mouse anti-CD45 antibodies (BD Pharmigen, Abcam) for 1 hour at room temperature in PBS+2%DS. Following primary antibody incubation, chips were washed again in PBS. Secondary antibody incubation was carried out in PBS+2%DS for 1 hour at room temperature using a cocktail of AlexaFluor-488 donkey anti-rabbit (Invitrogen), AlexaFluor-647 goat anti-chicken (Invitrogen), and AlexaFluor-555 donkey anti-mouse (Invitrogen). Chips were again washed in PBS and then attached to microscope slides with adhesive and mounted with cover slides using a DAPI mounting solution (Molecular Probes, Life Technologies). Mounted slides were allowed to dry for 1 hour prior to chip fluorescent scanning.

Chip scanning and CTC enumeration

Automated chip scanning with a Nikon Eclipse90i fluorescent microscope utilizing NIS Elements4.1 software was performed at 40x power to identify candidate cells. Higher (400x) power manual imaging of candidate cells was then performed, and the results plotted on a fluorescence scatter plot to verify identity and count. When viewing, WBCs were defined as round/ovoid, DAPI+/CD45+/CK-/CEA-. CTCs were defined as round/ovoid, size≥6-μm, DAPI+/CD45-, and CK+ or CEA+ (**Figure 6B**). Importantly, any CD45 positivity visible above background discounted a cell as being a CTC. Final CTC counts are represented as a total count per 4 mL VB. CTC enumeration was carried out by the same, blinded researcher.

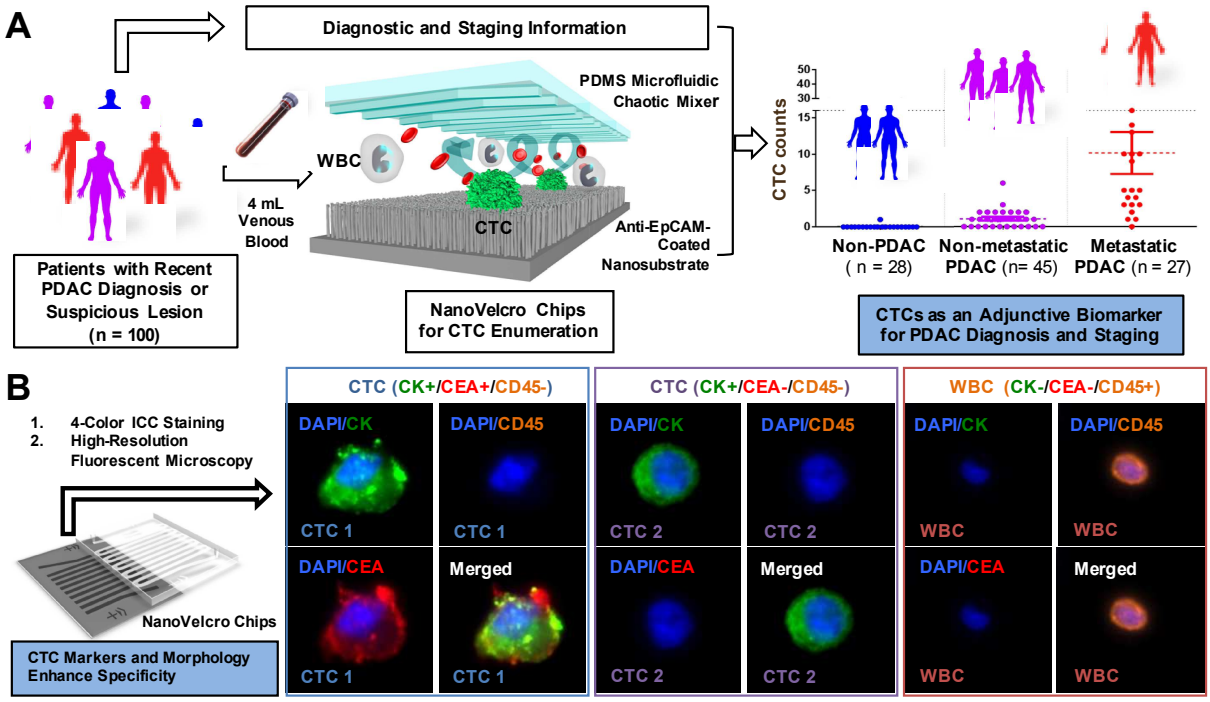


Figure 6. Study design and CTC detection principles. **(A)** Flow diagram depicting venous blood draws from our study cohort followed by CTC enumeration on NanoVelcro Chips and subsequent correlation with diagnostic and staging information. **(B)** Schematic depicting CTC identification via a 4-color ICC approach in conjunction with high-resolution fluorescent microscopy. Representative images of 2 common CTC and WBC staining patterns are shown at 400x magnification.

Statistical methods

All CTC numbers are reported as whole numbers in 4-mL of blood. Additional outlier analysis was performed with the Iterative Grubb's method (α=0.01). Differences in CTC number or CA19-9 level between non-adenocarcinoma and PDAC patients and those within the PDAC cohort were evaluated with a Mann-Whitney U-test given non-normal distribution. Comparisons across more than 2 groups were evaluated with a Kruskal-Wallis one-way ANOVA and the multiple comparisons test. Diagnostic performance of CTCs was evaluated using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and positive likelihood ratio (+LR) calculations in addition to the use of receiver operating characteristic curves (ROCs) for determination of the area under the curve (AUROC) and overall

discriminatory ability. A two-tailed p-value<0.05 was considered to be statistically significant. All statistical manipulations and calculations were performed with the assistance of GraphPad Prism6.0 (GraphPad Software).

Table 2. Clinical and pathologic characteristics associated with CTC counts.

Characteristic	Number (n)	0 CTC	≥1 CTC	≥2 CTC	≥3 CTC	≥5 CTC
Adenocarcinoma	72	18(25.0%)	54 (75.0%)	39 (54.2%)	29 (40.3%)	18 (25.0%)
Stage						
I	3	3 (100%)	0	0	0	0
II	28	11 (39.3%)	17 (60.7%)	9 (32.1%)	2 (7.1%)	1 (3.6%)
III	14	3 (21.4%)	11 (78.6%)	6 (42.9%)	4 (28.6%)	2 (14.3%)
IV	27	1 (3.7%)	26 (96.3%)	24 (88.9%)	23 (85.2%)	15 (55.6%)
Node Status						
-	7	5 (71.4%)	2 (28.6%)	2 (28.6%)	1 (14.3%)	1 (14.3%)
+	20	7 (35.0%)	13 (65.0%)	6 (30.0%)	1 (5.0%)	0
N/A	45					
Grade						
1	6	2 (33.3%)	4 (66.7%)	3 (50.0%)	2 (33.3%)	2 (33.3%)
2	22	7 (31.8%)	15 (68.2%)	9 (40.9%)	4 (18.2%)	3 (13.6%)
3	11	3 (27.3%)	8 (72.7%)	6 (54.5%)	4 (36.4%)	1 (9.1%)
N/A	33					
Tumour Size (cm)						
< 2	10	4 (40.0%)	6 (60.0%)	6 (60.0%)	3 (30.0%)	3 (30.0%)
2-3	17	5 (29.4%)	12 (70.6%)	9 (52.9%)	6 (35.3%)	3 (17.6%)
3-4	17	3 (17.6%)	14 (82.4%)	8 (47.1%)	8 (47.1%)	6 (35.3%)
> 4	18	5 (27.8%)	13 (72.2%)	8 (44.4%)	5 (50.0%)	1 (5.6%)
Surgery (PDAC)						
Whipple	22	9 (40.9%)	13 (59.1%)	8 (36.4%)	1 (4.5%)	1 (4.5%)
Distal Panc	5	3 (60.0%)	2 (40.0%)	0	0	0
Ex Lap Stage III	6	0	6 (100%)	4 (66.7%)	4 (66.7%)	2 (33.3%)
Ex Lap Stage IV	8	0	8 (100%)	7 (87.5%)	7 (87.5%)	4 (50.0%)
Non-Adenocarcinoma	28	27 (96.4%)	1 (3.6%)	0	0	0
Pathology/Clinical Dx						
IPMN (SB/MD)[a]	15 (4/11)	15 (100%)	0	0	0	0
MCN[b]	2	2 (100%)	0	0	0	0
Chronic Panc	1	1 (100%)	0	0	0	0
Serous Cyst	4	4 (100%)	0	0	0	0
Benign Pancreas	2	2 (100%)	0	0	0	0
PNET[c] (G1)	2	2 (100%)	0	0	0	0
Pseudocyst	1	1 (100%)	0	0	0	0
Complex Cyst	1	0	1 (100%)	0	0	0

2.2.4 Results

Study cohort

We successfully enrolled 108 patients as described in the methods section (**Figure 7**). Eight patients were excluded (3-duodenal cancer, 3-concurrent second malignancy, 1-

exacerbation of inflammatory bowel disease, 1-refused informed consent). Thus, our study cohort consisted of 100 patients: 28 with non-adenocarcinoma diagnoses and 72 with a diagnosis of PDAC (Table 2). Of the 72 patients with PDAC, 4.2% had AJCC Stage-I disease, 38.9% had Stage-II disease, 19.4% had Stage-III disease, and 37.5% had Stage-IV disease. Fourteen patients were initially clinical stage I-II but were upstaged following surgical exploration secondary to finding locally advanced disease in 6 cases (Stage-III) and metastatic (Stage-IV) disease in the other 8 cases. There were no statistically significant differences in patient sex or age between PDAC and non-adenocarcinoma diseases or in the differing PDAC stage groups.

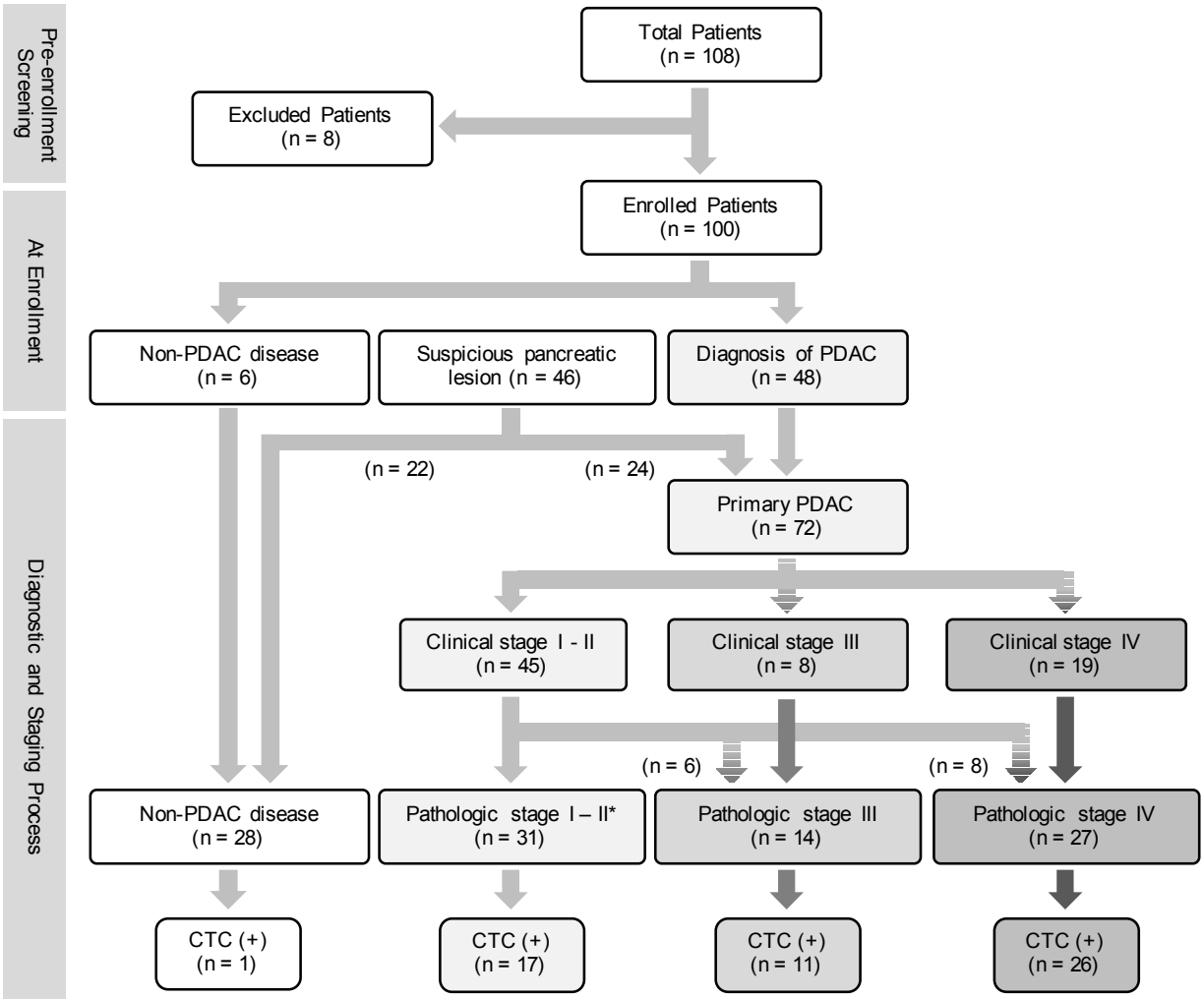


Figure 7. Study cohort characteristics. Diagnostic and staging flowchart of enrolled patients in the study.

CTCs in primary PDAC versus non-adenocarcinoma diseases

CTC enumeration was performed on VB from each of the patients enrolled in the study. A series of images depicting our multi-marker staining for typical PDAC CTCs is depicted in (**Figure 6B**). In the PDAC group, CTCs were found to be present in 54/72 (75.0%) patients (**Figure 8A**) (median = 2, range = 0–48).

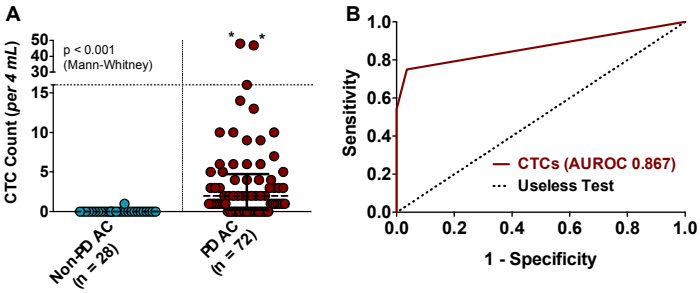


Figure 8. CTCs as a diagnostic biomarker. **(A)** Comparison of CTC enumeration in PDAC and non-adenocarcinoma diseases. **(B)** ROC curve for illustration of CTC performance in the discrimination of PDAC from non-adenocarcinoma diseases. CTC AUROC = 0.867 (95% CI 0.798 – 0.935, p < 0.001).

In the non-adenocarcinoma group, 1/28 (3.6%) patients was found to have a single CTC in 4-mL VB (**Figure 8A**) (median = 0, range=0–1). This patient had a large, complex cyst on imaging and EUS, lack of malignant cells on FNA, and cyst fluid analysis with CEA<192 ng/mL and amylase of 274 units/L. No confirmatory surgical pathology was available and no additional CTCs could be isolated from additional VB for *KRAS* mutational analysis

CTCs as a diagnostic biomarker in PDAC

Calculation of Youden's J-statistic (J) at various CTC cut-off values established an optimal diagnostic performance for PDAC at 1 CTC in 4-mL VB. CTC presence thus demonstrated 75.0%-sensitivity, 96.4%-specificity, 98.2%-PPV, 60.0%-NPV, and a positive likelihood ratio (+LR) of 21.00 for the diagnosis of primary PDAC. A ROC curve was then constructed and the AUROC was 0.867 (95%CI = 0.798–0.935, p<0.001), thus illustrating the overall ability of CTCs to discriminate PDAC from non-adenocarcinoma (**Figure 8B**).

CTCs correlation with PDAC stage

CTCs were found in 0.0%, 60.7%, 78.6%, and 96.3% of AJCC Stage I, II, III, and IV patients, respectively (Table1). For these respective stage groups, median CTC numbers were 0 (range=0), 1 (range=0–6), 1 (range=0–9), and 5 (range=0–48) per 4 mL VB (Fig. 5A). In comparing CTC counts by stage group, a statistically significant difference between Stage-IV PDAC and all other stages of PDAC ($p<0.0001$) was found on the multiple comparisons test, but no differences were noted between AJCC Stage I, II, or III groups. In the Stage-II patients with available node status, there was no statistically significant difference in CTC count between node positive and node negative patients (**Figure 9**). Additionally, CTC counts were not significantly different between patients with different tumor sizes or grades.

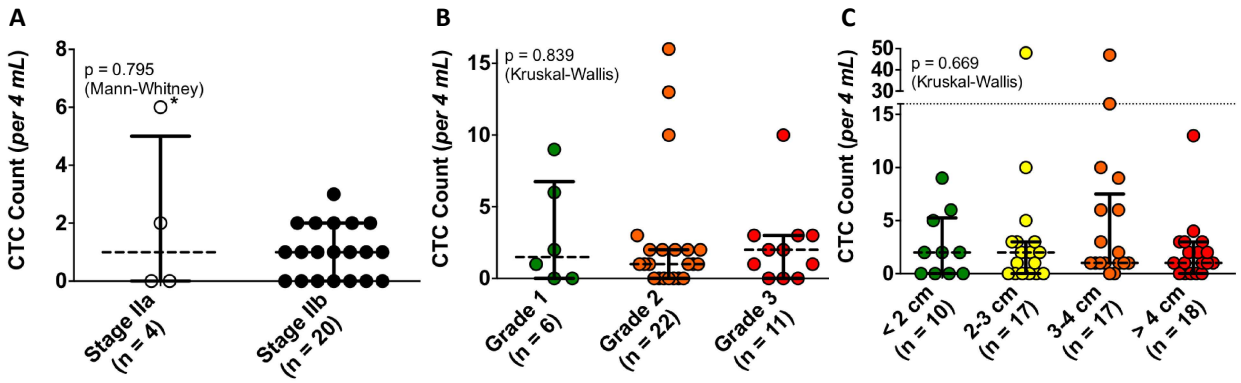


Figure 9. Additional statistical analysis. **(A)** Comparison of CTC counts between node-negative (Stage IIa) and node-positive (Stage IIb) patients reveals no significance ($p = 0.795$). **(B)** Comparison of CTC counts between patients with different tumor grades reveals no significance ($p = 0.839$). **(C)** Comparison of CTC counts and tumor size reveals no significance ($p = 0.669$).

CTCs and pre-treatment staging

To test the ability of CTCs to differentiate clinically meaningful groups of patients, we compared those with local/regional (Stage I-III) and metastatic (Stage-IV) disease, and found a significant difference in CTC enumeration between these groups ($p<0.001$) (**Figure 10B**). Based

on J, a CTC value of ≥3 CTCs in 4-mL blood was the optimum cut-off for association with metastatic disease in patients with PDAC. At this cut-off value, CTCs showed a sensitivity of 85.2%, specificity of 86.7%, PPV of 79.3%, and NPV of 90.7% in identifying patients with systemic disease and patients with ≥3 CTCs per 4-mL blood were 6.39 times more likely to harbor metastatic disease at the time of diagnosis. Overall, CTCs had an AUROC of 0.885 (95%CI = 0.800–0.969, p<0.001) for discriminating local/regional from metastatic disease (**Figure 10)**. By comparison, in the 59/72 patients with available CA19-9 levels, use of the optimum cut-off of ≥ 3000 units/mL yielded a sensitivity of 34.8%, specificity of 100%, PPV of 100%, NPV of 70.6%, and an AUROC of 0.690 (95%CI = 0.551–0.829, p=0.014) for the identification of PDAC patients with metastatic disease (**Figure 10C**).

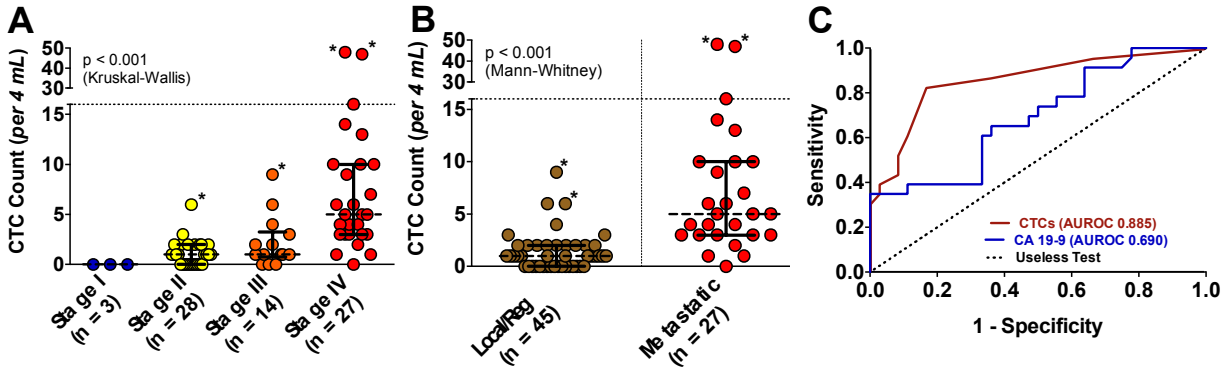


Figure 10. CTCs as a staging biomarker. **(A)** CTC enumeration showing correlation with PDAC stage. **(B)** CTC enumeration in local/regional (stage I-III) and metastatic (stage IV) disease. **(C)** Comparison of the performance of CTCs and CA19-9, in discriminating local/regional from metastatic disease. The CTC AUROC was 0.885 (95% CI = 0.800 – 0.969, p < 0.001), the CA19-9 AUROC was 0.690 (95% CI = 0.551 – 0.829, p = 0.014).

Outlier analysis of CTCs for diagnosis and staging of PDAC

Finally, we performed an outlier analysis of our CTC enumeration in patients with PDAC to determine if our results were significantly impacted by extreme values. These values are marked with an asterisk in our CTC enumeration graphs (**Figure 8** and **Figure 10**). For diagnosis of PDAC, exclusion of outliers did not significantly affect our data. The median CTC count of

the PDAC group remained at 2, and the AUROC was still 0.863 (95%CI = 0.793–0.933, p<0.001). Similarly, the exclusion of outliers actually improved the AUROC for local/regional versus metastatic disease, new AUROC 0.902 (95%CI = 0.818–0.986, p<0.001). These results further support our finding of the ability of CTC presence and enumeration to differentiate patients with cancer from those without, and to identify which PDAC patients are more likely to harbor advanced AJCC stage disease.

2.2.5 Discussion

Pancreatic cancer is a devastating disease that is incurable for approximately 95% of those diagnosed.[39] Current diagnostic methods utilizing image-guided biopsy are inconvenient, have associated morbidity, are relatively expensive and require local expertise. Furthermore, clinical staging based on cross-sectional imaging is insensitive to small-volume or micrometastatic disease and potentially results in under-staging at the time of diagnosis. This often results in these patients undergoing surgery as first line therapy when, in fact, they have systemic disease. Surgery as first line therapy carries significant morbidity and has been shown to cause delays in administration of systemic therapy, the most effective first-line therapy for metastatic disease.[47] Thus, there is a need for biomarkers that can help efficiently establish diagnosis and provide information regarding stage/prognosis at the time of disease presentation to better inform initial therapeutic decisions.

In the present study, we investigated CTCs as an adjunct biomarker for diagnosis and staging of PDAC at the time of disease presentation. We developed a sensitive and specific method for PDAC CTC detection using the NanoVelcro platform in conjunction with high-resolution fluorescent microscopy and a multi-color immunocytochemistry (ICC) approach. We

found the addition of CEA immunostaining to be of little value on its own, as only 1.6% of our CTCs isolated were CEA+ / CK- / CD45- / Nucleus+. However, as an additional marker, we found it to be helpful for reaffirming the identity of CK+ CTCs. We validated our CTC definition using NanoVelcro/LCM by mutational analysis of *KRAS* codon 12, demonstrating concordance of *KRAS* mutation subtype between CTCs and primary tumor tissue. We then demonstrated the ability of CTCs to function as an adjunctive biomarker for the diagnosis and staging of PDAC.

In regards to diagnosis of PDAC, our study found CTC presence (≥ 1 CTC / 4-mL VB) performed with an overall sensitivity of 75.0% and specificity of 95.7%. Studies by Kurihara *et. al.*, Khoja *et. al.*, and Bidard *et. al*., all utilized the CellSearch platform to study CTC enumeration in only advanced (Stage III or IV) PDAC patients and found CTC presence in 42.3%, 39.6%, and 11% of patients, respectively.[48,49,54] Our reported sensitivity was significantly higher than these studies, as we detected CTCs in 11/14 (78.6%) Stage III patients and 26/27 (96.3%) Stage IV patients. In fact, the sensitivity for our entire study cohort was higher, even with analysis of smaller blood volumes (4 mL versus 7.5 mL) and inclusion of early Stage I/II patients (43.1% of our study cohort). This may be secondary to the higher sensitivity of microfluidic capture techniques for PDAC CTCs. For example, similar to our finding from Stage IV patients, studies using the microfluidic CTC-Chip platform detected CTCs in 15/15 (100%) Stage IV PDAC patients.[55,56] Another study utilizing microfluidics detected high numbers of CTCs in both resectable and metastatic PDAC patients; however, they only reported the summary statistics of CTCs found, not the percentage of patients with CTCs.[57] Using a different microfluidic device, Rhim *et. al.,* found circulating pancreatic cells from 8/11 (73%) PDAC patients with stage I-IV, similar to our 75% sensitivity.[20] Interestingly, they also found

circulating pancreatic cells in 7/21 (33%) of patients with cystic pancreatic lesions, raising the possibility of EMT cells circulating in the blood as well.[58] Another promising technique is the Isolation by Size of Epithelial Tumor Cell (ISET) method. A study by Khoja et al found CTCs in ≥ 90% of Stage IV patients using ISET, similar to our sensitivity of 96.3% for Stage IV patients.[49] All studies, including ours, had a reported specificity of >90%, similar to research of CTCs in other solid tumors that has consistently found CTCs to have high specificity.[33] Utilizing mutational analysis as a means of confirming our CTC definition represents an orthogonal approach to ICC validation, and gave us confidence in our results. This is especially important for diagnostic tests, as a single CTC can only be considered diagnostic of a cancer if its tumor origin can be inferred.

The ability to accurately discriminate advanced disease is of obvious importance; informing first-line therapy, improving prognostication, and allowing trial stratification. In our study, CTC enumeration correlated with AJCC stage-groups, and was further found to outperform CA19-9 as a biomarker for differentiating local/regional from metastatic disease. While further studies are needed, CTC enumeration may have potential as a prognostic biomarker signifying likely metastatic disease.

Our identification of *KRAS* mutations in CTCs, but not WBCs, makes us confident in our ability to distinguish CTCs from other circulating hematopoietic cells. Furthermore, we used *KRAS* mutational analysis of CTCs and matched tumor tissue to provide us with insight into the origin of the CTCs found. Despite these confirmatory studies, we still had a single false-positive result. This false-positive result occurred in a patient with a large (6cm), complex cyst, non-diagnostic cyst fluid analysis, and lack of malignant cells on FNA. Although the patient may have a benign cyst, it is also possible that this patient may have malignant cystic disease. In

addition to the risk from the cyst itself, studies have shown that 2.8 - 9.3% of patients with benign pancreatic cysts have synchronous or metachronous PDAC [59,60]. There is no available confirmatory pathology at the time of this publication as the patient chose not to pursue surgery. Of note, a recent study [20] also demonstrated CTC presence in approximately 30% of patients with pre-malignant pancreatic cystic disease utilizing a different microfluidic technology and CTC identification criteria.

An important limitation of our study was the use of an epithelial surface marker (EpCAM) for CTC capture that potentially led to decreased sensitivity secondary to loss of CTCs expressing non-epithelial surface markers.[49] Given that PDAC cells have been shown to undergo epithelial to mesenchymal transition after entering the circulation in a mouse model, we hope to capture mesenchymal-type CTCs in the future as demonstrated in a study by Rhim *et. al.*.[58] These CTCs may provide different biological insights for pancreatic cancer. However, to our knowledge, our cohort is the largest to date looking at CTC enumeration as a biomarker for diagnosis and staging. This study is still on-going, and we hope to validate our findings in a larger cohort of patients. Furthermore, we have continued to follow all enrolled patients, and hope to analyze CTCs as a prognostic biomarker for recurrence and survival in future studies. Additionally, our confirmation of tumor origin of captured CTCs demonstrates the potential for CTCs to function as a "liquid biopsy" in PDAC. Future studies utilizing technologies such as NanoVelcro/LCM should allow for comprehensive CTC molecular analysis, which may provide more biological insight into the intravasated tumor cell population as well as tumor heterogeneity and the mechanism of metastasis.[56] Thus, future studies will hopefully not only confirm CTCs utility as a biomarker, but also demonstrate their potential to provide actionable information

about the tumor's biology, which together holds great promise with respect to realizing "personalized" treatment of PDAC.

2.3 Circulating Tumor Cells Predict Occult Metastatic Disease and Prognosis in Pancreatic Cancer

2.3.1 Abstract

Background: Occult metastatic tumors, below imaging thresholds, are a limitation of staging systems that rely on cross-sectional imaging alone, and a cause of the routine under-staging of pancreatic adenocarcinomas (PDAC). We investigated circulating tumor cells (CTCs) as a preoperative predictor of occult metastatic disease and as a prognostic biomarker for PDAC patients. **Methods:** A total of 126 patients were enrolled in our study and CTCs were identified and enumerated from 4-mL of venous blood using the microfluidic NanoVelcro assay. CTC enumeration was correlated with clinicopathologic variables and outcomes following both surgical and systemic therapies. **Results:** CTCs were identified in 78% of PDAC patients and CTC counts correlated with increasing stage ($\rho = 0.42$, $p < 0.001$). Of the 53 patients taken for potentially curative surgery, 13(24.5%) were found to have occult metastatic disease intraoperatively. Patients with occult disease had significantly more CTCs than patients with local disease only (median: 7 vs. 1 CTC, $p < 0.0001$). At a cutoff of ≥ 3 CTCs/4mL, CTCs correctly identified patients with occult metastatic disease preoperatively (AUROC: 0.82, 95%CI: 0.76–0.98, $p < 0.0001$). CTCs were a univariate predictor of recurrence-free survival following surgery (HR: 2.36, 95%CI: 1.17–4.78, $p = 0.017$) as well as an independent predictor of overall survival on multivariate analysis (HR: 1.38, 95%CI: 1.01–1.88, $p = 0.040$). **Conclusion:** CTCs show promise as a prognostic biomarker in PDAC patients at all stages of disease being treated both medically and surgically. Furthermore, CTCs demonstrate utility as a preoperative biomarker for identifying patients at high risk of occult metastatic disease.

2.3.2 Introduction

Pancreatic ductal adenocarcinoma (PDAC) ranks 4th in cancer-related mortality in the US, due to late presentation and resistance to systemic therapy; however, the recent success of multidrug regimens has begun to change this paradigm.[39,61] Tumor staging by cross-sectional imaging alone, the current practice standard, routinely understages PDAC patients as evidenced by high recurrence rates and the low survival benefit of curative-intent surgery.[62-64] Due to the historic inadequacies of systemic therapy in PDAC, a "surgery first" mentality for all patients with even a remote chance of cure could be justified. However, the success of multidrug regimens calls in to question the survival benefit of pancreatic resections for patients who almost certainly have occult metastatic disease, below the detection limit of cross-sectional imaging. Recent studies have even demonstrated a survival benefit for neoadjuvant therapy in patients with very early stage disease.[62,65] These studies highlight the need for better pre-treatment biomarkers for staging and prognostication in PDAC to inform first-line therapy decision-making.

Circulating tumor cells (CTCs), cells of tumor origin circulating in the blood, are thought to represent the intravasated tumor stage, between the formation of an invasive cancer and its eventual distant metastasis.[4,30] CTCs are found in the blood of patients with all epithelial tumors, and have been studied as a biomarker in numerous cancer types.[30] We have previously used the microfluidic NanoVelcro CTC assay's single cell sequencing ability to validate our immunocytochemical (ICC) CTC definition by confirming the tumor origin of PDAC-CTCs through demonstration of mutational congruence between the primary tumor and the isolated CTCs.[66] We then used our PDAC CTC assay to show that CTCs are present in the majority of

PDAC patients but not patients with non-malignant pancreatic disease, and that CTCs may have utility as a diagnostic biomarker at the time of disease presentation.[67]

In this study, we evaluated CTC enumeration as a pre-treatment biomarker for prognosis and therapeutic planning in PDAC patients. Specifically, we investigated if CTCs could help identify patients with occult metastatic disease, defined as the presence of metastatic disease that is not detected by cross-sectional imaging. Separately, we investigated the strength of association between CTC enumeration and survival; both alone and in conjunction with other variables previously identified as prognostic.

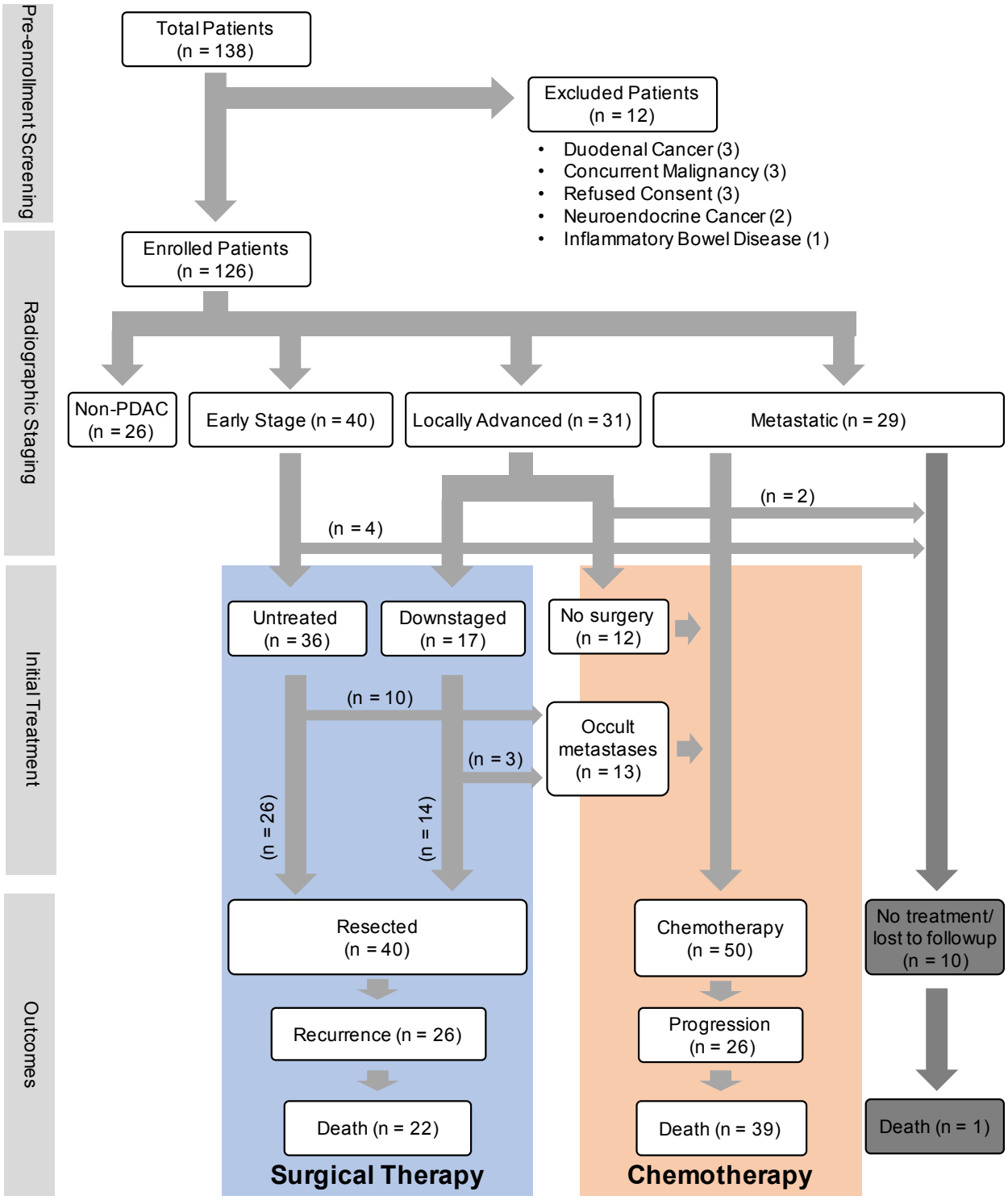


Figure 11. Flowchart of patient enrollment, radiographic staging, initial treatment decisions and outcomes for all patients.

2.3.3 Materials and methods

The materials and methods for project 2.3 have been extensively documented in **project 2.2.3**, any variation from these methods is documented below. Between December 2012 and March 2015, we approached patients presenting to our multi-specialty pancreatic clinic with suspicion for, or diagnosis of, malignant pancreatic disease and enrolled them under IRB#11-002112. Exclusion criteria included any malignancies within the past 5-years, active inflammatory bowel disease, or suspicion of pancreatic neuroendocrine tumor. A database of demographic, laboratory and relevant clinicopathologic variables was prospectively maintained. Diagnosis was confirmed by biopsy or surgical pathology whenever possible and radiographically when a tissue diagnosis was not available. Patients were assessed as having early-stage, locally-advanced, or metastatic disease preoperatively using cross-sectional imaging via a pancreas protocol, contrast-enhanced computerized tomography (CT) or magnetic resonance imaging (MRI) scan. TNM stage was assessed using the AJCC staging manual.[68]

Statistical methods

Differences in CTC number or CA19-9 levels were evaluated with the non-parametric Mann-Whitney *U*-test and cutoff points for CTC enumeration were evaluated using Cutoff Finder.[69] Diagnostic performance of CTCs was evaluated using receiver operating characteristic curves (ROCs) for determination of the area under the curve (AUROC) in addition to sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) calculations. Survival curves were computed using Kaplan-Meier methods and compared by log-rank (Mantel-Cox) tests. Univariate analysis of predictors was performed using a Cox regression model, and significant ($p < 0.15$) factors were subsequently entered into a multivariate Cox

model to identify independent predictors of overall survival (OS) and recurrence-free survival (RFS). CA19-9 and CTC values had skewed distributions originally, and were log-transformed to a normal distribution for use in survival analysis. A two-tailed p-value <0.05 was considered statistically significant throughout the study. All statistical analyses and calculations were performed with the assistance of R (version 3.3.2).

Table 3. Clinical, laboratory, radiologic, treatment and pathologic characteristics of the patients (n = 126).

Characteristics		Data
Patient		
Age, median (IQR)		67 (61 – 73)
Female, n (%)		59 (46.8)
Diagnosis, n (%)		
	PDAC	100 (79.4)
	IPMN [SB/MD]	13 [5/8] (10.3)
	Serous Cyst	4 (3.2)
	Chronic Pancreatitis	3 (2.4)
	MCN	2 (1.6)
	PANIN	2 (1.6)
	Pseudocyst	1 (0.8)
	Complex Cyst	1 (0.8)
AJCC Stage at draw, n (%)*		
	1	9 (9)
	2	31 (31)
	3	31 (31)
	4	29 (29)
CA19-9, median (IQR)**		139 (24 – 1010)
*Surgical characteristics****		
Downstaged, n (%)		18 (36.7)
Planned Procedure		
	Whipple, n (%)	37 (69.8)
	Distal Pancreatectomy, n (%)	16 (30.2)
Occult metastatic disease, n (%)		13 (24.5)
Metastatic location, n (%)		
	Liver	6 (46.2)
	Intraperitoneal	7 (53.9)
*Pathologic Characteristics*****		
Tumor Diameter, median (IQR)		3.5 (2.2 – 4.0)
LVI, n (%)		21 (52.5)
PNI, n (%)		36 (90.0)
Grade		
	Well	4 (10.3)
	Moderate	27 (69.2)
	Poor	8 (20.5)
	Indeterminate	1 (2.5)

2.3.4 Results

Patient characteristics

We approached 138 patients and enrolled 126 patients successfully as outlined in the Methods section (**Figure 11**). The clinical, laboratory, radiologic, and treatment characteristics of the 126 patients are summarized in **Table 3.** For the 100 patients with a diagnosis of PDAC, on cross-sectional imaging at the time of blood draw, the median tumor diameter was 3.5 cm (IQR: 2.2–4.0 cm). Of patients with local disease only (n=71), 31 (43.7%) demonstrated involvement of the mesenteric vessels and were classified as locally-advanced. Thus, by cross-sectional imaging, 9% had AJCC stage 1 disease, 31% had stage 2 disease, 31% had stage 3 disease, and 29% had stage 4 disease (**Figure 11**).

Of the 53 patients taken to the operating room for potentially curative surgery, 17 (32.1%) had received chemotherapy prior to the operation. The blood draws for all surgical patients occurred on the day of surgery prior to induction of anesthesia. The operative plan included a Whipple in 37 69.8%) patients and a distal pancreatectomy in 16 (30.2%) patients. Of the 53 patients, 13 (24.5%) were found to have metastatic disease not detected on cross-sectional imaging intraoperatively; referred to as occult metastatic disease (**Figure 11**). Of these 13 patients, 3 (23.1%) were discovered during diagnostic laparoscopy and the rest during exploratory laparotomy. The site of metastasis was the liver in 6 (46.2%) patients and other peritoneal sites in 7 (53.9%) patients. All 13 patients subsequently underwent chemotherapy. Thus, 50 patients were treated with palliative chemotherapy as their primary treatment, 25 with metastatic disease, 13 with occult metastatic disease and 12 with locally advanced disease.

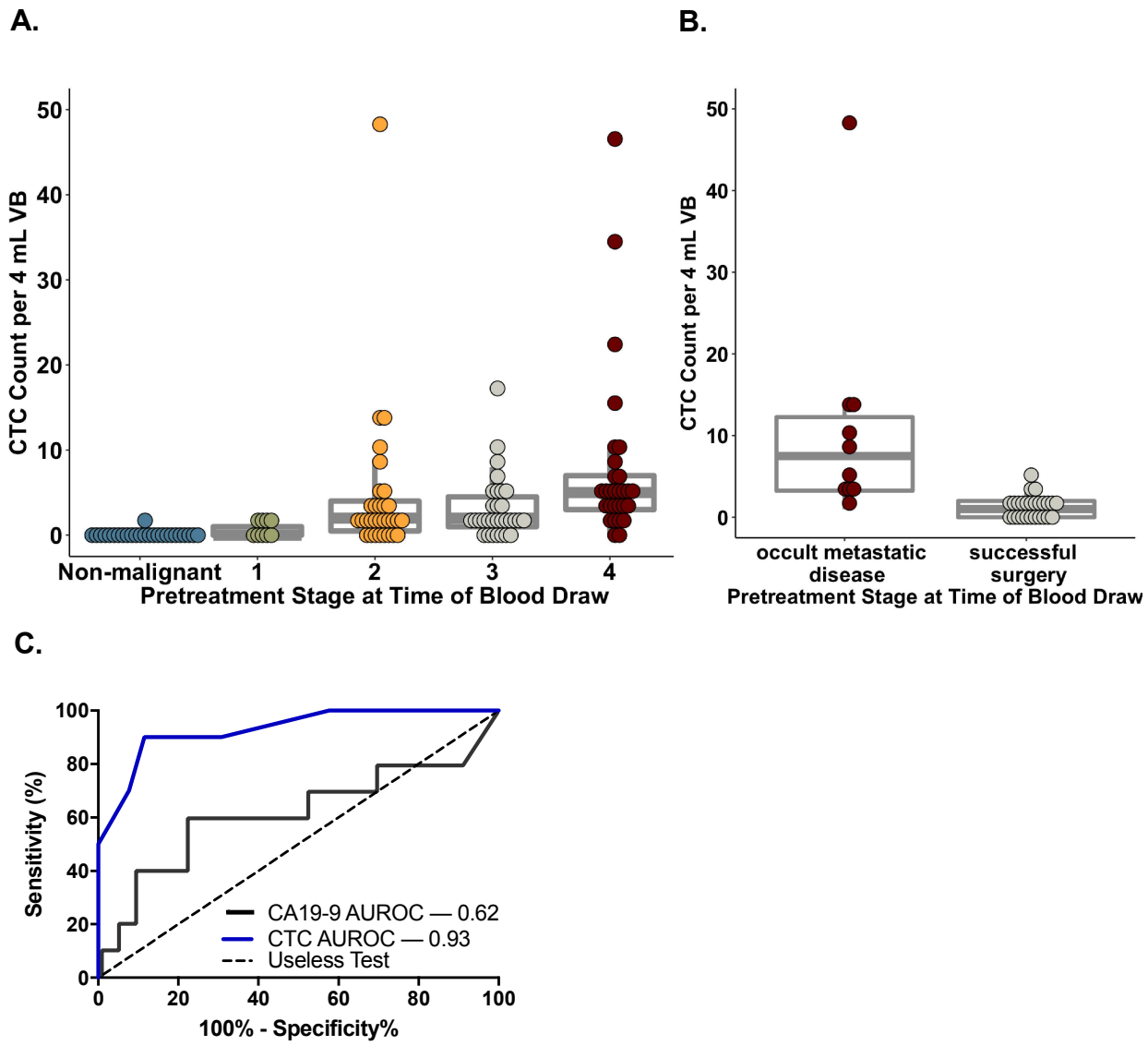


Figure 12. (A) CTC count by pretreatment radiographic AJCC stage of disease. Non-malignant refers to the 26 patients with non-malignant pancreatic diseases outlined in Table 1. **(B)** Performance of CTCs as a preoperative predictor of occult metastatic disease in the subset of patients with untreated tumors (n = 36). CTC enumeration is displayed for patients with successfully resected tumors versus those found to have occult metastatic disease intraoperatively. **(C)** Comparison of the performance of CTCs and CA19-9 in the preoperative detection of patients with occult metastatic disease in the subset of patients with untreated tumors (n = 36).

Circulating tumor cells enumeration

CTC counts were assessed from 4-mL VB for all patients. CTCs were found in 78/100 (78%) PDAC patients, and one or more CTCs per 4-mL VB were detected in 4/9 (44.4%) Stage I, 23/31 (74.2%) Stage II, 24/31 (77.4%) Stage III, and 27/29 (93.1%) Stage IV PDAC patients

(**Figure 12A**). CTC count correlated with increasing stage of disease ($\rho = 0.42$, $p < 0.001$) For patients with non-malignant pancreatic disease, 1/26 (3.8%) patients was found to have a single CTC. This patient had a 6cm complex cyst, non-diagnostic cyst fluid analysis, and a lack of malignant cells on FNA. He was followed for over 4 years and did not demonstrated any sign of invasive disease. No other patients with benign pancreatic disease developed PDAC over the course of the study. CTC counts did not correlate with CA19-9 levels ($\rho=0.20$, $p=0.064$), nor with age or gender (data not shown).

CTCs as a predictive biomarker for surgical resection

We evaluated CTC enumeration as a preoperative biomarker predicting the presence of occult metastatic disease among the 53 patients taken for potentially curative surgery. Occult metastatic disease was found in 13 (24.5%) patients intraoperatively, and these patients had a median of 7 CTCs (IQR: 3–13) versus 1 CTC (IQR: 0–2) in patients with localized disease ($p<0.0001$). Among the 36 patients without prior chemotherapy, a median of 8 CTCs (IQR: 3–12) were found in the 10 patients with occult metastatic disease versus 1 CTC (IQR: 0–2) in the 26 patients with localized disease ($p<0.0001$) (**Figure 12B**). For patients who received downstaging chemotherapy the results were similar, with a median CTC count of 7 (IQR: 4–12) for patients with occult disease versus 1 CTC (IQR: 1–5) for those with localized tumors.

For all surgical patients, at the optimized cutoff of ≥ 3 CTCs/4-mL VB, CTCs were able to preoperatively distinguish patients with occult metastatic disease from those with potentially curable, localized tumors with a sensitivity of 85%, specificity of 80%, PPV of 94%, NPV of 58%, and AUROC of 0.82 (95%-CI: 0.76–0.98, $p<0.0001$). Similarly, for the untreated patients only (n=36), ≥ 3 CTCs/4-mL VB distinguished patients with occult metastatic disease with a

sensitivity of 90%, specificity of 88%, PPV of 96%, NPV of 75%, and AUROC of 0.93 (95%-CI: 0.83–1.00, p<0.001) (**Figure 12C**). For comparison, CA19-9, at the optimum cutoff of 500 U/mL, distinguished patients with potentially curable tumors from those with advanced disease with a sensitivity of 60.0%, specificity of 77.3%, PPV of 81.0%, NPV of 54.5%, and AUROC of 0.62 (95%-CI: 0.37–0.87, p=0.084) (**Figure 12C**).

Outcomes and survival analysis

Patients were followed for a minimum of 24 months after blood draw, and the median survival time for all PDAC patients was 16.8 months (IQR: 7.4–33.3mos.). At the time of analysis, 38/100 (38.0%) remained alive. Overall survival for patients with occult metastatic disease was nearly identical to patients with radiologically apparent metastatic disease, and was significantly shorter than that of patients who underwent successful resection (**Figure 13A**, HR=3.6, 95%-CI: 1.75–7.73, p<0.001). We performed a univariate analysis of potential preoperative indicators of prognosis including CTC count, CA19-9 level, as well as radiographic indicators (**Table 4**). For all patients regardless of stage, CTC count was strongly associated with survival from the time of blood draw (HR=1.69, 95%-CI: 1.28–2.25, p<0.001,**Figure 13B**). Additional factors associated with shorter survival on univariate analysis included age, presence of metastatic disease and CA19-9 level (**Table 4**). To determine the subset of factors that provided independent information on survival time, a Cox proportional hazards model was developed. We used stepwise Cox regression analysis at a stringency level of $p < 0.15$ to include factors. The factors that produced the strongest independent association with survival time were age (per year, HR=1.04, 95%CI: 1.01–1.07, p=0.009), presence of metastatic disease (HR=2.04, 95%CI: 1.11–3.74, p=0.022), and CTC count (per log unit, HR=1.38, 95%CI: 1.01–1.88,

p=0.040) (**Table 4**). We next limited our analysis to just the 53 patients taken for potentially curative surgery. Again, CTCs were found to be an independent predictor of overall survival (HR=2.48, 95%CI: 1.32–4.67, p=0.0048).

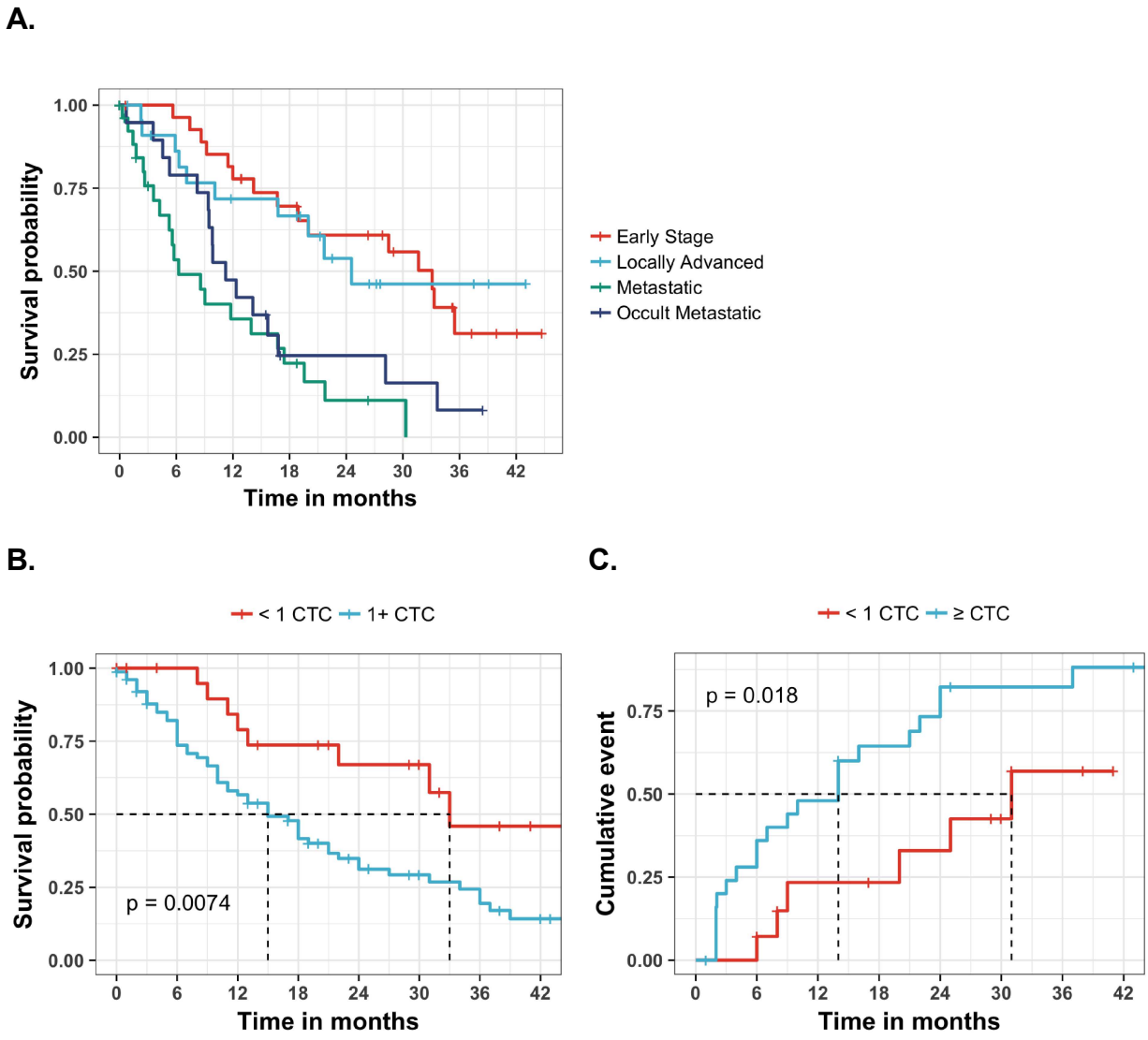


Figure 13. (A) Overall survival for all patients diagnosed with PDAC in the study, separated into the subset of patients with resected tumors (n = 40, red), locally-advanced disease treated palliatively (n = 12, light blue), occult metastatic disease (n = 13, purple), or metastatic disease (n = 25, green). Patients with occult metastatic disease had survival similar to patients with radiologically-visible metastatic disease, and significantly different from patients with early stage or locally advanced disease. **(B)** Overall survival for all patients stratified by CTC count (n = 100). **(C)** Time to recurrence for patients after surgical resection stratified by CTC count (n = 40).

Recurrence

Of the 40 patients successfully resected, 26(65%) developed a recurrence (**Figure 13C**) and 22(55%) died, with the median survival for all resected patients being 28.5 months. Median survival was 33.1 months in preoperatively untreated patients versus 21.7 months in patients receiving neoadjuvant chemotherapy prior to resection; however, the difference in median survival was not statistically significant (p=0.33). Univariate preoperative predictors of PDAC recurrence when accounting for the competing risk of non-cancer related death are shown in **Table 5**. On univariate analysis, tumors were more likely to recur in patients who received downstaging chemotherapy (HR=3.34, p=0.001), and patients with higher CTC counts (per log unit, HR=2.65, p=0.017). When the analysis was limited to the subset of early stage patients who did not receive neoadjuvant therapy, CTC counts were no longer associated with shorter time to recurrence (HR=2.57, p=0.08). On multivariate analysis, CTC counts were not found to be independently predictive of shorter time to recurrence (**Table 5**).

Table 4. Univariate and multivariate analysis for predictors of overall survival (n = 100).

	Univariate analysis			*Multivariate analysis*		
Predictor	**Hazard Ratio**	**95% CI**	**P-value**	**Hazard Ratio**	**95% CI**	**P-value**
Clinical predictors						
Age, per year	1.03	1.01 - 1.06	0.011	1.04	1.01-1.07	0.009
Male gender	0.89	0.54-1.47	0.64	–	–	–
Pre-operative predictors						
Radiographic						
Mesenteric vessel involvement	2.25	0.51 - 9.84	0.28	–	–	–
Metastatic disease	6.61	1.55 - 28.12	0.011	2.04	1.11 - 3.74	0.022
Tumor Location						
Head	–	–	–	–	–	–
Tail	1.48	0.66 - 3.30	0.34	–	–	–
Tumor Diameter, per cm*	1.00	0.86 - 1.17	0.96	–	–	–
Laboratory						
CA19-9, per log unit**	1.09	0.97 - 1.21	0.14	1.09	0.98 - 1.22	0.12
CTC count, per log unit	1.69	1.28 - 2.25	< 0.001	1.38	1.01 - 1.88	0.040

* - n = 91; ** - n = 90

Table 5. Univariate and multivariate analysis for predictors of recurrence-free survival.

All Patients (n = 40)	***Univariate Analysis***			***Multivariate Analysis***		
Predictor	**Hazard Ratio**	**95% CI**	**P-value**	**Hazard Ratio**	**95% CI**	**P-value**
Clinical predictors						
Age, per year	1.03	1.0 – 1.06	0.12	1.10	1.04 – 1.17	0.002
Male gender	0.99	0.53 – 1.88	0.99	–	–	–
Pre-operative predictors						
Radiographic						
Mesenteric vessel involvement	3.34	1.79 – 6.23	0.001	6.98	2.18 – 22.37	0.001
Tumor Location						
Tail	–	–	–	–	–	–
Head	1.09	0.38 – 3.17	0.87	–	–	–
Tumor Diameter, per cm*	1.10	0.91 – 1.32	0.32	–	–	–
Laboratory						
CA19-9, per log unit	1.08	0.88 – 1.32	0.46	–	–	–
CTC count, per log unit	2.36	1.17 – 4.78	0.017	1.5	0.71 – 3.17	0.29
Early Stage Only (n = 26)	***Univariate Analysis***			***Multivariate Analysis***		
Predictor	**Hazard Ratio**	**95% CI**	**P-value**	**Hazard Ratio**	**95% CI**	**P-value**
Clinical predictors						
Age, per year	1.09	1.01 – 1.17	0.02	1.08	1.00 – 1.16	0.04
Male gender	1.25	0.42 – 3.76	0.69	-	-	-
Pre-operative predictors						
Radiographic						
Tumor Location						
Tail	-	-	-	-	-	-
Head	0.91	0.25 – 3.31	0.89	-	-	-
Tumor Diameter, per cm	1.15	0.92 – 1.44	0.22	-	-	-
Laboratory						
CA19-9, per log unit	1.30	0.95 – 1.78	0.10	1.19	0.81 – 1.74	0.37
CTC count, per log unit	2.57	0.88 – 7.46	0.08	1.76	0.48 – 6.48	0.40

* - Stage III patients downstaged prior to surgery (n = 13)

2.3.5 Discussion

The combination of late presentation and resistance to systemic therapy makes PDAC one of the most deadly malignancies; however, the recent successes of multidrug regimens, and the anecdotal success of early-phase trials of new targeted- and immune-therapies, may finally change this paradigm.[70] Tumor staging by cross-sectional imaging alone routinely understages PDAC patients as evidenced by high recurrence rates and low survival benefit of curative-intent surgery. Furthermore, multiple studies have demonstrated that a "surgery first" paradigm for borderline-resectable, and even early-stage, patients does not result in a survival benefit and may, in fact, decrease survival.[62,65] Additionally, for borderline-resectable patients who have undergone neoadjuvant chemotherapy, studies have demonstrated that cross-sectional imaging is no longer an accurate predictor of the extent of local disease.[71] Thus, there is a pressing need for new biomarkers to complement cross-sectional imaging, in particular for the identification of patients likely to benefit from non-surgical treatments first.

In the present study, we investigated CTCs as a biomarker for staging and prognostication of PDAC. We enrolled patients at all stages of disease undergoing a variety of treatment modalities and determined how CTC enumeration correlated with clinicopathologic variables, surgical therapy, and survival outcomes. The current study builds on our previous investigation of CTCs as a diagnostic biomarker[67] by demonstrating the prognostic and predictive utility of CTCs for PDAC patients. We believe that our most significant finding was the predictive ability of CTCs for occult metastatic disease in the preoperative setting. Using the cutoff of ≥3 CTCs, we could differentiate patients with occult metastatic disease from those with localized, potentially curable tumors (AUROC=0.93 for untreated patients). This represents a significant improvement over the only widely available biomarker in PDAC, serum CA19-9

(AUROC=0.619). Radiographic criteria including tumor size and mesenteric vessel involvement was also not predictive of occult metastatic disease. Importantly, the survival curves of patients with occult metastatic disease reflected that of patients with radiographically detectable metastatic disease and not that of early stage patients or even locally-advanced patients who underwent downstaging chemotherapy (**Figure 13A**). While our study is small, the rate of occult metastatic disease in our patient cohort (13/53, 24.5%) was similar to the average rate for all patients taken to surgery at our institution between 2008–2013 (53/192, 27.6%). Thus, we believe our cohort is representative, and that CTCs may prove useful as a predictive biomarker in the difficult decision of pursuing local versus systemic therapy as an initial treatment modality. In addition to their predictive significance for occult metastatic disease, we found CTCs to have prognostic value at all stages of disease and for both surgical and systemic therapies. Like previous studies in PDAC, as well as other solid tumors, CTCs were associated with worse outcomes; and were an independent predictor of overall survival for all patients in our cohort.[49,54,72,73]

To our knowledge, our study is the largest to date utilizing a microfluidic platform for CTC enumeration in PDAC; however, its size is still limited for subgroup analyses of surgical outcomes. Our median CTC count for PDAC patients was 2(IQR: 1–6), lower than that found using other techniques such as filters or flow-cytometry.[33,49,57] One possibility is that our use of only 4-mL of venous blood may be too low to accurately establish the number of CTCs, especially in early stage patients. In future studies we will assess the utility of larger blood volumes.

In conclusion, CTCs detected using the NanoVelcro platform show promise as a prognostic biomarker for PDAC patients at all stages of disease. Furthermore, CTCs demonstrate

utility as an adjunctive biomarker to detect potential occult metastatic disease in patients with local disease being considered for surgical resection. Given the high recurrence rate after surgical resection, we see a potential application for surgical decision-making, especially in the subset of patients already at increased risk of recurrence such as those with borderline-resectable tumors at high risk of having occult metastatic disease.

2.4 CTC phenotyping as a biomarker in hepatocellular carcinoma

2.4.1 Abstract

Introduction: Current clinicopathologic staging systems and serum biomarkers poorly discriminate tumor biology in hepatocellular carcinoma (HCC), as evidenced by high recurrence rates following curative-intent surgical and locoregional therapy. We sought to develop a novel "liquid biopsy" assay capable of detecting HCC circulating tumor cells (CTCs), and characterizing phenotypic subpopulations with prognostic significance. **Methods:** Utilizing HCC cell lines, a tissue microarray, and human HCC blood samples, an antibody cocktail targeting the cell-surface markers asialoglycoprotein receptor (ASGPR), Glypican-3, and epithelial cell adhesion molecule (EpCAM) was optimized for HCC-specific CTC capture utilizing the NanoVelcro microfluidic assay. HCC-CTCs, as well as a subpopulation of vimentin(+) CTCs expressing an epithelial-to-mesenchymal phenotype, were identified. The ability of HCC-CTCs and vimentin(+) CTCs to accurately discriminate tumor stage, recurrence, progression, and overall survival was evaluated in a prospective study of 80 patients. **Results:** Multimarker CTC capture resulted in the detection of significantly greater numbers of CTCs than any individual antibody alone for both cell line and patient samples ($p<0.05$). HCC-CTCs accurately discriminated between HCC and non-HCC patients (AUROC = 0.92, $p<0.0001$; sensitivity=84.2%, specificity=88.5%). Vimentin(+) CTCs accurately discriminated early-stage versus locally advanced/metastatic HCC (AUROC = 0.89, $p<0.0001$; sensitivity=87.1%, specificity=90.0%), and predicted overall survival for all patients (HR 2.21, $p=0.001$), as well as faster recurrence after curative-intent surgical or locoregional therapy in potentially curable early stage HCC (HR 3.14, $p=0.002$). **Conclusion:** We have developed a novel multimarker CTC enrichment assay that detects HCC-CTCs with high efficiency and accuracy. A phenotypic

subpopulation of vimentin(+) CTCs appears to signify the presence of aggressive underlying disease and occult metastases, and may have important implications for treatment selection.

2.4.2 Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer and the second most common cause of cancer related death worldwide.[74] In the United States, the incidence of HCC has nearly doubled over the last two decades, and deaths attributable to liver cancer are increasing at the highest rate of all common cancers.[75] This problem is further compounded by the poor discrimination of current clinicopathologic staging systems and serum biomarkers (e.g. alpha-fetoprotein, AFP) for both early stage patients amenable to surgical therapy, where postoperative recurrence remains a significant challenge, and advanced stage patients receiving systemic chemotherapy or targeted therapy, where predictors of response remain unavailable.[76,77] For patients with this deadly malignancy, the development of better biomarkers to aid in prognostication and treatment selection is a critically urgent, unmet need.

Circulating tumor cells (CTCs) are thought to originate from the primary tumor or metastatic sites, can be detected in the peripheral blood, and have been implicated as a potential cause of post-surgical recurrence and metastases.[66,78] While CTCs can serve as prognostic biomarkers in most solid tumors, studies evaluating CTCs in HCC have found limited utility.[79,80] One important reason is that most CTC enrichment assays, including the FDA-approved CellSearch™ CTC assay, rely on the use of antibodies against the epithelial cell surface marker EpCAM to "capture" CTCs by antigen specific immunomagnetic separation from the remaining blood leukocytes. As only 20-35% of HCCs express EpCAM, methods based on EpCAM alone have resulted in low CTC detection rates and hence limited utility for HCC.[34] Alternative CTC

capture methods utilizing antibodies directed at hepatocyte-specific cell surface markers,[36,80] CD45-depletion,[81] and flow sorting or microfluidic[82,83] systems have all demonstrated increased efficiency in isolating HCC-CTCs. Furthermore, these non-EpCAM based methods have allowed for capture of distinct CTC subpopulations with more mesenchymal properties in HCC.[36,83,84]

The identification and significance of CTC subpopulations expressing a mesenchymal phenotype is an area of active investigation in many solid tumors due to their potential role in metastases.[85] Previous studies in HCC have demonstrated that epithelial-to-mesenchymal transition (EMT), associated with losing expression of cell-cell adhesion markers and gaining the migratory and potentially invasive properties of a mesenchymal cell, is an important step in the metastatic cascade for HCC.[86] Furthermore, several studies in HCC have demonstrated that the overexpression of mesenchymal markers such as vimentin, an intermediate filament, is associated with more advanced tumors and worse prognosis.[87,88] Thus, identifying CTCs that demonstrate an EMT phenotype holds great promise in identifying patients likely to harbor more aggressive underlying disease.

In this study, we sought to investigate the use of a microfluidic, antibody-based CTC capture assay to efficiently capture HCC-CTCs and characterize CTC phenotypes of prognostic importance in HCC. Unlike existing technologies, the NanoVelcro CTC Assay features the combination of a microfluidic system with enhanced topographic interactions and CTC-capture antibody coated nanostructured substrates to allow for the highly efficient separation and capture of HCC-CTCs from background WBCs. The working principle of the NanoVelcro CTC Assay has been validated by its performance to isolate CTCs for many types of solid tumors, including prostate cancer, melanoma, and pancreatic cancer.[37,89] To optimize the use of the NanoVelcro Assay for detecting HCC-CTCs from patient blood, we investigated HCC specific CTC capture

and immunostaining antibodies[80,90,91] followed by validation of their efficacy using a HCC tissue microarray (TMA), HCC cell lines, and a pilot group of HCC patients. Subsequently, utilizing this optimized NanoVelcro Assay in a prospective study of healthy controls, patients with non-malignant liver disease (NMLD), and HCC patients, we 1) identified, characterized, and enumerated HCC-CTCs and the subpopulation of vimentin(+) CTCs, 2) evaluated their ability to discriminate between non-HCC and HCC patients, as well as early stage and advanced stage HCC patients, and 3) evaluated their prognostic utility for cancer progression, recurrence, and survival.

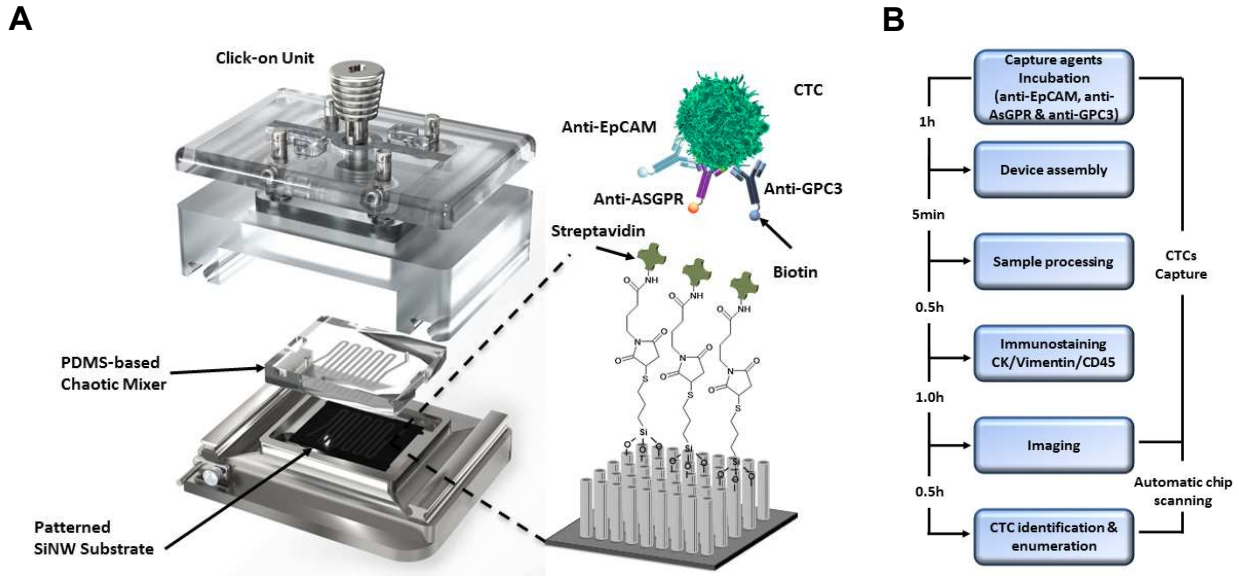


Figure 14. (A) Schematic diagram of the components of the NanoVelcro CTC assay. The click-on unit compresses the PDMS-based chaotic mixer against the serpentine channels of the silicone nanowire chip to create a microfluidic channel which optimizes cellular interactions with the chip surface. In the cutout, the nanowire substrate can be seen functionalized with biotinylated capture antibodies adjoined using streptavidin binding. These capture antibodies are responsible for "capturing" the CTCs on the chip. **(B)** General workflow diagram for CTC capture, identification, and enumeration on SiNS NanoVelcro Chips. The schematic shows the individual components of SiNS NanoVelcro Chips, chip holding device, and elements/duration of the workflow for processing a typical blood sample. SiNW, silicone nanowire; PDMS, Polydimethylsiloxane; EpCAM, epithelial cell adhesion molecule; GPC-3, glypican-3; ASGPR, asialoglycoprotein receptor; CTC, circulating tumor cell.

2.4.3 Materials and methods

Antibody selection for the capture and identification of HCC-CTCs

The capture of CTCs relies on the antigen-specific immobilization of CTCs on the NanoVelcro surface by the use of antibodies directed against HCC-specific cell surface markers. Once immobilized on the NanoVelcro surface, the CTCs are identified and phenotyped based on their immunohistochemical staining characteristics (**Figure 14**). A wide selection of antibodies for both CTC capture, as well as CTC identification and phenotyping, were evaluated using an 8 HCC cell line panel as detailed in the supplemental methods. Antibodies demonstrating optimal staining and specificity across the cell line panel were then tested for suitability on a 114 patient HCC tissue microarray (TMA) generated from archived, resected HCC specimens at the University of California, Los Angeles.[92] Antibodies against the cell surface markers EpCAM (Cell Signaling, Danvers, MA), asialoglycoprotein receptor (ASGPR; Abcam, Cambridge, UK), and glypican-3 (GPC-3; Santa Cruz Biotechnology, Santa Cruz, CA), as well as the cytoplasmic marker vimentin (Abcam), were used to stain the TMA using the EnVision™+ Dual Link Kit (DAKO, Glostrup, DK) according to the manufacturer's protocol (**Figure 15**). Staining intensity was assessed on a 4-point scale (none, weak, moderate, and strong staining) by a single pathologist (S.W.F.), and tumors were considered to have stained positive if they displayed moderate or strong staining (**Figure 16**). TMA staining results were summarized as shown in **Figure 15**.

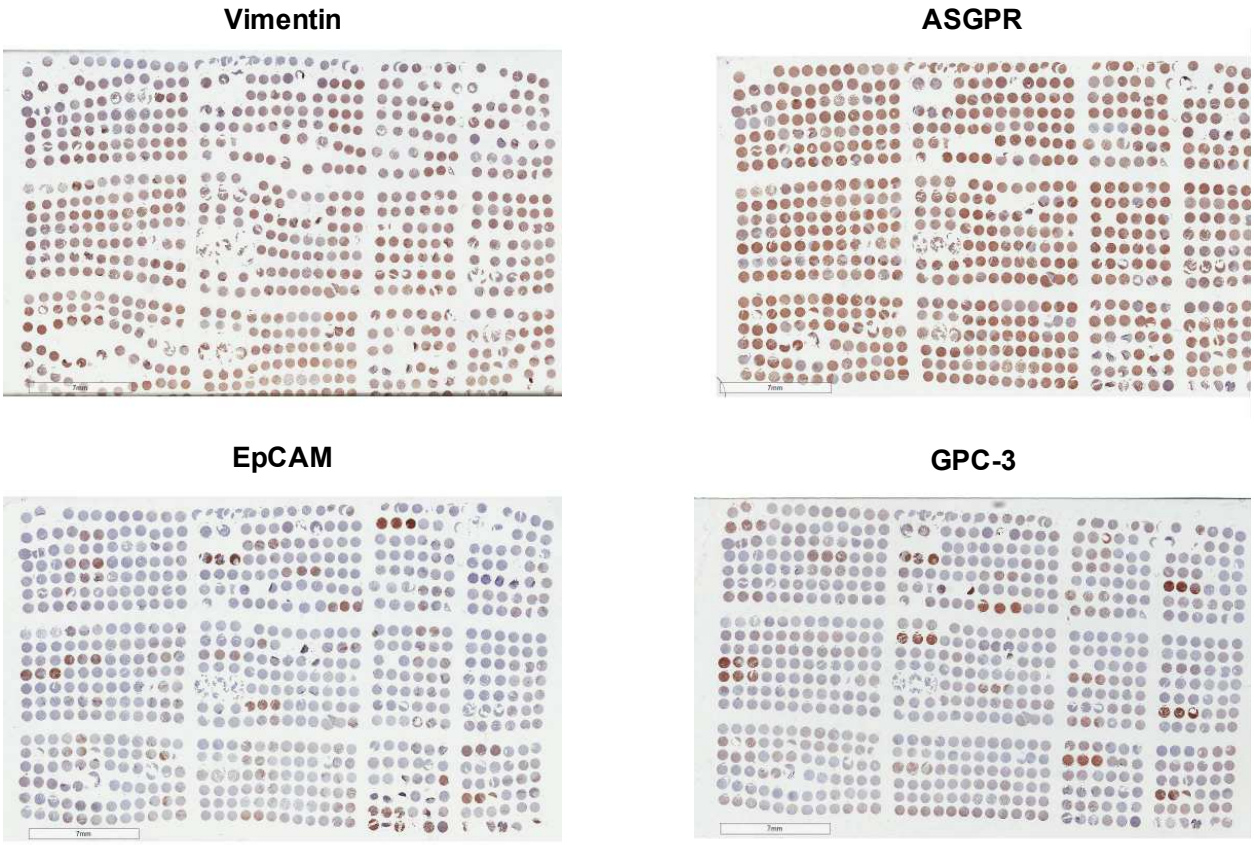


Figure 15. Low power images of the entire tissue microarray stained with vimentin, ASGPR, EpCAM, and GPC-3. Many of the tumors appear positive for vimentin at low power due to staining of the sinusoidal endothelial lining and not the tumor itself. EpCAM, epithelial cell adhesion molecule; GPC-3, glypican-3; ASGPR, asialoglycoprotein receptor.

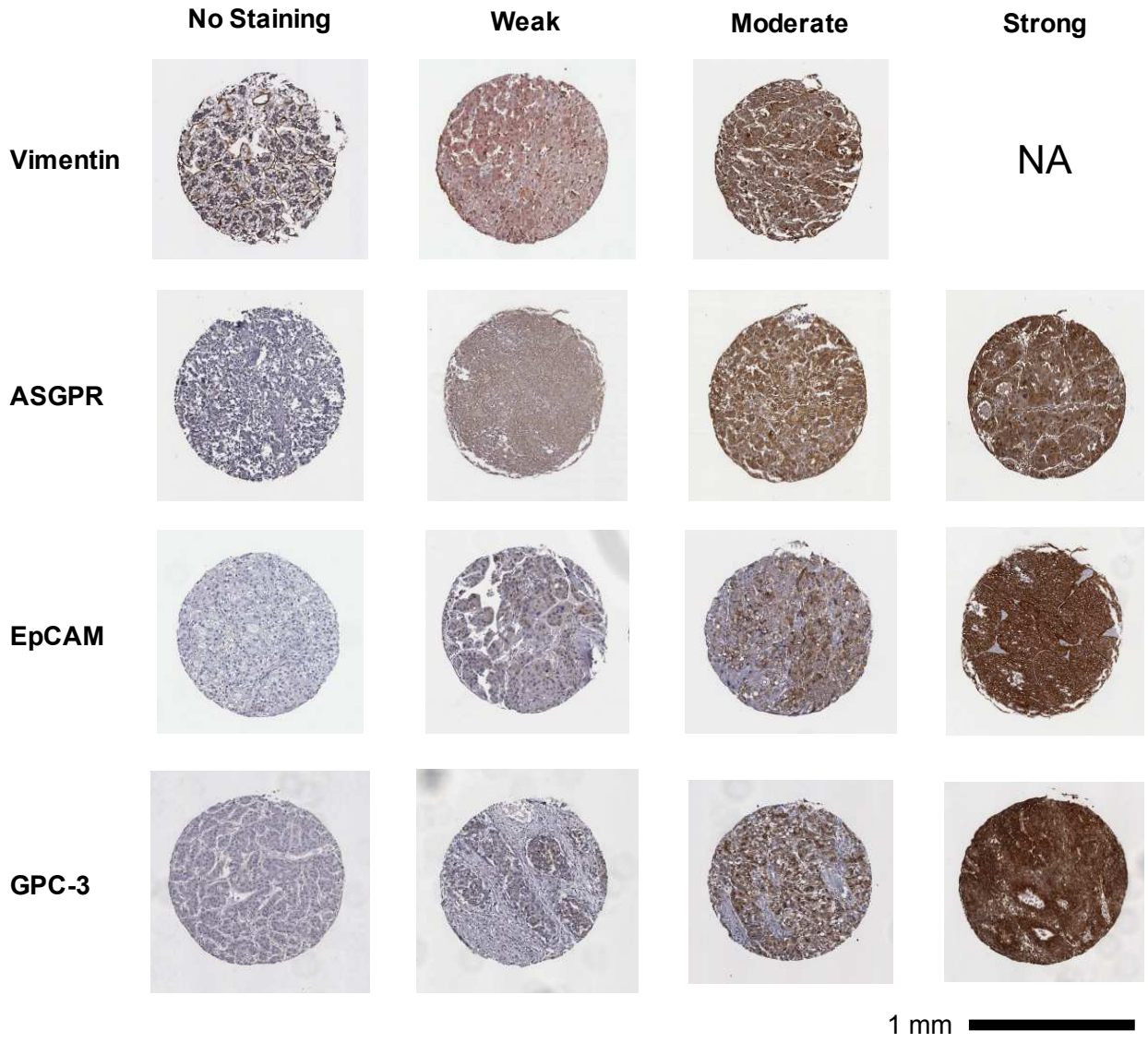


Figure 16. Examples of staining intensity determinations for each of the antibodies used on the 114 patient TMA. None of the tumors stained strongly for vimentin. EpCAM, epithelial cell adhesion molecule; GPC-3, glypican-3; ASGPR, asialoglycoprotein receptor

Optimization of multimarker CTC capture

The NanoVelcro CTC assay utilizes a microfluidic chaotic mixer to enhance CTC interactions with the capture antibody coated nanosubstrate surface of the chip to enhance CTC capture rates (**Figure 14**). Following CTC capture, CTCs are immunostained and then identified and phenotyped using multi-color ICC and cytometric assessment (**Figure 17** & **Figure 18**).

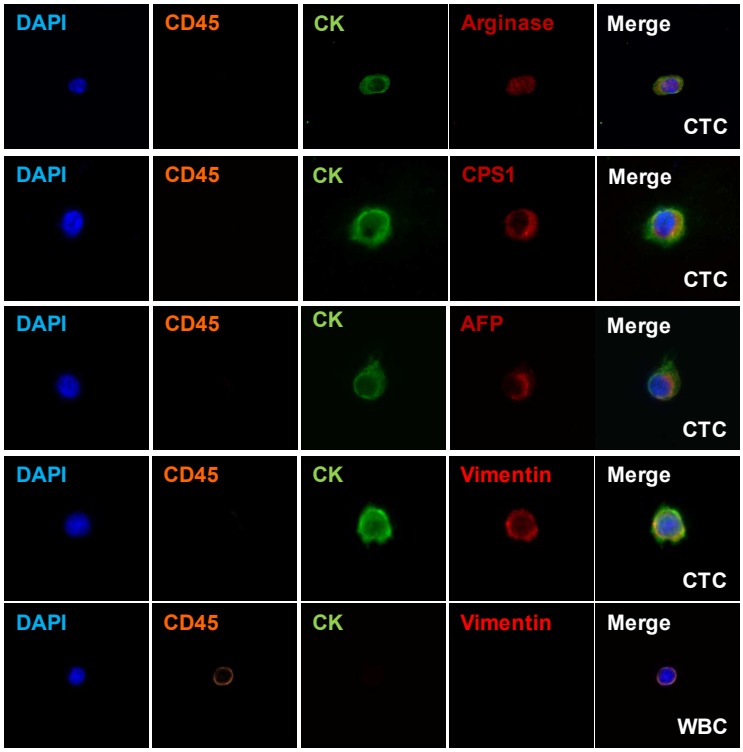


Figure 17. Confirmation of hepatocyte origin of HCC CTCs. Hepatocyte specific markers considered for HCC CTC identification (Arginase, CPS1[Hep-Par1], AFP) were found to only stain CK+ CTCs, not CD45 WBCs. Additionally, EMT (vimentin) markers were tested. All cells were obtained from the blood of patients with HCC and identified on a 4-color immunocytochemical approach in conjunction with high-resolution fluorescent microscopy. DAPI, 4',6-diamidino-2-phenylindole; CK, cytokeratin; CPS1, carbamoyl-phosphate synthase 1; AFP, alpha-fetoprotein.

While the performance of the NanoVelcro CTC assay has previously been reported for many solid tumors, we first validated the methodology for HCC using cell lines as previously described and detailed in the supplemental methods.[67] Briefly, HCC cells were spiked into healthy donor blood to create artificial blood samples. These were run on NanoVelcro CTC chips according to the method outlined below in the Sample Processing section and in the supplemental methods. Each chip is capable of processing 2 mL of blood. For calibration experiments, only 2 mL of blood was utilized; however, for clinical specimens all assays are run as parallel duplicate samples. Thus, CTC counts are recorded per 4 mL of blood except as noted otherwise. All experiments were performed at the optimum flow rate of 1 mL per hour. (**Figure 19**). Calibration experiments using both HCC cell lines (**Figure 20B&C**) and HCC patient samples (**Figure 21A&B**) were performed to compare the efficiency of the multimarker capture cocktail (EpCAM, ASGPR, GPC-3) to each antibody individually.

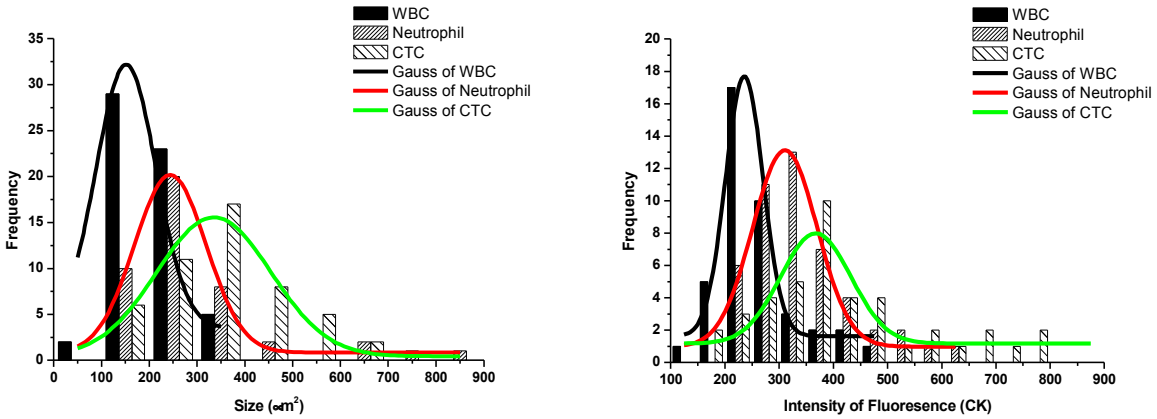


Figure 18. Development of CTC size criteria cutoff. **(A)** Frequency of CTCs, WBCs, and neutrophils by cross-sectional size. **(B)** Frequency of CTCs, WBCs, and neutrophils by CK RFU.

Patient recruitment and blood collection

Between April 2015 and September 2016, healthy controls, patients with non-malignant liver disease (NMLD – cirrhosis without HCC, adenoma, focal nodular hyperplasia), and patients with HCC were enrolled (University of California, Los Angeles IRB #14-001932). Inclusion criteria included pathologic or radiographic (LIRADS-5) diagnosis of HCC, with patients having synchronous or past (within 5 years) extrahepatic malignancies excluded from enrollment. A database of demographic and clinicopathologic information was maintained in a prospective manner, with clinical staging assigned based on either pathologic or radiographic assessment (Milan criteria [MC] [93] and University of California, San Francisco [UCSF] [94] transplant criteria) (Table 6) at the time of study enrollment. Patients were categorized as early stage if their tumors were within UCSF transplant

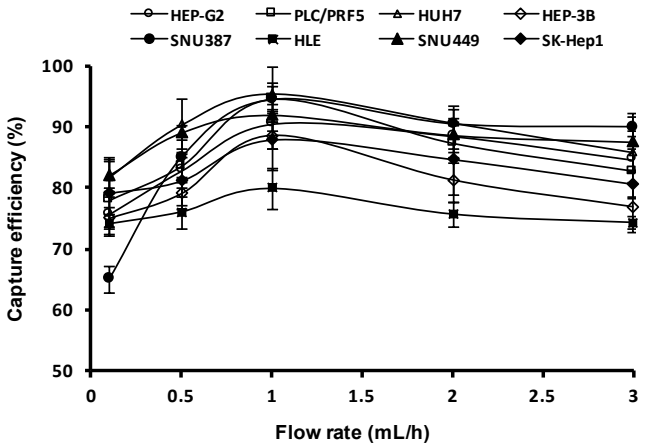


Figure 19. Calibration of optimum microfluidic flow rate for CTC capture by NanoVelcro CTC assay. Using 8 HCC cell lines, the cell capture efficiency of the NanoVelcro Chip was assessed at flow rates of 0.1, 0.5, 1, 2, and 3 mL/h. Data represents the mean ± SD (n = 3).

Table 6. Comparison of radiographic size criteria.

	Milan Criteria	UCSF Criteria
Size Criteria	Single lesion ≤ 5 cm or up to 3 lesions all ≤ 3 cm	Single lesions ≤ 6.5 cm or 2-3 lesions ≤ 4.5 cm and a total tumor diameter ≤ 8cm
Additional Criteria	No vascular or extrahepatic involvement	No vascular or extrahepatic involvement

criteria, locally advanced if they were outside of UCSF but without extrahepatic disease, and metastatic if there was evidence of distant metastases. The modified Response Evaluation Criteria In Solid Tumors (mRECIST) guidelines were used to evaluate patient's disease status as stable or progressing.[95]

Sample processing, chip scanning and CTC enumeration

Blood samples were processed as previously described and detailed in the supplemental methods using the parameters determined by the optimization experiments described above. Captured CTCs were imaged using immunocytochemistry (ICC), allowing for both cytometric and immunofluorescent identification parameters, and criteria for CTC identification were developed by a trained cytopathologist. Our ICC and cytometric criteria for CTC identification were optimized for HCC-CTCs using the methods presented in **Figure 17** & **Figure 18** and described in the methods. Chips were first scanned at 40x power by an automated chip scanning protocol using the NIS Elements 4.1 software (Nikon, Tokyo, Japan) on an Eclipse 90i fluorescent microscope to identify candidate cells. Higher magnification manual imaging of candidate cells was then performed at 400x power to verify results. When analyzing the resulting multi-channel ICC image, WBCs were defined as round/ovoid cells, DAPI+/CD45+/CK-, with size ≤ 6-μm; and HCC-CTCs are defined as round/ovoid cells, DAPI+/CD45-/CK+, with size ≥ 6-μm. Epithelial-to-mesenchymal phenotype, vimentin(+) CTCs are the subpopulation of HCC-CTCs defined as round/ovoid events, DAPI+/CD45-/CK+/vimentin+, with size ≥ 6-μm. Any CD45 positivity greater than 2x background discounted a cell as being a CTC. CTCs were

enumerated by the same blinded researcher (S.H.) and CTC counts are represented as a total count per 4 mL VB (**Figure 22B**).

Statistical methods

Continuous variables were compared using the Wilcoxon rank sum test, while categorical variables were compared using the χ^2 test/Fisher exact test. The Fine and Gray competing risks Cox regression model was used to identify univariate predictors of overall survival (OS), progression-free survival (PFS), and time to recurrence (TTR). Factors significant ($p < 0.2$) on univariate analysis were entered into a multivariate model to identify independent predictors. A two-tailed p-value < 0.05 was considered to be statistically significant.

2.4.4 Results

Antibody selection for the capture and identification of HCC-CTCs

CTCs are first incubated with biotinylated antibody cocktail (EpCAM, ASGPR, GPC-3) and then "captured" by the NanoVelcro assay through the specific interaction of biotin on the CTCs with the streptavidin coated surface of the NanoVelcro chips. To optimize CTC capture we utilized a TMA developed

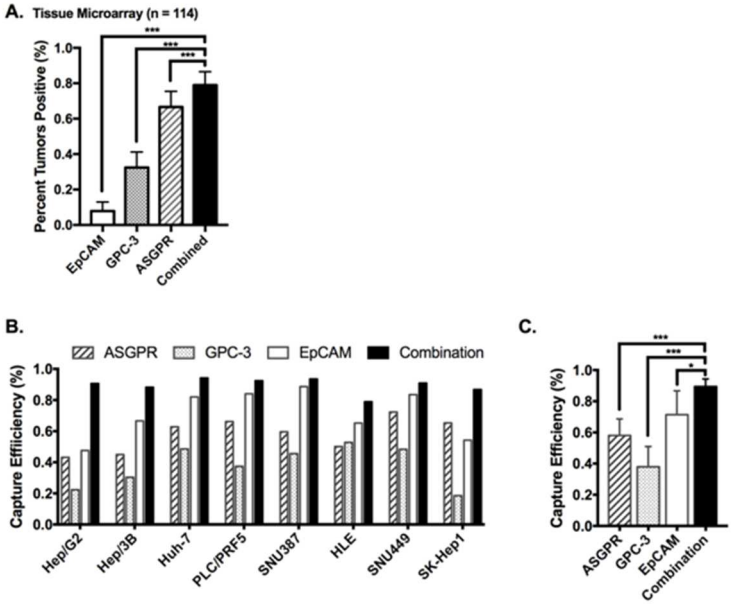


Figure 20. (A) A 114-tumor tissue microarray was used to test antibodies against markers of interest for the study. Percentage of tumors with moderate to strong staining for the given cell surface marker is shown and compared with the percentage of tumors staining for at least 1 of the 3 markers (Combined). **(B)** Capture efficiency of the NanoVelcro CTC capture assay using 8 HCC cell lines. Each individual antibody, as well as the multimarker combination of all 3 antibodies, was compared. The percentage of spiked cells that were captured and identified is reported. **(C)** Summary of capture efficiency data from **(B)** all 8 cell lines. The combination of all 3 antibodies demonstrated superior capture efficiency compared with the use of any single individual antibody alone. * – $P < 0.05$; *** – $P < 0.001$

from 114 resected human HCC samples to screen the potential HCC specific cell-surface capture markers ASGPR and GPC-3, in addition to the widely-utilized cell surface marker EpCAM. Of the 114 tumors, 89 (78.9%) stained for at least 1 of the 3 antibodies, with 65.8% of tumors staining positive for ASGPR, 31.6% for GPC-3, and only 7.9% for EpCAM (**Figure 20A**). However, the combination of all 3 antibodies stained significantly more tumors than any single antibody alone for the entire TMA ($p = 0.0004$ for ASGPR alone and $p < 0.0001$ for EpCAM or GPC-3 alone).

Optimization of the NanoVelcro CTC assay utilizing multimarker CTC capture

Capture antibody optimization in HCC cell lines

Utilizing artificial blood samples, the CTC capture efficiency for each individual capture antibody (EpCAM, ASGPR, GPC-3) was compared to the combination of all 3 antibodies for 8 HCC cell lines (**Figure 20B**). Use of the triple-antibody multimarker capture cocktail resulted in the highest CTC capture efficiency (> 80%), significantly greater than any single capture antibody alone ($p=0.0175$ for EpCAM, $p<0.0001$ for ASGPR and GPC-3, **Figure 20C**).

Capture antibody optimization in human HCC patients

In a preliminary cohort of 10 patients with HCC, we evaluated the total CTCs enumerated using EpCAM alone, ASGPR alone, GPC-3 alone, and the triple multimarker antibody cocktail (**Figure 21A**). Similar to our cell line results, we found that the combination of all 3 capture antibodies resulted in significantly greater CTCs captured compared with the use of any single antibody alone (vs. multimarker capture, $p=0.0408$ for EpCAM, $p=0.0065$ for ASGPR, and $p=0.0066$ for GPC-3, **Figure 21B**).

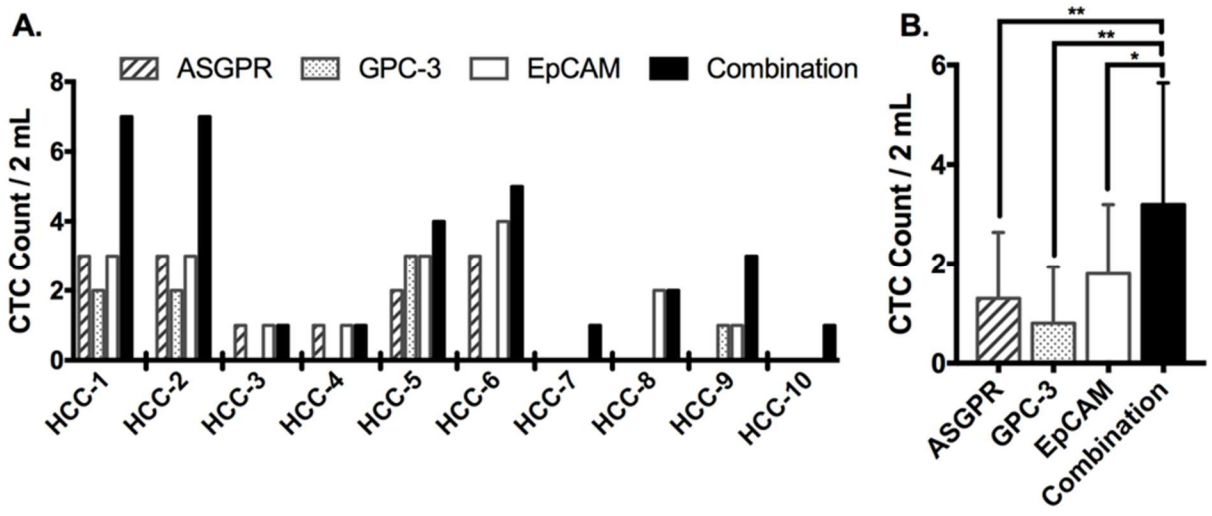


Figure 21. (A) Capture efficiency of NanoVelcro chips functionalized with each antibody individually as well as with the combination of all 3 antibodies for pilot study patient samples (n = 10). **(B)** Summary of capture efficiency data from **(A)** for each cell surface antibody demonstrating superior capture with the combination of all 3 antibodies (vs. each antibody alone, $p < 0.05$). * – $P < 0.05$; ** – $P < 0.01$

Defining HCC-CTC Phenotypes

To ensure that our immunocytochemical staining of HCC-CTCs was specific for hepatocytes, selected patient samples were stained with CK as the primary epithelial marker, and co-stained with hepatocyte-specific markers AFP, arginase, and hep-par-1. AFP, arginase, and hep-par-1 staining was only noted in the CK+ cells and never in the CD45+ leukocytes, confirming the specificity of our ICC criteria for HCC-CTCs (**Figure 17**). Additionally, we discovered a subpopulation of vimentin(+) CTCs with an epithelial-to-mesenchymal phenotype. Based on these initial experiments, all prospective HCC patients enrolled in the study underwent enumeration of HCC-CTCs and the subpopulation of vimentin(+) CTCs (**Figure 22)**.

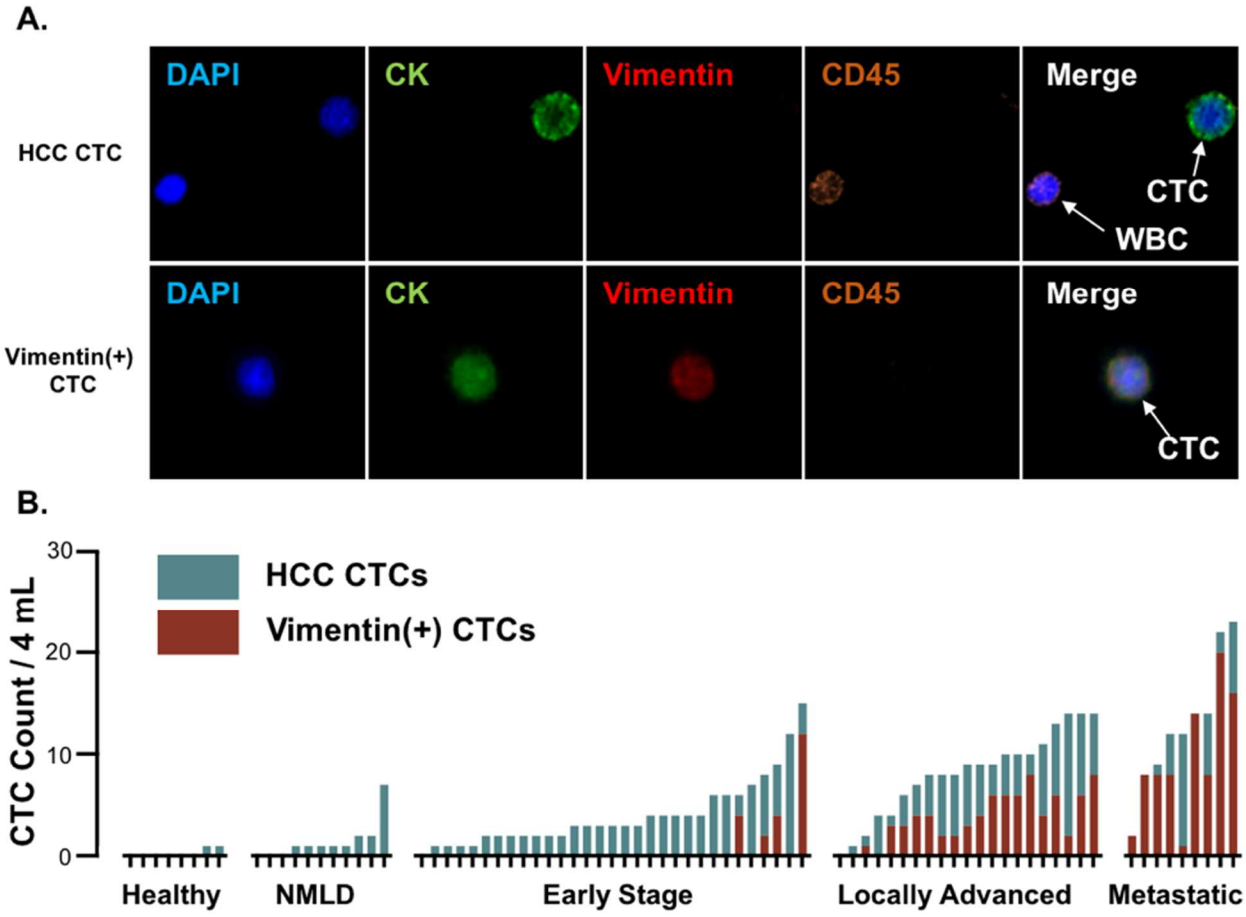


Figure 22. (A) Schematic depicting CTC identification via a 4-color ICC approach in conjunction with high-resolution fluorescent microscopy. Representative images of a HCC CTC and the subpopulation of vimentin(+) CTCs shown at 400x magnification. The WBC staining pattern can be seen in the cell in the lower left corner of the HCC CTC image. **(B)** CTC counts for both HCC CTCs and vimentin(+) CTCs for all patients enrolled in the study (n = 80). Patients are divided into groups based on staging criteria and sorted within groups based on total CTC count.

Prospective study evaluating CTC enumeration and phenotype utilizing optimized multimarker HCC NanoVelcro assay

Of 84 patients approached for study participation, 80 patients underwent peripheral blood draw, CTC enumeration, and phenotyping using our optimized multimarker NanoVelcro HCC-CTC assay (**Table 1**). Four patients were excluded (2–refused informed consent, 1–synchronous cancer, 1– insufficient blood draw). Of the 80 enrolled patients, 61 had HCC, 11 had NMLD (7

cirrhosis without HCC, 4 with benign liver lesions), and 8 were healthy controls (**Figure 23**). Among patients with HCC, 31/61 (50.8%) were early stage (defined as within UCSF transplant eligibility criteria),[94] 21/61 (34.4%) were locally advanced with extensive liver involvement, and 9/61 (14.8%) were metastatic. Patients were followed for a median of 325 days after blood draw. Of the 61 patients with HCC, 28 (45.9%) progressed and 20 (32.8%) died.

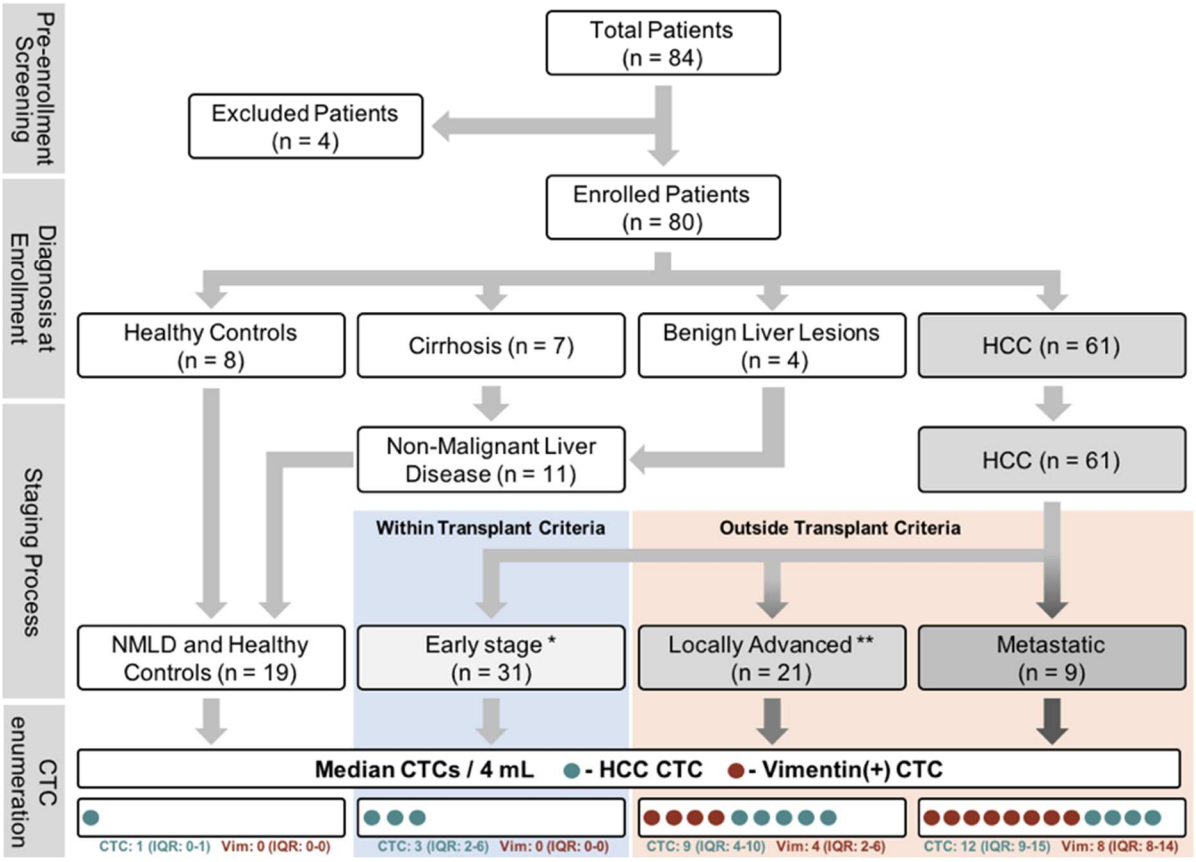


Figure 23. Diagnosis and staging flowchart for patients enrolled in the study as well as associated CTC counts listed as median (IQR) for HCC CTCs and Vimentin+ CTCs. * - Early Stage defined as patients whose lesions are within the UCSF radiographic staging criteria; ** - Locally Advanced defined as patients whose lesions are larger than UCSF radiographic staging criteria but without evidence of metastatic disease.

HCC-CTCs

HCC-CTCs were found in 59/61 (96.7%) patients with HCC (median=6, range=0-23). Occasional patients with NMLD were found to have low numbers of CTCs (median:1, range:0-

7), particularly those with inflammatory adenomas (**Figure 22B**). A single CTC was found in 2/8 (25%) healthy control patients. Among patients with HCC, CTC count correlated with stage, with a median of 3 CTCs in early stage patients (range:0-15), 9 CTCs in locally advanced HCC (range:0-14), and 12 CTCs in patients with metastatic HCC (range:2-23) (**Figure 23**).

At the optimum cutoff of ≥ 2 CTCs/4 mL VB, HCC-CTC enumeration accurately discriminated among patients with and without HCC (NMLD or healthy controls) with a sensitivity of 84.2%, specificity of 88.5%, PPV 69.6%, NPV 94.7% (**Figure 24A**) and an area under the ROC curve (AUROC) of 0.92 (95%CI = 0.86–0.99, p<0.0001), illustrating the ability of CTCs to discriminate patients with cancer from those without (**Figure 24B**).

Univariate analysis of all HCC patients revealed that HCC-CTCs were associated with worse PFS (HR:4.91, 95% CI: 2.52 – 9.54, p < 0.0001) and OS (HR: 1.96, 95% CI = 1.12 – 3.42, p = 0.018). In the subset of non-metastatic, potentially curable patients (n = 30) who underwent locoregional therapy (resection, n=9; transplantation, n=5; RFA, n=11; or TACE, n=5), HCC-CTCs were associated with shorter TTR (HR 9.7, 95% CI = 2.08 – 45.19, p = 0.004). In patients with incurable, locally advanced (n= 15, diffuse infiltration or macrovascular invasion) or metastatic disease (n=8), HCC-CTCs were associated with worse PFS (HR 2.09, 95% CI = 1.11 – 3.96, p = 0.023). On multivariate analysis, including the covariates Age, AFP, MELD score and tumor size, HCC-CTCs were again found to be significantly associated with PFS (HR 3.11, 95%CI = 1.45-6.65, p=0.003), but not TTR or OS.

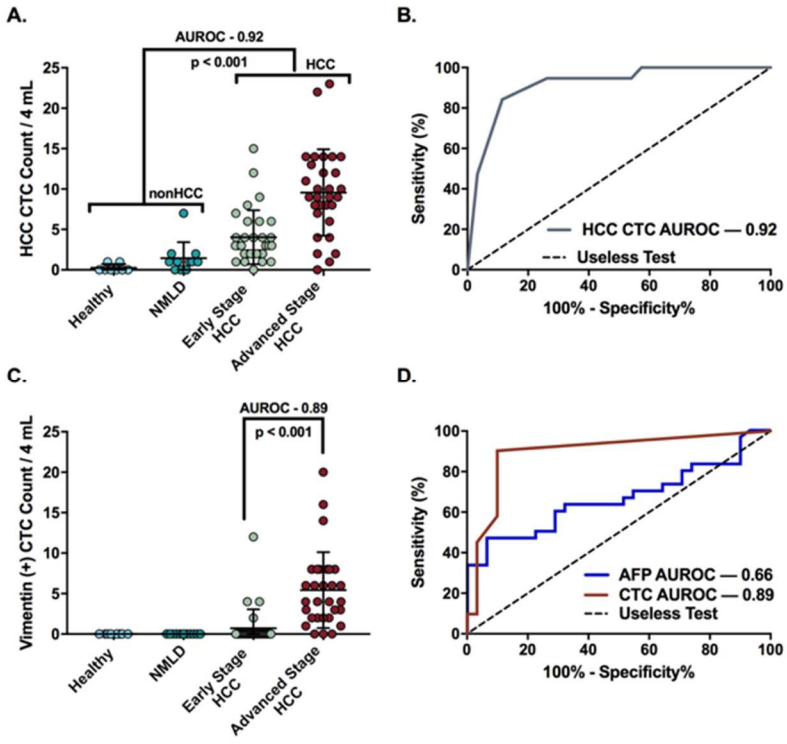


Figure 24. (A) CTC enumeration in patients with local or advanced HCC versus NMLD or healthy controls. Advanced stage patients include patients with locally advanced disease and metastatic disease**. (B)** ROC curve for illustration of HCC CTC performance in the discrimination of HCC from NMLD or healthy controls. At a cutoff of ≥ 2 CTCs / 4 mL VB, HCC CTC AUROC = 0.92 (95% CI: 0.86 – 0.99 , p < 0.001). **(C)** Vimentin(+) CTC enumeration in patients with early or advanced stage HCC versus nonHCC patients. **(D)** Comparison of the performance of vimentin(+) CTCs and AFP in discriminating early, potentially curable, HCC from advanced stage HCC. At a cutoff of ≥ 1 vimentin(+) CTCs / 4 mL VB, AUROC 0.89 (95% CI: 0.74 – 0.95, p < 0.001) versus AFP (cutoff: ≥ 1600 ng/mL), AUROC 0.66 (95% CI: 0.53 – 0.78, p = 0.021).

Vimentin(+) CTCs

Vimentin(+) CTCs were found in 31 (50.8%) patients with HCC (median:1, range:0-20) and never in patients with NMLD or healthy controls (**Figure 24C**). Among HCC patients, the number of vimentin(+) CTCs correlated with increasing tumor stage, with only 4/31 (12.9%) early stage patients demonstrating vimentin(+) CTCs (median:0, IQR:0-0, range:0-12), compared to 18/21 (85.7%) locally advanced patients (median:4, IQR:2-6, range:0-8), and 9/9 (100%) metastatic HCC patients (median:8, IQR:8-14, range:1-20).

At a cutoff of ≥ 1 CTC/4 mL VB, vimentin(+) CTCs accurately discriminated transplant eligible HCC patients with early stage, potentially curable disease from transplant ineligible patients with locally advanced or metastatic HCC (**Figure 24C**), with a sensitivity of 87.1%, specificity of 90.0%, PPV of 90.0%, NPV 87.1%, and an AUROC of 0.89 (95%CI = 0.74–0.95, p<0.0001) (**Figure 24D**). For comparison, we performed the same analysis for AFP, the only

widely available HCC biomarker. AFP discriminated early stage from advanced stage disease with a sensitivity of 93.5%, specificity of 46.7%, PPV of 65.6%, NPV of 65.5%, and AUROC of 0.66 (95% CI = 0.53 – 0.78, p = 0.021). Thus, vimentin(+) CTCs were a significantly better predictor of advanced stage disease than AFP (DeLong's test, Z = 2.6, p = 0.0092).[96]

Univariate analysis of all HCC patients revealed that vimentin(+) CTCs were highly associated with inferior OS (HR: 2.21, 95%CI = 1.38-3.52, p=0.001) and PFS (HR:3.91, 95%CI = 2.35-6.49, p<0.0001) (**Figure 25A&B**). Perhaps more notably, CTCs were able to discriminate outcomes in the subset of potentially curable patients undergoing locoregional therapy (resection, n=9; transplantation, n=5; RFA, n=11, and TACE, n=5). In this subset of 30 patients who radiographically demonstrated no residual disease following treatment, the presence of any pre-treatment vimentin(+) CTCs were associated with shorter TTR (HR:3.14, 95%CI = 1.50-6.57, p=0.002; **Figure 25C**). Similarly, in the subset of incurable locally advanced or metastatic patients (n=23), vimentin(+) CTCs were predictive of worse PFS (HR:1.81, 95%CI = 1.02-3.22, p=0.043; **Figure 25D**). On multivariate analysis, vimentin(+) CTCs were again found to be significantly associated with PFS (HR:2.16, 95%CI = 1.33-4.42, p=0.002) and OS (HR:2.21, 95%CI = 1.38-3.56, p=0.001)

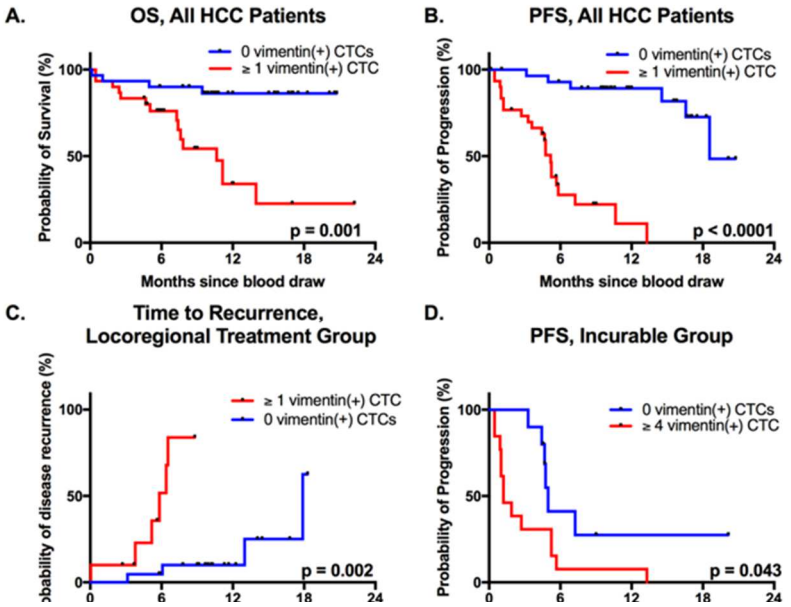


Figure 25. (A) Overall survival at the optimized cutoff of ≥ 1 vimentin(+) CTC, all HCC patients. **(B)** Progression-free survival at the optimized cutoff of ≥ 1 vimentin(+) CTC, all HCC patients. **(C)** Time to recurrence following potentially curative locoregional therapy at the optimized cutoff of ≥ 1 vimentin(+) CTC (n = 30). **(D)** Progression-free survival at the optimized cutoff of ≥ 4 vimentin(+) CTCs, for patients with locally advanced or metastatic disease not amenable to potentially curative locoregional therapy

with a trend towards inferior TTR (HR:2.45, 95%CI = 0.91-6.57, p=0.076).

Potential utility of HCC-CTCs

The value of HCC-CTCs and vimentin(+) CTCs as HCC blood biomarkers goes beyond initial prognosis, presenting potential utility for both longitudinal disease monitoring as well as appropriate treatment selection. 11 patients in the study underwent serial blood draws, CTC enumeration, and phenotyping over the course of treatment, with disappearance of CTCs following successful tumor resection and ablation, and a subsequent reappearance of CTCs prior to clinical recurrence. One such example is illustrated in **Figure 26A.** The patient was a 63-year-old gentleman with compensated cirrhosis and a 5.8cm arterially enhancing biopsy-proven HCC. His pre-resection blood draw revealed 8 vimentin(+) CTCs (14 total HCC-CTCs). He underwent a partial right hepatectomy which revealed an 8.5cm dominant lesion with a subcentimeter satellite lesion with microvascular invasion. CTC enumeration at 1- and 2-months post-resection revealed no CTCs, with MRI imaging revealing no evidence of recurrence. However, his 3rd post-resection blood draw revealed 6 vimentin(+) CTCs (12 total HCC-CTCs), with subsequent MRI demonstrating a segment 8 HCC recurrence.

The presence of vimentin(+) CTCs consistently portended faster time to recurrence following locoregional treatment **(Figure 25D)**, with an example illustrated in **Figure 26B.** The patient had a solitary 4cm right hepatic lobe HCC without evidence of metastases, but despite being radiographically staged as an early stage patient, was found to have 12 vimentin(+) CTCs (15 HCC-CTCs total). Despite successful locoregional therapy with TACE, she was found to have multifocal HCC on her 1-month post-procedure scan with rapid development of metastatic lung nodules at 3 months' post-procedure.

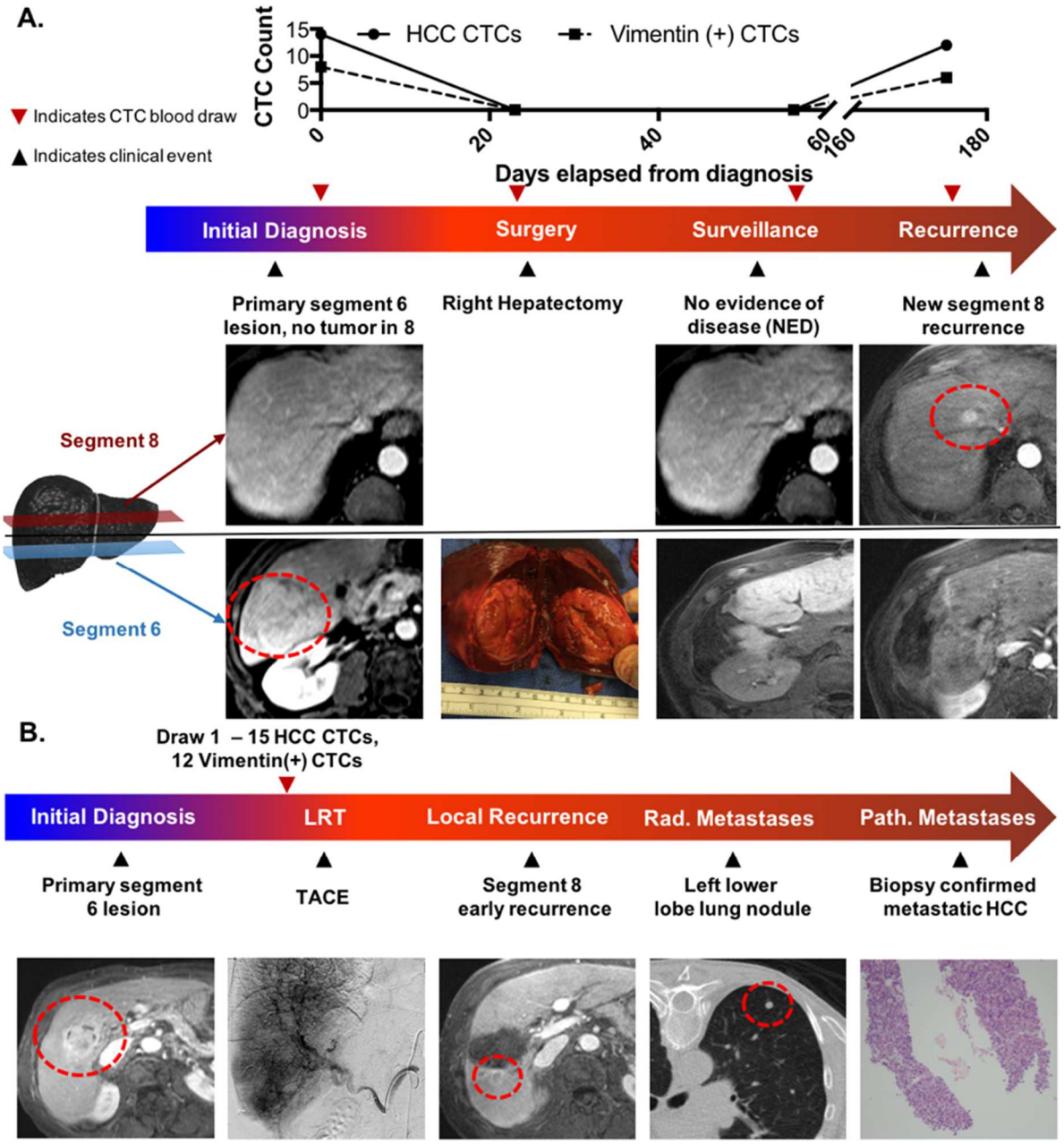


Figure 26. Demonstration of the potential utility of the HCC CTC Assay **(A)** Disease monitoring after surgical resection. Pre-surgical blood draw revealed 8 vimentin (+) CTCs (14 total HCC CTCs), with subsequent CTC enumeration revealing no CTCs at 1- and 2-months following resection of a 5.8 cm HCC, consistent with surveillance MRI imaging revealing no evidence of recurrence. A 3rd post-resection blood draw revealed 6 vimentin (+) CTCs (12 total HCC CTCs), presaging the subsequent MRI demonstration of a segment 8 HCC recurrence. **(B)** Vimentin (+) CTCs are highly predictive of earlier recurrence following potentially curative locoregional therapy, and appear to be a surrogate for patients who have tumors with aggressive underlying tumor biology. This patient had a solitary 4cm right hepatic lobe HCC without evidence of metastases. Despite being radiographically staged as an early stage patient, she was found to have 12 vimentin(+) CTCs (15 CTCs total) prior to any treatment. Despite successful locoregional therapy with TACE, she soon thereafter developed multifocal HCC on her 1-month post-procedure scan with rapid development of metastatic lung nodules at 3 months post-procedure.

2.4.5 Discussion

Hepatocellular carcinoma, a deadly malignancy typically arising in the setting of underlying liver disease, has now become the second leading cause of cancer deaths globally.[74] The limitations of existing clinicopathologic staging systems of HCC is evident in the high recurrence rate following locoregional therapies or surgical interventions such as resection or transplantation. CTCs are emerging as a promising biomarker for several cancer types, but their application to HCC has been limited when utilizing existing assays that rely on epithelial cell-surface markers alone for CTC capture. In this study, we present the development and optimization of a novel multimarker CTC capture platform that allows for the identification, enumeration, and analysis of HCC-CTCs with high sensitivity and specificity. Furthermore, we identify a phenotypic subpopulation of vimentin(+) CTCs, which are highly associated with advanced or metastatic HCC, increased recurrence after potentially curative therapy, and inferior progression-free and overall survival.

A unique strength of our study was the utilization of a multimarker capture antibody cocktail based on the combination of EpCAM, ASGPR, and Glypican-3 that allowed for the detection of CTCs across all HCC stages with high sensitivity and specificity. The majority of CTC enrichment platforms rely on antibodies to the cell-surface marker EpCAM alone, potentially limiting their utility in HCC where expression of EpCAM is reported in only 20-35% of tumors.[34] ASGPR, a transmembrane cell-surface protein highly expressed in well-differentiated HCCs, has been shown to yield high CTC capture rates in previous HCC studies, and was used in our antibody cocktail. [36,80,90] Based on our TMA studies, we chose to also utilize GPC-3 for CTC capture, especially since its expression has been associated with the presence of poorly differentiated HCC,[97] arguably the most important subset of tumors to be detected given

their poor prognosis. This unique three-marker antibody cocktail allowed us to detect CTCs from 96.7% of all patients with HCC, in comparison to the 20-50% typically captured by the use of epithelial markers alone.[79,98-101]

Arguably our most important finding was the identification of a subpopulation of vimentin(+) CTCs and their association with the presence of more advanced stage disease, faster HCC progression and inferior survival for both all patients and the subset of potentially curable patients undergoing locoregional treatments. Vimentin is an intermediate filament ubiquitously expressed in normal mesenchymal cells and is involved in multiple cellular processes including stress resistance and structural integrity.[102] Vimentin overexpression is widely regarded as the canonical marker of EMT in epithelial cancers [102], and tumors expressing vimentin demonstrate accelerated tumor growth and increased invasiveness. In HCC, vimentin overexpression, indicating the acquisition of an EMT phenotype in the primary tumor, has been linked to more aggressive tumor biology, inferior survival outcomes, and the establishment of the tumor initiating capacity critical for metastases.[86,87]

The importance of vimentin overexpression in CTCs as a marker of burden of disease and metastatic potential has been established previously in several cancer types, but data regarding vimentin in HCC-CTCs has been sparse.[85] Our finding that vimentin(+) CTCs were found almost exclusively in patients with advanced stage HCC corroborates previous reports in breast and prostate cancer that found the presence of vimentin(+) CTCs as a marker of metastatic disease and worse outcomes.[103] Prior studies investigating EMT in HCC-CTCs have demonstrated an association between vimentin(+) CTCs and both tumor size and vascular invasion; however, they have been of limited size or lacked sufficient follow up to assess the prognostic significance of

EMT-markers such as vimentin.[36,81] Our study supports these previous findings, demonstrating a correlation between vimentin(+) CTCs and tumor size as well as burden of disease.

Furthermore, our study demonstrates the potential clinical utility of vimentin(+) CTCs as both a prognostic biomarker for recurrence, progression, and survival, as well as an adjunct staging biomarker for identifying patients with more advanced stage disease. In this regard, we demonstrated that vimentin(+) CTCs vastly outperform the only existing serum biomarker, AFP, for both HCC staging and prognosis (**Figure 24D**). In particular, our finding that vimentin(+) CTCs are highly predictive of earlier recurrence and worse survival in otherwise indistinguishable early stage patients undergoing locoregional therapy highlights their potential utility as a biomarker for liver transplant candidate selection. It is precisely this limitation of the existing clinicopathologic staging systems in distinguishing cancer specific outcomes within the subset of potentially curable patients that may be overcome with the use of vimentin(+) CTCs as an adjunct biomarker.

Despite the limited size of our prospective cohort, the potential value of HCC-CTCs as a tool for both longitudinal monitoring of disease and identification of patients with aggressive underlying disease is apparent. Given the ease and repeatability of our non-invasive HCC-CTC assay, the reliable serial monitoring of a patient's response to treatment would be possible. In patients undergoing curative intent surgical resection and ablation, total HCC-CTCs dropped significantly after treatment, but increased in the subset of patients who subsequently proved to have recurrence on follow up imaging (**Figure 26**). In addition, Vimentin(+) CTCs in particular were highly predictive of recurrence and progression in the subset of potentially curable patients undergoing successful locoregional therapy who initially demonstrated no evidence of disease on post-treatment imaging. This highlights the potential role of vimentin(+) CTCs for the selection

of patients undergoing curative liver transplantation, where the identification of patients likely to have poor outcomes is a critical, unmet need to avoid the loss of scarce donor allografts. We plan to further investigate these potential roles with ongoing recruitment, enumeration, and phenotyping of HCC-CTCs in early stage patients undergoing curative intent treatment.

In conclusion, we developed a HCC-specific multimarker antibody-based CTC capture assay that allows for the identification of HCC-CTCs as well as a distinct subpopulation of vimentin(+) CTCs. We first optimized and validated the assay's functionality for capturing HCC-CTCs using both spiked cell line and clinical samples. In a subsequent prospective study of 80 patients, 61 of whom had HCC, our assay allowed for detection of HCC-CTCs from nearly all patients with HCC, with highly accurate discrimination between patients with HCC and those with NMLD or healthy controls. Most importantly, a phenotypic subpopulation of EMT-type, vimentin(+) CTCs allowed for accurate discrimination of patients with early stage, potentially curable HCC from those with advanced stage, incurable disease. The presence of any vimentin(+) CTCs was associated with earlier recurrence and inferior progression free survival in the subset of patients undergoing curative intent locoregional therapy. Our early results indicate that the novel NanoVelcro CTC assay is effective for HCC-CTC capture and phenotyping, with vimentin(+) CTCs showing great promise for identifying early stage patients with occult aggressive disease. Prospective validation of our findings in a larger cohort of patients may allow for the adoption of an HCC-CTC assay as an important biomarker in this deadly malignancy.

2.4 Conclusions

Our use of a tissue-specific CTC capture and phenotyping approach for GI cancers allowed us to capture CTCs from the majority of patients with PDAC (78% of 100 patients) and HCC (96.7% of 61 patients). These results are significantly higher than in prior studies using the CellSearch system, and are reflective of the advantages of both our microfluidic assay and tissue-specific approach.

For both PDAC and HCC we found that CTC counts correlated with increasing stage of disease (**Figure 12A & Figure 22B**), and that CTCs could accurately discriminate patients with non-malignant pancreatic or liver diseases from those with PDAC or HCC, respectively. For PDAC, we found that CTCs not only functioned well as an adjunct diagnostic biomarker, but also demonstrated prognostic significance for survival as well **(Figure 8 & Figure 13**). Similarly, for HCC patients, the presence of HCC-CTCs was associated with worse overall and progression-free survival (**Figure 22**). CTCs outperformed existing biomarkers, CA19-9 for PDAC and AFP for HCC, by wide margins, highlighting both the potential of CTCs and the limited utility of existing biomarkers in GI cancers.

Two of our results are of particular interest and are actively being pursued in larger studies. For PDAC, we found CTCs could predict patients likely to have occult metastatic disease, metastatic disease below the imaging threshold of cross-sectional imaging. At a cutoff of ≥ 3 CTCs/4mL, CTCs correctly identified patients with occult metastatic disease preoperatively (AUROC: 0.82, 95%CI: 0.76–0.98, $p < 0.0001$), significantly outperforming CA19-9 which had a sensitivity of 60.0%, specificity of 77.3%, PPV of 81.0%, NPV of 54.5%, and AUROC of 0.62 (95%-CI: 0.37–0.87, p=0.084). For HCC, we found that at a cutoff of ≥ 1 CTC/4 mL VB, vimentin(+) CTCs accurately discriminated transplant eligible HCC patients with early stage,

potentially curable disease from transplant ineligible patients with locally advanced or metastatic HCC, with a sensitivity of 87.1%, specificity of 90.0%, PPV of 90.0%, NPV 87.1%, and an AUROC of 0.89 (95%CI = 0.74–0.95, $p<0.0001$). This is a significant improvement over AFP for this metric, which had a sensitivity of 93.5%, specificity of 46.7%, PPV of 65.6%, NPV of 65.5%, and AUROC of 0.66 (95% CI = 0.53 – 0.78, $p = 0.021$). Thus, vimentin(+) CTCs were a significantly better predictor of advanced stage disease than AFP (DeLong's test, $Z = 2.6$, $p = 0.0092$). Both findings are of particular interest to surgical oncologists as they help differentiate patients with early stage disease that is potentially amenable to curative intent surgical resection from those with more advanced stage disease that are more likely to benefit from systemic therapies. Biologically, CTCs make sense as a biomarker for surgical oncology, as they are thought to represent the intravasated tumor state, between the formation of an invasive cancer and the development of distant metastases. Our finding that CTCs are found more often and at higher numbers in patients with occult metastatic disease in PDAC or advanced HCC supports the intravasated tumor state model and explains the clinical utility of CTCs.

We used immunocytochemistry (ICC), which allows for both IHC and cytometric criteria, for CTC identification, allowing us to develop a stringent CTC definition. Studies using different CTC technologies have reported vastly different CTC counts from the same volume of blood; from the 1-10 CTCs that we report and most other studies have found, to as high as 10-20,000 CTCs from a similar sized blood draw.[104] Part of this range is due to the subjective nature of ICC definitions. Prior studies have shown that CK^+ white blood cells have been found, either due to staining of macrophages or long sample storage times,[30] and several studies have reported CK^+ circulating epithelial cells in patients with inflammatory conditions such as inflammatory bowel disease or benign cystic pancreatic lesions.[19,20] If CTCs are going to translate into a useful

clinical assay, and especially as a form of "liquid biopsy," this potential for false positives must be addressed by confirming the tumor origin of the CTCs identified. In *Chapter 3*, we utilize our CTC assays developed in *Project 2*, to confirm the tumor origin of the CTCs we isolated as a first step in developing CTCs into a liquid biopsy for GI cancers.

Chapter 3 Confirmation of Tumor Origin of CTCs

3.1 Introduction

Circulating tumor cells (CTCs) are defined as cells of tumor origin that circulate in the blood; however, the subjective nature of microscopy, and in particular immunofluorescence and immunocytochemistry (ICC), means that CTC definitions can vary widely. When combined with vastly different blood input requirements, ranging from as low as 1 mL to as high as 1500 mL, the result is that comparisons of CTC studies and definitions is nearly impossible.[105] The resulting lack of a universal standard ICC definition of a CTC leads to questions regarding the possibility of false positives, especially given that prior studies have identified false positives from macrophages or prolonged sample processing time.[30,106,107] This uncertainty has led researchers to call for the validation of CTC definitions using an orthogonal approach such as molecular analysis.

We sought to confirm the tumor origin of the CTCs we isolated using our PDAC and HCC specific CTC assays molecularly. We did so by demonstrating the congruence of CTC sequencing results to that obtained from the primary tumor. This was made possible by the use of p-NanoVelcro chips, a modification to the NanoVelcro assay used in *Project 1* that allows for the isolation of identified CTCs by laser capture microdissection (LCM). Isolated CTCs can then be studied using molecular techniques such as Sanger sequencing or next generation sequencing (NGS). For PDAC, we did so by comparing the *KRAS* mutation subtype between the primary tumor and CTCs. The incidence of *KRAS* mutations in PDAC, greater than 95%, is among the highest for any point mutation in any cancer.[51] Moreover, 98% of the KRAS mutations found in PDAC are located in codon 12. Given the nearly uniform occurrence of KRAS mutations, tumor origin of captured CTCs can be inferred for most patients by sequencing of a single gene. For

HCC, we performed both aCGH and low-resolution whole genome sequencing to establish copy number profiles for CTCs, primary tumor tissue, peritumoral normal liver, and germline DNA from whole blood. We then utilized copy number profiling to demonstrate the tumor origin of CTCs by showing that the copy number profiles of CTCs are somatic in origin, seen in the primary tumor but not the peritumoral normal or whole blood. Overall, our methods allow us to confirm the accuracy and reliability of CTC molecular analysis for a variety of different sequencing assays and GI cancer types. In so doing, we confirmed that our ICC definitions of a CTC are highly specific, and that molecular analysis of the isolated CTCs reflects the genetic changes found in the primary tumor.

3.2 Methods

Cell line culture and preparation

Cell lines for both HCC and PDAC were prepared as outlined in **sections 2.2.3** and **2.3.3**. Artificial blood samples were made as outlined in those sections and consisted of an average of 1000 cells per mL of blood.

Silicon NanoVelcro chip fabrication and processing

NanoVelcro chips were fabricated and processed using the methods described in sections 2.2.3 and 2.3.3 for PDAC and HCC, respectively. All CTC counts are reported per 4 mL VB. Captured CTCs were imaged using immunocytochemistry (ICC), allowing for both cytometric and immunofluorescent identification parameters, and criteria for CTC identification were developed by a trained cytopathologist. Our ICC and cytometric criteria for CTC identification were optimized for PDAC-CTCs and HCC-CTCs using the methods described previously.

Materials for construction of PLGA nanofibers

Poly(lactic-co-glycolic acid) (PLGA) used in the present study was purchased from DURECT Corporation, Cupertino, CA (B6013-2, lactide : glycolide = 1 : 1, inherent viscosity range 0.55 – 0.75 dL g^{-1}). Hexafluoroisopropanol (≥ 99%) employed for dissolving PLGA was supplied by Sigma-Aldrich and used as received without alteration. To covalently conjugate streptavidin onto PLGA nanofibers through NHS chemistry, activation agents, i.e., 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide (EDC) and N-hydroxylsulfosuccinimide (sulfo-NHS), were obtained from Sigma-Aldrich. Laser microdissection (LMD) slides with pre-deposited 1.2-μm-thick poly(phenylene) sulfide PPS-membrane (Leica) were employed as the substrates for deposition of electrospun PLGA nanofibers.

Construction of p-NanoVelcro chips

We employed a previously described electrospinning method for the deposition of PLGA nanofibers onto a commercially available LMD membrane slide.[31] The PLGA working solution (10% w/v) was prepared by dissolving 1 g of PLGA pellets (DURECT Corporation) into 10 mL hexafluoroisopropanol (Sigma-Aldrich). The working solution was then loaded into a glass syringe (Hamilton) and continuously ejected through a nozzle at a constant flow rate (0.2 μL h^{-1}) using an infusion pump (TS2-60, Baoding Longer). Meanwhile, an electric field of 15 kv/cm was applied between the nozzle and the ground collector by a high voltage power supply (Tianjin Dongwen high-voltage power supply co., ltd, China). LMD membrane slides were placed in the middle of grounded collector for given durations to generate randomly orientated PLGA

nanofibers with the desired density and thickness. The resulting p-NanoVelcro substrates were dried overnight in a vacuum chamber at room temperature.

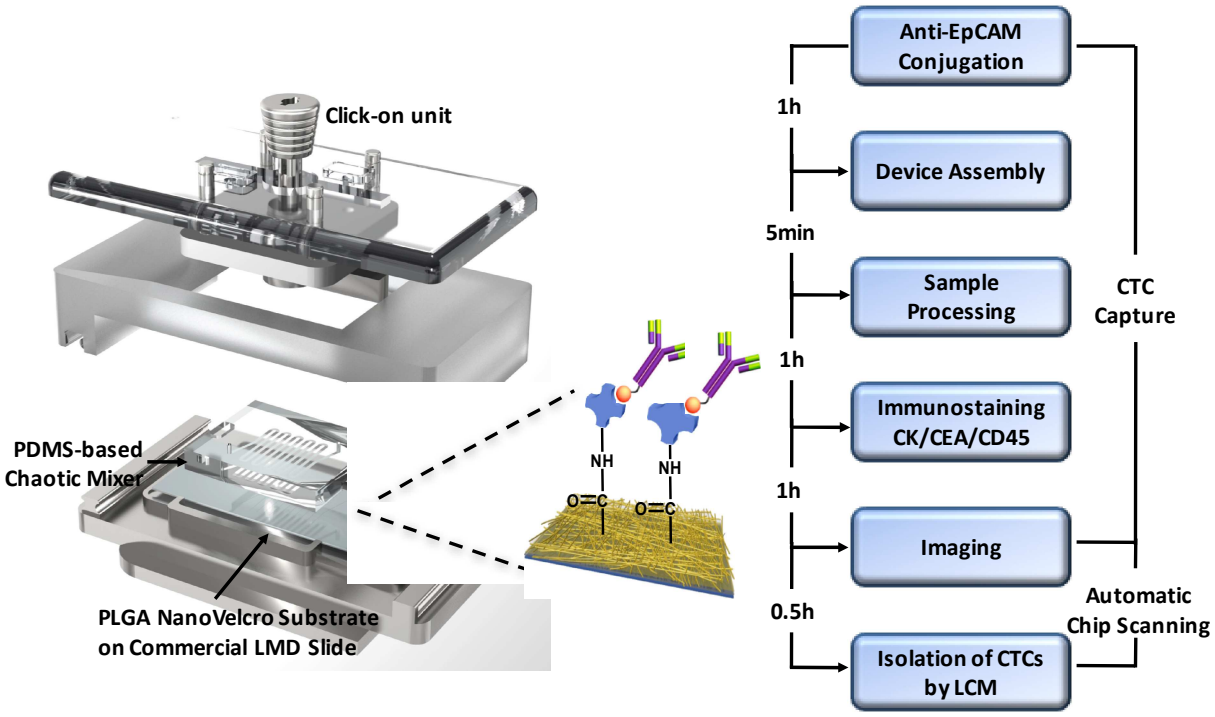


Figure 27. General workflow diagram for CTC capture, identification, and LCM of single CTCs on P-NanoVelcro/LCM Chips. Individual components of the NanoVelcro/LCM chips and holding device are shown in the schematic at the right.

Laser capture microdissection of single CTCs (cell lines), whole genome amplification, and directed *KRAS* PCR

The feasibility of utilizing p-NanoVelcro chips for single cell mutational analysis was first tested on 2 PDAC cell lines (AsPC-1 and CFPAC-1). p-NanoVelcro Chips utilized the same process for sample injection and CTC capture as that for SiNS NanoVelcro Chips. Of note, Hoechst stain and 100% ethanol were substituted in place of DAPI and 4% PFA to avoid DNA damage. CTCs immobilized on the p-NanoVelcro Chips were first identified and registered using the first fluorescent microscope (Nikon 90i) in conjunction with an auto-scan imaging software

(Nikon, Element). For more information, please see the section on NanoVelcro calibration and capture efficiency assessment. CTCs were stained with our 4-color ICC cocktail and were identified as those that were round/ovoid, Hoechst (Nucleus)+ / CK and/or CEA+ / CD45- / size ≥ 6 μm. In order to achieve the isolation of high purity CTCs, a second microscope, the ArcturusXT™ LCM System (Applied Biosystems™) was utilized to eliminate WBC contamination by selective laser dissection of identified CTCs. Briefly, CTCs identified with the imaging system were further confirmed by the ArcturusXT™ LCM system at three different fluorescent wavelengths (488nm, 546nm and 674nm). Afterwards, a CapSure™ HS Cap was placed on top of the region of the identified CTC. An 810 nm IR laser beam was then applied to melt the polymer membrane on the cap. The resultant conical polymer pillar, the so-called sticky finger, dropped down and adhered onto the p-NanoVelcro substrate. Subsequently, a 355 nm UV laser beam was utilized to cut through the p-NanoVelcro substrate in a designed route around the CTC of interest to exclude surrounding WBCs. In total, 20 single CTCs from each cell line (AsPC-1 and CFPAC-1) were isolated via LCM and mutational analysis. CTCs on the HS Caps were covered with 4 μL PBS. Finally, the HS Caps with dissected CTCs were stored at -20°C until whole genome amplification (WGA) was performed. Details of p-NanoVelcro sample processing and LCM are depicted in the workflow diagram in **Figure 27**.

Single PDAC cells attached to PLGA membrane on the HS Caps were freed from the sticky surface using a sterile needle under a microscope. Caps were then place into 0.5 mL microcentrifuge tubes and were briefly centrifuged at 10,000 rpm to bring cut cells to the bottom of the tube. WGA was then carried out on the 4 μL single cell solutions using the REPLI-g single cell kit (Qiagen) according to the manufacturer's protocol. Briefly, suspended cells underwent DNA denaturation with buffer D2 (Qiagen). Denaturation was terminated with a stop solution

(Qiagen). A master mix containing H_2O, reaction buffer, and scDNA polymerase (Qiagen) was added to each single cell sample. WGA was allowed to occur for 8 hours at 30°C. Following WGA, whole genome products were purified using the QIAquick PCR Purification Kit (Qiagen). The quality and integrity of the WGA product was further examined by Multiplex PCR on 8 housekeeping genes as described previously (Xu et al., *Cell 2012*, *148*, 886). Purified WGA products were then subjected to amplification of a 300 bp segment around *KRAS* codon 12 using Platinum PCR SuperMix High Fidelity Kits (Invitrogen). The primers (Integrated DNA Technologies) used were as follows: forward 5' – AAG GTA CTG GTG GAG TAT TTG–3' and reverse 5' – GTA CTC ATG AAA ATG GTC AGA G – 3'. Each PCR reaction underwent 35 cycles with a denaturing step of 94°C for 30 seconds followed by an annealing step at 55°C for 30 seconds and an elongation step at 68°C for 30 seconds. PCR products were then purified again with the QIAquick PCR Purification Kit. Presence of *KRAS* PCR products were then confirmed by gel electrophoresis on 2% agarose gels (**Figure 28 & Figure 29**). Purified PCR products were then sent to UCLA Genotyping and Sequencing Core for Sanger Sequencing (**Figure 28 & Figure 29**).

***KRAS* mutational analysis of single PDAC CTCs**

Venous blood was collected from 5 PDAC patients in which CTC counts were high on enumeration studies and primary tumor tissue was available. These samples were used for single CTC *KRAS* mutational analysis. Samples were collected and processed as described above for RBC lysis, anti-body incubation, assembly of NanoVelcro Chips, and ICC for identification (**Figure 27**). However, 100% ethanol replaced 4% PFA for fixation and Hoechst-stain replaced DAPI for nuclear staining. Triton-X was excluded from antibody incubations to prevent any

DNA damage. NanoVelcro/LCM was then used to allow for isolation of single CTCs and WBCs into individual Eppendorf tubes using an Arcturus^{XT}-LMD device attached to a Nikon Eclipse Ti fluorescent microscope (**Figure 28**).

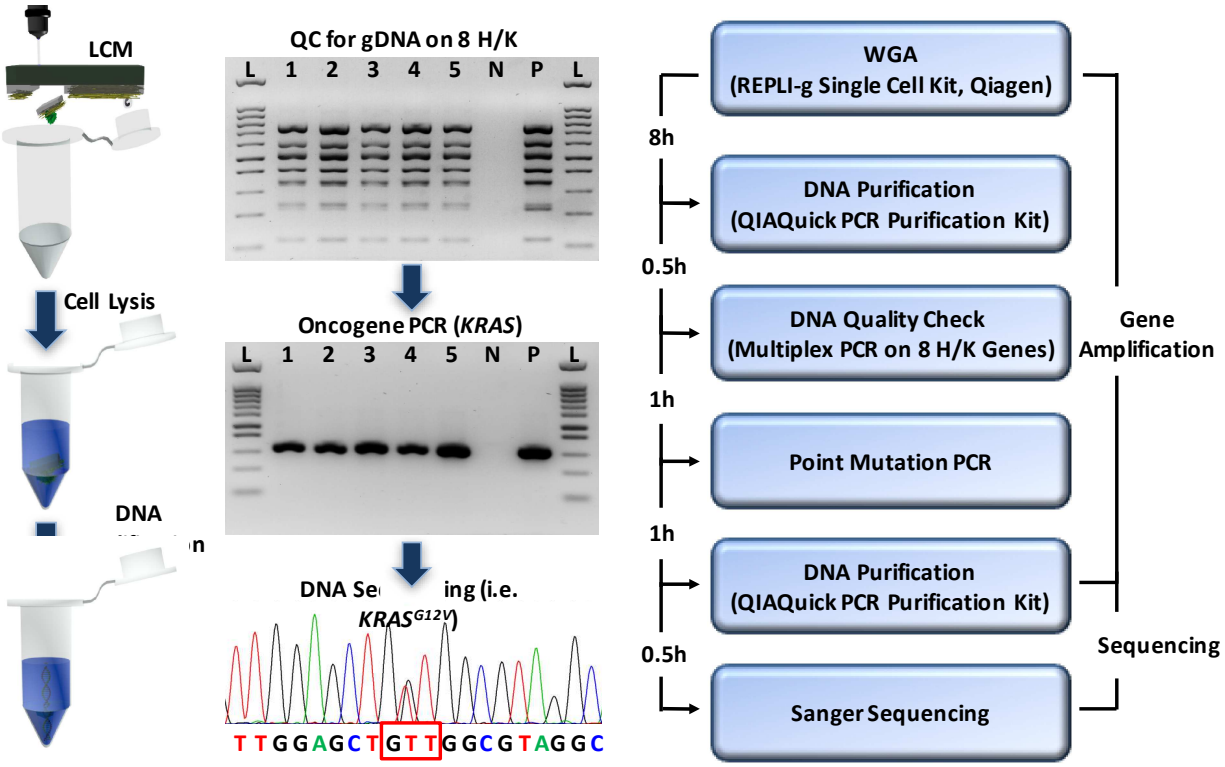


Figure 28. General workflow diagram for the process of single-CTC whole genome amplification, DNA quality assessment, and targeted mutational analysis of *KRAS* via PCR and Sanger Sequencing. H/K Genes – Housekeeping Genes

Cell samples then underwent whole genome amplification (WGA) using the REPLIg Single Cell Kit (Qiagen) per the manufacturers protocol. WGA samples were then subjected to PCR amplification of a 300-base pair fragment surrounding *KRAS* codon-12 using Platinum PCR SuperMix High Fidelity kits (Invitrogen). The primers (Integrated DNA Technologies) used were as follows: forward 5'–AAGGTACTGGTGGAGTATTTG–3' and reverse 5'–GTACTCATGAAAATGGTCAGAG–3'. Each PCR reaction underwent 35 cycles with a denaturing step of 94ºC for 30 seconds followed by an annealing step at 55ºC for 30 seconds and

an elongation step at 68ºC for 30 seconds. The presence of *KRAS* PCR product was confirmed by gel electrophoresis on 2%-agarose gels (**Figure 29**). Samples were then sent for Sanger Sequencing by the UCLA Genotyping and Sequencing Core on an Applied Biosystems 3730 DNA Analyzer. The forward primer listed above for *KRAS* amplification was used as the sequencing primer in all reactions. Sequencing results were then analyzed using FinchTV (Geospiza).

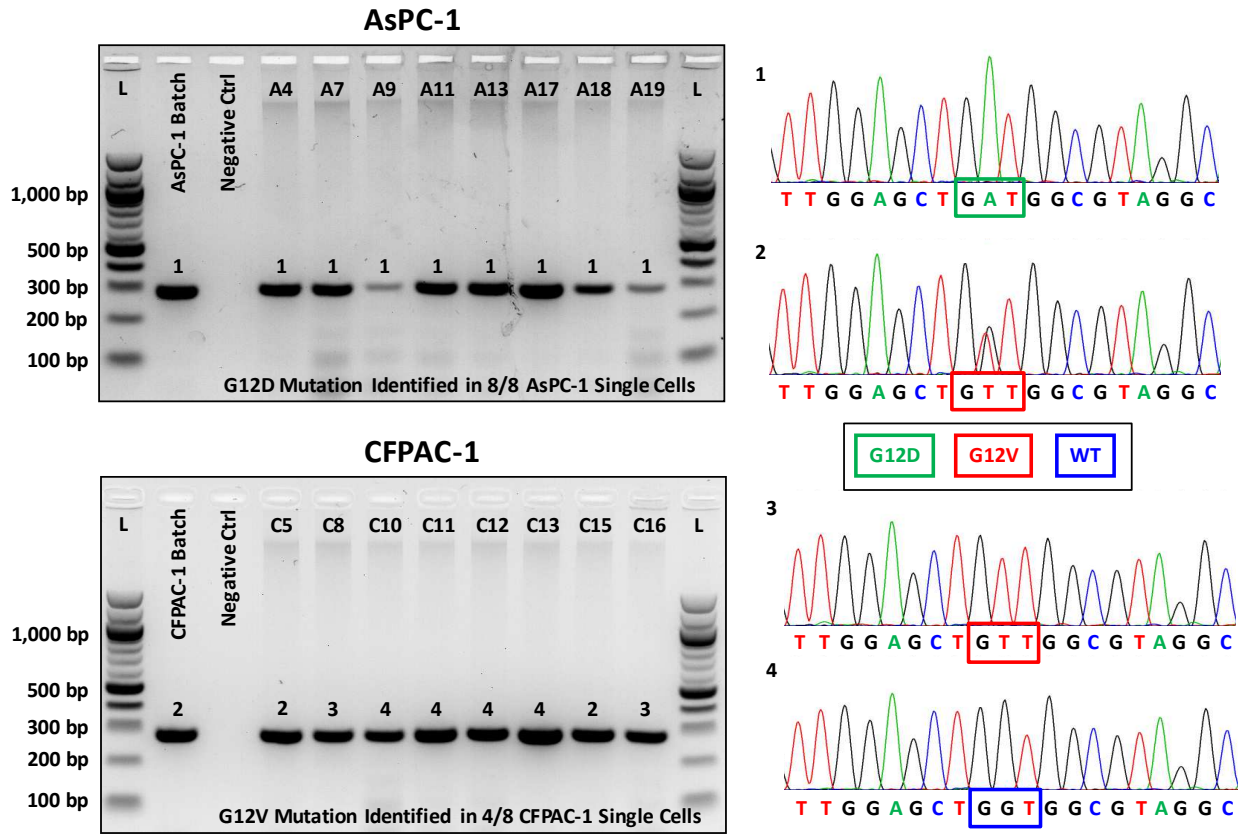


Figure 29. Targeted mutational analysis of *KRAS* codon 12 via PCR and Sanger Sequencing in PDAC cell lines. *KRAS* PCR of batch cellular DNA and WGA of single cells are shown in the agarose gels at the left for AsPC-1 (G12D Homozygote) and CFPAC-1 (G12V Heterozygote). The expected PCR product shows a band at ~300 bp. The accompanying chromatograms shown at the right are representations of the different Sanger Sequencing results obtained. *KRAS* mutation subtype is indicated by the colored boxes. The numeric labels over the bands in the agarose gels indicate the corresponding chromatogram/sequencing result shown on the right.

Isolation of primary matched tumor tissue for PDAC patients

Formalin fixed, paraffin embedded (FFPE) tumor blocks from the 5 patients with CTCs confirmed by Sanger sequencing were obtained. Ten serial sections of 5-μM thickness were cut and placed on PEN-membrane slides (Leica) followed by H&E staining by the UCLA Translational Pathology Core Laboratory. Areas of tumor were identified and marked by a pathologist. These areas were then laser micro-dissected using the Palm Micro-Beam laser micro-dissection system (Zeiss) to enrich the tumor content. DNA was then extracted and amplified using the REPLIg FFPE kit (Qiagen) according to the manufacturer's protocol. *KRAS* PCR and Sanger sequencing was then performed as previously described.

Isolation of primary matched tumor, peritumoral normal liver tissue, and whole blood DNA for HCC patients

Representative pieces of both primary tumor and peritumoral normal liver tissue were dissected from the resected specimen of all HCC patients undergoing surgery and flash frozen prior to storage at -80ºC. During processing of venous blood for CTC enumeration, the equivalent of 2 mL of whole blood was isolated and stored at -80ºC for germline DNA extraction. Genomic DNA from primary tumor tissue, peritumor normal liver tissue, and buffy coat samples was extracted using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturers recommended protocol.

Short tandem repeat (STR) analysis

To eliminate contamination as a confounding factor, we performed STR analysis of all CTC samples and compared it to that of the primary tumor and whole blood with the GenePrint

10 v1.1 system (Promega, Madison, WI) using the manufacturer's recommended protocol. Briefly, a 10ng aliquot of template DNA was added to the amplification master mix and amplified for 30 cycles on a GeneAmp PCR System 9700 thermal cycler (Thermo). Fragments were analyzed on a AB 3500 Genetic Analyzer with POP-4 Polymer (Applied Biosystems, Foster City, CA) and visualized using GeneMapper 5 software (Applied Biosystems).

Sequencing library preparation and sequencing

Purified WGA products were sheared to generate DNA fragments averaging 350bps by sonication (Covaris, Woburn, MA). Sonicated DNA was cleaned, end-repaired, ligated, and amplified using the KAPA DNA Library Preparation Kit (KAPA Biosystems, Wilmington, MA) according to the manufacturer's protocol. Sequencing was performed on an Illumina NextSeq 500 (Illumina, San Diego, CA) using 75 bp paired end reads (2 x 75 bp).

Array comparative genomic hybridization

Sample DNA (CTC, whole blood, peritumoral normal liver, and tumor tissue) and reference DNA (Agilent, Santa Clara, CA) were differentially labeled with cyanine-3 (CY3) and cyanine-5 (Cy5) dyes using the GenetiSure Amplification and Labeling Kit (Agilent) according to the manufacturer's protocol. Purified labeled DNA samples were prepared for hybridization, which took place on Agilent 8x60K CGH microarray slides at 67 ^{o}C for 6 hours. Following the hybridization, the slides were scanned using the Agilent SureScan Microarray Scanner (Agilent). Microarray images were analyzed using the Agilent CytoGenomics software (Agilent) and the Microarray text files were analyzed using R version 3.3.2 and the packages rCGH, limma, agilp, and snapCGH.

Copy number variation analysis from WGS data

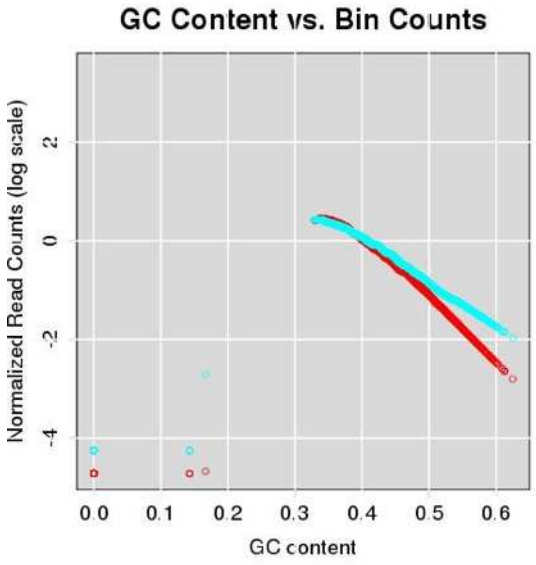


Figure 30. Example of GC content bias vs read count for CTC and primary tumor samples from patient H195 demonstrating the increased GC bias seen in CTC samples due to WGA. This bias is easily corrected using loess smoothing as previously described.

Fastq sequences were filtered and trimmed using Trimmomatic [108] and alignment was performed using bowtie2 [109] with settings "—very-sensitive-local." Coverage statistics and BAM to BED file conversion was performed using Bedtools.[110] Copy number variation and single cell NGS quality scores were analyzed by Ginkgo using a variable bin size of 250 kbps and simulating bins of 76 bp reads mapped with bowtie.[111] Segmentation was performed using normalized read counts, and sex chromosomes were masked for mixed-gender analyses. CTC samples demonstrated amplification bias due to GC content and mappability that was corrected using loess smoothing as previously described (**Figure 30**).[10,111] While the majority of the CTC samples demonstrated a relatively uniform amplification reaction as evidenced by the narrow distribution of normalized read counts around a copy number of 2, several samples demonstrated significant bias and noise, making copy number determination difficult (**Figure 31**). A matrix of segmented copy number values at each bin for every sample was thus created and used for further analysis.

Table 7. Characteristics of PDAC cell lines used for NanoVelcro Chip calibration, efficiency assessment, and feasibility of single cell mutational analysis.

PDAC Cell Line	Source	Differentiation	EpCAM Expression	*KRAS* Mutation	Zygosity
CFPAC-1	Metastasis	Well	High	G12V	Heterozygous
BxPC-3	Primary	Moderate	Moderate	WT*	WT*
PANC-1	Primary	Poor	Low	G12D	Heterozygous
AsPC-1	Metastasis	Poor	Moderate	G12D	Homozygous

*WT = Wild-type

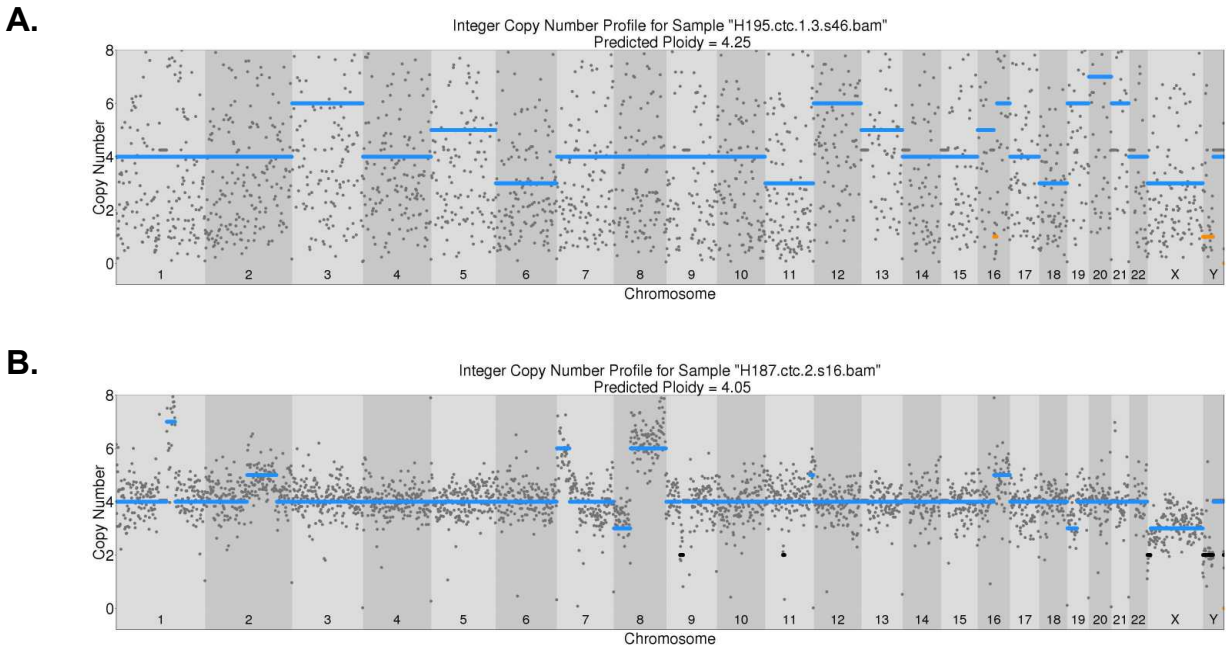


Figure 31. Examples of noisy CTC SCNA profiles resulting from either amplification bias or contamination. The CTC sample in (A) did not demonstrate a useful SCNA profile despite multiple attempts. In comparison, the CTC sample in (B) could be used after further normalization.

3.3 Results

Calibration and validation of NanoVelcro system and single cell KRAS analysis

The NanoVelcro and NanoVelcro/LCM platforms were optimized and calibrated to detect PDAC CTCs utilizing 4 PDAC cell lines (**Table 7**). Calibration and capture efficiency

data for each of the 4 PDAC cell lines are depicted previously (**Figure 4**). ICC definitions and *KRAS* mutational analysis were validated on artificial PDAC CTC samples (healthy donors' blood spiked with PDAC cell lines). AsPC-1, a homozygous PDAC cell line with a *KRAS* G12D mutation, and CFPAC-1, a heterozygous PDAC cell line with *KRAS* G12V mutation were utilized for *KRAS* mutational analysis. Batch cell DNA from each cell line was directly amplified for *KRAS* codon-12 and sequenced to verify published mutations and zygosity (**Figure 29**). Twenty single cells from each cell line were then isolated via LCM. We were able to identify the G12D mutation in single AsPC-1 cells, and the G12V mutation in single CFPAC-1 cells (**Figure 29**).

Development of a PDAC-specific ICC definition

We first set out to optimize our ICC techniques to ensure the specificity of our PDAC CTC definition. During our investigation of CTC enumeration using the NanoVelcro assay as a diagnostic biomarker in PDAC we noted a subset of cells that stained positive for both CK and CD45 (**Figure 32A**). These "double positive" cells, previously identified by other CTC studies as well, represented a diagnostic dilemma and potential source of false positive results.[106,29] In order to increase the specificity of our assay, we added a 4th fluorescent channel for CEA, a glycoprotein tumor marker over-expressed in 80% of PDACs (**Figure 32A, Row 4**).[38] Upon further review, we also noted that the double positive cells had horseshoe or multinucleated nuclear morphology more consistent with WBCs (**Figure 32B**). To address this we worked with a cytopathologist to develop nuclear morphology criteria to incorporate in our CTC definition.[112] Our final CTC definition is presented in **Figure 32C**. We then applied our CTC definition to 12 PDAC patients. We identified a total of 119 CTCs and at least 1 CTC / 4mL of blood in all 12 patients (range: 1-34; average: 9.9; median: 8). Additionally, we identified 37 double positive

cells in the 12 patients (range: 0-13; average: 3; median: 0). Of these 37 cells, 5 had low CD45 staining, and were primarily discounted as CTCs due to nuclear morphology.

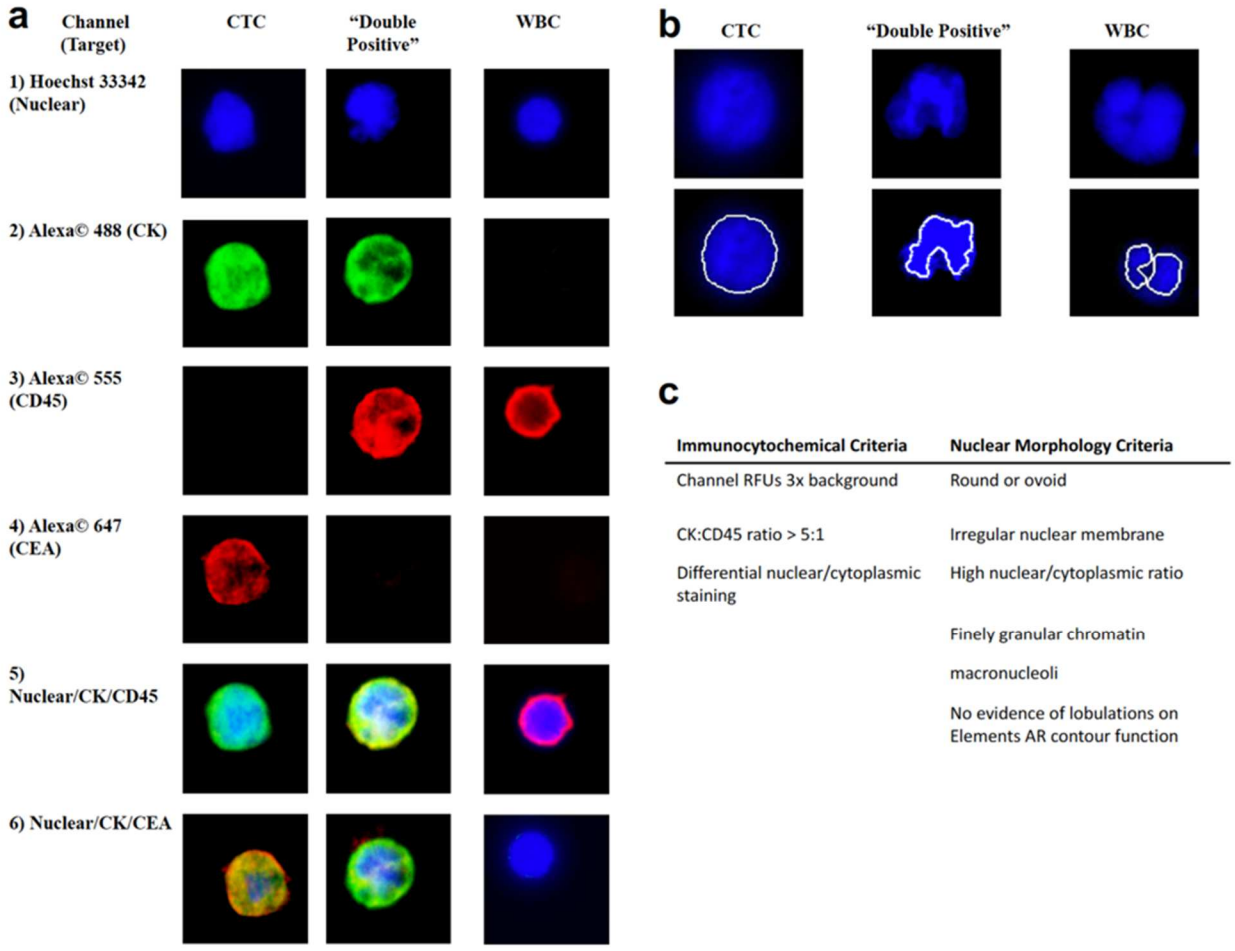


Figure 32. Development of a stringent definition of a pancreatic cancer circulating tumor cell. We modified the standard immunocytochemical definition of a CTC to include the tumor specific marker CEA as well as a stringent nuclear morphologic definition. **A.** Separate images of channels used in our immunocytochemical definition for a CTC, Nuc+/CK+/CD45+, and WBC. Rows 1-4 show typical fluorescence seen in each channel for the each of the three cell types. Rows 5-6 are overlays of the above channels demonstrating the images that are typically used to distinguish CTCs from other cells that are found. **B.** Examples of differing nuclear morphology between CTCs and other cells found in the blood. CTCs have round or ovoid nuclei without significant lobulation or multinucleation. **C.** ICC and nuclear morphology definitions used in distinguishing CTCs from other mononuclear cells.

KRAS mutational analysis of PDAC-CTCs and tumor tissue

A selection of 5 PDAC patients were utilized for *KRAS* mutational analysis to molecularly validate our stringent ICC criteria (round/ovoid, Nucleus+/CK+ and/or

CEA+/CD45-/size ≥ 6 μm). For the 5 patients, *KRAS* mutational analysis was performed on both single CTCs and WBCs. We were able to identify codon 12 activating *KRAS* point mutations in captured CTCs from all 5 patients, G12V (35G to T) in 1 patient and G12D (35G to A) in 4 patients. In contrast, *KRAS* mutations were not identified in any of the captured WBCs from the same 5 patients. Five patients with CTCs isolated using the NanoVelcro/LCM platform had primary tissue available for comparison. Of the patients, 2 had primary tumor tissue available and 3 had metastases available. *KRAS* mutational analysis was performed on both the single CTCs and matching primary tissue in order to provide evidence of pancreatic tumor origin of isolated candidate cells that met CTC criteria based on our ICC definition. We were able to identify codon 12 activating *KRAS* point mutations in captured CTCs from all 5 patients, G12V (35G to T) in 1 patient and G12D (35G to A) in 4 patients. We then performed *KRAS* mutational analysis on DNA isolated from matched FFPE primary tumor tissue to compare *KRAS* mutation subtype in CTCs to those found in the matched primary tumor (**Figure 33A**). For all 5 patients, the *KRAS* mutation subtype of the primary tissue matched that of the CTCs (**Figure 33B**). This confirmed the pancreatic tumor origin of captured CTCs and gave us confidence that the application of our PDAC ICC CTC definition would provide high specificity as a diagnostic biomarker.

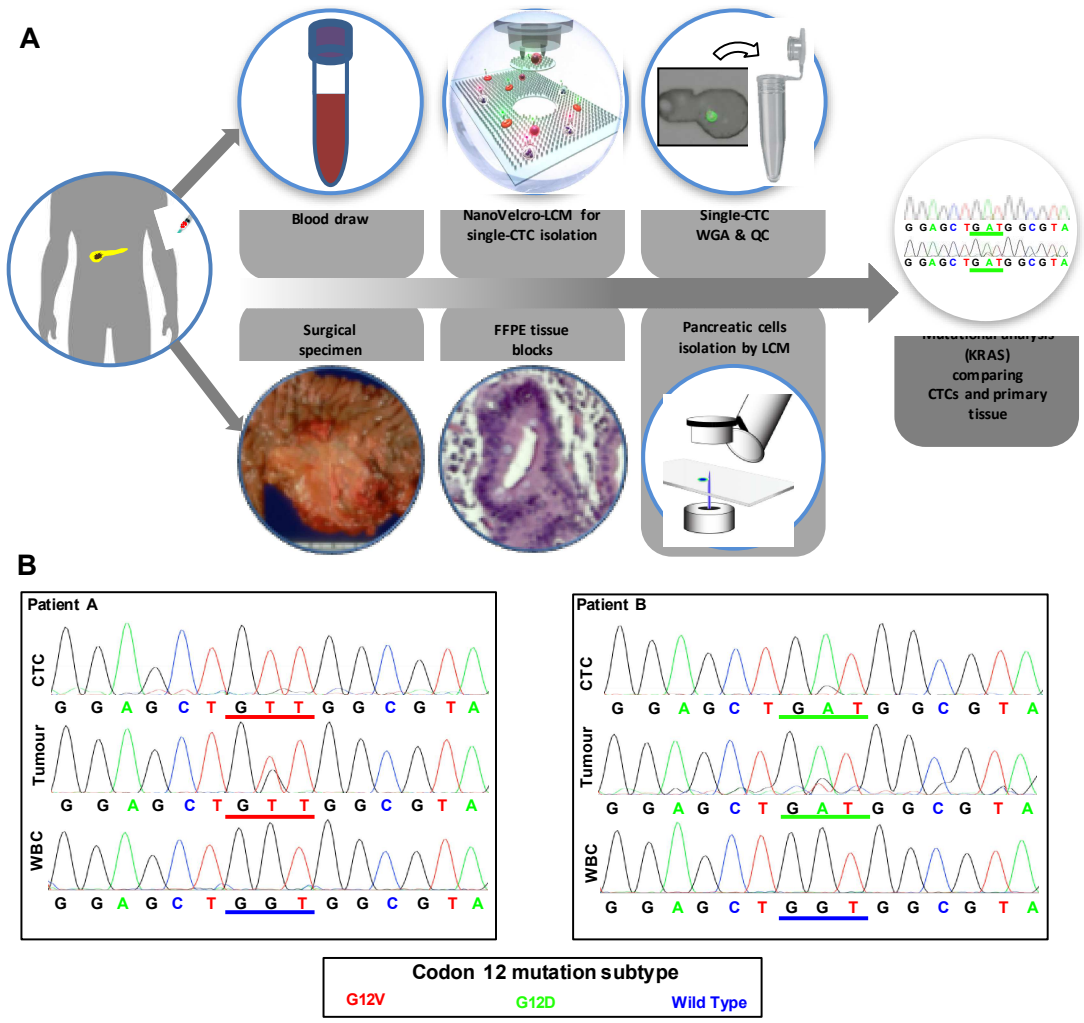


Figure 33. Validation of ICC and evaluation of tumor origin for CTCs by *KRAS* mutational analysis. **(A)** Flow diagram depicting the confirmation of tumor origin by isolating and sequencing CTCs and patient matched primary tumor tissue. **(B)** Sanger sequencing results for *KRAS* codon 12 mutations. CTCs, WBCs, and primary tumor tissue for 2 of the patients are depicted. Patient A has a G12V mutation and patient B has a G12D mutation. Both patients' WBCs were found to have wild type *KRAS* sequences.

CTC sequencing confirms tumor origin of HCC-CTCs

We identify HCC-CTCs using 4-channel ICC as previously described.[66] As ICC definitions of a CTC are subjective and vary widely between studies,[30,113] we sought to confirm the tumor origin of the HCC-CTCs to ensure that our CTC sequencing technique provides an accurate representation of the tumor. Each cell contains approximately 4 picograms of DNA, and the success of the WGA reaction largely depends on the quality of the sample and the potential

contamination during sample processing.[114] Furthermore, WGA, and especially MDA-based WGA, demonstrates stochastic amplification bias that cannot be accounted for systematically.[115] Therefore, we first sought to confirm that we could reliably sequencing the isolated HCC-CTCs without prohibitive contamination or amplification bias, and that the HCC-CTCs were in fact of tumor origin and reflective of the tumor. We began by performing STR typing of the blood, peritumoral normal liver, primary liver, and CTC to confirm that contamination had not occurred. While STR typing confirmed that CTCs had originated from the same individual for 9/10 patients, it revealed that the CTCs for patient H169 were likely from a different source (**Figure 34**). We next compared the binned copy number profiles of all CTC and primary tumor samples using a spearman correlation coefficient matrix (**Figure 35**). Unsupervised hierarchical cluster analysis demonstrated that CTCs consistently clustered with their respective primary tumor samples and not with other CTCs. This demonstrates that CTC sequencing not only accurately recapitulates the SCNA profile of the primary tumor, but also that the CTC SCNAs observed are due to the tumor origin of the DNA and not due to amplification bias or contamination introduced during WGA.

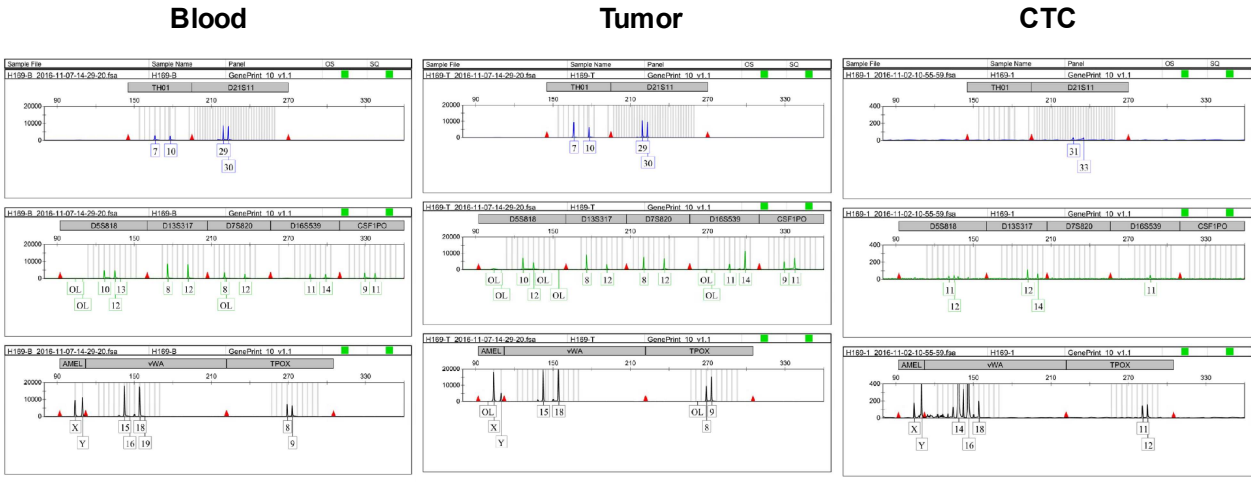


Figure 34. STR analysis for patient H169 demonstrating that the CTC sample genotype does not match that of the primary tumor or blood.

Using a 1 million bp fixed bin size, we visualized the SCNA profiles for the blood, peritumoral normal liver, primary tumor, and CTC samples. Inspection of the copy number profiles clearly demonstrates the recapitulation of the somatic changes in the primary tumor by the HCC-CTCs. Furthermore, it validates that our ICC definition of a HCC-CTC is accurate, and that the cells are not false positive CK^+ endothelial or mesenchymal cells as has been described previously.[116] Examination of the spearman correlation matrix of all CTCs and primary tumor samples (**Figure 35**) reveals that CTCs shared an average of 81.3% (range: 54.9 – 98.5%) of the gain and loss SCNA profiles found in the primary tumor.

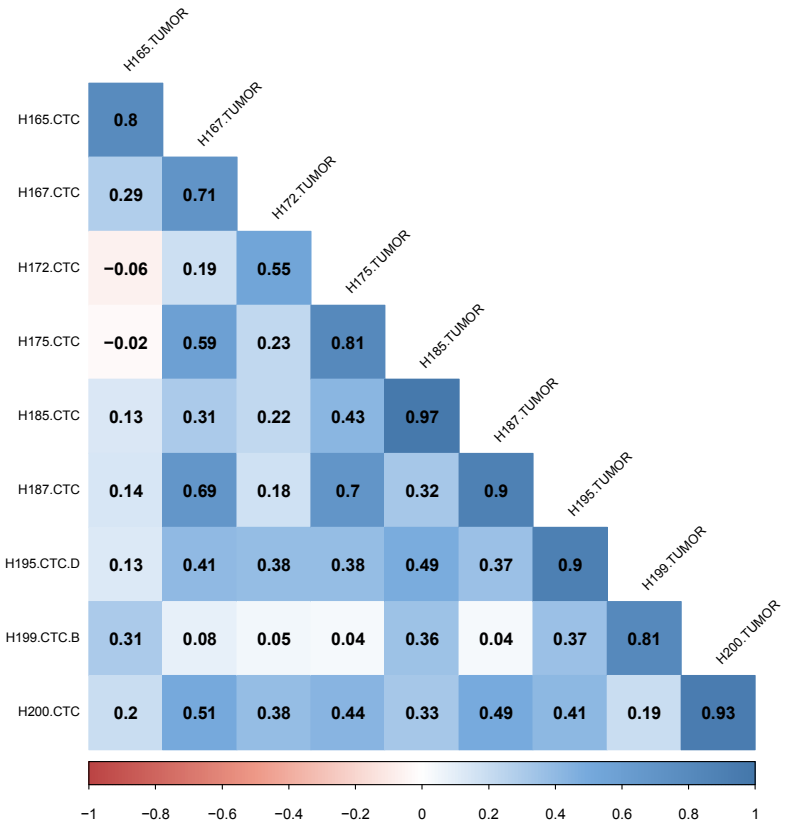


Figure 35. Spearman correlation plot for the low-resolution whole genome copy number profiling reveals the close correlation of CTCs with their respective primary tumors for 9/10 patients. Samples are ordered according to hierarchical clustering order with the 11 rectangles highlighting the final branches in the clustering results. Patient H169 is the only patient without clustering of CTC and primary tumor. This patient's CTC sample had previously failed STR analysis as well.

3.4 Conclusion

CTCs show enormous promise as a biomarker in all solid tumors and are of special interest for GI tumors which are notoriously difficult to biopsy. Due to the lack of standardization between CTC assays and definitions, there is a concern about the reproducibility, validity, and generalizability of prior CTC studies.[117] Of particular concern is the possibility of false positive CTC calls due to lenient ICC definitions, possibly due to misidentification of macrophages or prolonged sample processing time.[30] The accuracy of CTC definitions is only becoming more important with the introduction of CTC molecular analysis and liquid biopsy-type applications, in which false positives could affect treatment-decisions and even patient outcomes.

In the current project, we sought to confirm the accuracy of our CTC definitions for both HCC-CTCs and PDAC-CTCs using molecular analysis. For PDAC-CTCs, we were able to demonstrate the accuracy of our stringent CTC definition by finding *KRAS* mutations only in cells that we called CTCs and never in any of the cells identified as WBCs or "double-positive" cells from 12 PDAC patients. Furthermore, we used Sanger sequencing to demonstrate *KRAS* subtype congruence between CTCs and the primary tumor for 5 patients. For HCC-CTCs, we demonstrated that somatic copy number changes found in the primary tumor were seen in the CTCs but not in the peritumoral normal liver or whole blood sequencing samples. Together, these results give us confidence that the CTCs isolated and identified using our CTC assays are from the primary tumor, and that using our CTC sequencing pipeline allows for the identification of genetic alterations found in the primary tumor. Based on this groundwork, we felt confident moving forward that CTC sequencing would be an effective biomarker in GI cancers.

Chapter 4 Cellular limits and optimization of CTC molecular analysis

4.1 Overview

Chapter 4 focuses on our attempts to improve our whole genome amplification (WGA) protocol to allow for reliable sequencing from as few as 3-5 CTCs (the number of CTCs we were typically finding from 2 mL of blood). It can be divided into 3 sequential parts:

Section 4.2 – Determining the number of cells needed for reliable CTC sequencing.

Section 4.3 – Developing a quality control metric that would allow us to compare different WGA techniques.

Section 4.4 – Use the QC metric to optimize a CTC sequencing protocol to allow us to obtain reliable sequencing data from only 3-5 CTCs (the number of CTCs we are usually able to find in a given patients sample

In **Section 4.2** we looked at PDAC cell line sequencing to determine the minimum number of cells needed for reliable CTC sequencing. Using cell lines with *KRAS* mutations demonstrating imbalanced alleles (e.g. 4:1 mutant to wild-type gene fractions) we looked at how many cells were required to accurately reproduce the mutant allele fraction of the batch cell line DNA. We also used an 8-oncogene multiplex PCR method to determine how many cells were required to reliably produce all 8 bands. As will be demonstrated, samples with less than 5-10 cells did not reproduce the KRAS sequencing nor the oncogene PCR results of the batch DNA whereas samples with more than 10 cells did so reliably.

To better understand how to improve our WGA methods we developed a simple qPCR quality control metric that would let us know if a given sample was likely to produce reliable sequencing results. This work is discussed in **Section 4.3**. The 8 gene multiplex qPCR looked at

8 important oncogenes (*BRAF, EGFR, KRAS, NRAS, PIK3CA, PTEN, TP53*). We investigated if the results of this panel could be used to predict the quality of the sequencing obtained from these samples. As can be seen in the figure below, the qPCR score (1-8 depending on the number of oncogenes that successfully amplified) correlates with the quality of both Sanger sequencing and array CGH. Similar results were found for NGS-based SCNA sequencing.

Using the qPCR QC metric, we developed in **Section 4.3**, we tested a variety of modifications to the standard WGA method to create a modified WGA protocol that allowed for reliable sequencing from 3-5 CTCs in **Section 4.4**. The final method involved 4 major modifications: Cell resuspension in PBS to prevent cells sticking to the cap, addition of a mechanical cell lysis step to the enzymatic cell lysis to increase DNA accessibility, performing MDA in multiple small volumes to decrease amplification bias, and a post-amplification debranching step to address the hyperbranching that occurs with multiple displacement amplification. Overall, we found modest gains from our improvements in the WGA protocol as shown below for whole exome sequencing. Similar results were found when performing array CGH, NGS SCNA profiling, or Sanger sequencing. However, the MDA debranching step is a post-amplification process and showed potential as a "rescue" method for poorly performing MDA-amplified samples.

4.2 Cellular limits and reality of CTC sequencing

4.2.1 Abstract

To understand the potential and limitations of CTC sequencing in the context of molecular diagnostics, we investigated the feasibility of identifying the ubiquitous KRAS mutation in single CTCs from pancreatic cancer (PDAC) patients. We utilized the NanoVelcro/LCM CTC platform combined with whole genome amplification (WGA) and KRAS Sanger sequencing. We assessed both our ability to obtain KRAS codon 12 coverage and the degree to which allele dropout during WGA affected the detection of KRAS mutations from single CTCs. We isolated 385 single cells, 163 from PDAC cell lines and 222 from the blood of 12 PDAC patients, and obtained KRAS sequence coverage in 218/385 (56.6%) of single cells. For the PDAC cell lines with known KRAS mutations we were able to detect a mutation in 67% of homozygous cells but only 37.4% of heterozygous single cells, demonstrating that both coverage and ADO are significant causes of mutation detection failure from single cells. We were able to detect KRAS mutations in CTCs from 11/12 (92%) patients and 33/119 single CTCs sequenced, resulting in a KRAS mutation detection rate of 27.7%. Importantly, KRAS mutations were never found in the 103 white blood cells sequenced. We then sequenced groups of cells containing between 1-100 cells and determined that at least 10 CTCs are likely required to reliably assess KRAS mutation status from CTCs.

4.2.2 Introduction

The development of molecularly targeted agents in oncology has rapidly advanced in recent years, with over 800 drugs in clinical trials as of 2013.[118] Thus, therapeutic decisions are increasingly being made based on actionable genetic data obtained from a patient's tumor, resulting in precision treatment strategies. While biopsy or surgical excision has been the traditional means of obtaining tissue from solid tumors, these procedures are costly, painful, and not without risk to the patient. This is especially true for difficult to biopsy tumors such as pancreatic cancer (PDAC), in which multiple biopsies or even multiple procedures are required to obtain sufficient tumor tissue for genetic testing.[44] Due to these limitations, alternatives to traditional biopsy techniques are an area of active research, and will be crucial to realizing the potential of precision medicine for many tumor types.

Circulating tumor cells (CTCs), which are thought to originate from primary or metastatic sites and circulate in the blood, represent one such alternative source of tumor tissue for the diagnosis, staging, management, and prognostication of solid cancers.[30,119] The enumeration of CTCs utilizing the CellSearch™ platform currently has FDA approval as an adjunctive prognostic marker during therapy in many solid cancers.[4] The development of CTC assays that allow for the capture and isolation of CTCs for further analysis has the potential to transform CTCs into a "liquid biopsy" for cancers. Such a liquid biopsy has innumerable applications, from diagnosing patients with difficult to biopsy tumors to the real-time monitoring of a cancer's genotypic changes in response to therapy.[33] However, for CTCs to realize their potential as a convenient source of tumor tissue, important methodological issues must be addressed.

The primary issue that CTC assays must overcome is the rarity of CTCs in circulating blood, as an average 7.5 mL blood draw typically contains only 1-50 CTCs versus several million white blood cells (WBCs).[120] Differentiating CTCs from the vast hematopoietic cell background requires highly sensitive and specific assays, as well as a means of confirming that the cells captured are actually of tumor origin. One means of doing so is by using molecular analysis in addition to immunocytochemistry (ICC) to confirm the cells are of tumor origin. However, for molecular analysis, the small number of CTCs available is further hindered by the limited amount of DNA per cell. Thus, whole genome amplification (WGA) methods are required to bridge the gap between the picogram quantities of DNA available in a single cell and the nanogram to microgram amounts required for most molecular studies.

Studies of WGA from single cells have demonstrated the capability to achieve more than 10,000-fold amplification with up to 96% genome coverage.[121] Furthermore, WGA has allowed researchers to successfully perform whole exome and even whole genome sequencing of single cells.[122] However, WGA amplification is not linear, and amplification bias, especially in GC rich regions, is an inherent limitation. Unequal amplification can even result in the total loss of one allele over multiple rounds of amplification, termed allele dropout (ADO).[121] Thus, the limitations of WGA affect CTC sequencing in two major ways: by failure of the WGA reaction to provide coverage of the gene of interest, and by ADO of the mutant allele resulting in only the wild type sequence amplifying when in fact a mutant allele is present. It is important to differentiate these limitations because they affect the clinical implications of CTC sequencing studies in dramatically different ways. For example, for heterozygous mutations coverage alone is not in fact representative of the mutation detection rate (MDR) due to the effects of ADO. Thus, successful sequence coverage and mutation detection need to be considered separately in

characterizing CTC sequencing results. How these limitations affect CTC sequencing results have not been well characterized to date.[122]

In this study, we sought to better understand the potential and limitations of single cell sequencing of PDAC CTCs as it relates to their development as a clinical biomarker for diagnostic and pharmacogenomic applications. We did so by studying Kirsten rat sarcoma viral oncogene homolog (KRAS) sequencing of single cells from both PDAC cell lines and patient CTC samples. We chose to study pancreatic cancer because greater than 95% of patients have an early activating driver mutation in the KRAS oncogene, and 98% of those mutations are located in codon 12.[123,51] Thus, sequencing of a single codon can provide molecular evidence distinguishing CTCs from hematopoietic cells. For CTC isolation we utilized the NanoVelcro/LCM platform, which captures CTCs on a microfluidic chip using anti-EpCAM antibodies. We then identified captured CTCs by immunocytochemistry (ICC) and isolated them using laser capture microdissection (LCM). We developed an ICC-based CTC definition for PDAC that incorporated nuclear morphology and anti-carcinoembryonic antigen (CEA) staining, a glycoprotein tumor marker that is overexpressed in 80% of PDAC, to increase the specificity of the assay.[38] We then studied our sequencing methods on single cells from 5 PDAC cell lines to assess the limits of our platform both in terms of KRAS codon 12 coverage and KRAS mutation detection rate. We applied these findings to CTCs from 12 PDAC patients to determine our mutation detection rate for KRAS mutations from single CTCs. Finally, we utilized our results to better understand the number of CTCs required for reliable KRAS mutational analysis.

3.2.3 Methods

Pancreatic cancer cell lines

Pancreatic cancer cell lines CFPAC-1, AsPC-1, Panc-1, BxPC-3, and HPAF-II were obtained from American Type Culture Collection (ATCC), and grown using RPMI 1640 Medium (ATCC) supplemented with 10% fetal bovine serum (FBS) and 100 U/mL penicillin-streptomycin (P/S). All cell lines were grown at 37ºC with 5% CO_2 and were routinely passaged at 80% confluence using an iso-osmotic sodium citrate solution for cell release (Thermo). When preparing cells for single cell isolation and analysis, cells were released from the culture plates using the same sodium citrate solution. Following a wash with the culture medium, each cell line was diluted to a density of 1000 cells per 100 µL. Approximately 1000 cells (100 µL) were smeared on PEN membrane slides (Leica), air-dried for 10 minutes, and fixed with 100 µL of 100% Ethanol. Cells were then isolated by laser microdissection as outlined below. Batch genomic DNA was prepared from the entire contents of a 75 mm^2 plate at 80% confluence. DNA was extracted using the Qiagen Blood and Cell Culture DNA Mini Kit according to the manufacturer's protocol.

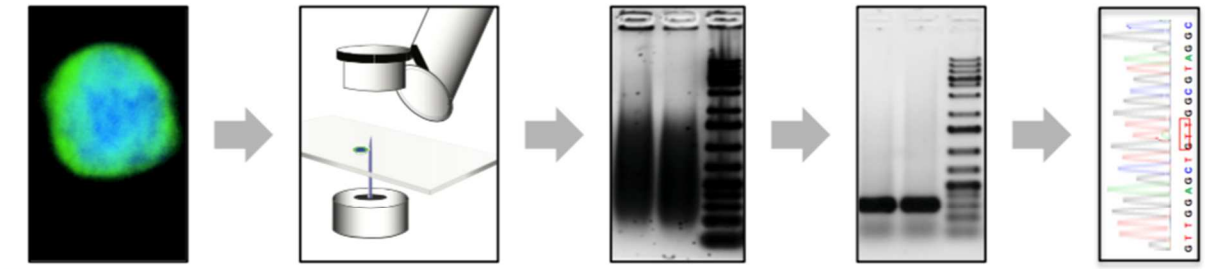
Patients samples

Twelve patients with pathology confirmed pancreaticobiliary cancers were voluntarily enrolled in the study under UCLA IRB #11-002112. All 12 patients underwent a venous blood draw that began with a 5 mL waste to prevent epithelial cell contamination followed by collection of at least a 10 mL sample into two 8.5 mL ACD Solution A tubes (BD). Following collection, the blood was inverted at least 10 times followed by transport to the lab on ice. Samples were stored for less than 24 hours at 4ºC prior to processing.

Sample processing

Our workflow for sample processing is outlined in **Figure 36A**. Density gradient centrifugation of patient samples was performed using 12 mL LeucoSep centrifuge tubes (vWR) according to the manufacturer's protocol. The buffy coat layer was washed with 5mL of wash medium (RPMI with 5% FBS, Gibco), centrifuged at 300 x *g* for 10 minutes at 4ºC, and the pellet resuspended in 200 µL of wash medium. 6 µL of biotinylated anti-EpCAM antibody (R&D Systems) was added and the mixture was placed on a shaker at 4ºC for 1 hour. Following an additional wash step, the pellet was re-suspended in 200 µL of PBS and immediately processed on the NanoVelcro/LCM CTC platform.

A



B

Single PC Cell Line Cells	LCM Transfer	WGA	KRAS PCR	Sanger Seq	KRAS codon 12 coverage
163 (100%)	161 (98.8%)	112 (68.7%)	106 (65.0%)	103 (63.2%)	103 / 163 (63.2%)

Figure 36. (A) Overview of the single cell sequencing workflow using NanoVelcro/LCM. Following CTC or PDAC cell line cell identification by ICC, individual cells are isolated by LCM. Transfer efficiency of LCM was determined based on identification of the cell isolated in the Eppendorf tube cap. Single cells then underwent WGA and success was assessed by demonstration of a smear band. Next, KRAS PCR was performed and was considered successful if a band of the proper size (295 bp) was found with gel electrophoresis. Finally, Sanger sequencing success was determined by visual inspection of the trace file. **(B)** Stepwise workflow analysis of cell line single cell sequencing. A total of 163 single cells were isolated and sequenced and the success of the individual steps in the workflow from (A) were analyzed, revealing that WGA was responsible for the majority of sequencing failures.
LCM – laser capture microdissection, MDA – multiple displacement amplification, PCR – polymerase chain reaction, Sanger – Sanger sequencing

NanoVelcro/LCM CTC chip processing

NanoVelcro/LCM CTC chips were assembled and operated as described previously.[37,124] Briefly, poly(lactic-co-glycolic acid) (PLGA) nano-spun chips, manufactured via an electro-spin method in our nano-materials lab, were assembled onto a laser microdissection slide (Leica) with an overlaid polydimethylsiloxane (PDMS) microfluidic component and attached to a syringe based microfluidic pump (KD Scientific). Prepared samples were injected through the device at the previously established optimal flow rate of 0.5 mL/hr. and were then fixed using 100% ethanol.

Immunocytochemistry (ICC) and nuclear staining

Following fixation, processed chips were washed in PBS. Chips were then blocked using PBS with 2% donkey serum (DS, Jackson Immunoresearch) for 1 hour. Chips were then incubated with a primary antibody cocktail containing two rabbit anti-pancytokeratin antibodies (Invitrogen, Abcam), two mouse anti-CD45 antibodies (BD Pharmigen, Abcam) and one chicken anti-CEA antibody (Abcam) overnight at 4ºC. Following primary antibody incubation, chips were washed again in PBS. Secondary antibody incubation was carried out for 1 hour at room temperature using a cocktail of AlexaFluor 488 donkey anti-rabbit (Invitrogen), AlexaFluor 555 donkey anti-mouse (Invitrogen), and AlexaFluor 647 goat anti-chicken (Invitrogen). Incubation took place in PBS with 2% DS. Chips were again washed in PBS with 2% DS and then stained with Hoechst 33342 (Life Technologies) for nuclear visualization.

Chip scanning, CTC identification, and laser micro-dissection

Following immunostaining, automated chip scanning was performed at 40x power with a Nikon Eclipse 90i scanning fluorescent microscope using NIS Elements 4.1 software. Manually acquired 400x power images of candidate cells were then obtained. For ICC, we added carcinoembryonic antigen (CEA) and nuclear morphology characteristics to the standard ICC CTC definition in an attempt to increase our specificity. Thus, WBCs were defined as Nuclear+/CD45+/CK-/CEA- and CTCs as Nuclear+/CD45-/and CK+ or CEA+ (**Figure 37**). Nuclear morphology was assessed by a trained cytopathologist, and included in our CTC definition. The positional coordinates of cells to be isolated were then input into the PALM MicroBeam laser microdissection system (Zeiss). After locating cells using the pre-programmed coordinates on the LCM device, single cells were laser dissected and collected into 200 µL tube caps (Zeiss) using the laser pressure catapult function. Cell transfer to the tube cap was confirmed by imaging the cap prior to cap closure.

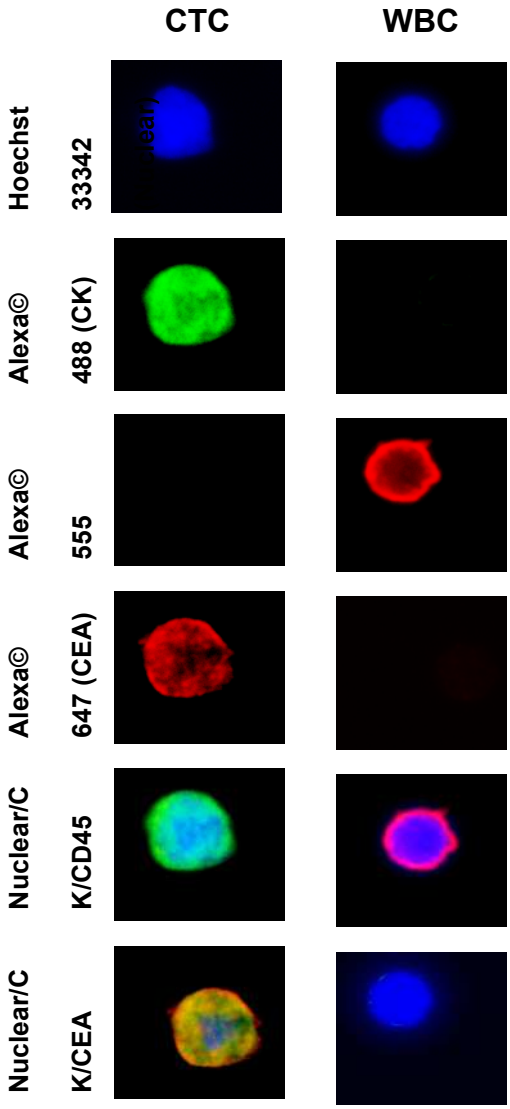


Figure 37. PDAC immunocytochemical definitions for CTCs and WBCs using the NanoVelcro/LCM assay. Representative images under 40x magnification of the different immunofluorescent channels used to discriminate CTCs from WBCs are shown. Rows 5-6 are overlays of the above channels demonstrating the images that are typically used to distinguish CTCs from other circulating cells.

Whole genome amplification, *KRAS* PCR

Isolated cells were then subjected to genomic DNA isolation and whole genome amplification (WGA) using the REPLI-g Single Cell Kit (Qiagen) according to the manufacturer's protocol. Genomic amplification occurred at 30°C for 8 hours. WGA products were cleaned using the QIAquick PCR Purification Kit (Qiagen). Cleaned WGA products underwent PCR amplification of KRAS exon 2 using the following primers (Integrated DNA Technologies): forward 5' – AAG GTA CTG GTG GAG TAT TTG – 3' and reverse 5' – GTA CTC ATG AAA ATG GTC AGA G – 3', resulting in a predicted amplicon length of 295 bp. Melting temperatures were 51.9°C and 51.0°C for the forward and reverse primers, respectively. PCR reactions were carried out on an Arktik Thermal Cycler (Thermo Scientific) using the Platinum PCR SuperMix High Fidelity Kit (Invitrogen) using total volumes of 50 µL per reaction and 200ng of template DNA according to the manufacturer's protocol. The reaction conditions were as follows: denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 68°C for 45 seconds for a total of 35 cycles. PCR success was confirmed by demonstration of the expected 295 bp band on 2% agarose gel electrophoresis.

Sanger sequencing

The PCR products were then purified using the QIAquick PCR Purification Kit (Qiagen) and eluted into 25 uL of nuclease-free water (Qiagen). DNA was diluted to a concentration of 15 ng/uL based on Nanodrop quantification of the PCR product. Automated dideoxy terminator sequencing was performed by capillary electrophoresis by the UCLA GenoSeq Core on an ABI 3730 DNA analyzer using Big Dye Terminator chemistry (Applied Biosystems). Bidirectional sequencing was used to confirm the specificity of the target in selected samples using the primers

listed above. The forward primer was then used for subsequent sequencing runs. Sequences were analyzed by two methods. All sequences were analyzed by manual inspection of the individual trace files using Four Peaks (Nucleobytes), as well as using ab1 Peak Reporter (Thermo Fisher Scientific). A minor allele peak height of >10% by the 7-scan window of ab1 Peak Reporter was used to define heterozygous samples.

Digital PCR

Digital PCR (dPCR) was performed on a QuantStudio™ 3D digital PCR System (Thermo) following the digital MIQE Guidelines.[125] The TaqMan KRAS_521 (Assay ID: AH6R5PI) assay was used for primers and probes. 18-µl reaction mixes were prepared, containing 9-µl twofold QuantStudio 3D Digital PCR Master Mix (Thermo), 0.9-µl 20-fold TaqMan KRAS_521 primer-probe mix, 1.8-µl diluted gDNA, and 6.3-µl nuclease-free water (Qiagen). The reaction mixture was loaded onto QuantStudio 3D Digital PCR 20 K Chips (Thermo) using an automatic chip loader according to the manufacturer's protocol. Loaded chips underwent thermocycling on the Gene Amp 9700 PCR System using the following conditions: initial denaturation at 96^{o}C for 10 minutes, followed by 39 cycles of 98^{o}C for 30 seconds and 60^{o}C for 2 minutes. VIC and FAM imaging of each chip was performed by the instrument, which assesses raw data and calculates the estimated concentration of the nucleic acid sequence targeted by the FAM and VIC dye labeled probes by Poisson distribution. The raw data was then uploaded to QuantStudio 3D AnalysisSuite Cloud Software for analysis (**Figure 38**).

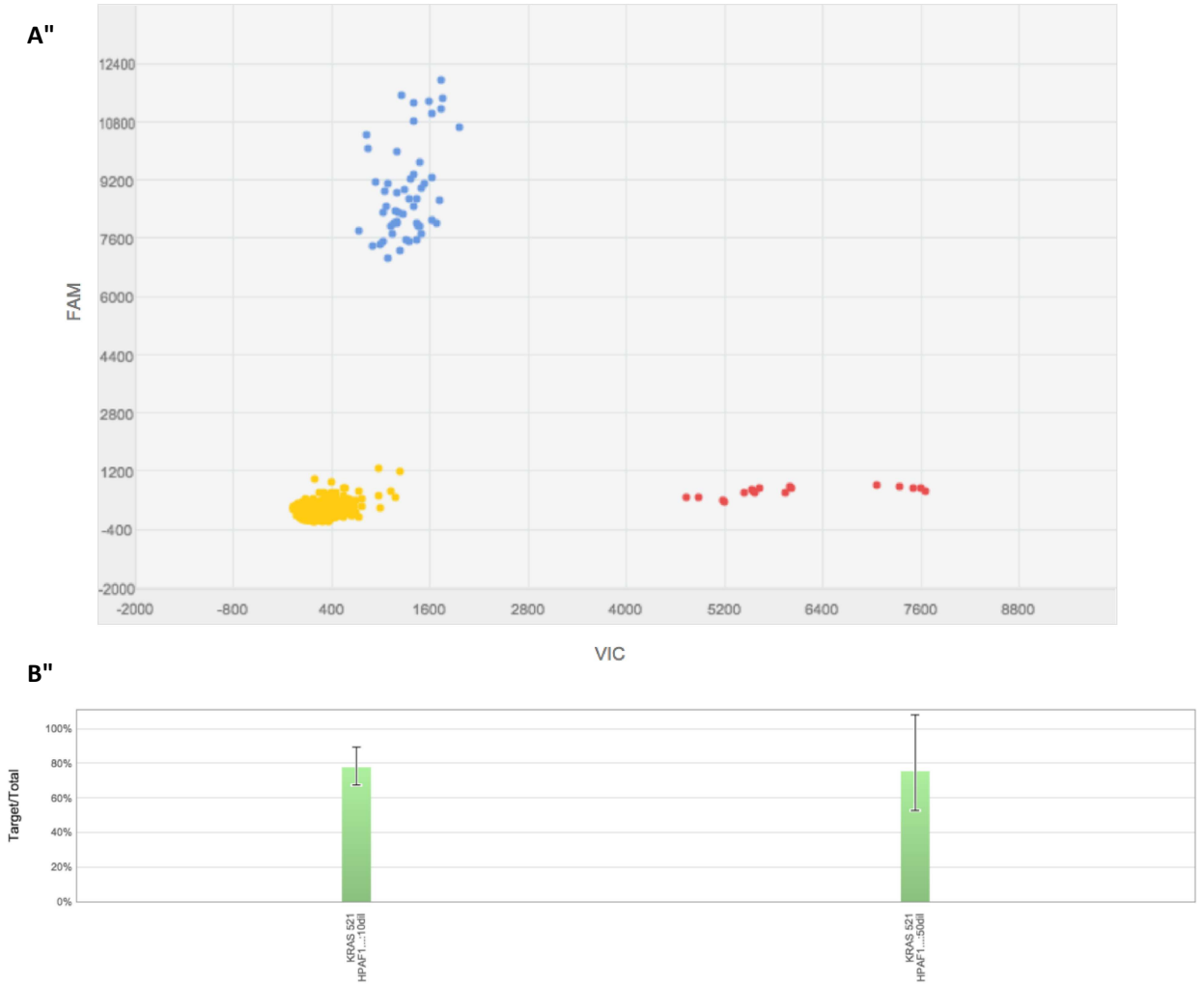


Figure 38. Digital PCR to confirm allelic ratio for KRAS G12D mutations for the HPAF-II cell line. **A.** The figure is the graphical output of the fluorescent intensities of the individual wells analyzed using the QuantStudio AnalysisSuite software. The blue FAM dye correlates with the mutant allele probe (G12D) and the red VIC dye correlates with the wild type probe. **B.** The bar graph is interpreted from the figure above, and is the ratio of mutant signal to the total signal. It demonstrates the expected 75:25 mutant (G12D):wild type ratio for the HPAF-II cell line from two different batches of cells.

Statistical Methods

Statistical analyses were performed with the Rstudio (version 0.99.489) environment of the open source software R available at www.rstudio.com. Descriptive statistics were used to summarize cell line and patient sample sequencing results. Pearson's Chi-squared test with Yates' continuity correction was used to compare KRAS codon 12 coverage and mutation

detection rates (MDR). The likelihood of detecting mutations was calculated by single event binomial distribution using the equation $P = n^{Cx} * p^x * q^{n-x}$, with $n^{Cx} = n! / (x! * (n – x)!)$, where p = the probability of a detecting a mutation, and n = the number of single CTCs sequenced.

4.2.4 Results

Single Cell KRAS Sequencing of PDAC Cell Lines

To test the feasibility of single cell sequencing using the NanoVelcro/LCM platform, we used five PDAC cell lines with differing KRAS mutations and zygosity. We captured and isolated a total of 163 single cells utilizing the NanoVelcro/ LCM platform, including 60 CFPAC-1 (heterozygous G12V) cells, 44 PANC-1 (heterozygous G12D) cells, 27 HPAF-II (heterozygous G12D) cells, 21 AsPC-1 (homozygous G12V) cells, and 13 BxPC-3 (wild type) cells. Each cell then underwent WGA, KRAS PCR amplification, and Sanger sequencing. To better understand the factors that contributed to the successful coverage of KRAS codon 12, we performed a stepwise workflow analysis of our methods to determine the success rate of each step (**Figure 36B**). Confirmation of LCM transfer of single cells was determined by visualization of a cell in the cap on inverted microscopy. The success of WGA and KRAS PCR was determined by visualization of a smear or band by gel electrophoresis, respectively. KRAS codon 12 coverage was determined by manual inspection of the individual trace files. Using this method, we determined which step failed for the 60 single cells that did not successfully show an interpretable Sanger trace file. We discovered that failure occurred at all steps in our methodology. However, the vast majority (81.7%) occurred during WGA (n=49) followed by KRAS PCR (n=6; 10%), Sanger sequencing (n=3; 5%), and LCM transfer (n=2; 3.3%). Overall, we obtained KRAS codon 12 coverage in 103/163 (63.2%) of the single cells sequenced.

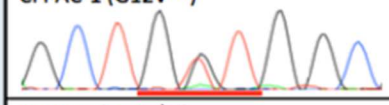
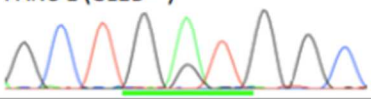
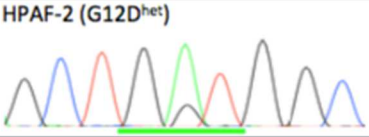
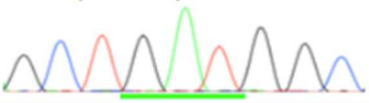
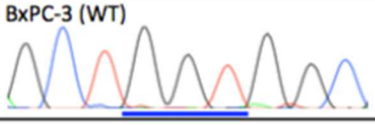
	Cell Line (KRAS Mutation subtype) G12V G12D WT	Cells Cut (#)	Mutation only (%)	Wild Type only (%)	Both alleles (%)	Overall ADO (%)	Mutation Detection (%)	Mutation Detection Rate (MDR)
HETEROZYGOUS	CFPAC-1 (G12V^{het}) 	60	15/39 (38%)	16/39 (41%)	8/39 (21%)	31/39 (79%)	23/39 (59%)	23 / 60 (38.3%)
	PANC-1 (G12D^{het}) 	44	12/23 (52%)	8/23 (35%)	3/23 (13%)	20/23 (87%)	15/23 (65%)	15 / 44 (34. 1%)
	HPAF-2 (G12D^{het}) 	27	9/19 (47%)	8/19 (42%)	2/19 (10%)	17/19 (89%)	11/19 (58%)	11 / 27 (40.7%)
	All Heterozygous Cells	**131**	**36/81 (44%)**	**32/81 (40%)**	**13/81 (16%)**	**68/81 (84%)**	**49/81 (61%)**	**49 / 131 (37.3%)**
HOMOZYGOUS	AsPC-1 (G12D^{homo}) 	21	14/14 (100%)	0/14 (0%)	0/14 (0%)	0/14 (0%)	14/14 (100%)	**14 / 21 (67.7%)**
	BxPC-3 (WT) 	13	0/8 (0%)	8/8 (100%)	0/8 (0%)	0/8 (0%)	0/8 (0%)	**0 / 13 (0%)**

Figure 39. Comparison of single cell KRAS sequencing results from different PDAC cell lines. Cell lines with both heterozygous and homozygous codon 12 mutations were analyzed, and the batch gDNA trace files for the different cell lines are depicted with codon 12 underlined. Based on the KRAS sequence base calls, individual cells were classified as demonstrating the mutation only, the wild type only, or both alleles, indicative of a heterozygous sequence. ADO rate was calculated as the number of single cell sequences that demonstrated only one of the two alleles in the heterozygous samples. Mutation detection rate (MDR) was defined as the percentage of single cells demonstrating a mutant allele over the total number of single cells isolated for sequencing. Homozygous cell lines were used as positive (AsPC-1) and negative (BxPC-3) controls.
ADO – allele dropout, MDR – mutation detection rate, het – heterozygous, homo – homozygous

KRAS mutation detection rate (MDR)

After establishing the success rate for amplification and sequencing of KRAS exon 2 from a single cell, we next investigated our ability to detect a mutation if it is known to be present. To do so we calculated and analyzed our mutation detection rate (MDR), defined as the percentage of single cells that demonstrated a mutant KRAS codon 12 allele over the total cells isolated. By comparing sequencing results from heterozygous and homozygous cell lines, we

were able to quantify the effects of allele dropout (ADO) on the MDR. We first sequenced batch gDNA to confirm the published KRAS mutations as well as their allele frequencies (**Figure 39**). Next, sequencing results from the 81 heterozygous single cells were interpreted using the peak-7 window function of ab1 Peak Reporter, with a threshold of >10% for base calling of the minor allele peak. The sequences were then classified as mutation only, both mutation and wild-type (heterozygous or double peak), or wild-type only (**Figure 39**). ADO rate was calculated as the percentage of single cells demonstrating only a single peak at the heterozygous site. We determined that ADO occurred in 68/81 (85%; range: 79-89%) single cells sequenced from the three heterozygous cell lines.

Given the high rate of ADO when sequencing single cells, we then investigated our ability to detect KRAS mutations from single cells. Of the 81 heterozygous single cells that demonstrated KRAS codon 12 coverage, only 49 (60.5%) demonstrated a mutant allele (**Figure 39**). As expected, all 14 (100%) of the homozygous AsPC-1 cells demonstrated a mutant allele, and no mutations were found in the 8 wild-type KRAS BxPC-3 cells. Overall, considering the entire workflow from cell capture to Sanger sequencing, we detected a KRAS mutation in 49/131 (37.4%) heterozygous cells and 14/21 (67%) homozygous cells. Thus, ADO during WGA was the single biggest limitation to successful single cell KRAS sequencing, greater than the error rate attributable to all other steps of our methods combined.

Performance of NanoVelcro/LCM on clinical samples

We developed an ICC definition for PDAC CTCs utilizing the NanoVelcro/LCM system that incorporates anti-CEA staining and nuclear morphology characteristics to the commonly used 3-channel ICC definition to improve the specificity for identifying true circulating cells of

tumor origin. (**Figure 37**). We then studied the performance of the NanoVelcro/LCM assay on peripheral blood samples from 12 PDAC patients. We identified a total of 119 CTCs from the 12 patients, and at least 1 CTC was found per patient (range: 1-34; average: 9.9; median: 8). Additionally, we identified and isolated a total of 103 WBCs from the 12 patients to serve as negative controls. We used LCM to isolate the 222 single cells, and performed single cell WGA and KRAS Sanger sequencing as outlined above. We found a mutant KRAS sequence in at least one CTC from 11/12 (92%) patients. Importantly, no mutant KRAS sequences were found in any of the WBCs isolated and sequenced, demonstrating the specificity of our CTC definition for the NanoVelcro/LCM platform.

Table 8. *KRAS* Sequencing of Single Cells from Patient Samples Demonstrating *KRAS* Mutations in CTCs but Not WBCs.

Cell type	Cells cut, *n*	*KRAS* coverage, *n*	Mutation detected, *n*
CTC	119	60	30
WBC	103	55	0
Total	222	115	33

CTC, circulating tumor cell; WBC, white blood cell.

We then analyzed our results to determine the cause of error for the single cells that failed sequencing. We found that the type of cell did not significantly influence our single cell sequencing results: we obtained KRAS codon 12 coverage in 60/119 (50.4%) CTCs and 55/103 (53.4%) WBCs (**Table 1**)**.** Of the 60 patient CTCs with KRAS coverage, a mutation was identified in only 33 (55%). Thus, considering the entire workflow from cell capture to Sanger sequencing, our MDR for isolated CTCs was 27.7%, as we detected a mutant KRAS sequence in 33/119 CTCs (**Table 9**).

Table 9. Stepwise workflow analysis for the 119 CTCs studied.

Cell type	PC CTCs, *n* (%)
LCM transfer	119 (100)
WGA	77 (64.7)
KRAS PCR	61 (51.2)
Sanger sequencing	60 (50.4)
Mutation detected	33/60 (55)
Mutation detection rate	33/119 (27.7)

CTC, circulating tumor cell; LCM, laser capture microdissection; PC, pancreatic cancer, WGA, whole genome amplification.

Overall, we found our CTC sequencing performed similarly to our cell line sequencing studies. While our CTC MDR was significantly lower than that found for the homozygous cell lines (χ^2 = 10.45, $p < 0.01$), it was similar to that of the heterozygous single cells ($\chi^2 = 2.26$, $p = 0.14$). On analysis of individual steps in the workflow, the only significant difference found between our patient CTC samples and single cells isolated from PDAC cell lines was the rate of successful PCR amplification, 79.2% vs. 94.6% respectively ($\chi^2 = 9.11$, $p < 0.01$).

Comparison of Multiple Cell Sequencing Results

As ADO was determined to be a significant source of error for single cell sequencing, we investigated ways to minimize it. Since ADO is a stochastic process, one means of overcoming it is by sequencing multiple single cells. We therefore investigated the number of single CTCs we would need to sequence in order to guarantee detection of a KRAS mutation. We first modeled our prediction based on our CTC sequencing results using single event binomial probability. We determined that if we were able to isolate 5 or 10 CTCs from a patient, we had an 80.2% or 96.3% chance of detecting a KRAS mutation in at least one CTC, respectively.

Another method for minimizing ADO is to increase the amount of template DNA by using more cells. To test if increasing the amount of starting template improved the detection rate, we performed our CTC sequencing protocol on groups of differing numbers of heterozygous HPAF-II cells (**Figure 40**). HPAF-II has a known mutant allele amplification, allowing us to also study how WGA amplification bias affects allelic fraction determination.[126] We first confirmed the

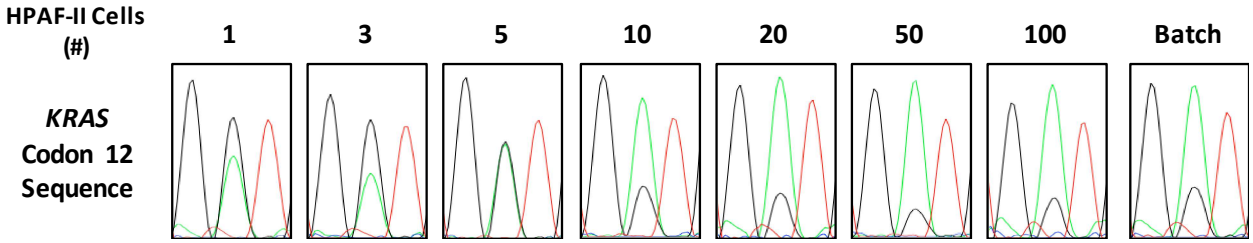


Figure 40. Determination of the number of cells needed to ensure reliable *KRAS* sequencing. By analyzing the allelic ratio of the sequencing results from different numbers of starting cells, the minimum cellular input for reliable *KRAS* codon 12 Sanger sequencing can be analyzed. The batch gDNA demonstrates the expected 75:25 G12D mutation: wild type allelic ratio for HPAF-II cells. For each group of differing numbers of cells, the replicate with the best sequence (i.e. with an allelic ratio closest to that of batch gDNA) from amongst the six replicates is depicted. As the number of cells sequenced increases, the allelic ratio becomes more similar to that of batch gDNA. Analysis of all replicates indicates a significant increase in mutation detection reliability above the 5-10 cell threshold.

75:25 mutant:wild type allele frequencies by digital PCR (**Figure 38**). Then, a total of 6 replicates were run for each group of differing cell numbers. KRAS mutations were detected in 4/6 (67%) single cells, 4/6 (67%) of the 3 cell groups, 6/6 (100%) of the 5 cell groups, 5/6 (83%) of the 10 cell groups, and 6/6 (100%) of the 20, 50, and 100 cell groups. A heterozygous sequence, demonstrating both alleles without ADO, was obtained in 17% (1/6) single cells, 17% (1/6) of the 3 cell groups, 33% (2/6) of the 5 cell groups, 67% (4/6) of the 10 cell groups, 67% (4/6) of the 20 cell groups, and 100% (6/6) of the 50 and 100 cell groups. An accurate allelic ratio of the mutant allele to the wild type allele (75:25) was not obtained from any of the 1, 3, or 5 cell groups, 17% (1/6) of the 10 cell groups, 50% (3/6) of the 20 cell groups, and 100% (6/6) of the 50 and 100 cell groups.

4.2.5 Discussion

CTCs have the potential to allow clinicians access to tumor tissue from a simple peripheral blood draw. Such a liquid biopsy has innumerable applications, from diagnosing patients with difficult to biopsy tumors to the real-time monitoring of a cancer's genotypic changes. These applications are only going to become more important as precision medicine advances and companion diagnostics become an essential tool for making treatment decisions. However, for CTCs to realize their potential, important limitations of sequencing data from single cells must be understood and addressed.[119] The first is ensuring that the CTC identification methods used successfully discriminate CTCs from other circulating cells with high specificity. This high specificity is required for successful adoption of CTCs as a realistic alternative to traditional biopsies. Secondly, the isolation and sequencing of CTCs is difficult, requiring multiple independent techniques that all introduce possible sources of error that can lead to inaccurate representation of the tumor genotype. These errors must be studied and addressed if CTC sequencing is to have clinical relevance. Finally, the detection efficiency of CTC sequencing for actionable mutations must be established in order to assess the adequacy of a sample for genetic analysis, and prevent false negative results from interfering with patient care. In this study, we attempted to better understand the potential and limitations of the NanoVelcro/LCM platform for single cell KRAS sequencing of CTCs in PDAC. We utilized PDAC cell lines to test our methodology and found that most sequencing failures occurred during WGA. We compared the sequencing results from heterozygous and homozygous single cells and found that ADO was the primary limiting factor of mutation detection from heterozygous samples. We then used the NanoVelcro/LCM platform to identify CTCs and other hematopoietic cells for isolation and single cell KRAS sequencing from 12 PDAC patients. We

demonstrated mutant KRAS sequences in CTCs from 11/12 (92%) patients, and never found a mutant KRAS sequence in any of the other hematopoietic cells sequenced from the same patient. By analyzing our sequencing results, we were able to demonstrate that at least 10 CTCs are likely required to reliably determine KRAS mutations. We then confirmed this finding by sequencing groups of varying numbers of cells and again found a distinct drop in mutation detection for heterozygous alleles with less than 10 CTCs.

PDAC is a model cancer for the study of CTC sequencing due to the ubiquity of KRAS codon 12 mutations in PDAC, an early driver mutation found in over 95% of patients. Thus, by sequencing a single codon, the cancerous origin of CTCs can be demonstrated molecularly. To ensure that our results would model the realities of clinical CTC sequencing, we chose to use Sanger sequencing, which has a lower sensitivity than newer methods, but is still the gold standard for clinical sequencing.[127] In analyzing why sequencing failed, we found that the majority of failures occurred during WGA, either from complete failure of the WGA reaction or, more commonly, due to ADO. Our ADO rate of 85% was near the upper limit of previously reported single cell MDA-based WGA studies, which have ranged from as low as 10% to as high as 90%.[121,122,128,129] This high ADO rate may be at least partially attributed to our use of Sanger sequencing, which is less sensitive than newer methods to low mutant allele fractions.[130]

For clinical samples, we found CTCs in 100% of the 12 patients, and confirmed a KRAS mutation in at least one CTC from 92% of the patients. Since the zygosity of KRAS mutations in our clinical samples was not known, we compared our CTC sequencing results to that from both homozygous and heterozygous cell lines. Our MDR for KRAS in clinical samples of 27.7% was significantly lower than that found from homozygous single cells, but statistically similar to that of heterozygous single cells that we sequenced. However, previous studies estimate that around

70% of KRAS mutations in PDAC are heterozygous, indicating that the similarity of our CTC sequencing results to the heterozygous cell lines is to be expected.[131,132] Overall, our SNP detection rate from CTCs is comparable or better to previous studies utilizing Sanger sequencing, which have reported rates as low as 10%.[29,133,134]

Our findings are important for two aspects of CTC research. First, for CTC technologies to move from an enumeration biomarker to a molecular diagnostic, different parameters are likely to become important. Previous research had frequently attempted to liberalize CTC definitions in an attempt to increase the number of CTCs found from patients.[8,135] However, for CTC sequencing studies every effort must be made to ensure the specificity of CTC identification methods, as the introduction of any normal cells within the population would increase the risk of false negatives to the detriment of patients. To that end, we utilized the modular nature of the NanoVelcro/LCM CTC platform to incorporate anti-CEA staining and nuclear morphology into our CTC definition to hopefully increase its specificity. Furthermore, we used CTC sequencing to validate our ICC definition, finding KRAS mutations only in cells we called CTCs, and never in other hematopoietic cells. We found using an orthogonal method, such as molecular analysis, to confirm ICC definitions for CTCs important for demonstrating the tumor origin of the cells captured. We think it prudent for new CTC platforms to employ such a method of verification to address the known limitations of ICC alone in defining CTCs.

Secondly, our study provides insight into the reality of CTC sequencing for clinical applications as well as the quantity of CTCs required for reliable mutation detection. While previous single cell CTC sequencing experiments have demonstrated the ability to perform multigene panels and even whole exome analysis of individual CTCs, the parameters required for reproducible, accurate sequencing for clinical use have not been studied.[27,28,136] Our data

indicated that WGA was the most error-prone step in performing single CTC sequencing, and that ADO was the most common reason for sequencing failure. However, we also demonstrated that these limitations could be overcome if at least 10 individual CTCs are sequenced. Furthermore, we found a similar cutoff of at least 20 cells for accurately assessing KRAS allelic fractions in a PDAC cell line with a known mutant allele amplification. Overall, our study provides further evidence that the amplification bias present at the single cell level responsible for ADO can be reduced if at least 10 CTCs are sequenced together.

In conclusion, we demonstrated the mutation detection rate of the NanoVelcro/LCM platform for KRAS sequencing of single CTCs. We investigated the shortcomings inherent in work at the single cell level, and analyzed ways to account for and minimize sources of error. Finally, we established a cutoff of at least 10 CTCs as a lower limit for reliable KRAS mutational analysis from CTCs. Future studies looking at other actionable genetic targets, as well as other molecular analyses such as gene panels, exome sequencing and RNASeq, will need to consider the sources of error that we have highlighted in our study. We are hopeful that future studies will further confirm the ability of CTCs to function as a liquid biopsy, an important step in advancing precision cancer treatment.

4.3 Single Cell Precision Oncology: Selection and Generation of Optimal Whole Genome Amplification Products for Sequencing Applications

4.3.1 Abstract

Introduction: Sequencing analysis of circulating tumor cells (CTCs) enables "liquid biopsy" to guide precision oncology strategies. However, this process requires low-template WGA that is prone to errors and biases from uneven amplifications. Currently, quality control methods for WGA products, as well as the number of CTCs needed for reliable downstream sequencing, remain poorly defined. We sought to define strategies for selecting and generating optimal WGA products from low-template input as it relates to their potential applications in precision oncology strategies. **Methods:** Pancreatic cancer cells (HPAF-II) were isolated using laser microdissection. WGA was performed using multiple displacement amplification (MDA), multiple annealing and looping based amplification (MALBAC) and PicoPLEX. Quality of amplified DNA products were assessed using a multiplex/RT-qPCR based method that evaluates for 8-cancer related genes. A quality control (QC) score of 0-8 was assigned for each WGA product. We further developed a modified-MDA protocol that differs from the conventional protocol in 2 key aspects: 1. extended cell lysis and 2. small volume reaction (3ulx16 parallel amplifications vs. single 50ul bulk amplification). WGA products were subjected to Sanger sequencing, array comparative genomic hybridization (aCGH) and next generation sequencing (NGS) to evaluate their performances in respective downstream analyses. **Results:** Single-cell WGA products exhibited a significant sample-to-sample variability in amplified DNA quality as assessed by our 8-gene QC assay. Increasing the number of cellular input resulted in improved QC-scores overall, but a reliable WGA reaction that consistently passed the QC step (QC-

score=8) required a starting cellular input of 20-cells. Modified-MDA reaction effectively reduced this number, achieving reproducible high-quality WGA reactions using 5-cells as a starting template. Single-cell WGA products that passed the pre-analysis QC had lower amplification bias and improved aCGH/NGS performance metrics when compared to single-cell WGA products that failed the QC. A starting cellular input of 5 to 10-cells amplified using the modified-MDA achieved aCGH and NGS results that closely matched that of unamplified, batch genomic DNA. **Conclusion:** The 8-gene QC assay and the modified-MDA protocol serve as effective strategies with potential to enhance workflows involving low-template WGA reactions. Furthermore, a threshold number of 5-10 cells are likely needed for a reliable WGA reaction with high fidelity to the original starting template.

4.3.2 Introduction

"Liquid biopsy" of circulating tumor cells (CTCs) has been suggested in many recent studies as an ideal biopsy technique for precision oncology applications.[137-141] CTCs are thought to arise from both primary and metastatic lesions, allowing for a more comprehensive representation of the tumor genomic make-up. [142] Furthermore, the need for only a simple peripheral blood draw in "liquid biopsy" makes it amenable to repeated samplings without incurring significant costs or risks to patients. Although successful CTC enrichment, capture and downstream molecular analysis has been described, major obstacles still remain prior to its clinical translation (**Figure 41**).[142,143] One major challenge is the limited number of CTCs available for molecular analysis. The number of CTCs obtainable from a single peripheral blood remains highly limited, with most studies showing <5 CTC/mL from a single peripheral blood.[139,144] For many GI cancers, especially pancreas ductal adenocarcinoma, CTCs are even

more limited.[67,105] Thus, for molecular analysis to be performed using CTCs, the limited amount of genomic materials available from CTCs must undergo whole genome amplification (WGA) to generate adequate quantities of DNA for downstream sequencing analysis.

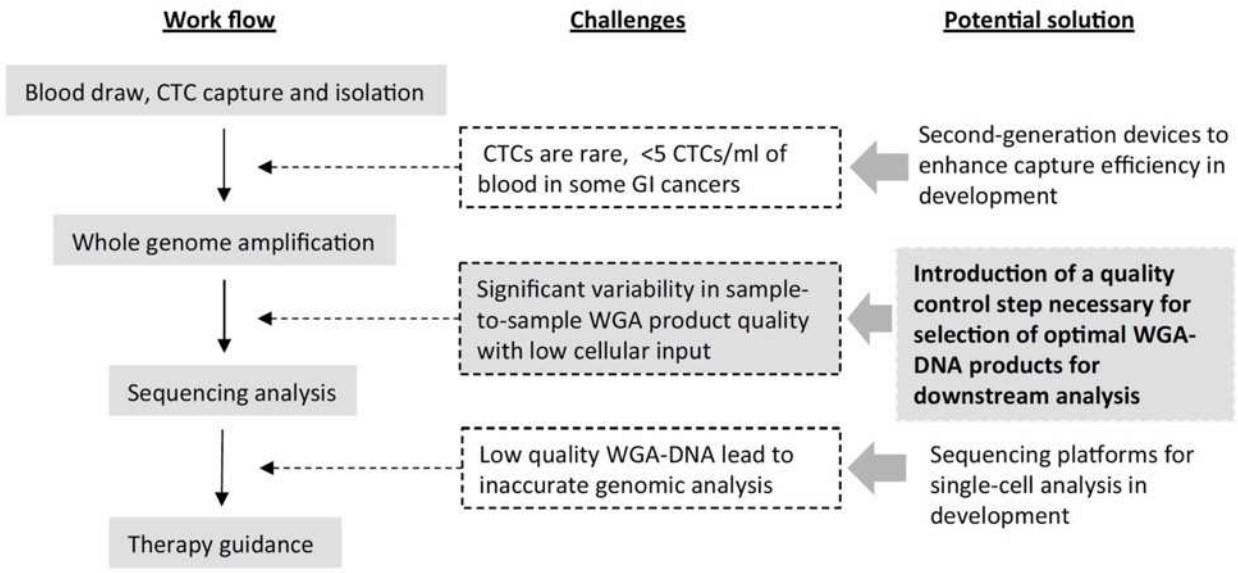


Figure 41. Overall workflow for a "liquid biopsy" in solid tumors.

Despite recent advances in WGA techniques, amplification processes remain prone to uneven amplifications, resulting in amplification bias.[145-148] For heterozygous sites, this can result in a total loss of one allele, a phenomenon called allele drop out (ADO).[149,150] This is especially problematic when working with a small number of cells, as in CTC analysis, as the stochastic variation in the WGA process is exacerbated when starting with low copies of genomic input.[151,152] This results in considerable variability in sample-to-sample quality when working with CTCs.[149] Low-quality WGA products with significant degrees of amplification bias and ADO are inappropriate for precision oncology applications as they fail to accurately represent the original genomic template. Therefore, one must be able to accurately differentiate between high- and low- quality WGA products in order to ensure accurate sequencing results for guiding cancer therapy.

Currently, however, quality control metrics and selection criteria for high-quality WGA products from single cells have not been sufficiently defined. Much of the existing literature utilizing whole genome amplified DNA lack analysis of the quality of WGA samples being used. Given the known sample-to-sample variability in minimal template WGA products, WGA quality must be defined in order to accurately interpret and compare WGA-DNA derived data. Furthermore, although prior studies have shown improved WGA quality with increasing amount of genomic template input, the number of CTCs needed to generate a WGA product suitable for sequence analysis to be utilized in precision oncology strategies remain unknown.[153,154]

In the current study, we sought to define strategies for selecting and generating optimal WGA products from samples ranging from one to twenty cells as it relates to their potential applications in precision oncology strategies. To this end, we developed a quality control (QC) assay to help facilitate the selection of high-quality WGA product suitable for use in downstream sequencing applications, including point mutation detection, array comparative genomic hybridization (aCGH) and next generation sequencing (NGS). In order to better understand the key determinants of WGA quality, we compared various WGA methods in addition to the number of input cells to determine their influence upon amplification reactions. We then used our findings to develop a modified multiple displacement amplification (MDA) protocol with a notable improvement in amplified DNA quality over the conventional MDA protocol. Ultimately, we utilized these findings to determine the "threshold" number of cells needed for reliable molecular analysis that could be utilized in precision oncology strategies.

4.3.3 Methods

Cell Lines

Pancreatic cancer cell line HPAF-II was obtained from American Type Culture Collection (ATCC, Virginia, USA), and grown using EMEM medium (ATCC) supplemented with 10% fetal bovine serum (ATCC) and 100 U/mL penicillin-streptomycin (ATCC). All cell lines were grown at 37^{o}C with 5% CO_2 and were routinely passaged at 80% confluence using an iso-osmotic sodium citrate solution for cell release (Thermo, Massachusetts, USA).

Laser Micro dissection

In preparation for laser microdissection, cells were released from the culture plates using the iso-osmotic sodium citrate solution (Thermo). Following a wash with the culture medium, each cell line was diluted to a density of 1000 cells per 100 µL. Approximately 1000 cells (100 µL) were smeared on PEN membrane slides (Leica, Wetzlar, Germany), air-dried for 10 minutes, and fixed with 100 µL of 100% ethanol. Cells were then isolated using the PALM MicroBeam laser microdissection system (Zeiss, Oberkochen, Germany). 1, 5, 10 or 20 cells were laser microdissected and collected into 200 µL opaque tube caps (Zeiss) using the laser pressure catapult function. Cell transfer to the tube cap was confirmed by imaging the cap prior to cap closure using the cap-check function.

Whole genome amplification

Isolated cells were subjected to genomic DNA isolation and WGA using one of the three commercially available single-cell WGA kits according to the manufacturer's protocol: REPLI-g

Single-Cell Kit (Qiagen, California, USA), Multiple Annealing and Looping Based Amplification Cycles (MALBAC) Single-cell WGA Kit (Yikon Genomics, Beijing, China) and PicoPlex Single-Cell WGA Kit (Rubicon, Michigan). Reactions were performed a minimum of 3 times for all conditions tested. WGA products were purified using the QIAquick PCR Purification Kit (Qiagen) and quantified with NanoDrop 2000 (Thermo).

Modified MDA protocol was performed using the same reagents available from the REPLI-g Single Cell Kit (Qiagen), with modifications made in the cell lysis step and the final amplification step. Cell lysis was performed over a course of 30 minutes to ensure complete lysis of genomic material from isolated cells (as opposed to 10 minutes recommended by the manufacturer). Prior to the final amplification step, the 50-ul MDA reaction mix was mixed for 30 seconds by pipetting up and down and then partitioned into 16 individual reactions (approximately 3ul each) and the amplification occurred at 30^{o}C for 8 hours. This resulted in 16 individual MDA reactions with reduced individual reaction volumes (3ul instead of 50ul), all taking place in parallel. Following the amplification step, contents within the 16-wells were collected into one tube, followed by the purification and quantification step as described above.

Development of a quality control (QC)-score for WGA-DNA

Prior literatures have described the role of multiplex PCR in evaluating DNA quality isolated from formalin-fixed paraffin embedded (FFPE) tissues prior to downstream analysis using aCGH.[155] Based on this concept, we developed an 8-gene multiplex PCR/quantitative PCR (qPCR) based QC assay for evaluation of WGA-DNA quality. We selectively chose for 8 cancer-related genes that are considered molecular targets for targeted cancer therapy and therefore highly implicated in guiding therapeutic decisions. This way, the QC assay identifies

WGA products suitable for precision oncology applications by evaluating for the coverage and "accessibility" of these important genomic locations within the WGA-DNA. Genes evaluated by the QC assay included BRAF, EGFR, KIT, KRAS, NRAS, PIK3CA, PTEN and P53.
Score of 0-8 was assigned based on the number of genes successfully amplified and detected using multiplex PCR and qPCR. Failure to detect one or more of the 8-gene from the WGA product signifies lack of coverage or potential ADO during the WGA process, both of which indicate a poor-quality WGA product and jeopardizes accurate representation of the original starting genome. Thus, we only gave WGA-DNA products with 8 out of 8 score a "pass" and deemed them fitting for further downstream analysis.

A secondary QC-score was generated to confirm the internal validity of the original QC-score based on the 8-cancer related genes. The secondary QC-score was generated from a distinct set of 8 primer pairs representing 8 housekeeping genes: NDUFA, UQCRC, ACTG, CYB5A, GABA-RAPL, MIF, MYC, PRPH.

Multiplex PCR and quantitative PCR (qPCR)

WGA products were subjected to multiplex PCR using primer sets representing genes mentioned above. Primer sets used for the primary QC assay included: BRAF (forward 5' – TAC TGC TCT TTC TTC TCC AAC AC – 3'; reverse 5' – CCT GAT TGT ATT TGA GAT CTA GTA GGG – 3') EGFR (forward 5'-CAG CCT TCT CCG TAA TTA GCA T – 3'; reverse 5' – TGA CAC AGA TAA TTG TCC CAC AG – 3') , KIT (forward 5' – GGC ATT GAG GAG GGA TAG TAA AT – 3'; reverse 5' – CTG AAC AAT TTG CTT GAA TGT TGG – 3'), KRAS (forward 5' – GTG TTA CTT ACC TGT CTT GTC TTTG – 3'; reverse 5' – GCC TTC TAG AAC AGT AGA CAC AA – 3'), NRAS (forward 5' – AAT GGA ATC CCG TAA CTC TTG G

– 3’; reverse 5’ – GAT GAT GTA CCT ATG GTG CTA GTG – 3’), PIK3CA (forward 5’ – AGG GCA AAT AAT AGT GGT GAT CT – 3’; reverse 5’ – CAG CAA TTA CTT GTT CTG GTA CAC – 3’), PTEN (forward 5’ – CTT TCT CTA GGT GAA GCT GTA CT – 3’; reverse 5’ – GGT TCA TTG TCA CTA ACA TCT GG – 3’) and P53 (forward 5’ – AAG AGA AGC AAG AGG CAG TAA G – 3’; reverse 5’ – CTT AGG CTC CAG AAA GGA CAA G – 3’). Primer sets used for the secondary QC assay included: NDUFA7 (forward 5’ – TGC TCT GGA TGT GAA GAT GCC A – 3’; reverse – 5’ – TTC CAG GTA AAT CCA GCC CAG G – 3’), UQCRC1 (forward 5’ – CAG CCA GTC AGC ATC ATC CAA C – 3’; reverse 5’ – GAA AGC CGG ATT GCG GTA ACA T - 3’), ACTG1 (forward 5’ - GCT CAA TGG GGT ACT TCA GGG T – 3’; reverse 5’ – GTG GAC GTT ACG TAA AAG GCC C – 3’) , CYB5A (forward 5’ – GGC AAC GCT TAG ACT CTG TGT G – 3’; reverse 5’ – CTG CCC TTG GCC TAA CTA ACC T – 3’),, GABARAPL2 (forward 5’ – CCA GCC AAT TCA TGA GTC GGT G – 3’; reverse 5’ – CCT GAC AAC TCG CAA GTA GCA C – 3’), MIF (forward 5’ – AGA AGT CAG GCA CGT AGC TCA G – 3’; reverse 5’ – GGC ACG TTG GTG TTT ACG ATG A – 3’), MYC (forward 5’ – GGA TAG CTC TGC AAG GGG AGA G – 3’; reverse 5’ –TCG TCG CAG TAG AAA TAC GGC T – 3’), PRPH (forward 5’ – GTT CCT CAA GAA GCT GCA CGA G – 3’; reverse 5’ – CGT TAG ACT CTG GAT CTG GCG T – 3’). PCR reactions were carried out on a C1000 Thermal Cycler (Bio-Rad) with the Multiplex PCR Plus Kit (Qiagen) using total volumes of 17 μL per reaction. The reaction conditions were as follows: denaturation at 94^{o}C for 30 seconds, annealing at 64^{o}C for 30 seconds, and extension at 72^{o}C for 30 seconds for a total of 10 cycles.

Following the pre-amplification step by multiplex PCR, the resulting amplified products were analyzed and detected using qPCR. Reactions took place on BioRad CFX-96 real time

system (BioRad) using the QuantiTect SYBR Green PCR Kit (Qiagen). A 25-ul reaction mixtures were prepared, which contained 12.5 ul of the SYBR Green PCR Master Mix (Qiagen), 9.5 ul of RNA-grade water, 1 ul of individual primers sets (10uM) and 2ul of the multiplex PCR product. The reaction conditions were as follows: 95 ºC for 15 minutes, followed by 35 cycles of 94 ºC for 15 seconds, 64 ºC for 20 seconds, and 72 ºC for 20 seconds. The plate was read following the extension step at 72 ºC. Melting curve analysis was performed between 70-95 ºC at 0.5*C intervals. Real-time PCR data were reviewed and analyzed using the CFX manager (BioRad). Specificity of the PCR amplification product was determined using melting curve analysis. PCR products with melting-temperature (Tm) matching the expected value based on primer sequence, and threshold cycle (Ct) <30 were counted as reliable amplification and detection.

Array-CGH

Sample WGA-DNA and reference DNA were differentially labeled with cyanine-3 (CY3) and cyanine-5 (Cy5) dyes using the GenetiSure Amplification and Labeling Kit (Agilent) according to the manufacturer's protocol. Briefly, a 15.5 ul reaction mixture containing sample or reference DNA, Random Primer Mix and water was placed in 98C for 3 minutes for DNA denaturation and transferred to 4C for 3 minutes. Next, 9.5 ul of Labeling Master Mix containing 5ul of 5x Reaction Buffer, 2.5ul of 10x dNTP Mix, 1.5 ul of Cy5-dUTP/Cy3-dUTP and 0.5 Exo (-) Klenow were added to each tube containing the sample/reference. DNA labeling reaction was performed in 37C for 45 minutes followed by inactivation step in 65C for 10 minutes. Labeled DNA was purified using the Post Labeling Purification Columns (Agilent) according to the manufacturer's protocol. Purified labeled DNA samples were prepared for hybridization, which

took place on Agilent 8x60K CGH microarray slides at 67C for 6 hours. Following the hybridization, the slides were washed per manufacturer protocol, and prepared for scanning using the Agilent SureScan Microarray Scanner (Agilent). Microarray image was prepared and analyzed using the Agilent CytoGenomics software (Agilent).

Sequencing library preparation and sequencing

Purified WGA products were sheared to generate DNA fragments of 350bps using the Covaris sonicator (Covaris). Following cleanup of the sonicated DNA, end-repair and ligation were performed using the KAPA DNA Library Preparation Kit (KAPA Biosystems) according to the manufacturer's protocol, followed by library amplification by PCR. Sequencing was performed on an Illumina NextSeq 500 using random primers and pair-end reads of 75bps (2X75bps).

Sequencing analysis and visualization

The sequencing read data was analyzed using Ginkgo, an open-source platform for the analysis of single-cell somatic copy-number alteration (SCNA). Detailed description of the analysis methodology is as provided by Garvin et al.[156] Briefly, mapped reads were binned using variable length-binning and underwent GC bias normalization prior to segmentation using circular binary segmentation. Quality metrics data, including Lorenz curve, histogram of read count distribution and index of dispersion were obtained as a part of the Ginkgo analysis pipeline. Sequencing data visualization was performed using Nexus software (Biodiscovery, Inc.)

KRAS PCR and Sanger sequencing

PCR amplification of KRAS exon 2 was performed using the following primer (Integrated DNA Technologies): Forward 5' – AAG GTA CTG GTG GAG TAT TTG – 3' and Reverse 5' – GTA CTC ATG AAA ATG GTC AGA G – 3', with expected amplicon length of 295 bps. PCR reactions were carried out on a C1000 Thermal Cycler (Bio-Rad) with Platinum PCR SuperMix High Fidelity Kit (Invitrogen) using total volumes of 50 μL per reaction according to the manufacturer's protocol. The reaction conditions were as follows: denaturation at 94^{o}C for 30 seconds, annealing at 55^{o}C for 30 seconds, and extension at 68^{o}C for 45 seconds for a total of 40 cycles. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and eluted into 50 uL of nuclease-free water (Qiagen). DNA was diluted to a concentration of 10 ng/uL based on Nanodrop quantification of the PCR product. Automated dideoxy terminator sequencing was performed by capillary electrophoresis by the UCLA GenoSeq Core on an ABI 3730 DNA analyzer using Big Dye Terminator chemistry (Applied Biosystems). All sequences were analyzed by manual inspection of the individual trace files using Four Peaks (Nucleobytes).

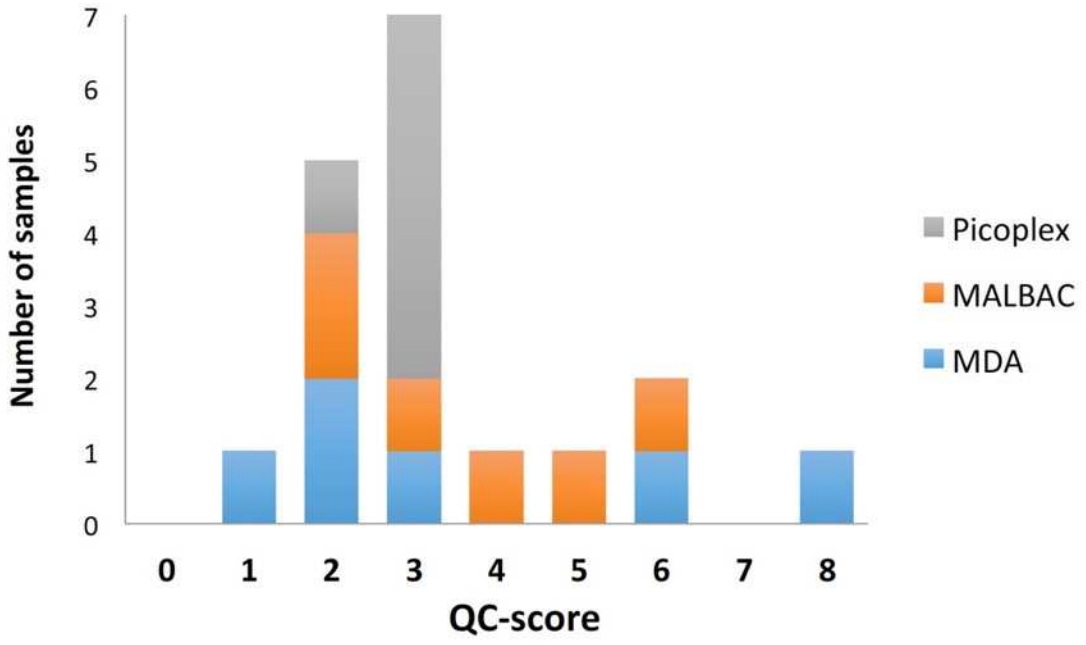


Figure 42. 8 oncogene quality control score assessment of 3 different WGA methods. DNA from 6 single HPAF-II cells were amplified using each of the 3 methods and the QC score assessed. Only MDA resulted in a perfect 8/8 score.

4.3.4 Results

WGA product variability in single-cells

We applied our 8-gene QC assay to single-cell WGA-DNA samples. We tested three different WGA reaction methods: MDA, MALBAC and PicoPlex. Total of 18 individual HPAF-II cells were isolated using laser micro-dissection and used in the respective WGA methods. The experiment was repeated 6 times in order to account for the variability expected with the single-cell WGA process. QC assay was performed and the QC-score was assigned for each WGA product.

Despite using a clonally expanded cell line and performing WGA reactions in a parallel fashion under the same condition, we noted significant variability in sample-to-sample WGA-DNA quality as assessed by the 8-gene QC assay (**Figure 42**). QC-scores for MDA amplified

single-cells ranged between 1 and 8. Similar variability in amplified DNA quality was noted for MALBAC, with QC-scores ranging between 2 and 6. Although PicoPlex resulted in less quality disparity, majority of samples achieved scores of only 3. QC-score profiles for the MDA and MALBAC methods were overall similar, but the WGA product that passed our QC criteria (QC-score = 8) was only found in the single-cell sample amplified with using the MDA method.

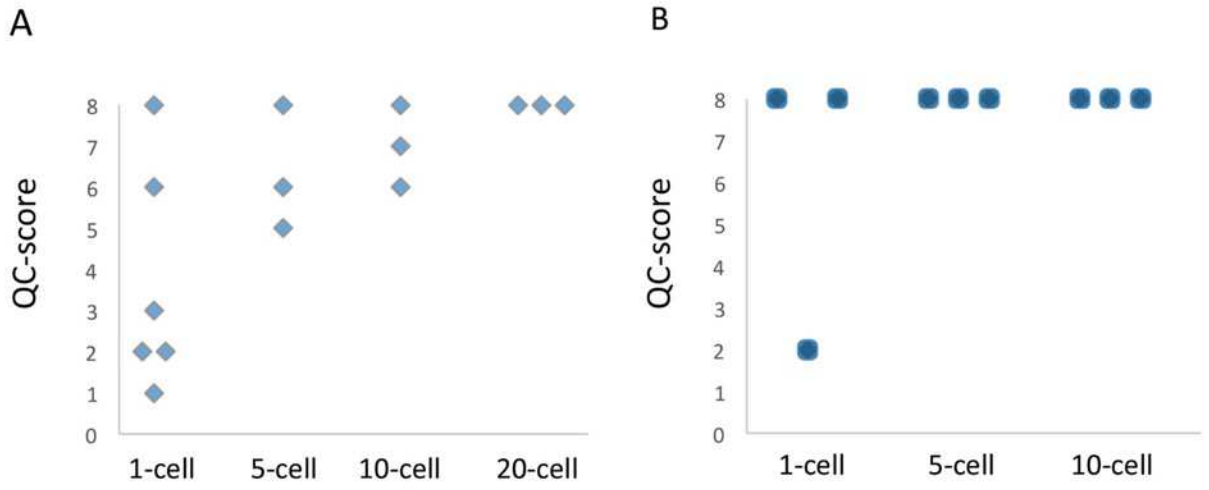


Figure 43. Assessment of the starting number of cells needed to reliably produce high quality amplified DNA based on 8-gene QC panel results. **(A)** Using the standard repli-G MDA protocol. **(B)** Using the modified WGA protocol that includes more thorough lysis and smaller MDA reaction volumes.

Determination of number of cells needed to achieve reproducible WGA product quality

We noted significant random variations in WGA product quality when using single-cells as starting template (**Figure 42**). Increasing the amount of starting genomic template has been shown to improve WGA product quality.[153,154] Thus, we hypothesized that there would be a certain "threshold" number of cellular input above which a reliable and reproducible WGA process is possible, i.e. a number of cells which will achieve a "pass" in our QC step on a consistent basis.

In order to test this hypothesis, 5-cell, 10-cell and 20-cell samples of HPAF-II cells were cut and isolated using a laser micro-dissector. WGA was performed using the MDA method, followed by the QC assay and assignment of the QC-scores (**Figure 43A**). Less variability in amplified DNA quality was noted when 5 or more cells were used as starting genomic template. All samples persistently achieved a QC-score of ≥5 (5-cell) or ≥6 (10-cell), as opposed to single-cell WGA reactions with scores ranging from 1 to 8. Furthermore, at least one of the triplicates in 5-cell and 10-cell group passed the QC-step (QC-score = 8). However, a highly reliable WGA process with all of the WGA products passing the QC step could not be achieved until 20 cells were used as starting genomic template input.

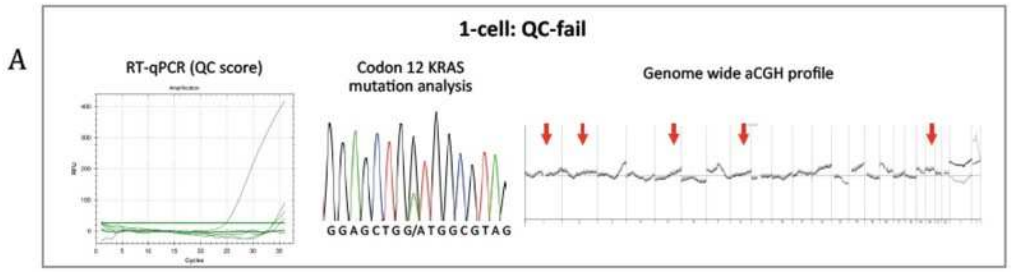
Modified MDA reaction to improve quality of WGA product from limited template samples.

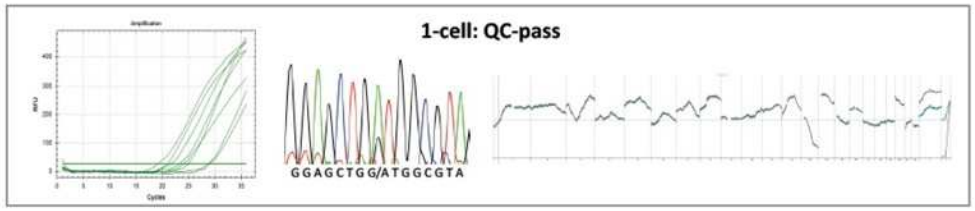
With the existing MDA protocol, multiple single-cell WGA reactions must be performed in order to obtain one high-quality WGA product suitable for downstream analysis. Even when more than a single-cell was used, a reliable WGA reaction with its amplified DNA product passing the QC step on a consistent basis could not be achieved until 20 cells were used. (**Figure 43A**) For many GI cancers that are known to generate only few CTCs (i.e. pancreas ductal adenocarcinoma), obtaining 20 CTCs from a single peripheral blood draw may be unrealistic.

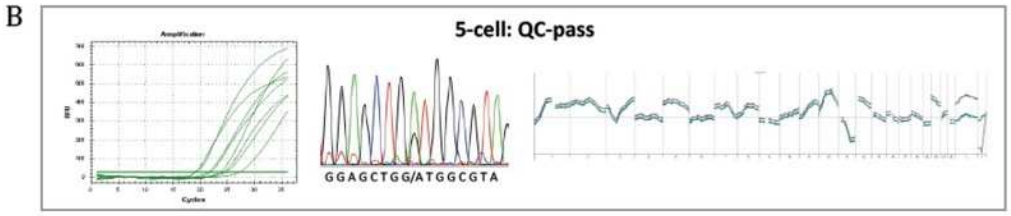
Given these problems, we sought to develop a modified WGA protocol that would reduce the sample-to-sample variability in DNA quality and lower the number of starting cells needed to achieve a reliable WGA process. Multiple prior reports have described the benefit of performing MDA by partitioning the reaction into parallel smaller volume reactions.[150,152,154] The small reaction volumes and template DNA partitioning restricts the degree of aberrant preferential

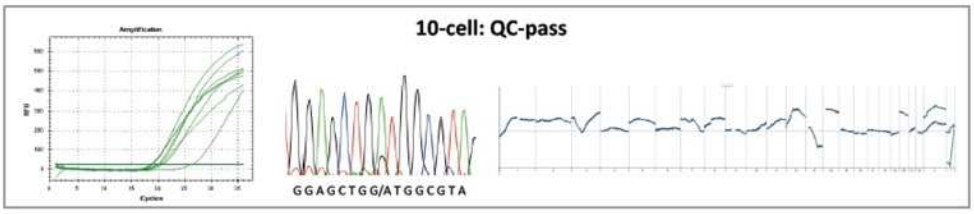
amplification, leading to a more uniform WGA process overall. Based on this principle, we developed a modified MDA protocol with key changes in two aspects: (1) increasing the cell lies step from 10 minutes to a minimum of 30 minutes to ensure adequate release of genomic materials from cells, and (2) partitioning the final MDA reaction into 16 individual reactions containing ~3ul each on a 96-well plate, prior to the final isothermal amplification step at 30ºC for 8 hours.

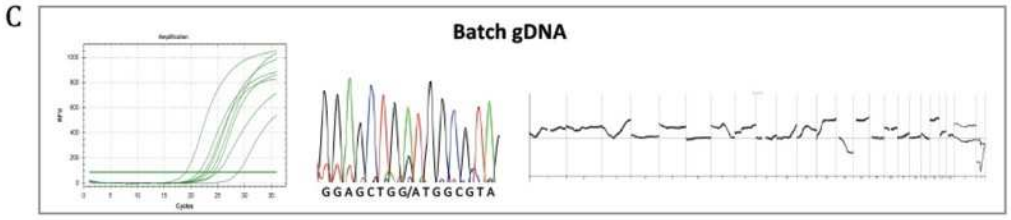
WGA products obtained using this modified MDA protocol resulted in improved reproducibility and higher QC-scores overall (**Figure 43B**). Amplification reaction gain was overall lower for the modified MDA products compared to the conventional MDA, with reduction of approximately 50% on average for each sample undergoing the modified MDA protocol. For single-cells, 2 of 3 WGA reactions resulted in a product with the perfect QC-score of 8, compared to 1 in 6 WGA reactions using the conventional MDA protocol. Importantly, for 5-cell and 10-cell samples, all WGA reactions generated products that passed our QC criteria (**Figure 43B**). The modified MDA protocol in our hands effectively reduced the "threshold" number of cells needed for a reliable WGA reaction down to 5 cells from 20 cells, well within the number of CTCs attainable from a single peripheral blood draw.











D

DNA source	DLRS	Signal to noise ratio
1-cell WGA (QC-fail)	1.3	3.96
1-cell WGA (QC-pass)	0.72	53.6
5-cell WGA (QC-pass)	0.57	34.5
10-cell WGA (QC-pass)	0.40	48.1
Batch gDNA	0.30	33.1

Figure 44. Correlation of 8-gene QC assay score with sequencing results. **(A)** Comparison of single cell samples that either passing or failing QC-score demonstrating that a failing QC score results in unreliable sequencing results. **(B)** 5- and 10- cell samples routinely passed QC and reliably reproduced the batch gDNA results for Sanger sequencing and aCGH. **(C)** Batch genomic DNA **(D)** comparison of aCGH QC parameters for samples illustrated above.

WGA product QC assay using 8- vs. 16-gene multiplex PCR

Our proposed 8-gene QC assay evaluates 8 genomic locations within a WGA product. Evaluating more loci in theory provides more comprehensive evaluation of the amplified DNA product. We tested whether there was any benefit to evaluating more genetic loci beyond the 8-cancer genes during the QC process. To test for this, we performed a secondary QC assay using a different set of 8-housekeeping genes on the same single-cell conventional MDA products, as well as the 1/5/10-cell modified MDA products. The original (8-cancer gene) and the secondary (8-housekeeping genes) QC assays generated highly concordant QC-scores for all of the samples tested. All of the samples that passed the QC-step based on the original QC assay also passed using the secondary QC assay. For the sole modified MDA sample with the low QC-score (sample 1-C), the secondary QC also resulted in a similarly low QC-score indicating poor quality WGA-DNA. Thus, evaluation of more genetic loci beyond the original 8-cancer genes did not provide any additional QC information and did not change QC results for any of the samples tested.

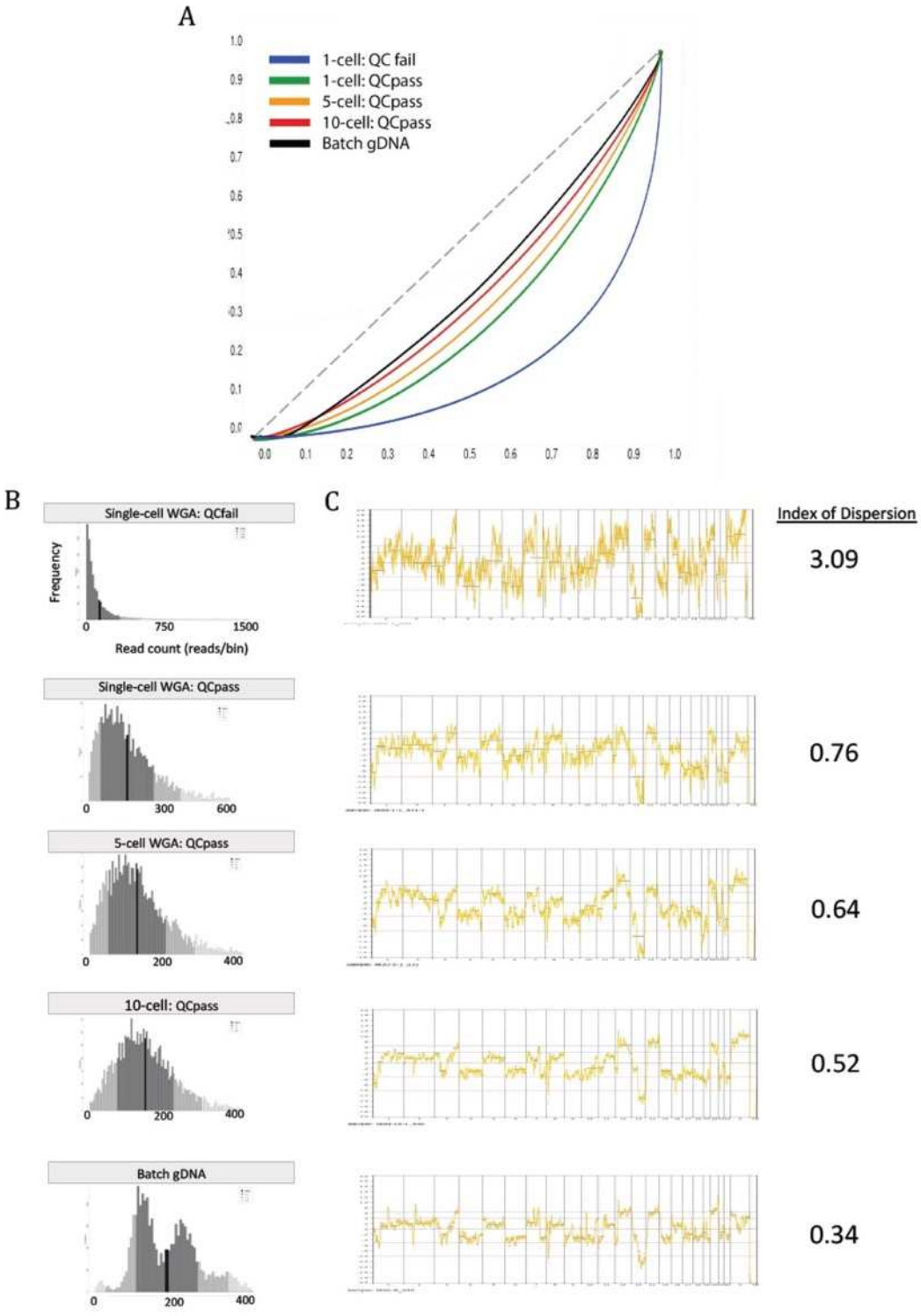


Figure 45. Low-resolution whole genome sequencing copy number profiling. **(A)** Lorenz curve demonstrates the importance of the QC-score as a pre-sequencing quality control metric. **(B)** Read count frequency. **(C)** Low-resolution whole genome copy number profiles and index of dispersions for cell groups sequenced as well as batch gDNA.

Using 8-cancer gene QC assay to select samples for downstream aCGH and NGS applications

We performed point mutation detection, aCGH and NGS using single-cell WGA products that either passed (WGA-QCpass) or failed (WGA-QCfail) the QC-step. **Figure 44** and **Figure 45** illustrate the relationship between QC results and performances in various downstream applications.

Point mutation detection within the KRAS gene was successful in both QCpass and QCfail WGA products (**Figure 44A**). However, while the QCpass WGA sample exhibited a mutant to wild-type allelic ratio closely matching that of the batch gDNA (HPAF-II cells are known to have 4 to 1 mutant:wild-type allelic ratio), the QCfail WGA sample exhibited evidence of amplification bias, demonstrated by an altered allelic ratio.

When we compared aCGH performances between the two single-cell WGA products, WGA-QCpass resulted in a notably better derivative log2 ratio spread (DLRS) and signal-to-noise ratio compared to WGA-QCfail (**Figure 44D**). DLRS is a key quality metric for aCGH data measuring the point-to-point consistency or "noisiness" in data, with high values indicating poor signal-to-noise relationship and difficulty in assessing true somatic copy number alteration (SCNA) status. The poor-quality metrics (DLRS: 1.3, signal-to-noise ratio: 3.9) associated with WGA-QCfail renders its aCGH data unfit for meaningful analysis and interpretation. On the other hand, DLRS of 0.72 and signal-to-noise ratio of 53.6 associated with WGA-QCpass meets the quality threshold for single-cell derived aCGH set forth by Agilent Technologies.[157] Gain and loss profiles generated from aCGH data are as shown in **Figure 44**. Comparison between WGA-QCpass and WGA-QCfail reveals disparities in single-cell aCGH profiles, even though clonally expanded HPAF-II cells were used for both. When compared to the unamplified batch gDNA,

we found multiple areas of alterations that were detected with WGA-QCpass, but not with WGA-QCfail (**Figure 44A**, red arrows).

The same two single-cell WGA products were also analyzed by massive multiplex short read sequencing. **Figure 45** shows the quality metrics of sequencing data associated with each WGA-DNA sample and unamplified batch gDNA. When we compare the two single-cell WGA products (WGA-QCpass and WGA-QCfail), Lorenz curves, histograms of read count frequency and indexes of dispersion all indicated a superior NGS data quality with higher coverage uniformity for WGA-QCpass compared to WGA-QCfail. The Lorenz curve provides information on the uniformity of the sequencing reads distribution. Perfect coverage results in a straight line with slope of 1 (y=x). The wider the curve below the line of y=x, the lower the coverage uniformity and greater the amplification bias. As can be seen in **Figure 45A**, WGA-QCpass resulted in more uniform distribution of read depth compared to WGA-QCfail. The histogram of read count frequency also provides information on coverage dispersion (**Figure 45B**). Histograms with a wide range of distribution, as seen in the sequencing data obtained using WGA-QCfail, indicates a greater degree of amplification bias. Genomic profiles generated from NGS sequencing data are as shown in **Figure 45C**. Comparison between WGA-QCpass and WGA-QCfail reveals significantly more "noise" in the genomic profile derived from the latter, with a greater index of dispersion compared to that derived from WGA-QCpass.

Determination of threshold number of cells needed for optimal performances in aCGH and NGS analysis

Although the single-cell WGA product passing our QC criteria (WGA-QCpass) performed well in point mutation detection, aCGH and NGS, room for improvement still existed

when compared to the performance metrics of unamplified batch gDNA (**Figure 44 & Figure 45**). Increasing the number of starting cells has been shown to improve WGA product quality.[149,153,154] However, the number of cells needed for generating a WGA product that faithfully represents the unamplified, original DNA in downstream molecular analysis remains to be defined.

Using our modified MDA protocol, we lowered the numbers of cells needed to reliably pass our QC-step down to 5 cells from 20 cells. However, it remained to be answered how closely 5-cell and 10-cell WGA products approximated the unamplified DNA in aCGH and NGS analysis, and whether obtaining 5 to 10 CTCs from a single liquid biopsy is truly sufficient for sequencing applications within precision oncology strategies. (**Figure 44B** and **Figure 45** illustrate the performances of 5-cell and 10-cell modified MDA products (QC score=8 for both) in aCGH and NGS. As the number of cells used for starting genomic template increased from 1 to 5 to 10 cells, progressive improvement in all quality metrics of aCGH and NGS were noted. Notably, the Lorenz curves for WGA-DNA and unamplified batch gDNA nearly overlapped by 10-cells (**Figure 45A**). Furthermore, by 10-cells, WGA-DNA derived aCGH and NGS genomic profiles appeared highly concordant compared to that of unamplified batch gDNA, with progressive reduction in signal "noise" and index of dispersion as the number of cells used increased (**Figure 44B & Figure 45C**). Using 10 cells as the starting template for our modified MDA reaction generated WGA-DNA that closely approximated the unamplified gDNA in both aCGH and NGS performances.

4.3.5 Discussion

CTC analysis offers unprecedented potentials for furthering precision oncology. Realization of a "liquid biopsy" through CTC sequencing helps avoid invasive and costly traditional biopsy procedures. Moreover, it allows for a dynamic monitoring of evolution in tumor genome in response to cancer therapy.[158] However, most downstream applications (i.e. aCGH, NGS) necessitate a whole genome amplification (WGA) step prior to the analysis of single/few-cell, which is known to introduce errors and biases.[145-148,159] In the current study, we described our strategic approach for selecting and generating optimal WGA products for analysis using aCGH and NGS. We presented and validated a quality control assay for WGA product quality, and introduced a modified MDA protocol that helped improve the reproducibility and reliability of the existing WGA process. Finally, we showed that by combining our QC-criteria, modified MDA protocol and using as little as 5-10 cells as a starting template, a WGA product with high fidelity to the unamplified template DNA could be obtained.

Multiple groups have published on performance comparison of different WGA methods over the recent years, with significant differences in results and conclusions existing between these studies.[148,150,160] However, the majority of these studies lack analysis of the quality of WGA samples being used. Given the known high sample-to-sample variability in single-cell WGA quality, no meaningful interpretation and comparison of WGA-DNA derived data can be made without a well-defined quality standard. Our proposed 8-cancer gene QC assay has the potential to fulfill this gap, as it successfully predicts the performances of WGA-DNA in downstream analysis by aCGH and NGS, in addition to providing limited sequencing information.

Prior reports have described multiplex PCR based QC assay for formalin-fixed paraffin embedded (FFPE) DNA and *Ampli1* WGA-DNA.[155,161] The *Ampli1* WGA is a PCR-based amplification process using non-random primers, which is fundamentally dissimilar to the MDA method of non-PCR based isothermal amplification process using random primers. To date, a QC assay specific to the MDA method remains to be defined. Furthermore, no study to date has described a QC assay designed specifically to evaluate for genes implicated in cancer therapy. The eight genes included in our QC assay are all considered important molecular targets for cancer therapy, and under active investigation in the National Cancer Institute – Molecular Analysis for Therapy CHoice (NCI-MATCH) trial.[162 163-165] Thus, evaluation for WGA-DNA quality using this QC assay helps identify samples suitable for potential use in precision oncology strategies.

Along with the 8-cancer gene QC-assay, we have also developed a modified MDA protocol in order to help reduce the number of cells needed for a reliable WGA reaction. Our modified MDA protocol based on the principle of small volume MDA required only a 96-well plate and negated the need for labor-intensive protocols and costly special equipment described in previous works. Although this meant that we could not achieve the nanoliter reaction volumes described in prior studies, we still observed a meaningful improvement in sample-to-sample reproducibility and overall improvement in our defined QC-scores. Furthermore, we noted reduced amplification reaction gains associated with our modified MDA protocol, consistent with observations made in prior literature on small volume MDA.[150,152] This is important, as the degree of amplification reaction gain has been shown to correlate with WGA product quality, with excessive gains associated with worsening amplification bias.[150] In particular, amplification

bias in MDA was found to be a direct function of increasing reaction gains, with notable deterioration in WGA product quality starting at around gain of $>10^7$.

Lastly, by combining our QC-criteria and the modified MDA protocol, we found that approximately 5-10 cells were needed to generate amplified DNA with aCGH and NGS performance approximating that of batch gDNA. Our current finding is in concordance with prior reports suggesting that 5-10 CTCs are likely needed to achieve reliable point mutation detections and aCGH analysis.[149,161] As the interest for "liquid biopsy" and CTC analysis grows, 5-10 cells may be viewed as the minimal number of CTCs needed for reliable molecular analyses and applications in precision oncology strategies.

4.4 Improved performance of multiple displacement amplification for limited template samples

4.4.1 Abstract

Background: Limited template samples, such as fine needle aspirates or circulating tumor cells, require whole genome amplification for molecular analyses such as array comparative genomic hybridization (array CGH) and next generation sequencing (NGS). Multiple displacement amplification (MDA) is a commonly used amplification method but is hampered by amplification failure and biases that can limit its reliability. In this study, we attempt to address these limitations both by modifying the MDA protocol itself, and by developing a post-MDA technique that can help "rescue" samples that demonstrate suboptimal amplification.

Results: Using cell lines, we tested different modifications to the standard MDA protocol and assessed their performance using an 8-oncogene real time quantitative polymerase chain reaction (RT-qPCR) assay. We found that performing a combination of mechanical and alkaline cell lysis in multiple smaller reaction volumes significantly improved our RT-qPCR results compared to the standard WGA protocol. We then developed a post-MDA debranching protocol to help increase DNA accessibility by resolving the hyperbranched MDA amplification product back to linear DNA. This post-MDA debranching technique improved the RT-qPCR results for WGA DNA amplified using both the standard MDA protocol and our modified MDA protocol. We further investigated the potential utility of MDA-debranching for improving results from both array CGH and NGS assays and found that it made significant improvements in the performance of some samples while never decreasing a samples' performance. **Conclusions:** We provide simple and effective modifications to the standard MDA protocol and demonstrate the potential

for MDA-debranching to serve as an adjunct technique to help rescue amplified samples with poor QC performance or problematic amplification biases.

4.4.2 Introduction

The introduction of massively parallel next generation sequencing (NGS) has revolutionized the field of molecular biology and resulted in breakthroughs in almost all fields of biology.[166] However, the quantity of DNA required for most NGS assays is in the nanogram to microgram range, the equivalent of thousands to millions of cells.[25] For many fields, such as studies utilizing fine needle aspirations or circulating tumor cells (CTCs), often only a single cell is available for analysis.[30] Therefore, techniques are required to bridge the gap from the approximately 4-6 picograms of DNA in a single cell to the nanogram to microgram quantities required for NGS assays.

Whole genome amplification (WGA) is one such technique for generating sufficient DNA from a single cell for NGS assays. Studies of WGA from single cells have demonstrated the ability to achieve 10^3 to 10^9 fold increase in DNA quantity with up to 96% genome coverage.[121] Furthermore, researchers have successfully performed a multitude of NGS assays on single cells using various WGA techniques, ranging from whole exome[167] and whole genome sequencing[133] to somatic copy number alteration (SCNA)[168,169] and epigenetic profiling[170]. While the polymerase chain reaction (PCR) is the most widely used technique for DNA amplification, WGA techniques based on PCR such as primer extension preamplification PCR, degenerate oligonucleotide primed PCR (DOP-PCR), and multiple annealing and looping based amplification cycles (MALBAC) are noted for their high error rate due to the error-prone nature of *Taq* and other PCR-capable DNA polymerases.[115,121,171]. An alternative method known as multiple displacement amplification (MDA) has sought to overcome the limitations of PCR-based methodologies (**Figure 46A**). MDA utilizes Φ29 DNA polymerase from bacteriophage Φ29 of *Bacillus subtilis* and exonuclease-resistant random hexamer primers to perform

isothermal DNA amplification.[172] The single subunit proofreading Φ29 DNA polymerase is able to

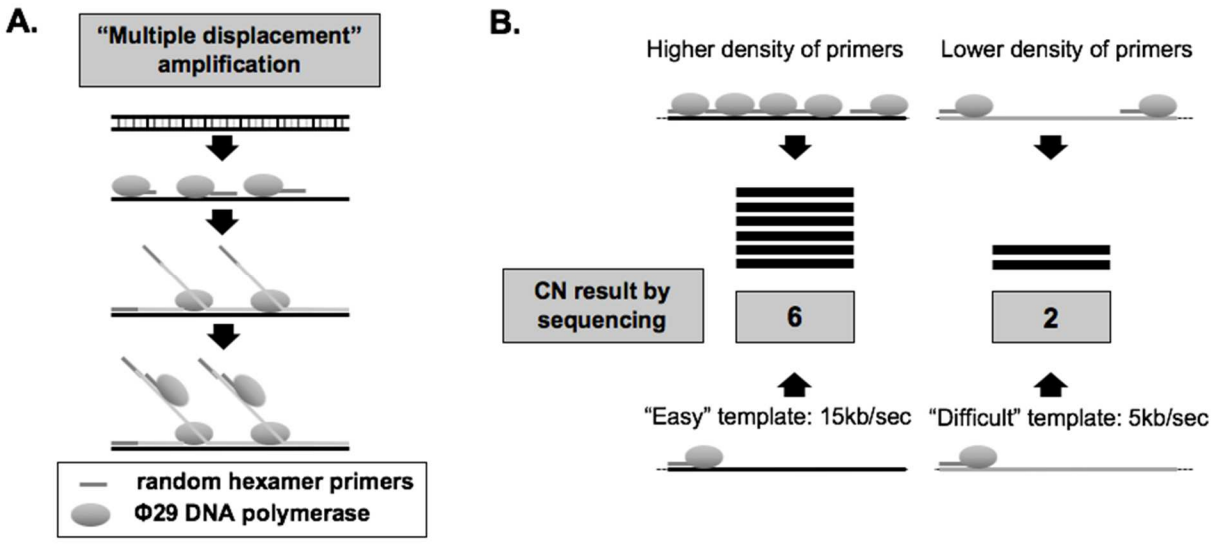


Figure 46. A. Overview of the multiple displacement amplification process (MDA). Random hexamer primers bind throughout the genome and allow amplification by the highly processive Φ29 DNA polymerase. 5' — 3' strand displacement allows for multiple rounds of amplification during a single isothermal reaction. **B.** Reasons for the introduction of amplification biases during MDA include differences in the density of random hexamer primer sites and DNA template accessibility. The resulting copy number (CN) found by sequencing is found to be falsely amplified (CN = 6) for areas with increased primer sites or DNA accessibility when compared to other areas (CN = 2).

incorporate >70,000 nucleotides per binding event and has excellent strand displacement activity.[173] Importantly, due to the intrinsic proofreading activity of Φ29, its error rate is 1×10^{-6} – 10^{-7}, significantly lower than the 2×10^{-4} rate of *Taq DNA* polymerase.[174]

Despite the advantages of Φ29 DNA polymerase, MDA is not a linear amplification method, and the resulting amplification bias is an inherent limitation. While some causes of amplification, such as GC-content, are known and can be accounted for systematically, the majority cannot. As WGA of the 3 billion base pairs of the human genome mandates priming and amplification by millions of polymerases at different sites simultaneously, small differences in random hexamer primer site density or DNA accessibility can result in significant biases at the start of the MDA reaction when relatively few DNA templates are available (**Figure 46B**). These small biases are then exponentially magnified during hyperbranched replication, resulting in

preferential amplification of some areas of the genome in a stochastic manner that cannot be accounted for systematically.[115] MDA has thus traditionally demonstrated much greater variance in amplification bias than other WGA methods like MALBAC or DOP-PCR for applications like somatic copy number alteration (SCNA) detection, with some samples performing adequately while others do not.[175] Additionally, MDA is known to generate chimeric DNA rearrangements in the amplified DNA due to the hyperbranched structure that develops during amplification.[176-178] Both amplification bias and chimeric DNA rearrangements limit the potential of MDA for single-cell NGS applications like SCNA detection, forcing researchers and clinicians to choose a WGA method based on what molecular assays are suspected to be most important for a given sample. For precious samples such as fine needle aspirate biopsies or archival material, for which repeat WGA is not possible, this presents a difficult dilemma. Thus, methods to improve MDA-based WGA for applications that have traditionally been limited by MDA's amplification bias are an area of active research.[171,178-180]

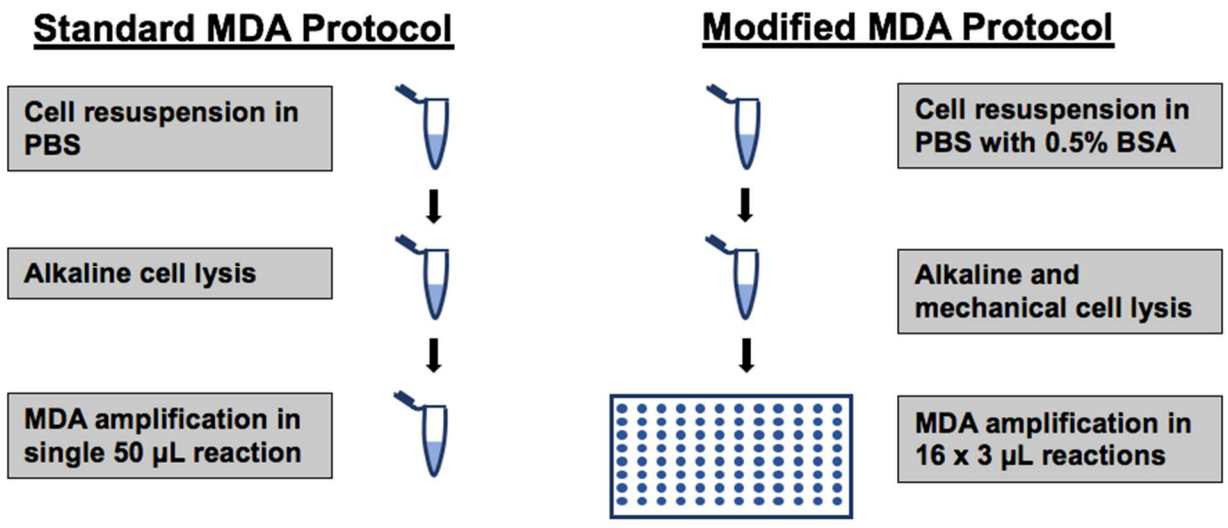


Figure 47. Overview of the 3 changes to the standard MDA protocol used in the modified MDA protocol.

In this study, we sought to develop a method for improving the results of sequencing limited template (defined as single or small numbers of cells) samples amplified using MDA. We did so by investigating 2 interrelated methods. The first was modifying the MDA protocol itself in 3 simple ways (**Figure 47**). First, by including bovine serum albumin (BSA) in the resuspension solution to assist with resuspension of the cell from the cap and to help improve MDA efficacy by neutralizing amplification inhibitors.[181,182] Second, including a mechanical freeze-thaw cycle to the standard alkaline cell lysis to improve cell lysis and DNA release. Third, we decreased the reaction volume used for WGA to 3 μL from 50 μL in order to limit amplification bias. Prior studies utilizing microfluidic and droplet-based techniques have demonstrated significant decreases in amplification bias, but require the use of specialized and expensive equipment.[121,154,183] In addition to modifying the MDA protocol, we modified a debranching and DNA repair method to transform the hyperbranched MDA-amplified DNA into linear double-stranded DNA in an effort to overcome both amplification bias and the chimeric DNA rearrangements introduced during MDA (**Figure 48**).[176] We then compared the performance of standard MDA to both the modified MDA-based WGA protocol and post-MDA debranching technique for multiple molecular analyses including qPCR, array CGH, and NGS sequencing.

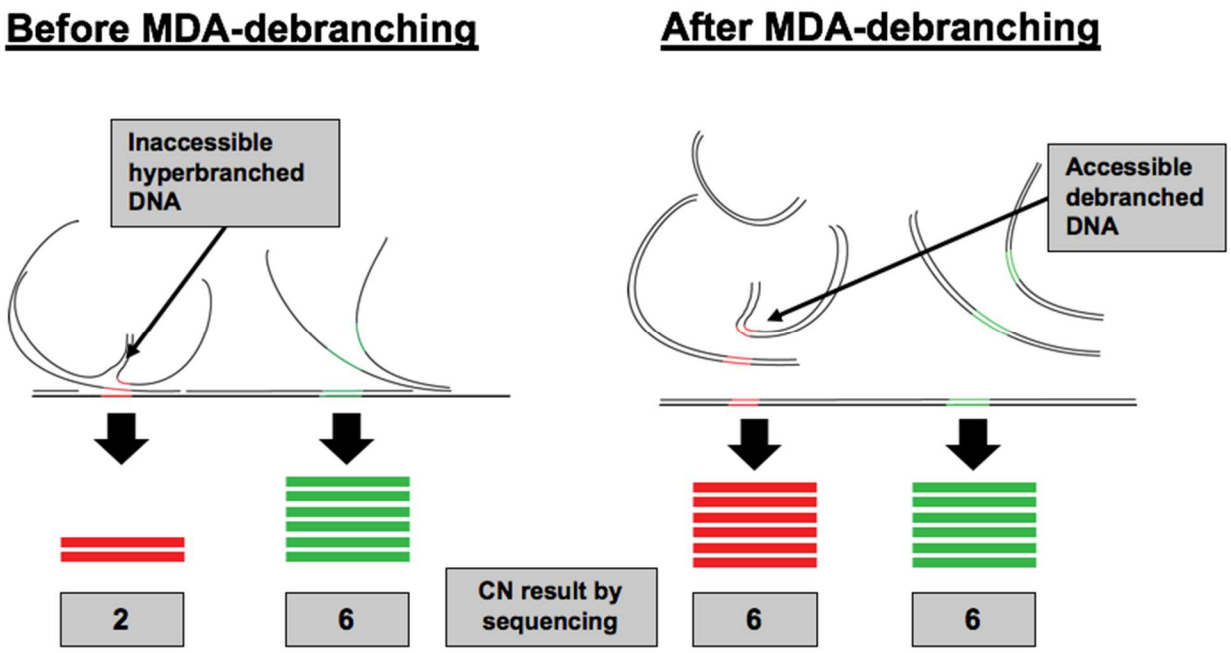


Figure 48. DNA branch points introduced during MDA are complex and more difficult to access during downstream molecular analysis, resulting in differences in the resulting copy number found between branch points (CN = 2, red) and other areas (CN = 6, green). By debranching the DNA prior to analysis, the bias between regions can be resolved.

4.4.3 Methods

Cell culture and DNA isolation

Pancreatic cancer (PDAC) cell lines HPAF-II and AsPC-1, as well as hepatocellular carcinoma (HCC) cell line Huh-7 were obtained from American Type Culture Collection (ATCC, Manassas, VA). All cell lines were grown using RPMI 1640 Medium (ATCC) supplemented with 10% fetal bovine serum and 100 U/mL penicillin-streptomycin (Thermo Fisher Scientific, Waltham, MA). All cell lines were grown at 37°C with 5% CO_2 and were routinely passaged at 80% confluence using an iso-osmotic sodium citrate solution for cell release. When preparing cells for single cell isolation and analysis, cells were released from the culture plates using the same sodium citrate solution. Following a wash with the corresponding culture medium, each cell line was diluted to a density of 1000 cells per 100 µL and 100 µL were smeared on PEN membrane slides (Leica, Wetzlar, Germany), air-dried for 10 minutes, and

fixed with 100 µL of 100% Ethanol. Cells were then isolated by laser microdissection as outlined below in the section **Laser micro-dissection**. Batch DNA was prepared by releasing all the cells on a 75 mm^2 plate and extracting the DNA using the Qiagen Blood and Cell Culture DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

Ethics, consent and permissions

Patients were approached in our multidisciplinary liver cancer clinic and consented for enrollment as part of our prospective studies of CTCs in gastrointestinal cancers.[184,185] The study was approved by the University of California, Los Angeles Institutional Review Board (IRB #14-001932 and #11-002112).

Patient recruitment and sample processing

Following discard of 5 mL of venous blood to prevent epithelial contamination, 10 mL of venous blood was collected into anticoagulant citrate dextrose (ACD) solution A tubes (BD Pharmigen, Franklin Lakes, NJ), and stored at 4°C until processed. All samples were processed within 24 hours of collection. Blood samples were diluted 1:1 with phosphate-buffered saline (PBS, Sigma, St. Louis, MO) and transferred to 12 mL LeucoSep centrifuge tubes (VWR, Radnor, PA) with 3mL of Histopaque-1077 (Sigma) below the porous barrier. Samples were centrifuged at 400 *x g* at 4°C for 30 minutes with the break off. The buffy coat layer was then transferred to a new 15 mL tube and washed with 5 mL of wash medium (RPMI with 5% FBS, Gibco, Carlsbad, CA), centrifuged at 300 *x g* for 10 minutes at 4°C, and the pellet resuspended in 200 µL of wash medium. 6 µL of biotinylated capture antibody cocktail (either epithelial cell adhesion molecule (EpCAM) alone for PDAC or EpCAM, asialoglycoprotein receptor, and

glypican-3 for HCC) was added and the mixture was incubated on a shaker for 30 minutes at room temperature. Following a final wash step, the pellet was re-suspended in 200 μL of PBS and immediately processed on the NanoVelcro platform.

P-NanoVelcro CTC chip processing, immunocytochemistry, and chip scanning

The assembly, operation, and staining of P-NanoVelcro CTC chips has previously been described and is detailed in the supplemental methods.[66] Briefly, poly(lactic-co-glycolic acid) (PLGA) nano-spun chips, manufactured via an electro-spin method in our nano-materials lab, were assembled onto a laser micro-dissection slide (Leica) with an overlaid polydimethylsiloxane (PDMS) microfluidic component and attached to a syringe based microfluidic pump (KD Scientific, Holliston, MA). Chips are stained as described previously and CTCs identified via scanning fluorescent microscopy on a Nikon Eclipse 90i using immunocytochemistry (ICC) and NIS Elements 4.1 software. For ICC, WBCs were defined as Nuc+/CD45+/CK-, and CTCs as Nuc+/CD45-/CK+.

Laser micro-dissection

CTC chips and cell line slides were transferred to a PALM MicroBeam laser microdissector (Zeiss, Oberkochen, Germany). After locating cells, single cells or groups of cells were isolated using the laser catapult function. Cell transfer to the opaque cap Eppendorf tubes (Zeiss) was confirmed by visualization of the cell in the cap using the CapCheck function.

Whole Genome Amplification

Isolated cells were re-suspended in 4 µL of PBS with or without 0.5% molecular grade BSA (Sigma, St. Louis, MO) and then subjected to a freeze-thaw cycle at -80°C for 30 minutes followed by warming to room temperature depending on the protocol used (**Figure 47**). All cells were then lysed using the REPLI-g Single Cell Kit (Qiagen) using the manufacturer's recommended protocol. MDA-based WGA was performed using the manufacturer's recommended protocol with the modification that prior to isothermal amplification, the 50 µL MDA reaction mixture was mixed for 30 seconds by pipetting and then divided into 16 individual reactions (approximately 3 µL per well). Following the amplification step, contents from all wells were collected into a single tube and purified using a 1.8:1 ratio of AMPure XP beads (Beckman Coulter, Brea, CA) using the manufacturer's recommended protocol resulting in 30 µL of purified WGA product.

MDA-debranching treatment

WGA products then underwent branch reduction. The protocol is based on the work of Zhang et al for polymerase cloning but modified for human DNA sequencing.[176] 15 µL of WGA product was added to a mix containing 2.5 µL of 10x Φ29 Reaction Buffer (NEB, Ipswich, MA), 1 µL Φ29 DNA polymerase (NEB), 2.5 µL of 10 mM stock deoxyribonucleotide triphosphates (dNTPs, Sigma), and 4 µL molecular grade H_2O on ice. The mix was incubated for 2 hours at 30°C followed by heat inactivation of the Φ29 DNA polymerase at 65°C for 10 minutes. Following debranching, a mixture of 15 µL 5x S1 Nuclease Reaction Buffer (Sigma), 1 µL of 100 U / µL S1 Nuclease (Sigma), and 34 µL H_2O were added. The resulting 75 µL reaction was incubated at 37°C for 30 minutes followed by S1 nuclease inactivation at 70°C for 10 minutes.

S1-treated WGA products underwent column cleanup using the QIAquick spin columns (Qiagen) with elution in 34 µL of H_2O. Next, a mix of 4 µL T4 DNA Ligase Reaction Buffer (NEB) and 2 µL T4 DNA Ligase was added, and the resulting mixture incubated for 2 hours at room temperature. Following heat inactivation at 65°C for 10 minutes, a final QIAquick column (Qiagen) cleanup was performed and DNA quantified by Nanodrop (Thermo).

Table 10. Primers used for the 8-oncogene qPCR assay

Gene	Forward primer	Reverse Primer
BRAF	5′ – TAC TGC TCT TTC TTC TCC AAC AC – 3′	5′ – CCT GAT TGT ATT TGA GAT CTA GTA GGG – 3′
EGFR	5′-CAG CCT TCT CCG TAA TTA GCA T – 3′	5′ – TGA CAC AGA TAA TTG TCC CAC AG – 3′
KIT	5′ – GGC ATT GAG GAG GGA TAG TAA AT – 3′	5′ – CTG AAC AAT TTG CTT GAA TGT TGG – 3′
KRAS	5′ – GTG TTA CTT ACC TGT CTT GTC TTTG – 3′	5′ – GCC TTC TAG AAC AGT AGA CAC AA – 3′
NRAS	5′ – AAT GGA ATC CCG TAA CTC TTG G – 3′	5′ – GAT GAT GTA CCT ATG GTG CTA GTG – 3′
PIK3CA	5′ – AGG GCA AAT AAT AGT GGT GAT CT – 3′	5′ – CAG CAA TTA CTT GTT CTG GTA CAC – 3′
PTEN	5′ – CTT TCT CTA GGT GAA GCT GTA CT – 3′	5′ – GGT TCA TTG TCA CTA ACA TCT GG – 3′
TP53	5′ – AAG AGA AGC AAG AGG CAG TAA G – 3′	5′ – CTT AGG CTC CAG AAA GGA CAA G – 3′

Multiplex real-time quantitative PCR (RT-qPCR)

WGA DNA was diluted to a concentration of 200 ng/µL as measured by optical density at 260/280nm (Nanodrop 2000, Thermo) and analyzed by multiplex PCR using 8 primer sets representing frequently mutated cancer genes spread across the genome (**Table 10**). Pre-amplification PCR reactions were carried out on a C1000 Thermal Cycler (Bio-Rad) with the Multiplex PCR Plus Kit (Qiagen) using total volumes of 17 µL per reaction. The reaction

conditions were as follows: denaturation at 94ºC for 30 seconds, annealing at 64ºC for 30 seconds, and extension at 72ºC for 30 seconds for a total of 10 cycles. Following the pre-amplification step, the resulting amplified products were detected using real time quantitative PCR (RT-qPCR). Reactions took place on BioRad CFX-96 real time system (BioRad, Hercules, CA) using the QuantiTect SYBR Green PCR Kit (Qiagen). Reaction mixtures were prepared containing 12.5 µL of the SYBR Green PCR Master Mix (Qiagen), 9.5 µL of RNA-grade water, 1 µL of individual primers sets (10uM) and 2 µL of the multiplex pre-amplification PCR product. The reaction conditions were as follows: 95 ºC for 15 minutes, followed by 35 cycles of 94 ºC for 15 seconds, 64 ºC for 20 seconds, and 72 ºC for 20 seconds. The plate was read following the extension step at 72 ºC. Melting curve analysis was performed between 70-95 ºC at 0.5 ºC intervals. Real-time PCR data were reviewed and analyzed using the CFX manager (BioRad). Specificity of the PCR amplification product was determined using melting curve analysis. PCR products with melting-temperature (Tm) matching the expected value based on primer sequence, and threshold cycle (Ct) <30 were counted as reliable amplification and detection.

Array-CGH

Sample WGA-DNA and reference DNA were differentially labeled with cyanine-3 (Cy3) and cyanine-5 (Cy5) dyes using the GenetiSure Amplification and Labeling Kit (Agilent, Santa Clara, CA) according to the manufacturer's protocol and as described in the supplemental methods. Purified labeled DNA samples were prepared for hybridization, which took place on Agilent 8x60K CGH microarray slides at 67 ºC for 6 hours. Following the hybridization, the slides were scanned using the Agilent SureScan Microarray Scanner (Agilent). Microarray

images were analyzed using the Agilent CytoGenomics software (Agilent) and the Microarray text files were analyzed using R (R Foundation, Vienna, Austria) version 3.3.2 and the packages rCGH, limma, agilp, and snapCGH.

Sequencing library preparation and sequencing

Purified WGA products were sheared to generate DNA fragments averaging 350bps by sonication (Covaris, Woburn, MA). Sonicated DNA was cleaned, end-repaired, ligated, and amplified using the KAPA DNA Library Preparation Kit (KAPA Biosystems, Wilmington, MA) according to the manufacturer's protocol. Sequencing was performed on an Illumina NextSeq 500 (Illumina, San Diego, CA) using 75 bp paired end reads (2 x 75 bp).

Data Analysis

Fastq sequences were filtered and trimmed using Trimmomatic [108] and alignment was performed using bowtie2 [109] with settings "—very-sensitive-local." Coverage statistics and BAM to BED file conversion was performed using Samtools and Bedtools.[110] Whole exome sequencing data was analyzed using Samtools stats function. Sequencing statistics were normalized by the raw sequencing count except where noted. To avoid conflating read errors with chimeras, chimera rate was assessed as only the portion of reads that had pairs mapped to a different chromosome divided by the total number of mapped reads. Copy number variation and single cell NGS quality scores were analyzed by Ginkgo using a variable bin size of 100 kbps and simulating bins of 76 bp reads mapped with bowtie.[111] Segmentation was performed using normalized read counts, and sex chromosomes were masked for mixed-gender analyses.

Continuous variables were summarized as medians and interquartile ranges (IQR) and categorical variables were summarized as frequencies and percentages. A p-value of < 0.05 was used for significance throughout the manuscript. qPCR results were compared using nonparametric tests, the Wilcoxon-Pratt signed rank Z test for 2 sample comparisons and the Kruskal-Wallis test for more than 2 samples. The effect of MDA-debranching on sequencing and array CGH QC metrics was assessed using paired t-tests.

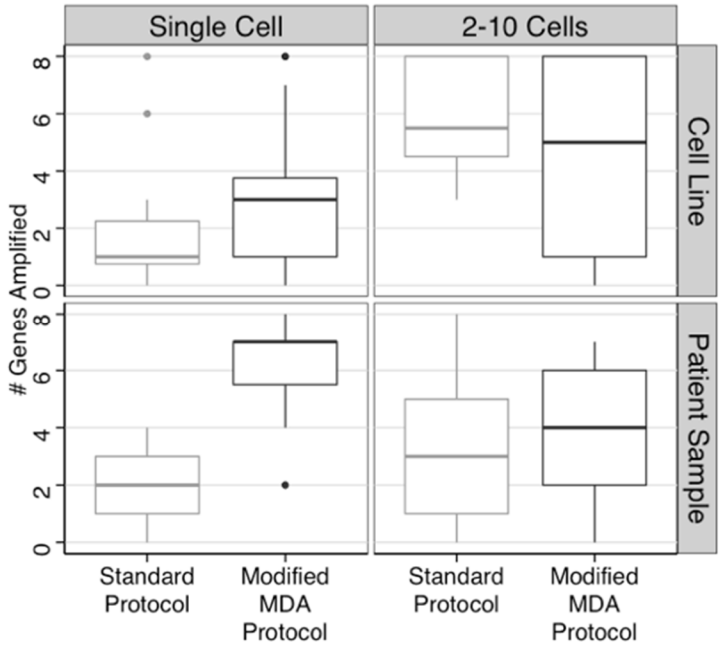


Figure 49. Comparison of the performance of the standard and modified MDA protocol for the 8 gene RT-qPCR assay. The results are separated between cell line and patient samples (CTC samples) and between single and small groups of cells. The improvement found for the modified MDA protocol was most pronounced for single cell samples.

4.4.4 Results:

Design of Experiments

A total of 93 samples were included in the experiment, 59 from cell lines and 34 from patient samples (**Table 11**). Of the samples, 59 (63%) were single cells, 11 (12%) included

between 2-4 cells, and 23 (25%) included between 5-10 cells. We compared the efficacy of our modified MDA protocol to that of the standard MDA protocol using a multiplex RT-qPCR assay for 8 important cancer-related genes on different chromosomes.[186] Multiplex RT-qPCR has previously been used as a quality control step prior to NGS applications as it is representative of both the quality and accessibility of the DNA. Using the standard WGA protocol, we were able to amplify an average of 3.0/8 genes (IQR: 1 - 5) versus 3.7/8 genes (IQR: 2 – 6) with our modified WGA protocol, a significant increase (Wilcoxon-Pratt Signed-Rank Test, $Z = 6.47$, $p < 0.0001$). The results were most pronounced for single cell patient samples (**Figure 49**).

Table 11. Protocols used for the 93 samples that underwent WGA in the experiment. Numbers in parentheses are the number of samples from each group that subsequently underwent MDA debranching.

MDA protocol used	Total Samples (debranch treated)	Cell Line Samples (debranch treated)	CTC Samples (debranch treated)
Standard Protocol	38 (29)	20 (11)	18 (18)
Modified MDA Protocol	55 (49)	39 (33)	16 (16)

MDA-Debranching

In addition to the modifications to the WGA reaction above, we developed a post-MDA debranching protocol to resolve the hyper-branched structure resulting from MDA-based amplification (**Figure 48**). We first assessed the efficacy of MDA-debranching by comparing the results of our multiplex RT-qPCR assay before and after treatment for the 78 samples with sufficient DNA available for MDA-debranching treatment. Prior to debranching, the samples averaged 3/8 genes amplified (IQR: 1-5). Following MDA-debranching, the samples averaged 5/8 genes amplified (IQR: 2-7), significantly more than the pre-treatment samples (Wilcoxon-Pratt Signed-Rank Test, $Z = 7.73$, $p < 0.0001$).

We then investigated the potential of combining the WGA modifications with the MDA-debranching as a means of optimizing a WGA protocol. As seen in **Figure 50,** we further found that while both the modified MDA protocol and MDA-debranching resulted in improved WGA performance, the combination of both modifications together results in significantly improved RT-qPCR results (Kruskal-Wallis rank sum test, $\chi^2 = 18.31$, df = 3, p = 0.0004).

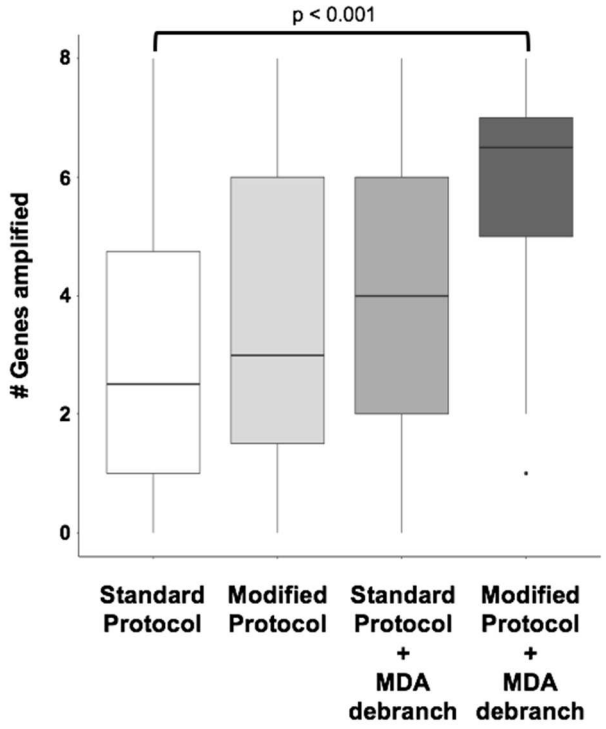


Figure 50. Comparison of RT-qPCR assay performance for samples undergoing amplification with the standard MDA protocol versus modified MDA protocol both with and without MDA-debranching. The combination of both modified MDA and MDA-debranching resulted in significantly higher number of genes amplified by RT-qPCR (Kruskal-Wallis rank sum test, $\chi 2 = 18.31$, df = 3, p = 0.0004)

Performance of MDA-Debranching for Circulating Tumor Cell Molecular Analyses.

To test if the improvement in qPCR results for MDA-debranching hold true for other molecular analyses with real patient samples, we tested its utility on 15 circulating tumor cell (CTC) samples from 13 patients with liver cancer. Patients had a diagnosis of either hepatocellular carcinoma (n = 10) or intrahepatic cholangiocarcinoma (n = 3), and the CTC samples contained between 1 and 7 circulating tumor cells each (Error! eference source not found.).

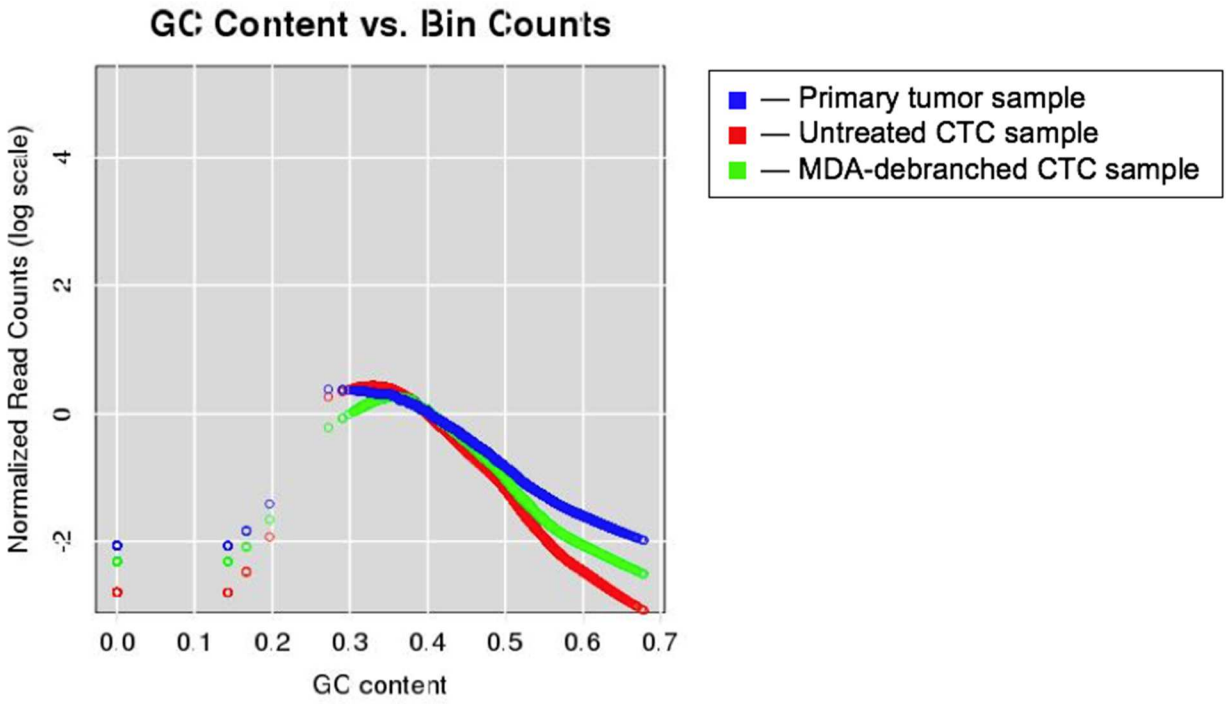


Figure 51. Comparison of NGS WGA quality control metrics for bulk sequencing from the primary tumor, an untreated MDA-amplified CTC sample, and the same sample after MDA-debranching from a single patient: Lowess fit of GC content across all normalized bin counts.

Copy Number Variation Analysis

In order to assess the effects of our modified WGA protocol for copy number analysis, we compared the results of pre and post MDA-debranching samples for both NGS-based SCNA detection and array CGH. For NGS analysis we used a low-resolution whole genome SCNA profiling method. Average sequencing depth was 0.014x with an average of 907960 reads per sample (IQR: 825949 – 1057806; range: 583040 – 1905465). Overall, 93.74% of reads were successfully mapped to the reference genome.

The effect of MDA-debranching was assessed for the quality control metrics utilized by the Ginkgo single cell sequencing pipeline.[111] The index of dispersion (IOD), the ratio between the read-coverage variance and the mean coverage of the genome, is a measure of the degree of

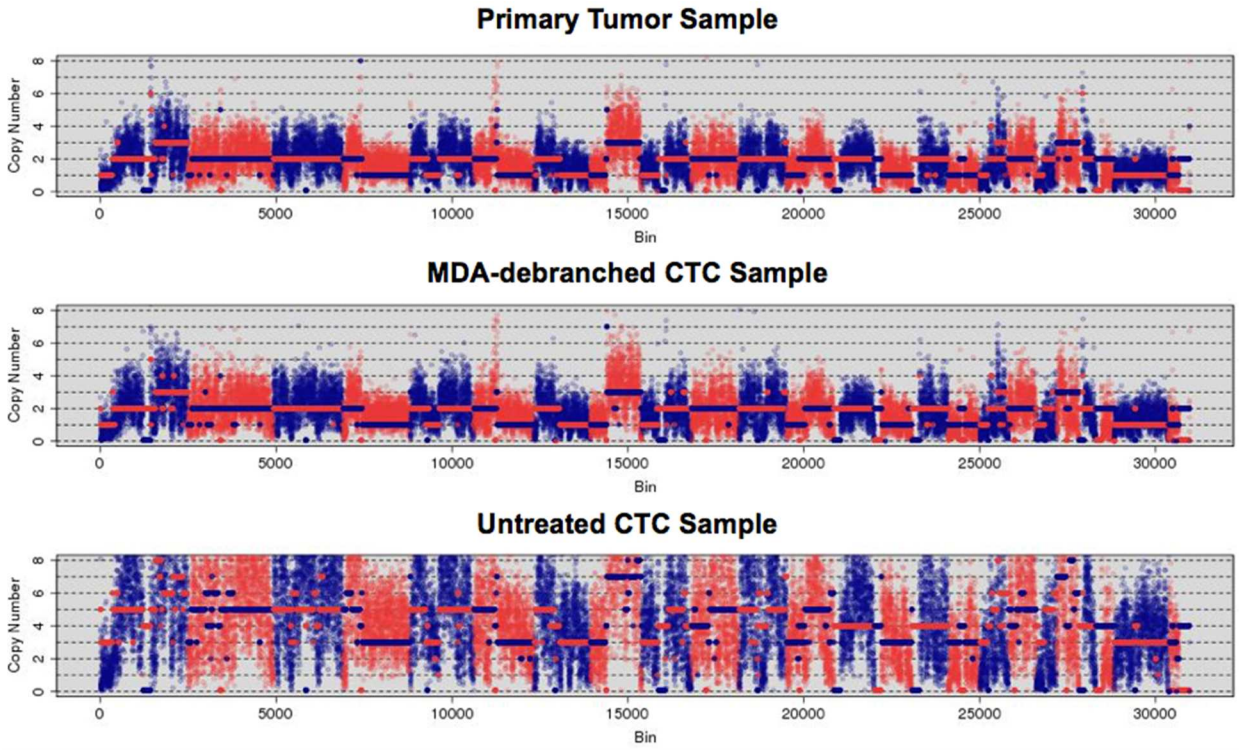


Figure 52. Comparison of NGS WGA quality control metrics for bulk sequencing from the primary tumor, an untreated MDA-amplified CTC sample, and the same sample after MDA-debranching from a single patient: Integer copy number at each 100 kbp bin across the whole genome.

noise in the SCNA data. MDA-debranched DNA had significantly lower IOD values compared to pretreatment DNA (paired t-test, $t = 2.27$, $p = 0.041$) GC-content bias was notably decreased by debranching for some but not all samples (**Figure 51**). Comparison of the SCNA profiles pre and post MDA-debranching clearly demonstrated the reduction in the dispersion (**Figure 52**).

A total of 4 samples had sufficient DNA to allow for array CGH comparison testing. We analyzed the performance of MDA-debranching on two standard array CGH quality control metrics (**Figure 53A**). The first is the derivative log ratio standard deviation (DLRSD), a measure of the degree of probe-to-probe noise. MDA-debranching significantly improved DLRSD values overall ($p = 0.02$), with all 4 samples demonstrating improvement after treatment. The signal to noise ratio was also higher after MDA-debranching; however, it was not statistically significant (**Figure 53B**).

Table 12. Comparison of WES quality control metrics for CTC samples before and after MDA-debranching.

Metric	Untreated DNA	MDA-debranched DNA	p-value
Mapped reads (%)	99.7	99.6	0.09
Properly paired reads (%)	90.6	94.1	0.0026
Failed reads (MQ0, %)	5.2	6.4	0.002
Base Mismatch (%)	3	2.6	<0.0001
Chimera rate (%)	2.5	1.2	0.035

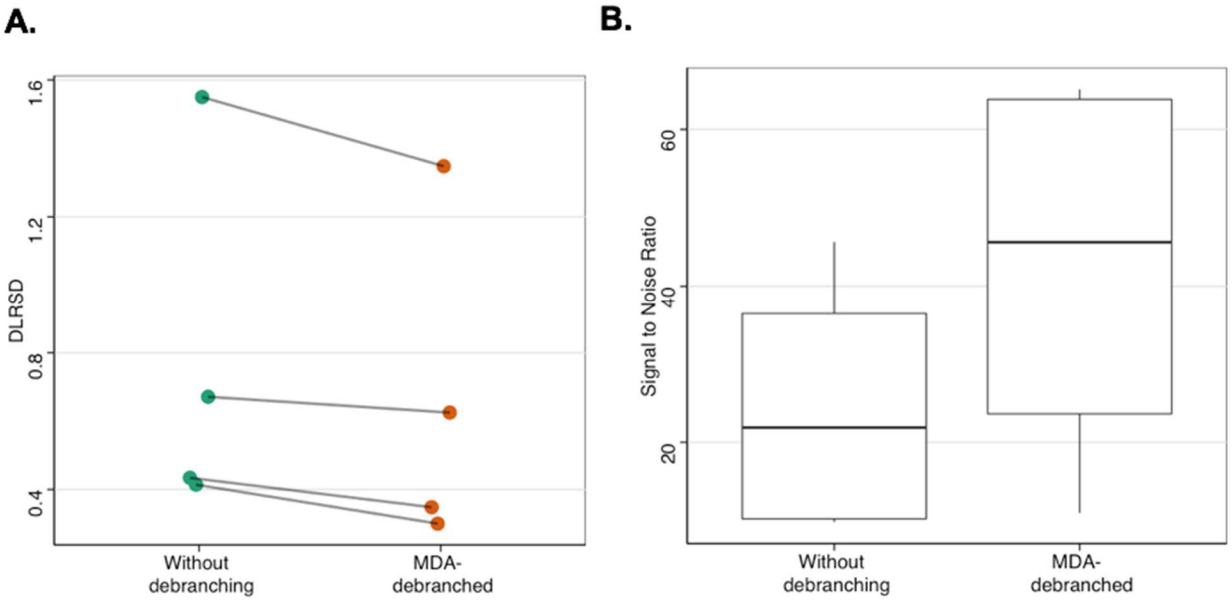


Figure 53. Comparison of arrayCGH quality control metrics for paired pre- and post- MDA-debranching samples (n = 4). **A.** Derivative log ratio standard deviation, a measure of the degree of probe-to-probe noise. **B.** Signal to noise ratio.

Table 13. Overview of the tumor of origin, number of cells present, and molecular analyses performed for all circulating tumor cell patient samples.

Sample	**Tumor type**	**Number of CTCs**	**NGS SCNA**	**Array CGH**	**WES**
H118-1	CCA	6	Y	N	Y
H160-1	CCA	6	Y	N	N
H165-1	HCC	1	Y	N	Y
H167-1	HCC	3	Y	N	Y
H172-1	HCC	4	Y	N	Y
H175-1	HCC	5	Y	N	Y
H185-1	HCC	7	Y	Y	Y
H187-1	HCC	4	Y	Y	Y
H195-1	HCC	3	Y	Y	Y
H195-2	HCC	4	N	N	Y
H195-3	HCC	3	N	N	N
H199-1	HCC	3	Y	Y	Y
H199-2	HCC	3	N	N	N
H200-1	HCC	4	N	N	Y
H201-1	CCA	3	N	N	Y

Whole Exome Sequencing

We performed whole exome sequencing (WES) on 11 CTC samples from 11 of the 13 patients with liver cancer Error! Reference source not found.). A median of 75 million reads per ample was obtained (IQR: 67 – 93 million reads; range: 51 million – 112 million reads). There was no difference in mapping rate between treated and untreated DNA (**Table 12**); however, MDA-debranched DNA had a higher percentage of properly paired reads (94.1% versus 90.6%

for untreated DNA, $p = 0.0026$). The base mismatch rate was also higher in untreated DNA (3.0%) versus MDA-debranched DNA (2.6%, $p < 0.0001$). We estimated the rate of chimera formation during MDA

by looking at the number of paired end reads that mapped to different chromosomes. More pairs mapping to different chromosomes (2.5% chimera rate) were found then would be expected from sequencing genomic DNA, likely due the hyperbranched nature of the amplified DNA. MDA-debranching resulted in significantly less pair discordance (1.2% chimera rate, $p = 0.035$).

4.4.5 Discussion:

Genetic analysis of single cell and limited template samples has numerous potential applications in cancer research and clinical practice, but the current limitations of WGA methods must first be addressed.[120] For Φ29-based multiple displacement amplification these limitations are primarily amplification failures, introduction of chimeric sequences, and preferential amplification caused by the exponential nature of MDA. In this study, we present a novel method for performing MDA-based WGA which demonstrates improved performance for multiple different downstream molecular analyses. Our method relied on 4 modifications to the standard MDA protocol: 1) improving resuspension and limiting DNA damage from DNases by the addition of BSA to the cell resuspension medium, 2) Improving cell lysis effectiveness by adding a mechanical freeze-thaw cycle to the standard alkaline lysis step, 3) Limiting the degree of exponential amplification that a given amplicon can undergo by performing MDA in multiple 3 μL reactions instead of a single 50 μL reaction, and 4) improving DNA accessibility and quality for downstream molecular analysis by debranching the hyperbranched MDA product.

Previous studies comparing MDA-based WGA to other WGA methods consistently demonstrated that MDA allows for the highest degree of amplification with the lowest error rate and highest genome coverage.[25,115,121,175,187-189] Furthermore, the length of MDA amplicons allows amplification of genomic regions with long repeats or difficult templates not easily covered by methods relying on less processive DNA polymerases.[179] However, MDA has also demonstrated greater variability in coverage/amplification biases and chimeric sequence formation when compared to other WGA methods.[111,178,188] While some of these biases, such as GC-content, can be accounted for systematically, biases introduced by hexamer priming distribution, replication speed, or chimera formation are poorly characterized and random in nature.[115] Thus, either multiple samples must be analyzed to overcome these stochastic effects, which is not possible for the majority of limited template applications, or improvements to the standard MDA protocol are required to address these shortcomings.

We consistently found an improvement in the results of molecular analyses from DNA amplified using our modified MDA protocol when compared with the standard one. While our qPCR assay is primarily for quality control purposes, we anticipate that similar improvements would be found for other PCR-based clinical assays such as hotspot mutation detection[190] or multiplex PCR characterization of bacterial pathogens.[191]

An important advantage of our MDA-debranching method is that it is used on the post-amplification DNA and can thus potentially improve the quality of previously amplified DNA. In our study, MDA-debranching improved the qPCR results from DNA amplified using both the standard and modified WGA protocols (**Figure 50**). Similar improvements were found when MDA-debranching was applied to array CGH and NGS analyses of CTC samples from liver cancer patients. Overall, we found that the improvements provided by MDA-debranching for

array CGH and NGS were more modest than those found for our qPCR-based assay. This may be due to the sonication that occurs prior to array CGH and NGS, which has a similar but less specific effect on improving the DNA accessibility of the WGA product.

For array CGH, all post-debranching samples had lower DLRSD than untreated WGA samples, which was statistically significant ($p = 0.02$). While the signal to noise ratio of the array CGH results improved dramatically for 2 post-debranching samples, it was unchanged in one and decreased slightly for the final sample but the overall improvement was not significant. For both low-resolution copy number profiling and whole exome sequencing, post-debranching samples again showed improvements in key quality control and analytical measures. Importantly, debranching treatment never resulted in a decrease in commonly used sequencing QC metrics for any of the 15 CTC samples. While many of the samples showed no or modest improvements, several samples demonstrated dramatic improvements in multiple QC metrics. We hypothesize that this is to be expected given the random nature of the sequencing biases we are trying to correct by debranching. Thus, we found that while not all samples were improved by MDA-debranching, no samples were compromised either.

4.4 Conclusions

In conclusion, we presented a strategic workflow for obtaining and selecting optimal WGA products for aCGH and NGS from minimal template samples. We showed the utility of our 8-cancer gene QC assay in selecting high-quality WGA samples suitable for further molecular analysis, and demonstrated ways to improve the reproducibility and overall quality of MDA reactions. Lastly, we showed that a threshold number of 5-10 cells are likely needed to reliably and accurately represent of the original template genome using aCGH and NGS. These findings contribute to the much-needed quality control criteria for minimal template WGA reactions as we explore the potential of CTCs to provide tumor specific sequence data for precision oncology strategies.

Using our quality control criteria, we then developed a modified MDA-based WGA protocol and post-MDA debranching technique that together demonstrate improved performance for multiple downstream molecular analyses including qPCR, array CGH, and NGS. MDA has emerged as an important WGA method for both research and clinical applications of single cell genetic analysis,[192] but the wide variation in amplification bias, resulting in some samples demonstrating uniform amplification while others show irresolvable biases, limits its utility for applications like SCNA detection. Our post-MDA debranching technique demonstrated one potential technique for overcoming some of these amplification biases. Importantly, post-MDA debranching never worsened the results from either array CGH or NGS studies, and dramatically improved the accuracy and uniformity of several samples. As post-MDA debranching is an adjunct technique, it offers particular promise as a potential "rescue" method for WGA products of precious samples that demonstrate significant MDA-induced amplification bias for which additional cells are not available.

Chapter 5 CTC Sequencing as a Biomarker in HCC

5.1 Abstract

Background: Somatic copy number alterations (SCNAs) are important genetic drivers of many cancers and are increasingly recognized as targetable somatic changes. We investigated the feasibility of obtaining SCNA profiles from circulating tumor cells (CTCs) as a molecular liquid biopsy for hepatocellular carcinoma (HCC).

Methods: CTCs were isolated and low-resolution whole genome sequencing performed to establish SCNA profiles. Primary tumor, peritumoral normal liver, and germline genomic DNA was sequenced for comparison. The Cancer Genome Atlas (TCGA) data was used to develop cancer-specific SCNA profiles.

Results: Sequencing of 18 CTC samples (median 4 CTCs per sample) from 10 HCC patients using a low-resolution whole genome sequencing strategy (median 0.88 million reads per sample) revealed frequent copy number changes in previously reported HCC regions such as 8q amplifications and 17p deletions. Analysis of SCNA profiles for primary tumor and CTC samples revealed that CTCs share a median of 80% concordance with the primary tumor and demonstrate the majority of important somatic SCNAs found in the primary tumor. Sequencing of multiple CTC samples from a single patient demonstrated the reproducibility of the assay. CTCs had SCNAs not seen in the primary tumor, some of which have prognostic implications. Furthermore, specific SCNA patterns were found for certain tumor types in the TCGA database, and were found to correctly classify the tumor type for both HCC and lung cancer CTCs.

Conclusion: SCNA profiling of HCC-CTCs is feasible and accurately recapitulates SCNAs seen in the primary tumor. The use of CTC-derived SCNA profiles demonstrates potential utility and should be explored further.

5.2 Introduction

Somatic copy number alterations (SCNAs) are found in 90% of solid tumors and are increasingly recognized as playing a vital role in activating oncogenes and inactivating tumor suppressors through changes in gene dosage and structure.[193] Newer sequencing methods and large-scale genetic studies indicate that SCNAs affect a larger fraction of the genome in cancer than any other somatic alteration.[193] Similarly, a recent paper found that SCNAs provided the largest contribution to a pan-cancer tumor classification model, greater than that provided by transcriptome and methylome alterations.[194] Research has also found that larger SCNAs, such as whole chromosome, arm, and cytoband length events, are likely more important than focal SCNAs in the development of cancer.[193,195] These larger SCNAs are easily detectable using next generation sequencing (NGS) techniques, and result in a very robust and reproducible signal.[10] These favorable characteristics make NGS-based SCNA profiling an ideal molecular study for limited template samples like fine needle aspirates (FNAs) or circulating tumor cells (CTCs).

Circulating tumor cells (CTCs), cells of tumor origin that circulate in the blood, are a promising new biomarker for many solid tumors.[30,105] As potential metastatic precursors,[196] they are thought to represent the subclones of the primary tumor that are more invasive and metastatic.[197,198] Advances in CTC isolation and sequencing has evolved to the point that CTCs may soon serve as a form of "liquid biopsy" for cancer patients.[197]

The majority of studies on the molecular characterization of CTCs in solid tumors have focused on the detection of actionable somatic point mutations through targeted or exome sequencing. However, due to the combination of cost, analysis time, and risk of false positives at typical exome sequencing depths, exome sequencing of CTCs has not seen widespread usage. In contrast, SCNA profiling of CTCs via low resolution NGS-based whole genome sequencing

(WGS) has recently been described and has the advantage of offering a robust signal at a significantly lower cost than exome sequencing.[10] Given the recent advances in our understanding of the importance of SCNAs for cancer prognosis and treatment, SCNA analysis of CTCs has significant potential as a biomarker. The few studies available to date have demonstrated that CTC SCNA analysis has potential significance as a prognostic and predictive biomarker for colorectal and lung cancer.[27,29,104,199]

To understand the potential of molecular characterization of CTCs as a clinically useful liquid biopsy, we attempted to address several important methodological questions. These include validating the tumor origin of the isolated CTCs, as well as demonstrating the reliability and reproducibility of the assays that are utilized for CTC characterization.[67] We recently developed a HCC-specific CTC isolation method and demonstrated its efficacy for isolating and enumerating CTCs in hepatocellular carcinoma (HCC).[184] We have now developed a modification to that assay that allows for low resolution whole genome sequencing of the isolated CTCs using whole genome amplification (WGA) and next generation sequencing (NGS). Using this methodology, our current study investigates NGS-based SCNA profiling of CTCs as a potential molecular biomarker for HCC patients.

To that end, we performed SCNA profiling of HCC CTCs isolated immediately prior to surgical resection, and compared them to the SCNA profiles from the surgically resected HCC tumor tissue, peritumoral non-cancerous liver, and genomic DNA from pooled white blood cells. With a robust analysis of multiple DNA sources from individual patients, we aimed to extensively validate the tumor origin of the isolated CTCs, to establish that CTC-derived SCNA profiles can serve as surrogates of the underlying molecular alterations in the primary HCC

tumor, and to demonstrate the utility of a CTC-derived SCNA assay as a potentially clinically useful liquid biopsy in HCC.

5.3 Methods

Patient recruitment and sample processing

We prospectively enrolled patients undergoing surgical resection of HCC under our Institutional Review Board (IRB) approved protocol at the University of California, Los Angeles (IRB #14-001932).[66] Following discard of 5 mL of peripheral venous blood to prevent epithelial contamination, 10 mL of venous blood was collected in the operating room immediately prior to surgical resection into ACD solution A tubes (BD Pharmigen, Franklin Lakes, NJ), and stored at 4°C until processed. All samples were processed within 24 hours of collection as previously described [66]. Briefly, following initial density gradient centrifugation, the buffy coat is incubated with a cocktail of biotinylated CTC capture antibodies against the HCC cell surface markers asialoglycoprotein receptor (ASGPR; Abcam, Cambridge, UK), glypican-3 (GPC-3; Santa Cruz Biotechnology, Santa Cruz, CA), and epithelial cell-adhesion-molecule (EpCAM; Cell

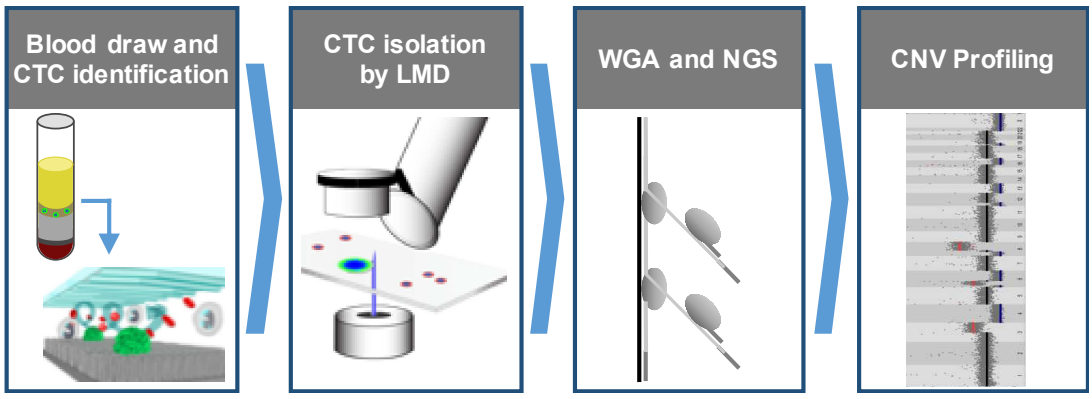


Figure 54. Workflow used in the study for CTC isolation and SCNA profiling using the NanoVelcro assay. Following gradient centrifugation, HCC-CTCs are selectively captured on the NanoVelcro surface for identification and subsequent isolation by LMD. CTC samples undergo WGA and QC prior to low resolution whole genome sequencing for SCNA profiling.

Signaling, Danvers, MA). Following capture antibody incubation, cells are washed and

resuspended in PBS (Gibco, Carlsbad, CA) and processed on the NanoVelcro platform (**Figure 54**).

In addition to peripheral blood samples, all patients had a single radiographically apparent lesion, and had both a section of the primary tumor and a section of the peritumoral normal liver isolated and flash frozen for subsequent molecular analysis. All patients underwent a blood draw prior to surgery as described in the methods section, with a portion of the venous blood sample being processed to obtain germline genomic DNA while the remainder was used for CTC isolation. Thus, all patients had DNA from whole blood, peritumoral normal liver, primary HCC tumor tissue, and CTCs available for molecular analysis (**Figure 55**).

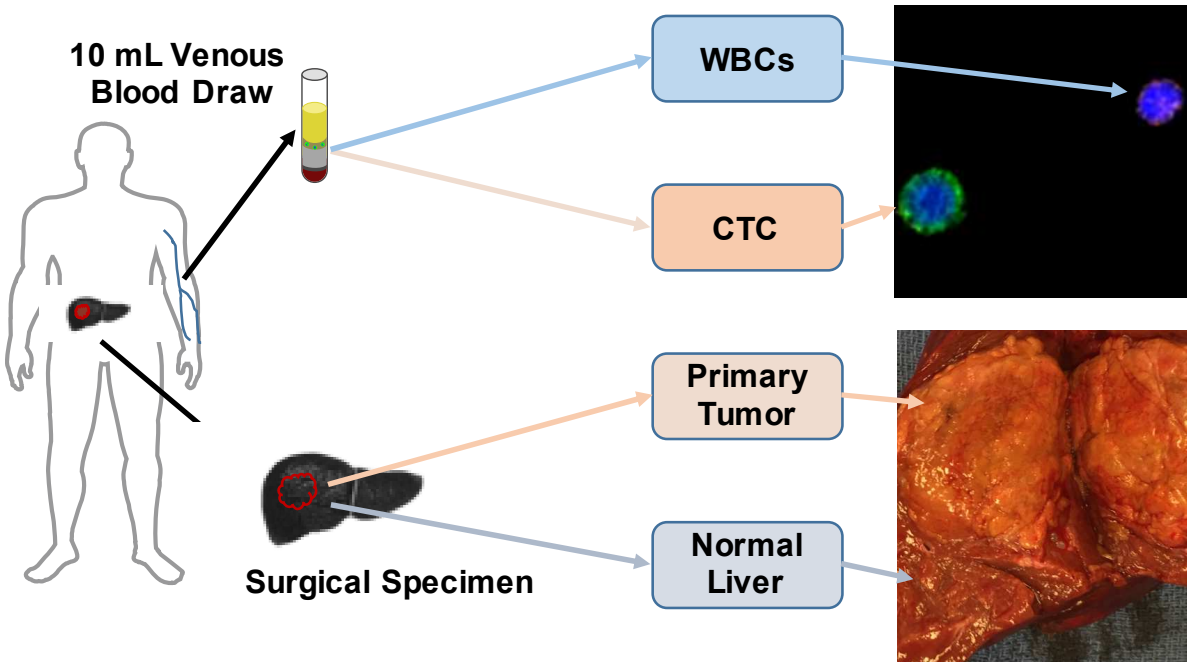


Figure 55. A total of 4 specimens are obtained from each patient. From peripheral blood, germline genomic DNA is obtained from the bulk buffy coat layer. Circulating tumor cells are obtained and sequenced as described in **Figure 54**. From the surgically resected specimen, a sample of both the primary tumor and the peritumoral normal liver are obtained and sequenced.

P-NanoVelcro CTC Chip Processing, Immunocytochemistry, and Chip Scanning

The assembly, operation, and staining of P-NanoVelcro CTC chips has previously been described.[37,66] An electro-spin method was used to assemble the Poly(lactic-co-glycolic acid) (PLGA) nano-spun chips onto a laser micro-dissection slide (Leica, Wetzlar, Germany) with an overlaid custom polydimethylsiloxane (PDMS) microfluidic component and attached to a syringe based microfluidic pump (KD Scientific, Holliston, MA). Chips are stained as described previously and CTCs identified via scanning fluorescent microscopy on a Nikon Eclipse 90i using immunocytochemistry (ICC) and NIS Elements 4.1 software. Chips are first scanned at 40x power followed by higher magnification manual imaging of candidate cells at 400x power for verification (**Figure 56**). For the resulting multi-channel ICC image, CTCs are defined as round/ovoid cells, $DAPI^{+}/CD45^{-}/CK^{+}/Vimentin^{+/-}$, with size $\geq$ 6-µm. WBCs are defined as round/ovoid cells, $DAPI^{+}/CD45^{+}/CK^{-}$, with size $\leq$ 6-µm. Any cells displaying CD45 positivity greater than 2x background were excluded as CTCs. CTC enumerations are reported as total counts per 4-mL venous blood, and were performed by the same blinded researcher (S.H.).

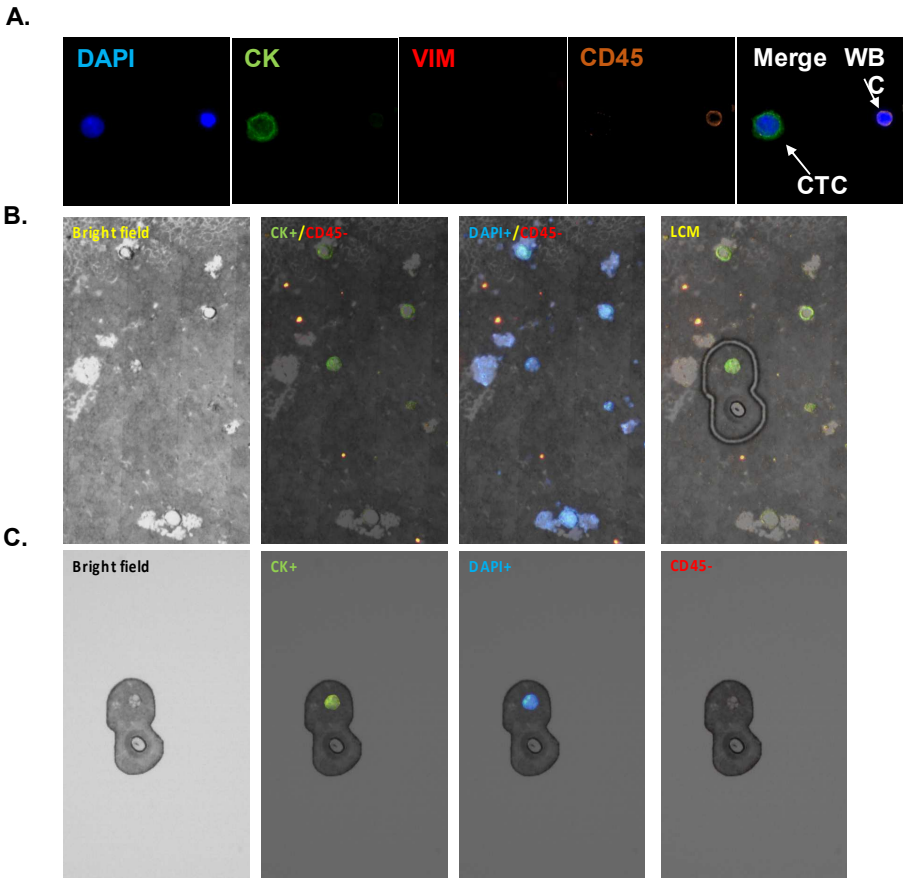


Figure 56. Representative CTC images. **(A)** CTCs imaged on the silicon wire chip used for identification. **(B)** CTCs imaged on the PLGA chip used for laser microdissection. **(C)** Cells isolated by LMD are imaged on the cap to ensure complete isolation.

Laser Micro-dissection

After CTC identification as outlined above, CTC chips were transferred to an ArcturusXT laser capture microdissection system (Thermo Fisher Scientific, Waltham, MA) attached to a Nikon Eclipse Ti microscope, and the CTCs were isolated into CapSure HS Caps (Thermo). Cell transfer to the cap was confirmed by light microscopy, and cells re-suspended in 4-µL PBS using a sterile pipette tip (**Figure 56C**). All candidate CTC cells from a single PLGA slide (equivalent to 2-mL of whole blood) were re-suspended on to a single cap with the exception of patients who had > 5 CTCs per slide (n = 2).

Whole Genome Amplification

Re-suspended cells were subjected to a freeze-thaw cycle at -80°C for 30 minutes followed by warming to room temperature. Cells were then lysed using the REPLI-g Single Cell Kit (Qiagen) using the manufacturer's recommended protocol. Whole genome amplification (WGA) was performed using the manufacturer's recommended protocol. Amplified DNA was purified by AMPure XP beads (Beckman Coulter, Brea, CA) using the manufacturer's recommended protocol resulting in 25 μL of purified WGA product.

Short Tandem Repeat (STR) Analysis

To eliminate contamination as a confounding factor, we performed STR analysis of all CTC samples and compared it to that of the primary tumor and whole blood with the GenePrint 10 v1.1 system (Promega, Madison, WI) using the manufacturer's recommended protocol. Briefly, a 10ng aliquot of template DNA was added to the amplification master mix and amplified for 30 cycles on a GeneAmp PCR System 9700 thermal cycler (Thermo). Fragments were analyzed on a AB 3500 Genetic Analyzer with POP-4 Polymer (Applied Biosystems, Foster City, CA) and visualized using GeneMapper 5 software (Applied Biosystems).

Sequencing library preparation and sequencing

Purified WGA products were sheared to generate DNA fragments averaging 350bps by sonication (Covaris, Woburn, MA). Sonicated DNA was cleaned, end-repaired, ligated, and amplified using the KAPA DNA Library Preparation Kit (KAPA Biosystems, Wilmington, MA) according to the manufacturer's protocol. Sequencing was performed on an Illumina NextSeq 500 (Illumina, San Diego, CA) using 75 bp paired end reads (2 x 75 bp).

Array Comparative Genomic Hybridization

Sample DNA (CTC, whole blood, peritumoral normal liver, and tumor tissue) and reference DNA (Agilent, Santa Clara, CA) were differentially labeled with cyanine-3 (CY3) and cyanine-5 (Cy5) dyes using the GenetiSure Amplification and Labeling Kit (Agilent) according to the manufacturer's protocol. Purified labeled DNA samples were prepared for hybridization, which took place on Agilent 8x60K CGH microarray slides at 67 ºC for 6 hours. Following the hybridization, the slides were scanned using the Agilent SureScan Microarray Scanner (Agilent). Microarray images were analyzed using the Agilent CytoGenomics software (Agilent) and the Microarray text files were analyzed using R version 3.3.2 and the packages rCGH, limma, agilp, and snapCGH.

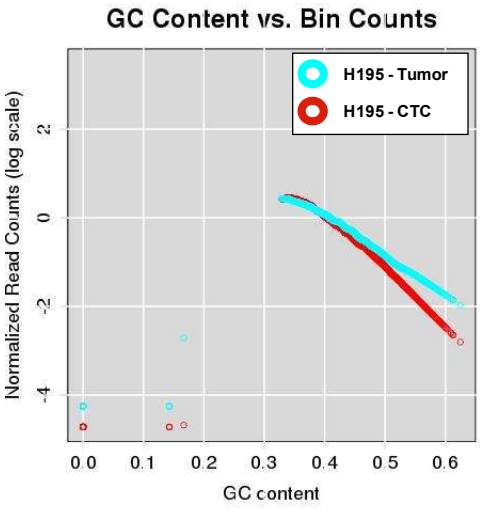


Figure 57. Example of GC content bias vs read count for CTC and primary tumor samples from patient H195 demonstrating the increased GC bias seen in CTC samples due to WGA. This bias is easily corrected using loess smoothing as previously described.

Copy Number Variation Analysis from WGS data

Fastq sequences were filtered and trimmed using Trimmomatic [108] and alignment was performed using bowtie2 [109] with settings "—very-sensitive-local." Coverage statistics and BAM to BED file conversion was performed using Bedtools.[110] Copy number variation and single cell NGS quality scores were analyzed by Ginkgo using a variable bin size of 250 kbps and simulating bins of 76 bp reads mapped with bowtie.[111] Segmentation was performed using normalized read counts, and sex chromosomes were masked for mixed-gender analyses. CTC samples demonstrated amplification bias due to GC content and

mappability that was corrected using loess smoothing as previously described (**Figure 57**).[10,111] While the majority of the CTC samples demonstrated a relatively uniform amplification reaction as evidenced by the narrow distribution of normalized read counts around a copy number of 2, several samples demonstrated significant bias and noise, making copy number determination difficult (**Figure 58**). A matrix of segmented copy number values at each bin for every sample was thus created and used for further analysis.

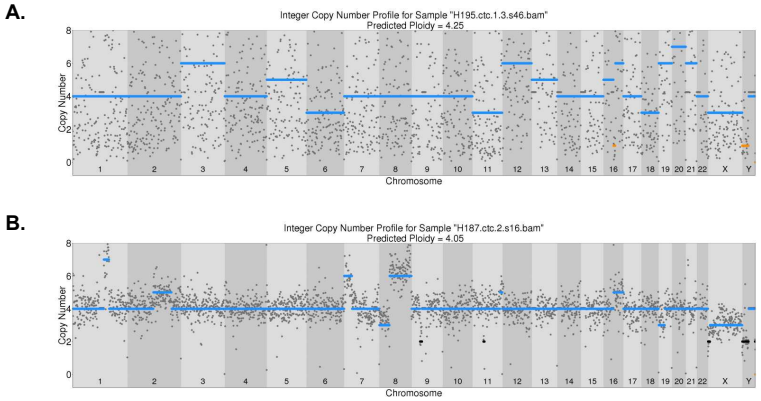


Figure 58. Examples of noisy CTC SCNA profiles resulting from either amplification bias or contamination. The CTC sample in **(A)** did not demonstrate a useful SCNA profile despite multiple attempts. In comparison, the CTC sample in **(B)** could be used after further normalization.

To evaluate the ability of CTC SCNA profiles to identify the tumor type of origin, additional lung cancer CTC data from the study by Ni et. al. was obtained and processed using the same methodology as for the CTC samples from our study.[29] Binned SCNA profiles for CTC and tumor samples from both our experiment and the Ni et. al. paper were first scaled so that the distribution of copy number states matched that of the TCGA dataset using the sweep and scale functions in R. The resultant matrix of binned SCNA values was similarly transformed to a SCNA per gene matrix based on gene – bin overlap using biomaRt.[200] SCNA copy numbers were transformed to the 5 value copy number scale used by the TCGA.

TCGA Copy Number Profiling

TCGA copy number profiles for all cancer types listed in **Table 16** were obtained from the Broad Institute's Firebrowse TCGA data version 2016_01_28 using FirebrowseR.[201] The "all_thresholded.by_genes.txt" file was used in all cases. All TCGA SCNA data was obtained from microarray data with the addition of low-pass NGS SCNA data when available. A comparison of the available microarray and low-pass SCNA data confirmed the close correlation between these data types for the "by gene" SCNA data (data not shown). Only primary tumor samples (TCGA barcode sample code "01") were included. A total of 10478 samples from 32 tumor types were used (**Table 16**). The final gene list was then limited to genes with adequate coverage for all CTC samples (n = 14351 genes). The high dimensional patient x gene data frame was reduced by eliminating genes demonstrating a > 95% correlation resulting in a total of 1049 genes. The 1049 genes were further reduced by transforming the data into a patient x cytoband data frame using Biomart, taking the mean of all duplicate entries.[200] The resulting 556 cytobands were then reduced to just 274 cytobands that were previously associated with global and cancer type specific SCNAs based on prior pan-cancer SCNA analysis.[193]

Two pan-genomic variables were created to assist with classification. First, a copy number instability (CIN) score was created from the TCGA SCNA data. The CIN score was defined as the absolute value of all SCNAs for each sample. Next, the TCGA SCNA data was analyzed visually by 2-dimensional T-distributed Stochastic Neighbor Embedding (t-SNE) using the "Rtsne" package.[202,203] Both the copy number instability score and the two t-SNE values for all samples was added to the final dataset used in the classification model.

The TCGA data was divided 80/20 into training and testing sets and a random forest classifier was trained on the training set (n = 8382 samples) using the R packages "randomForest" and "caret".[204] Class representation was equalized via the upSample function of the caret package. Cancer types demonstrating a balanced class accuracy < 80% were excluded from further analysis. Additionally, cancer types were grouped based on the results of prior pan-cancer studies as well as known similar cancer types such that the final model included 21 cancer types and 11 cancer classes (**Table 15**).

The resulting 11 cancer types (**Table 15**) were then used to train a random forest model for cancer classification based on the TCGA SCNA data. Parameter tuning was performed by a repeated cross-validation approach and the final model verified for accuracy using the test set. The classifier was trained using the caret package as a modeling wrapper and the R packages "randomForest" for the random Forest model.[204-207] This final model was then used to classify the CTC samples from both this study as well as the previously published lung cancer CTC study.[29]

Statistical Analysis

Statistical analysis and visualization was performed in R (version 3.3.2). Categorical variables were summarized as frequencies and percentages while continuous variables were summarized as medians and interquartile ranges (IQR). All CTC numbers are reported as whole numbers in 4-mL of venous blood.

The "gplots" package was used for unsupervised hierarchical clustering of SCNA profiles from CTC and primary tumor samples using complete linkage and Euclidean distance metric.[208]

The correlation of the SCNA pattern of the CTCs to that of the matched primary tumor samples was calculated using Pearson's correlation.

5.4 Results

Patient characteristics and sample collection

Ten consecutive patients undergoing surgical resection of HCC from our ongoing prospective HCC biomarker protocol between March, 2016 and January, 2017 were included for analysis. The clinical, laboratory, radiologic, and treatment characteristics of the 10 patients are summarized in **Table 14**. Nine of the patients underwent resection of a primary liver tumor and one patient underwent resection of a retroperitoneal metastasis that developed 10 years after liver transplantation for HCC (patient H167). All patients had DNA from whole blood, peritumoral normal liver, HCC tumor tissue, and CTCs available for molecular analysis (**Figure 55**).

Table 14. Clinical, Laboratory, Radiologic, Treatment and Pathologic Characteristics of the Patients (n = 10)

Characteristic	Data
Clinical	
Age, median (IQR)	65 (60 – 73)
Female, n (%)	3 (30)
Diagnosis, n (%)*	
HCV	7 (70)
HBV	3 (30)
Cryptogenic	1 (10)
Laboratory	
Physiologic MELD, median (IQR)	7 (6 – 7)
AFP, most recent, median (IQR)	4.3 (3.6 – 8.7)
AFP, maximum, median (IQR)	4.9 (3.9 – 9.3)
HCC CTC count, median (IQR)	4 (2 – 11)
Vimentin (+) CTC count, median (IQR)	3 (0 – 4)
PD-L1 (+) CTC count, median (IQR)	0 (0 – 2)
Radiologic tumor characteristics	
Tumor Location, n (%)	
Right Lobe	5 (50)
Left Lobe	2 (20)
Central	2 (20)
Retroperitoneum**	1 (10)
Maximum tumor diameter, median (IQR)	5.3 (3.8 – 7.5)
Pathologic Characteristics	
Tumor Diameter, median (IQR) ***	4.8 (3.6 – 7.9)
Microvascular invasion, n (%)***	5 (56)
Macrovascular invasion, n (%)***	1 (11)
PNI, n (%)***	0 (0)
Grade, n (%)	
Well	0 (0)
Moderate	6 (60)
Poor	4 (40)

* n = 1 patient with both HCV and HBV; ** n = 1, post-transplant recurrence undergoing retroperitoneal mass excision; *** n = 9 patients undergoing primary liver resection

CTC enumeration and sequencing

CTC enumeration revealed a median of 4 (IQR: 2 – 11, range: 1 – 16) CTCs per 4-mL VB. Multiple displacement amplification-based WGA was successfully performed on CTC samples, and low-resolution whole genome sequencing performed on a NextSeq 500. Sequencing reads for the CTC samples were processed using a variable bin algorithm specifically designed for low-resolution single cell copy number determination.[10,111] We obtained an average of 0.88 million uniquely mapped reads (range: 0.51 – 1.9 million uniquely mapped reads) from the CTC samples, with an average depth of coverage of 0.022x and a minimum read count of 0.5 million reads for CTC samples. Blood, peritumoral normal liver, and HCC tumor tissue DNA was extracted as detailed in the methods before being subjected to the same sequencing and analysis pipeline as the CTC samples, resulting in a median of 0.99 million uniquely mapped reads (IQR: 0.86 – 1.2 million reads, range: 0.38 – 2.1 million reads).

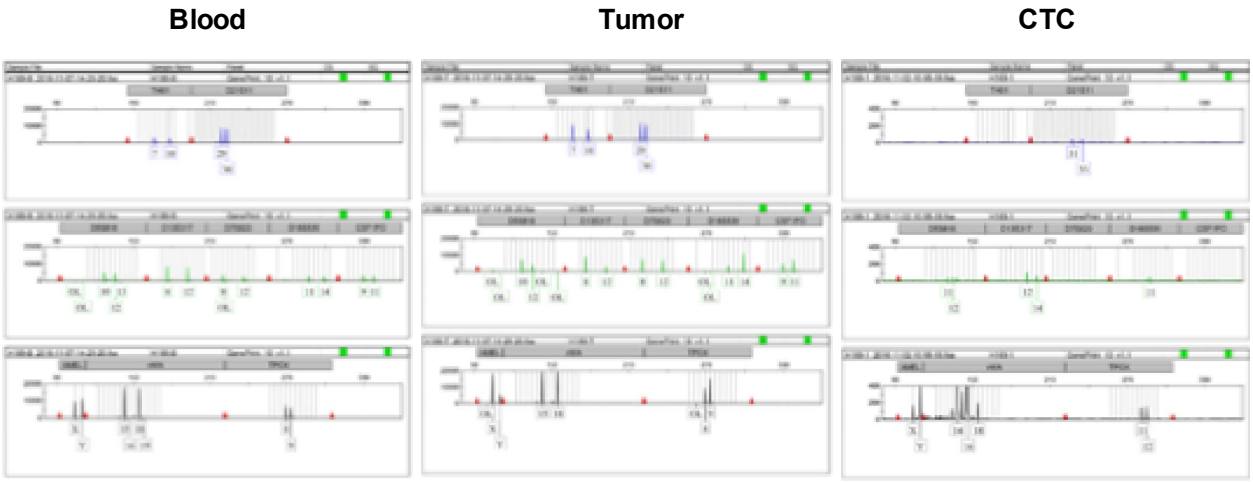


Figure 59. STR analysis for patient H169 demonstrating that the CTC sample genotype does not match that of the primary tumor or blood.

CTC versus primary tumor SCNA profiles

We performing STR typing of the blood, primary tumor, and CTCs to confirm that contamination had not occurred during WGA processing. While STR typing confirmed that CTCs had originated from the same individual for 9/10 patients, it revealed that the CTCs for patient H169 were likely from a different source and this patient was excluded from further analyses (**Figure 59**). Next, whole genome SCNA profiles were obtained by visualizing the copy number state at each 250-kbps bin along the entire genome for the DNA from whole blood, normal liver, the primary tumor and CTC samples (**Figure 60**). For the 9 patients with confirmatory STR typing, the whole genome SCNA profiles were compared to determine if CTCs exhibit the somatic SCNAs found in the primary tumor. Inspection of the copy number profiles clearly demonstrates the recapitulation of the somatic changes of the primary tumor by the CTCs, helping to establish the tumor origin of the CTCs. Most of the SCNAs seen in **Figure 60** are in regions of the genome previously identified as somatic SCNAs associated with HCC (**Table 17)**. For example, for patient H167 both the primary tumor and CTC sample demonstrate 1q and 8q amplifications as well as chromosome 13, 15, and 18 loss (**Figure 60**). Similar results

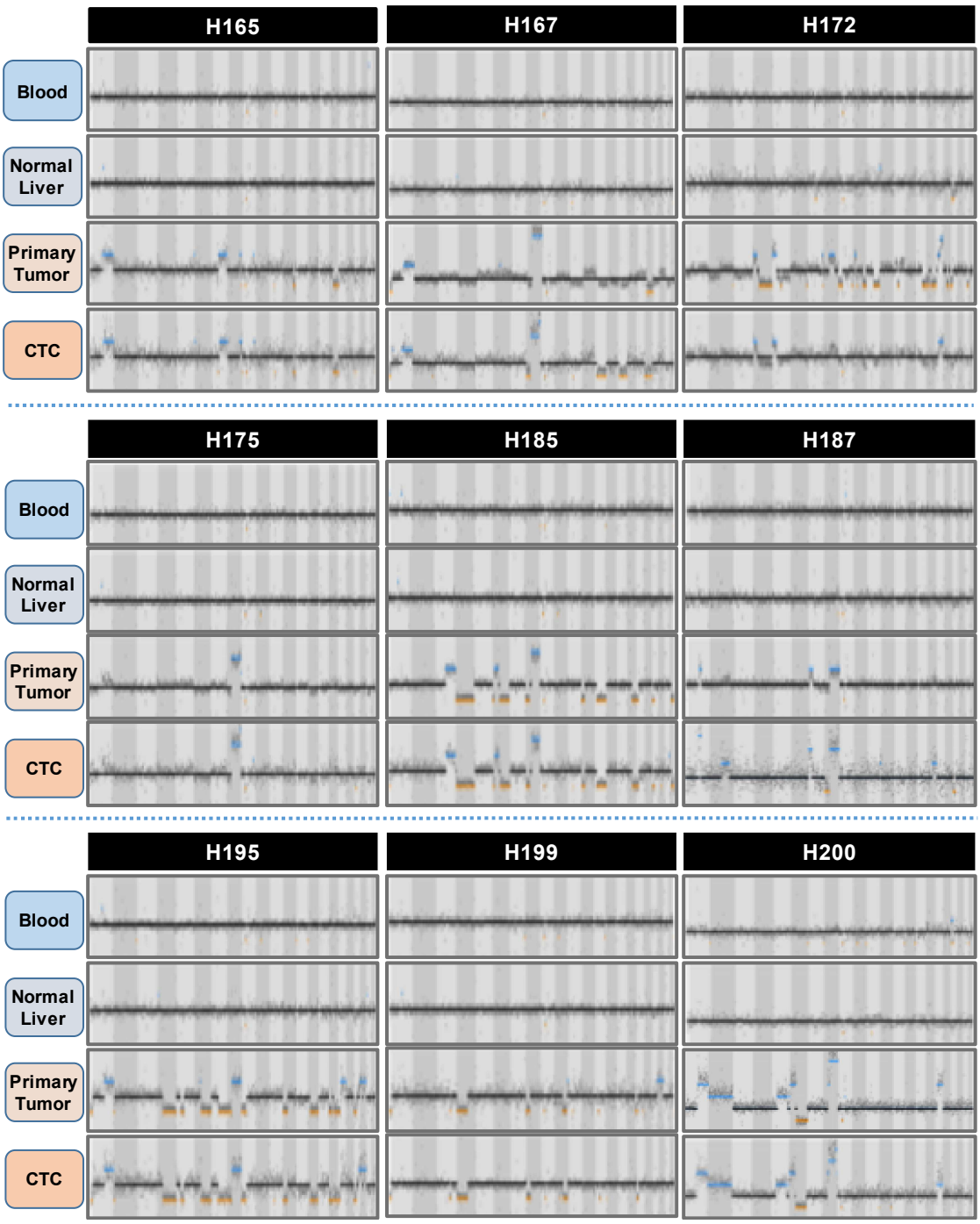


Figure 60. Global copy number profiles for blood, peritumoral normal liver, CTCs, and primary tumor (n = 8) or solitary metastasis (n = 1, H167) demonstrates the recapitulation of somatic SCNAs from the tumor in CTC samples, supporting the potential of CTCs to act as a liquid biopsy of the tumor.

were found for all 9 samples, with the majority of somatic SCNAs seen in the primary tumor also present in the CTC samples. We confirmed the SCNA profiles found by NGS using array CGH

for a subset of 6 patients with sufficient DNA available, and again found that the CTCs recapitulated the somatic changes found in the primary tumor (**Figure 61**).

Figure 61. Array CGH of tumor and CTC samples for a subset of patients confirming the results of the NGS SCNA profiling.

CTC clustering analysis

To statistically verify the visualized results of **Figure 60**, we developed a 59 loci panel of previously identified regions or genes amplified or deleted in HCC (**Table 17**).[209-216] We

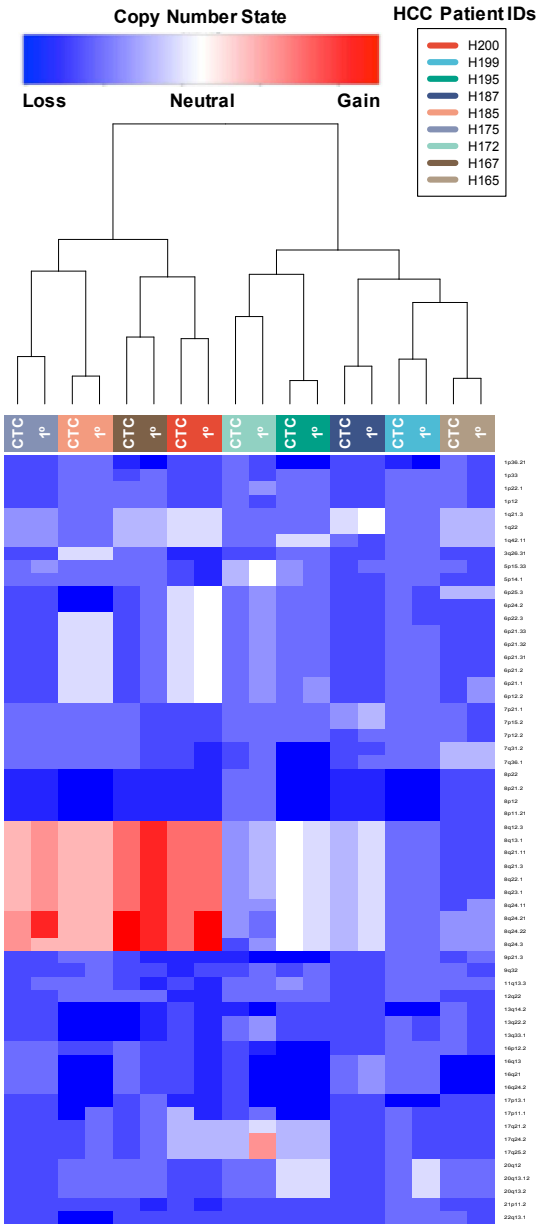


Figure 62. Heatmap showing copy number profiles at 59 loci frequently amplified or lost in HCC. Gain (red) or loss (blue) at each of the 59 cytobands (identified on the right of the heatmap) in both CTC and tumor samples from all 9 patients that passed STR analysis are shown. Unsupervised hierarchical clustering demonstrates clustering of CTC – tumor pairs for all 9 patients. CTC — Circulating tumor cell, 1° — Primary tumor.

compared the segmented, normalized copy number data of CTCs and primary tumors at the 59 loci in the panel, and found frequent somatic gains and losses for all patients, consistent with previous studies of HCC tumors (**Figure 62**). Unsupervised hierarchical clustering of the CTC and tumor copy number values demonstrated that CTC samples clustered with their respective primary tumor for all 9 patients. Analysis of the global copy number profiles by a Pearson correlation matrix (**Figure 63)** demonstrated that the CTC SCNA profiles shared an average of 95% (IQR: 86 – 97%) of the gains and losses found in the primary tumor.

CTC SCNA sequencing as a molecular biomarker

To investigate the potential clinical relevance of our CTC SCNA profiling to act as a surrogate for the primary tumor of interest, we explored our ability to detect frequently found prognostic or actionable copy number changes identified by prior studies of HCC genetics (**Table 17**). One such example is illustrated in **Figure 64**. Patient H199 is a 64-year-old male with cryptogenic cirrhosis and a 9.0 cm segment 2-3

AFP-nonproducing lesion who underwent a left hepatectomy. On pathologic examination, his tumor was found to be moderately differentiated with evidence of microvascular invasion but no macrovascular invasion.

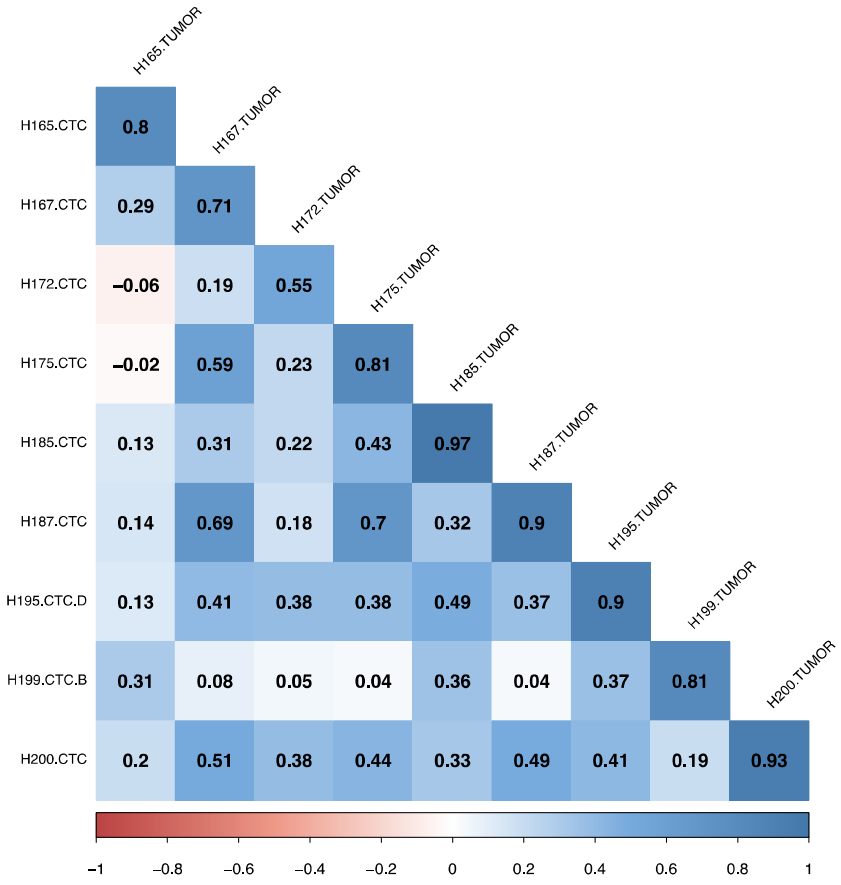


Figure 63. Spearman correlation matrix of CTC and primary tumor samples of the low-resolution whole genome sequencing copy number profiles for 250 kbp bins.

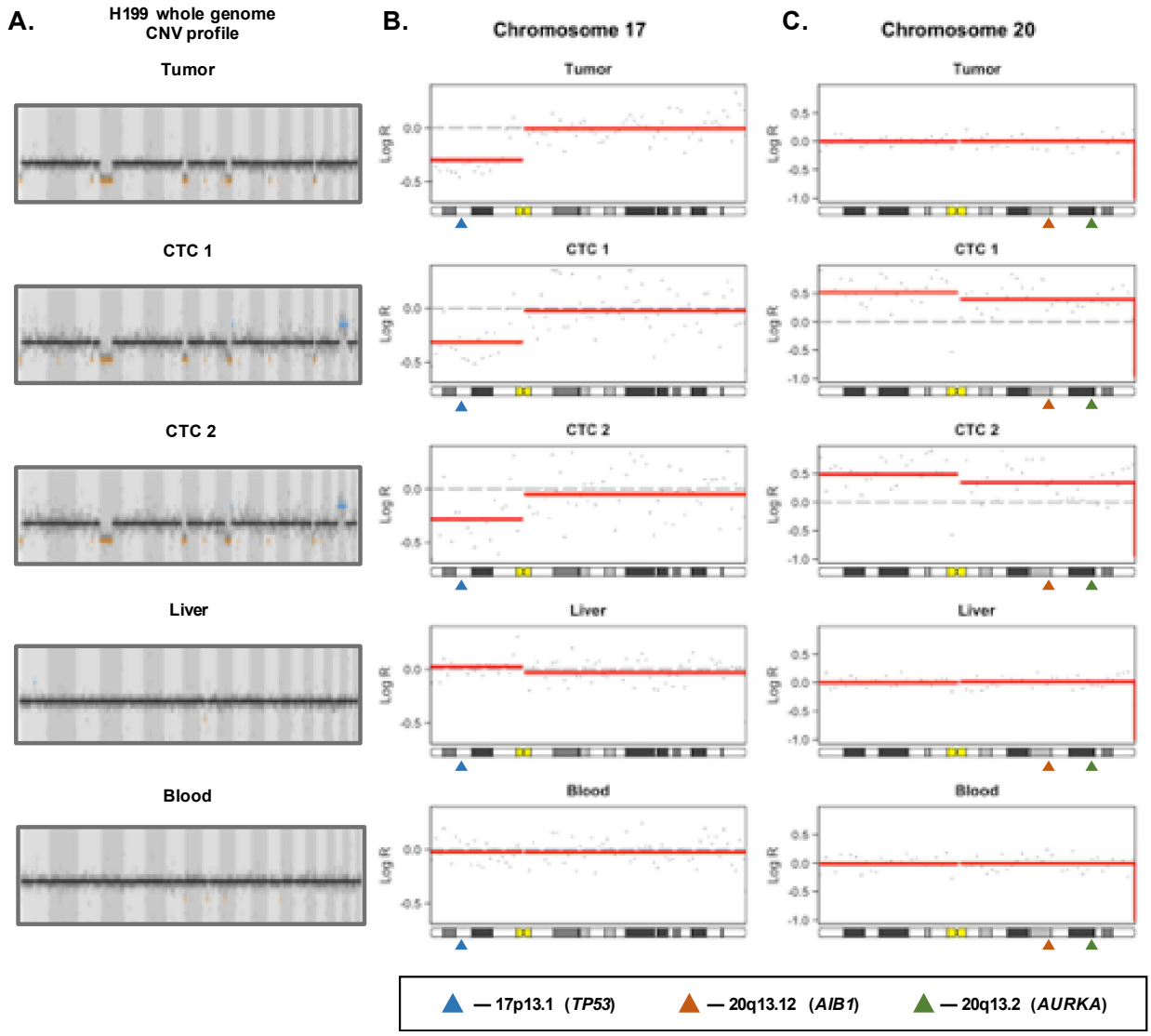


Figure 64. (A) Comparison of low-resolution whole genome copy number profiles for primary tumor, CTCs 1 and 2, normal liver, and whole blood reveals reproduction of the majority of somatic SCNAs from the primary tumor in both CTC samples. **(B)** Chromosome 17p loss is seen in the primary tumor and both CTC samples, but not the normal liver or whole blood. The location of the tumor suppressor TP53 gene, the most frequently mutated or lost gene in HCC, is indicated by the blue triangle. **(C)** Chromosome 20 amplification was seen in both CTC samples but not in the primary tumor. Chromosome 20 amplifications are a recurrent somatic copy number alteration in HCC associated with 2 important oncogenes. Overexpression of *AIB1* (orange triangle) is frequently described in HCC, and has previously been associated with invasiveness. Additionally, recent research has demonstrated that the oncogenic effects of *MYC* dysregulation, a common occurrence in HCC, require overexpression of *AURKA* (green triangle) for stabilization. Furthermore, in p53-altered HCC patients, the MYC-AURKA complex is an actionable drug target based on preclinical studies.

Examination of the copy number profile from the primary tumor revealed copy number loss at chromosome 1p, 4p, 8p, 10q and 17p (**Figure 64A**). Of particular note is the chromosome 17p loss, as it contains the well-known tumor suppressor *TP53* gene, the most frequently mutated

or lost gene in HCC (**Figure 64B**).[209] When examining the two CTC samples from this patient, all of the CN losses found in the primary tumor were detected; however, an additional amplification of chromosome 20 was detected from both CTC samples (**Figure 64C**). Chromosome 20 amplifications are a recurrent somatic SCNA in HCC associated with 2 important oncogenes.[217-221] Overexpression of *AIB1* is frequently described in HCC, and has previously been associated with invasiveness and sensitivity to sorafenib therapy.[217,221] Additionally, recent research has demonstrated that the oncogenic effects of *MYC* dysregulation, a common occurrence in HCC, require overexpression of *AURKA* for stabilization.[220] Furthermore, in p53-altered HCC patients, the MYC-AURKA complex is an actionable drug target in preclinical studies.[218]

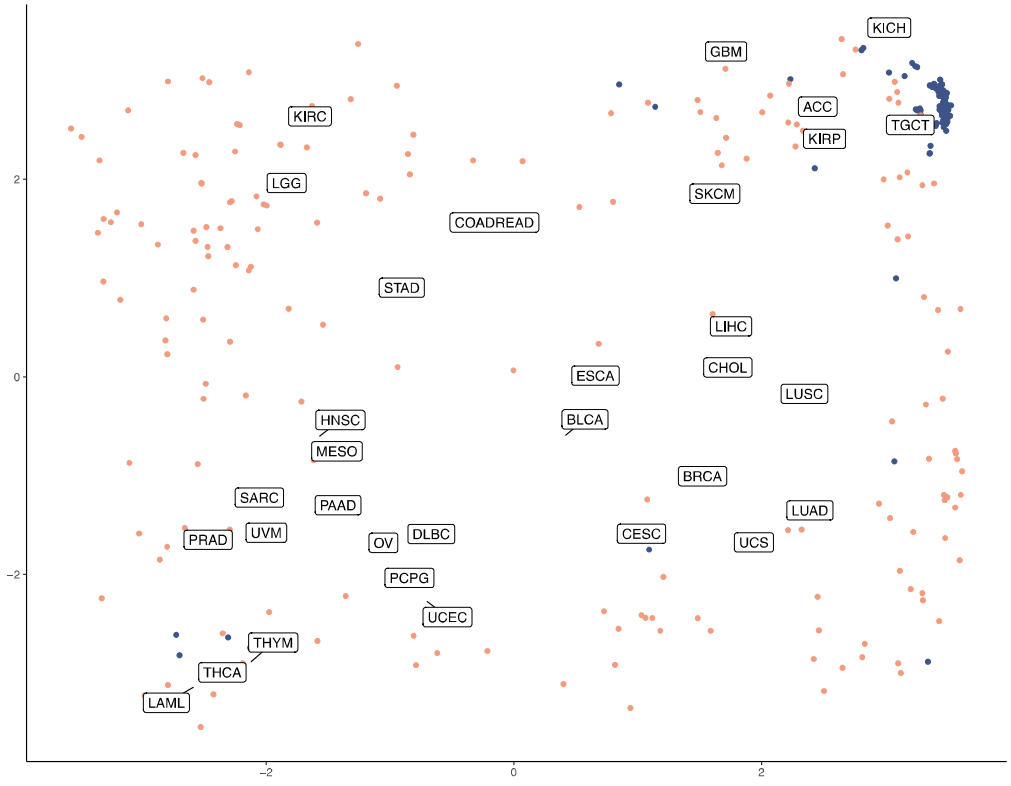


Figure 65. SCNAs at all genes for the 32 cancer types were transformed using t-SNE into a 2-dimensional space. The centroid mean of all samples for each cancer type is plotted as well as the individual samples for the best clustering (TGCT — Blue) and worst clustering (ESCA — Pink) cancer types. All individual samples are plotted by cancer type in **Figure 66**.

Cancer Type Classification Using SCNA Profiles

As CTCs are universally shed by all tumor types, and CTCs demonstrate similar SCNA patterns to the primary tumor, we investigated the ability of CTC SCNAs to determine the site of origin of the primary tumor. To do so we obtained whole genome copy number data for 10478 samples from 31 tumor types available in the TCGA dataset.[201] The copy number state at 274 cancer-associated cytobands for all samples of the 31 cancer types was evaluated visually through 2-dimensional transformation using t-SNE (**Figure 65; Figure 66**).[202] While some cancers such as glioblastoma multiforme (GBM) or testicular germ cell tumors (TGCT) demonstrated clear clustering, others such as pancreatic adenocarcinoma (PAAD) or mesothelioma (MESO) had samples scattered across the t-SNE space with no clear clustering identified. We used the copy number state of each sample to calculate 3 whole genome metrics: a chromosome instability score (CIN) as well as the two t-SNE dimension variables. We then trained a random forest model to predict the tumor site of origin on the training set (80% of samples) and obtained a model accuracy of 0.59 for the test set. Analysis of the misclassification rate revealed that many misclassifications were occurring between expected classes such as low-grade glioma (LGG) and GBM or between different types of kidney tumors (KICH, KIRC, and KIRP). Based on the combination of the class accuracy of the random forest model, the degree of clustering on t-SNE, and prior studies on cancer subtype classification[194], a total of 21 cancer

types were divided into 11 cancer classes (**Table 15**).

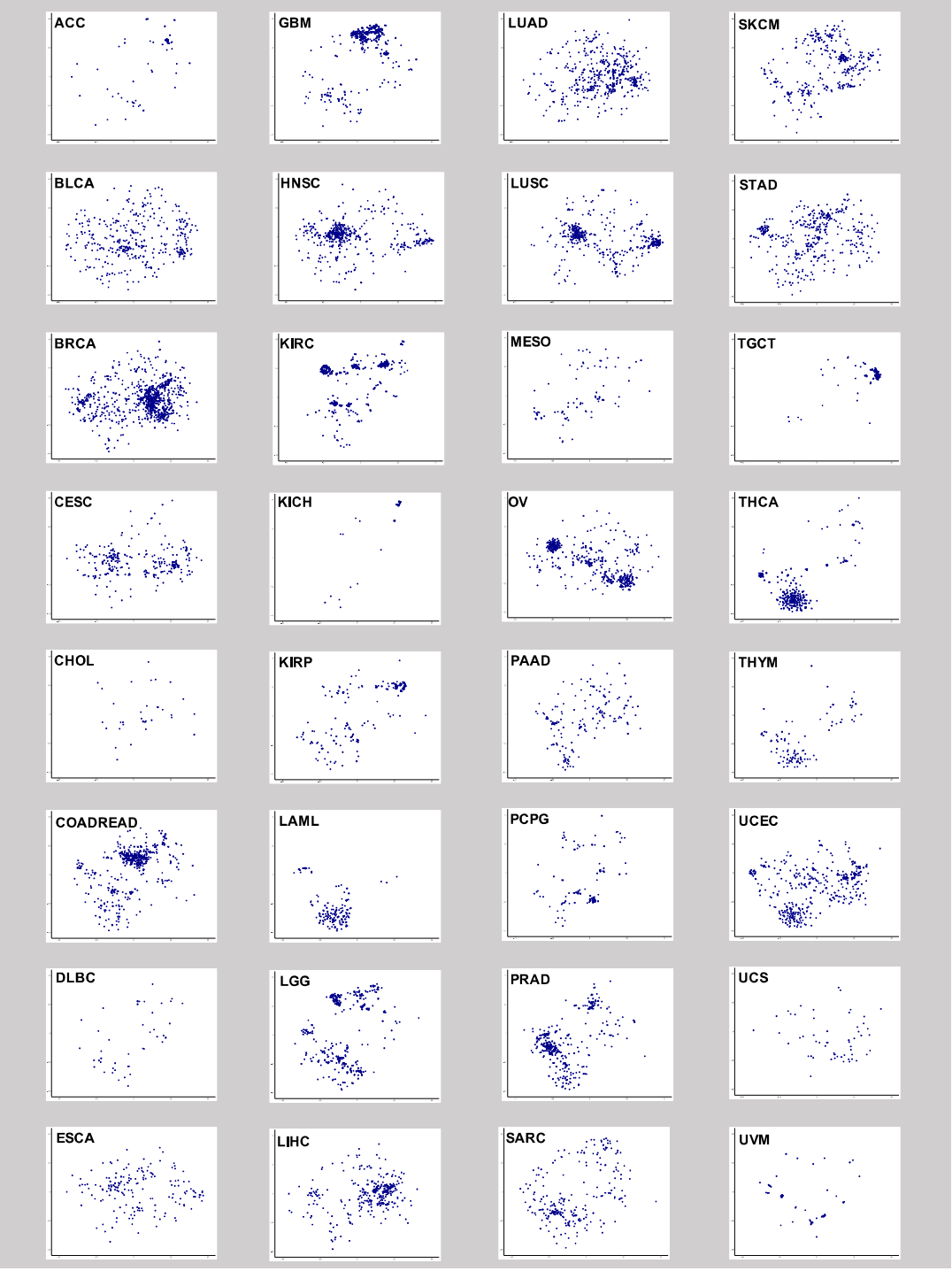


Figure 66. All samples for each of the 32 cancer types plotted onto the 2-dimensional t-SNE space. While some cancer types such as testicular germ cell tumor (TGCT) appear to cluster closely together, others such as esophageal cancer (ESCA) demonstrate no clustering at all.

We used a random forest algorithm for the 11-cancer class model. The overall model accuracy was 0.79 (95% CI: 0.77 – 0.81), with individual cancer type balanced accuracies ranging from 0.75 – 0.98 (**Figure 67**).

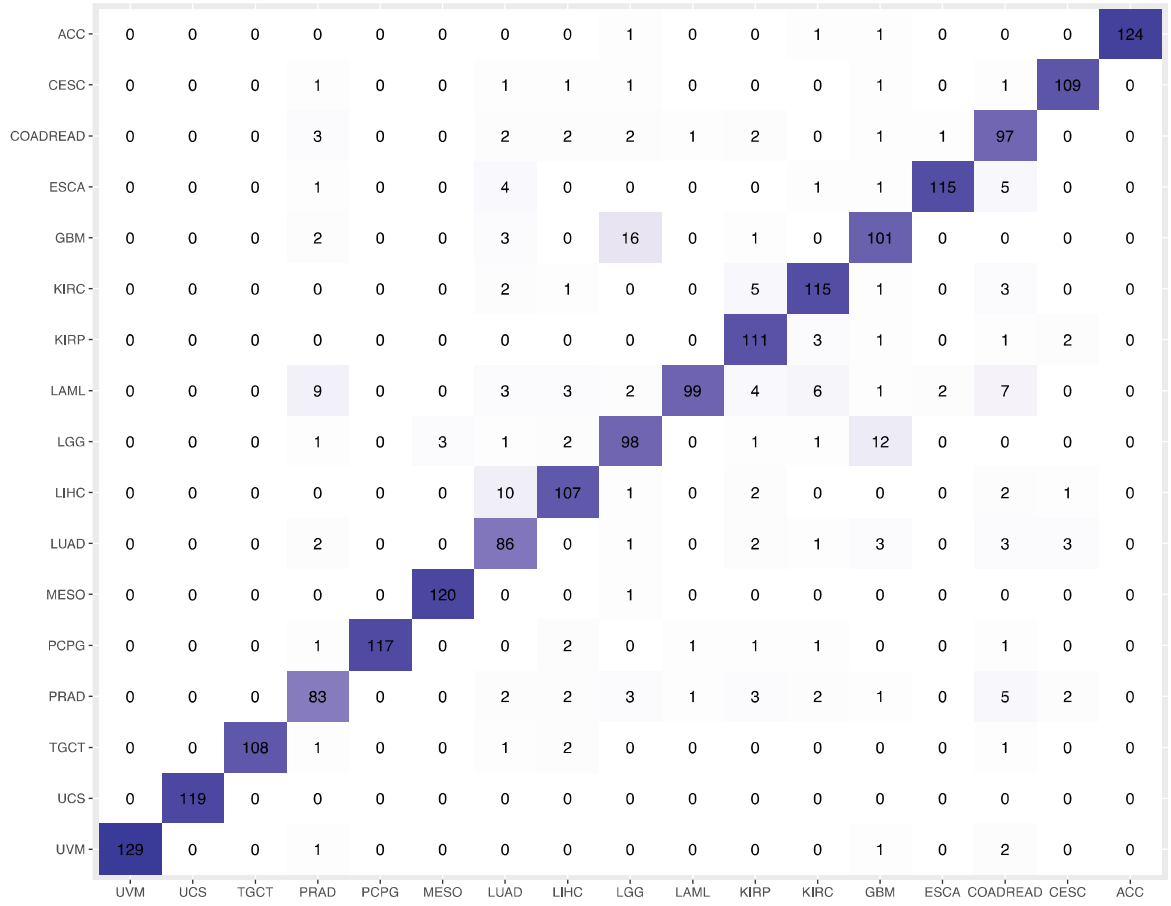


Figure 67. TCGA class prediction confusion matrix for the final cancer classification model.

CTC Cancer Type Classification

Given the circulating nature of CTCs we next sought to determine if CTC SCNA profiles could be used to determine the tissue of origin from which the CTCs originated. From a previously published study we acquired the CTC SCNA data of 29 CTC samples obtained from 7 lung cancer patients[29]. Our model correctly identified the cancer type for 33/44 (75%) CTCs. A total of 9/15 (60%) HCC CTCs were correctly classified, and 5/9 (56%) patients had at least 1

CTC sample identify HCC as the tissue of origin. For the lung cancer CTCs, the model identified the tissue of origin correctly for 23/29 (79%) CTCs and all 7 patients had at least 1 CTC sample identifying lung adenocarcinoma as the cancer type correctly (**Figure 68**; **Figure 69**).

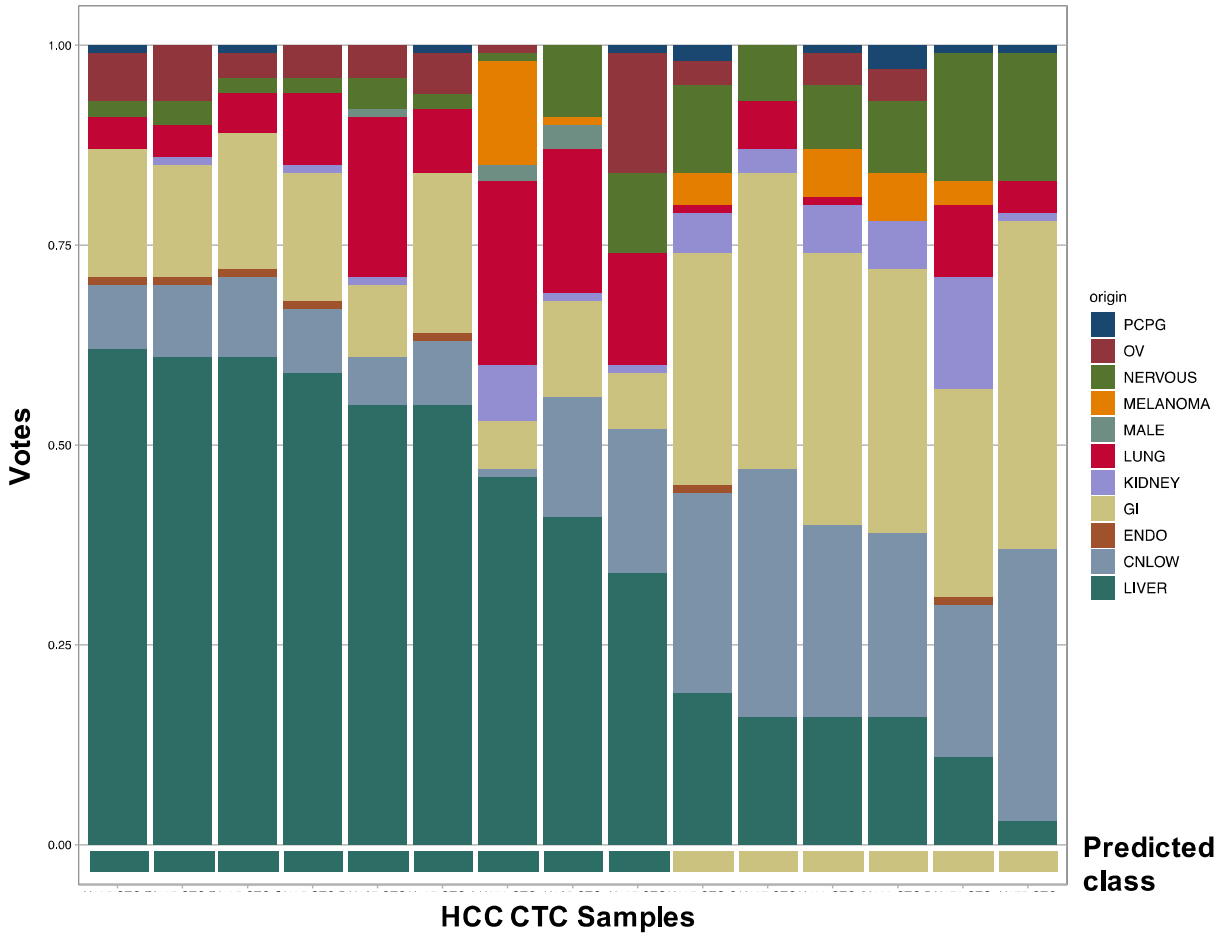


Figure 68. Cancer class prediction based on SCNA profile for each of the 15 HCC CTC samples using the classification model. 9/15 (60%) of CTCs were correctly classified as being HCC while 6/15 (40%) were classified as being from a GI source. 5/9 (56%) patients had at least one CTC classified as being from a HCC tumor.

5.5 Discussion

Blood-based "liquid" biopsies hold numerous potential benefits over traditional percutaneous or surgical biopsies such as reduced risk, reduced cost, and reduced patient discomfort. Furthermore, they are increasingly recognized as a necessary component of a precision oncology treatment strategy, given the ongoing need for tumor tissue as the tumor adapts to new therapies.[222] Despite these benefits, liquid biopsies have yet to enter clinical practice due to issues including reproducibility and applicability for molecular analysis.[223] One reason may be that most molecular liquid biopsy studies to date have investigated detection of point mutations with mixed results due to the difficulty of making accurate mutation calls from limited starting DNA template.[224,225] In contrast, whole genome NGS SCNA profiling of liquid biopsies results in a robust signal that is highly reproducible from as few as a couple hundred thousand reads. [10,226] There are fewer targetable SCNAs than somatic mutations currently, but this may change as the functional impact of SCNAs has been less well studied than somatic mutations. For example, recent studies have found that SCNAs represent a significant portion of the somatic genetic changes for many tumors, and may be more important driver events than somatic mutations for many cancer types.[227] To that end, we sought to investigate the validity of NGS-based SCNA-profiling of CTCs and to demonstrate the potential utility of such an assay for different clinical applications.

Previous research into low-resolution copy number profiling has demonstrated that as few as 250-350k reads are sufficient for calling copy number events larger than 500 kbp.[10] Thus, our sequencing method which resulted in less than a million reads per sample, potentially limits our detection of smaller focal SCNAs. However, current research indicates that whole chromosome, arm, and cytoband length SCNAs are the more important in the development of cancer, making

our depth sufficient.[193,195] When run in bulk, our assay is also surprisingly affordable; a prior study using a similar protocol showed that the total reagent and sequencing costs for such a SCNA profiling assay would be as low as $30 USD, a significant improvement over older array CGH based SCNA assays.[10]

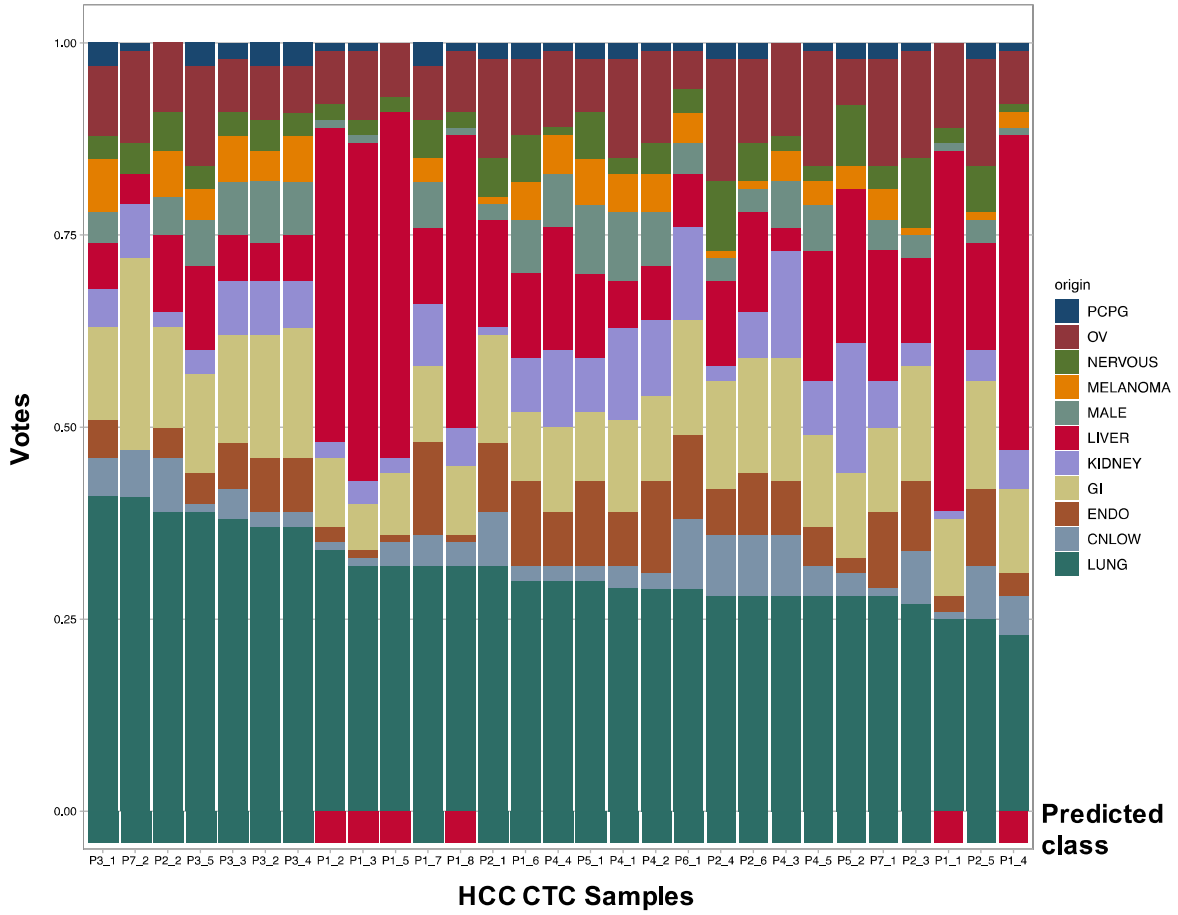


Figure 69. Cancer class prediction based on SCNA profile for each of the 29 Lung CTC samples using the classification model. 23/29 (79%) of CTCs were correctly classified as being HCC while 6/29 (21%) were classified as being from a GI source. 7/7 (100%) patients had at least one CTC classified as being from a lung tumor.

We tested our SCNA sequencing method of CTCs from 10 patients with HCC recruited for our ongoing study of CTCs in liver cancer.[184] We found that CTC SCNA profiles consistently demonstrated that the copy number changes seen in the primary tumor are also found in the CTCs. This finding makes us confident that the CTCs we sequenced likely originated from the primary tumor, and that CTCs can act as molecular surrogates of the primary tumor for SCNA

profiling. We further tested the reproducibility of the assay by sequencing multiple CTC samples from a single patient and found similar CTC profiles from almost all of the CTC samples. Of the 16 CTC samples sequenced, only 3 (19%) of them demonstrated problems. One sample did not pass STR typing analysis indicating potential contamination, one sample demonstrated decreased signal and missing SCNAs indicative of contamination with germline DNA, and one sample had an uninterpretable SCNA profile indicative of failure at some step of the protocol. These failures are all common problems when working with single cell sequencing and the failure rate could likely be greatly reduced by implementing clean room controls and automation.[121]

We demonstrated two potential applications of our CTC SCNA profiling assay: identification of prognostic or targetable SCNAs and determining the tumor site of origin for CTCs. Many prognostic and targetable SCNAs were found in our CTC samples. For example, patient H199's 2 CTC samples both demonstrated the chromosome 17p loss seen in the primary tumor. However, they also both showed chromosome 20 amplifications, a prognostic and potentially actionable finding.[218,221] As prior studies have demonstrated that metastases tend to arise from a single subclone of the primary tumor,[228] and that CTCs have been shown to be oligoclonal precursors of metastases in breast cancer,[196] it is plausible that CTCs originate from the more aggressive subclones of the primary tumor. Further studies are needed to investigate this hypothesis. If true, it would lend credence to the idea that CTC-based liquid biopsies selectively sample aggressive subclones at increased risk of metastasizing and are thus able to overcome the abundant tumor heterogeneity that can limit the clinical utility of traditional percutaneous biopsies for some cancer types.

Prior studies investigating the classification of tumors by site of origin found that SCNAs are the most important genetic determinant in classification models, contributing more

information than mRNA expression, miRNA expression, or DNA methylation data.[194] Our finding that some cancer types were well classified by SCNAs while others were not was previously found in other studies investigating SCNA-based classification.[227,229] Overall, we found that SCNA data alone could correctly identify the tumor site of origin for most cancer types, and that most of the errors were due to known similarities between some cancers such as between head and neck squamous cell and esophageal squamous cell cancers. Applying our model to CTC SCNA data we again found that we could determine the site of origin of the CTCs for the majority of samples, both from our own study and from a prior study of lung cancer CTCs. This may indicate a potential use in identifying the site of origin for patients with tumors of unknown origin, or in cases of recurrence with multiple prior primary tumors.

In conclusion, we demonstrate the feasibility of low resolution NGS SCNA profiling of CTCs as a molecular liquid biopsy and demonstrate the potential applications of such an assay. Analysis of CTCs and primary tumor tissue demonstrated concordant alterations that were not present in the peritumoral normal liver or blood genomic DNA. This supports the potential use of CTC-derived SCNA profiling as a clinically relevant surrogate of the primary tumor. While our study was underpowered to address the clinical utility of the assay, we hope to use it in future studies as part of our ongoing work studying CTCs in gastrointestinal tumors.

Table 16. Number of samples for each of the 31 cancer types. Bolded cancer types are those that were included in the final classification model.

Cancer Type	Number of Samples
BRCA	1080
COADREAD	616
OV	579
GBM	577
UCEC	539
KIRC	528
HNSC	522
LUAD	516
LGG	513
LUSC	501
THCA	499
PRAD	492
STAD	441
BLCA	408
LIHC	370
CESC	295
KIRP	288
SARC	257
LAML	191
ESCA	184
PAAD	184
PCPG	162
TGCT	150
THYM	123
ACC	90
MESO	87
UVM	80
KICH	66
UCS	56
DLBC	48
CHOL	36

Table 15. Final cancer classification model classes and the types of cancers included in each of the classes. A total of 21 cancer types and 11 cancer classes were used.

ENDO (ACC)	**LIVER** (HCC, CHOL)	**GI** (CRC, PDAC, STAD)	**NERVOUS** (GBM, LGG)	**MALE** (TGCT)	**LUNG** (LUAD, LUSC)
OV	**PCPG**	**KIDNEY** (KICH, KIRC, KIRP)	**MELANOMA** (UVM)	**CNLOW** (THYM, LAML, THCA, PRAD)	

Table 17. List of the 59 loci frequently amplified or lost in HCC as well as the associated study that reported the loci.

Cytoband	Study
11q13.3	Schulze, 2016
12q22	Kan, 2013
13q14.2	Schulze, 2016
13q22.2	Kan, 2013
13q33.1	Kan, 2013
16p12.2	Kan, 2013
16q13	Kan, 2013
16q21	Kan, 2013
16q24.2	Kan, 2013
17p11.1	Kan, 2013
17p13.1	Kan, 2013
17q21.2	Kan, 2013
17q24.2	Kan, 2013
17q25.2	Zhang, 2015
1p12	Guichard, 2012
1p22.1	Guichard, 2012
1p33	Guichard, 2012
1p36.21	Guichard, 2012
1q21.3	Woo, 2009
1q22	Woo, 2009
1q42.11	Woo, 2009
20q12	Kan, 2013
20q13.12	Li, 2016
20q13.2	Dauch, 2016
21p11.2	Kan, 2013
22q13.1	Kan, 2013
3q26.31	Guichard, 2012
5p15.33	Kan, 2013
5p14.1	Kan, 2013
6p12.2	Kan, 2013
6p21.1	Chiang, 2008
6p21.2	Chiang, 2008
6p21.31	Chiang, 2008
6p21.32	Guichard, 2012
6p21.33	Chiang, 2008
6p22.3	Kwon, 2013
6p24.2	Kwon, 2013
6p25.3	Guichard, 2012
7p12.2	Guichard, 2012
7p15.2	Guichard, 2012
7p21.1	Guichard, 2012
7q31.2	Guichard, 2012
7q36.1	Woo, 2009
8p11.21	Roessler, 2012
8p12	Roessler, 2012
8p21.2	Roessler, 2012
8p22	Roessler, 2012
8q12.3	Kan, 2013
8q13.1	Kan, 2013
8q21.11	Kan, 2013
8q21.3	Kan, 2013
8q22.1	Kan, 2013
8q23.1	Kan, 2013
8q24.11	Woo, 2009
8q24.21	Woo, 2009
8q24.22	Woo, 2009
8q24.3	Kan, 2013
9p21.3	Guichard, 2012
9q32	Guichard, 2012

Chapter 6 Conclusion

6.1 Summary of research

Rapid advances in our understanding and treatment of cancer over the last decade has begun to realize the long-held goal of effecting personalized, precision cancer care for patients. Fueled by the combination of next-generation sequencing with targeted and immune therapy, an increasing number of patients are being treated based on a detailed genetic and molecular understanding of their particular tumor, not just a gross understanding of where the cancer originated. These advances extend to surgical oncology as well, with numerous studies demonstrating that the decision of when and how to resect a tumor requires assessment of more than just feasibility. However, for precision medicine to realize its true potential, clinicians will need access to tumor tissue for analysis in a real-time, repeatable, safe and cost-effective manner. While true for all cancers, such a "liquid biopsy" holds particular promise for gastrointestinal (GI) cancers due to their anatomic location and difficulty of obtaining tumor tissue by traditional biopsy methods.

Circulating tumor cells (CTCs), cells of tumor origin that circulate in the blood, are one such potential liquid biopsy for GI tumors. Historically, relatively few CTCs have been found from GI cancers when using general CTC assays such as CellSearch, but newer techniques may change that. Furthermore, while advances in single cell sequencing techniques make CTC molecular analysis possible, important methodologic and feasibility questions remain unanswered.

In this thesis, we demonstrate the development and validation of an assay for performing molecular analysis of circulating tumor cells (CTCs) as a biomarker in GI cancers. We first used

the microfluidic NanoVelcro CTC platform to developed and optimized GI tumor CTC assays for pancreatic and liver cancer through incorporation of tissue-specific markers for CTC capture and identification. This approach allowed us to overcome the limitations of traditional EpCAM-only CTC assays such as low CTC yield and false-positive CTC calls. Using our tissue-specific CTC assays, we conducted prospective studies testing the utility of CTC enumeration alone as a biomarker in pancreatic (n = 100 patients) and liver (n = 61 patients) cancer. We found that the use of a tissue-specific assay resulted in higher CTC counts than in previous studies, and that CTC counts correlated with increasing stage and that CTC enumeration may have diagnostic and prognostic utility as a biomarker in GI cancers.

To ensure that the cells we were calling CTC were in fact of tumor origin we used a modified NanoVelcro assay that allows for the complete isolation of the CTCs identified to perform whole genome amplification and molecular analysis. We first confirmed the tumor origin of pancreatic cancer CTCs by showing *KRAS* mutation subtypes matched between the primary tumor and the CTCs captured. Similar results were found by comparing low-resolution whole genome copy number profiles of primary liver cancers and CTCs, with CTCs recapitulating the somatic copy number changes of the primary tumor, supporting the use of CTCs as clinically relevant surrogates of the primary tumor.

We next sought to develop CTC sequencing as a biomarker for GI cancers. We developed a quality control assay that allowed us to determine how many CTCs we would need to isolate to reliably perform CTC molecular analysis. Using this metric, we modified our CTC sequencing method so that we only needed 3-5 CTCs as opposed to 10-20 CTCs needed for CTC analysis using the traditional method. We then used our optimized CTC sequencing protocol in a prospective study to test the feasibility and potential utility of CTC molecular analysis as a

biomarker in liver cancer. In a pilot study of 10 patients with HCC we found that our assay could serve as a liquid biopsy for patients, providing clinicians access to tumor tissue from the blood for applications like SCNA profiling. Additionally, we demonstrated the potential of classifying CTCs by their site of tumor origin based solely on the CTC SCNA profile. While the number of patients involved in our study is small, it serves to demonstrate the potential future applications of our assay.

In this thesis, we developed, optimized, and tested tissue-specific CTC assays for GI cancers in prospective studies (Chapter 2), and confirmed the tumor origin of the CTCs that we isolated (Chapter 3). We then determined the number of CTCs required for reliable and accurate CTC molecular analysis using a novel whole genome amplification approach (Chapter 4), and demonstrated the potential of our CTC sequencing assay in a pilot study (Chapter 5). The research described in this thesis serves 2 primary goals. The first is as a validated method that allows for the use of CTCs in liquid biopsy type applications for both HCC and PDAC patients. The second is as a framework for validating CTC assays and definitions to ensure the reliability and reproducibility of CTC sequencing results going forward.

References

1. Jemal A, Bray F, Center MM, et al: Global cancer statistics. CA Cancer J Clin 61:69-90, 2011

2. Mayer RJ, Venook AP, Schilsky RL: Progress against GI cancer during the American Society of Clinical Oncology's first 50 years. J Clin Oncol 32:1521-30, 2014

3. Figg WD, Newell DR: Pharmacologic biomarkers in the development of stratified cancer medicine. Clin Cancer Res 20:2525-9, 2014

4. Allard WJ, Matera J, Miller MC, et al: Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clinical cancer research : an official journal of the American Association for Cancer Research 10:6897-6904, 2004

5. Ashworth T: A case of cancer in which cells similar to those in the tumours were seen in the blood after death. Aust Med J 14:146-149, 1869

6. Parkinson DR, Dracopoli N, Petty BG, et al: Considerations in the development of circulating tumor cell technology for clinical use. J Transl Med 10:138, 2012

7. Satelli A, Mitra A, Brownlee Z, et al: Epithelial-mesenchymal transitioned circulating tumor cells capture for detecting tumor progression. Clin Cancer Res 21:899-906, 2015

8. Lustberg MB, Balasubramanian P, Miller B, et al: Heterogeneous atypical cell populations are present in blood of metastatic breast cancer patients. Breast Cancer Res 16:R23, 2014

9. Haber DA, Velculescu VE: Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer Discov 4:650-61, 2014

10. Baslan T, Kendall J, Ward B, et al: Optimizing sparse sequencing of single cells for highly multiplex copy number profiling. Genome Res 25:714-24, 2015

11. Liu S, Li N, Yu X, et al: Expression of intercellular adhesion molecule 1 by hepatocellular carcinoma stem cells and circulating tumor cells. Gastroenterology 144:1031-1041 e10, 2013

12. Alix-Panabieres C: EPISPOT assay: detection of viable DTCs/CTCs in solid tumor patients. Recent Results Cancer Res 195:69-76, 2012

13. Adams DL, Stefansson S, Haudenschild C, et al: Cytometric characterization of circulating tumor cells captured by microfiltration and their correlation to the cellsearch((R)) CTC test. Cytometry A 87:137-44, 2015

14. Tarin D, Price JE, Kettlewell MG, et al: Mechanisms of human tumor metastasis studied in patients with peritoneovenous shunts. Cancer Res 44:3584-92, 1984

15. Fidler IJ: Metastasis: guantitative analysis of distribution and fate of tumor embolilabeled with 125 I-5-iodo-2'-deoxyuridine. J Natl Cancer Inst 45:773-82, 1970

16. O'Flaherty JD, Gray S, Richard D, et al: Circulating tumour cells, their role in metastasis and their clinical utility in lung cancer. Lung Cancer 76:19-25, 2012

17. Luzzi KJ, MacDonald IC, Schmidt EE, et al: Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol 153:865-73, 1998

18. Jiao LR, Apostolopoulos C, Jacob J, et al: Unique localization of circulating tumor cells in patients with hepatic metastases. J Clin Oncol 27:6160-5, 2009

19. Pantel K, Deneve E, Nocca D, et al: Circulating epithelial cells in patients with benign colon diseases. Clin Chem 58:936-40, 2012

20. Rhim AD, Thege FI, Santana SM, et al: Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. Gastroenterology 146:647-51, 2014

21. Saxena M, Christofori G: Rebuilding cancer metastasis in the mouse. Mol Oncol 7:283-96, 2013

22. Gall TM, Jacob J, Frampton AE, et al: Reduced dissemination of circulating tumor cells with no-touch isolation surgical technique in patients with pancreatic cancer. JAMA Surg 149:482-5, 2014

23. Eloubeidi MA, Chen VK, Eltoum IA, et al: Endoscopic ultrasound-guided fine needle aspiration biopsy of patients with suspected pancreatic cancer: diagnostic accuracy and acute and 30-day complications. Am J Gastroenterol 98:2663-8, 2003

24. Bournet B, Selves J, Grand D, et al: Endoscopic ultrasound-guided fine-needle aspiration biopsy coupled with a KRAS mutation assay using allelic discrimination improves the diagnosis of pancreatic cancer. J Clin Gastroenterol 49:50-6, 2015

25. Navin NE: Cancer genomics: one cell at a time. Genome Biol 15:452, 2014

26. Milos PM: Emergence of single-molecule sequencing and potential for molecular diagnostic applications. Expert Rev Mol Diagn 9:659-66, 2009

27. Heitzer E, Auer M, Gasch C, et al: Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer research 73:2965-2975, 2013

28. Lohr JG, Adalsteinsson VA, Cibulskis K, et al: Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. Nat Biotechnol 32:479-84, 2014

29. Ni X, Zhuo M, Su Z, et al: Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci U S A 110:21083-8, 2013

30. Alix-Panabieres C, Pantel K: Challenges in circulating tumour cell research. Nat Rev Cancer 14:623-31, 2014

31. Hou S, Zhao L, Shen Q, et al: Polymer nanofiber-embedded microchips for detection, isolation, and molecular analysis of single circulating melanoma cells. Angewandte Chemie (International ed.in English) 52:3379-3383, 2013

32. Polzer B, Medoro G, Pasch S, et al: Molecular profiling of single circulating tumor cells with diagnostic intention. EMBO Mol Med 6:1371-86, 2014

33. Court CM, Ankeny JS, Hou S, et al: Improving pancreatic cancer diagnosis using circulating tumor cells: prospects for staging and single-cell analysis. Expert Rev Mol Diagn 15:1491-504, 2015

34. de Boer CJ, van Krieken JH, Janssen-van Rhijn CM, et al: Expression of Ep-CAM in normal, regenerating, metaplastic, and neoplastic liver. J Pathol 188:201-6, 1999

35. Khan MS, Tsigani T, Rashid M, et al: Circulating tumor cells and EpCAM expression in neuroendocrine tumors. Clin Cancer Res 17:337-45, 2011

36. Li YM, Xu SC, Li J, et al: Epithelial-mesenchymal transition markers expressed in circulating tumor cells in hepatocellular carcinoma patients with different stages of disease. Cell Death Dis 4:e831, 2013

37. Lin M, Chen JF, Lu YT, et al: Nanostructure embedded microchips for detection, isolation, and characterization of circulating tumor cells. Acc Chem Res 47:2941-50, 2014

38. Girgis MD, Olafsen T, Kenanova V, et al: Targeting CEA in Pancreas Cancer Xenografts with a Mutated scFv-Fc Antibody Fragment. EJNMMI Res 1:24, 2011

39. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2015. CA Cancer J Clin 65:5-29, 2015

40. Rahib L, Smith BD, Aizenberg R, et al: Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 74:2913-21, 2014

41. Yachida S, Jones S, Bozic I, et al: Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467:1114-1117, 2010

42. Hawes RH, Xiong Q, Waxman I, et al: A multispecialty approach to the diagnosis and management of pancreatic cancer. Am J Gastroenterol 95:17-31, 2000

43. Cooper M, Newman NA, Ibrahim AM, et al: Unnecessary tests and procedures in patients presenting with solid tumors of the pancreas. J Gastrointest Surg 17:1218-23, 2013

44. LeBlanc JK, Ciaccia D, Al-Assi MT, et al: Optimal number of EUS-guided fine needle passes needed to obtain a correct diagnosis. Gastrointest Endosc 59:475-81, 2004

45. Diehl SJ, Lehmann KJ, Sadick M, et al: Pancreatic cancer: value of dual-phase helical CT in assessing resectability. Radiology 206:373-8, 1998

46. Cameron JL, He J: Two thousand consecutive pancreaticoduodenectomies. J Am Coll Surg 220:530-6, 2015

47. Merkow RP, Bilimoria KY, Tomlinson JS, et al: Postoperative complications reduce adjuvant chemotherapy use in resectable pancreatic cancer. Ann Surg 260:372-7, 2014

48. Kurihara T, Itoi T, Sofuni A, et al: Detection of circulating tumor cells in patients with pancreatic cancer: a preliminary result. J Hepatobiliary Pancreat Surg 15:189-95, 2008

49. Khoja L, Backen A, Sloane R, et al: A pilot study to explore circulating tumour cells in pancreatic cancer as a novel biomarker. Br J Cancer 106:508-16, 2012

50. Cristofanilli M, Budd GT, Ellis MJ, et al: Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 351:781-91, 2004

51. Bryant KL, Mancias JD, Kimmelman AC, et al: KRAS: feeding pancreatic cancer proliferation. Trends Biochem Sci 39:91-100, 2014

52. Wang S, Liu K, Liu J, et al: Highly efficient capture of circulating tumor cells by using nanostructured silicon substrates with integrated chaotic micromixers. Angew Chem Int Ed Engl 50:3084-8, 2011

53. Chen JF, Ho H, Lichterman J, et al: Subclassification of prostate cancer circulating tumor cells by nuclear size reveals very small nuclear circulating tumor cells in patients with visceral metastases. Cancer 121:3240-51, 2015

54. Bidard FC, Huguet F, Louvet C, et al: Circulating tumor cells in locally advanced pancreatic adenocarcinoma: the ancillary CirCe 07 study to the LAP 07 trial. Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO 24:2057-2061, 2013

55. Nagrath S, Sequist LV, Maheswaran S, et al: Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature 450:1235-9, 2007

56. Yu M, Ting DT, Stott SL, et al: RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. Nature 487:510-513, 2012

57. Kamande JW, Hupert ML, Witek MA, et al: Modular microsystem for the isolation, enumeration, and phenotyping of circulating tumor cells in patients with pancreatic cancer. Anal Chem 85:9092-100, 2013

58. Rhim AD, Mirek ET, Aiello NM, et al: EMT and dissemination precede pancreatic tumor formation. Cell 148:349-61, 2012

59. Lafemina J, Katabi N, Klimstra D, et al: Malignant progression in IPMN: a cohort analysis of patients initially selected for resection or observation. Ann Surg Oncol 20:440-7, 2013

60. Tanno S, Nakano Y, Sugiyama Y, et al: Incidence of synchronous and metachronous pancreatic carcinoma in 168 patients with branch duct intraductal papillary mucinous neoplasm. Pancreatology 10:173-8, 2010

61. Von Hoff DD, Ervin T, Arena FP, et al: Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med 369:1691-703, 2013

62. Gillen S, Schuster T, Meyer Zum Buschenfelde C, et al: Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages. PLoS Med 7:e1000267, 2010

63. Shubert CR, Bergquist JR, Groeschl RT, et al: Overall survival is increased among stage III pancreatic adenocarcinoma patients receiving neoadjuvant chemotherapy compared to surgery first and adjuvant chemotherapy: An intention to treat analysis of the National Cancer Database. Surgery 160:1080-96, 2016

64. Riediger H, Keck T, Wellner U, et al: The lymph node ratio is the strongest prognostic factor after resection of pancreatic cancer. J Gastrointest Surg 13:1337-44, 2009

65. Mokdad AA, Minter RM, Zhu H, et al: Neoadjuvant Therapy Followed by Resection Versus Upfront Resection for Resectable Pancreatic Cancer: A Propensity Score Matched Analysis. J Clin Oncol, 2016

66. Court CM, Ankeny JS, Sho S, et al: Reality of Single Circulating Tumor Cell Sequencing for Molecular Diagnostics in Pancreatic Cancer. J Mol Diagn 18:688-696, 2016

67. Ankeny JS, Court CM, Hou S, et al: Circulating tumour cells as a biomarker for diagnosis and staging in pancreatic cancer. Br J Cancer 114:1367-75, 2016

68. Amin M, Edge S, Greene F: AJCC cancer staging manual, (ed 8th), 2016

69. Budczies J, Klauschen F, Sinn BV, et al: Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS One 7:e51862, 2012

70. Garrido-Laguna I, Hidalgo M: Pancreatic cancer: from state-of-the-art treatments to promising novel therapies. Nat Rev Clin Oncol 12:319-34, 2015

71. Donahue TR, Isacoff WH, Hines OJ, et al: Downstaging chemotherapy and alteration in the classic computed tomography/magnetic resonance imaging signs of vascular involvement in patients with pancreaticobiliary malignant tumors: influence on patient selection for surgery. Arch Surg 146:836-43, 2011

72. Hoffmann K, Kerner C, Wilfert W, et al: Detection of disseminated pancreatic cells by amplification of cytokeratin-19 with quantitative RT-PCR in blood, bone marrow and peritoneal lavage of pancreatic carcinoma patients. World J Gastroenterol 13:257-63, 2007

73. de Albuquerque A, Kubisch I, Breier G, et al: Multimarker gene analysis of circulating tumor cells in pancreatic cancer patients: a feasibility study. Oncology 82:3-10, 2012

74. Lozano R, Naghavi M, Foreman K, et al: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380:2095-128, 2012

75. El-Serag HB: Hepatocellular carcinoma. N Engl J Med 365:1118-27, 2011

76. Agopian VG, Harlander-Locke M, Zarrinpar A, et al: A novel prognostic nomogram accurately predicts hepatocellular carcinoma recurrence after liver transplantation: analysis of 865 consecutive liver transplant recipients. J Am Coll Surg 220:416-27, 2015

77. Llovet JM, Hernandez-Gea V: Hepatocellular carcinoma: reasons for phase III failure and novel perspectives on trial design. Clin Cancer Res 20:2072-9, 2014

78. Baccelli I, Schneeweiss A, Riethdorf S, et al: Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. Nat Biotechnol 31:539-44, 2013

79. Fan JL, Yang YF, Yuan CH, et al: Circulating Tumor Cells for Predicting the Prognostic of Patients with Hepatocellular Carcinoma: A Meta Analysis. Cell Physiol Biochem 37:629-40, 2015

80. Xu W, Cao L, Chen L, et al: Isolation of circulating tumor cells in patients with hepatocellular carcinoma using a novel cell separation strategy. Clinical cancer research : an official journal of the American Association for Cancer Research 17:3783-3793, 2011

81. Nel I, Baba HA, Ertle J, et al: Individual profiling of circulating tumor cell composition and therapeutic outcome in patients with hepatocellular carcinoma. Transl Oncol 6:420-8, 2013

82. Fan ST, Yang ZF, Ho DW, et al: Prediction of posthepatectomy recurrence of hepatocellular carcinoma by circulating cancer stem cells: a prospective study. Annals of Surgery 254:569-576, 2011

83. Yang ZF, Ngai P, Ho DW, et al: Identification of local and circulating cancer stem cells in human liver cancer. Hepatology 47:919-28, 2008

84. Chen L, Wu LL, Zhang ZL, et al: Biofunctionalized magnetic nanospheres-based cell sorting strategy for efficient isolation, detection and subtype analyses of heterogeneous circulating hepatocellular carcinoma cells. Biosens Bioelectron 85:633-40, 2016

85. Yu M, Bardia A, Wittner BS, et al: Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science 339:580-4, 2013

86. Giannelli G, Koudelkova P, Dituri F, et al: Role of epithelial to mesenchymal transition in hepatocellular carcinoma. J Hepatol 65:798-808, 2016

87. Hu L, Lau SH, Tzang CH, et al: Association of Vimentin overexpression and hepatocellular carcinoma metastasis. Oncogene 23:298-302, 2004

88. Yang MH, Chen CL, Chau GY, et al: Comprehensive analysis of the independent effect of twist and snail in promoting metastasis of hepatocellular carcinoma. Hepatology 50:1464-74, 2009

89. Chen JF, Ho H, Lichterman J, et al: Subclassification of prostate cancer circulating tumor cells by nuclear size reveals very small nuclear circulating tumor cells in patients with visceral metastases. Cancer, 2015

90. Li J, Chen L, Zhang X, et al: Detection of circulating tumor cells in hepatocellular carcinoma using antibodies against asialoglycoprotein receptor, carbamoyl phosphate synthetase 1 and pan-cytokeratin. PloS one 9:e96185, 2014

91. Mu H, Lin KX, Zhao H, et al: Identification of biomarkers for hepatocellular carcinoma by semiquantitative immunocytochemistry. World journal of gastroenterology : WJG 20:5826-5838, 2014

92. Buzzanco A, Gomez A, Rodriguez E, et al: Digital quantitation of HCC-associated stem cell markers and protein quality control factors using tissue arrays of human liver sections. Exp Mol Pathol 97:399-410, 2014

93. Mazzaferro V, Regalia E, Doci R, et al: Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N Engl J Med 334:693-9, 1996

94. Yao FY, Ferrell L, Bass NM, et al: Liver transplantation for hepatocellular carcinoma: expansion of the tumor size limits does not adversely impact survival. Hepatology 33:1394-403, 2001

95. Lencioni R, Llovet JM: Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis 30:52-60, 2010

96. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837-45, 1988

97. Libbrecht L, Severi T, Cassiman D, et al: Glypican-3 expression distinguishes small hepatocellular carcinomas from cirrhosis, dysplastic nodules, and focal nodular hyperplasia-like nodules. Am J Surg Pathol 30:1405-11, 2006

98. Sun YF, Xu Y, Yang XR, et al: Circulating stem cell-like epithelial cell adhesion molecule-positive tumor cells indicate poor prognosis of hepatocellular carcinoma after curative resection. Hepatology (Baltimore, Md.) 57:1458-1468, 2013

99. Schulze K, Gasch C, Staufer K, et al: Presence of EpCAM-positive circulating tumor cells as biomarker for systemic disease strongly correlates to survival in patients with hepatocellular carcinoma. International journal of cancer.Journal international du cancer 133:2165-2171, 2013

100. Sanchez-Lorencio MI, Ramirez P, Saenz L, et al: Comparison of Two Types of Liquid Biopsies in Patients With Hepatocellular Carcinoma Awaiting Orthotopic Liver Transplantation. Transplant Proc 47:2639-42, 2015

101. Ogle LF, Orr JG, Willoughby CE, et al: Imagestream detection and characterisation of circulating tumour cells - A liquid biopsy for hepatocellular carcinoma? J Hepatol 65:305-13, 2016

102. Satelli A, Li S: Vimentin in cancer and its potential as a molecular target for cancer therapy. Cell Mol Life Sci 68:3033-46, 2011

103. Armstrong AJ, Marengo MS, Oltean S, et al: Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers. Mol Cancer Res 9:997-1007, 2011

104. Carter L, Rothwell DG, Mesquita B, et al: Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. Nat Med 23:114-119, 2017

105. Court CM, Ankeny JS, Sho S, et al: Circulating Tumor Cells in Gastrointestinal Cancer: Current Practices and Future Directions, in Bentrem D, Benson AB (eds): Gastrointestinal Malignancies. Cham, Springer International Publishing, 2016, pp 345-376

106. Coumans FA, Ligthart ST, Uhr JW, et al: Challenges in the enumeration and phenotyping of CTC. Clin Cancer Res 18:5711-8, 2012

107. Coumans FA, Doggen CJ, Attard G, et al: All circulating EpCAM+CK+CD45- objects predict overall survival in castration-resistant prostate cancer. Ann Oncol 21:1851-7, 2010

108. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20, 2014

109. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-9, 2012

110. Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-2, 2010

111. Garvin T, Aboukhalil R, Kendall J, et al: Interactive analysis and assessment of single-cell copy-number variations. Nat Methods 12:1058-60, 2015

112. Lin F, Staerkel G: Cytologic criteria for well differentiated adenocarcinoma of the pancreas in fine-needle aspiration biopsy specimens. Cancer 99:44-50, 2003

113. Kalinich M, Bhan I, Kwan TT, et al: An RNA-based signature enables high specificity detection of circulating tumor cells in hepatocellular carcinoma. Proc Natl Acad Sci U S A 114:1123-1128, 2017

114. Ning L, Li Z, Wang G, et al: Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons. Sci Rep 5:11415, 2015

115. Zhang CZ, Adalsteinsson VA, Francis J, et al: Calibrating genomic and allelic coverage bias in single-cell sequencing. Nat Commun 6:6822, 2015

116. El-Heliebi A, Kroneis T, Zohrer E, et al: Are morphological criteria sufficient for the identification of circulating tumor cells in renal cancer? J Transl Med 11:214, 2013

117. Allan AL, Keeney M: Circulating tumor cell analysis: technical and statistical considerations for application to the clinic. J Oncol 2010:426218, 2010

118. Mendelsohn J: Personalizing oncology: perspectives and prospects. J Clin Oncol 31:1904-11, 2013

119. Gold B, Cankovic M, Furtado LV, et al: Do circulating tumor cells, exosomes, and circulating tumor nucleic acids have clinical utility?: a report of the association for molecular pathology. J Mol Diagn 17:209-24, 2015

120. Navin NE: The first five years of single-cell cancer genomics and beyond. Genome Res 25:1499-507, 2015

121. de Bourcy CF, De Vlaminck I, Kanbar JN, et al: A quantitative comparison of single-cell whole genome amplification methods. PLoS One 9:e105585, 2014

122. Gawad C, Koh W, Quake SR: Single-cell genome sequencing: current state of the science. Nat Rev Genet, 2016

123. Almoguera C, Shibata D, Forrester K, et al: Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. Cell 53:549-54, 1988

124. Zhao L, Lu YT, Li F, et al: High-purity prostate circulating tumor cell isolation by a polymer nanofiber-embedded microchip for whole exome sequencing. Adv Mater 25:2897-902, 2013

125. Huggett JF, Foy CA, Benes V, et al: The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. Clin Chem 59:892-902, 2013

126. Hamidi H, Lu M, Chau K, et al: KRAS mutational subtype and copy number predict in vitro response of human pancreatic cancer cell lines to MEK inhibition. Br J Cancer 111:1788-801, 2014

127. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al: ACMG clinical laboratory standards for next-generation sequencing. Genet Med 15:733-47, 2013

128. Ballantyne KN, van Oorschot RA, Muharam I, et al: Decreasing amplification bias associated with multiple displacement amplification and short tandem repeat genotyping. Anal Biochem 368:222-9, 2007

129. Van Loo P, Voet T: Single cell analysis of cancer genomes. Curr Opin Genet Dev 24:82-91, 2014

130. Tsiatis AC, Norris-Kirby A, Rich RG, et al: Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. J Mol Diagn 12:425-32, 2010

131. Soh J, Okumura N, Lockwood WW, et al: Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. PLoS One 4:e7464, 2009

132. Aoki Y, Hosaka S, Tachibana N, et al: Reassessment of K-ras mutations at codon 12 by direct PCR and sequencing from tissue microdissection in human pancreatic adenocarcinomas. Pancreas 21:152-7, 2000

133. Zong C, Lu S, Chapman AR, et al: Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338:1622-6, 2012

134. Yu Z, Lu S, Huang Y: Microfluidic whole genome amplification device for single cell sequencing. Anal Chem 86:9386-90, 2014

135. Chinen LT, de Carvalho FM, Rocha BM, et al: Cytokeratin-based CTC counting unrelated to clinical follow up. J Thorac Dis 5:593-9, 2013

136. Hou Y, Song L, Zhu P, et al: Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell 148:873-85, 2012

137. Millner LM, Strotman LN: The Future of Precision Medicine in Oncology. Clinics in Laboratory Medicine 36:557-573, 2016

138. Alix-Panabières C, Pantel K: Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. Cancer Discovery 6:479, 2016

139. Court CM, Ankeny JS, Hou S, et al: Improving pancreatic cancer diagnosis using circulating tumor cells: prospects for staging and single-cell analysis. Expert Review of Molecular Diagnostics 15:1491-1504, 2015

140. Maheswaran S, Sequist LV, Nagrath S, et al: Detection of Mutations in EGFR in Circulating Lung-Cancer Cells. New England Journal of Medicine 359:366-377, 2008

141. Ignatiadis M, Dawson SJ: Circulating tumor cells and circulating tumor DNA for precision medicine: dream or reality? Annals of Oncology 25:2304-2313, 2014

142. Wills QF, Mead AJ: Application of single-cell genomics in cancer: promise and challenges. Human Molecular Genetics 24:R74-R84, 2015

143. Hong B, Zu Y: Detecting Circulating Tumor Cells: Current Challenges and New Trends. Theranostics 3:377-394, 2013

144. Paterlini-Brechot P, Benali NL: Circulating tumor cells (CTC) detection: Clinical impact and future directions. Cancer Letters 253:180-204, 2007

145. Pinard R, de Winter A, Sarkis GJ, et al: Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics 7:216, 2006

146. Lasken RS, Egholm M: Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends in Biotechnology 21:531-535, 2003

147. Hughes S, Arneson N, Done S, et al: The use of whole genome amplification in the study of human disease. Progress in Biophysics and Molecular Biology 88:173-189, 2005

148. Huang L, Ma F, Chapman A, et al: Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annual Review of Genomics and Human Genetics 16:79-102, 2015

149. Court CM, Ankeny JS, Sho S, et al: Reality of Single Circulating Tumor Cell Sequencing for Molecular Diagnostics in Pancreatic Cancer. The Journal of Molecular Diagnostics 18:688-696, 2016

150. de Bourcy CFA, De Vlaminck I, Kanbar JN, et al: A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. PLoS ONE 9:e105585, 2014

151. Raghunathan A, Ferguson HRJ, Bornarth CJ, et al: Genomic DNA amplification from a single bacterium. Appl Environ Microbiol 71, 2005

152. Marcy Y, Ishoey T, Lasken RS, et al: Nanoliter Reactors Improve Multiple Displacement Amplification of Genomes from Single Cells. PLoS Genet 3:e155, 2007

153. Bergen AW, Qi Y, Haque KA, et al: Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance. BMC Biotechnology 5:24, 2005

154. Rhee M, Light YK, Meagher RJ, et al: Digital Droplet Multiple Displacement Amplification (ddMDA) for Whole Genome Sequencing of Limited DNA Samples. PLoS ONE 11:e0153699, 2016

155. van Beers EH, Joosse SA, Ligtenberg MJ, et al: A multiplex PCR predictor for aCGH success of FFPE samples. Br J Cancer 94:333-337, 2005

156. Garvin T, Aboukhalil R, Kendall J, et al: Interactive analysis and assessment of single-cell copy-number variations. Nat Meth 12:1058-1060, 2015

157. Möhlendick B, Bartenhagen C, Behrens B, et al: A Robust Method to Analyze Copy Number Alterations of Less than 100 kb in Single Cells Using Oligonucleotide Array CGH. PLOS ONE 8:e67031, 2013

158. Hayes DF, Cristofanilli M, Budd GT, et al: Circulating Tumor Cells at Each Follow-up Time Point during Therapy of Metastatic Breast Cancer Patients Predict Progression-Free and Overall Survival. Clinical Cancer Research 12:4218, 2006

159. Lasken RS, Egholm M: Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends Biotechnol 21, 2003

160. Barker DL, Hansen MS, Faruqi AF, et al: Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. Genome Res 14, 2004

161. Polzer B, Medoro G, Pasch S, et al: Molecular profiling of single circulating tumor cells with diagnostic intention. EMBO Molecular Medicine, 2014

162. Mullard A: NCI-MATCH trial pushes cancer umbrella trial paradigm. Nat Rev Drug Discov 14:513-515, 2015

163. Hadd AG, Houghton J, Choudhary A, et al: Targeted, High-Depth, Next-Generation Sequencing of Cancer Genes in Formalin-Fixed, Paraffin-Embedded and Fine-Needle Aspiration Tumor Specimens. The Journal of Molecular Diagnostics 15:234-247, 2013

164. Dancey Janet E, Bedard Philippe L, Onetto N, et al: The Genetic Basis for Cancer Treatment Decisions. Cell 148:409-420, 2012

165. MacConaill LE, Campbell CD, Kehoe SM, et al: Profiling Critical Cancer Gene Mutations in Clinical Tumor Samples. PLoS ONE 4:e7887, 2009

166. Koboldt DC, Steinberg KM, Larson DE, et al: The next-generation sequencing revolution and its impact on genomics. Cell 155:27-38, 2013

167. Xu X, Hou Y, Yin X, et al: Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148:886-95, 2012

168. Wang J, Fan HC, Behr B, et al: Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell 150:402-12, 2012

169. Baslan T, Kendall J, Rodgers L, et al: Genome-wide copy number analysis of single cells. Nature protocols 7:1024-1041, 2012

170. Adey A, Shendure J: Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res 22:1139-43, 2012

171. de Vega M, Lazaro JM, Mencia M, et al: Improvement of phi29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proc Natl Acad Sci U S A 107:16506-11, 2010

172. Dean FB, Hosono S, Fang L, et al: Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 99:5261-6, 2002

173. Dean FB, Nelson JR, Giesler TL, et al: Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11:1095-9, 2001

174. Esteban JA, Salas M, Blanco L: Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. J Biol Chem 268:2719-26, 1993

175. Treff NR, Su J, Tao X, et al: Single-cell whole-genome amplification technique impacts the accuracy of SNP microarray-based genotyping and copy number analyses. Mol Hum Reprod 17:335-43, 2011

176. Zhang K, Martiny AC, Reppas NB, et al: Sequencing genomes from single cells by polymerase cloning. Nat Biotechnol 24:680-6, 2006

177. Lasken RS, Stockwell TB: Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnol 7:19, 2007

178. Tu J, Guo J, Li J, et al: Systematic Characteristic Exploration of the Chimeras Generated in Multiple Displacement Amplification through Next Generation Sequencing Data Reanalysis. PLoS One 10:e0139857, 2015

179. Alsmadi O, Alkayal F, Monies D, et al: Specific and complete human genome amplification with improved yield achieved by phi29 DNA polymerase and a novel primer at elevated temperature. BMC Res Notes 2:48, 2009

180. Li J, Lu N, Shi X, et al: 1D-Reactor Decentralized MDA for Uniform and Accurate Whole Genome Amplification. Anal Chem 89:10147-10152, 2017

181. Kwon YM, Cox MM: Improved efficacy of whole genome amplification from bacterial cells. Biotechniques 37:40, 42, 44, 2004

182. Svec D, Andersson D, Pekny M, et al: Direct cell lysis for single-cell gene expression profiling. Front Oncol 3:274, 2013

183. Marcy Y, Ishoey T, Lasken RS, et al: Nanoliter reactors improve multiple displacement amplification of genomes from single cells. PLoS Genet 3:1702-8, 2007

184. Court CM, Hou S, Winograd P, et al: A novel multimarker assay for the phenotypic profiling of circulating tumor cells in hepatocellular carcinoma. Liver Transpl 24:946-960, 2018

185. Court CM, Ankeny JS, Sho S, et al: Circulating Tumor Cells Predict Occult Metastatic Disease and Prognosis in Pancreatic Cancer. Ann Surg Oncol 25:1000-1008, 2018

186. Sho S, Court CM, Winograd P, et al: Precision oncology using a limited number of cells: optimization of whole genome amplification products for sequencing applications. BMC Cancer 17:457, 2017

187. Borgstrom E, Paterlini M, Mold JE, et al: Comparison of whole genome amplification techniques for human single cell exome sequencing. PLoS One 12:e0171566, 2017

188. Huang L, Ma F, Chapman A, et al: Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annu Rev Genomics Hum Genet 16:79-102, 2015

189. Pinard R, de Winter A, Sarkis GJ, et al: Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics 7:216, 2006

190. Zuo Z, Chen SS, Chandra PK, et al: Application of COLD-PCR for improved detection of KRAS mutations in clinical samples. Mod Pathol 22:1023-31, 2009

191. Perez-Perez FJ, Hanson ND: Detection of plasmid-mediated AmpC beta-lactamase genes in clinical isolates by using multiplex PCR. J Clin Microbiol 40:2153-62, 2002

192. Panelli S, Damiani G, Espen L, et al: Towards the analysis of the genomes of single cells: further characterisation of the multiple displacement amplification. Gene 372:1-7, 2006

193. Zack TI, Schumacher SE, Carter SL, et al: Pan-cancer patterns of somatic copy number alteration. Nat Genet 45:1134-40, 2013

194. Hoadley KA, Yau C, Hinoue T, et al: Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 173:291-304 e6, 2018

195. Beroukhim R, Mermel CH, Porter D, et al: The landscape of somatic copy-number alteration across human cancers. Nature 463:899-905, 2010

196. Aceto N, Bardia A, Miyamoto DT, et al: Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell 158:1110-22, 2014

197. Miyamoto DT, Ting DT, Toner M, et al: Single-Cell Analysis of Circulating Tumor Cells as a Window into Tumor Heterogeneity. Cold Spring Harb Symp Quant Biol, 2017

198. Hata AN, Niederst MJ, Archibald HL, et al: Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. Nat Med 22:262-9, 2016

199. Heitzer E, Ulz P, Geigl JB, et al: Non-invasive detection of genome-wide somatic copy number alterations by liquid biopsies. Mol Oncol 10:494-502, 2016

200. Durinck S, Spellman PT, Birney E, et al: Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4:1184-91, 2009

201. Deng M, Bragelmann J, Kryukov I, et al: FirebrowseR: an R client to the Broad Institute's Firehose Pipeline. Database (Oxford) 2017, 2017

202. van der Maaten L: Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15:3221-3245, 2014

203. Krijthe JH: Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation, 2015

204. Kuhn M: caret: Classification and Regression Training, (ed R package version 6.0-77), 2017

205. Venables WNR, B.D.: Modern Applied Statistics with S (ed 4th), Springer-Verlag New York, 2002

206. Ridgeway G: gbm: Generalized Boosted Regression Models., (ed R package version 2.1.3), 2017

207. Meyer DD, Evgenia; Hornik, Kurt; Weingessel, Andreas; Leisch, Friedrich e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), (ed R package version 1.6-8). TU Wien, 2017

208. Warnes GRea: gplots: Various R Programming Tools for Plotting Data., 2016

209. Guichard C, Amaddeo G, Imbeaud S, et al: Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 44:694-8, 2012

210. Kan Z, Zheng H, Liu X, et al: Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res 23:1422-33, 2013

211. Chiang DY, Villanueva A, Hoshida Y, et al: Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 68:6779-88, 2008

212. Zhang J, Wen B, Cong W, et al: [Association of chromosome 17q copy number variation with overall survival of patients with hepatocellular carcinoma and screening of potential target genes]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 32:615-9, 2015

213. Kwon SM, Kim DS, Won NH, et al: Genomic copy number alterations with transcriptional deregulation at 6p identify an aggressive HCC phenotype. Carcinogenesis 34:1543-50, 2013

214. Roessler S, Long EL, Budhu A, et al: Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology 142:957-966 e12, 2012

215. Woo HG, Park ES, Lee JS, et al: Identification of potential driver genes in human liver carcinoma by genomewide screening. Cancer Res 69:4059-66, 2009

216. Schulze K, Nault JC, Villanueva A: Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol 65:1031-1042, 2016

217. Xu Y, Chen Q, Li W, et al: Overexpression of transcriptional coactivator AIB1 promotes hepatocellular carcinoma progression by enhancing cell proliferation and invasiveness. Oncogene 29:3386-97, 2010

218. Dauch D, Rudalska R, Cossa G, et al: A MYC-aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. Nat Med 22:744-53, 2016

219. Tong Z, Li M, Wang W, et al: Steroid Receptor Coactivator 1 Promotes Human Hepatocellular Carcinoma Progression by Enhancing Wnt/beta-Catenin Signaling. J Biol Chem 290:18596-608, 2015

220. Lu L, Han H, Tian Y, et al: Aurora kinase A mediates c-Myc's oncogenic effects in hepatocellular carcinoma. Mol Carcinog 54:1467-79, 2015

221. Li M, Wang W, Dan Y, et al: Downregulation of amplified in breast cancer 1 contributes to the anti-tumor effects of sorafenib on human hepatocellular carcinoma. Oncotarget 7:29605-19, 2016

222. Kumar-Sinha C, Chinnaiyan AM: Precision oncology in the age of integrative genomics. Nat Biotechnol 36:46-60, 2018

223. Webb S: The cancer bloodhounds. Nat Biotechnol 34:1090-1094, 2016

224. Sundaresan TK, Sequist LV, Heymach JV, et al: Detection of T790M, the acquired resistance EGFR mutation, by tumor biopsy versus noninvasive blood-based analyses. Clin Cancer Res, 2015

225. Park SM, Wong DJ, Ooi CC, et al: Molecular profiling of single circulating tumor cells from lung cancer patients. Proc Natl Acad Sci U S A 113:E8379-E8386, 2016

226. Gao Y, Ni X, Guo H, et al: Single-cell sequencing deciphers a convergent evolution of copy number alterations from primary to circulating tumor cells. Genome Res 27:1312-1322, 2017

227. Ciriello G, Miller ML, Aksoy BA, et al: Emerging landscape of oncogenic signatures across human cancers. Nat Genet 45:1127-33, 2013

228. Gerlinger M, Rowan AJ, Horswell S, et al: Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366:883-92, 2012

229. Molparia B, Nichani E, Torkamani A: Assessment of circulating copy number variant detection for cancer screening. PLoS One 12:e0180647, 2017