Application of Ultra-low Circulating Biomarkers and Single Cell Analysis from Solid Tumors with Single Molecule Resolution

by

Sarah Owen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2021

Doctoral Committee:

Associate Professor Sunitha Nagrath, Chair
Professor Erdogan Gulari
Professor Joerg Lahann
Professor Nils Walter

Sarah Owen

snowen@umich.edu

ORCID iD: 0000-0002-4102-7707

Dedication

To my parents, Dave and Linda Owen, who supported me throughout my entire, unconventional academic journey.

Acknowledgements

There are so many people and organizations that have helped and supported me throughout my Ph.D. I feel so grateful, and this thank you doesn't feel like enough.

First, I would like to thank my advisor, Dr. Sunitha Nagrath. I feel so fortunate to have worked with her over these years. She has always been incredibly supportive and compassionate. She not only allowed but encouraged me to explore different research projects. It was abundantly clear that the motivation for doing our work was never about self-recognition, but about making a meaningful impact in the medical field and to those suffering from aggressive and devastating diseases. Her perspective was the driving force that brought me to the University of Michigan for my PhD to work with her. I am so thankful I got work alongside her. Because of her support, I became a stronger and independent researcher. Thank you.

I would like to thank my committee members, Dr. Joerg Lahann, Dr. Erdogan Gulari and Dr. Nils Walter for their scientific insight and support. Their diverse expertise and outside perspectives helped me achieve the highest quality research products possible.

I would also like to thank the entire research lab. Sunitha's kindness extended to our entire lab culture. Despite the long days and odd hours, it was a place of support and encouragement, and homemade baked goods on the infamous "Sharing Table" with Molly Kozminsky, Ting-Wen Lo, Emma Purcell, Kaylee Smith, Zeqi Niu, Brittany Rupp, and Harrison Ball. The gatherings around the Sharing Table were home to both countless in-depth scientific discussions and conversations

filled with laughter. Some of my lab mates became some of my closest friends, and I couldn't be more grateful.

I would also like to thank all my collaborators. I was fortunate enough to work with many exceptional clinicians and hospital staff. I would like to thank Dr. Nithya Ramnath for all of her help and support to provide the clinical perspective and sitting through countless meetings to provide invaluable insight to our lab's work. I would also like to thank Dr. Vaibhav Sahai and Valerie Gunchick for their support on providing clinical samples. They both went above and beyond to provide hard to come by samples and share their clinical expertise.

I would like to thank all the patients who selflessly donated their blood for these studies. Your contribution to science is incredibly appreciated and I hope you believe I have used these samples well. I would also like to thank all my lab mates, friends and peers who trusted me to draw their blood and use for experiments and optimization.

I would also like to thank all the incredible funding sources and on-campus resources I used during my Ph.D. I would like to thank both the National Institute of Health (NIH) Microfluidics in Biomedical Science Training Program (MBSTP) and National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) for funding me during my research, allowing me focus my time, energy, and effort on my research. I feel so blessed to be the recipient of two prestigious awards. I would also like to thank the Rackham Graduate School (RGS) for the numerous Rackham Travel Grants enabling me to attend research conferences. These conferences allowed me to both highlight my research as well as learn so much groundbreaking research from others. I would also like to thank RGS for their support with the Rackham Research Grant, which provided a route to purchase new equipment intended to increase research throughput and reproducibility. I was also very fortunate to have access to world-class

resources on campus. I would like to thank the Lurie Nanofabrication Facility (LNF), which houses all the tools necessary to create the devices used in my research projects. I would also like to thank the Single Cell Resource, which has sophisticated single cell technologies enabling my research goals to become a reality. Further, I would also like to thank Biointerfaces Institute. I am so grateful that my lab space was housed within this research institute. Being in an open floorplan surrounded with people from such diverse research expertise organically facilitated meaningful research conversations, collaboration, and ultimately friendships that wouldn't have otherwise happened.

I would also like to thank the friends and support I received in and outside of the university. They made Ann Arbor feel like home. I would like to thank Katie Leguizamon for always being my go-to gal. Thank you for taking the chance and getting lunch at Songbird when we all moved to Michigan. Thank you for countless coffees, "arb walks" and Gilmore Girls episodes. I would like to thank Stephen Vicchio for never asking, but rather informing me I was coming along for any and all events. Thank you for convincing me to go on crazy adventures like a day trip to Sleeping Bear Dunes. Thank you Eshita Khera for always being up for a work session filled with a lot of coffee that far too often turned to a fun hangout. I have made so many great friends in graduate school that have supported and encouraged me, and the list is too long to name everyone, but I would especially also like to thank you Marcus Deloney, Lucy Pinheiro, Alison Banka, and Vyas Ramasubramani.

I would like to thank all my friends from Huron Hills Church, especially Katie Cavanagh, Beth Kuhl, and Amanda Dowker. From the weekend retreats to Muskegon, MI, to the summer evenings at Silver Lake to long weekends at the Red River Gorge learning how to rock climb and eating half a large pizza for dinner, to Bruce Peninsula, there are too many adventures to count. I

would also like to thank my friends at Blue Lion Fitness, especially Natalie Dostie, who held me accountable to show up to the 7am workouts, especially on deadlift day.

I would like to thank all of my friends back home for their encouragement. Thank you for always picking up where we left off every time I came home to visit. I would especially like to thank Ali Tucker, Marley Olson, Lindsay Ulbrickson and Hannah Bauer.

Lastly, and most importantly, I would like to thank my parents for their continual support and encouragement and going above and beyond to let me follow a truly one of a kind educational path. It started when they overanalyzed to make sure I ended up in the best preschool. Then they spent numerous hours in the car driving me to participate in academic programs outside of my local school. In elementary school they drove me once a week to participate in an accelerated program, Qwest. For four years of junior high, they packed the car full for carpool to go to a choice school, Environmental and Adventure School (EAS). Sixth-ninth grade gave me once in a lifetime opportunities, to go camping with 140 other students three times a year, including a trip to Mexico, learn practical life skills, contribute to society with week-long service projects. EAS gave me some of my best, lifelong friends. Thank you to my parents for letting me use one of the cars to commute to Bellevue Community College, to spend my junior and senior years of high school there while simultaneously earning my Associates degree. It has been a journey. Thank you for all your love and support.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables xii

List of Figures xiv

List of Appendices xvii

Abstract xviii

Chapter 1 Introduction 1

1.1 Cancer Statistics 1
1.2 Current Cancer Care 2
1.2.1 Traditional therapy 2
1.2.2 Advanced Molecular therapies 3
1.3 Liquid Biopsies 6
1.3.1 CTCs 6
1.3.2 Circulating tumor DNA (ctDNA) 8
1.3.3 Extracellular Vesicles (EVs) and Exosomes 8
1.4 Isolation and Analysis Techniques for Liquid Biopsies 9
1.4.1 CTCs in Tumor Monitoring 9
1.4.2 CTC Culture and Expansion 19
1.4.3 EV Isolation and Profiling 22
1.4.2 Digital PCR Technologies 23
1.5 Hypotheses and scope of this thesis 27

Chapter 2 Simultaneous Single Cell Gene Expression and EGFR Mutation Analysis of Circulating Tumor Cells Reveals Distinct Phenotypes in NSCLC 29

2.1 Publication Information 29
2.2 Abstract 29
2.3 Introduction 30
2.4 Materials and Methods 33
2.4.1 Cell Culture 33
2.4.2 RNA Extraction & cDNA Synthesis 33
2.4.3 Experimental Protocol for Labyrinth (Patient Sample Processing) 33
2.4.4 Immunofluorescent Staining and CTC Enumeration 35
2.4.5 Mutation Detection using dPCR 37
2.4.6 Statistical Analysis 37
2.5 Results 38
2.5.1. Single Cell Co-Analysis Workflow 38
2.5.2. Validation of EGFR Mutation Detection Using Digital PCR (dPCR) 41
2.5.3. Validation of Robust Single Cell Gene Expression and EGFR Mutation Co-Analysis 44
2.5.4. Demographics of Patient Cohort 48
2.4.5. Single Cell Characterization of NSCLC CTCs 49
2.6 Discussion 52
2.7 Conclusion 54
Chapter 3 EGFR Mutations Carried in Extracellular Vesicle-derived Cargo Mirror Disease Status in Metastatic Non-small Cell Lung Cancer 55

3.1 Abstract 55
3.2 Introduction 55
3.3 Materials and Methods 57
3.3.1 Cell Culture 57
3.3.2 Patient Enrollment 58
3.3.3 Blood Preparation 58
3.3.4 Extracellular Vesicle Isolation using Differential and Ultracentrifugation 58
3.3.5 Extracellular Vesicle Quantification 59

3.3.6 RNA Extraction and Reverse Transcription 59
3.3.7 RT-qPCR 59
3.3.8 Cell-free DNA Extraction 60
3.3.9 Mutation Detection by Droplet Digital PCR (ddPCR) 60
3.3.10 Protein Extraction, Quantification, and Normalization 61
3.3.11 Statistical Analysis 63
3.4 Results 63
3.4.1 EV Cargo Carries Mutations from Cells of Origin 63
3.4.2 Study Design and Patient Cohort 66
3.4.3 Quantification and Longitudinal Monitoring of EGFR Mutations in EV-RNA 71
3.4.4 Detection and Longitudinal Monitoring of EGFR Mutations in EV-protein 77
3.4.5 Correlation between Sensitizing EGFR Mutations in EV-RNA, EV-protein, and ctDNA Circulating Biomarkers 78
3.5 Discussion 81
Chapter 4 Synergistic Analysis of Circulating Tumor Cells Reveal Prognostic Signatures in Treatment-Naïve Metastatic Pancreatic Cancer Patients 84
4.1 Abstract 84
4.2 Introduction 84
4.3 Materials and Methods 86
4.3.1 Cell Culture 86
4.3.2 Labyrinth Fabrication 87
4.3.3 Patient Enrollment 87
4.3.4 Sample Collection & CTC Isolation 88
4.3.5 Immunofluorescence and CTC Enumeration 88
4.3.6 RNA Extraction 89
4.3.7 cDNA Synthesis, and droplet digital PCR (ddPCR) 90
4.3.8 Affymetrix microarray processing and data analysis 91
4.3.9 Statistical analysis 91
4.4 Results 91
4.4.1 Patient Characteristics 91
4.4.2 Detection of CTCs for Phenotypic and Molecular Profiling 94
3.4.3 CTCs Display Heterogeneous EMT Phenotypes and KRAS Mutation Burden 95

4.4.4 Monitoring Treatment-related CTC Dynamics 98
4.4.5 Evaluating the Influence of Treatment Type on CTC Dynamics 105
4.5 Discussion 106
Chapter 5 Natural killer Cells Offer Potential Therapeutic Approach to Control Metastasis through Circulating Tumor Cell Vulnerabilities 109
5.1 Abstract 109
5.2 Introduction 110
5.3 Materials and Methods 111
5.3.1 Cell Culture 111
5.3.2 Sample Collection and Labyrinth Processing 112
5.3.3 Immunofluorescence and Fluorescent Microscopy 112
5.3.4 CTC Analysis and Enumeration 114
5.3.5 Quantitative Fluorescent Intensity Data Export 114
5.3.6 NK Cytotoxicity Assays and data analysis 115
5.3.7 Statistical Analysis 116
5.4 Results 116
5.4.1 Assessment of NK Susceptibility in Cancer Cells Based on NK Activating and Inhibiting Ligand Expression 116
5.4.2 Patient Demographics and Study Design 120
5.4.3 Circulating Tumor Cell Characterization Workflow 121
5.4.4 CTC Characterization and Inter-patient Heterogeneity 122
5.4.5 Correlation between Expression of NK Activating and Inhibiting Ligands and Patient Clinical Characteristics 125
5.5 Discussion 129
Chapter 6 Conclusions and Future Directions 132
6.1 Research Summary 132
6.1.1 Development of Single Cell Gene Expression and Mutation Detection Workflow 132
6.1.2 Mutations Carried in EV Cargo as a Prognostic Signature 133
6.1.3 Prognostic Signatures of Patient Survival in PDAC CTCs 134
6.1.4 Correlation Between NK Sensitivity and Patient Outcomes 135
6.2 Limitations and Future Directions 136

6.2.1 Multiplexed Mutation Detection using ddPCR 136
6.2.2 Mutation Burden in EV-cargo: Beyond Correlation to Prognostic Signature 138
6.2.3 CTCs to Predict Treatment Response 139
6.2.4 NK Sensitivity: from Correlation to Clinical Impact 140
6.3 Conclusions 142
Appendices 143
References 162

List of Tables

Table 1.1 Summary of dPCR systems. 26

Table 2.1 96 gene panel used for BioMark™ HD qPCR for cell lines. 40

Table 2.2 96 gene panel used for BioMark™ HD qPCR for Lung CTC Patient Samples. 41

Table 2.3 Lung cancer patient demographics. 48

Table 2.4: Lung cancer patient demographics. 48

Table 3.1 TaqMan gene expression assay for RT-qPCR. 60

Table 3.2 TaqMan EGFR mutation detection assays. 61

Table 3.3 Antibodies used for EV-protein detection and EGFR mutation profiling. 62

Table 3.4 Patient demographics and clinical information. 67

Table 3.5 Summary of the patient EV-RNA and EV-protein tested for L858R, T790M and exon 19 del mutations. 69

Table 3.6 Patient EGFR mutation ddPCR and extracellular vesicle characterization. 70

Table 4.1 Patient demographics, treatment, and clinical status. 93

Table 4.2 Complete patient demographics. 94

Table 4.3 Key differentially expressed genes in pre-treatment and on-treatment CTC samples. 102

Table 4.4 Differentially expressed genes from top 20 deregulated pathways in pre-treatment and on-treatment CTC samples based on patient survival. 104

Table 5.1 Staining panels for NK activating and inhibiting ligands. 113

Table 5.2 Patient clinical characteristics. 120

Table 5.3 Patient demographics. 121

Table D.1 Summary of mutation detection assays validated for different sample types using the RainDrop dPCR system (RainDance Technologies).. 153

Table D.2 Common mutations found in solid cancers of the breast, lung, pancreas and prostate 158

List of Figures

Figure 1.1 Single cell profiling reveals intra-sample subpopulations that are lost in bulk sample profiling..... 15

Figure 1.2 Survey of liquid biopsy landscape used in this thesis. 28

Figure 2.1 Circulating tumor cell sample processing schematic for single cell analysis. 39

Figure 2.2 Validation of EGFR mutation detection using RainDrop PlusTM dPCR system..... 43

Figure 2.3 Validation of single cell workflow for gene expression and EGFR mutation analysis using lung cancer cell lines. (Caption shown on next page)..... 46

Figure 2.4 Patient characteristics and CTC analysis..... 49

Figure 2.5 Patient tumor-matched mutations detected in CTCs. 51

Figure 3.1 EGFR mutations carried in RNA and protein from cell line derived EVs H1975 (L858R/T790M), H3255 (L858R), H1650 (exon 19 del). 64

Figure 3.2 Workflow to profile EVs and ctDNA, including isolation, RNA characterization, and protein characterization. 66

Figure 3.3. Changes in EGFR mutation burden in EV-RNA mirror disease status..... 72

Figure 3.4 Clinical data timeline mapped to the mutant EV-RNA at each visit..... 76

Figure 3.5 Detection of EGFR mutations in EV-protein. 77

Figure 3.6 Comparison of sensitizing mutations, exon 19del and L858R, in EV-RNA, EV-Protein, and ctDNA. 79

Figure 4.1 Multipronged CTC enumeration and transcriptome analysis from metastatic PDAC patients. 95
Figure 4.2 CTC enumeration, phenotype and mutant KRAS burden profiling. 96
Figure 4.3 Optimization of KRAS G12/13 ddPCR assay. 98
Figure 4.4 CTC dynamics based on patient survival. 100
Figure 4.5 CTC differential gene expression and pathway analysis based on patient survival.. 101
Figure 4.6 Effect of treatment on CTC transcriptome. 105
Figure 5.1 Expression of NK activating and inhibiting ligands and NK sensitivity. 119
Figure 5.2 Project overview for evaluating NK sensitivity of CTCs. 122
Figure 5.3 Representative immunofluorescent images of CTCs evaluated for NK inhibitors or NK activators from NSCLC patients. 123
Figure 5.4 Inter-patient heterogeneity of NK activating and inhibiting ligand expression in CTCs across patients and time points. 124
Figure 5.5 CTCs display heterogeneous expression profiles of NK activating and NK inhibiting ligands. 126
Figure 5.6 Correlation between NK ligand expression and clinical metrics. 128
Figure 6.1 Cancer cells exhibit different NK sensitivity. 141
Figure A.1 EGFR L858R titration……...144
Figure A.2 T790M titration……..144
Figure A.3 Exon19 titration……145
Figure A.4 EGFR L858R representative single cells……...146
Figure A.5. EGFR T790M representative single cells……147
Figure A.6 EGFR Exon 19 deletion representative single cells……..147

Figure B.1 EGFR point mutations, L858R and T790M, detected in lung cancer cell line derived EV-RNA. ……148

Figure C.1 Comparison of plasma preparations for EV analysis…….149

Figure D.1 PIK3CA dPCR plots from breast cancer cell lines……154

Figure D.2 PIK3CA H1047R dPCR plot from single T47D cells……...154

Figure D.3 TP53 R273H dPCR plots from lung cancer cell line H1975…….155

Figure D.4 Multiplex EGFR mutation detection…….157

List of Appendices

Appendix A: Development of digital PCR (dPCR) for EGFR Mutant Screening from Single Cells 144

Appendix B: Development of dPCR for EGFR Mutant Screening from EVs 148

Appendix C: Comparison of EVs using Different Plasma Generation Methods 149

Appendix D: Development of digital PCR (dPCR) for Oncogenic and Tumor Suppressor Mutations Broadly Found Across Solid Tumor Types 152

D.1 Development of dPCR Methods for SNP Detection 152

D.2 Considerations for Multiplexed Mutation Detection 156

Abstract

Circulating tumor cells (CTCs) and extracellular vesicles (EVs) carry tumor-derived cargo in the circulation and provide the potential for minimally invasive patient monitoring. These circulating biomarkers have both been shown to be critical for cancer progression and metastasis. However, their clinical utility has been limited due to the rarity of CTCs in the blood, whereas EVs, while abundant, only carry miniscule amounts of cancer-derived cargo. This thesis shows the development and application of novel characterization methods for profiling these two circulating biomarkers, which both require ultra-sensitive approaches for characterization.

First, a single cell analysis workflow was developed for mutation and gene expression profiling from the same cell. This multistep process was subsequently used for mutation detection using droplet digital PCR (ddPCR) and gene expression profiling using the Fluidigm Biomark HD. Epidermal growth factor receptor (EGFR) mutations were used as a model to optimize and validate this workflow. EGFR mutations confer either sensitivity or resistance to tyrosine kinase inhibitors (TKIs). Significant cell-to-cell variability of total EGFR expression and relative mutant-to-wildtype allele expression were observed, even within a clonal cell line. This workflow was applied to CTCs from non-small cell lung cancer (NSCLC) patients. CTCs were isolated using the Labyrinth, a high-throughput microfluidic CTC isolation technology. CTCs showed variable mutation status and revealed intra-patient heterogeneity. All CTCs carrying point mutations showed higher wildtype than mutant allele expression, which may be a mechanism for avoiding cell death from TKIs.

Next, I present the first longitudinal patient monitoring for mutations carried in EVs. Mutation profiling in EVs is a new field, with most groups reporting combining EV-derived RNA and DNA with circulating tumor DNA. Here I screened EV-derived RNA (EV-RNA) using ddPCR and protein (EV-protein) using western blot for the presence of EGFR mutations from NSCLC patients. The change in EGFR mutation carried in EV-RNA compared to the previous visit mirrored disease trajectory. Neither EV-protein presence nor change showed a correlation with disease trajectory or detection with EV-RNA.

Then, I applied a multi-pronged CTC profiling method to compare patient survival in metastatic pancreatic adenocarcinoma. CTCs were isolated from treatment naïve patients as well as after beginning therapy. CTCs were screened for their abundance and phenotype, KRAS mutation status and expression, and transcriptome profiles. Patients with poor survival showed higher levels of total CTCs and epithelial-to-mesenchymal transition (EMT)-CTCs at pre-treatment samples. Mutant KRAS was higher at both time points in deceased patients. Significant deregulation of genes and related pathways was observed including those known to be involved in RAS signaling.

Lastly, I evaluated the expression of natural killer (NK) cell activating and inhibiting ligands on CTC-derived cell lines and freshly isolated CTCs. NK cells have been implicated in controlling and halting metastasis, however, little is known about the regulation balancing activating and inhibiting signals. I showed the NK-sensitivity varied based on the relative expression of these ligands. In freshly isolated CTCs, I showed that increased expression of human leukocyte antigen (HLA)-A/B/C and decreased expression of CADM1 were correlated with disease progression.

Overall, in this thesis I developed novel circulating biomarker analysis methods. Using these methods I showed that patients can be longitudinally monitored for signatures of treatment resistance and patient survival, which I posit can lead to improved patient care.

Chapter 1 Introduction

1.1 Cancer Statistics

Cancer remains the second leading cause of death in the United States (US). Cancer-associated mortality is responsible for nearly 600,000 deaths in the US alone each year, accounting for over 20% of total deaths[1]. One of the main factors is the prevalence of cancer, approximately 40% of the population will be diagnosed with cancer at some point in their lifetime, with only a 5-year survival probability of 67%[2]. The high overall mortality rate of cancer is due to late detection. Metastatic disease is responsible for over 90% of cancer-related deaths each year[3]. While late diagnosis is seen across many cancer types, this is highlighted by two of the most lethal cancers, lung cancer and pancreatic cancer[4,5]. These two cancers suffer from late diagnosis and demonstrate similar disease outcome patterns.

Lung cancer is the second most common cancer diagnosis, but accounts for the most cancer-associated deaths each year. Each year there are approximately 230,000 new cases, and nearly 143,000 deaths[5]. The high death rate, with a 19.4% 5-year survival rate, is largely due to the late diagnosis. Only 16% of lung cancer cases are diagnosed while the disease is still localized, but these patients have a relatively good prognosis with a 5-year survival rate of 57%. Once the cancer has started to spread into the lymph nodes, 22% of new cases, or has overtly metastasized, 57% of new cases, the 5-year survival drops to 31% and 5%, respectively[5]. Non-small lung cancer (NSCLC) is one subset of lung cancer, which accounts for approximately 75% of all lung cancer cases[6] and is further investigated in this thesis.

Pancreatic cancer has less cases each year than lung cancer, with around 57,000 new cases, and 46,000 deaths annually[4]. However the 5-year survival rate is only 9%. Similar to other cancer types, pancreatic cancer survival probability is dependent on how early it is diagnosed. Only about 10% of new cases are caught when the tumor is localized, and those patients have a modest 5-year survival rate of 37%. Patients who are diagnosed when it has spread to the lymph nodes, 29% of cases, drop to a 5-year survival of 12%. For the 53% of patients diagnosed with metastatic disease, there is an abysmal 5-year rate of only 2.9%[4,7]. Pancreatic cancer is particularly difficult to diagnose early, due to the presentation of generic symptoms, which could be indicative of many other morbidities in combination with the pancreas being difficult to image due to the being hidden by other internal organs[4,8–10]. Pancreatic ductal adenocarcinoma accounts for approximately 95% of pancreatic cancer cases[11], and is further studied in this thesis.

1.2 Current Cancer Care

Current standards of care are vastly different across cancer types[12]. Some cancers have very limited treatment options and are limited to the conventional options of surgery, radiation, and chemotherapy. While other cancer types, such as breast and lung cancer, have numerous additional treatment options such as targeted therapy and immunotherapy, which are designed to target a specific molecular feature of the cancer cells to kill them.

1.2.1 Traditional therapy

Traditional therapies include surgery, radiation, and chemotherapy. These approaches have several limitations and disadvantages. Only a small fraction of cancer patients qualify for surgical removal of the tumor. The disease must be caught at an early stage before it has started to spread; in pancreatic cancer, this only accounts for about 15% of patients[13]. Surgery offers a potentially

curative treatment for cancer by removing the tumor, however many people eventually relapse, due to the incomplete removal of the entirety of the primary tumor or the presence of micrometastases[14,15].

Radiation is used to locally target and destroy cancer cells and shrink the tumor. Radiation breaks the DNA inside the cancer cells, and inhibits the ability for cell growth and division. While the radiation is directed to the tumor, there is the potentially to damage the surrounding healthy tissue[16].

Chemotherapy is a cytotoxic compound that leads to arrest in cell division. While the other two traditional therapies are localized, chemotherapy is a systemic treatment. It can kill cancer cells throughout the body. There are many different chemotherapies, and can be delivered using many different mechanisms but commonly include oral and intravenous (IV) options. Due to the rapid proliferation of cancer cells, these cells are particularly effected by chemotherapy. However, other rapidly dividing cells in the body can are also vulnerable to chemotherapy[17,18].

Sometimes a combination of these approaches are used to treat cancer[19]. For example, radiation is used as a neoadjuvant therapy to shrink a tumor before surgical removal. Or radiation will be used as an adjuvant therapy after surgical removal, either to the primary tumor site or sites where the tumor is likely to metastasize, and micrometastases may be present, with the goal to prevent recurrence. Similarly chemotherapy can be given before or after treatment with radiation or surgery[20].

1.2.2 Advanced Molecular therapies

In recent years there have been many therapies designed to target specific features of the tumor. The use of these therapies aren't as universal compared to traditional therapies, but they do offer improved outcomes for a subset of patients, and can have reduced side effects due to reduced

off target damage. The main two categories of molecular therapies are targeted therapy and immunotherapy[17]. Target therapy directly targets a specific characteristic of the cancer cell and leads to cancer death. Conversely, immunotherapy leverages the immune system to destroy the cancer cells.

1.2.2.1 Targeted therapy

Targeted therapy were the first advanced molecular therapies. These targeted therapies can fall into three main categories: (1) small molecule, (2) antibody, or (3) antibody drug conjugates (ADC)[17]. Small molecule targeted therapies target specific protein residues on the tumor cells. They include tyrosine kinase inhibitors (TKIs), commonly used in mutated non-small cell lung cancer patients (NSCLC) with various mutations such as epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK) and ROS1 rearrangements[17]. Antibodies bind to a specific antigen on the cancer cell. Depending on the mechanism of action, the antibody can either be characterized as a targeted therapy or an immunotherapy[21]. If the antibody directly leads to cancer cell death, it is considered a targeted therapy, however, if the antibody causes or enhances an immune response, it is considered an immunotherapy. One of the most well-known targeted therapy antibodies is trastuzumab (Herceptin), which is used for Her2-positive breast cancer[17].

EGFR TKI therapy

While there are many forms of targeted therapy, one of the most relevant targeted therapies for the work presented in this thesis are EGFR TKIs. The power of targeted therapy was first shown in NSCLC, where patients harboring activating mutations in EGFR responded to TKIs compared to patients with wildtype (unmutated EGFR)[22]. EGFR TKIs are small molecule drugs that bind to and inhibit EGFR. Activating EGFR mutations, such as the two most common exon 21 L858R point mutation and exon 19 deletion, lead to the tyrosine kinase existing in an active state, therefore

have increased binding to adenosine triphosphate (ATP), resulting in increased tumor cell growth and cell survival. However, first and second generation TKIs – erlotinib, gefitinib, afatinib - have increased binding to EGFR harboring activating mutations compared to ATP, and can inhibit kinase activity[23]. All three of these treatments are approved as first line therapy for EGFR mutant NSCLC patients. While a majority of patients show a strong initial response to treatment, many develop resistance in as little as nine months, commonly through a secondary EGFR mutation T790M[24–26]. The mechanism of resistance has been debated. One theory is that the T790M mutation leads to an increased binding affinity between EGFR and ATP. The T790M mutation has a 10-15 times stronger binding affinity to ATP compared to activating EGFR mutations, therefore the TKIs aren't able to displace the EGFR-bound ATP[23]. It is also possible that the EGFR mutations result in differences in protein stability, and that TKI therapy result in rapid protein degradation for the activating mutant EGFR but not the T790M mutation. There are now third generation TKIs – such as osimertinib- which is also effective against T790M mutant EGFR[27].

1.2.2.2 Immunotherapy

Immunotherapy has been an important component of cancer treatment for many years, but recent advances have shifted immunotherapy to the forefront of oncology[28]. Immunotherapies are treatments designed to induce or enhance an immune response[29], either through immune stimulation or through the use of *in vitro*-derived immune system components[30]. Critical recent progress includes the approval of durvalumab in lung cancer (April 2020), and initial success on applying adoptive T cell therapy in solid tumors[31]. However, T cell-based therapies are often ineffective because of the low major histocompatibility complex (MHC) expression environment[32]. MHC class I molecules are expressed on all nucleated cells in the human body, and are a marker of "self". They are essential for antigen presentation during T cell activation and

effector T cell function. This downregulation or loss of MHC class I is found in 40-90% of cancers such as bladder, breast, lung and colorectal cancer[33,34]. Nature kill (NK) cell immunotherapies, which don't rely on MHC class I presentation, are currently in development and clinical trials[30].

1.3 Liquid Biopsies

A major hurdle, especially in molecular and immunotherapies, is determining which patients will likely respond to a given therapy. Currently, a patient must get a tumor biopsy, an invasive procedure, in which a small portion of the tumor is removed for molecular profiling. The biopsy sample will guide the treatment course for that patient. However, through a routine blood draw, it is possible to obtain disease information for patient monitoring[35,36]. The most common analytes of liquid biopsies are circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), and extracellular vesicles, however other markers have been studied as well, such as CA19-9 for pancreatic cancer. CA19-9 primarily suffers from low sensitivity and high false negative rates[9,37–39].

1.3.1 CTCs

CTCs are tumor cells which have left the tumor microenvironment and have entered into the bloodstream. During the onset of metastasis, tumor cells undergo the epithelial to mesenchymal transition (EMT), a transient phenotypic change, in which the cells become migratory[40–42]. The cells intravasate into the bloodstream and circulate throughout the body. While in the bloodstream these cancer cells are referred to as CTCs[43,44]. A fraction of these circulating tumor cells (CTCs) survive the circulation long enough to reach a secondary site, and extravasate into the tissue. Those tumor cells undergo the reverse process, mesenchymal to epithelial transition (MET) where the cells return to a proliferative state and can create a metastatic tumor[45,46].

EMT drives cancer metastasis through a transient dedifferentiation process where cells acquire stem-like characteristics and an aggressive phenotype. EMT cells lose their tight cell-to-cell adhesion and have an increase in cell motility[47]. They also have increased survival through anti-apoptosis mechanisms and chemoresistance[47,48].

EMT is characterized by an elongated cell morphology, the downregulation of epithelial genes, such as E-cadherin (E-cad), and upregulation of mesenchymal genes, such as vimentin, N-cadherin (N-cad), Zeb1 and Zeb2[48]. E-cad is a cell-surface protein that facilities cell-cell interaction between epithelial cells. Reduced expression of E-cad promotes an elongated cell morphology and increased migration capacity. Zeb1 and Zeb2 are part of a subset of transcription factors that mediate EMT/MET and act as E-cad repressors. Other transcription factors that induce EMT include TWIST1, FOXC1, Snail, and Slug. The transcription factors are also associated with stem cell-like properties, including increased survival through drug resistance and anti-apoptosis mechanisms[49]. Vimentin is an intermediate filament that is expressed at sites of cellular elongation and is associated with a migratory phenotype[50]. Because EMT is linked to metastasis and drug resistance, two of the most important causes of patient mortality, understanding and controlling EMT is crucial in improving patient prognosis.

This metastatic cascade is a highly inefficient process. A majority of these CTCs are cleared from the body due to either the physical or biological environment the CTCs endure while in circulation. Within 24 hours, only 0.1% of CTCs remain viable while in circulation[51], with only 0.01% of CTCs ultimately leading to a metastasis[51,52]. While in the circulation, CTCs must survive harsh sheer stressed and immune surveillance, such as from NK cells[51,53,54].

1.3.2 Circulating tumor DNA (ctDNA)

ctDNA are short DNA fragments that make up a small portion, typically on 0.1-10%, of total cell-free DNA content (cfDNA)[55]. cfDNA is generated during apoptosis-mediated cell death. cfDNA fragments tend to be 70-200 base pairs in length[44,56]. The use of ctDNA for clinical tracking suffers from many limitations including poor stability in the blood, rapid clearance and degradation, and the bias for dead cells[44]. Groups have reported that ctDNA is rapidly cleared from the circulation through multiple mechanisms including nuclease degradation and renal clearance. The reported clearance rates have ranged from 10 minutes to 6 hours, and a half live of mintues to 1-2 hours[57–59].

Despite these challenges, ctDNA is the most widely used liquid biopsy biomarker. Likely due to the minimal specialized equipment for isolation and testing. ctDNA has been used across many cancer types including pancreatic[60,61] and lung[24,62,63] cancers.

Currently, the The Cobas EGFR Mutation Test v2, is the only FDA approved liquid biopsy test for EGFR mutation testing of ctDNA from NSCLC patients. While the ctDNA test is approved for EGFR mutation screening in NSCLC patients, it still has many limitation. While the test has high sensitivity and specificity for activing EGFR mutations, L858R and exon 19 del, it still suffers from only a modest sensitivity and specificity for the resistance mutation T790M, at 58.4% and 80.4% respectively. Due to this limitation, the test is only approved as a "rule-in" test for the presence of these mutations, but a repeat biopsy is recommended to confirm a negative result[64].

1.3.3 Extracellular Vesicles (EVs) and Exosomes

Extracellular vesicles (EVs) are a broad class of membrane-bound vesicles released by cells, typically ranging from 30-2,000nm in size[65,66]. They are divided into three main subclasses – (1) exosomes, (2) microvesicles, and (3) apoptotic bodies[67]. EVs are released by most cells types,

cancer and healthy, including immune cells, platelets, and stem cells[67]. They are functional vesicles which carry information from their cells of original including proteins, RNA and DNA.

Exosomes, nanovesicles, approximately 30-150 nm in diameter, and are bound in a lipid bilayer remain one of the most commonly studied extracellular vesicles (EVs) in cancer[68–70]. Exosomes are actively secreted from live cells with prepackaged, intact cargo, offering current information about a patient's cancer[71–73]. Several groups have shown an increase in circulating exosomal concentration in plasma from various cancer types, including pancreatic ductal adenocarcinoma, breast, and ovarian cancers, when compared to healthy controls[74–76]. The increase in tumor-derived exosomal load offers unique potential of exosomes and other EVs to serve as biomarkers in cancer monitoring, diagnostics, and prognostics. Additionally, the lipid bilayer of the exosomes provides excellent protection from proteases and nucleases for their cargo of protein, microRNA (miRNA), messenger RNA (mRNA), and DNA fragments, rendering them stable in both in circulation as well as in collected specimens. It is demonstrated previously that the stability of exosomes allows for the transfer of information from parent cell to distant cells[77–80]. Additionally, several groups have characterized specific miRNA and protein compositions in exosomes that have the potential as diagnostic and prognostic markers[76,81–84].

1.4 Isolation and Analysis Techniques for Liquid Biopsies

1.4.1 CTCs in Tumor Monitoring

While CTCs are extremely rare in the body, typically on the order of 10 CTCs/mL, the presence and abundance of CTCs has been identified as an independent prognostic factor for patient outcome. CTCs were first discovered in 1869[56,85] during an autopsy, and the concept of a "seed and soil" theory for metastisis was first proposed in 1889 by Stephen Paget, a European

Surgeon[51]. However, it wasn't until the development of sophisticated isolation technologies, most commonly through microfluidics, that CTC research really expanded.

Currently, CellSearch® is the only FDA-approved system for CTC enumeration for use in metastatic breast, colorectal and prostate cancers[86–91]. The CellSearch® system uses an antibody-labeled ferrofluid to capture CTCs based on the expression of the epithelial cell adhesion molecule (EpCAM). The CellSearch® system has demonstrated that CTC abundance can be used to serially monitor patients. Originally, CTCs were simply enumerated and their abundance was correlated with disease prognosis. This demonstrated that patients with elevated CTCs at baseline had worse prognosis than those with lower CTCs levels. However, as patients began being monitored throughout treatment, it became clear that a single time point was insufficient for patient monitoring, and patient could be further stratified based on if and how their CTC levels changed throughout treatment. For example, patients with low CTCs at baseline and throughout treatment had good prognosis; similarly patients with elevated CTCs at baseline but showed decreasing CTC abundance throughout treatment also had improved prognosis. Conversely, not only patients with sustained elevated CTC levels from baseline throughout treatment, but also patients who ad low CTC levels at baseline but increased CTC burden throughout treatment had poor prognosis[90]. While CTC burden, and changes in CTC burden have shown clinical utility, it has been shown that further characterization of CTCs can yield important insights into tumor monitoring.

While the CellSearch® has demonstrated its clinical value, it also has many disadvantages. First, the system relies on the expression of EpCAM on the CTCs for capture. It has been shown that EMT CTCs have down-regulated or even no EpCAM expression, and may not be captured by the system. Additionally, the system requires the use of special blood collection tubes, which

preserve the cells with fixatives. The cells are no longer viable and it limits the other analyses that can be performed such as transcriptomic or functional profiling.

1.4.1 CTC Isolation

One of the main challenges of studying CTCs is their rarity in the blood, typically on the order of 1-10CTCs/mL of blood surrounded by millions of white blood cells (WBCs) and billions of red blood cells (RBCs)[87,92,93]. Due to this rarity of CTCs in the blood, sophisticated isolation strategies have been developed to facilitate CTC isolation and characterization. While there many approaches, they fit into two main categories: immunoaffinity and label-free. Initial work in this space relied on macro approaches by leveraging the physical (size, density, deformability) or biologic (expression of cancer or CTC specific surface markers) differences between CTCs and other blood cells.

Physical macro approaches include filtration systems, such as ISET (Isolation by Size of Epithelial Tumor cells) and the Nucleopore assay, take advantage of the size difference of CTCs (15-20μm) from white blood cells (WBCs) (7-12 μm) and red blood cells (RBCs) (5-6μm) to selectively capture CTCs, and density-gradient techniques, such as Ficoll gradient centrifugation and Oncoquick separate mononucleated cells (WBCs and CTCs) from the other blood components[87,93–95]. Biologic macro approaches, known as immunoaffinity, have been employed to separate CTCs based on their unique surface markers. The most widely used capture target is the epithelial cell adhesion marker (EpCAM). The CellSearch System discussed above relies on macro immunoaffinity based CTC capture. The labeled CTCs are then captured using a magnetic field for enumeration by staining for cancer markers[96,97].

Many researchers have tried to improve on these isolation systems by using microfluidic devices[98]. The CTC-chip[99,100], Herringbone (HB)-chip[101], graphene oxide (GO)-chip[102] and

Oncobean[103] are examples of immunoaffinity microfluidic devices that contain structures coated with antibodies for CTC capture. These methods have high CTC purity due to antibody-antigen specificity. A major limitation of these devices is that they cannot efficiently capture EMT CTCs due to their reduced EpCAM expression.

Another type of microfluidic technology, inertial microfluidics, is a label-free isolation technology that exploits of size differences between CTCs and blood cells to focus them at different channel cross-sections in a size dependent manner[96,104]. Focusing occurs through the combination of lift forces, imposed by the shear gradient, and the Dean forces, which arise from secondary flows due to curvature in a channel. Size-based separation in the inertial sorter is based on established physical phenomena occurring in curved microfluidic channels. For (nearly) neutrally buoyant particles, such as cells, the dominant forces acting on the particles are the inertial lift forces, F_L, and the Dean drag force, F_D. The inertial lift forces, F_L, are a combination of shear-induced and wall-induced lift forces. The shear-induced inertial lift force points outwards from the center of the channel towards the channel walls due to the shear gradient. The wall-induced inertial lift points from the channel walls towards the center of the channel[105–107].

The Dean drag force arises from secondary flows that are due to fluid being confined to a curved channel geometry. The fast moving fluid in the center of the channel has a higher inertia and tends to continue along its trajectory rather than following the curvature of the channel. This displaces the fluid near the outer wall towards the top and bottom of the channel, resulting in two symmetric, counter-rotating vortices, called Dean vortices[105,108].

A particle flowing through a curved channel will experience both Dean drag forces and inertial lift forces. The combination of these forces determines where particles of a given size will

focus to a single position within the channel[105,108]. By collecting different outlet streams from the inertial sorter, it is possible to isolate different sized particles[105].

Inertial microfluidics channels are frequently spirals, although introducing corners into the geometry has been shown to enhance focusing. This improves CTC recovery and WBC depletion, as highlighted by the development of the Labyrinth within the Nagrath Lab[109]. The size dependent separation of cells enables heterogeneous CTCs, such as EMT cells to be efficiently captured. These methodologies tend to be higher throughput, but sample purity tends to be lower compared to immunoaffinity methods. Hybrid technologies have also been developed to pre-enrich the sample using size-based separation following by immunoaffinity capture using magnetic beads coated in anti-EpCAM antibodies such as the CTC-iChip[110] and the Integrated device[111].

1.4.1.2 CTC Analysis and Characterization

After isolation, CTCs can be enumerated and characterized for phenotype, gene expression and mutation status. CTC enumeration has been used as an independent indicator of patient prognosis for both disease-free and overall survival in many clinical studies[42,89,91,112,113]. CTCs are enumerated most commonly through immunocytochemistry (ICC) to stain for different cell markers.

Beyond enumeration, ICC can be used to investigate intra- and inter-patient heterogeneity using additional markers. CTCs have been found to not only span the entire EMT spectrum[87,114], but also have been found to travel as both single cells and as CTC-clusters[46,113,115]. The presence and size of CTC-clusters is being evaluated as a potential prognostic indicator[113]. The proliferative capacity of CTCs has been assessed using the proliferation marker Ki67[116] indicating not all CTCs have metastatic potential, as some are apoptotic or damaged.

With the emergence of improved isolation and characterization technologies, more sophisticated analysis has been used to characterize CTCs. CTC samples have been investigated for the presence of specific mutations or using gene expression analysis to create a predictive CTC profile of patient outcome[87,117,118]. These approaches leveraged either polymerase chain reaction (PCR) or RNA sequencing (RNA-seq). While informative, bulk analysis of pooled CTCs can impair observing the differential gene expression between different CTC subpopulations (Figure 1.1)[87,94]. The use of developing single cell analysis techniques has enabled high sensitivity analysis solutions to circumvent this. These gene expression techniques include single cell RNA-seq[119] (scRNA-seq), Fluidigm C1[120], and the 10x Genomics systems, however many other single cell analysis techniques also exist, including DNA sequencing (DNA-seq), assay for transposase-accessible chromatin sequencing (ATAC-seq), and chromatin immunoprecipation sequencing (ChIP-seq). Single cell analysis techniques have gained rapid popularity, however they still suffer from many technical limitations. While single cell techniques promise the ability to identify intra-patient heterogeneity, there are still many limitations which can arise due to the ultra-low input of material from a single cell. These techniques and their limitations are further discussed below.

However, one of the biggest challenges of single cell CTC analysis is isolating intact single cells[121]. While, there are many methods for isolating single cells, many rely on random-chance and don't allow the precise selection of particular cells of interest or involve time consuming methods for cell picking. These approaches include Drop-Seq[122], Seq-Well[123], 10x Genomics, the Fluidigm C1[120], and the DEPArray[124].

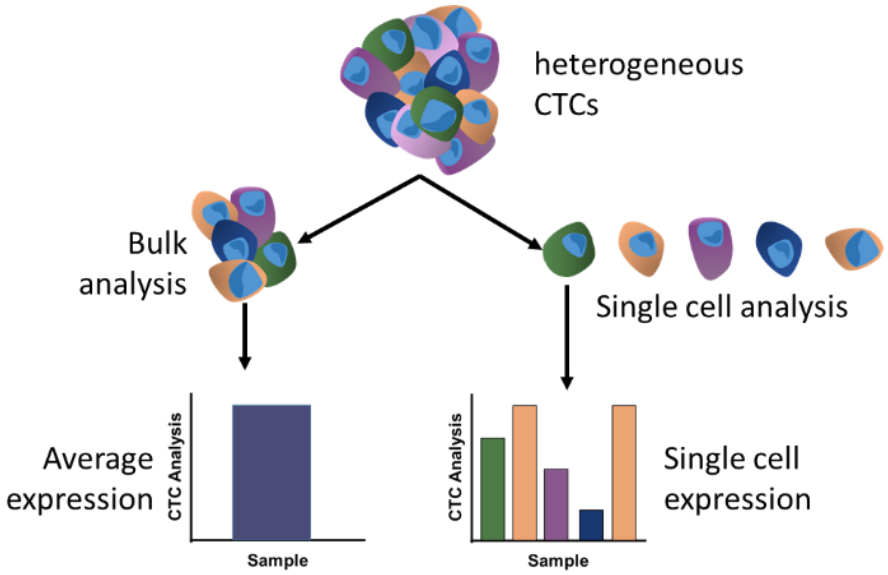


Figure 1.1 Single cell profiling reveals intra-sample subpopulations that are lost in bulk sample profiling.

Single cell DNA-seq (scDNA-seq)

scDNA-seq is a method to analyze the entire genome from of multiple single cells to identify and track the presence of heterogeneous genetic abnormalities of different cancer cells; this analysis includes different somatic mutations such as substitution and single nucleotide polymorphisms (SNPs), insertions and deletions (INDELs), copy number variants (CNV) and rearrangements using next-generation sequencing (NGS)[125–127]. These different mutations could be used to identify the transient clonal heterogeneity of the tumor as well as identify both driver mutations as well as targetable mutations[121,125,128]. A single cell contains less than 10 picogram of DNA and therefore requires whole genome amplification (WGA) prior to analysis using sequencing or other techniques such as microarray[126]. WGA relies on nucleic amplification techniques PCR or multiple displacement amplification (MDA), and sometimes a combination of both. Both techniques suffer from notable limitations, such as non-uniform coverage, bias for cytosine and guanosine-rich sequences (GC-bias), preferential allelic amplification, allelic

dropout, and base pair copy errors. Each of these limitations can introduce mutations and artifacts in the sample[121], however certain error types are more prevalent in each WGA method. PCR-based WGA tends to achieve uniform genome amplification, however there is limited coverage. Due to the uniform amplification, PCR-based WGA methods are well suited for CNV studies, but due to the limited coverage, is not suitable for small variants such as SNPs. Conversely, MDA-based WGA achieves superior coverage, but non-uniform amplification. Therefore, MDA-based WGA is best suited for small variant detection, such as SNPs, but is not suitable for CNV analysis due to the possibility of false positive arising due to non-uniform amplification[126]. However, methods have been developed to minimize the limitations of each of these approaches such as Multiple annealing and loop-based amplification cycles (MALBAC), which consists of pre-amplification using strand-displacement followed by amplification PCR, which achieved both high genome coverage and low allele dropout[126,128].

Lohr et al. showed in prostate cancer patients CTCs whole genome amplification using MDA and applied whole exosome sequencing (WES) to achieve the most cost effect method of a data-rich dataset, with the goal to study the select regions of the genome mostly likely to provide the most meaningful insights. This was applied to a cohort of five prostate cancer patients with high CTC counts (20 or more CTCs) to identify SNPs. In one patient the authors compared the mutation profiles of CTCs and metastatic tumors to mutations identified in nine spatial regions in the preserved primary tumor that had been resected a few years prior. One region of the primary tumor contained a mutational landscape most similar to those of the CTCs and metastatic tumor. Further, CTCs contained nine of the ten early lineage mutations (trunk mutations)[129]. Ni et al. showed both SNPs and INDELs as well as CNV profiles in lung cancer CTCs using MALBAC to allow both analysis types. Using WES, Ni et al. showed large intra- and inter-patient heterogeneity

of SNPs and INDELs. Using whole genome sequencing (WGS) the authors found a striking observation; despite the large heterogeneity of SNPs and INDELs across CTCs, the majority of CTCs from a single patient retained similar CNV profiles. The authors then found that while SNPs and INDELs found in CTCs are transient and change throughout a patient's treatment, the CNV profiles remain consistent[130]. De Luca et al. isolated single CTCs from metastatic breast cancer patients using the DEPArray and identified large inter- and intra-patient heterogeneity of mutations found in these CTCs as well as changes in the detectable mutations before and during treatment, with many only being present at one of the time points[131].

Single cell RNA-seq (scRNA-seq)

scRNA-seq has emerged as one of the most prevalent methods of single cell CTC analysis. scRNA-seq tries to quantifies the transcriptional profile of the entire cell. While scDNA-seq can identify genetic abnormalities, scRNA-seq can identify aberrantly expressed genes, and using computational techniques, it is possible to infer some genetic mutations and CNVs[132]. In this technique, the RNA is reverse transcribed (RT) into complementary DNA (cDNA), amplified using whole transcriptome amplification (WTA) and then sequenced using NGS[127]. Some of the major concerns with using scRNA-seq include RNA capture and RT efficiency, degradation, and changes in cell state during processing[132]. In order to improve processing, many techniques have been developed to combine single cell isolation with the cell lysis and RT steps, such as Seq-Well[123], Drop-Seq[122], and 10x Genomics[132].

Ramsköld et al. developed a full-length scRNA-Seq method, Smart-Seq, to improve read coverage across the entire length of the transcripts. The authors applied this technique to melanoma CTCs and identified gene expression profiles that could serve as biomarker candidates in melanoma. Of particular interest, the authors identified five downregulated genes implicated in immune evasion, aggressiveness, and metastsis[133].

Miyamota used scRNA-seq to profile CTCs from prostate cancer patients. The authors uncovered inter- and intra-patient heterogeneity. The authors compared the CTC profiles to those found in primary and metastatic tumor sites. The authors also detected androgen receptor splice variants expressed across cells and strikingly even within the same CTC[119].

While scRNA-seq has enabled intra-patient heterogeneity to be studied, it still suffers from limitations such a gene-dropout and incomplete transcript coverage based on "end-counting" scRNA-seq techniques[134]. Therefore, Rodriguez-Meira *et al* developed a combination approach to evaluate both gene expression profiles using scRNA-seq in parallel with mutation analysis[134]. This downstream analysis technique would be a powerful tool to gain both transcriptomic and genomic information from the same cell and overcome the low cell numbers of CTCs.

Microarray

Microarrays were the first method to allow rapid and thorough transcriptome coverage and analysis[135]. Mircoarrays contain patterned probes of DNA sequences that are complementary to the RNA transcripts in the sample based on the known genetic sequence of the sample host. The RNA sample is labeled with a fluorescent dye, loaded onto the microarray cartridge where it interacts with and binds to the complementary probes. The fluorescent intensity can then be read and used to measure gene expression[135].

Smirnov *et al.* used mircoarray profiling to show the first global gene expression profiles of CTCs. CTCs were isolated from metastatic colorectal, metastatic prostate, and metastatic breast cancer patients. Candidate genes were identified to distinguish cancer patients from healthy donors, but also to distinguish between cancer types[136].

Obermayr *et al.* screened gene expression profiles from CTCs from female patients across various cancer types (primary breast, ovarian, cervical, or endometrial cancer, and advanced breast cancer) to identify a six gene panel to detect CTCs using an RT-qPCR method[137].

Based on the observed CTC heterogeneity, it is evident future single cell studies are required to classify which CTC subpopulations can survive the entirety of the metastatic cascade. Ultimately, these analysis techniques can be applied to identify the presence and relative quantity of these highly aggressive CTCs during treatment which can be used as a prognostic factor and indicator of treatment efficacy.

1.4.2 CTC Culture and Expansion

Due to the rarity of CTCs in the blood, only a few CTCs can be isolated from a patient at a time. This limits the number and type of assays, such as drug screening, that can currently be conducted. To circumvent this limitation, the ability to reliably expand CTCs is an emerging field of study. To date, only a handful of groups have successfully expanded CTCs *ex vivo*. The main efforts towards CTC culture have focused on using a combination of soluble growth factors, co-culture with tumor-associated cells, hypoxia or spheroid culture to encourage cell proliferation[138–141]. These efforts, described below, have not shown a reliable method to consistently expand CTCs and is an ongoing research effort in the field. Below is a summary of the in vitro CTC expansion reports including methods and characterization approaches.

Cayrefourcq *et al*. reported the establishment of CTC cultures from two of 71 colon cancer patients. These two patients had abnormally high CTC frequencies (>300CTCs/mL). Culture conditions used hypoxia, non-adherent spheroid culture and media supplemented with growth factors including insulin, epithelial growth factor (EGF), L-Glutamine, fibroblast growth factor (FGF) and 2% serum. From these two patients, one sample was established into a cell line that eventually was grown in adherent culture conditions under normoxia. This CTC derived cell line was used for copy number variation (CNV) analysis and mouse xenograph model. The second sample was grown for about 2 months but eventually died off[139].

Recently, follow up work using the aforementioned CTC line was published[142]. This long term CTC-derived colorectal cancer cell line was the first of its kind in this cancer type[142]. The CTC line was established from a metastatic, treatment naïve patient. This cell line, CTC-MCC-41, did not carry mutations in the oncogenes: PIK3CA, AKT, PTEN, KRAS, or NRAS, but did contain a BRAF V600E mutation - consistent with both the primary tumor and lymph node metastasis. However, Smit *et al*. found that the CTC line expressed a unique gene expression profile not typically seen in established colorectal cancer cell lines. This transcriptional profile may reveal unique signatures that allow some CTCs to survive in circulation[142]. They also showed that established CTCs could be used in the same way established, commercially available cancer cell lines are used for functional and molecular analysis.

Zhang *et al*. successfully cultured 14 of 19 lung CTCs for short-term on-chip expansion. CTCs were captured using an immunoaffinity microfluidic chip. After CTC isolation, the chip was filled with collagen I and Matrigel, and GFP-labeled CAFS. CTCs were expanded on-chip for 7 days before the contents of the chip were released and cultured in a well plate for an addition 7-14 days[138]. Zhang *et al*. showed rapid expansion of the lung CTCs using this method, with an average 54-fold increase in cell number. These cultured CTCs were compared to the mutation status of the primary tumor in cancer-related genes. Matched mutations were identified in three of the eight paired tumor and cultured CTC samples[138].

Hamilton *et al.* established two small cell lung cancer (SCLC) CTC lines[143]. Interestingly, the CTC lines were established from spheroids generated due to overgrowth of a 2D adherent culture. The CTC lines were profiled for EMT and stem cell markers as well as growth factors. The CTC lines expressed both epithelial and mesenchymal marks, and both expressed high levels of E-Cadherin.

Cegan *et al.* showed the first culture of urinary bladder cancer CTCs[144]. The CTCs were isolated on using MetaCell filtration, and the CTCs were cultured on the filter for 14 days under normoxic conditions[144].

Yu *et al*. established six breast CTC lines from 45 patients using non-adherent spheroid culture, using serum free media supplemented with EGF and FGF under hypoxia. For three of the patients, multiple cultures were established from CTCs isolated at multiple time points during treatment. Of particular interest, all the patients that generated multiple CTC lines were also progressing on treatment. Three of the CTC lines were screened for drug sensitivities to multiple anticancer therapies. CTC line drug sensitivities demonstrated high concordance with the clinically observed drug sensitives to these treatments[140]. This work highlights the potential of using cultured CTCs to predict treatment response for a patient.

Khoo *et al.* demonstrated short-term expansion of breast CTCs for up to 14 days using a microwell-based culture method for suspension cluster culture. Copy number variation (CNV) of six key driver mutations was studied in the CTC cultures using DNA fluorescence *in situ* hybridization (FISH)[145].

Koch *et al.* established a breast cancer CTC line from a metastatic estrogen receptor (ER) positive breast cancer patient[146]. The CTC line retained ER positivity, matching the primary tumor and fresh CTCs. The patient showed high CTC levels at the time point the CTC line was established, with 1,547 CTCs/mL as determined used the CellSearch system.

Rivera-Baez *et al.* established three long term CTC lines from locally advanced, treatment naive pancreatic patients, which were used for functional and xenograft studies[147]. CTC lines were established using a fibronection coated culture area, and used media only supplemented with 10% fetal bovine serum (FBS). All three CTCs lines showed heterogeneous distribution of EMT

phenotypes. Interestingly, compared to other CTC expansion attempts there was a very high success rate, 30% (3/10 samples). Further, there were no striking unique features about the three successful cultures compared to the other samples in terms of CTC abundance or phenotype.

Soler *et al.* derived nine sequential CTC lines from a colon cancer patient throughout different time points of receiving therapy: before treatment, after second and third relapses[148]. The CTC lines expressed high levels of epithelial phenotype makers, but shared heterogeneous levels of EMT and cancer stem cell markers.

1.4.3 EV Isolation and Profiling

There are numerous EV isolation technologies, however, they fall into three main categories: (1) ultracentrifugation[23], microfluidic devices, and commercially available kits[149,150]. This thesis uses ultracentrifugation for EV isolation, therefore is additional information is provided here. Ultracentrifugation is currently considered the gold standard for EV isolation, however requires the use of specialized equipment, an ultracentrifuge, which is capable of producing forces of 100,000xg to pellet the EVs[151]. Ultracentrifugation relies on multiple rounds of centrifugation to pellet and wash the EVs, which is extremely time and labor intensive and suffers from low yield[151,152]. However it is compatible the wide range of downstream analysis approaches[149].

While EVs carry both protein and nucleic acids from their cells of origin, microRNA (miRNA) remains the mostly widely studied. Several groups have characterized specific miRNA and protein compositions in exosomes that have potential as diagnostic and prognostic markers[76,81–84].

Screening EVs for oncogenic mutations is a new but growing field therefore the current state of the field is discussed in greater detail here. Thakur *et al.* reported NSCLC cell line derived exosomes harbored driver EGFR mutations in their exosomal double-stranded DNA that reflected

the original cells[153]. Similarly, groups have found EGFRvIII variants in exosomal DNA from the cerebrospinal fluid of glioblastoma patients and KRAS mutations in exosomal DNA from early stage pancreatic cancer patients[154–156]. Chen *et al.* reported IDH1 mutation in mRNA in EVs derived from serum and cerebrospinal fluid in glioma patients[157]. Recently, Krug *et al.* analyzed a cohort of advanced NSCLC patients at baseline for activating and resistance EGFR mutations. Using exosomal RNA in combination with ctDNA increased the detection of activating and resistance EGFR mutations compared to using ctDNA alone[62]. Similarly, Castellanos-Rizaldos *et al.* identified the EGFR T790M mutation in NSCLC patients using a combined exosomal RNA/DNA and cfDNA isolation protocol[158]. Dong *et al.* used a microfluidic chip to isolate EVs from two NSCLC, which were screened for EGFR T790M mutation and ROS1 rearragnement[159]. This was the first report to show a proof of concept of using mutations carried in EVs for patient monitoring.

1.4.2 Digital PCR Technologies

Digital PCR (dPCR) revolutionized the capabilities of PCR-based assays by dramatically improving assay sensitivity compared to standard PCR and real-time PCR techniques[160,161]. dPCR is unique to other PCR techniques, in that it divides the PCR reaction into many partitions, each acting as an individual, smaller PCR reaction, either based on predesigned chambers or through microfluidics to produce droplets in a water-in-oil emulsions. Dividing the PCR reaction into smaller, discrete reactions allows for each partition to act as its own reaction vessel. Each partition can then be measured based on fluorescent intensity to detect PCR amplification-positive and PCR amplification-negative partitions[162,163]. The PCR positive droplets can then be counted sample concentration can be determined without a standard curve.

This method of counting relies on the governing rules of droplet loading, following the Poisson distribution. The partitions are designed to be in great excess to the target gene copy number in the sample so that during the sample portioning, some of the droplets will contain no target molecules, some will contain one[160], and ideally little to no droplets will contain two or more target molecules. This enables absolute quantification simply by counting the number of PCR amplification positive partitions.

One of the most important metrics for dPCR is λ, which is derived from the Poisson distribution, and defines the average number of target molecules per partition:

$$\lambda = -ln\left(1 - \frac{k}{n}\right)$$

Where k is the number of positive partitions detected, and n is the total number of partitions[164].

dPCR gives single molecule resolution making it an ideal platform for both rare target detection and comparative analysis. dPCR has been instrumental in a wide range of different applications including (1) copy number variation in genetically modified organism and in diseases, such as cancer, (2) in plant and prenatal karyotyping, and in human diseases to address (3) gene amplification, (4) gene expression, and (4) epigenetic regulatory control as well as rare signal detection such as (5) antimicrobial resistance, (6) organ transplant rejection and (7) liquid biopsy mutation detection[165].

While the dPCR systems have been extensively used in liquid biopsy applications, it has typically been used in the context of rare mutation detection in the background of relatively abundant wildtype signal. There are minimal publications using dPCR for single cell profiling. Trifunov showed single cell profiling from muscle cells using dPCR.[166] The authors evaluated deletions in the mitochondrial DNA from patients suffering from mitochondrial diseases. Zhang et al. showed the first single cell profiling in CTCs using dPCR[167]. However, in this study, the

authors used amplified DNA product from CTCs, and didn't capture the true sensitivity of dPCR for single cell profiling.

One of the major limitations of PCR-based experiments is that the sample is consumed during the experiment and cannot be probed for other markers. Therefore many systems are being designed to combat this limitation. Newer technologies are increasing the multiplexing capabilities through the addition of more fluorescent channels, incorporating three to four fluorescent channels such as the Naica system[168,169] by Stilla or the Absolute Q system[170,171] from COMBiNATi, respectively (Table 1.1).

The Naica digital PCR system, however, has further combatted this limitation with more innovation. The cartridge design allows for droplet recovery, the DNA extracted and used for NGS. There is an 98% droplet recovery rate and 70% DNA recovery rate[172]. This allows for samples to be probed using dPCR for specific characteristics, and then only those samples are used for NGS, saving time and costs.

Table 1.1 Summary of dPCR systems.

System	Manufacturer	Method	Total Volume	Partitions	Volume/Reaction	Throughput	Multiplexing Channels	Ref
RainDrop	RainDance	Droplet	25µL or 50µL	4million or 8 million	4.39pL/droplet	8 samples/cartrdige 10-15 min to read sample	Duplex	[173,174]
QX200	Bio-Rad	Droplet	20µL	20,000	1nL	8 samples/cartridge	Duplex	[173,174]
Quantstudio 3D Digital PCR System	Applied Biosystems	Chambers	14.5 µL	20,000	0.725pL	1 rxn/chip Up to 24 chips at a time 30sec to read chip	Duplex	[173]
Naica™	Stilla	Droplet	25µL	30,000	0.59nL/droplet	Up to 12 samples 4 samples/cartridge 2.5 hours start to finish	Triplex	[168]
Absolute Q	Combinati	Chambers	10µL	20,480	500pL/chamber	16 samples/plate 90 min start to finish	Quadplex	[170,171]
Biomark™ /EPI1™	Fluidigm	Chambers	4µL/2 µL	9,180/36,960	6nL/0.85nL	48 samples in 4 hours	Five-plex	[173–175]

1.5 Hypotheses and scope of this thesis

Circulating biomarkers, CTCs and EVs, carry clinically relevant information about active disease status that can be accessed through a minimally invasive blood draw, termed a liquid biopsy. However, their use in the clinic has remained limited due to insufficient sensitivity of current technologies to detect these trace amounts of cancer-derived cargo in the blood. I hypothesized that the development of innovative characterization methods specifically designed for these ultra-low input samples would enable new liquid biopsy applications and bring them one step closer to use in the clinic.

In this thesis, I develop, optimize and validate highly sensitive methods specifically designed for ultra-low input samples. I applied these techniques to characterize CTCs and EVs isolated from metastatic NSCLC and PDAC patients to demonstrate the clinical potential of using these circulating biomarkers for patient tumor characterization and monitoring. Further, I highlight the breadth of clinical utility that circulating biomarkers possess. First, I show through the development of a single cell gene expression and mutation co-analysis workflow that CTCs enable the discovery of new potential mechanisms of therapeutic resistance. This was demonstrated in NSCLC CTCs which carried signatures that may result in a TKI resistance mechanism. Next I discover a novel biomarker using mutations carried in EV-derived RNA for patient monitoring and present the first longitudinal study of its kind. Then I show in PDAC that CTCs carry unique signatures based on patient survival which could be used to stratify patients and be used in treatment regime guidelines. Finally, I show that CTCs can extend beyond a circulating biomarker for patient prognosis and monitoring and can be a therapeutic target in rational treatment design. I applied this in the context of NK cells which have already been implicated in immune surveillance in the circulation and provide evidence that NK cell-based therapies may aid in halting metastasis

and improve patient outcomes. Taken together, this thesis explores the landscape of CTCs for clinical utility and application (Figure 1.2) and contributes to the field with the development of advanced characterization approaches that can be widely applied to different cancers and disease stages.

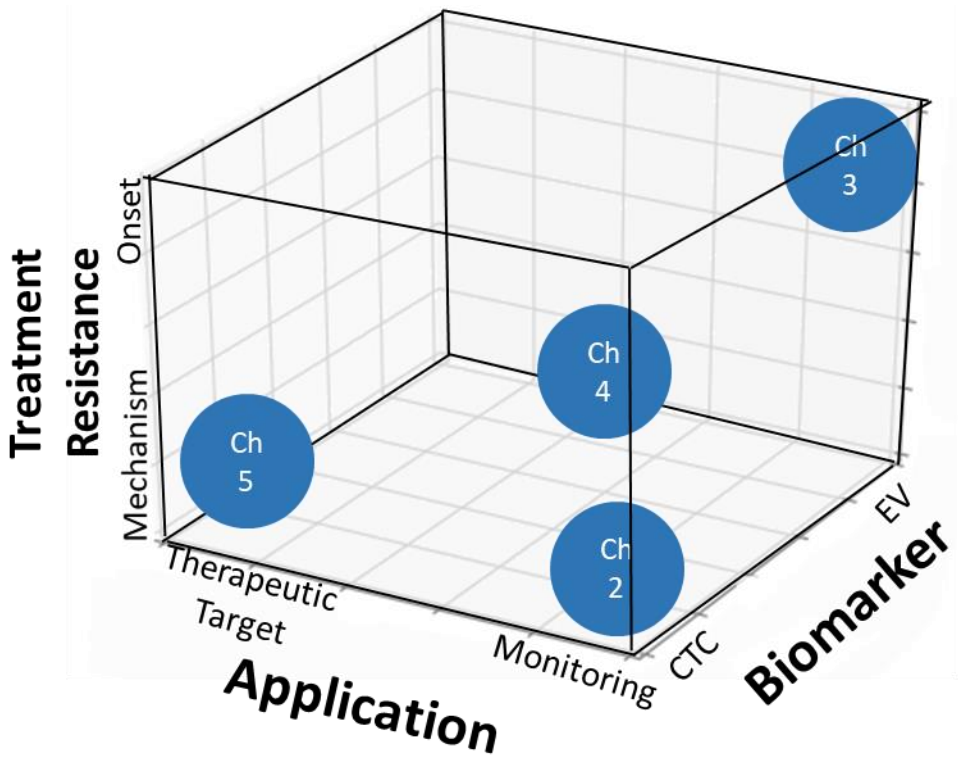


Figure 1.2 Survey of liquid biopsy landscape used in this thesis.

Chapter 2 Simultaneous Single Cell Gene Expression and EGFR Mutation Analysis of Circulating Tumor Cells Reveals Distinct Phenotypes in NSCLC

2.1 Publication Information

Owen, S., Lo, T.W., Fouladdel, S., Zeinali, M., Keller, E., Azizi, E., Ramnath, N. and Nagrath, S., 2020. Simultaneous Single Cell Gene Expression and EGFR Mutation Analysis of Circulating Tumor Cells Reveals Distinct Phenotypes in NSCLC. *Advanced Biosystems*, *4*(8), p.2000110.

The accepted manuscript has been published in Advanced Biosystems. Changes have been made to the text to fit the context off this text.

2.2 Abstract

While cancer cell populations are known to be highly heterogeneous within a tumor, the current gold standard of tumor profiling is through a tumor biopsy. These biopsies are invasive and prone to missing these clones due to spatial heterogeneity, and this bulk analysis approach can miss information from rare subpopulations. To noninvasively investigate tumor cell heterogeneity, a streamlined workflow was developed to characterize circulating tumor cells (CTCs), for simultaneous analysis of mutations and gene expression profiles at the single cell level. This powerful workflow overcomes low-input limitations of single cell analysis techniques. The utility of this multiplexed workflow to unravel inter- and intra-patient heterogeneity is demonstrated using non-small-cell lung cancer (NSCLC) CTCs (n=58) from six epidermal growth factor receptor (EGFR) mutant positive NSCLC patients. CTCs are isolated using a high-throughput microfluidic technology, the Labyrinth, and their EGFR mutation status and gene expression profiles are characterized. Using the workflow, a novel mechanism of tyrosine kinase inhibitor

(TKI) resistance is proposed. CTCs displayed reduced mutant allele expression in combination with gene expression profiles of aggressive phenotypes.

2.3 Introduction

The power of targeted therapies designed to target specific molecular vulnerabilities, was shown in non-small cell lung cancer (NSCLC) patients with activating mutations in the epidermal growth factor receptor (EGFR) gene, who responded favorably to tyrosine kinase inhibitors (TKIs) compared to patients with wildtype EGFR[176]. A patient's eligibility to receive targeted therapy is determined by molecular information obtained from tumor biopsies. However, false-negatives resulting from spatial tumor heterogeneity and the risk of missing mutant tumor clones from regional tissue biopsies can occur. Additionally, tissue biopsies are not favorable for monitoring the evolution of clonal heterogeneity and resistance with treatment. The ability to monitor patients for the onset of resistance is important for improving patient care. NSCLC patients harboring activating mutations, L858R or exon 19 deletions, in EGFR have a dramatic response to TKI therapy, but develop resistance after about 10 months through a secondary mutations, most commonly T790M[24–26].

Alternatively, circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA) present in the blood can be accessed through a routine blood draw, termed a "liquid biopsy", which may better capture tumor heterogeneity[128]. Both ctDNA and CTCs have been used to monitor disease progression and have shown clinical significance across many cancer types.[42,89–91,113] ctDNA, present in low levels in the blood, typically ranging from <0.1% to >10% of total cell-free DNA content,[177] is frequently below the detection limit of technologies such as Sanger sequencing and

quantitative PCR (qPCR), therefore many groups have focused on next-generation sequencing and more recently digital PCR (dPCR) to detect tumor-specific mutations[24,177].

Currently, the Roche Cobas EGFR mutation test v2 is the only FDA-approved liquid biopsy test, used for screening EGFR mutations from ctDNA as a companion diagnostic. However, the sensitivity and specificity of the system is only 58.4% and 80.4% respectively for the T790M mutation.[64] Due to the relatively low efficiency of this test, it is used as a "rule-in" test, meaning if the mutation is not detected, it isn't considered absent, and rather a clinician may consider recommending a repeat tumor biopsy before modifying treatment[64,178].

While ctDNA has shown promise, there is still debate about using ctDNA for tumor monitoring. ctDNA arises from lysed cells, and therefore represents a bulk snapshot of the tumor. Alternatively, CTCs may provide a real-time view of active disease status, and will likely be enriched for highly aggressive, treatment-resistant live cells. While CTCs are rare in the blood, typically on the order of tens of CTCs per milliliter of blood, many isolation techniques have been developed, most commonly using microfluidic technologies[87,92,93]. This enables whole cell analysis, which can include genomic, transcriptomic, and proteomic analysis.

Initial work in CTCs focused on enumeration and bulk analysis for gene expression using techniques such as qPCR or RNA sequencing (RNA-seq). This approach is limited to characterizing the sample as an aggregate of the bulk population, therefore missing rare phenotypes and may be biased by sample purity from leukocytes remaining in the sample.[87,94]

As a result, several groups have developed workflows for single cell analysis to study tumor cell heterogeneity, commonly through gene expression or genetic profiling. These analysis techniques tend to be labor intensive, time consuming, and expensive. Due to the low starting material many single cell analysis platforms require pre-amplification steps such as whole genome amplification

(WGA)[179,180] for genomic aberrations and whole transcriptome amplification (WTA)[133,181]. An inherent limitation of pre-amplification is the predisposition for PCR bias and gene dropout[125]. Nonetheless, single cell RNA-seq (scRNA-seq) has emerged as a common method for transcriptomic analysis, important in exploratory studies for prognostic markers. scRNA-seq has revealed intra-patient CTC heterogeneity and signatures of highly metastatic cells[125,128]. Despite its promises, groups have reported many inefficiencies and technical limitations, leading to few recovered cells proceeding through the entirety of the workflow from recovery, library preparation, and sequencing[125]. It has been reported that after all these limitations, sequencing only achieves coverage of 15-50% of total transcripts[18].

Alternatively, dPCR is a highly sensitive approach facilitating precise and accurate detection and quantification of specific nucleic acid target sequences without the requirement of pre-amplification. In dPCR, individual DNA segments are partitioned into discrete reaction droplets through a water-in-oil emulsion following the Poisson distribution[162,163]. Each droplet acts as an individual reaction for PCR amplification. The fluorescent intensity of each droplet is then measured in parallel with end-point PCR analysis. This converts data analysis into a binary positive/negative result based on fluorescent intensity, leading to simple data analysis without the need for standard curves for quantification, which could be quickly reported back to a treating physician[162,163].

Here, we have established a workflow for rapid single cell profiling of CTCs for simultaneous gene expression and mutation detection to evaluate how tumor subpopulations evolve over time. We screened 58 CTCs from six NSCLC patients with known EGFR mutations and identified intra-patient heterogeneous mutation profiles in these CTCs.

2.4 Materials and Methods

2.4.1 Cell Culture

Cells were maintained at 37°C under normoxic conditions. Cells were grown to 70-80% confluence before subculturing using 0.05% Trypsin-EDTA (Gibco). H1975 (EGFR L858R/T790M mutant), H1650 (exon 19 deletion) and T47D (EGFR wildtype) cells were grown in RPMI-1640 (Gibco), and A549 (EGFR wildtype) were grown in F-12 (Gibco), each supplemented with 10% FBS (Sigma) and 1% Antibiotic-antimycotic (Gibco). Media was exchanged every 48-72 hours between subculturing. Cell lines were routinely tested and reported negative for mycoplasma contamination (Lonza).

2.4.2 RNA Extraction & cDNA Synthesis

For cell line experiments, total RNA was purified using miRNeasy mini kit (Qiagen) following the manufacturer's protocol. RNA concentration and purity was evaluated using a NanoDrop ND-1000 spectrophotometer. For each sample, 2000 ng of total RNA was loaded into each reverse transcription reaction. cDNA was synthesized using SuperScript IV VILO Master Mix with ezDNase Enzyme (Invitrogen) following the manufacturer's protocol. All purified RNA and cDNA products were handled in a PCR workstation (AirClean Systems) to prevent nuclease contamination.

2.4.3 Experimental Protocol for Labyrinth (Patient Sample Processing)

The experimental protocol was approved by the Ethics (Institutional Review Board) and Scientific Review Committees of the University of Michigan and all patients gave their informed consent to participate in the study. All patients had a diagnosis of EGFR mutant lung adenocarcinoma.

Briefly, blood samples were collected in EDTA tubes and processed through the Labyrinth within 2 hours of collection. RBCs in the blood samples were removed using density separation with Ficoll-Paque™ PLUS Media (GE Healthcare) following the manufacturer's protocol prior to processing in the Labyrinth.

The plasma and blood mononuclear cells (PBMCs) layers were collected and diluted with PBS (1:5). The diluted sample was collected and processed through the Labyrinth at 2500 μL/min, and the product from outlet 2 was collected. To achieve a higher purity, the second outlet's products of the Labyrinth (single) were processed through another Labyrinth (double).[109] To further purify the sample, we used a glass bottom graphene oxide (GO) chip for WBC depletion. The GO Chip is made of a gold patterned silicon substrate and a PDMS top layer.

Previously described fabrication and protocol[102] for sample processing of the GO Chip were followed with a few modifications. Briefly, the PDMS top layer was bonded to a standard 1 in x 3 in glass slide (Fisher) with plasma surface activation of oxygen. The device was immediately injected and incubated with 3-mercaptopropyltrimethoxysilane (Gelest) for 2 hour at room temperature, followed by rinsing with ethanol and adding N-gamma-Maleimidobutyryloxy-Succinimide (GMBS) (ThermoScientific). After a 30 min incubation with GMBS, the device was then washed with ethanol and treated with NeutrAvidin (Invitrogen). The device was then stored at 4˚C until future use. Before the experiments, cocktails of primary antibodies, including anti-CD11b, anti-CD15, and anti-CD45, were incubated on-chip for 1 hr at room temperature. Following antibody incubation, the devices were blocked with 3% bovine serum albumin (BSA) for 0.5 hr at room temperature. The second outlet's products of the Labyrinth were first stained with Hoechst, and processed through the GO Chips at 1 mL/hr by a syringe pump. The samples were collected from the outlet of the chip for single-cell sorting using Fluidigm C1. Prior to loading

onto the C1 system (Fluidigm), CTC sample cell concentrations were evaluated using trypan blue and manually counted using a hemocytometer. After sample processing, the immunoaffinity chip with captured WBCs was washed with PBS, fixed with 4% paraformaldehyde (PFA), and stored at 4°C until imaged with fluorescent microscope.

2.4.4 Immunofluorescent Staining and CTC Enumeration

The product of single Labyrinth from outlet 2 was processed using a Thermo Scientific™ Cytospin Cytocentrifuge. A poly-lysine coated slide was placed into the cytospin funnel and 250 µL of sample was added to each cytospin funnel and cytocentrifuged at a speed of 800 rpm for 10 min. Samples were fixed on the cytoslides using 4% PFA and cytocentrifuged at the same conditions as described above. Slide samples were permeabilized by applying 0.2% Triton X-100 solution for 3 min. Slides were then washed with PBS (X 3) for 5 min and blocked using 10% donkey serum for 30 min at room temperature.

The panel of antibodies (anti-human CD45 (mouse IgG2a) (Bio-Rad), anti-human Pan-cytokeratin (CK) (mouse IgG1) (Bio-Rad), anti-human EpCAM, biotinylated (goat IgG) (R&D Systems), and anti-human vimentin (rabbit IgG) (Abcam) were used. The slides were then incubated with a cocktail of primary antibodies (anti-CD45, anti-PanCK, anti-EpCAM, and anti-Vimentin) overnight at 4˚C, followed by PBS wash (X 3) for 5 min the following day. Slides were incubated in the dark with secondary antibodies secondary antibodies goat anti-mouse IgG2a Alexa Fluor 488 (AF 488) (Invitrogen), goat anti-mouse IgG1 Alexa Fluor 546 (AF 546) (Invitrogen), goat anti-rabbit Alexa Fluor 647 (AF 647) (Invitrogen), and Stepavadin, Alexa Fluor 750 conjugate (Invitrogen) for 1.5 hour at room temperature. Finally, slides were washed with PBS (X 3) for 5 min and mounted using Prolong Gold Antifade Mountant with DAPI (Invitrogen). The

stained slides were imaged using a Nikon TI inverted fluorescent microscope at 20X magnification for enumeration.

The tiled images generated from the scans were manually viewed and CTCs were determined based on their fluorescent signals in each channel. A CTC was counted as DAPI+/CK+ (AF 546)/CD45- (AF 488). The CTC phenotype was determined based on the presence/absence of phenotype markers. EpCAM was used an epithelial marker, and vimentin was used as a mesenchymal marker. CTCs (DAPI+/CK+/CD45-) were considered epithelial if EpCAM+/vimentin-, mesenchymal if EpCAM-/vimentin+, and EMT if EpCAM+/vimentin+.

Fluidigm C1 & Biomark™ HD: Cell suspensions were loaded onto the C1 Single-Cell Auto Prep IFC for Preamp (10-17μm) (Fluidigm) following the company's protocol with on-chip cell staining. The cells were stained with FITC pre-conjugated anti-human CD45 (Biolegend), as a negative marker. After loading, each cell capture site was manually imaged.

For targeted pre-amplification, a pre-designed panel of 96 genes implicated in cancer progression, phenotype, and aggressiveness was used to characterize the cells. After the C1 run was complete, the targeted pre-amplified from the C1 IFC is harvested into a 96-well plate (Applied Biosystems). The sample was diluted using C1 DNA dilution reagent. This diluted cDNA product is split for single cell analysisusing dPCR (10 μL) and Biomark™ HD systems (2 μL). For the dPCR data generated using lung cancer cell lines, the C1 product was diluted to 28 μL, based on the manufacturer's protocol. For the lung CTC samples, the C1 product was diluted to only 12-15 μL to keep the sample more concentrated for dPCR testing. 2 μL of this sample was further diluted to 4 μL using the C1 DNA dilution reagent and was used for gene expression analysis on the Biomark™ HD system (Fluidigm) following the manufacturer's protocol, while the remainder was used for mutation detection. A limit of detection of 40 cycles was use for data analysis of gene

expression profiles of single cell profiling of the lung cancer cell lines and lung CTC samples. Water and no cell controls (harvested from the C1 pre-amplified product) were used as negative controls to ensure proper sample processing. The negative controls did not show any non-specific PCR amplification signal.

2.4.5 Mutation Detection using dPCR

The RainDrop Plus™ dPCR system (RainDance Technologies) was used for dPCR mutation detection. In brief, the PCR mix was prepared using TaqMan™ SNP Assay (Life Technologies), TaqMan™ Genotyping Master Mix (Applied Biosystems), and droplet stabilizer (RainDance Technologies). cDNA was mixed with the PCR mix in PCR tubes to generate 25 μL reactions and loaded onto the Source Chip (RainDance Technologies). The PCR reaction is emulsified with Carrier Oil (RainDance Technologies) into approximately 4 million, 5 pL sized droplets with single molecule loading, and collected into an 8-tube PCR strip (Axygen). After droplet generation, the PCR tubes were transferred to the thermocycler for 45 rounds of PCR amplification. The TaqMan™ SNP assays contain two probes, one for wildtype EGFR sequence the other for the mutant EGFR sequence, with VIC and FAM probes, respectively. The TaqMan™ exon 19 deletion assay contains probes for 19 common exon 19 deletion variants, all with a FAM probes. The PCR tubes containing the amplified samples were then transferred onto the Sense Machine (RainDance Technologies) where the endpoint fluorescence intensity of each droplet is measured. Gating templates were generated using positive and negative cell line controls.

2.4.6 Statistical Analysis

Statistical analyses were done using Prism, and error bars were generated based on average and calculated standard deviation. Gene expression analysis was conducted using the SINGuLAR Analysis Toolset (Fluidigm), which is operated through R. Data was compared and visualized

using a Gene z-score, which normalizes the expression of each gene in terms of mean and standard deviation.

2.5 Results

2.5.1. Single Cell Co-Analysis Workflow

The development of a highly sensitive approach could overcome the limitations of liquid biopsies, such as rarity of the targets and low volume constraints, facilitating early detection of resistance mutations. Here, our workflow is optimized for ultra-low input and can consistently detect the presence of these EGFR mutations at the single cell level (Figure 2.1). To interrogate CTCs at the single cell level, we combined our previously developed inertial microfluidic isolation technology, the Labyrinth[109] and the commercial Fluidigm C1 integrated fluidic chip (C1 IFC).[109] First, CTCs are enriched from peripheral blood using the Labyrinth, a high-throughput, inertial microfluidic device, which isolates CTCs from blood cells based on cell size differences. This label-free technology relies on curved channels and sharp corners to efficiently focus both CTCs and white blood cells (WBCs) into separate streamlines. It has been previously shown to yield >90% recovery of CTCs and >90% WBC removal.[109] The label-free isolation enables efficient isolation of heterogeneous CTC subpopulations. Further, without an antibody-based capture, this platform does not involve any additional CTC release step for downstream analysis. To improve CTC purity and facilitate better compatibility with current single cell isolation technologies, the remaining WBCs are depleted using modified immunoaffinity capture microfluidic device[102] functionalized with a cocktail of antibodies against common WBC targets including CD45, CD15, and CD11b. Finally, the ultra-pure CTC suspension is loaded onto the C1 IFC, which streamlines on-chip single-cell capture, lysis, reverse

transcription (RT), and targeted PCR pre-amplification of up to 96 single cells. For pre-amplification and gene expression, we chose a targeted analysis because it has been previously reported that single cell PCR offers all the benefits of bulk qPCR analysis including high sensitivity, specificity, and reproducibility.

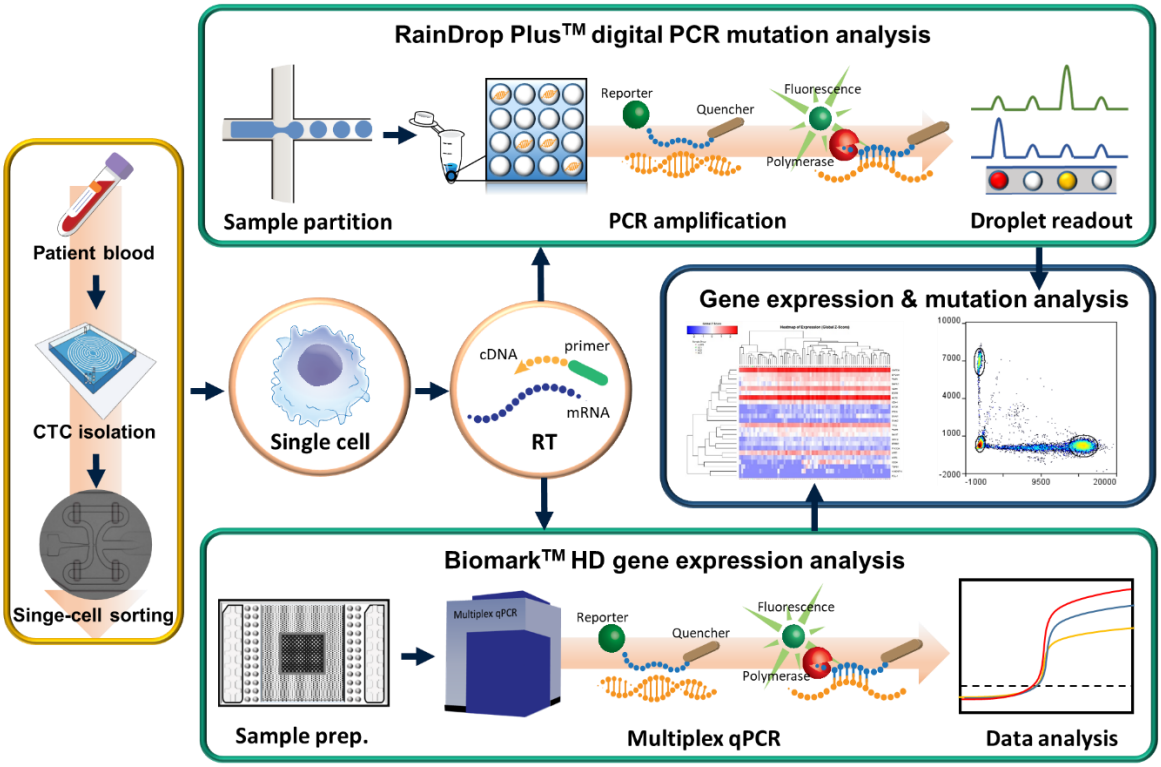


Figure 2.1 Circulating tumor cell sample processing schematic for single cell analysis. Blood is collected from patients with EGFR mutant NSCLC. The blood sample is processed through the Labyrinth, a size-based , inertial sorting microfluidic technology to isolate CTCs (~15-20μm) from WBCs (~10-12μm). The CTC sample is loaded onto the Fluidigm C1 for single cell capture and processing. The resulting single cell cDNA sample is then used for gene expression and EGFR mutation profiling co-analysis.

Due to the C1 IFC processing protocol, the generated complementary DNA (cDNA) is fragmented due to RT with random primers. This is optimal for the gene expression and EGFR mutation co-analysis because it allows for better coverage of the entire transcript during the reverse transcription step. The pre-amplification on the C1 IFC is comprised of a targeted gene panel (Table 2.1 and 2.2), using TaqMan™-based assays. The amplified region of the gene consists of a short amplicon, while the regions of the EGFR mutations (exons 19-21) remain un-amplified. This provides an untampered view of not only the EGFR mutational burden but also the relative transcription of wildtype and mutant alleles for cells with heterozygous mutations.

To achieve the simultaneous gene expression and mutation detection from a single cell, a small portion (~10%) of the pre-amplified cDNA was used for gene expression profiling of the pre-designed, targeted 96 gene panel (Tables 2.1 and 2.2) with highly multiplexed qPCR via the Biomark™ HD Dynamic Array (Fluidigm, USA), as previously shown.[109] The remaining single cell sample is analyzed for mutations using a duplex assay for the wildtype and mutant-specific sequences on the RainDrop Plus™ dPCR system (RainDance Technologies, USA).

Table 2.1 96 gene panel used for BioMark™ HD qPCR for cell lines.

Cell Line 96 Gene Panel											
GAPDH	CD133	KRT5	MMP2	JUP	ELF3	SERPINB6	TMPRSS2	IL6	CXCR1	MKI67	CD3D
ACTB	ALDH1A1	KRT7	MMP9	EVPL	CHP1	DSP	CXCL16	IL8	IGFBP5	XIST	CD11B
EPCAM	ALDH1A3	KRT8	TIMP1	NTRK2	PKP2	MLPH	ETV1	MTOR	ERCC1	HOTAIR	CD20
VIMENTIN	CDH1	KRT14	TIMP2	MUC1	CDH11	TROP2	ERG	ALK	KLF4	SPARC	CD33
ERBB2	CDH2	EMP2	PD-1	PSME3	BCL-xL	CTNND1	KLK3	EGFR	MAPK1	CCND1	CD34
CD24	TGFB1	TP53	PDL-1	XBP1	XIAP	CTNNA1	FOLH1	BMi1	STAT3	COL3A1	CD45
CD44	ABCG2	PTEN	COL1A2	FOXC1	CASP3	NFKB1	PTCH1	PIK3CA	ZEB1	SNAI1	CD146
CD44v6	CTNNB1	RB1	LGALS3BP	FOXC2	UBB	AR	PTPRN2	KRAS	ZEB2	SNAI2	FGF18

Table 2.2 96 gene panel used for BioMark™ HD qPCR for Lung CTC Patient Samples.

Lung CTC 96 Gene Panel											
GAPDH	CD133	KRT5	MUC1	CTNNB1	MMP2	SPARC	HK1	GPX4	IL6	CCND1	CD3D
LDHA	ALDH1A1	KRT7	ANXA2	STAT3	MMP9	LGALS3BP	HK2	XIST	IL8	CCNE1	CD11B
EPCAM	ALDH1A3	KRT8	ACTA2	TWIST1	TIMP1	COL14A1	PFKP	SHARP	MTOR	CDK4	CD20
VIMENTIN	CDH1	KRT14	EGFR	MCL1	TIMP2	COL1A2	PKM2	HOTAIR	MKI67	CDK6	CD33
ERBB2	CDH2	EMP2	PIK3CA	BAX	PD-1	JUP	G6PD	ESR1	PCNA	CDKN1A	CD34
CD24	TGFB1	TP53	AKT1	BCL2	PDL-1	ZEB1	TXN	ERa36	G0S2	CDKN2A	CD45
CD44	ABCG2	PTEN	ATG7	BMi1	NFKB1	ZEB2	TXNRD1	PGR	CPXM1	LOXL1	CD146
CD44v6	18SrRNA	RB1	GATA3	GSK3B	BRAC1	GSR	PRDX3	AR	CACNA1A	GPC3	FGF18

2.5.2. Validation of EGFR Mutation Detection Using Digital PCR (dPCR)

Three major EGFR driver mutations, L858R, T790M and exon 19 deletions, were chosen for this study because of their implications on patient sensitivity to TKI therapy. To validate the performance of the dPCR platform, we detected and quantified the mutant transcripts of bulk cDNA derived from H1975 (L858R/T790M) and H1650 (exon 19 deletion) and A549 (wildtype) lung cancer cell lines. Representative dPCR plots generated using each cell line are shown in Figure 2.2A.

As shown, for point mutations, L858R and T790M, gates were used to identify mutant-positive and wildtype-positive droplets, based on H1975 positive control. In these two point mutation assays, there were no false-positive mutant droplets detected in cell lines with wildtype EGFR or containing a different EGFR mutation.

For the exon 19 deletion, the assay used only screens for the mutation and contains a pool of 19 common exon 19 deletion variants, which caused a larger population spread, even in control cell lines. This lead to there being a small number of false positive-droplets being consistently generated in negative control cell lines. To maximize droplets identified in positive controls, and

minimize false-positive droplets counted, a quadrant-based gating was used, as shown in Figure 2.2A.

To determine the dynamic range and sensitivity of the system to detect low input samples, cDNA from positive control cell lines was analyzed using dPCR for the presence of EGFR L858R, T790M and exon 19 deletions at loadings ranging from 0.05ng-50ng (Figure 2.2B, Appendix A, Figure A.1-3). Linear regression analysis of the fraction of positive droplets versus cDNA loading exhibited a linear relationship (R^2=0.99) for each mutation, demonstrating a linear dynamic range across three orders of magnitude and approaching single cell sensitivity (Figure 2.2B).

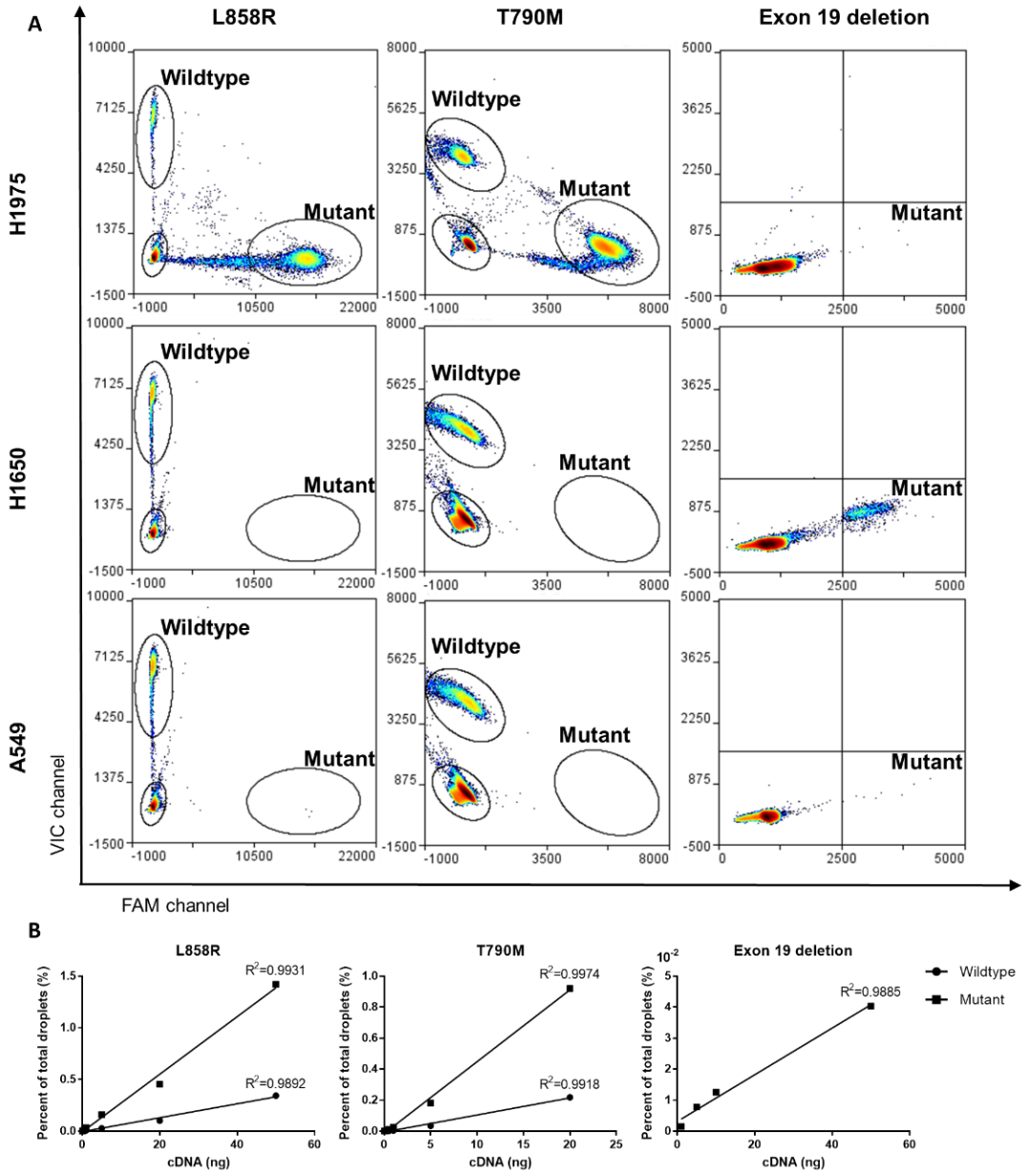


Figure 2.2 Validation of EGFR mutation detection using RainDrop PlusTM dPCR system.

A) Representative dPCR plots of lung cancer cell line controls, H1975 (L858R/T790M), H1650 (exon 19 deletion) and A549 (wildtype), for the three tested EGFR mutations, L858R (left), T790M (middle), and exon 19 deletion (right). For the point mutations (L858R and T790M), TaqManTM assays detect the wildtype (VIC channel) and mutant (FAM channel) variants. For exon 19 deletion, the assay screens for 19 common deletions (FAM channel). B) Dynamic linear range of positive droplet counts using serial dilutions of cDNA for L858R (left), T790M (middle) and exon 19 deletion (right) (Range = 0.05ng-50ng). Droplets counts for L858R and T790M results used cDNA generated from H1975 cells and exon 19 deletion results used H1650 cells.

2.5.3. Validation of Robust Single Cell Gene Expression and EGFR Mutation Co-Analysis

To establish and validate the single cell co-analysis workflow, H1975 and H1650 cells were processed on the aforementioned C1 IFC chip followed by multiplexed qPCR and dPCR analysis. EGFR mutations were identified on dPCR using the single-cell cDNA product, as shown in Figure 2.3A-C, Appendix A Figure A1.4-6. In 13/13 (100%) H1975 single cells EGFR wildtype and L858R mutant transcripts and in 14/14 (100%) H1975 single cells wildtype and T790M mutant transcripts were detected (Figure 2.3D,E). A heterogeneous expression of total EGFR was observed in H1975 cells with an average of 57 EGFR droplets per cell (range 17-88) for the L858R mutant duplex assay and 41 positive EGFR droplets per cell (range 13-72) for the T790M mutant duplex assay (Figure 2.3D,E). The H1650 cells tested with the exon 19 deletion assay showed an average of 11 positive droplets per cells (range 3-14). Due to the larger background signal in the assay, a threshold of 11 droplets was used to distinguish between positive and negative exon 19 deletion signal (Figure 2.3F).

In H1975 single cells, the mutant population contained approximately 5 times higher droplet counts than the wildtype population in both the L858R (ratio range 2.5-9.3) and T790M (ratio range 2.0-13.4) duplex assays (Figure 2.3G,H). The variations in the mutation-wildtype ratio demonstrate cell-to-cell heterogeneity, highlighting the utility of single cell analysis. However, the average of the H1975 single cells matched the bulk population for both mutation assays (Figure 2.3I). This higher expression of the mutant transcript is consistent with an increased EGFR mutant copy number in H1975 cells.[182,183]

Even within a cell line, heterogeneous gene expression profiles with multiplexed qPCR (Figure 3J; Table S1, Supporting Information) and EGFR transcript counts based on dPCR were observed (Figure 2.3D-F), highlighting the need to dissect heterogeneity of tumor cells at a single-

cell resolution. As can be seen, the expression of common NSCLC phenotype markers, such as EGFR, CD44, and epithelial cell adhesion molecule (EpCAM), showed cell to cell variability in each H1975 and H1650 single cells and could be used to characterize phenotypic subpopulations (Figure 2.3J).

To further confirm the results, the EGFR gene expression measured by qPCR was compared to the detected EGFR droplet count in the same H1975 single cells. The inferred EGFR expression using these two methods demonstrated a strong concordance (Figure 2.3K).

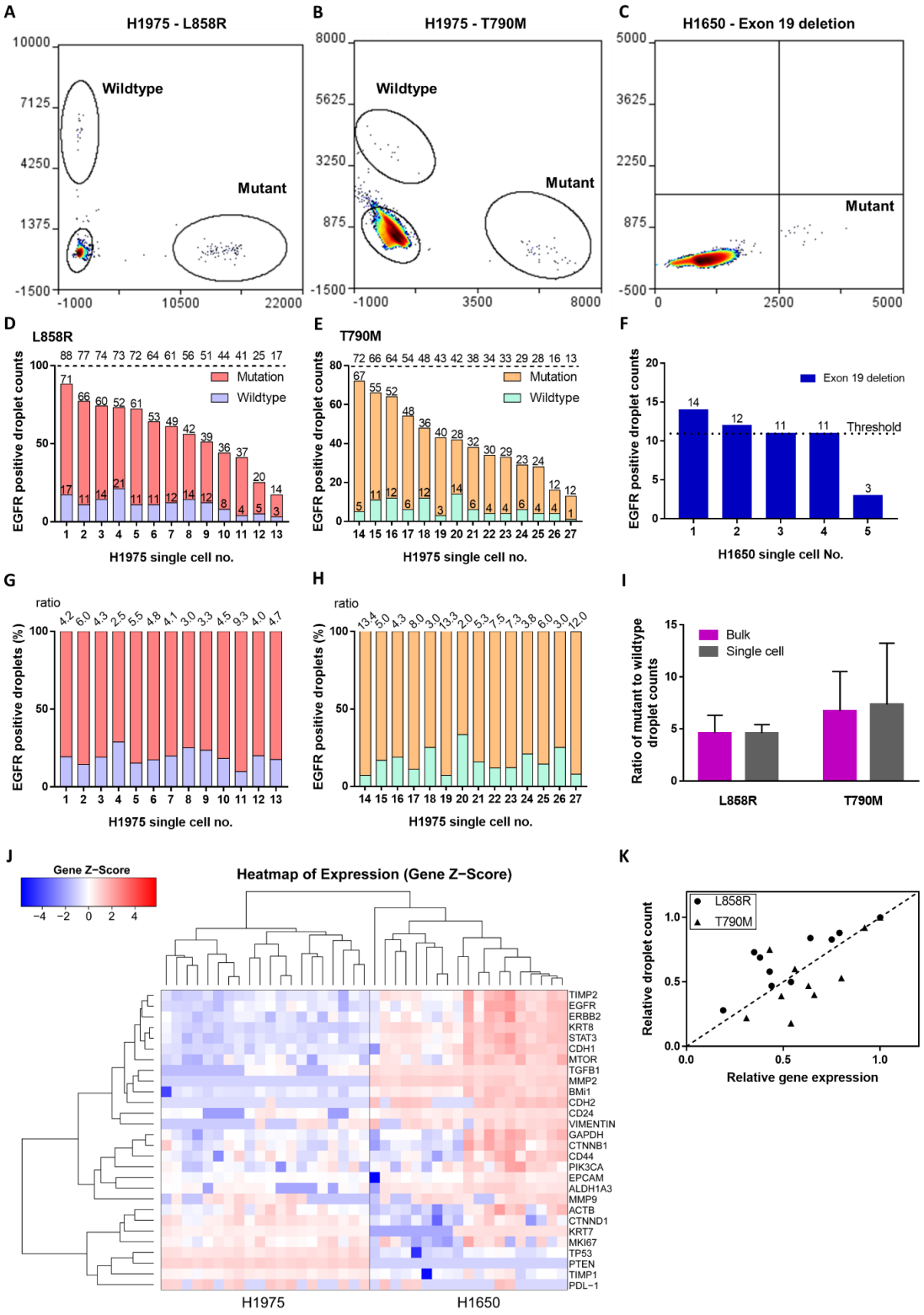


Figure 2.3 Validation of single cell workflow for gene expression and EGFR mutation analysis using lung cancer cell lines. (Caption shown on next page)

Figure 2.3 Validation of single cell workflow for gene expression and EGFR mutation analysis using lung cancer cell lines.

A-C) Representative dPCR plots of H1975 and H1650 single cells for A) EGFR L858R and B) T790M point mutations and C) exon 19 deletion respectively. D-E) H1975 single cells express heterogeneous total EGFR levels based on the combined mutant and wildtype droplet counts based on D) L858R (n= 13 H1975 cells) and E) T790M assays (n= 14 H1975 cells) F) H1650 single cells express heterogeneous mutant EGFR levels based on the exon 19 deletion assay (n= 5 H1650 single cells). G-H) Comparison of relative mutant and wildtype EGFR expression in single cell droplet counts in H1975 single cells using G) L858R (n= 13 H1975 cells) and H) T790M (n= 14 H1975 cells). Ratio of mutant:wildtype droplet counts shown above each cell. I) The average ratio of mutant:wildtype droplet counts in H1975 single cells compared to bulk cells in the L858R and T790M assays. L858R - bulk: n=12, single cells: n= 13. T790M - bulk: n=9, single cells: n= 14. J) Hierarchical clustering of H1975 and H1650 single cells. Distinct gene expression profiles between the two cell lines are identified based on 28 genes from Biomark^{TM} HD multiplexed qPCR. K) The correlation between gene expression and total droplet counts of mutation and wildtype in the same single cells. Each data point denotes a single cell (L858R: n= 10 H1975 cells, T790M: n= 10 H1975 cells). The gene expression and mutation droplet count data of each single cell are normalized to a cell that has the highest gene expression and mutation droplet counts.

2.5.4. Demographics of Patient Cohort

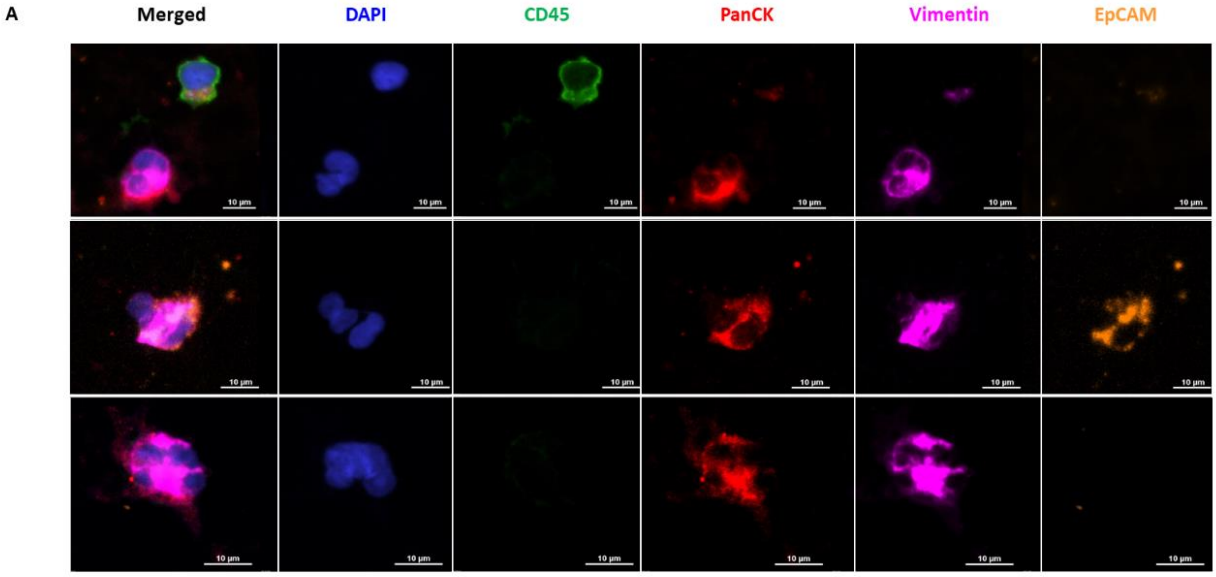
After establishing this single cell workflow using cell lines, we applied this system to analyze CTCs. NSCLC patients were enrolled at the University of Michigan Rogel Cancer Center, under an IRB approved protocol (Table 2.3). Six patients with clinical stage IV were enrolled in this study and contained known EGFR mutation based on primary tumor biopsy. Four patients had a single EGFR mutation, one with T790M, and three with exon 19 deletions. The other two patients each contained two mutations, one with L858R and T790M mutations, and the other with exon 19 deletion and T790M. The median age of the cohort was 66 years old (range: 45-78 years) and was evenly divided male and female and ranged from never, former, and current smokers.

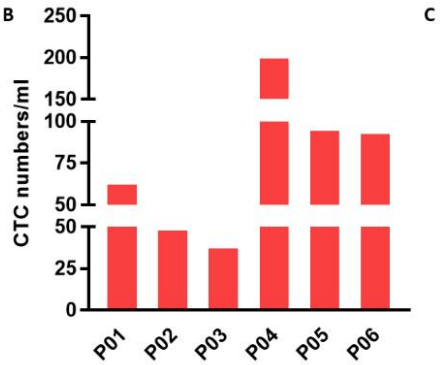
Table 2.3 Lung cancer patient demographics.

Lung Cancer Patient Demographics		
Total patients		6
Sex		
	Male	3
	Female	3
Age, years median (range)		66 (45-78)
EGFR mutation type		
	L858R	0
	T790M	1
	Exon19 deletion	3
	L858R & T790M	1
	Exon19 deletion & T790M	1
Smoking status		
	No	2
	Yes	1
	Former	3
Clinical Stage		IV
Number of Metastatic sites		NA

2.4.5. Single Cell Characterization of NSCLC CTCs

CTCs were isolated from six metastatic NSCLC patients with known EGFR mutations (Figure 2.4) using the previously described Labyrinth microfluidic technology.[109] A small portion of the CTC sample was used for enumeration with immunocytochemistry (ICC). CTCs were identified as being cytokeratin (CK) positive and CD45 negative (Figure 2.4A) and had high





C

Patient	EGFR Mutation	Treatment at time of blood draw	Status
P01	T790M	Osimertinib	Stable
P02	Exon 19 deletion	Osimertinib	Stable
P03	L858R, T790M	Pemetrexed	Progressing
P04	Exon 19 deletion	Afatinib	Progressing
P05	Exon 19 deletion, T790M	Osimertinib	Progressing
P06	Exon 19 deletion	Osimertinib	Stable

Figure 2.4 Patient characteristics and CTC analysis.
A) Representative images of heterogeneous CTCs and different EMT phenotypes based on EpCAM (epithelial) and vimentin (mesenchymal) ICC. B) CTC enumeration across the 6 patients based on immunofluorescence. (Range = 38.5-201.4 CTC/ml). C) Summary of patient characteristics including EGFR mutation status based on tumor biopsy, treatment information, and disease status at time of blood draw for CTC isolation.

heterogeneous expression of epithelial to mesenchymal transition EMT (EMT) markers. Some CTCs showed exclusive expression of epithelial marker, EpCAM, while others only showed expression of mesenchymal marker, vimentin (vim). A subset of CTCs showed dual expression of EpCAM and vim, suggesting an intermediate state within EMT (Figure 4A). We observed a wide range of CTC numbers across the patients (range 38.5-201.4 CTCs/mL blood) (Figure 2.4B), and it was noted that the patients who were progressing at the time of blood draw tended to have higher CTC numbers than those who have stable disease (Figure 2.4C).

From the six patients, the remainder of the CTC sample was used for single cell analysis via the C1 platform. The sample was stained on the C1 IFC for CD45, and only CD45 negative cells were processed for single cell analysis. The CTCs showed heterogeneous gene expression profiles (Figure 2.5A). Interestingly, the genes most commonly expressed in the CTCs were regulators of cell proliferation, such as estrogen receptor 1 (ESR1) and anti-apoptosis, such as b-cell leukemia/lymphoma 2 (BCL2), or differentiation, such as transforming growth factor beta 1 (TGFβ1). TGFβ1, known to be involved in EMT, expression showed a positive correlation with the expression of estrogen receptor (ER) in the CTCs. Interestingly, the CTCs with the highest ER and TGFβ1 also had BCL2 expression, an inhibitor of apoptosis.[184]

Additionally, individual CTCs were tested for the presence of tumor tissue-matched EGFR mutations (Figure 2.4C and 2.5B). In 5/6 patients, matched EGFR mutations were identified in at least one CTC (Figure 5B). In patients P01 and P03-P06, the CTCs contained patient-matched EGFR mutations, whereas for patient P02, dPCR signals were within the level of uncertainty to confidently classify the CTCs as containing exon 19 deletion. From patient P03, 3/4 (75%) (Figure 2.5C) of CTCs were positive for L858R mutation, but only 1/4 (25%) of cells tested positive for T790M mutation (Figure 2.5D). Similar heterogeneity of detection of different tumor-matched

mutations was seen in patient P05, 1/2 (50%) of CTCs were exon 19 deletion positive, while 3/12 (25%) (Figure 2.5E) of CTCs were T790M positive (Figure 2.5D). We did not observe any correlation between the total CTCs/mL and the percent of CTCs that tested positive for EGFR mutations.

For the subset of patients with either point mutations, L858R and T790M, in all mutation-positive CTCs we observed exclusively a heterozygous EGFR mutation expression, with variable mutation to wildtype expression ratios. For all tested cells, there was a higher abundance of wildtype expression compared to the mutant allele (Figure 2.5C,D).

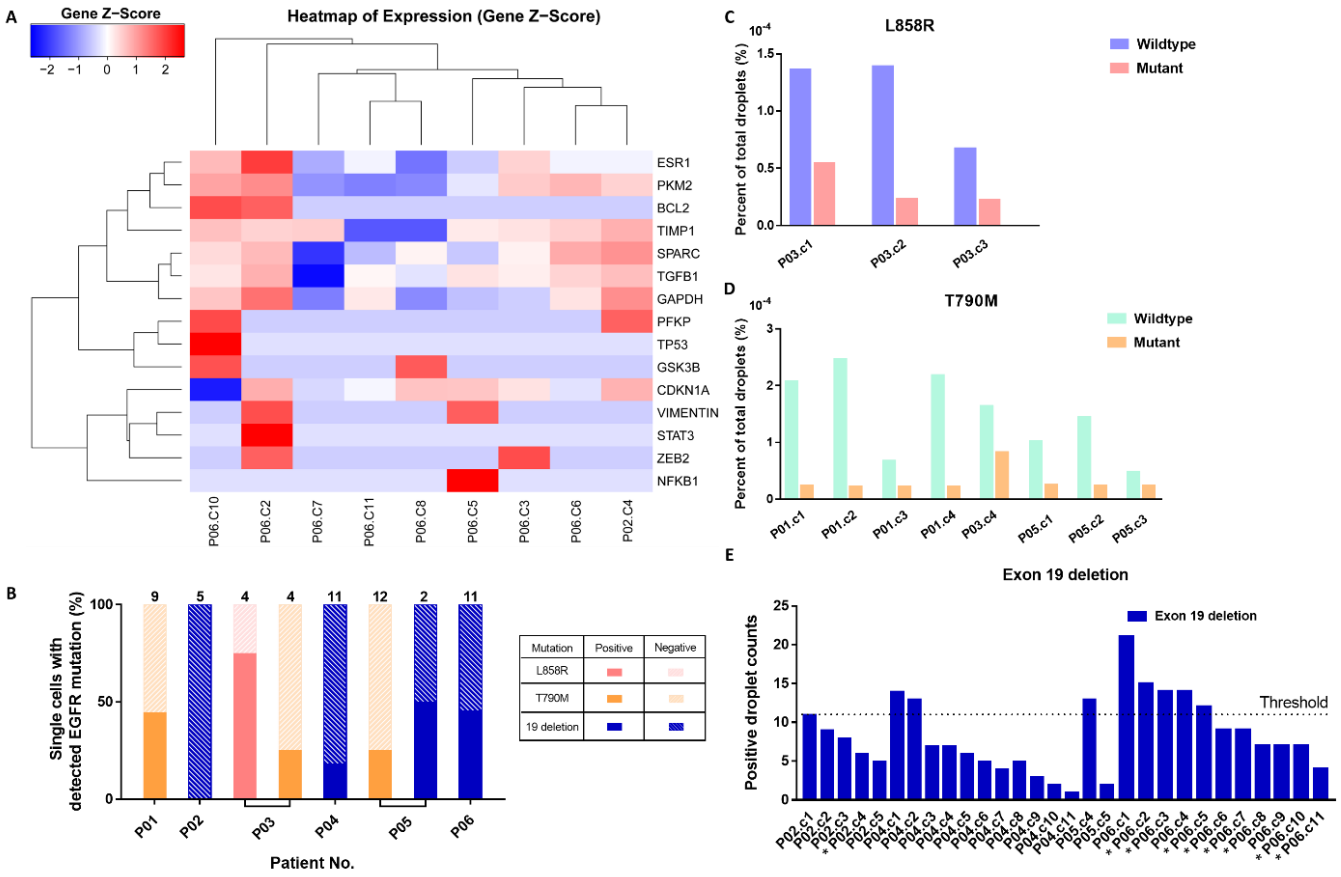


Figure 2.5 Patient tumor-matched mutations detected in CTCs.
A) Hierarchical clustering of single CTCs shows heterogeneous gene expression from two NSCLC patients (n=9 single cells). B) Summary of percent EGFR mutant positive single CTCs identified in each of the six NSCLC patients (n=58). C-E) Primary tumor matched mutations detected in patients' single CTCs for C) L858R (n=3), D) T790M (n=8), and E) Exon 19 deletion (n=8). *indicates matched gene expression data.

EGFR expression was below the limit of detection on the qPCR analysis in the CTCs, but was detected using the dPCR (Figure 2.5). Notably, for patient P06, a subset of the CTCs demonstrated a high level of exon 19 deletion EGFR expression, and also exhibited gene expression profiles consistent with aggressive phenotypes.

2.6 Discussion

In this work, we describe a multiplexed, single-cell analysis method targeting gene expression and mutation detection from a single cell through the integration of ultra-sensitive technologies. We validated our method to simultaneously quantify mutant transcripts and profile the gene expression of a single cell using lung cancer cell lines. Within a cell line, we observed both heterogeneous gene expression profiles and ratio of mutant to wildtype EGFR expression. This heterogeneity would have been lost using bulk analysis techniques.

We then applied this method to analyze CTCs from EGFR-mutated NSCLC patients. Compared to current clinical testing, this multiplexed method enables rapid turnaround from sample collection to result, in as little as two days. While others have reported the use of single cell dPCR to screen CTCs for EGFR mutations, this work required WGA, and able to identify EGFR mutations in the DNA of single CTCs.[167] Here, we highlight our workflow to screen single CTCs for mutant EGFR transcripts without the need for pre-amplification, and may reveal clinically-relevant, actionable information. In the case of NSCLC, co-analysis could reveal if specific clones harboring EGFR mutations are primarily utilizing EGFR-related pathways, or if another driver mechanism may be being utilized.

In all tested CTCs positive for L858R or T790M mutations, we observed higher wildtype than mutant EGFR expression. This could be caused by the altered copy number variation profiles

within the CTCs or differential gene expression regulation, which have been demonstrated previously in lung cancer cell lines and tumors.[183,185] This demonstrated the significance of single cell analysis to identify homozygous and heterozygous mutations, which would be lost in bulk analysis. Further, with single cell analysis we could also evaluate the relative expression of each allele. Interestingly, this mutant allele specific imbalance (MASI) is known to occur commonly in EGFR in lung cancer cell lines and tumors, usually with the mutant allele expression favored, but rarely with the wildtype allele expression being favored (reverse MASI).[183] In cancer cells lines it was shown that those with MASI trended with sensitivity to gefitinib, while the cell lines with no MASI or reverse MASI tended to be resistant to gefitinib, although this relationship needed to be further studied.[183] Other studies have found that in NSCLC patients with EGFR MASI on TKI therapy had a longer progression free survival, although it wasn't statistically significant.[186] In this study, the authors did not include reverse MASI in their analysis. As we found in our study that reverse MASI is more prevalent in CTCs, it is possible that the CTCs with reverse MASI were able to survive longer in the circulation to be isolated, despite the patient receiving TKI therapy, although this would need to be further investigated.

We found some CTCs had increased exon 19 deletion expression in combination with high ESR1 expression. It has been previously shown that ESR1 expression and EGFR mutations tends to occur more frequently together in NSCLC.[187,188] ESR1 expression in lung cancer has been associated with poor patient prognosis, while ESR2 correlation with patient prognosis appears to be dependent on cellular localization. Estrogen, through estrogen receptor signaling, can activate the signaling pathway downstream of EGFR, phosphoinositide 3-kinases/protein kinase B (PI3K/AKT). This can lead to EMT and promote cancer metastasis.[188] Additionally, there have been clinical trials evaluating the efficacy of EGFR TKI therapy in combination with estrogen

receptor antagonists and in a pilot study treatment was well tolerated and showed efficacy.[30] Future work in designing combination therapies could include the co-analysis of gene expression and mutational burden of CTCs to measure the induced tumor changes.

2.7 Conclusion

In this study, we present a combined gene expression and mutation detection of single CTCs from EGFR-mutation positive NSCLC patients. Future studies should include this analysis approach of patients across multiple visits to investigate the changes in the tumor landscape as well as incorporating the screening of additional and resistance mutations through multiplexing the dPCR analysis. Due to the genetic instability and dynamic changes a tumor undergoes throughout treatment, real-time CTC monitoring of tumor evolution could help continually optimize a patient's treatment plan, improving patient outcome. The single-cell resolution enables the early detection of emerging rare clones that could lead to therapeutic resistance from a routine blood draw, allowing for more predictive analysis of targeted therapy response. The coupled gene expression and mutation profiling using a simple workflow that doesn't require complex computational analysis could be easily integrated into a clinical setting, enabling real-time monitoring and could ultimately facilitate more timely treatment modification.

Chapter 3 EGFR Mutations Carried in Extracellular Vesicle-derived Cargo Mirror Disease Status in Metastatic Non-small Cell Lung Cancer

3.1 Abstract

Extracellular vesicles (EVs) offer a stable, abundant biomarker in the blood to serially profile molecular characteristics of patient tumors through a non-invasive liquid biopsy. In non-small cell lung cancer (NSCLC), identifying the presence of activating and resistance epidermal growth factor receptor (EGFR) mutations informs treatment. This study demonstrates the detection of EGFR mutations in both extracellular vesicle-derived RNA (EV-RNA) and extracellular vesicle-derived protein (EV-protein). Further, this pilot study reports the longitudinal study of EV-mutations for patient monitoring. EV-RNA from NSCLC patients is screened for sensitizing (exon 19 del and L858R) and resistance (T790M) EGFR mutations, and EV-protein is screened for sensitizing mutations. Through comparison to clinical data, changes in sensitizing mutations in EV-RNA strongly mirrored disease trajectory. However, EV-protein did not reveal an apparent clinical application. Extracellular vesicles, carrying both RNA and protein cargoes, may provide multiple, independent insights, from a single sample.

3.2 Introduction

Detecting the onset of resistance to cancer therapy is challenging due to the invasiveness of repeat tumor biopsies and spatial tumor heterogeneity. However, advancements in liquid biopsies have led to the clinical use of blood-based biomarkers, mostly commonly circulating tumor DNA (ctDNA), to monitor changes in the tumor non-invasively and longitudinally[189]. Yet,

ctDNA assays suffer from several notable technical challenges; ctDNA is shed only during cell death[190] and suffers from low abundance[191] and rapid clearance from circulation.[58,59]

Extracellular vesicles (EVs), lipid bilayer bound nanovesicles approximately 30-150 nm in diameter, are a promising alternative blood-based biomarker. These abundant, functional vesicles are actively secreted from live cells as a mechanism of cell-cell communication.[77]. They contain cargo from their originating cells, including DNA, RNA and protein, which is protected by the lipid bilayer from exogenous degradation while in circulation. Extracellular vesicles carry current information about a patient's cancer, and many groups have reported their utility in identifying cancer-specific markers, commonly using microRNA or protein to develop diagnostic and prognostic signatures[192,193].

Mutation profiling in extracellular vesicles is an emerging field, with several groups having previously investigated specific mutations[153,155,194]. Mutations carried in extracellular vesicles have been shown in glioblastoma, pancreatic cancer, and non-small cell lung cancer (NSCLC), with groups commonly combining ctDNA with exosomal DNA and RNA to increase detection rates[62,158]. While there have been no studies using mutations carried in extracellular vesicles to monitor a cohort of patients, initial evidence suggests this tracking may enable patient monitoring[159]. Furthermore, the detection of mutations carried in extracellular vesicles has largely been in the context of nucleic acid analysis, with few groups reporting the detection of mutant proteins in extracellular vesicles[192]. However, this was limited to cell line-derived extracellular vesicles, and was not expanded to include patient samples. Taken together, these works have shown that extracellular vesicles carry mutant nucleic acids and proteins from their cells of origin. Having a robust method to track mutations may allow for an improved and non-invasive method of monitoring patient responses compared to the current gold-standard, radiographic imaging.

A subset of NSCLC patients, those harboring sensitizing (L858R and exon 19 deletion (exon 19 del)) epidermal growth factor receptor (EGFR) mutations have seen significantly improved survival due to tyrosine kinase inhibitors (TKIs); yet resistance often occurs in as few as nine months, commonly through the secondary *EGFR* T790M mutation[195–197]. The development of a method for non-invasive longitudinal monitoring may improve their care.

In this retrospective study, a cohort of metastatic NSCLC patients with known EGFR mutations were longitudinally monitored for EGFR mutation burden carried by EV-RNA and EV-protein. To date, to the best of our knowledge, there have be no previous studies tracking exclusively extracellular vesicle mutations in a cohort of cancer patients using extracellular vesicle-derived RNA (EV-RNA) or protein (EV-protein) cargo. Distinct roles for EV-RNA and EV-protein were revealed. EV-RNA mirrored disease trajectory. However, the clinical implications of EV-protein remain unclear.

3.3 Materials and Methods

3.3.1 Cell Culture

H1975, H3255 and H1650 cells were grown in RPMI-1640 (Gibco) supplemented with 10% FBS (Sigma-Alrich) and 1% Antibiotic-antimytotic (Gibco). Cells were grown to 80% confluence before subculturing using 0.05% Trypsin-EDTA (Gibco). To prepare extracellular vesicles, cell were seeded at 3,000,000 cells/100 mm dish (Sarstedt) in complete media. 24 hours after seeding, cells were washed three times with phosphate-buffered saline (PBS) pH 7.4 (Gibco) and incubated for 72 hours in serum-free RPMI-1640 media (Gibco). Cell culture media (CCM) was centrifuged at 2,000 x g for 15 min and frozen at -20°C.

3.3.2 Patient Enrollment

All blood was collected following IRB (HUM00119934) approval, and all patients gave their informed consent to participate in the study. All patients had metastatic lung adenocarcinoma. The cohort of patients in this study had known EGFR mutations.

3.3.3 Blood Preparation

Plasma prep 1 (Ficoll): Samples were prepared using Ficoll-Paque^{TM} PLUS (GE Healthcare) following the manufacturer's protocol. The plasma and leukocyte layers were collected for CTC isolation and effluents were centrifuged following the plasma prep 3 protocol.

Plasma prep 2 (Dextran): 1 mL of 6% dextran solution (w/v) was mixed into 5 mL whole blood. The sample sat 1-1.5 hours at room temperature to allow the red blood cells sedimentation. The supernatant was collected for CTC isolation and the effluent was centrifuged following the plasma prep 3 protocol.

Plasma prep 3 (Plasma): Whole blood collected in EDTA tubes was centrifuged at 2,000xg for 15 minutes at room temperature. The plasma supernatant was collected and frozen at -20°C for up to 30 days.

3.3.4 Extracellular Vesicle Isolation using Differential and Ultracentrifugation

Plasma or CCM was centrifuged at 12,000xg for 20 minutes. The supernatant was ultracentrifuged at 100,000xg for 90 min to pellet the EVs using 36 mL Polyethylene terephthalate (PET) tubes (Thermo Fisher). Excess tube volume was filled with sterile PBS pH 7.4 (Gibco). The extracellular vesicle-pellet was washed with PBS and centrifuged at 100,000 x g for 90 min. Extracellular vesicles were suspended in PBS pH 7.4 or RIPA with protein inhibitor cocktail and frozen at -20°C.

3.3.5 Extracellular Vesicle Quantification

Following isolation, extracellular vesicles were quantified for size and concentration using NTA using Malvern's Nanosight. Quantification was performed using five-30 second runs at a flow rate of 20 using the brightness setting of 15. The camera detection was set to a level of 4 for all runs.

3.3.6 RNA Extraction and Reverse Transcription

Ultracentrifuged extracellular vesicles were lysed using Trizol™ Reagent (Trizol) (Invitrogen) at a 1:10 ratio of extracellular vesicle suspension to Trizol and incubated at room temperature for 5 minutes. A 1:5 ratio of cholorform (Sigma-Aldrich) to Trizol was added and briefly vortexed to mix, then incubated for 2-3 minutes at room temperature. The sample was centrifuged at 12,000xg for 15 minutes. The aqueous phase was collected and mixed in a 1:1 ratio with 70% ethanol (Sigma-Aldrich). Total RNA was purified using the Norgen Single Cell RNA isolation kit (Norgen Biotek Corp.). cDNA was prepared using SuperScript IV VILO Master Mix with ezDNase Enzyme (Invitrogen) following the manufacturer's protocol. All purified RNA and cDNA products were handled in a PCR workstation to prevent contamination.

3.3.7 RT-qPCR

20µL TaqMan™ gene expression PCR reactions were prepared using TaqMan™ Fast Advanced Master Mix (Applied Biosystems) in 96-well MicroAmp Fast Optical Plates (Applied Biosystems) and processed on a QuantStudio 3 (Applied Biosystems) using fast cycling conditions. Each mRNA:sample pair was analyzed in technical triplicates.

Table 3.1 TaqMan gene expression assay for RT-qPCR.

TaqMan™ gene expression assay IDs	
Gene	Assay ID
ACTB	Hs01060665_g1
GAPDH	Hs03929097_g1

3.3.8 Cell-free DNA Extraction

Cell-free DNA (cfDNA) was isolated from the plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) following the manufacturer's processing protocol. cfDNA was eluted into 15 μL for ddPCR mutation detection.

3.3.9 Mutation Detection by Droplet Digital PCR (ddPCR)

EGFR mutations were identified by using droplet digital PCR (RainDance Technologies). In brief 25 μL reactions were prepared using TaqMan™ SNP Assay (Life Technologies), 2x TaqMan Genotyping Master Mix (Applied Biosystems), and droplet stabilizer (RainDance Technologies). Maximum cDNA was loaded into each dPCR reaction. The PCR reaction was loaded onto the Source Chip (RainDance Technologies) to for droplet generation and collected into an 8-tube PCR strip (Axygen). The PCR tubes were transferred to the thermocycler for 45 rounds of PCR amplification (Bio-Rad). The PCR tubes, containing the samples, were then transferred onto the Sense Machine (RainDance Technologies) where the fluorescence intensity of each droplet was measured.

For the point mutations, L858R and T790M, mutations were considered present based on the detection of one or more positive droplets within the pre-established gates based on positive

EV-RNA controls. For exon 19 del, deletions were considered present based on the detection of one or more positive droplets above the threshold. A threshold for detection was determined based on the number of false positive droplets detected using EV-RNA negative controls. The maximum number of false positive droplets detected in any negative control (16 droplets) was used as the threshold for detection. All presented data is represented as the threshold subtracted from the total number of mutant positive droplets counted and any further normalization specified in the respective figure.

Table 3.2 TaqMan EGFR mutation detection assays.

TaqMan™ EGFR mutation detection assay IDs	
Gene	**Assay ID**
L858R	AHRSRSV
T790M	AHRSROS
Exon 19 deletion	Hs00000228_mu

3.3.10 Protein Extraction, Quantification, and Normalization

Extracellular vesicles were isolated from ultracentrifugation into 150 µL RIPA buffer (Thermo Scientific, cat #89900), protein concentration was measured by microBCA™ Protein Assay Kit (Thermo Scientific cat #23235). Western blot loading was normalized by using 5mL blood volume for extracellular vesicle isolation, loading the maximum protein (37.5 µL) in each lane, and using Bio-Rad's Stain Free gels to allow normalization. Briefly, the protein was separated at 250V for 30 minutes. Semi-dry transfer was then performed using Trans-Blot Turbo Transfer System (Bio-Rad) to a high fluorescence PVFD membrane (Bio-Rad, cat #1620261). The membrane was imaged using Bio-Rad's ChemiDoc to quantify total protein per lane. The

membrane was blocked and incubated overnight with primary antibody in 5mL of 5% bovine serum albumin (Sigma-Alrich) in Tris Buffered Saline (TBS) (Bio-Rad). with 1% Tween 20 (Sigma-Aldrich) (TBST). The membrane was then washed thoroughly before incubating with HRP-secondary antibody in 3% nonfat milk in TBST for 90 minutes followed again by additional washes. Measurement was performed using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Bio-Rad, Cat#34579) and SuperSignal West Femto (Thermo, cat #34096) and imaged on the ChemiDoc.

Table 3.3 Antibodies used for EV-protein detection and EGFR mutation profiling.

Antibodies used for Western Blot		
Target	**Dilution**	**Catalog number (Cell Signaling)**
CD9	1:1000	#13174
ACTB	1:1000	#4970
GAPDH	1:1000	#5174S
Calnexin	1:1000	#2679
EGF Receptor L858R Mutant Specific	1:1000	#3197
EGF Receptor exon 19 E746-A750del specific	1:1000	#2085
Anti-rabbit IgG, HRP-linked Antibody	1:1500	#7074S

Normalization was performed following Bio-Rad's Stain Free Gel analysis protocols. Briefly, following protein separation and transfer, the blot is imaged using Bio-Rad's Stain Free Blot imaging setting to capture the total protein per lane. Using Bio-Rad's Image Lab 6.0.1 software, the total protein in each lane is compared and a normalization coefficient determined. After blotting for specific proteins, each band is compared to the total protein of the lane, adjusted using the normalization coefficient, and quantified as a Normalized Protein Intensity.

3.3.11 Statistical Analysis

All statistics were performed using GraphPad Prism 8.4.2. For population comparisons, p-values were calculated using unpaired, parametric t-tests with two-tailed p-values. To test correlations, Pearson's correlation was performed, with Pearson's r value is reported.

3.4 Results

3.4.1 EV Cargo Carries Mutations from Cells of Origin

To establish the presence of EGFR mutations in EVs, EVs secreted from lung cancer cell lines with known EGFR mutations were tested for their matching sensitizing and resistance EGFR mutations. EV-RNA and EV-protein were tested using droplet digital PCR (ddPCR) and western blot, respectively. Matching the cells of origin, EV-RNA derived from H3255 carried a heterozygous L858R mutation (Figure 3.1A, **Fig. S1)**, H1975 carried heterozygous L858R and T790M mutations (Figure 3.1A), and exon 19 deletion in H1650 (Figure 3.1b). Additionally, EV-RNA isolated from healthy controls (HC1-3) was analyzed for each mutation (Figure 1A,B**,** Appendix B Figure B.1). In healthy controls, no mutant droplets detected for either the L858R or T790M ddPCR assays. However, the exon 19 del ddPCR assay showed an average background of

14.5 ± 2.12 (n=2) droplets. This assay simultaneously screens for 19 different deletion variants, resulting in increased background signal. A threshold for positive detection was determined based on the highest background signal observed among all negative control samples. This threshold for positive detection was used for all subsequent analysis.

Cell line derived EVs also carried mutant EGFR protein, as demonstrated using western blot. Specific identification of the two activating mutant proteins was achieved. L858R was detected exclusively in H3255 and H1975 derived EVs (Figure 3.1C,D), and exon 19 del only in H1650 derived EVs (Figure 3.1D). Unlike the exon 19 del ddPCR assay, the exon 19 del antibody screens solely for the E746-A750 deletion, the most common exon 19 deletion[198]. A validated

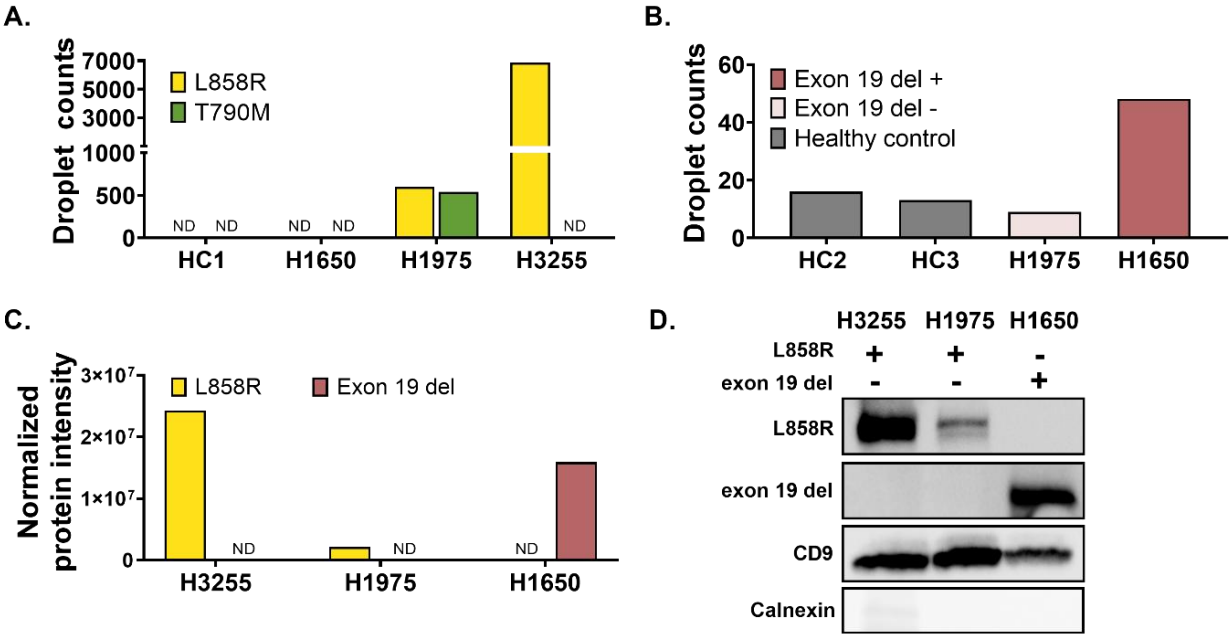


Figure 3.1 EGFR mutations carried in RNA and protein from cell line derived EVs H1975 (L858R/T790M), H3255 (L858R), H1650 (exon 19 del).
A-B) EV-RNA ddPCR droplet counts of lung cancer cell line-derived and healthy plasma prep 1 for A) L858R and T790M point mutations and B) exon19 del C) Normalized protein intensity for L858R and exon 19 of cell line derived extracellular vesicles from western blots. Normalized protein intensity was calculated using BioRad's StainFree Blot technology to compare specific bands to the total protein of each lane. D) Western blot of cell line derived extracellular vesicles tested for L858R, exon 19 del, CD9, and calnexin. HC = healthy control. ND= not detected.

T790M antibody is not yet commercially available, and therefore the samples were not tested for this mutation. Each EV-protein sample was additionally profiled for the EV marker CD9, and was shown to be free of cellular contamination based on calnexin, Figure 3.1C,D.

Due to the lack of validated loading controls in extracellular vesicles, protein intensity for each marker was normalized to the total protein loaded into each lane. This was done using BioRad's StainFree Gel technology. Protein was loaded in equal volumes across EV samples that were generated from identical cell culture conditions. After the total protein was separated and transferred, the membrane was imaged to quantify the total protein per lane. Each specific band was normalized to the total protein per lane, eliminating the need for housekeeping genes which are not equally present in all EVs due to their loading mechanisms and cells of origin[199]. StainFree Gels have been found to be more consistent than housekeeping proteins or Ponceau staining as a loading control and provide the added benefit of controlling for differential loading[200,201]. This is especially critical for this EV study, where signal is normalized to 5mL of blood for comparison across biomarker types.

3.4.2 Study Design and Patient Cohort

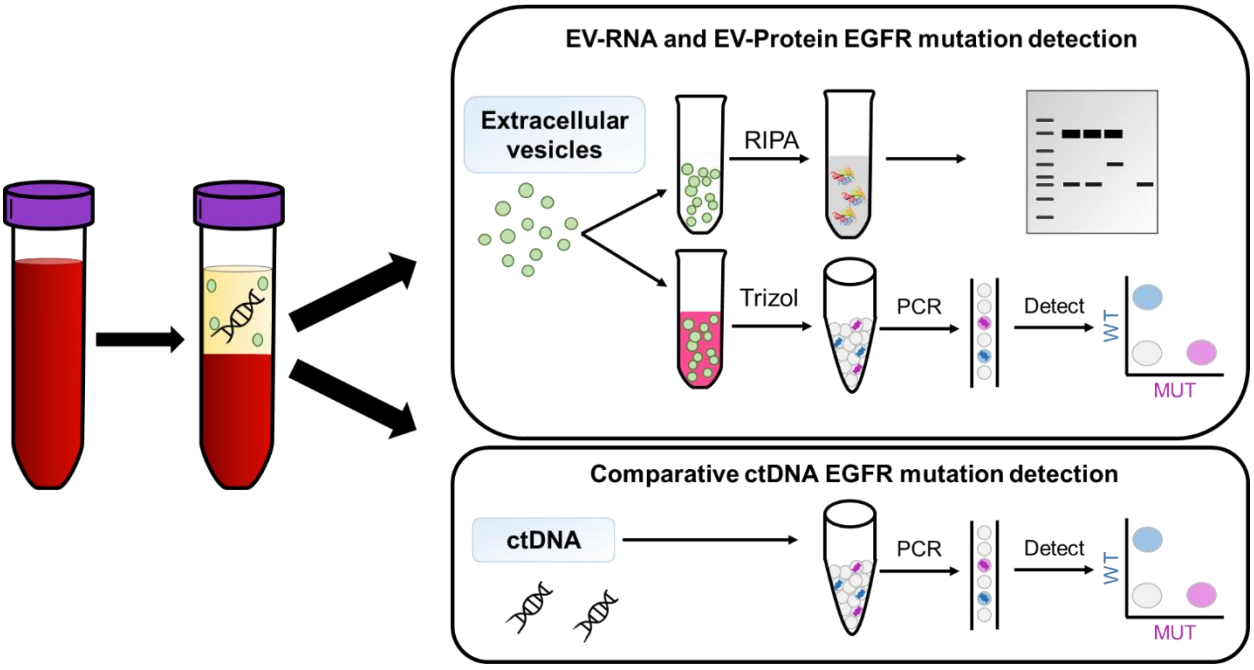


Figure 3.2 Workflow to profile EVs and ctDNA, including isolation, RNA characterization, and protein characterization.

Dual EV-RNA and EV-protein mutational profiling was then applied to a cohort of 10 metastatic NSCLC patients with at least one sensitizing EGFR mutation based on primary tumor biopsy. Patients were enrolled and blood was collected under IRB approval. Median age was 64 years (range, 45-82 years) and was well distributed between male and female. Full patient demographics are found in Table 3.4.

EVs were isolated from the effluent material generated in a tangential circulating tumor cell (CTC) characterization study[202]. The blood samples were preprocessed to isolate plasma either by red blood cell depletion with (1) Ficoll-Paque^{TM} PLUS (Ficoll) or (2) dextran, then

Table 3.4 Patient demographics and clinical information.

Patient	Gender	Age	Race	Stage	Node status	Tumor	EGFR Mutation	Treatments Used	Current patient status
L1	Female	45	White	IV	N3	T4	Exon 19 deletion, w/ PDL1 10%	Osimertinib, carboplatin + pemetrexeb + pembrolizumab, gencitabine	Deceased
L2	Male	72	White	IV	NA	NA	Exon 19 deletion, T790M	Osimertinib, carboplatin + pemetrexeb	Deceased
L3	Male	62	White	IV	N2	T3	Exon 19 deletion	Osimertinib	Stable
L4	Male	70	White	IV	N0	T2a	Exon 19 deletion	Osimertinib	Stable
L5	Male	66	White	IV	N0	T1a	Exon 19 deletion	Osimertinib	Stable
L6	Female	61	White	IV	N0	T2	Exon 19 deletion	Erlotinib	Stable
L7	Female	70	White	IV	N3	T1a	Exon 19 deletion	Osimertinib	Stable
L8	Female	82	White	IV	NA	NA	Exon 19 deletion	Osimertinib	Stable
L9	Female	61	White	IV	N3	T2	L858R, T790M	Osimertinib	Progressing
L10	Female	58	White	IV	N0	T2	L858R, T790M	Osimertinib and pemetrexed	Stable

centrifugation, or (3) centrifugation alone, referred to as plasma prep 1, 2, and 3. Validation experiments are shown in Appendix C Figure C.1**.** Due to the nature of repurposing the effluent from the CTC study, not every sample was originally processed using both Ficoll and dextran. However, to be consistent within this study, plasma prep 1 samples were processed for extracellular vesicle size, concentration, and EV-RNA mutation profiling (n=33), and plasma prep 2 samples were processed for EV-protein mutation analysis (n=21), Fig. 3.2, Table 3.5.

While a plethora of EV isolation methods have been widely developed, including ultracentrifugation[23], microfluidic devices for EV capture[203,204], and commercially available kits such as ExoQuick (System Biosciences), EVs in this study were isolated using ultracentrifugation Microfluidic devices, magnetic beads, and commercial kits are all designed to work on relatively

low volumes of plasma, typically 100 µL – 10mL, and can limit downstream analysis options through protein or microvesicle contamination or the presence of PEG or other substances[149,205]. Ultracentrifugation, however, offers the widest array of downstream applications and high purity[205,206], and is compatible with large sample input volumes. Due to dilution steps prior to CTC isolation, plasma prep volumes ranged from 1mL to 120 mL. After isolation, extracellular vesicle concentration and size was determined using nanoparticle tracking analysis (NTA), Table 3.6. Additionally, matching ctDNA samples from five patients and a total of eight samples, four with exon 19 del and four with L858R/T790M, were tested for the corresponding EGFR mutations.

Table 3.5 Summary of the patient EV-RNA and EV-protein tested for L858R, T790M and exon 19 del mutations.

Patients are identified by their original EGFR mutation as tested by needle biopsy. The EV-RNA droplet counts for L858R, T790M, and exon 19 del mutations were measured using ddPCR and normalized to 5mL of blood volume. EV-protein was characterized by western blot. *Samples tested for two mutations were only counted once towards the total number of samples processed, but each test was counted separately when counting number of total tests run. ND = not detected.

					Mutation concentration ($5mL^{-1}$)		
					EV-RNA	ctDNA	EV-Protein
Patient	EGFR mutation	Treatments used	Current disease status	Visit	exon 19 del	exon 19 del	exon 19 del
L1	exon 19 del	osimertinib, carboplatin/ pemetrexed/ pembrolizumab, gemcitabine	deceased	1	0	—	—
				2	0	—	ND
				3	3.29	—	ND
				4	10.63	—	ND
				5	12.5	—	ND
L2	*exon 19 del T790M*	osimertinib, carboplatin/ pemetrexed	deceased	1	0	—	—
				2	4.75	—	—
				3	6.64	138.2	—
L3	exon 19 del	osimertinib	stable	1	28.03	—	Yes
				2	10.27	—	Yes
				3	7.81	0	ND
				4	3.47	0	Yes
L4	exon 19 del	osimertinib	stable	1	8.06	—	—
				2	5.59	—	ND
				3	6.94	—	ND
				4	2.33	0	ND
L5	exon 19 del	osimertinib	stable	1	14.68	—	Yes
				2	5.94	—	ND
				3	—	—	Yes
				4	2.56	—	Yes
L6	exon 19 del	erlotinib	stable	1	0	—	—
				2	7.59	—	—
L7	exon 19 del	osimertinib	stable	1	53.75	—	—
L8	exon 19 del	osimertinib	stable	1	0	—	—
Summary of exon 19 del detection		No.of patient = 8		No. of samples tested = 24	n = 18/23 exon 19 del+ EV-RNA	n =1/4 exon 19 del+ ctDNA	n = 6/15 exon 19 del+ EV-Protein

					EV-RNA		ctDNA		EV-Protein
Patient	EGFR mutation	Treatments used	Current disease status	Visit	L858R	T790M	L858R	T790M	L858R
L9	L858R, T790M	osimertinib	progressing	1	0	0	—	—	Yes
				2	0.99	0	—	—	—
				3	1.97	2.63	25.78	152.97	Yes
				4	1.64	64.8	262.17	1529	—
L10	L858R T790M	osimertinib, pemetrexed	stable	1	0	0	—	—	Yes
				2	0	0	—	—	—
				3	0	0.26	—	—	Yes
				4	0.46	0	0	0	Yes
				5	0.29	0	1.47	1.47	Yes
				6	2.24	0	0	0	—
L2	*exon 19 del T790M*	*see above*	*see above*	1	—	0	—	—	—
				2	—	0	—	—	—
Summary of L858R/T790M detection		No. of patients = 3		No. of samples tested =12	n = 6/10 L858R+ EV-RNA n = 3/12 T790M+ EV-RNA		n = 3/5 L858R+ ctDNA n = 3/5 T790M+ ctDNA		n = 6/6 L858R+ EV-Protein

Summary of samples tested:

No. of patients = 10	**Total no. of samples* = 34**	**n = 27/45 (60%) EGFR mut EV-RNA**	**n = 7/14 (50%) EGFR mut ctDNA**	**n = 12/21 (57%) EGFR mut EV-Protein**

Table 3.6 Patient EGFR mutation ddPCR and extracellular vesicle characterization.

							EV Characteristics	
Patient	Visit	Mutation	Blood volume (mL)	Raw droplet count	Raw droplets without noise	Normalized droplet counts (5mL blood)	Concentration (per mL blood)	Size (nm)
L1	1	exon 19 del	8.6	9	-7	0.00	2.64E+09	141.7 ± 2.2
	2	exon 19 del	18.2	16	0	0.00	1.42E+09	157.3 ± 5.5
	3	exon 19 del	15.2	26	10	3.29	4.66E+09	140.4 ± 10.2
	4	exon 19 del	8	33	17	10.63	1.05E+10	130.4 ± 2.4
	5	exon 19 del	15.2	54	38	12.50	2.43E+09	164.2 ± 2.1
L2	1	exon 19 del	13.6	4	-12	0.00	—	—
	2	exon 19 del	20	35	19	4.75	1.80E+09	149.4 ± 2.6
	3	exon 19 del	12.8	33	17	6.64	2.40E+09	150.0 ± 6.0
L3	1	exon 19 del	6.6	53	37	28.03	—	—
	2	exon 19 del	11.2	39	23	10.27	—	—
	3	exon 19 del	12.8	36	20	7.81	1.58E+09	140.9 ± 4.4
	4	exon 19 del	7.2	21	5	3.47	—	—
L4	1	exon 19 del	6.2	26	10	8.06	—	—
	2	exon 19 del	15.2	33	17	5.59	1.99E+09	139.3 ± 3.3
	3	exon 19 del	10.8	31	15	6.94	2.78E+10	139.0 ± 6.5
	4	exon 19 del	17.2	24	8	2.33	1.22E+10	107.2 ± 4.3
L5	1	exon 19 del	12.6	53	37	14.68	4.06E+09	137.2 ± 11.7
	2	exon 19 del	10.1	28	12	5.94	2.26E+09	158.2 ± 7.5
	4	exon 19 del	19.5	26	10	2.56	2.23E+09	120.3 ± 1.7
L6	3	exon 19 del	6	12	-4	0.00	1.27E+09	157.7 ± 15.0
	4	exon 19 del	11.2	33	17	7.59	1.17E+09	134.1 ± 6.2
L7	1	exon 19 del	4	59	43	53.75	3.98E+09	138.5 ± 4.8
L8	1	exon 19 del	13.4	16	0	0.00	2.01E+09	138.8 ± 1.9

				Raw droplet counts		Normalized droplet counts (5mL blood)		EV Characteristics	
Patient	Visit	Mutation	Blood volume (mL)	Mutant	Wildtype	Mutant	Wildtype	Concentration (per mL blood)	Size (nm)
L9	1	L858R	14	0	4	0.00	1.43	3.26E+09	130.8 ± 2.5
		T790M	14	0	19	0.00	6.79		
	2	L858R	15.2	3	24	0.99	7.89	8.55E+08	167.9 ± 5.5
		T790M	15.2	0	8	0.00	2.63		
	3	L858R	7.6	3	24	1.97	15.79	—	—
		T790M	7.6	4	41	2.63	26.97		
	4	L858R	15.2	5	1	1.64	0.33	—	—
		T790M	15.2	197	326	64.80	107.24		
L10	1	L858R	10.9	0	1	0.00	0.46	3.17E+09	145.9 ± 3.5
		T790M	10.9	0	50	0.00	22.94		
	2	L858R	14.4	0	76	0.00	26.39	6.11E+08	142.5 ± 6.0
		T790M	14.4	0	65	0.00	22.57		
	3	L858R	19.2	0	0	0.00	0.00	—	—
		T790M	19.2	1	29	0.26	7.55		
	4	L858R	10.9	1	3	0.46	1.38	—	—
		T790M	10.9	0	6	0.00	2.75		
	5	L858R	17.4	1	6	0.29	1.72	—	—
		T790M	17.4	0	7	0.00	2.01		
	6	L858R	13.4	6	3	2.24	1.12	1.98E+09	141.3 ± 3.5
		T790M	13.4	0	110	0.00	41.04		
L2	2	T790M	13.6	0	6	0.00	2.21	1.80E+09	149.4 ± 2.6
	3	T790M	20	0	9	0.00	2.25	2.40E+09	150.0 ± 6.0

3.4.3 Quantification and Longitudinal Monitoring of EGFR Mutations in EV-RNA

EV-RNA samples were then isolated from 10 patients and tested for EGFR mutations based on the patient's original tumor biopsy result. Eight patients had exon 19 del, with one also having a concurrent T790M mutation. The other two patients co-harbored L858R and T790M mutations. EGFR mutant transcripts were detected in 9/10 patients, and at 25/33 (76%) visits (Table 3.5). Due some patients possessing two EGFR mutations, the overall detection rate was 60% (27/45 tests). Exon 19 del was detected in 7/8 patients, and in 78% of samples. The detection rate of L858R was modest at 60%, and the detection rate of T790M, which is known to be challenging to detect, was 25% Figure 3.3A, Table 3.5**.**

The potential clinical utility of longitudinally monitoring targetable mutations carried in EVs was evaluated. Transient mutant EGFR burden was compared to the clinical outcomes of seven patients across up to six time points. The mutation burden was defined as the number of mutant EGFR droplets detected in EV-RNA normalized to 5 mL of starting blood volume. At each time point, the patients were determined to be either stable or progressing based on clinical data. At each time point response to therapy is classified as either a stable (n=17) or progressing (n=22) based on available clinical data corresponding to each blood draw time point. The absolute quantity of mutant EV-RNA was not significantly different between stable and progressing time points for any of the mutations tested, Figure 3.3B.

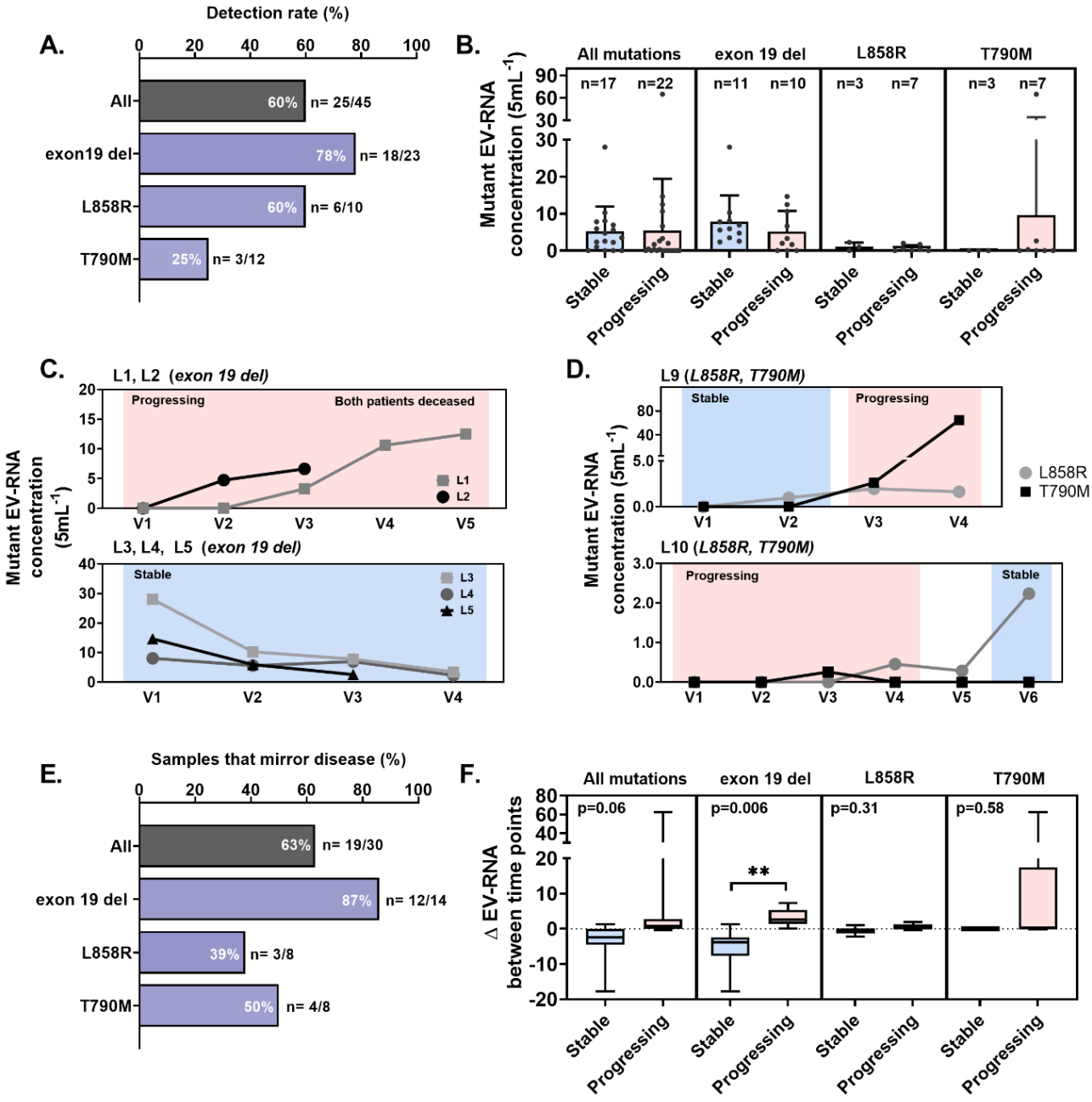


Figure 3.3. Changes in EGFR mutation burden in EV-RNA mirror disease status.
A) Percent detection rate of mutant EV-RNA samples based on mutation. B) Mutant EV-RNA concentration from stable or progressing time points for different EGFR mutations. C) Mutant EV-RNA concentration for exon 19 del patients across multiple visits for patients with progressive (top) or stable (bottom) disease. D) Mutant EV-RNA concentration for L858R/T790M patients, L9 (top) and L10 (bottom) across multiple visits for patients with progressive. E) Percent of ΔEV-RNA samples that mirror disease status based on mutation. F) Changes in EV-RNA (ΔEV-RNA) for samples with progressive or stable disease based on mutation. **indicates $p<0.01$

Patients were then classified into one of three categories based on trend in EV-RNA burden over time: progressing with exon 19 del (n=2), stable with exon 19 del (n=3), and L858R/T790M (n=2). The two patients with progression and exon 19 del, L1 and L2, had increasing EV-RNA burden that mirrored disease progression, Figure 3.3C, Figure 3.4A,B. Both patients had no detectable EV-RNA in their first visits, however as their disease progressed, the EV-RNA burden increased with each visit. While qualitative, the rate of increase seemed to be linked with progression severity. For L1 between visits 2 and 3, there was an increase in lung nodules, which correlated with the onset of detectable EV-RNA, Figure 3.4A. Between visits 3 and 4, the size and number of lung nodules both increased, corresponding with a in a significant EV-RNA increase of 223%. Similarly, for L2, the patient had progressing disease between visit 1 and 2, which was reflected at the onset of EV-RNA detection, Figure 3.4B. While the patient was considered clinically stable between visit 2 and 3, only a modest increase of 40% in EV-RNA burden was observed. However, L2's disease further progressed after visit 3. Although, this patient was also originally positive for the T790M mutation, this was never detected in the EV-RNA, Table 3.5, Table 3.6. Ultimately, both patients were placed in hospice shortly before their final time point in this study and are now both deceased, Figure 3.4A,B.

The second category, L3-L5, had sustained, clinically stable disease and showed a downward trend in their exon 19 del EV-RNA burden at nearly every time point, Figure 3.3C, Figure 3.4C-E. L5 initially had brain metastases and high levels of EV-RNA, however each subsequent time point demonstrated a decrease, in parallel with radiologically monitored resolution of brain metastases, Figure 3.4E. This observation of sustained decreased EV-RNA burden is notably different than what was observed in L2, which showed a sustained increase in mutant EGFR burden, despite temporarily being labelled as having clinically stable disease. All

three patients have demonstrated sustained clinical stability, L3-5 have been stable for 191, 182, and 204 days, for an average of 192 ± 9 days since the final EV-RNA time point, Figure 3.4C-E.

Finally, the two patients with L858R and T790M mutations, L9 and L10, showed opposite clinical trajectories. L9 had undetectable levels of L858R and T790M EV-RNA at the initial visit and was clinically stable. However, as the patient progressed, EV-RNA burden of L858R and T790M increased between visits 3 and 4, 83% and 2460% respectively (Figure 3.3D). At 87 days after the final blood draw, L9 has continued to show progression, and has brain metastasis. Conversely, L10 was progressing while on TKI therapy (osimertinib) but did not have any detectable L858R or T790M EGFR burden. However, when the patient's therapy was switched to chemotherapy (pemetrexed) and the disease stabilized, the L858R mutation became detectable and increased by 672% between visits 5 and 6 in EV-RNA, Figure 3.3D. This patient has continued to be clinically stable for 133 days since the final blood draw.

During this study it was observed that the change in mutant EV-RNA (ΔEV-RNA) best correlated with disease trajectory compared to absolute EV-RNA signal at a given time point. To further investigate this, the change in mutant EV-RNA burden between successive time points (ΔEV-RNA) was compared across patients with progressive and stable disease. ΔEV-RNA mirrored disease trajectory when both EV-RNA and disease severity increased (progression) or, similarly, when both EV-RNA and disease severity were stable or decreasing. With all three mutations combined, 60% of time points had mirrored ΔEV-RNA-disease trends, Figure 3.3E. Exon 19 del ΔEV-RNA performed the strongest with an 87% correlation rate, while L858R had 38%, and T790M had 50%.

Finally, ΔEV-RNA was quantitatively compared to disease trajectory where clinical data guided the classification of each time point as progressive or stable disease. As shown in Figure

3.3F, independent of mutation, ΔEV-RNA mirrored clinical trajectory; however, the correlation was not statistically significant ($p = 0.06$) across all samples, potentially due to the sample size or patients receiving different treatments. Patients with progressive disease had an average ΔEV-RNA of 4.8 ± 14.0, while patients with stable disease had -3.6 ± 5. Exon 19 del samples continued to perform the strongest, with a statistically significant correlation between ΔEV-RNA and clinical status ($p= 0.006$). A decrease in EV-RNA mirrored stable disease, while an increase in EV-RNA mirrored progressing disease.

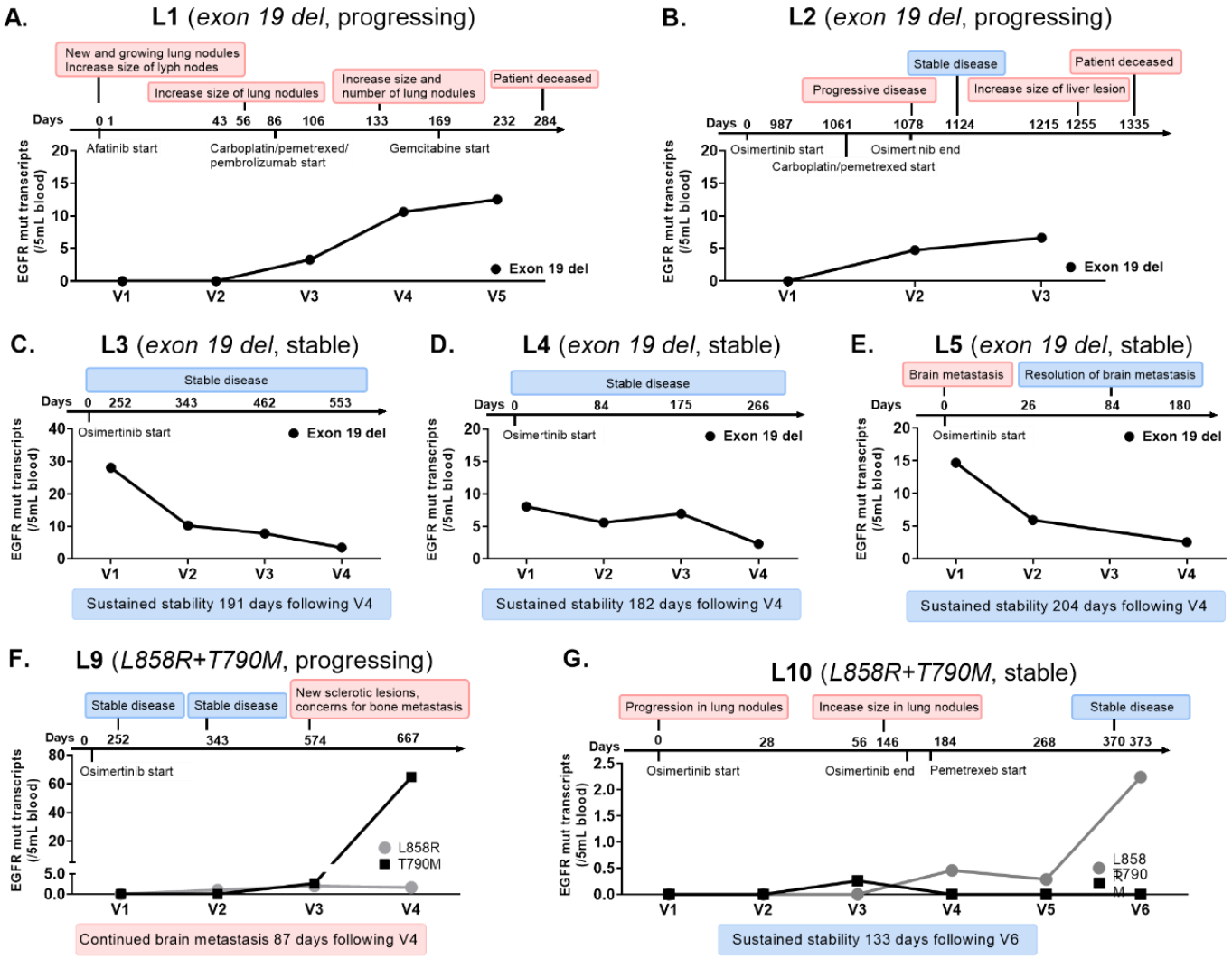


Figure 3.4 Clinical data timeline mapped to the mutant EV-RNA at each visit.
A-G) Normalized droplet counts for EGFR mutations in EV-RNA using ddPCR. A timeline with each patient's clinical data is displayed in days from start of treatment. Clinical data showing disease progression is shown in red text boxes, while stable disease is shown blue text boxes.

3.4.4 Detection and Longitudinal Monitoring of EGFR Mutations in EV-protein

A critical benefit of using EVs for liquid biopsies is their multiple cargoes. As such, in this study EVs were profiled for EV-protein longitudinally. EV-protein was tested for sensitizing mutations exon 19 del or L858R based on initial biopsy. The detection rate for both mutations was 57% (n=12/21 samples), while the detection rate of exon 19 del was only 40% and the L858R detection rate was 100%, Figure 3.5A. Similar to EV-RNA, EV-protein and clinical trajectory trends were compared. The presence of exon 19 del was highly patient dependent. Patients L3 and L5 were each positive for the mutation at 3/4 time points, while L1 and L4 were never positive,

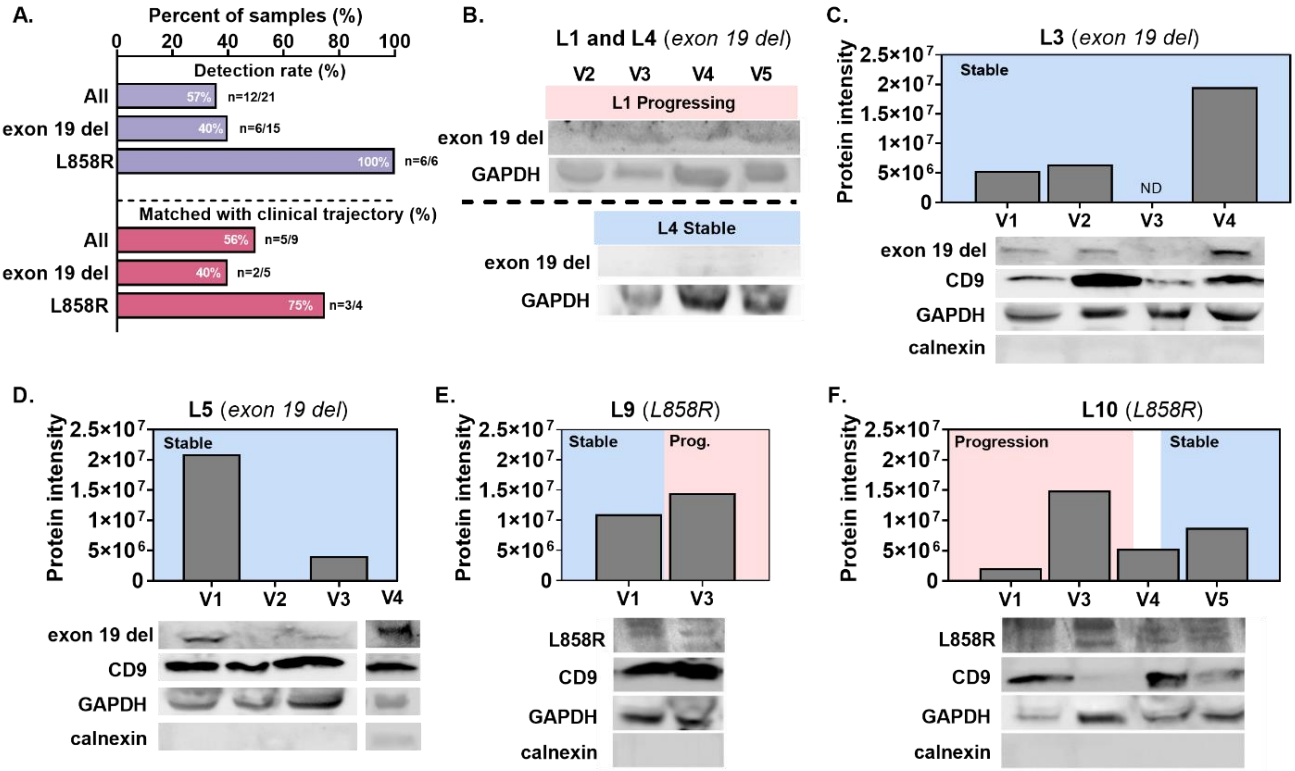


Figure 3.5 Detection of EGFR mutations in EV-protein.
A) Percent detection rate of mutant EV-protein samples and percent of samples that match clinical status based on mutation. B) EV-protein with no detectable mutant EGFR from exon 19 del patients. C-F) EV-protein with detected mutant EGFR from patients across multiple visits from C) L3 with exon 19 del, D) L5 with exon 19 del, note that one additional sample, visit 4, was tested by western blot and did not have a matching EV-RNA sample. E) L9 with L858R, and F) L10 with L858R. Samples were additionally screened for CD9, GAPDH, and calnexin. Normalized mutant EV-protein is quantified above each western blot using BioRad's StainFree Gel technology to normalize to the total protein as quantified by imaging the Stain Free Gel after transfer to a PVDF membrane. ND = not detected.

despite all four patients having different disease trajectories, Figure 3.5B-D. L858R was found in all samples tested, and quantity of L858R EV-protein mirrored disease in 3/4 time points, Figure 3.5E,F. Although the relative quantity of EV-protein does not mirror disease trajectory, this finding marks the first demonstration of EGFR mutations detected in patient-derived EV-protein.

3.4.5 Correlation between Sensitizing EGFR Mutations in EV-RNA, EV-protein, and ctDNA Circulating Biomarkers

While the ability to robustly detect multiple EGFR mutations independently in these circulating biomarkers was demonstrated, only the sensitizing EGFR mutation were evaluated for co-detection comparisons. This allowed the most robust analysis across sample types due to the lack of available T790M antibody to screen EV-protein samples. EV-RNA and EV-protein detection rates were additionally compared to the liquid biopsy gold standard, ctDNA. Sensitizing EGFR mutations were detected in 33% (n=3/7) of ctDNA, 73% (n=24/33) of EV-RNA, and 52% (n=12/23) of EV-protein samples, Figure 3.6A. ctDNA and EV-protein were least likely to be co-detected, at 25% (n=2/8). An EGFR mutation was only found in all three sample types in 25% (n=2/8) of tests.

To explore the relationship between these circulating biomarkers, the presence of these biomarkers for sensitizing mutations were correlated by comparing their co-detection (Figure 3.6A) using a Pearson's correlation matrix (Figure 3.6B). Since the EV-RNA and EV-protein has different blood pre-processing steps, only detection rates and not relative abundance were compared to mitigate a potential bias introduced due to sample prep differences. Interestingly, EV-RNA and ctDNA, which have the highest co-detection rate of any two samples at 44% (n=4/9), were not found to be correlated (r=0). This is due to mutations in EV-RNA being found in all matched samples independent of the presence of mutant ctDNA, Figure 3.6A,B. ctDNA and EV-

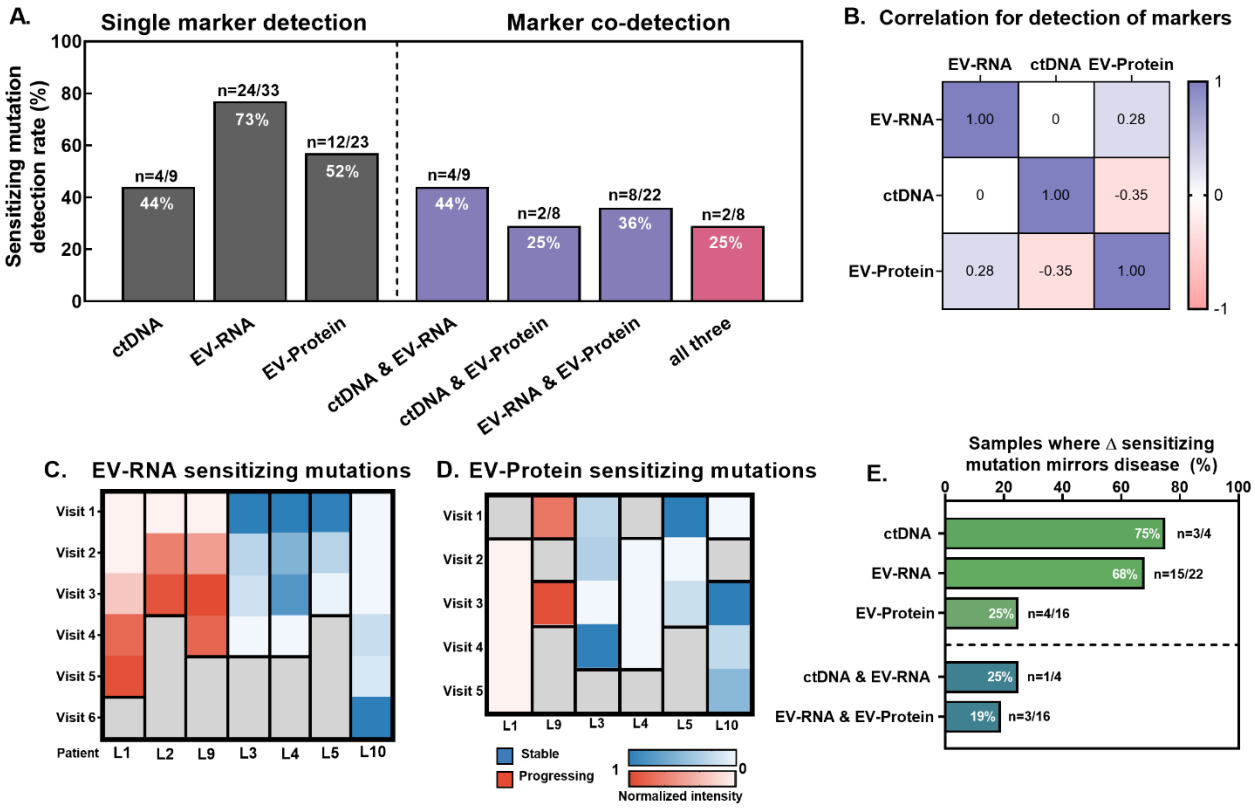


Figure 3.6 Comparison of sensitizing mutations, exon 19del and L858R, in EV-RNA, EV-Protein, and ctDNA.
A) Percent detection rate of sensitizing mutations in each sample type independently or co-detected across multiple sample types. B) Pearson's correlation matrix for the three sample types. Values are Pearson's r. C-D) Heatmaps showing C) EV-RNA and D) EV-protein intensity (normalized per patient) across multiple visits. E) Percent of samples types alone or in combination where the change in sensitizing mutation signal that mirror disease status.

protein were negatively correlated (r=-0.35); the likelihood of detecting both biomarkers was low, and had a co-detection rate of only 25% (2/8), Figure 3.6A,B.

Interestingly, the presence of EV-RNA and EV-protein were only slightly correlated, with a co-detection rate of 36% (n=8/22) and a correlation value of r=0.28, (Figure 3.6B), which may be indicative of selective packaging of cargo into the EVs. Alternatively, this mismatched detection could arise from the difference in the EV-RNA and EV-protein screening capacity. Nineteen exon 19 del variants are simultaneously screened for in the EV-RNA, while the E746-A750 deletion is specifically screened for in the EV-RNA, and other deletions would be missed.

The stark differences between the detection EGFR mutations in EV-RNA and EV-protein suggest they may represent independent biomarkers, but further studies are needed to understand this relationship.

Shown together, sensitizing mutation signal intensities for both EV-RNA and EV-protein were normalized per patient, Figure 3.6C,D. The patients with progressive disease showed an increase in the abundance of EV-RNA over time, Figure 3.6C. Similarly, the patients with stable disease, demonstrated an overall decrease in EV-RNA compared to their previous visit, with the exception of patient L10. Conversely, there was no observed trend in the amount of EV-protein carried by the extracellular vesicles with patient disease status, Figure 3.6D. This parallel analysis most clearly demonstrates that although extracellular vesicles carry mutations in both EV-RNA and EV-protein, the two biomarkers are packaged independently and likely convey different information.

To compare the utility of measuring mutant EV cargo against ctDNA for patient monitoring, the change in sensitizing mutation burden was compared to the disease trajectory between each succeeding time point. Changes in ctDNA mutation burden, the current standard, matched disease trajectory in 75% (n=3/4) time points. ΔEV-RNA of sensitizing mutations achieved a similarly high mirroring rate of 68% (n=15/22), while exon 19 del alone had a significantly higher mirroring rate of 87% (n=12/14), Figure 3.3E, 3.6E. ΔEV-protein was notably less likely to match disease status, with only 25% of samples matching. When assessed in combination, both ctDNA and EV-RNA matched disease trajectory in only 25% (n=4/16) of samples, and both EV-RNA and EV-protein matched in 19% (n=3/16) samples, Figure 3.6E.

3.5 Discussion

In this pilot study, the utility of EV-RNA and EV-protein is demonstrated to not only screen for the presence of mutations, but to dynamically monitor patient disease status. This finding highlights the power of liquid biopsies and demonstrates that a single test result is insufficient to assess patient status. In this work, liquid biopsies were collected at multiple time points from each patient, highlighting the ease and the non-invasive nature of a blood draw. The heterogeneity of baseline mutant EV-RNA levels between patients demonstrates the necessity of serial sampling for patient monitoring.

In a cohort of 10 NSCLC patients, mutant EV-RNA was detected in 9/10 patients. For 6/7 patients who were longitudinally monitored, EV-RNA burden mirrored clinical trajectory. Increasing exon 19 del EV-RNA burden mirrored disease progression in two patients, while decreasing burden mirrored stable disease in three patients. The three patients who saw a decrease in exon 19 del burden have remained clinically stable for an average of 192 ± 9 days after the final blood draw, while those who had progressive disease are now deceased. This indicates the possibility that trend in EV-RNA mutant burden was an early indicator of clinical stability.

Of the two patients with L858R/T790M mutations, one patient's EV-RNA burden mirrored disease trajectory, while the other did not. L9 continued to have progressive disease 87 days after the final blood draw. Correspondingly, L9 had an 83% and 2460% increase in L858R and resistance T790M mutations in EV-RNA respectively, Figure 3.3D, 3.4F. Conversely, between visits 4 and 5, L10 switched from TKI to chemotherapy and saw an increase in L858R EV-RNA while remaining clinically stable for 133 days after the final blood draw, Figure 3.3D, 3.4G. The change from TKI therapy to chemotherapy during this study could impact the use of EV-RNA for patient monitoring. Without the use of TKIs specifically targeting mutant EGFR, an increase

mutant burden may not indicate treatment resistance in the tumor. Therefore, while EV-RNA may mirror disease trajectory for patients receiving targeted therapy, this may not extend to patients receiving other treatment types, such as chemotherapy and further studies are needed to investigate this.

As a first step to investigate clinical utility, EV-RNA and EV-protein were compared to clinical data and showed that increasing EV-RNA mirrored disease progression. Additional studies with larger cohorts are needed to validate these findings, however, this study presents initial evidence that increase in EV-RNA indicates progression for patients receiving targeted therapy. Future studies are needed to determine if EV-RNA can be used to detect progression prior to current techniques. Of interest, L6, only had two time points but saw an upward trend in their burden despite being clinically stable thus far, Table 3.4, Table 3.6. The preliminary findings presented here warrant a recommendation that the clinical trajectory of this patient should closely be monitored for indicators of disease progression.

While EV-protein was detected in in 12/21 (57%) of samples, there wasn't an observed correlation between EV-protein burden with either EV-RNA burden or disease trajectory. In many cases, the trajectory of EV-RNA and EV-protein were opposite. Previous studies in cell lines have shown that when exposed to TKIs, EGFR mutations result in differential protein stability compared to wildtype[30]. Specifically, it was shown that treatment with erlotinib led to protein degradation in a mutant dependent manner, without significantly changing the transcriptomic expression. Additionally, osimertinib has been suggested to affect protein stability in both wildtype and T790M mutant EGFR. Further studies are needed to evaluate if this observation is consistent in larger cohorts of patients. However, it is possible that the reduced detection rate of EV-protein compared to EV-RNA can be attributed to EGFR protein degradation in the cells due to treatment

with TKIs, and therefore is not packaged into EVs[23]. This finding would indicate that EV-RNA and EV-protein may serve as independent biomarkers, harvested from a single source.

Liquid biopsies hold the potential to address spatial heterogeneity limitations, however current options, such as ctDNA, suffer from rapid degradation and low abundance in the blood. Therefore liquid biopsy mutation screening is currently considered a "rule in" test, with the recommendation of a tissue biopsy to confirm a mutant negative result[64]. Compared to the current standard blood-based biomarker, ctDNA, EVs demonstrated a more robust detection of exon 19 del, and similar detection rates for point mutations L858R and T790M. As an added benefit, EVs offer the ability to characterize both RNA and protein cargo, allowing for multi-faceted analysis of a patient's disease. By using EVs, liquid biopsy mutation screening may advance beyond the current "rule in" test.

In this novel proof of concept study, EVs were screened for previously identified EGFR mutations carried by each patient. Changes in EV-RNA correlated with disease trajectory, however the clinical implications of EV-protein remain unclear. However, the dual analysis of EV-derived cargo has the potential to go beyond monitoring and be used in lieu of a tumor biopsy for non-invasive screening for both driver and resistance mutations in EGFR across a patient's treatment course. This may enable more rapid identification of treatment resistance and allow for timely treatment changes, overall improving patient care.

Chapter 4 Synergistic Analysis of Circulating Tumor Cells Reveal Prognostic Signatures in Treatment-Naïve Metastatic Pancreatic Cancer Patients

4.1 Abstract

Pancreatic cancer is typically diagnosed at late stages, with metastatic disease having a median survival of 12 months. Here we use circulating tumor cells (CTCs) as minimally invasive approach to characterize the dynamic changes in the tumor. We identify signatures based on survival and treatment response. Patients with poor survival had increased mutant KRAS expression and deregulation of connected pathways such as PI3K-AKT and MAPK signaling. Further, in these patients, expression patterns of gemcitabine resistance mechanisms were observed, even prior to initiating treatment. This work highlights the need for identifying patients with these resistance profiles and designing treatment regimens to circumvent these mechanisms.

4.2 Introduction

Pancreatic cancer remains one of the most deadly cancer types, with a five-year survival rate of <3%[7]. Despite significant increases in survival of other cancers, largely through early detection and the advent of targeted therapy, pancreatic cancer care has had limited improvement[207]. Pancreatic cancer still suffers from late diagnosis, due to the lack of 1) available early screening methods, 2) symptoms at early stages and 3) disease-specific symptoms, even at later disease stages[4,8–10]. Additionally, current pancreatic cancer treatment options are typically limited to traditional therapies including chemotherapy, radiation, and surgery, or a combination of these approaches. These treatment options have limited clinical success and patients experience

many harsh side effects from treatment[13]. Surgery is the only potentially curative options for these patients. Unfortunately, only patients diagnosed early, before the cancer has spread, qualify for surgery, accounting for only about 15% of patients. The remaining patients are left with few options: chemotherapy and/or radiation, or entering a clinical trial. However, chemotherapy has only limited success whether alone or in combination with radiation treatment[13].

Due to the location of the pancreas within the body, shielded by other organs, tumor biopsies are particularly challenging[208]. Easily accessible biomarkers, such as from a liquid biopsy, are critical to aid in detection and monitoring[8]. Currently the clinical standard biomarkers include carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9). In cohort studies CA19-9 has been shown to correlate with presence and stage of pancreatic cancer[39], however these can suffers from low sensitivity coupled with high false negative rates[37,38] as well as a lack of specificity for pancreatic cancer[9].

Circulating tumor cells (CTCs) offer the potential to noninvasively access tumor cells through a minimally invasive blood draw. CTCs are shed from the primary tumor into the blood stream, with some reaching secondary sites, and eventually leading to the development of metastatic tumors. Despite their rarity in the blood, CTCs abundance has been shown to be an independent prognostic factor for patient outcome in many cancer types[56,209–214]. However further characterization of these CTCs may allow the detection of aggressive genotypes and phenotypes and allow better patient stratification[208,215,216].

The epithelial to mesenchymal transition (EMT) is one of the first steps in the metastatic cascade. EMT is implicated in both tumor metastasis and drug resistance[49,117,217]. During EMT, cells acquire an aggressive phenotype hallmarked by increased invasion and drug resistance. Recent studies have shown in early stage pancreatic ductal adenocarcinoma (PDAC), that CTC

phenotypes may provide different prognostic information. The detection of vimentin, a mesenchymal marker, in CTCs showed an increase risk of recurrence whereas the presence epithelial CTCs was correlated to worse overall survival[215]. Understanding and controlling the cell regulation involved in EMT may present a way to improve cancer treatment.

Additionally, oncogenic KRAS is known to be mutated in greater than 90% of PDAC, the most common type of pancreatic cancer[218], cases. KRAS mutations aid in the cellular transformation to cancer, uncontrolled cell growth, and promote an invasive cell phenotype[219]. Additionally, KRAS mutations contribute to resistance to targeted therapies such as epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) erlotinib[219,220]. While KRAS mutations have been shown to be detected in PDAC CTCs[147,221–223], the impact of KRAS mutations on CTC biology has not been studied.

In this proof of concept study, we characterize CTCs from metastatic, treatment naïve PDAC patients. CTCs are enumerated and their EMT phenotype is characterized. A secondary sample was obtained from patients after the initiation of therapy, and CTC changes were compared. Additionally, we evaluate not only the presence of KRAS mutations in CTCs, but also the mutant burden before and during therapy. We apply comprehensive CTC transcriptome analysis, to identify prognostic gene expression signatures in CTCs based on microarray profiling, revealing signatures that correlate with patient prognosis to treatment.[208,221]

4.3 Materials and Methods

4.3.1 Cell Culture

All cell lines were maintained at 37°C, 5% CO2 under normoxic conditions. Panc-1 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco) supplemented with 10%

fetal bovine serum (FBS) (Sigma), and 1% antibiotic-antimycotic (anti-anti) (Gibco). Mia-PaCa-2 were grown in DMEM supplemented with 10% FBS, 2.5% horse serum and 1% anti-anti. AsPC-1, bxPC-3, and H1650 were grown in RMPI-1650 (Gibco) supplemented with 10% FBS and 1% anti-anti. Cells were grown to 70-80% confluence before sub-culturing using 0.05% Trypsin-EDTA (Gibco). Between cell passages, media was replaced every 48-72 hours. All cell lines were tested and reported negative for mycoplasma using MycoAlert™ Mycoplasma Detection Kit (Lonza).

4.3.2 Labyrinth Fabrication

Labyrinths were fabricated using silicon wafer molds using soft lithography. Polydimethylsiloxane (PDMS) is prepared by prepared by mixing the elastomer base with a cross-linking agent at a ratio of 10:1 (Ellsworth Adhesive). The mixture is poured onto the molds, degassed in a desiccator for 1-2 hours, and cured at 75°C overnight. The hardened PDMS is cut from the mold using a razor blade and inlet (1) and outlet (4) holes are punched through the PDMS using a 0.75 mm biopsy punch. The PDMS is then bonded to a 2”x3” glass slide via plasma bonding. After visual inspection for flaws under a microscope, the labyrinth is tubed with 10” strands of 0.030” outer diameter/0.010” inner diameter Tygon® tubing (Cole-Parmer). The tubing is cut at a 45° angle and placed so that it does not impact the flow of device.

4.3.3 Patient Enrollment

The protocol was reviewed and approved by the University of Michigan Medicine Institutional Review Board. All patients gave informed consent to participate in the research study. Patients were enrolled based on the following criteria. Patients had metastatic pancreatic ductal adenocarcinoma (PDAC) and were treatment naive at the time of initial blood draw.

4.3.4 Sample Collection & CTC Isolation

Blood was collected into EDTA tubes and processed within 2 hours of sample collection. Blood was processed as previously described[224]. In brief, blood volume was measured and divided into conicals. Red blood cells (RBCs) depletion was done using HetaSep™ (STEMCELL Technologies), centrifuged at 90xg for 1 min at room temperature, with the centrifuge brake off. After centrifugation, the sample incubated at room temperature for an additional 10 min to allow further RBC sedimentation. The supernatant, containing nucleated cells, was collected and diluted with phosphate buffered saline (PBS) pH7.4 (Gibco) to 3x the original blood volume. The sample was processed through the Labyrinth at a processing rate of 2500 µL/min. Before sample collection, flow was stabilized for 1 mL, then the CTC outlet (Outlet 2) was collected in a separate conical. The final sample was divided for CTC enumeration and transcriptomic analysis.

4.3.5 Immunofluorescence and CTC Enumeration

A portion of the CTC-enriched sample after Labyrinth processing was used for immunofluorescence. Polysine™ microscope slides (Thermo Scientific) were prepared in EZ Cytofunnel (Thermo scientific) and loaded with the 200 µL of CTC sample each. The samples were spun 800rpm for 10 minutes at room temperature. The cells were fixed using 200µL of 4% paraformaldehyde (PFA), which was added to the Cytofunnel and spun again using the same conditions. Slides were covered in PBS and stored at 4°C for up to 2 weeks before immunofluorescent staining.

Slides were permeabilized with 0.2% Triton X-100 for 3 minutes at room temperature, washed 3x with PBS, then blocked with 10% goat serum (Life Technologies) for 30 minutes at room temperature. Slides were then incubated with primary antibodies diluted in 10% goat serum overnight at 4°C. The following day, slides were washed 3x using PBS, each with a 5 min

incubation. Slides were incubated, in the dark with secondary antibodies diluted in 10% goat serum for 45 minutes at room temperature. The slides were then washed 3x using PBS, each with a 5 min incubation. Coverglass was mounted using Prolong Gold Antifade Mountant with DAPI (Invitrogen).

Slides were imaged using Nikon TI microscope at 20x magnification. The entire sample area was scanned and the resulting tiled images were manually analyzed. CTCs were identified based on the following criteria. CTCs are defined as DAPI+/CD45-/CK+. The CTC phenotype was further characterized based on vimentin protein expression. Ep CTCs are defined as DAPI+/CD45-/CK+/VIM- and EMT CTCs are defined as DAPI+/CD45-/CK/VIM+.

Table 4.1 Antibodies used for CTC ICC.

Primary Antibody	**Catalog number**	**Host (Isotype)**		**Secondary Antibody**	**Catalog number**	**Host (Isotype)**
Pan Cytokeratin	BioRad MCA1907	Mouse (IgG1)		anti-Mouse IgG1, AF546	Thermo/Invitrogen A21123	Goat (IgG)
CD45	BioRad MCA87GA	Mouse (IgG2a)		anti-Mouse IgG2a, AF488	Thermo/Invitrogen A21131	Goat (IgG)
Vimentin	Cell signaling 5741	Rabbit (IgG)		Anti-rabbit IgG (H+L), AF647	Thermo/Invitrogen A21245	Goat (IgG)

4.3.6 RNA Extraction

A portion of the sample was used for transcriptomic analysis. The CTC-enriched sample was centrifuged at 300xg for 10 min to pellet the cells. The cell pellet lysed in 700 µL of TRIzol™ Reagent (Life Technologies) an incubated for 5 min at room temperature. The lysed sample was immediately frozen at -20°C for up to 2 weeks before RNA isolation and purification.

RNA was purified using an in-house, modified protocol using the Total RNA Purification kit (Norgen Biotek Corp.). The protocol was utilized TRIzol™ Reagent to lyse the sample to increase RNA yield in the presence of the remaining blood components, mainly RBCs, compared to the provided lysis reagent. After the sample in TRIzol™ Reagent was thawed, 140 µL of chloroform was added and vortexed to mix. After 3 min of incubation at room temperature, the sample was centrifuged at 12,000xg for 15 minutes. The aqueous layer (RNA layer) was collected and mixed with equal volume 70% ethanol and added to the provided columns (Norgen Biotek Corp.). The columns were washed 3x with the provided Wash Solution A (Norgen Biotek Corp.), and eluted into 30 µL of provided Elution Solution A (Norgen Biotek Corp.).

All purified RNA, cDNA, and PCR reagents were handled in a PCR workstation to prevent nuclease contamination.

4.3.7 cDNA Synthesis, and droplet digital PCR (ddPCR)

cDNA was prepared using SuperScript IV VILO with ezDNase™ Enzyme (Invitrogen) following the manufacturer's protocol.

The 20x ddPCR KRAS G12/G13 Screening Multiplex Assay (Bio-Rad Laboratories) was used with a modified protocol. 25 µL reactions were prepared combining (1) the provided master mix, (2) the primer/probes, and (3) the cDNA sample. A modified primer/probe concentration was optimized yielding a reduced primer concentration of 432 nM, and probe concentration of 120 nM. CTC samples were analyzed for mutation status using the RainDance RainDrop dPCR system. cDNA was partitioned into aqueous reaction droplets with single molecule loading, where each droplet represents a single cDNA molecule. Analysis was done via the RainDrop Analyst II Software (RainDance Technologies). Gating templates were generated using positive and negative cell line controls.

4.3.8 Affymetrix microarray processing and data analysis

Total RNA was processed using Affymetrix GeneChip® WT PLUS Expression Arrays. Total RNA concentration was evaluated by nanodrop spectrophotometer and processed either at the University of Michigan Advanced Genomics core or at the Affymetrix Bioinformatics Services core (at Thermo Fisher) following standard protocol. Data was returned and processed using Transcriptome Analysis Console (TAC) software (Thermo Fisher). Data was normalized for batch effects. All analysis was performed using the criteria of a fold change >2 or <-2 and p-val <0.05 between the two specified groups.

4.3.9 Statistical analysis

Statistical analysis was performed using GraphPad Prism 8.4.2. Significance and p-values for non-matched samples were determined using a two-sided unpaired t-tests assuming a Gaussian distribution, with a 95% confidence level. Significance and p-values for matched samples (pre-treatment vs on-treatment) were determined using a two-sided paired t-tests assuming a Gaussian distribution, with a 95% confidence level. Plotted error bars represent standard deviation.

4.4 Results

4.4.1 Patient Characteristics

All patients gave informed consent to participate in this study and blood was collected under an institutional approved IRB protocol. All patients were diagnosed with metastatic, treatment naïve PDAC, Table 4.1. Blood samples were collected prior to the start of therapy, and after therapy had begun (n=25 samples). The mean time the patient was on therapy prior to follow-up sample

collection was 23 days. Patients were evenly split between male and female, with a median age of 67 years (range: 48-85 years). Full patient demographics are shown in Table 4.2.

Table 4.2 Patient demographics, treatment, and clinical status.

Patient	Diagnosis and Disease Stage	Visit	Treatment Status	All Treatments Used	Treatment at Time of Follow-up Visit	Time On Treatment (days)	Clinical Status	Time from Intervention until Death (days)
1	metastatic PDAC	1	Treatment-naïve	Gemcitabine	-	-	Deceased	21
2	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	9
3	metastatic PDAC	1	Treatment-naïve	Gemcitabine, Paclitaxel, & PEGPH20	-	-	Alive (587 days)	NA
		2	On-treatment		Gemcitabine, Paclitaxel, & PEGPH20	10		
4	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	132
		2	on-treatment		Gemcitabine & Paclitaxel	14		
5	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel; added capectiabine after 242 days	-	-	Alive - Progression (476 days)	NA
		2	On-treatment		Gemcitabine & Paclitaxel	21		
6	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	24
7	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	123
		2	On-treatment		Gemcitabine & Paclitaxel	28		
8	metastatic PDAC	1	Treatment-naïve	Gemcitabine, Paclitaxel, Oleclumab & Durvalumab	-	-	Deceased	100
		2	On-treatment		Gemcitabine, Paclitaxel, & PD-L1 antibody	14		
9	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	433
10	metastatic PDAC	1	Treatment-naïve	Gemcitabine, cisplatin & veliparib; after 370 days verliparb only	-	-	Alive - Continuing Treatment (593 days)	NA
		2	On-treatment		Gemcitabine, Cisplatin & Veliparib	20		
11	metastatic PDAC	1	Treatment-naïve	Gemcitabine; after 37 days Paclitaxel added	-	-	Deceased	138
12	metastatic PDAC	1	Treatment-naïve	Gemcitabine, Paclitaxel, Oleclumab & Durvalumab	-	-	Deceased	212
		2	On-treatment		Gemcitabine, Paclitaxel, & PD-L1 antibody	14		
13	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	147
		2	On-treatment		Gemcitabine & Paclitaxel	55		
14	metastatic PDAC	1	Treatment-naïve	Gemcitabine & Paclitaxel	-	-	Deceased	42
		2	stopped-treatment		Gemcitabine & Paclitaxel	26		
15	metastatic PDAC	1	Treatment-naïve	Gemcitabine, Cisplatin, 5-Fu; after 112 days Cisplatin replaced with Oxaliplatin; after 161 days Gemcitabine only	-	-	Deceased	317
		2	On-treatment		Gemcitabine, Cisplatin, 5-FU	28		

Table 4.3 Complete patient demographics.

Patient	Sex	Race	Ethnicity	Age	Alcohol Consumption	Pack-Years Smoking
1	M	White or Caucasian	Non-Hispanic	68	N	11
2	F	White or Caucasian	Non-Hispanic	54	Y	2.5
3	F	White or Caucasian	Non-Hispanic	67	N	3.75
4	M	White or Caucasian	Non-Hispanic	72	N	0
5	M	Black or African American	Non-Hispanic	74	Y	50
6	M	White or Caucasian	Non-Hispanic	55	Y	0
7	F	White or Caucasian	Non-Hispanic	73	Y	20
8	M	White or Caucasian	Non-Hispanic	75	Y	10
9	F	White or Caucasian	Non-Hispanic	48	N	30
10	F	White or Caucasian	Non-Hispanic	66	Y	0
11	M	White or Caucasian	Non-Hispanic	85	N	18
12	F	White or Caucasian	Non-Hispanic	67	N	45
13	F	White or Caucasian	Non-Hispanic	66	Y	0
14	F	White or Caucasian	Non-Hispanic	62	Y	0
15	M	White or Caucasian	Non-Hispanic	58	Y	0

4.4.2 Detection of CTCs for Phenotypic and Molecular Profiling

CTCs were isolated from 15 metastatic, treatment-naïve PDAC patients, with a repeat sample collected in ten patients once the patient began therapy. CTCs were isolated from the blood using the Labyrinth, an inertial microfluidic technology that isolates CTCs with >90% recovery across various solid tumor types[109,147,225] (Figure 4.1). The isolated CTCs were enumerated using immunocytochemistry (ICC) and gene expression and KRAS mutation status were profiled.

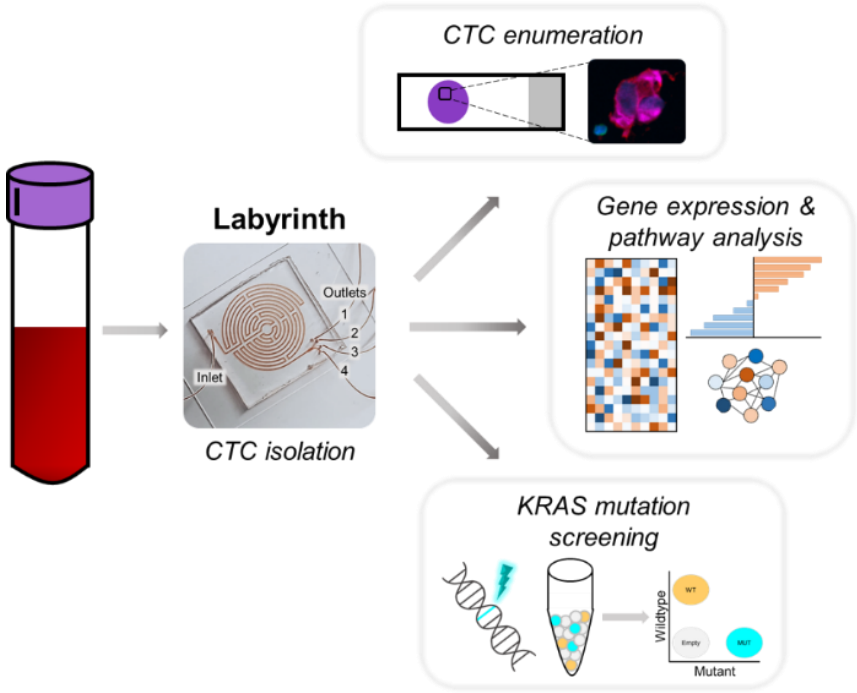


Figure 4.1 Multipronged CTC enumeration and transcriptome analysis from metastatic PDAC patients.
CTCs are isolated from the blood of metastatic PDAC patients using the Labyrinth for label-free isolation. The CTCs are then enumerated and phenotyped using ICC. CTC transcriptomes and mutant KRAS expression are profiled using microarray and digital PCR, respectively.

3.4.3 CTCs Display Heterogeneous EMT Phenotypes and KRAS Mutation Burden

CTCs were identified based on the expression of cytokeratin, a marker of epithelial origin, and absence of CD45, a leukocyte marker. EMT phenotype was determined based on the expression of vimentin in EMT-like CTCs (EMT-CTCs), while epithelial-like CTCs (Ep-CTCs) were vimentin negative (Figure 4.2A). CTCs were detected in 16/22 (73%) of samples. The CTC burden in the patients ranged from 0 CTCs/mL to 249 CTCs/mL blood, with an average of 37.8 ± 55.2 CTCs/mL blood.

Based on treatment status, CTCs were detected in 10/14 (71%) of pre-treatment samples, ranging from 0 CTCs/mL to 249 CTCs/mL blood, with an average of 46.1 ± 65.6 CTCs/mL (Figure

2B). While nearly all patients had both EMT-CTC and Ep-CTC populations, patients displayed the entire spectrum of proportions of EMT- and Ep-CTCs. Of the ten patients with detectable CTCs, seven had both Ep-CTCs and EMT-CTCs (70%), with two patients having only EMT-CTCs and one patient having only Ep-CTCs. However, the relative abundance of these two CTC subpopulations, was highly patient dependent, ranging from 3-96% (average: 45.7% ± 38.6) EMT-CTCs (Figure 4.2B,C).

The median time from beginning treatment to sample collection was 23 days. CTC detection rates of on-treatment samples were similar to pre-treatment samples; CTCs were detected in in 6/8 (75%) of on-treatment samples compared to 10/14 (71%) of pre-treatment samples. However, CTC levels tended to be lower during treatment, compared to pre-treatment samples. CTCs ranged from 0-68 CTCs/mL, with an average of 23.2 ± 22.3 CTCs/mL (Figure 4.2B).

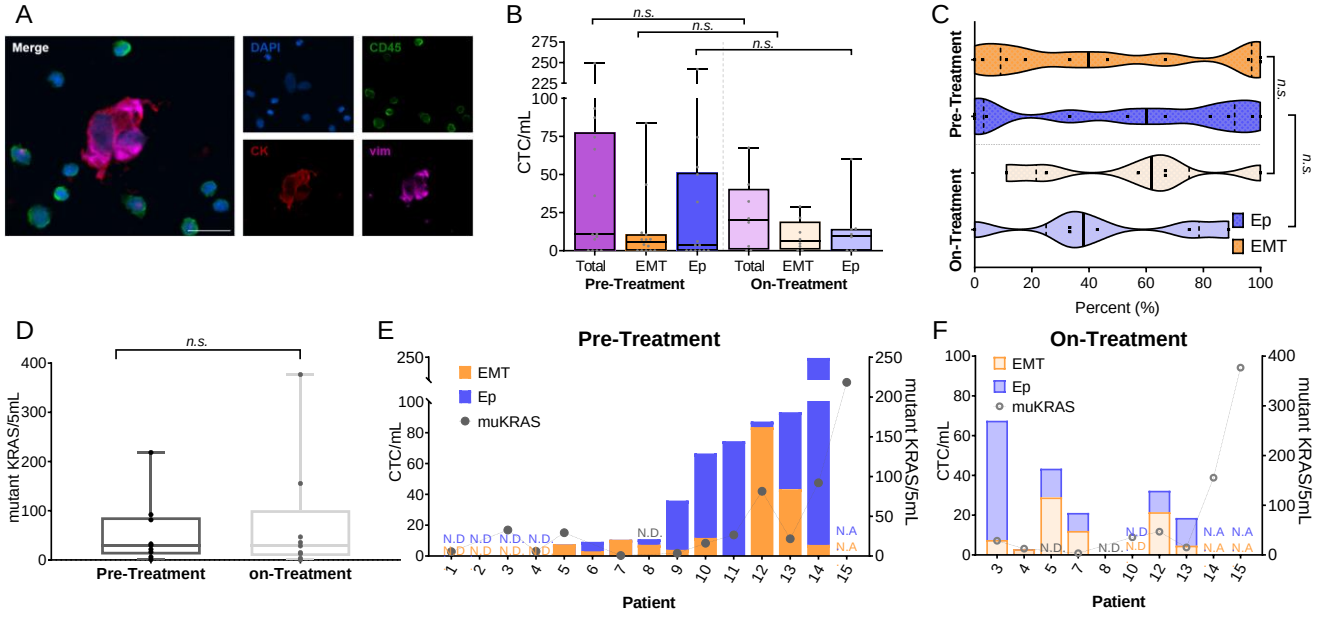


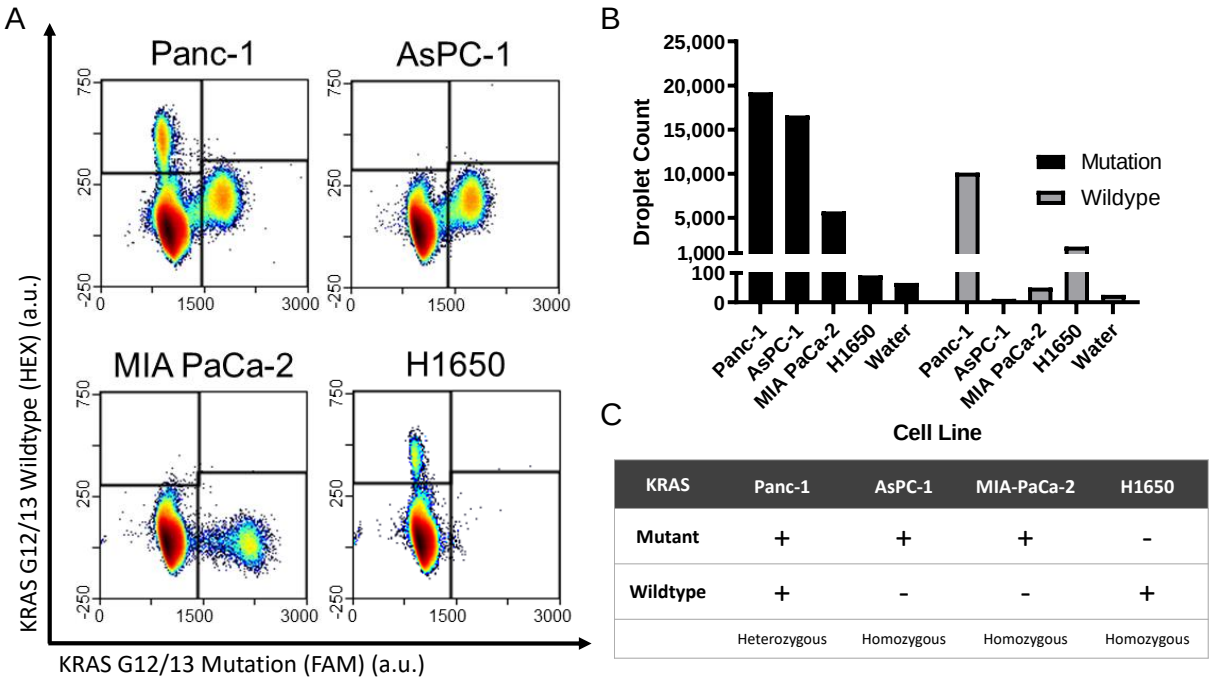
Figure 4.2 CTC enumeration, phenotype and mutant KRAS burden profiling.
A) Image of two EMT CTCs surrounded by white blood cells. Scale bar represents 20µm. B) CTC enumeration based on phenotype for pre-treatment and on-treatment samples. C) Distribution of CTC phenotype based on time point. D) Mutant KRAS expression in CTCs based on time point. E,F) Comparison of CTC counts based on phenotype with mutant KRAS (muKRAS) expression for E) pre-treatment and F) on-treatment CTC samples. N.A. = not available, N.D. = not detected.

Despite lower total CTCs, there was a shift towards a higher percentage of EMT-CTCs which we have previously reported[147] (Figure 4.2C). On-treatment samples showed an average of 58.1% ± 33.8 EMT-CTCs, nearly a 30% increase compared to pre-treatment CTC samples. Two patients had entirely EMT-CTCs, while the other four patients with detected CTCs had both EMT-CTCs and Ep-CTCs, ranging from 43-67%. At both time points, there was not an observed correlation between the total CTC abundance and proportion of a CTC phenotype subpopulation.

CTC samples were further tested for the presence and abundance of KRAS G12/13 mutations. RNA extracted from the CTCs was screened for KRAS mutations via droplet digital PCR (ddPCR). ddPCR is an ultrasensitive technique enabling single molecule detection, critical in low input samples[226]. The ddPCR RainDance system was optimized and validated for use with a KRAS mutation detection assay for CTCs, which simultaneously screens for the 7 most frequent clinically detected mutations in the G12/G13 region of the KRAS gene (Figure 4.3). Mutant KRAS was detected in 12/13 (92%) of pre-treatment CTC samples. Patients with KRAS mutations had an average of 48.0 ± 59.5 mutant KRAS transcripts/5mL (Figure 4.2D). Mutant KRAS was detected in 8/10 (80%) of on-treatment CTC samples with an average of 84.3 ± 119.4 mutant KRAS transcripts/5mL, despite overall lower CTCs (Figure 4.2D). These detection rates are in agreement with reported literature that 90% of PDACs contain a KRAS mutation[218]. Previous studies in CTCs for KRAS mutation have primarily focused on evaluating KRAS mutations in the DNA and only at a single time point[221,222,227,228], however the level of mutant KRAS expression in CTCs may be influenced by treatment status (Figure 4.2D). With the exception of patient 4, all patients maintained their KRAS mutation status despite changes in mutant KRAS expression across time points. Patient 8 had no detectable mutant KRAS at both time points, and the remaining

patients were positive at both time points. Patient 8 had no detectable mutant KRAS at both time points, and the remaining patients were positive at both time points.

In most patients, the abundance of mutant KRAS expression trended with the number of EMT-CTCs, most prominently in pre-treatment CTC samples, but this trend was still mildly observed in on-treatment samples (Figure 4.2E,F).



KRAS	Panc-1	AsPC-1	MIA-PaCa-2	H1650
Mutant	+	+	+	-
Wildtype	+	-	-	+
	Heterozygous	Homozygous	Homozygous	Homozygous

Figure 4.3 Optimization of KRAS G12/13 ddPCR assay.
A) Control cell lines were used for KRAS ddPCR assay optimization. Panc-1, AsPC-1 and MiaPaca-2 were used as positive controls, while H1650 was used as a negative control. Due to noise in ddPCR results, a signal intensity threshold was set beyond the empty droplet population. Gates were drawn 100 intensity units away from the least dense layer of the empty droplet population to remove the background. B) Raw droplet counts from 5ng of cDNA from the tested cell line samples. In patient samples droplet counts were normalized by subtracting our negative control sample droplet counts. C) Table showing the observed KRAS mutation status of the four cell line control samples.

4.4.4 Monitoring Treatment-related CTC Dynamics

Ten patients had matched pre-treatment and on-treatment CTC samples, eight with quantified CTCs, and all ten with KRAS mutation profiling. The pre-treatment and on-treatment

samples were compared for each patient. The CTC burden in response to therapy showed significant patient variability, with four patients having increased CTCs (patients 3-5,7), and four with decreased CTCs once receiving therapy compared to pre-treatment samples (patients 8,10,12,13) (Figure 4.4A). Interestingly, of the four patients who had increased total CTC counts, three also had increases in both EMT-CTCs and Ep-CTCs, with one patient showing only an increase in EMT-CTCs. Strikingly, all four of these patients had the lowest pre-treatment CTC abundance. Similarly, the four patients with decreased CTCs, all showed a decrease in EMT-CTCs and with three of the four patients also having decreased Ep-CTCs, while only one had an increase (Figure 4.4B,C). Of the nine patients who had detectable mutant KRAS levels at at least one time point, five had increased levels after beginning therapy, while four had decreased, one of which completely lost mutant KRAS signal (Figure 4.4D).

To evaluate the clinical relevance of these dynamics, the patients with samples analyzed at both time points were stratified based on their current clinical status, alive or deceased. While each patient demonstrated a unique trajectory (Figure 4.4A-D) there was a trend that patients who are now deceased demonstrated a higher total CTC burden (Figure 4.4E) as well as higher EMT-CTCs. Ep-CTC abundance didn't show a correlation with disease trajectory (Figure 4.4G). Mutant KRAS levels tended to be higher at both time points in patients who are now deceased compared to those that are still alive, however it wasn't significant, which could be due to the small size of this pilot cohort (Figure 4.4H). The CTC samples were further profiled for prognostic, differentially expressed genes and dysregulated pathways. Broad gene expression profiling was performed using GeneChip® Expression Arrays for whole transcriptome profiling. Differential expression was defined as fold change > 2 and p-val <0.05[229]. In the pre-treatment samples, there were 1,710 genes that were significantly differentially expressed between the patients who are deceased vs alive

(Figure 4.5A), compared to only 872 in the on-treatment samples (Figure 4.5B), with 561 of the genes being commonly differentially expressed across both time points (Figure 4.5C).

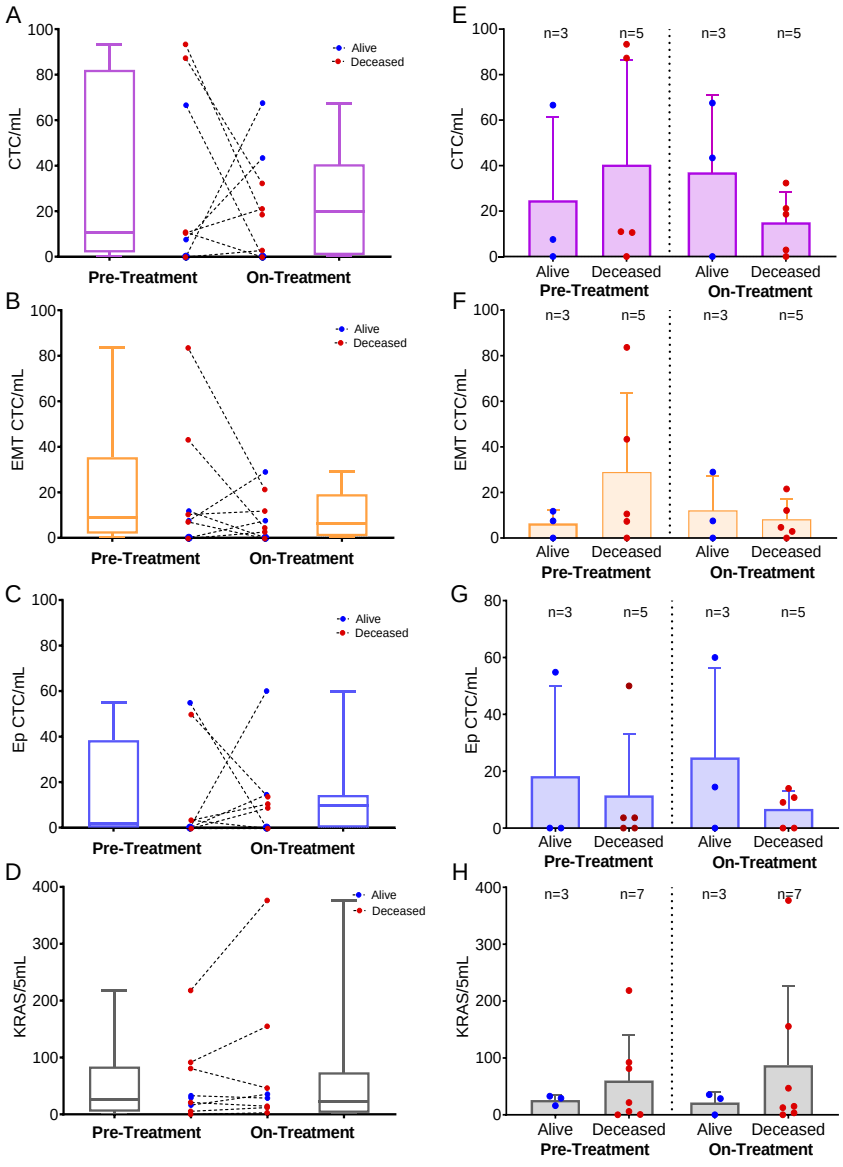


Figure 4.4 CTC dynamics based on patient survival.
A-D) Distribution of pre-treatment and on-treatment CTC samples with individual patient trajectories mapped for A) total CTCs, B) EMT CTCs, C) Ep CTCs, and D) mutant KRAS expression. Deceased patients are shown with red dots and alive patients are shown with blue dots. E-H) Comparison based on patient survival for E) total CTCs, F) EMT CTCs, G) Ep CTCs, and H) mutant KRAS. Error bars represent the standard deviation.

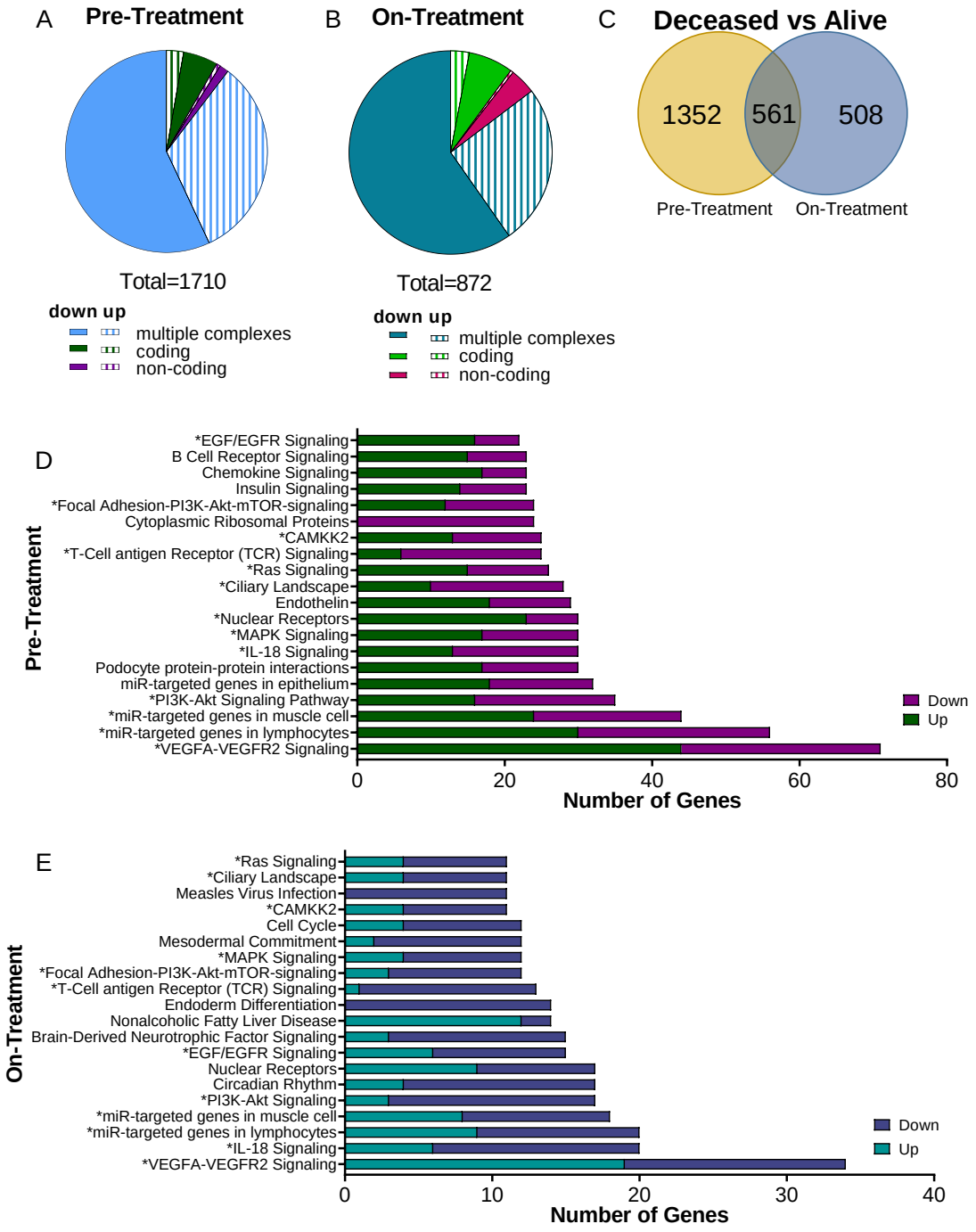


Figure 4.5 CTC differential gene expression and pathway analysis based on patient survival.

A,B) Differential gene expression based on patient survival for A) pre-treatment and B) on-treatment CTC samples. C) Comparison of differentially expression genes based on survival across time points. D,E) Top 20 deregulated pathways based on patient survival for D) pre-treatment and E) on-treatment CTC samples. *indicates the pathway was deregulated at both time points.

In pre-treatment samples, the fold change between differentially expressed genes in deceased versus alive patients ranged from 34.81 to -15.94, with some of the top differentially expressed genes based on patient survival were MMP9, Zeb2, S100P, MKI67, ALDHA2 and BCL2 (Table 4.3), which are known to be involved with cell phenotype, survival and proliferation, and disease aggressiveness[230–232]. Further, multiple S100 series A (S100A) genes, S100A8, S100A9, S100A11, and S100A12 where all upregulated in deceased patients (fold change, range = 4.73-34.81). S100A genes have shown growing interest in their involvement in pancreatic cancer progression, and in the epithelial to mesenchymal transition[233,234]. In on-treatment samples, the differentially expressed genes had fold changes ranging from 35.55 to -11.46, and displayed some overlapping top differentially expressed genes as the pre-treatment samples, such as MKI67, ALDHA2 and A disintegrin and metalloproteinase 9 (ADAM9) (Table 4.3).

Table 4.4 Key differentially expressed genes in pre-treatment and on-treatment CTC samples.

Gene	Fold Change		Change between On-Treament/Pre-Treatment
	Pre-Treatment	On-Treatment	
MMP9	22.98	--	↓
ALDH2	3.86	12.42	↑
ZEB2	3.43	--	↓
S100P	3.16	-2.97	↓
ADAM9	2.52	2.2	↓
MKI67	2.02	5.3	↑
MYC	-2.88	-2.38	↑
BCL2	-4.93	-4.63	↑
CTNNA1	3.95	4.02	↑

The ADAMs family of proteins is involved in several biological processes including the cell adhesion and migration, and proteolytic processing of other transmembrane proteins. ADAM9, has been previously reported to be upregulated in pancreatic cancer[235], and high ADAM9 expression in PDAC has been correlated with poor patient survival[236,237]. In these CTC samples, ADAM9 was shown to be upregulated at both time points in patients who are deceased. Further,

the CXC chemokine, CXCL16 was shown to be upregulated only in pre-treatment CTC samples of deceased patients. CXCL16 has been shown to be higher expressed in PDAC compared to healthy tissue, and is readily cleaved by ADAM proteins. Soluble CXCL16 has been shown to increase the invasiveness of pancreatic cancer cells[238] and ovarian cancer[239]

Beyond individual differentially expressed genes, many pathways were shown to be deregulated based on patient survival. The top 20 most dysregulated pathways based on survival are shown prior and during treatment (Figure 4.5D,E). Notably, known targets of oncogenic KRAS, mitogen-activated protein kinase (MAPK) signaling and phosphoinositide 3-kinase (PI3K)–AKT–mechanistic target of rapamycin (mTOR) signaling were shown to be differentially regulated based on patient survival[7]. Thirteen of the top 20 pathways appeared at both time points such as: VEGFA-VEGFR2, IL-18, Ras, PI3K-Akt signaling and EGF/EGFR signaling. (Figure 4.5D,E, Table 4.4). However the differentially expressed genes within each pathway only had moderate overlap across the two time points (Table 4.4). For genes that were differentially expressed at both time points, in every case it was differentially expressed in the same direction (i.e. up at both time points, or down at both time points).

Table 4.5 Differentially expressed genes from top 20 deregulated pathways in pre-treatment and on-treatment CTC samples based on patient survival.

Pathway	pre-treatment only		on-treatment only		both		pre-treatment only		on-treatment only		both	
	up	down	up	down	up	down	up	down	up	down	up	down
VEGFA-VEGFR2 Signaling	31	20	6	8	13	7	AP2A1, ARF4, ATP6V0D1, CAMKK2, CCRL2, CDC42BPB, EGR1, FOXO3, GAB1, LMO2, MAP2K3, MAPK1, MAPK14, MICAL2, MKNK1, PGD, PGK1, PRKCD, PTPRJ, QKI, RAC1, RAF1, SH3BGRL3, SIAH2, SLC8A1, STAT3, TAL1, TKT, TPP1, TUBA1C, TXN	ACKR3, ASCC3, CALR, DNAJB9, DSC1, EIF3H, EZR, FARSB, GIGYF2, HYOU1, LMAN1, NCL, NFATC2, PIK3R1, PRKCA, RPL10A, RPL5, RPL7, RPS6, TRPC1	ERG, GAPDH, HBD, HSPB1, LRRC59, SSR4	CREB1, FOXO1, NAP1L1, NFKB1, OCLN, PBXIP1, PRRC2C, STAT1	ADAM9, ALDOA, AP2S1, ATF4, BSG, CFL1, CTNNA1, FLII, GPX1, HDAC5, NAPA, PLA2G4A, TMOD1	APOLD1, BCL2, BIN1, ETS1, PLCG1, S1PR1, TNFRSF25
miR-targeted genes in lymphocytes	27	20	6	5	3	6	ANPEP, AP2A1, ARF4, ATP6V0A1, CORO1C, E2F3, IFRD1, MAPK14, METTL7A, MIR15A, MIR16-1, MTX1, NUCB1, PICALM, PPIF, PPP3R1, PTPRJ, RAB5C, RCOR1, RHOG, RTN4, SH3BGRL3, STX7, TM6SF1, UBE2J1, VAMP3, VTI1B	BCKDHB, CAND1, CHORDC1, GNPNAT1, GRPEL2, MATR3, MIR17, MIR20A, MSI2, MTRR, NCL, NT5E, PRIM1, RCN2, SLC25A32, SRPRB, SUCLG2, SYPL1, TAF9B, UTP15	ATP6V1F, BRI3BP, NEDD4, PPP2R4, RAB34, TYMS	KRAS, NUFIP2, PTBP2, SPCS3, SYNE1	CYP1B1, FADS2, G6PD	ANAPC1, BCL2, PDCD4, SLC38A1, SLC4A7, THEM4
miR-targeted genes in muscle cell	21	19	5	9	3	1	ANPEP, ARF4, ATP6V0A1, CORO1C, E2F3, IFRD1, LRP1, MAPK14, METTL7A, MIR15A, MIR16-1, NUCB1, PICALM, PPIF, PTPRJ, RHOG, RTN4, SH3BGRL3, STX7, TM6SF1, UBE2J1	ANAPC1, CAND1, CHORDC1, GNPNAT1, GRPEL2, MATR3, MIR17, MIR20A, MTRR, NCL, NT5E, PDCD4, PHLDB2, RCN2, SLC25A32, SLC38A1, SLC4A7, SRPRB, SYPL1	ERG, NEDD4, PPP2R4, RAB34, TYMS	ANAPC1, KRAS, NUFIP2, PDCD4, PTBP2, SLC38A1, SLC4A7, SPCS3, SYNE	CYP1B1, FADS2, G6PD	BCL2
PI3K-Akt Signaling	14	9	1	4	2	10	FLT3, FOXO3, GNB2, GNB4, HGF, IKBKG, JAK2, MAPK1, PHLPP1, PIK3AP1, PIK3CB, RAC1, RAF1, SYK	EIF4B, FASLG, FGF9, HSP90AB1, IL2RB, ITGA6, PIK3R1, PRKCA, RPS6	EIF4EBP1	CREB1, DDIT4, KRAS, NFKB1	ATF4, GNG5	BCL2, CCND2, FLT3LG, IL2RA, IL7R, ITGB7, LPAR6, MYC, RBL2, THEM4
IL-18 Signaling	9	6	2	3	4	11	CXCL16, GATA1, GSK3A, MAPK1, MAPK14, PRKCD, RAF1, STAT3, STK40	APBA2, FASLG, PARP1, PIK3R1, PRKCA, RPS6	HSPB1, PTX3	CREB1, IL18BP, NFKB1	BSG, CD36, GRN, IL18	BAZ1B, BCL2, BIN1, BIRC3, IFNG, IL2RA, LCK, LONP2, PLCG1, SLC4A7, TRAF1
MAPK Signaling	15	8	2	3	2	5	DUSP3, IKBKG, MAP2K3, MAPK1, MAPK14, MKNK1, PPM1A, PPP3R1, PRKCD, RAC1, RAF1, RAPGEF2, RASGRP4, STK3, ZAK	CACNA1I, FASLG, FGF9, HSPA8, MAP3K4, PPP3CC, PRKCA, RASGRF2	HSPB1, RRAS	KRAS, MAP4K1, NFKB1	ATF4, PLA2G4A'	CDC25B, DUSP16, ELK4, MYC, RASGRP
Nuclear Receptors	18	3	4	4	5	4	ABCC3, CPEB4, EGR1, FGD4, FTL, GSR, HGF, PGD, RXRA, S100P, SEC14L1, SLC6A6, SLC6A8, SRGN, STAT3, TSC22D3, TXN, VDR	HSP90AB1, MFGE8, SMC1A	GCLM, MGST3, PPP2R4, SLC2A13	BHLHE40, FOXO1, POU5F1, PPARD	BLVRB, CYP1B1, G6PD, MGST1, MGST2	BIRC3, IFNG, MYC, TGFBR3
Ciliary Landscape	7	16	1	5	3	2	AGPAT2, EXOC6, EXOC7, MKLN1, RAC1, RMND5A, WDR26	AIMP1, BBS9, CALM1, CLUAP1, EXOC6B, EXOSC2, EXOSC7, EXOSC9, MCM3, MCM7, MCM8, MSH2, NUP88, PGRMC2, TBC1D4, UQCC1	MCM4	APC, CNOT1, LRPPRC, NFKB1, PSMC6	GLA, LCN2, MYL6	CEP290, RNGTT
Ras Signaling	12	6	1	2	3	5	ETS2, FLT3, GAB1, GNB2, GNB4, IKBKG, MAPK1, PIK3CB, RAB5C, RAC1, RAF1, RASGRP4	CALM1, FASLG, PIK3R1, PRKCA, RASGRF2, ZAP70	RRAS	KRAS, NFKB1	CALM2, GNG5, PLA2G4A	ETS1, PLCG1, RASGRP1, SYNGAP1, TIAM1
T-Cell antigen Receptor (TCR) Signaling	6	10	1	3	--	9	IKBKG, MAPK1, MAPK14, PRKCD, RAF1, VAV3	CD247, CD3D, CD8A, FYN, GATA3, MALT1, NFATC2, PIK3R1, PRKCQ, ZAP70	DBNL	CREB1, MAP4K1, NFKB1	--	BLB, CD28, CD3E, CD3G, ICOS, ITK, LCK, PLCG1, SKAP1
CAMKK2	12	7	3	2	1	5	CAMKK2, CTSE, CYB5R3, ERMAP, FAM20C, HK2, MAP1LC3B, MAPK1, NEDD4L, RAC1, S100A8, TMEM176B	CALM1, NFATC2, PARP1, PLAC8, PRKCA, RPS6, SFI1	EIF4EBP1, HMBS, RHAG	CREB1, SELL	HK1	CAMK4, CCND2, CD28, MAF, RASGRP1
Focal Adhesion-PI3K-Akt-mTOR-signaling	10	7	1	4	2	5	FOXO3, GNB2, GNB4, HGF, IKBKG, JAK2, MAPK1, PHLPP1, PIK3CB, RAF1	EIF4B, FGF9, HSP90AB1, IL2RB, ITGA6, PIK3R1, RPS6	EIF4EBP1	CREB1, DDIT4, FOXO1, KRAS	ATF4, GNG5	IL2RA, IL7R, ITGB7, LPAR6, PIK3IP1
EGF/EGFR Signaling	12	3	2	6	4	3	AP2A1, ASAP1, GAB1, JAK2, MAPK1, MAPK14, PRKCD, RAC1, RAF1, REPS2, STAT3, VAV3	MAP3K4, PIK3R1, PRKCA	EIF4EBP1, NEDD4	CREB1, FOXO1, KRAS, MAP4K1, STAT1, STAT5B	AP2M1, AP2S1, AURKA, CFL1	CBLB, ELK4, PLCG1

4.4.5 Evaluating the Influence of Treatment Type on CTC Dynamics

Up to this point, all the samples have been stratified based on the time point or the patient status, alive or deceased. However, the patients in this cohort received diverse treatment regimens **(Table 4.1)**. The findings thus far highlight that a pre-treatment, single time point assessment may be insufficient to broadly predict patient prognosis across different treatment plans, and a more tailored approach is necessary. Patients were compared based on the treatments they received. Patients received one of five different treatments, all including gemcitabine (Table 4.1). On-treatment CTC gene expression profiles were compared using unbiased hierarchical clustering.

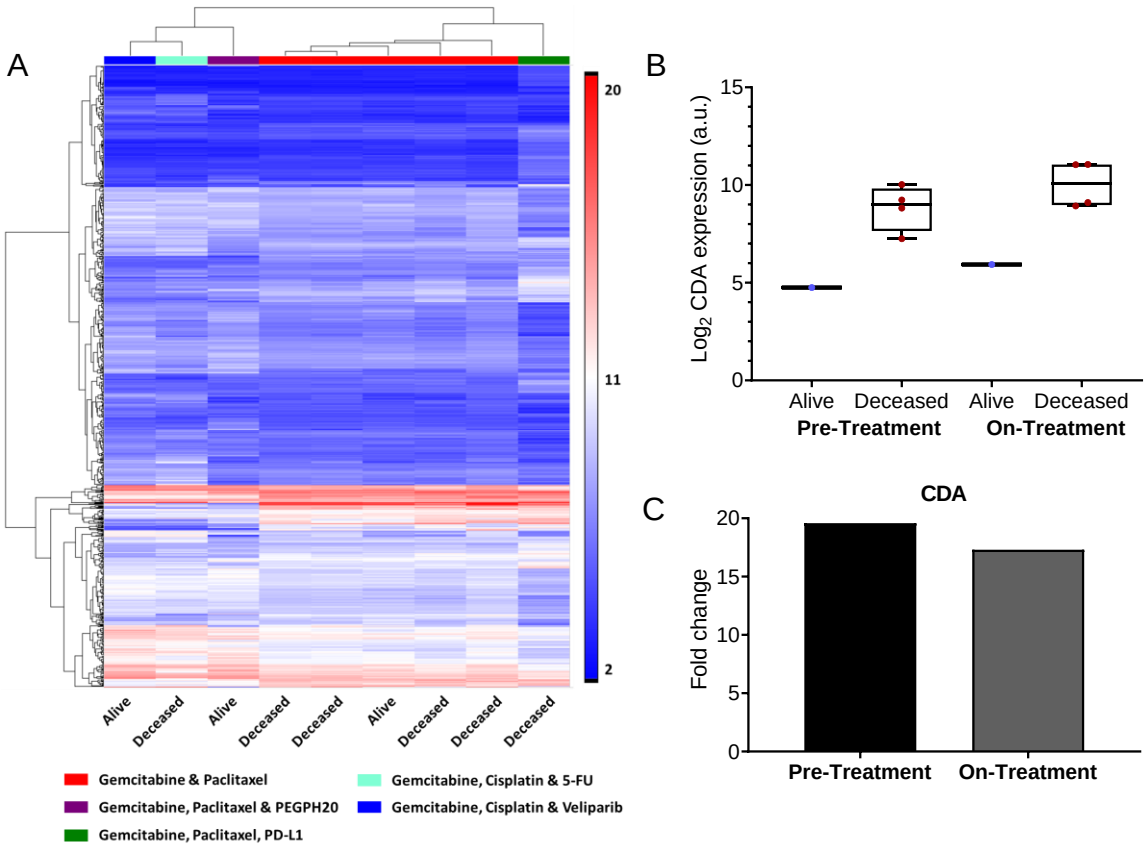


Figure 4.6 Effect of treatment on CTC transcriptome.
A) Hierarchical clustering of on-treatment CTC samples. Treatment is denoted with the colored bars at the top and patient status is shown below. B) Log_2 expression of CDA for patients received gemcitabine and paclitaxel therapy. C) Fold change in expression of CDA based on deceased vs alive at each time point.

CTC samples clustered based on the treatment each patient was receiving (Figure 4.6A), with similar treatment types also clustering closely together. Patients receiving cisplatin were differentiated from those receiving paclitaxel, independent of the specific combination therapy the patient was receiving. Strikingly, the differences induced by these two chemotherapy agents was stronger than those caused by the other treatments, such as the poly ADP ribose polymerase (PARP) inhibitor veliparib, programmed death ligand 1 (PD-L1) immunotherapies, or the hyaluronidase, PEGPH20.

In our cohort, only the gemcitabine and paclitaxel combination therapy had multiple patients for whom we had transcriptome profiling. Therefore, only these patients were further evaluated. In these patients there was a strong differential expression of cytidine deaminase (CDA) based on patient survival (Figure 4.6B,C). CDA is known to inactivate gemcitabine and can lead to drug resistance[240]. Interestingly, this differential expression was observed both prior to and while receiving therapy.

4.5 Discussion

Metastatic pancreatic cancer is extremely lethal with a median survival of 12 months, and 5-year survival rate of <3%[7]. In this study, we used a multipronged approach to evaluate CTC signatures of patient survival. CTCs were profiled based on their abundance, phenotype, KRAS mutation status, and gene expression signatures. In our initial cohort of metastatic, treatment naïve patients, the presence of EMT-CTCs, while not statistically significant, was higher in patients who are now deceased. Further, a shift towards more EMT phenotype was observed once the patient began therapy.

Previous CTC studies have identified KRAS mutant genotypes[221,222]. However, only a single time point assessment was used, and doesn't capture the temporal changes of the tumor cells. Our study builds upon this previous work finding 20/23 (87%) of CTC samples contain mutant KRAS. Concordant KRAS mutation status at both time points was observed in 9/10 patients , however one patient, patient 5, initially tested positive for a KRAS mutation, subsequently tested negative at the follow-up, on-treatment sample. Interestingly this patient is still alive 15 months after beginning therapy. Further, we observed increased mutant KRAS expression in patients with poor survival. This was associated with deregulation of known connected pathways such as PI3K-AKT-mTOR and MAPK signaling. Highlighting the need to not only screen for the presence, but also the expression levels of oncogenic mutations.

Additionally, increased MYC expression has been correlated with a more stem-like phenotype in other cancer types, however Sancho et al. found in pancreatic cancer that MYC expression was correlated with a differentiated phenotype[241]. These findings support our findings in this study, which observed downregulated MYC was found to be correlated with poor patient survival in our cohort.

CTCs revealed treatment-specific gene expression profiles. These signatures highlight the utility of CTCs to monitor the changes in the tumor cells over the course of therapy, which isn't possible through a traditional biopsy due to the invasiveness of the procedure and location of the pancreas within the body, making it particularly challenging to access. In the subset of patients being treated with gemcitabine and paclitaxel, we observed higher expression in deceased patients of CDA, a known inactivator of gemcitabine, even prior to the beginning of therapy. CTCs may provide a minimally invasive approach to predict patient response. Interestingly, this observation was conserved across the entire cohort in pre-treatment samples (deceased vs alive fold change =

22.5), but was lost when comparing on-treatment samples receiving different treatment plans. This suggests that the incorporation of other drugs into the treatment plan may circumvent CDA mediated gemcitabine resistance.

Effectively treating pancreatic cancer has been very challenging, believed largely to be due to the dense stroma. This desmoplasia limits the amount of cytotoxic agents that reach the tumor cells, therefore stoma modifying agents, such as PEGPH20, are being investigated to allow increased tumor penetration of the drug. However, phase III clinical trial evaluating PEGPH20 recently stopped, after showing no significant improvement compared to the control arm[242]. Interestingly, in our study, while only one patient, patient 3, who was receiving PEGPH20, saw a large increase of CTCs once beginning treatment. Prior to treatment, no CTCs were detected, however once beginning therapy, this number drastically rose to 67.5 CTC/mL. Further, nearly 90% of the CTCs were Ep-CTCs, the highest proportion of Ep-CTCs detected in on-treatment samples. It is possible that the use of PEGPH20 and decreased stroma allowed for extravasation of tumor cells into the bloodstream without undergoing EMT. While this is only one patient, it does suggest that simply improving tumor access for the drug may have unintended consequences such as easier tumor cell extravasation into the blood leading to increased CTCs and potentially driving metastasis. Further studies are needed to investigate the role of stroma modify agents on the presence of CTCs.

Collectively, we have shown that CTCs, despite their rarity in the blood, can enable multi-functional studies to evaluate signatures of patient survival in PDAC. Mutant KRAS expression and transcriptome analysis revealed mechanistic differences between patients that lead to poor patient outcomes.

Chapter 5 Natural killer Cells Offer Potential Therapeutic Approach to Control Metastasis through Circulating Tumor Cell Vulnerabilities

5.1 Abstract

Immunotherapy has rapidly grown in popularity and has been touted as a superior cancer therapy, yet it suffers from many limitations. Cancer cells have developed numerous mechanisms for immune escape, including physical exclusion and transformed gene expression, notably the down-regulation or loss of major histocompatibility complex I (MHCI) expression, limiting the success of T cell-based therapies. Circulating tumor cells (CTCs), separate the protection of the tumor microenvironment, present a unique opportunity to target and halt metastasis. In order to enter circulation, CTCs undergo the epithelial to mesenchymal transition (EMT), however, some of the genes involved in EMT are known to contribute to natural killer (NK) cell susceptibility. NK cells, don't rely on antigen presentation in MHCI and offer an alternative cell-based immunotherapy approach. While NK cells are known to rely on the balance between NK activating and inhibiting ligands expressed on cells, little is known about the relationship between these ligands that leads to NK activation. In this pilot study the expression of two NK inhibiting ligands – HLA-A/B/C and E-Cad – and two activating ligands – MICA/B and CADM1 – on cell lines are compared to their NK susceptibility. In a discovery cohort of eight metastatic non-small cell lung cancer (NSCLC) patients, we observed in CTCs that increased expression of NK activators correlated with better patient status compared to those with increased expression of NK inhibitors. This work provides a framework for CTC characterization to identify patients who would benefit from NK-based immunotherapy.

5.2 Introduction

Immunotherapy, which leverages the immune system to treat cancer, has led to striking successes in some patients[28]. Yet, tumor cells have developed a multitude of mechanisms for immune escape, including physical exclusion of immune cells, immune cell transformation, high levels of inhibitory cytokines and altered gene expression[29,243]. This has led to immunotherapy remaining ineffective in many patients. Identifying patients who would receive benefit from immunotherapy remains elusive and is an ongoing question.

Part of the innate immune system, natural killer (NK) cells don't rely on antigen presentation for activation; originally it was believed that the loss of major histocompatibility complex I (MHCI) was sufficient to lead to NK cell activation[244]. However, NK regulation relies on the complex balance of activating and inhibiting ligands on the surface of the cancer cells for activation. Upon activation, NK cells secrete cytolytic granules which mediate target cell apoptosis[29,245]. While activating and inhibiting ligands are well known, the requirements to shift the balance towards NK cell activation is poorly understood.

Circulating tumor cells (CTCs), rare cells shed from the tumor into the bloodstream en route to metastasize, present a non-invasive method to harvest a patient's tumor cells without the need of a tumor biopsy, and have been a prognostic factor in many cancer types[246–249]. Without the protection of the tumor microenvironment, CTCs may be more vulnerable to certain immune cell-based therapies, and are emerging as a target in rational therapy design[250,251]. During the first steps of metastasis, tumor cells undergo the epithelial to mesenchymal transition (EMT), in which they lose cell-cell adhesion and gain motility. However, some of the hallmark gene expression changes are also implicated in NK sensitivity presenting a potential therapeutic advantage[252].

The metastatic cascade is a highly inefficient process, with many CTCs being rapidly cleared from the circulation. It is currently estimated that up to 85% of CTCs are cleared within the first 5 minutes due to cell death, largely due to anoikis[253]. However, some CTCs are resistant to anoikis and continue to circulate throughout the bloodstream. Immune surveillance from natural killer (NK) cells have previously been shown to recognize and eliminate many of the remaining CTCs within the first 24 hours[254]. The use of NK cell based therapy may further improve the elimination rate of CTCs, thereby interrupting the metastatic cascade, and leading to better patient outcomes.

To identify the potential for using NK cell therapy to control metastasis, this study compares the expression of NK cell inhibitors[245,255–258] and NK cell activators[259–263] and their relationship to NK cell sensitivity using immortalized cancer cell lines and CTC-derived cell lines. In a discovery cohort of eight metastatic non-small cell lung cancer patients (NSCLC) patients, the relationship between CTC protein expression patterns of NK ligands is compared to clinical data. Further, multiple samples were analyzed from four of the patients and reveals preliminary evidence that the use of other treatments alters NK sensitivity and will require a sophisticated treatment plan to maximize patient benefit.

5.3 Materials and Methods

5.3.1 Cell Culture

Cells were maintained at 37°C under normoxic conditions. Cells were grown to 70-80% confluence before subculturing using 0.05% Trypsin-EDTA (Gibco). H1975, H1650 H3255, H441, A549, CTC-Lu1, and CTC-Lu2 cells were maintained in RPMI-1640 (Gibco) supplemented with 10% fetal bovine serum (FBS) (Sigma) and 1% Antibiotic-antimycotic (Gibco). NK92mi

cells were maintained in suspension culture in Minimum Essential Medium-alhpa (MEM-α) supplemented with 10% FBS (Sigma), 12.5% horse serum (Corning) and 1% Antibiotic-antimycotic (Gibco). Media was exchanged every 48-72 hours between subculturing. Cell lines were routinely tested and reported negative for mycoplasma contamination (Lonza).

5.3.2 Sample Collection and Labyrinth Processing

The experimental protocol was approved by the Ethics (Institutional Review Board) and Scientific Review Committees of the University of Michigan and all patients gave their informed consent to participate in the study. All patients had a diagnosis of metastatic EGFR mutant or ALK mutant lung adenocarcinoma.

Blood samples were collected in EDTA tubes and processed through the Labyrinth within 2 hours of collection. Red blood cells (RBCs) were removed with Ficoll-Paque™ PLUS Media (GE Healthcare) based on density separation principles, following the manufacturer's protocol prior to processing in the Labyrinth.

After RBCs depletion, the plasma and blood mononuclear cells (PBMCs) fractions were collected and diluted 1:5 with phosphate buffered saline (PBS) based on the original blood volume. The subsequent diluted sample was processed through the Labyrinth at a flow rate of 2500 μL/min. The resultant from outlet 2 was collected. The CTC-enriched sample was used for immunofluorescent staining of Panels 1 and 2.

5.3.3 Immunofluorescence and Fluorescent Microscopy

To perform immunofluorescent staining, cells either from culture or from the Labyrinth outlet 2 product were processed using Cytospin Cytocentrifuge (ThermoFisher Scientific). A poly-lysine coated slide was placed into the cytospin funnel (ThermoFisher Scientific), cell suspension was added to each cytospin funnel and cytocentrifuged at 800 revolutions/min (RPM) rpm for 10

min. Samples were then fixed on the cytoslides using 4% paraformaldehyde (PFA) and cytocentrifuged at 800 RPM for an additional 10 min. The cells on the glass slides were covered with PBS and stored at 4°C until used for immunofluorescence. Two slides were prepared from each patient sample or cell line, for the staining of panel 1 and panel 2 respectively.

Cells or samples on slides were permeabilized with 0.2% Triton X-100 solution for 3 min. Slides were then washed 3 times with PBS, followed by blocking using 10% goat serum for 30 min at room temperature (rt). Primary antibody cocktails with following concentrations (table) for each marker were prepared in blocking reagent (10% goat serum) and applied to cover the whole slide sample surface. Slides with primary antibodies are then incubated at 4°C overnight. On the following day, slides were washed with PBS for 3 times, with 5min incubation each time. Secondary antibody cocktails with concentrations shown below were prepared in blocking reagent and applied to cover the sample area for 1.5 hrs. This was followed by three 5-min PBS washes. The coverslips were then mounted onto the sample areas with Gold prolong antifade mountant reagent with DAPI.

The entire sample area of the stained slides were scanned using a Nikon Ti2 inverted fluorescent microscope at 20X magnification for CTC identification and characterization.

Table 5.1 Staining panels for NK activating and inhibiting ligands.

Staining Panel	Channel	Primary Antibody	Catalog number	Host (Isotype)	Staining cocktail dilution	Catalog number (All purchased through Thermofisher)	Staining cocktail dilution
Panel 1 & 2	Cy3	Pan Cytokeratin (panCK)	BioRad MCA1907T	Mouse (IgG1)	1:100	A21123	1:100
Panel 1 & 2	FITC	CD45	enQuire 5788-MSM9	Mouse (IgG2b)	1:100	A21141	1:100
Panel 1	Cy5	HLA-A/B/C	Thermo/eBioscience 14-9983-82	Mouse (IgG2a)	1:50	A21241	1:100
Panel 1	Cy7	E-Cadherin	Cell signaling 31955	Rabbit (IgG)	1:50	SA5-10035	1:100
Panel 2	Cy5	MICA/B	Abclonal A12622/A9802	Rabbit (IgG)	1:50	A21245	1:100
Panel 2	Cy7	CADM1	MBL International CM004-3	Chicken (IgY)	1:50	SA5-10075	1:100

5.3.4 CTC Analysis and Enumeration

The images from the scanned CTC slides were manually analyzed using the NIS-Elements software (Nikon). CTCs were defined as being panCK+/CD45-. CTCs were further characterized by their expression of NK inhibitors (Panel 1) or NK activators (Panel 2).

5.3.5 Quantitative Fluorescent Intensity Data Export

Regions of Interest (ROIs) were defined for each cell (cancer cell line or CTC) manually based on PanCK and DAPI staining using NIS Analysis software (Nikon). Due to the MHC class I expression in residual platelets on the slide, a background subtraction step was done on panel 1 (HLA-A/B/C, E-cad) using the subtract background function in ImageJ. Rolling ball radius for the algorithm was set to the pixel number of the largest cell of interest on each image. Then, a multidimensional dataset within defined ROI, including the mean intensity values for each channel, were exported as an excel file using the NIS-Elements software ROI statistics module. Intensity z-score $Zmarker$ for a single marker (inhibitor or activator) for each cell line or CTC was calculated using the following formula:

$$Zmarker = \frac{x - \mu}{\sigma}$$

For each cancer cell line, x is the average intensity of cells that were chosen randomly across the slide. μ is the average intensity between different cell lines. σ is the standard deviation of the mean intensity for different cell lines.

For each CTC, x is the mean intensity of defined ROI. μ is the average intensity of all CTCs found between different patients that were stained with the same marker. σ is the standard deviation of the intensity of all CTCs found between different patients that were stained with the same marker.

5.3.6 NK Cytotoxicity Assays and data analysis

Cancer cells were labeled with cell tracker red CMTPX (Invitrogen) for 45 min in serum free media, and washed with PBS 3 times. The cell concentration was counted by a hemocytometer. Cells were diluted to a concentration of 50,000 cells/mL. 5,000 cells were seeded into each well of a 96 well plate (Corning) and adhered to the dish overnight. The next morning NK92mi cells were collected, and resuspended in 1:1 mix of RPMI-1640 and MEM-α. Cells were counted by hemocytometer and diluted to the appropriate concentration to seed 100 μL of NK cell suspension to each well, containing the indicated ratio of NK:cancer cells. After preparing the NK cell dilutions, media was gently removed from each well in 96 well plate and replaced with NK cell suspension. The co-culture was incubated for four hours.

At the end of the four hour co-culture incubation period, each well was washed with PBS, then stained with a cocktail containing Calcein AM (BD bioscience, final concentration ?) for live cell staining and hoechst (Invitrogen, final concentration 10 μg/mL) for nuclei staining. After incubating in dark at 37 degree C for 20min, the staining reagents were gently removed from each well and replaced with PBS. The plate was then scanned with three channels (DAPI, FITC, Cy3) using a Nikon Ti2 inverted fluorescent microscope at 10X magnification.

The images are then batch analyzed using the General Analysis module using the NIS-elements software. For the layer in each channel, “Remove Average Background” function was applied as preprocessing. Appropriate intensity and size thresholding was then set based on each cell line image to ensure optimal cell counts. Post processing functions such as “Morpho Separate Objects”, “Fill Holes", “Smooth” were used accordingly. An intersection between the layers in each channel was then generated as a combined binary layer. Object number was then recorded

based on the combined layer as live cancer cell number in each well, as they represent Cy3+, FITC+ and DAPI+ cells.

Percent specific lysis is calculated using the following equation:

$$\%\, Specific\ Lysis\ = 100 - \frac{Nexp}{Nc} \times 100$$

$Nexp$ is the live cancer cell count in the experiment wells with the presence of NK cells. Nc is the mean of live cancer cell counts of the replicate control wells containing target cancer cells alone. Relationship between percent specific lysis (Y) and NK cell to target cancer cell ratio (X) was fitted using a built-in nonlinear hyperbolic regression (X is concentration) model in GraphPad Prism 8.4.0.

5.3.7 Statistical Analysis

Statistical analysis was performed using GraphPad Prism 8.4.2. Significance and p-values between samples were determined using a two-sided unpaired t-tests assuming a Gaussian distribution, with a 95% confidence level. Plotted error bars represent standard deviation.

Intensity z-score heatmap was generated using R package ComplexHeatmap[264]. Complete-linkage hierarchical clustering was performed between cell lines or patient sample CTCs.

5.4 Results

5.4.1 Assessment of NK Susceptibility in Cancer Cells Based on NK Activating and Inhibiting Ligand Expression

To enable the rapid assessment of CTCs for NK susceptibility, the protein expression patterns of activating and inhibitory signals needed to be correlated to NK susceptibility. Using immunofluorescence, the relative protein expression was characterized across six immortalized

cancer cell lines - five lung cancer, one pancreatic cancer - and five CTC-derived cell lines: two NSCLC and three pancreatic adenocarcinoma (PDAC) - for NK inhibiting ligands, human leukocyte antigen A/B/C (HLA-A/B/C) and E-cadherin (E-Cad), and the NK activating ligands, MHC class I polypeptide-related sequence A/B (MICA/B) and cell adhesion molecule 1 (CAMD1) (Figure 5.1A).

The eleven cell lines displayed diverse protein expression levels of the four markers, however, based on hierarchical clustering cell lines tended to cluster based on their original isolation site (Figure 5.5B). Lung cancer cell lines A549 and H1975 were isolated from the primary tumor, whereas lung cell lines H441, H1650, and H3255 were isolated from secondary sites: pericardial fluid and pleural effusions. While the pancreatic cancer cell lines, Capan-2, was isolated from the primary tumor, it clustered most closely with those from secondary sites. Interestingly, all five CTC lines, independent of cancer type, clustered most closely with cell lines isolated from primary sites.

The NK sensitivity of five lung cancer cell line was evaluated through NK:cancer cell co-cultures. Cancer cells were incubated with NK92mi cells at four different NK:cancer cell ratio for four hours, and the number of remaining viable cancer cells based was quantified. The % specific lysis was calculated using the number of viable cancer cells at each NK:cancer cell ratio compared to a no-treatment control (Figure 5.5C). The two lung CTC-derived cell lines were the some of the most sensitive, and A549 was the least sensitive. A549 has extremely low HLA-A/B/C expression and using this single marker would predict NK sensitivity, however A549 is known to have high E-Cad expression. The CTC-derived cell lines had low expression of both NK inhibitors HLA-A/B/C and E-Cad (nearly completely absent in CTC-Lu1), and moderate expression of NK activators MICA/B and CADM1, most prominently shown in CTC-Lu2.

H1650 and CTC-Lu2 showed similar NK susceptibility, despite each cell line having very different expression levels of each marker, highlighting the complexity of NK signaling regulating NK activation. NK activation wasn't dependent on the absolute level of NK activator or inhibitor expression, rather on the complex balance between them.

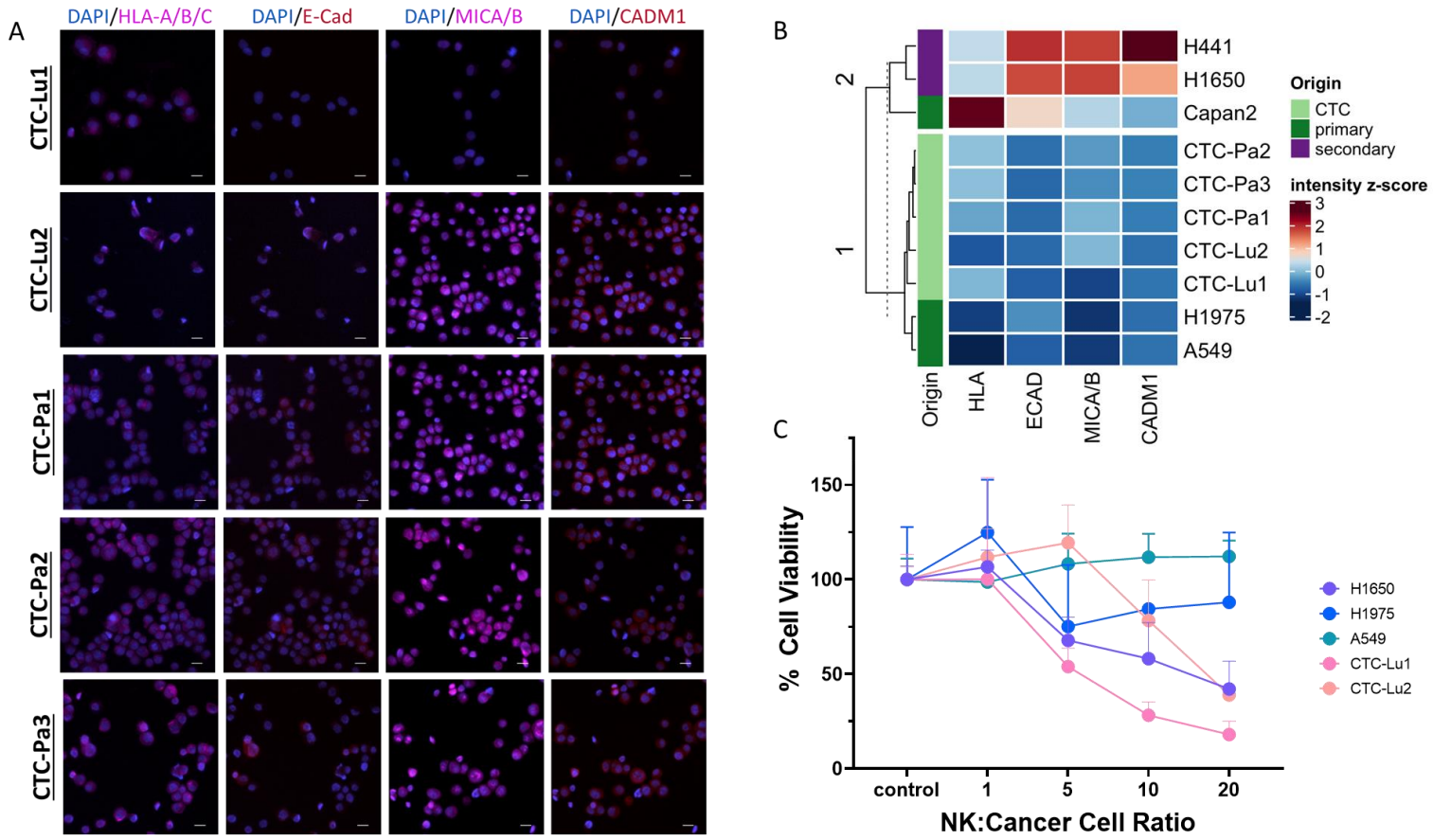


Figure 5.1 Expression of NK activating and inhibiting ligands and NK sensitivity.
A) Representative immunofluorescent images of expression of NK inhibitors (HLA-A/B/C and E-Cad) and NK activators (MICA/B and CADM1) in five CTC-derived cell lines. B) Hierarchical clustering of cell lines based on the median z-score expression per marker across cancer cell lines. Their original isolation site is shown. C) Percent viability of cancer cells remaining after co-culture for four hours with different ratios of NK cells. Error bars represent the standard deviation of triplicate experiments.

5.4.2 Patient Demographics and Study Design

After comparing NK sensitivity in cancer cell lines, the correlation between NK activating and inhibiting ligand protein expression was applied to analyze CTCs. Eight patients with metastatic NSCLC patients gave their informed consent and were enrolled in the study, Table 5.1. Four patients had two or more samples collected, while the remaining four only had one time point. The average patient age was 62 years old, and was well distributed between male and female, additional patient demographics are shown in Table 5.2.

Table 5.2 Patient clinical characteristics.

Patient	Mutation	Visit	Treatment Status	Therapy	Treatment Type	Status
1	ALK	1	pre-treatment	--	--	Baseline, Stable
		2	on-treatment	Alectinib	Kinase inhibitor	Stable
2	EGFR exon 19 deletion	1	on-treatment	Carboplatin & pemetrexed	Chemotherapy	Progression
		2	on-treatment	Osimertinib	Kinase inhibitor	Progression
3	EGFR exon 19 deletion	1	pre-treatment	--	--	Progression
		2	on-treatment	Osimertinib	Kinase inhibitor	Stable
		3	on-treatment	Osimertinib	Kinase inhibitor	Stable
4	EGFR exon 19 deletion	1	on-treatment	Osimertinib	Kinase inhibitor	Progression
		2	on-treatment	Osimertinib	Kinase inhibitor	Progression
5	ALK	1	on-treatment	Alectinib	Kinase inhibitor	Stable
6	EGFR L858R & T790M	1	on-treatment	Osimertinib	Kinase inhibitor	Progression
7	EGFR exon 19 deletion	1	on-treatment	Osimertinib	Kinase inhibitor	N.A.
8	EGFR exon 19 deletion	1	on-treatment	Osimertinib	Kinase inhibitor	N.A.

N.A. = not available.

Two patients carried mutations in Anaplastic lymphoma kinase (ALK) and six carried mutations in epidermal growth factor receptor (EGFR) based on screening of the primary tumor. Most patients were receiving kinase inhibitor treatments, however one patient (patient 2) was originally receiving chemotherapy until next generation sequencing results were available and identified an EGFR mutation, at which point the patient began kinase inhibitor therapy.

Table 5.3 Patient demographics.

Patient	Sex	Age (years)	Stage	Node status	Tumor	Tumor origin	Smoking status	Race	Ethnicity
1	Male	50	IV	N2	T2	LLL	No	White	Non-Hispanic
2	Female	45	IV	N3	T4	LLL	Former	White	Non-Hispanic
3	Male	70	IV	N0	T2a	LUL	Former	White	Non-Hispanic
4	Male	66	IV	N0	T1a	LLL	Never	White	Non-Hispanic
5	Male	58	IV	Nx	T3	RLL	No	White	Non-Hispanic
6	Female	61	IV	N3	T2	LUL	No	White	Non-Hispanic
7	Female	82	IV	N.A.	N.A.	LUL	Former	White	Non-Hispanic
8	Male	67	IIIB	N.A.	N.A.	RLL	Never	White	Non-Hispanic

5.4.3 Circulating Tumor Cell Characterization Workflow

Due to the rarity of CTCs in the blood, typically only the order of 10 CTCs/mL of blood, surrounded by millions of white blood cells (WBCs), many CTC isolation technologies have been developed, most commonly through microfluidics. Our lab has recently published the Labyrinth, a high throughput, inertial CTC isolation microfluidic device. It achieves >90% recovery of CTCs, and >90% depletion of WBCs, allowing for rapid isolation of CTCs from large starting blood volumes. Due to the label free isolation of CTCs from the Labyrinth, heterogeneous CTCs subpopulations of CTCs can be isolated. After CTC isolation, the CTCs can be used for multiple analysis including immunocytochemistry (ICC), culture and expansion, and functional analysis. To have ample cell numbers for functional analysis, we have previously generated five CTC-derived cell lines, two from non-small cell lung cancer (NSCLC) and three from pancreatic ductal adenocarcinoma (PDAC). In this cohort of eight NSCLC patients, freshly isolated CTCs from the Labyrinth were used for CTC characterization using ICC. The CTC sample was split across two slides to be characterized based on the two staining panels, panel 1 and panel 2 (Figure 5.2).

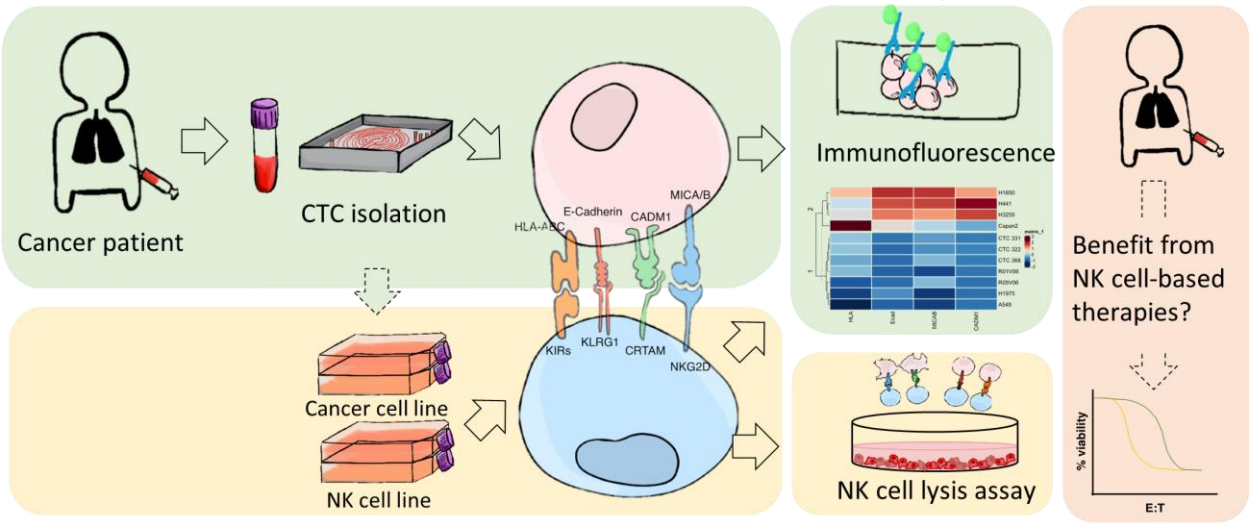


Figure 5.2 Project overview for evaluating NK sensitivity of CTCs.
CTCs are isolated from metastatic NSCLC patients using the Labyrinth, and inertial, label-free microfluidic device. The freshly isolated CTCs are profiled using ICC for protein expression of NK activating and inhibiting ligands. The expression profiles of these markers on CTCs is compared to the expression profiles of cell lines. Cell lines are also evaluated for NK sensitivity through co-culture experiments, which can be used to predict NK sensitivity of CTCs in patients.

5.4.4 CTC Characterization and Inter-patient Heterogeneity

CTCs were identified using ICC based on their expression of cytokeratin (CK) and lack of CD45, a common WBC marker. The CTCs were further profiled for their expression of either NK inhibitors (panel 1 - HLA-A/B/C (HLA) and E-Cad) or NK activators (panel 2 - MICA/B and CADM1). The CTCs showed a wide range of expression for the NK activators and inhibitors (Figure 5.3).

Due to the presence of HLA-A/B/C being expressed on WBCs and platelets, there was significant background signal observed that was both patient and spatially dependent. Therefore, the images were further processed using an ImageJ algorithm to remove the local average background prior to further characterization. After background subtraction, the mean fluorescent intensity of each marker was evaluated for each CTC. The expression of each marker across CTCs

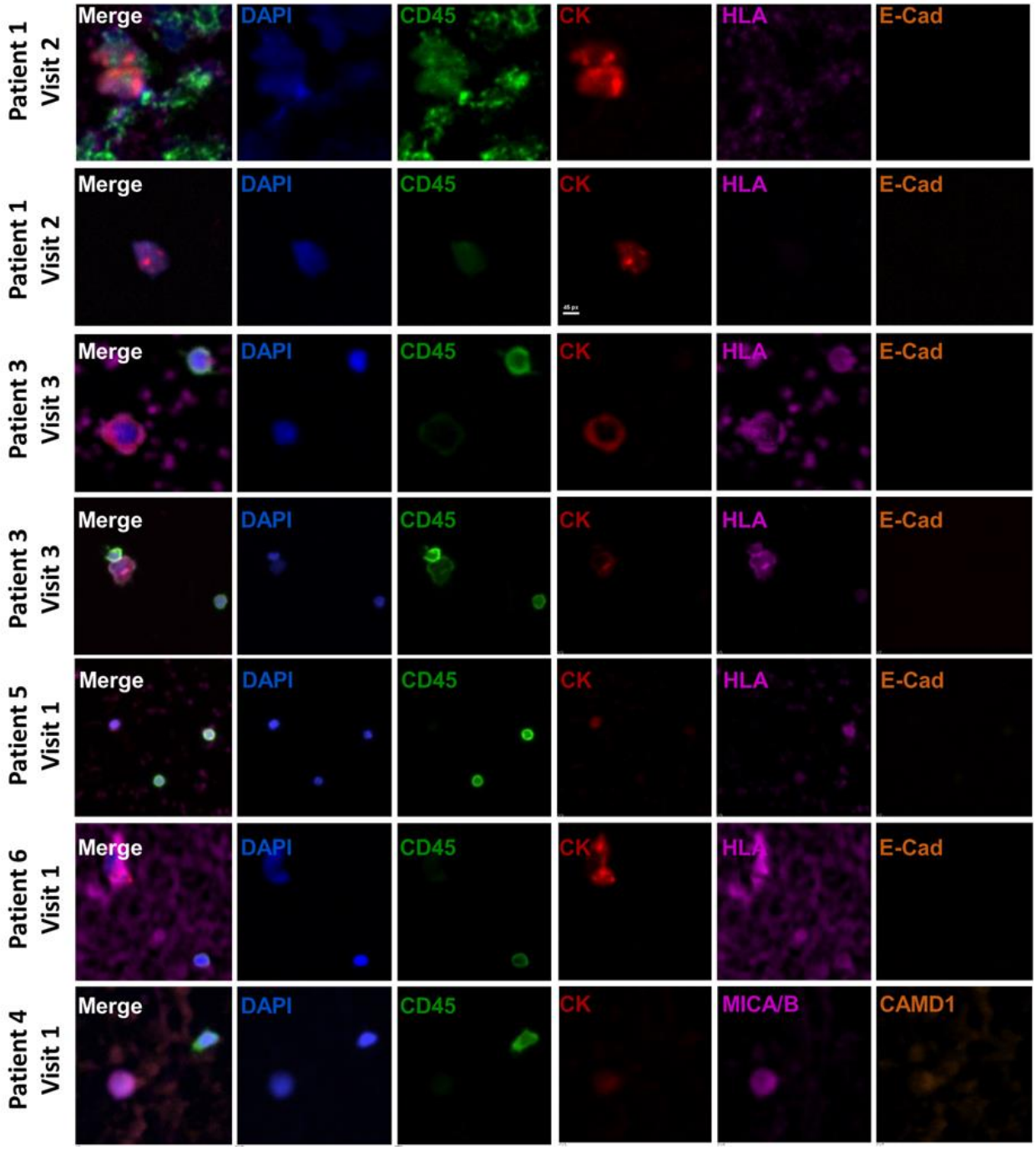


Figure 5.3 Representative immunofluorescent images of CTCs evaluated for NK inhibitors or NK activators from NSCLC patients.
CTCs are identified as cells being positive for CK (red) positive and CD45 negative (green). The expression of NK activating or inhibiting ligands are shown in pink (HLA-A/B/C (panel 1) or MICA/B (panel 2) and orange (E-Cad (panel 1) or CADM1 (panel 2)).

was compared on the z-score. To compare inter-patient heterogeneity the average z-score expression of each marker within a sample (average of all CTCs within a sample) was compared. The patients with multiple time points are plotted each with a unique color, while the remaining patients with only one point are shown in grey. With the exception of patient 3, the patients tend

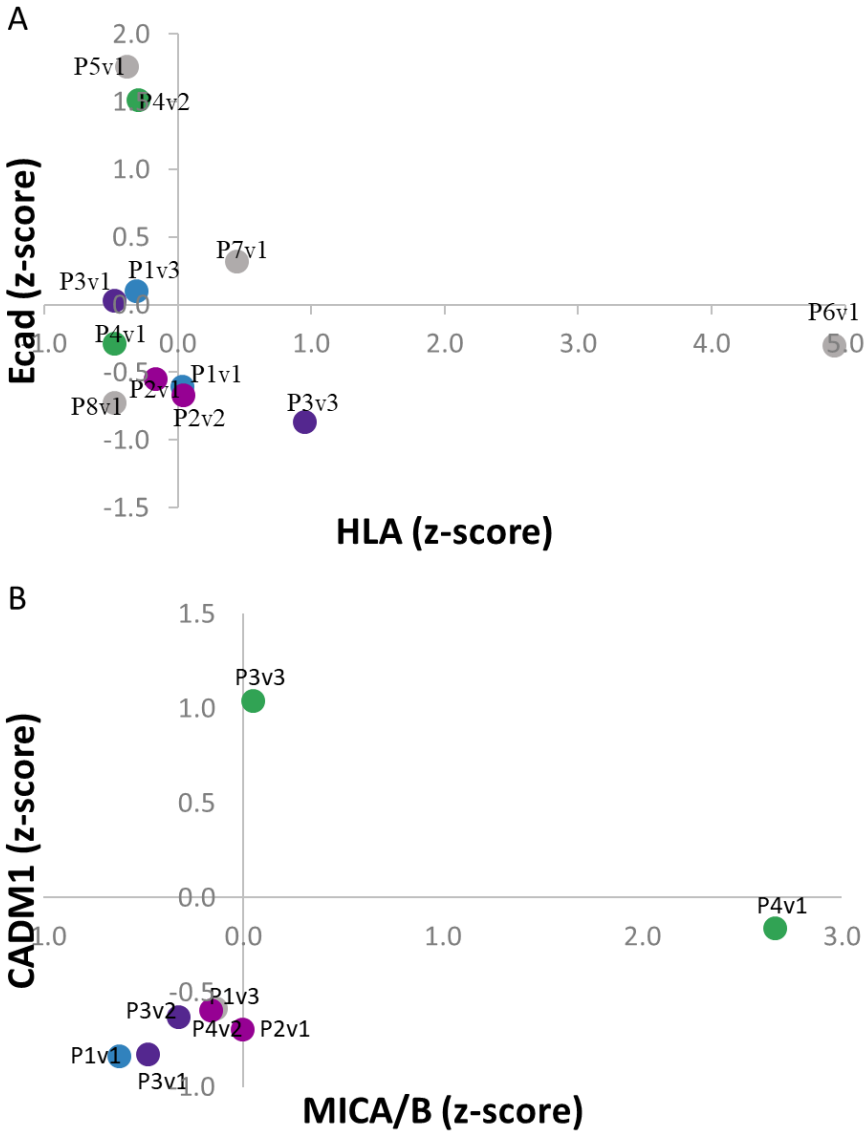


Figure 5.4 Inter-patient heterogeneity of NK activating and inhibiting ligand expression in CTCs across patients and time points.
All CTCs (across all patients and time points) were used to generate the z-score expression per marker per CTC. The average z-score per marker across all CTCs was calculated within in sample (one time point for a given patient) for A) NK inhibiting ligands and B) NK activating ligands. Patients with multiple time points are shown in unique colors, while patients with only visit are all shown in grey.

to have similar levels of NK inhibitor (Figure 5.4A) or NK activator (Figure 5.4B) expression. Patient 3 shows high levels of MICA/B coupled with moderate levels of CADM1 at the first visit, but this trend is reversed at the third visit.

5.4.5 Correlation between Expression of NK Activating and Inhibiting Ligands and Patient Clinical Characteristics

To explore the clinical relevance of the expression of these markers on CTCs, their expression was correlated with clinical data across multiple visits including the presence of specific mutations, treatment, and clinical status.

The CTCs, based on NK inhibitor expression, clustered into three main populations based on unbiased, hierarchical clustering; the first population with high HLA-A/B/C, low E-Cad expression, the second with low HLA-A/B/C, high E-Cad expression, and the third with moderate levels of both markers (Figure 5.5A). While each cluster contained diverse CTCs, they did appear to be enriched for certain subpopulations. Cluster 1 was enriched for earlier visits and patients with stable disease. Whereas clusters 2 and 3 were associated with later visits and progressive disease. As shown, the patients with stable disease tended to have reduced expression of both NK inhibitors (cluster 1).

For the NK activators, the CTCs clustered into two main clusters, one with high levels of MICA/B and/or CADM1 and one with moderate levels of both markers (Figure 5.5B). Similarly, cluster 1 had the most diverse CTCs, but was enriched for earlier visits and patients with progressive disease. Cluster 2 was associated with later visits and stable disease. The patients with stable disease tended to have high expression of at least one NK activator (cluster 2).

There appears to be a synergy between low NK inhibitor expression and high NK activator expression with favorable patient clinical status.

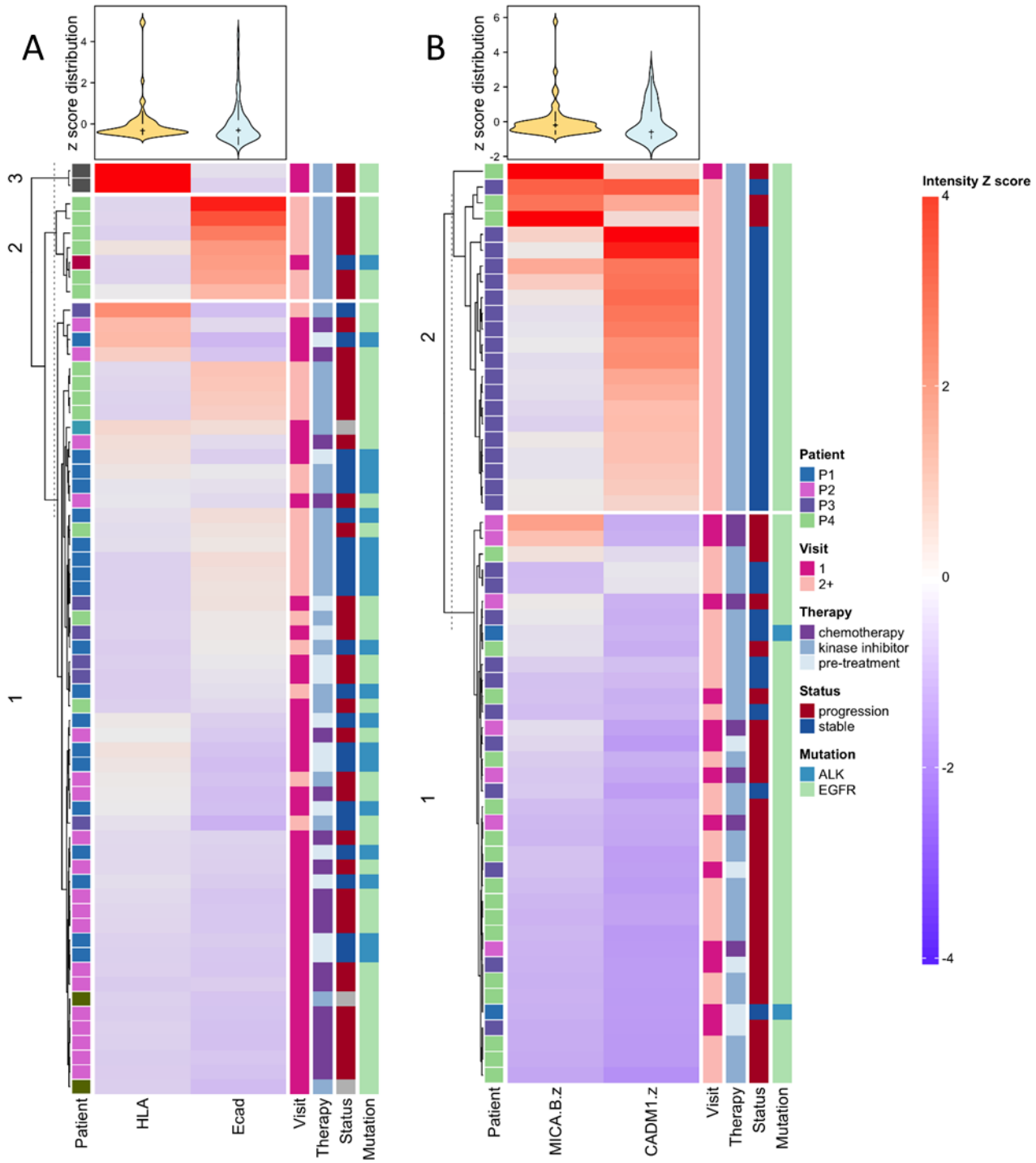


Figure 5.5 CTCs display heterogeneous expression profiles of NK activating and NK inhibiting ligands.

Hierarchical clustering of CTCs from eight NSCLC patients based on z-score expression of A) NK inhibitors or B) NK activators. Corresponding clinical data is shown as additional columns to the right of each heatmap. A violin plot of z-scores for each marker is shown above each heatmap and shows the expression distribution across all analyzed CTCs.

The clinical relevance of each marker was further explored based on inter- and intra-patient heterogeneity, intra-patient temporal changes, clinical status, treatment, and mutation status. The patients demonstrated heterogeneous expression levels across the four markers (Figure 5.6A). For the four patients with multiple samples, expression levels changed between visits (Figure 5.6B). Patient P1 became overall more NK sensitive between visits 1 and 2, as seen by decreased HLA-A/B/C, and increased MICA/B and CADM1. Conversely P4 became more NK resistant over time, indicated by increased HLA-A/B/C and E-Cad, and decreased MICA/B and CADM1. These observations correlated with their clinical status, with P1 having stable disease and P4 having progressive disease (Figure 5.6B, Table 5.2)

Further supporting the impact of NK cells on disease management, patients with stable disease had very different expression profiles than those with progressing disease (Figure 5.6C). Across the four markers, CADM1 expression had the strongest correlation with patient status. Patients with progressing disease had significantly lower levels of CADM1 compared to those with stable disease.

The vast majority of patients were receiving kinase inhibitor therapies, with only one patient originally receiving chemotherapy before switching to a kinase inhibitor. However, initial evidence points to a treatment-dependence on expression of E-Cad and CADM1 (Figure 5.6D). The mutations carried by the primary tumor did not impact the expression of any of the four markers (Figure 5.6E).

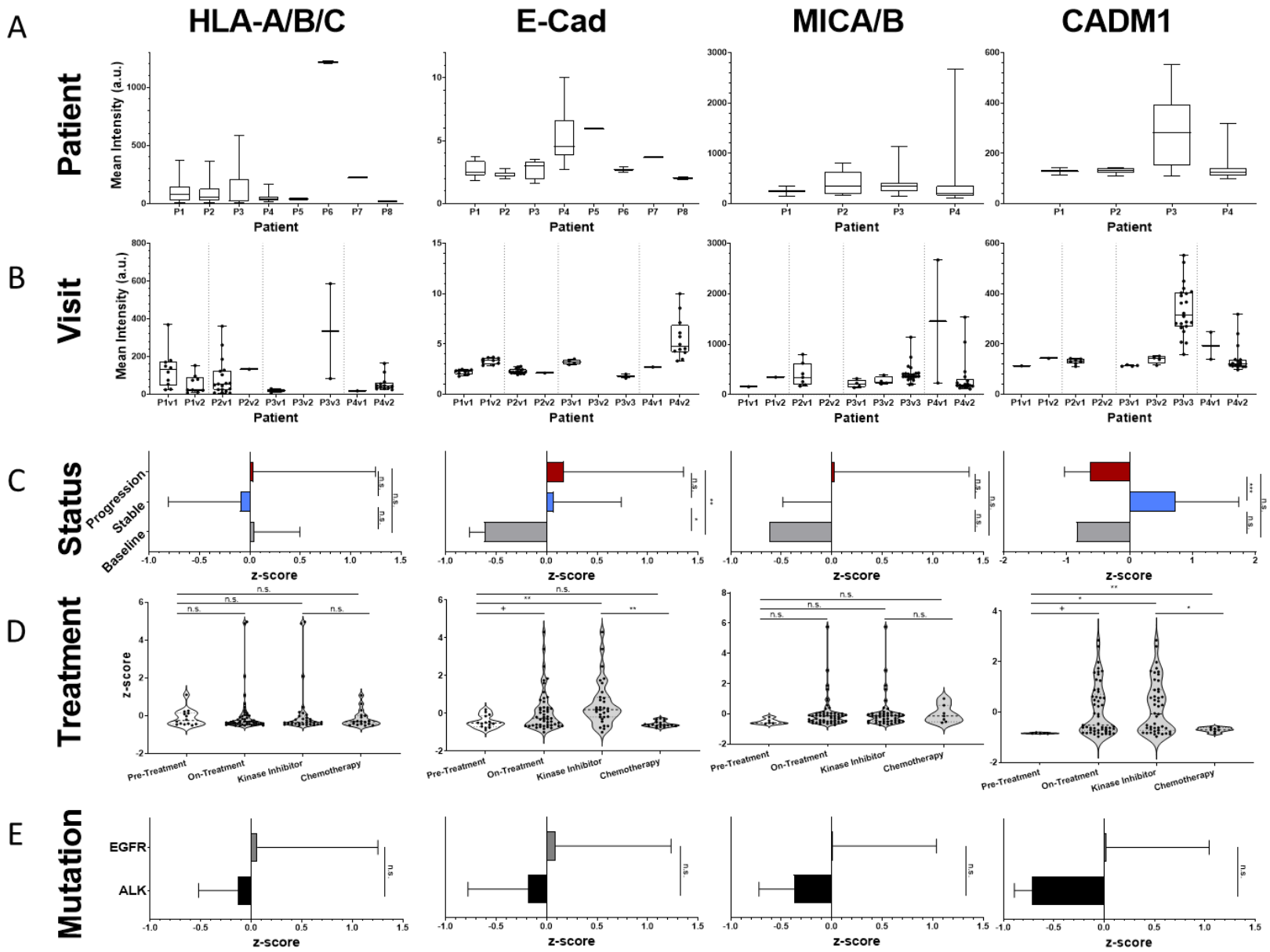


Figure 5.6 Correlation between NK ligand expression and clinical metrics.
Expression of each marker based on A) patient, B) visit for patients that have more than one time point, C) clinical status, D) treatment the patient is receiving, or E) mutation the tumor harbors based on tumor biopsy. Error bars represent standard deviation. $*p < 0.05$, $** p < 0.01$, $***p < 0.001$

5.5 Discussion

This study presents initial evidence that CTCs may be effectively targeted by NK cells and the use of a NK-based therapy may further improve their clinical impact, particularly patients who have CTCs with NK sensitive profiles, yet still have progressive disease. This study focused on the balance between two NK inhibiting ligands - HLA-A/B/C and E-Cad, and two activating ligands - MICA/B and CADM1. CTCs present a unique vulnerability to NK cells compared to the primary tumor cells. Not only do CTCs lack protection due physical exclusion of NK cells from the tumor microenvironment, in circulation they also are in TGF-**β**, a known NK cell inhibitor, scarce environment[184,259]. Further, during the first steps of the metastatic cascade tumor cells undergo EMT, typically associated with an aggressive and treatment resistant phenotype. However, during EMT E-Cad is downregulated and CADM1 is upregulated, therefore increasing NK sensitivity[256,259,265].

Despite the apparent intersection between NK cells and halting metastasis through CTC clearance, little work has been done to characterize the expression of the ligands on CTCs that regulate this immune response. Pantel et al. reported heterogeneous downregulation of MHCI expression in disseminated tumor cells (DTCs) found in the bone marrow from breast, colon, and stomach cancer patients[266]. The authors found that patients with poorly differentiated tumors had less frequent expression of MHCI on DTCs compared to those with moderately differentiated tumors. Additionally, Ramsköld et al. found low expression in melanoma CTCs of HLA-B, C, H and G compared to primary melanocytes and suggested this may be a mechanism for CTC escape from immune surveillance[133]. Finally, Huaman et al. showed in human xenograft mouse models using two cancer cell lines, the resulting CTCs had a 1.8-fold and 1.5-fold decrease in expression of MHCI expression compared to the primary tumor, respectively[267]. The reduced

expression in MHC1 in the CTCs was sustained over multiple passages when cultured. Lo et al. showed that CTCs traveling as clusters may be more resistant to NK-mediated death compared to single CTCs[268]. Further the authors show that EMT and expression profiles leading to NK cell activation are highly correlated, and using microRNA as a method to modulate EMT and confer NK sensitivity. These works lay the initial groundwork suggesting MHCI-low cells may survive the entirety of the metastatic cascade and develop into metastatic tumors.

The rarity of CTCs in the blood limits their ability to be compatible with traditional cytotoxicity assays which require large starting cell numbers. However, CTC expansion into stable cell lines has been limited, suffering from low success rates and long culture times, further limiting their immediate clinical impact[269,270]. Five previously established CTC lines were leveraged in this study. Their expression profiles most closely clustered together compared to other established cell lines, suggesting that CTCs carry unique gene expression signatures implicated in these pathways that can survive multiple passages, as supported by Huaman et al.[267], and can be used to characterize differences between cell lines isolated from the primary tumor, CTCs, and metastatic sites. Further work is needed to develop a quantitative method to predict NK sensitivity based on the relative expression of NK inhibitor and NK activators. The rapid assessment of CTCs using ICC allows for a robust and rapid assessment, without relying on time consuming expansion efforts, for personalized patient care.

In freshly isolated CTCs, from 8 NSCLC patients significant inter- and intra-patient heterogeneity was observed. This suggests not all patients may benefit from a NK-based therapy. Patients with progressive disease had significantly lower levels of CADM1 compared to those with stable disease. Treatment and sequential visits showed changes in CTC gene expression. Therefore, it is critical to perform a rational design of combination therapy to be synergistic and

improve patient care. The chemotherapy agent, gemcitabine, has been shown to increase MICA/B expression and therefore may improve NK efficacy. Conversely, the kinase inhibitor, gefitinib, has been shown to decrease MICA/B, and therefore may reduce NK efficacy[271]. Further, kinase inhibitors, erlotinib and afatinib, have been shown to increase HLA-A/B/C expression, however this effect may be influenced by the presence of EGFR mutations[272].

Due to their presence in the blood, CTCs present an opportunity to develop a method for identifying patients who would receive benefit from NK cell therapy which could be universally applied across diverse solid tumor types. Metastasis is responsible for 90% of cancer related deaths[273], therefore the paradigm shift from CTCs as a prognostic marker, to a therapeutic target has the potential to halt metastasis and improve patient survival.

Chapter 6 Conclusions and Future Directions

6.1 Research Summary

The work presented here has advanced the field of liquid biopsy characterization methods for single cells and EVs, both of which suffer from scarce starting amounts of cancer-derived cargo and require creative, tailored and innovative isolation and molecular characterization strategies. These characterization approaches were applied to two of the most deadly cancers, lung and pancreatic cancer, which are responsible for over 140,000[5] and 46,000[4] annual deaths, respectively.

6.1.1 Development of Single Cell Gene Expression and Mutation Detection Workflow

CTCs suffer from low abundance and are not compatible with traditional single cell analysis workflows which suffer from high cell loss. Here, a method for characterizing both gene expression and mutations from the same cell was developed. Using EGFR mutations, which dictate sensitivity to TKIs, as a model, I present the feasibility of this workflow to not only detect the presence of a mutation, but also variable expression levels. I showed cell-to-cell variability, even within a cell line, and relative expression of wildtype and mutant alleles, which would be lost in bulk analysis. This workflow was applied to the detection of EGFR mutations and gene expression profiles carried in NSCLC CTCs. In a cohort of six patients, 58 CTCs were profiled for EGFR mutations, which were detected in 5/6 (83%) patients and in 19/58 (33%) CTCs.

I observed intra-patient heterogeneity in gene expression profiles, and notably the presence of EGFR mutations. In this cohort of metastatic NSCLC patients, the patients were known to carry

EGFR mutations in the primary tumor, however in CTCs, I observed both CTCs that carried tumor-matched mutations and CTCs that did not. In the subset of CTCs which carried point mutations, L858R or T790M, I observed higher wildtype than mutant allele expression (known as reverse MASI). This is different than what is commonly observed in established cancer cells lines, however reverse MASI offers a potential mechanism for TKI-mediated cell death escape despite carrying a targetable mutation.

6.1.2 Mutations Carried in EV Cargo as a Prognostic Signature

EVs offer an abundant and stable circulating biomarker. In this work, I showed that EVs carry mutations from their cells of origin in both the protein and RNA. I screened EVs isolated from a cohort of 10 metastatic NSCLC patients, for a totally of 34 samples, for the presence of EGFR mutations at the RNA (EV-RNA) and protein (EV-protein) level. Mutations in EV-RNA were detected in 27/45 tests, since some patients were tested for multiple EGFR mutations. EGFR mutation detection rates in EV-RNA varied across mutations with 18/23 (78%) for exon 19 del, 6/10 for L858R, and 3/10 for T790M. EV-protein was screened only for sensitizing mutations, but also showed large variability in detection rates across the two mutations, with 6/15 (40%) for exon 19 del, and 6/6 for L858R.

Seven of these patients were longitudinally monitored, and each sample was screened for EGFR mutations in EV-RNA and EV-protein. The abundance of EGFR mutations in EV-RNA and EV-protein was correlated to patient status based on available clinical data. In EV-RNA, my results showed that the level of EGFR mutation in EV-RNA didn't correlate with patient status, rather the change in EV-RNA compared to the previous visit correlated with patient status. Patients with progression showed increased EGFR mutation burden in EV-RNA at each subsequent time point, while patients with stable disease showed decreasing levels. In one patient, L2, the patient

was considered clinically stable at the second time point, however I observed increased EGFR mutation in EV-RNA compared to the first time point; the patient went on to progress at the third time point. This suggested the potential to use EGFR mutation burden in EV-RNA as a more sensitive approach for patient monitoring than current clinical approaches, such as imaging.

EGFR mutations carried in EV-protein didn't reveal a correlation with clinical status in these patients, either based on total burden, or relative burden compared to the previous visit. Interestingly, the co-detection of EGFR mutations in both EV-RNA and EV-protein was only found in 8/22 (36%) samples.

This work presents the first application of using EGFR mutation burden in EVs for patient monitoring and also the first study to characterize mutations in both EV-RNA and EV-protein.

6.1.3 Prognostic Signatures of Patient Survival in PDAC CTCs

In this work from Chapter 4 I profiled CTCs from ten treatment naïve, metastatic PDAC patients from baseline/before treatment and after treatment time points. I evaluated the CTCs for their abundance, EMT phenotype, mutant KRAS presence and expression levels, and transcriptome. I observed that CTC levels were decreased in on-treatment samples compared to pre-treatment samples, and saw a shift towards a higher proportion of EMT-CTCs. I further stratified patients based on alive or deceased. Patients who are now deceased tended to have higher total CTCs and EMT-CTCs at pre-treatment time points than those who are still alive. Mutant KRAS expression tended to be elevated at both time points in CTCs from patients who are now deceased compared to patients who are still alive.

The CTCs' transcriptomes were profiled using microarrays. The CTCs from patients who are deceased versus alive showed significant differentially expressed genes and pathways, both at pre-treatment and on-treatment time points. Notably, 13 of the top 20 differentially regulated

pathways overlapped between both time points based on patient survival, including EGF/EGFR, VEGFA-VEGFR2, MAPK, focal adhesion-PI3K-AKT-mTOR and RAS signaling.

I also showed preliminary data from a subset of patients receiving combination gemcitabine and paclitaxel that the expression of CDA in CTCs may provide a method for predicting treatment response to this combination therapy. The fold change in expression of CDA based on patient survival was nearly 20 at both pre-treatment and on-treatment time points.

Taken together, this work highlights the diverse utility of CTCs which can be profiled beyond enumeration, but also phenotype and to predict treatment sensitivity based on protein and/or transcriptome signatures.

6.1.4 Correlation between NK Sensitivity and Patient Outcomes

NK cells are part of the innate immune system and don't rely on antigen presentation through MHCI, which is known to be downregulated in cancers. NK cells have been proposed to be involved in immune surveillance in the circulation to eliminate CTCs and halt metastasis.

In Chapter 5 I evaluated the expression levels of NK inhibitors (HLA-A/B/C and E-Cad) and NK activators (MICA/B and CADM1) in cancer cell lines, CTC-derived cell lines, and freshly isolated CTCs. I showed that cell lines carry heterogeneous expression levels of these four markers, which impacted their relative sensitivity to NK cells based on NK:tumor co-culture experiments. I further showed inter-patient and intra-patient heterogeneity of freshly isolated CTCs. The expression level of each marker was correlated to patient outcome. Overall, CTCs from patients with progressive disease showed increased levels of HLA-A/B/C and decreased levels of CADM1 compared to patients with stable disease, however significant diversity was observed. This work presented evidence that CTCs may be effectively targeted by NK cells and

suggests a subset of patient with CTCs that display NK sensitive profiles yet the patients aren't clinically stable may benefit from NK cell therapy.

6.2 Limitations and Future Directions

While this work has developed novel characterization approaches for patient monitoring and identifying prognostic signatures, this work suffers from small cohorts of patients and should be followed up with larger studies. Furthermore, all of this work was done in the context of metastatic disease, which has dismal 5-year survival rates. This type of analysis should be expanded to earlier stages of disease. Currently, there is a push in the medical field to stratify patients based on disease severity, such as risk of recurrence. The goal is to identify patients who need the most aggressive treatments to give them the best chance of progression-free survival and overall survival, while not over-treating patients who don't need such aggressive treatment measures. This would prevent the devastating physical and financial costs to those patients of undergoing unnecessary treatment and improve their quality of life. Applying these analysis approaches to earlier disease stage patients may help provide a method for patient stratification and allow for a more tailored treatment approach, thereby improving patient care for both cohorts.

6.2.1 Multiplexed Mutation Detection using ddPCR

Mutation detection was applied in three of these chapters (Chapters 2-4) and in each case revealed the importance of not only detecting the presence of a mutation, but also measuring the relative expression of that mutation. However, in each case, the results are limited to screening for a single mutation at a time. To screen for the presence of multiple mutations currently requires splitting the sample, which isn't always feasible, and can result in false negatives due to sampling errors or the signal being below the limit of detection. The development of multiplexed assays that

consistently detect more than one mutation in a single run is critical to empower more complex and comprehensive profiling of the sample.

In Chapter 2, a single cell could only be screened for one mutation, therefore it is possible that a CTC carried a different mutation either instead of or in addition to the mutation it was tested for. If multiplexed mutation profiling was made possible, the presence of sensitizing mutations, L858R and/or exon 19 deletion, and the resistance mutation, T790M, could be simultaneously screened for. By screening for both sensitizing and resistance mutations from the same cell, this would reveal if the cells carry both mutations, and is therefore resistant to TKIs, or if clones exist that only carry a sensitizing mutation, and would be sensitive to TKIs. Ideally, an assay would be able to screen for all three mentioned EGFR mutations, and would screen for mutations beyond what was detected based on the primary tumor biopsy profiling; it is possible that different regions in the tumor carry different mutations that were missed.

In Chapter 3, similar to Chapter 2, the EV samples were tested exclusively for primary tumor-matched mutations and would benefit from multiplexed profiling. Currently, the sample volumes were already sufficient for screening two mutations, however multiplexed EGFR mutation would increase the capacity to screen for additional mutations, such as KRAS, which is highly prevalent in lung cancers (Appendix D Table D.2).

Lastly, in Chapter 4, bulk pancreatic CTCs were profiled for their KRAS mutation burden. In this case, the assay screens for seven of the most common KRAS mutations in G12/13, however it isn't able to distinguish between these mutatant variants. Future work is critical to distinguish between these mutations, at least KRAS G12C. Targeted therapies against KRAS G12C are currently in development. The work presented in Chapter 4 would provide a minimally invasive

approach to both check for the presence of a KRAS G12C mutation, but also to evaluate its expression to determine if a patient might benefit from this therapeutic approach.

6.2.2 Mutation Burden in EV-cargo: Beyond Correlation to Prognostic Signature

The work I presented in Chapter 3 showed that changes in EGFR mutation burden in EV-RNA correlated with clinical status in a cohort of seven patients. However, future work, with larger studies over longer time scales are needed to determine if this correlation can be extended as a prognostic signature. Large cohorts, following patients from the onset of diagnosis and throughout treatment are needed to determine if changes in EGFR mutation burden in EV-RNA can detect the onset of therapeutic resistance to TKIs or detect progression sooner than is observed using current approaches. If so, EGFR mutations in EV-RNA may provide a more sensitive approach for patient monitoring.

My results qualitatively showed that the change in EGFR mutation burden in EV-RNA correlated with the severity of progression. Future work is needed to establish quantitative methods for this – such as comparisons such between EV-RNA and tumor volume.

EVs may also allow for improved patient care. First, due to their accessibility from a routine blood draw, they may be able to be collected more frequently and require less time from the patient than using traditional imaging modalities. Further, it is possible that the screening of EVs may be more cost-effective than current imaging approaches.

Lastly, this work should be extended to other mutations and other cancer types to determine if the correlation between mutant burden in EV-RNA extends beyond EGFR. Other common mutations in NSCLC include KRAS mutations, and ROS1 and ALK fusions. Other cancer types that could particularly benefit from this monitoring approach would be cancers in the abdomen, which are typically difficult to detect and monitor due to being shielded by other

organs in the body. This could include pancreatic, colorectal, ovarian and bladder cancers and many more.

6.2.3 CTCs to Predict Treatment Response

CTCs offer the potential for personalized treatment decisions based on a patient's own tumor characteristics. There has been a large effort to culture and expand CTCs for drug testing to identify which treatments a patient is most sensitive to, however, CTC cultures suffer from low success rates and are time consuming and labor intensive. Further, CTC cultures have the potential to enrich for the genotypes and/or phenotypes that are best adapted to survive ex vivo, and may not reflect the overall CTC population in vivo.

What if treatment response could instead be predicted based on the molecular profiles carried in CTCs? Robust molecular characterization approaches are established and are already accepted in oncology practice. For example, the presence of certain EGFR mutation genotypes guides the use of TKIs, similarly for ALK fusions. Can a similar approach be applied to gene expression profiles and phenotypes from circulating biomarkers?

In Chapter 2, I showed that CTCs with EGFR mutations all exhibited reverse MASI. After the establishment of a multiplexed EGFR mutation ddPCR assay is developed, one possible clinically relevant approach would be to correlate these findings with a patient status. Suggested analyses would be to compare (1) the fraction of CTCs containing a mutation, (2) the fraction of CTCs showing MASI versus reverse MASI, and (3) the extent of MASI/reverse MASI (i.e. ratio of mutant to wildtype transcripts).

In Chapter 4 I showed preliminary evidence that CTCs with increased CDA expression correlated to increased patient mortality when treated with combination gemcitabine and paclitaxel, however this effect was lost when including patients receiving other treatment

regimens. CDA is a known to inactivate gemcitabine, therefore rendering the treatment ineffective. PDAC is one of the most lethal cancers and suffers from limited treatment options, with relatively few guidelines when deciding the first treatment regime. Future studies are needed to determine if this finding is consistent in larger studies. If the findings remain consistent, this could provide new guidelines for treating PDAC patients.

6.2.4 NK Sensitivity: from Correlation to Clinical Impact

6.2.4.1 Extending to Early Stage and Other Cancer Types

The work presented in Chapter 5 was conducted in a cohort of metastatic NSCLC patients, however NK cells have been implicated in preventing and halting metastasis and may be generalizable to other cancer types. Applying this characterization method to CTCs isolated from patients who have not already developed metastases would be important in understanding the role of NK cells in disease progression. In my experience, it tends to be more challenging to receive early stage samples due to the large proportion of patients who aren't diagnosed until they already have metastatic disease. Therefore, this will be likely be a lengthy research project, but may better highlight the potential of NK-based therapies. Patients' CTCs should be collected at numerous time points and characterized for the expression of NK activators and inhibitors, and clinically monitored for the development of metastases. The patients should be stratified based on the development of metastases and their CTC profiles compared to determine if it was possible to predict which patients were likely to metastasize based on their NK sensitivity profiles. Furthermore, the patients who developed metastases would need to be further stratified to distinguish if there is a subset of these patients with CTCs that exhibited NK-sensitive profiles and therefore may have received benefit from NK therapy.

One of the biggest limitations of the CTC analysis used in Chapter 5 is that the CTCs were evaluated for either the NK inhibitors or NK activators, and therefore each CTC doesn't contain a complete profile. Rather the sample had to be divided in half, with half of the sample used for each set of NK ligands. Future work is needed to further multiplex these markers together to profile each CTC for both the NK inhibitors and NK activators.

6.2.4.2 Developing Mathematical Predictive Model for NK Sensitivity

The work presented in Chapter 5 showed preliminary data on the connection between protein expression levels of NK ligands and the resulting NK sensitivity based on NK:tumor cell co-cultures. Future work developing a mathematical model to predict NK sensitivity based on the

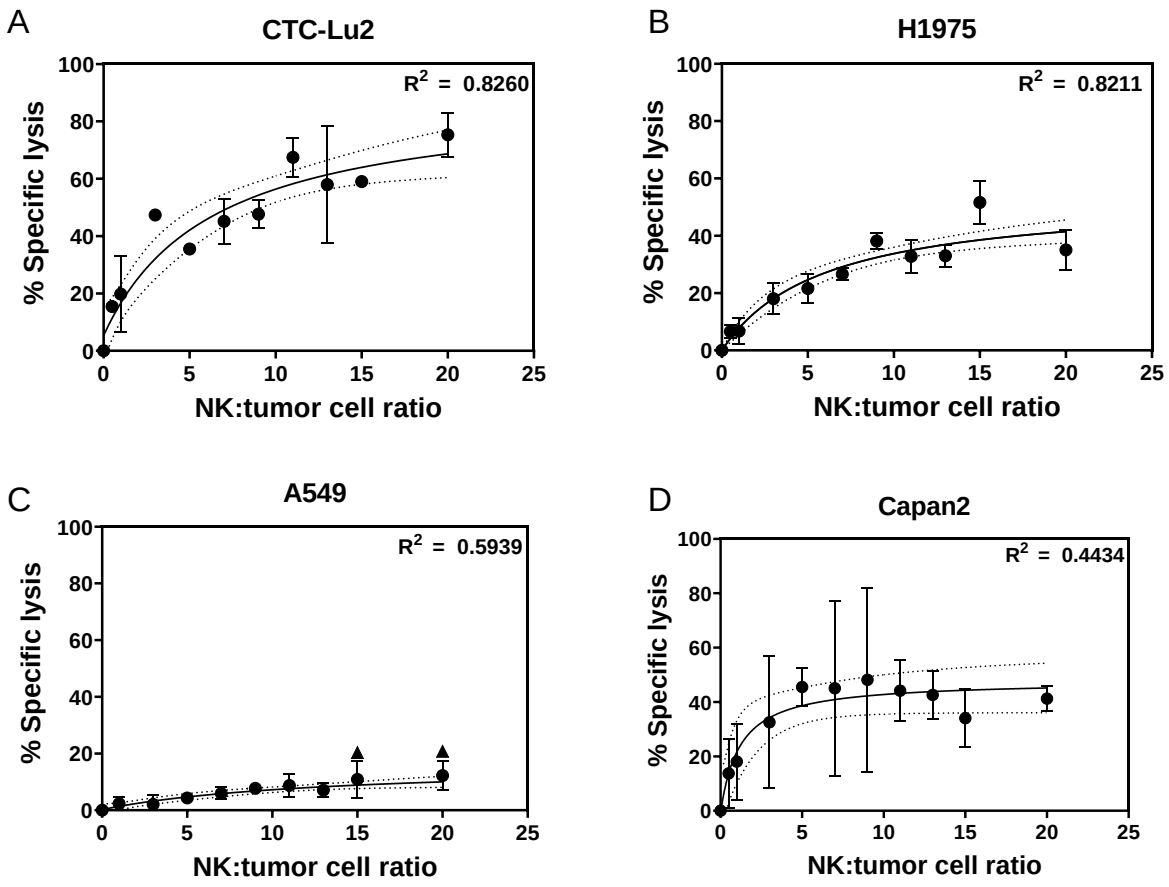


Figure 6.1 Cancer cells exhibit different NK sensitivity.
Preliminary data of NK sensitivity for four cancer cell lines, A) CTC-Lu2, B) H1975, C) A549, and D) Capan2. Cancer cells were co-cultured with NK cells for four hours at indicated ratios.

expression of these markers is currently in progress, as described by Scherbakova et al[274]. NK cells are known to rely on the balance of activating and inhibiting ligands, however not much is known about what causes the shift from inhibition to activation. This model will compare the relative NK sensitivity of CTCs based on their protein expression profiles. The model will be developed by comparing the expression levels of each marker across cell lines and correlating that with the observed NK sensitivity, as evaluated by the % specific lysis in NK:tumor cell co-culture models (Figure 6.1).

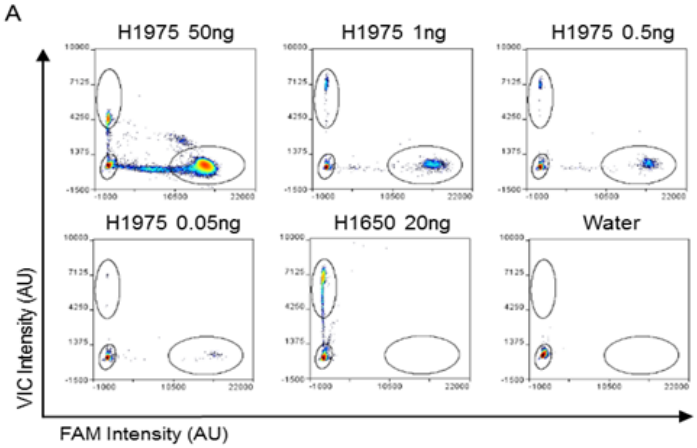
6.3 Conclusions

The work presented here has contributed to the field of liquid biopsy research and its clinical implications. I developed innovative single cell analysis methods (Chapter 2), discovered a novel potential method of treatment resistance through reverse MASI (Chapter 2), identified a novel biomarker for patient prognosis in EGFR mutations carried in EV-RNA (Chapter 3), compared CTC characteristics, such their abundance, phenotype, and transcriptome, based on patient survival (Chapter 4) and provided insights on the intersection of CTCs, NK cells – and their role in immune surveillance and halting metastasis – with clinical status. Taken together, liquid biopsies provide a minimally invasive method to characterize and monitor a patient's disease and can provide discernment in how to best care for and treat patients, personally.

Appendices

Appendix A: Development of Digital PCR (dPCR) for EGFR Mutant Screening from Single Cells

A

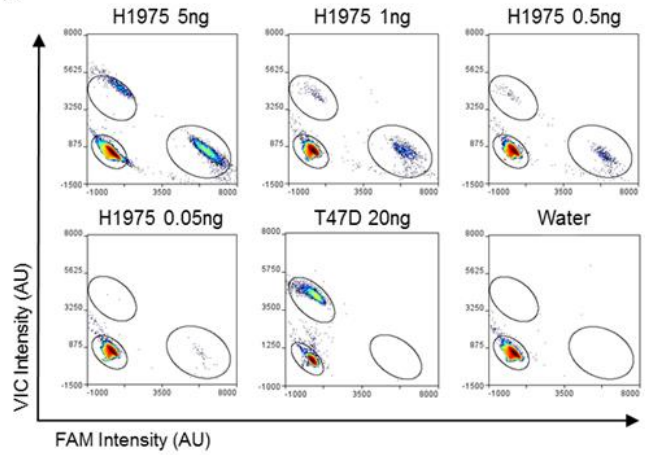


B

Sample	Loading (ng)	Droplet Counts	
		Wildtype	*Mutant*
H1975	50	14247	59117
H1975	20	7716	32487
H1975	5	1316	7060
H1975	1	312	1356
H1975	0.5	159	645
H1975	0.05	10	66
H1650	20	5605	0
T47D	20	4847	0
water	--	0	0

Figure A.1 EGFR L858R titration.
A) Representative dPCR plots of EGFR L858R mutation assay cDNA loading titration. cDNA generated from control cell lines was used to determine the sensitivity of the RainDrop Plus™ dPCR system using the TaqMan™ SNP assays. Decreasing amounts of cDNA were loaded onto to the dPCR platform ranging from 0.05-50ng of cDNA. B) Representative EGFR wildtype and mutant droplet counts at different cDNA loading.

A

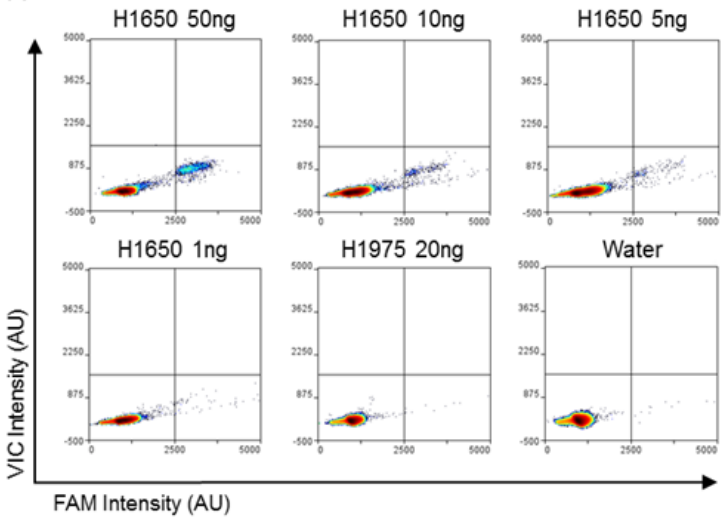


B

Sample	Loading (ng)	Droplet Counts	
		Wildtype	*Mutant*
H1975	20	7037	29619
H1975	5	738	3549
H1975	1	127	864
H1975	0.5	57	393
H1975	0.05	2	46
H1650	20	8860	0
A549	20	8263	0
T47D	20	3193	0
water	--	0	0

Figure A.2 T790M titration.
A) Representative dPCR plots of EGFR T790M mutation assay cDNA loading titration. cDNA generated from control cell lines was used to determine the sensitivity of the RainDrop Plus™ dPCR system using the TaqMan™ SNP assays. Decreasing amounts of cDNA were loaded onto to the dPCR platform ranging from 0.05-20ng of cDNA. B) Representative EGFR wildtype and mutant droplet counts at different cDNA loading.

A

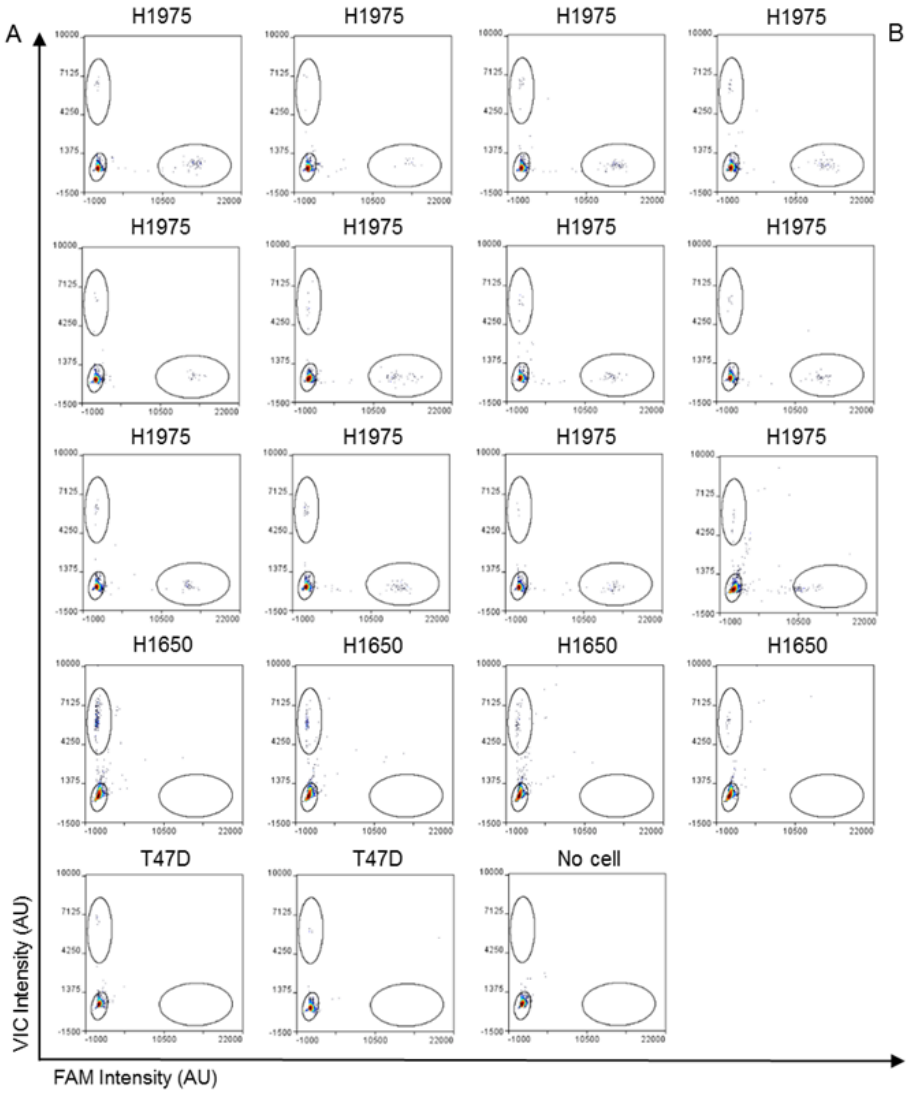


B

Sample	Loading (ng)	Droplet Counts *Mutant*
H1650	50	1598
H1650	10	415
H1650	5	232
H1650	1	61
H1975	20	12
A549	20	11
water	--	15

Figure A.3 Exon19 titration.

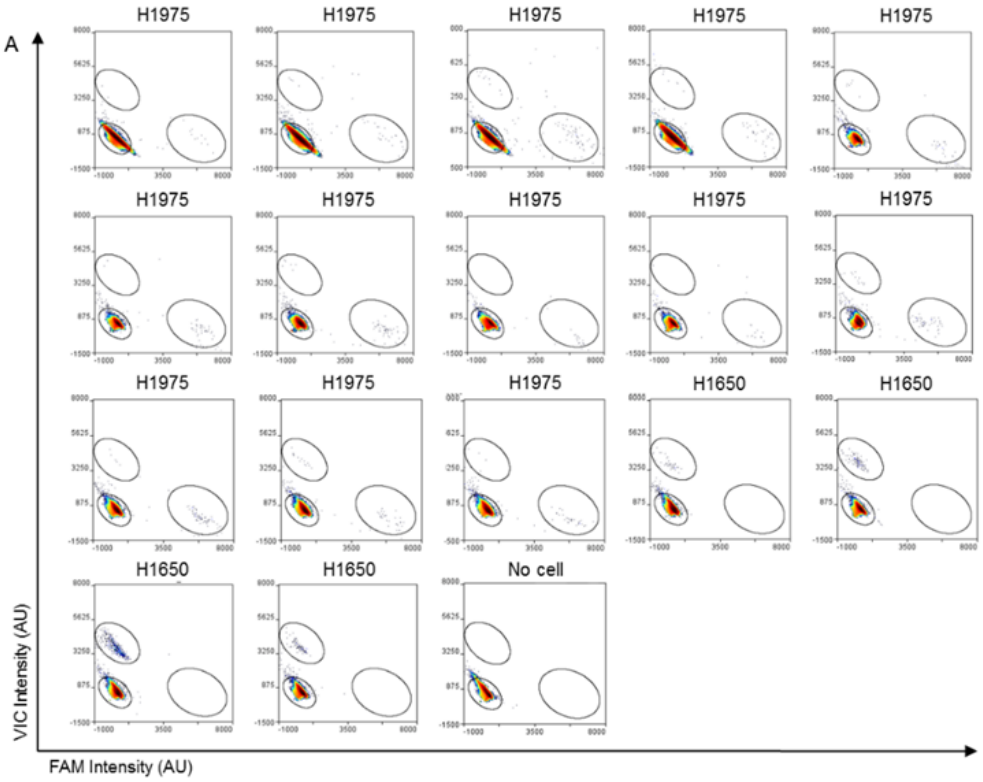
A) Representative dPCR plots of EGFR exon 19 deletion mutation assay cDNA loading titration. cDNA generated from control cell lines was used to determine the sensitivity of the RainDrop Plus™ dPCR system using the TaqMan™ deletion assays. Decreasing amounts of cDNA were loaded onto to the dPCR platform ranging from 1-50ng of cDNA. H1650 50 ng image duplicated from Figure 2A. B) Representative EGFR wildtype and mutant droplet counts at different cDNA loading.



B

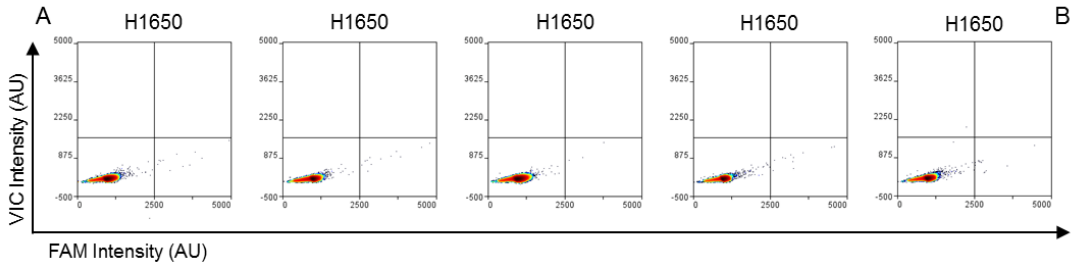
Sample	Droplet Counts	
	Wildtype	*Mutant*
H1975	11	66
H1975	3	14
H1975	14	60
H1975	12	49
H1975	5	20
H1975	11	53
H1975	12	39
H1975	8	36
H1975	14	42
H1975	21	52
H1975	4	37
H1975	11	61
H1650	190	0
H1650	80	0
H1650	46	0
H1650	17	0
T47D	8	0
T47D	5	0
No cell	0	0

Figure A.4 EGFR L858R representative single cells.
A) Single cells processed on the Fluidigm C1 were screened for the EGFR L858R point mutation. H1975 single cells were reproducibly identified to contain the L858R point mutation, while the control cell lines, T47D and H1650, only contained wildtype EGFR. Finally, no cell controls processed through the entire C1 and dPCR workflow were not found to contain false positive signals of wildtype or mutated EGFR. (n=12 H1975 single cells, n=4 H1650 single cells, n=2 T47D single cells, n=1 no cell control). B) Wildtype and mutant droplet counts for single cells.



Sample	Droplet Counts	
	Wildtype	*Mutant*
H1975	4	12
H1975	4	24
H1975	11	55
H1975	6	32
H1975	6	23
H1975	3	40
H1975	6	48
H1975	0	16
H1975	1	12
H1975	12	52
H1975	5	67
H1975	14	28
H1975	4	30
H1650	40	0
H1650	113	0
H1650	520	0
H1650	98	0
No cell	0	0

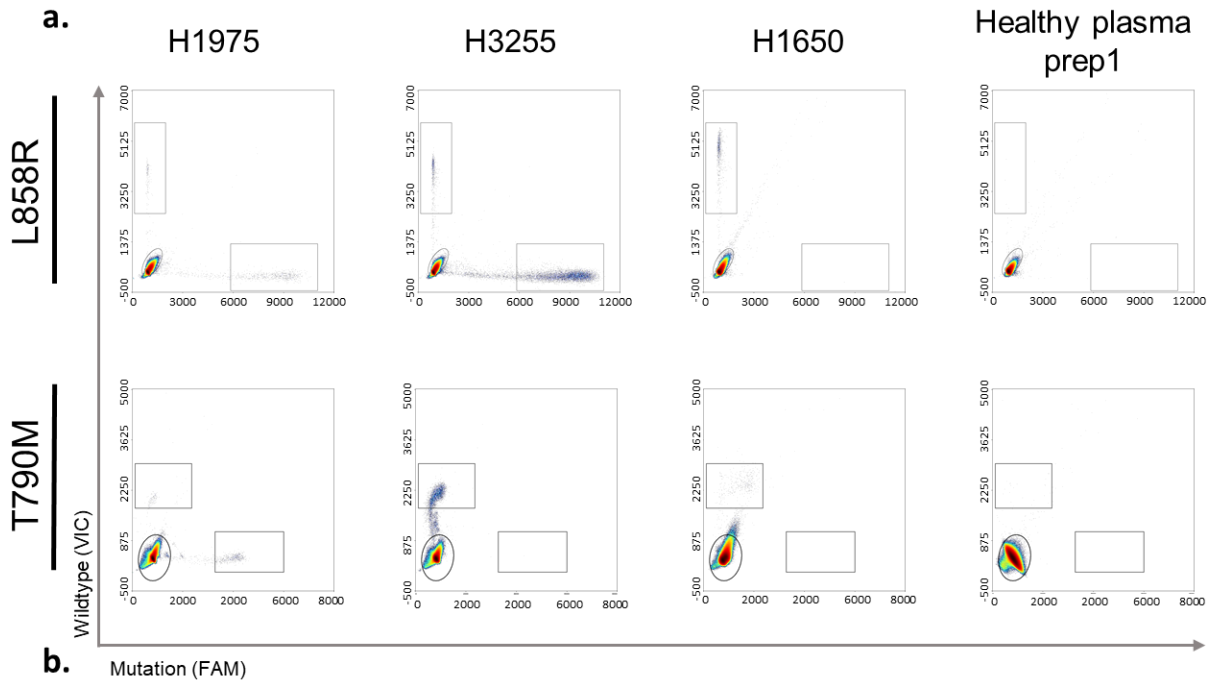
Figure A.5. EGFR T790M representative single cells.
A) Single cells processed on the Fluidigm C1 were screened for the EGFR T790M point mutation. H1975 single cells were reproducibly identified to contain the T790M point mutation, while the control cell line, H1650, only contained wildtype EGFR. Finally, no cell controls processed through the entire C1 and dPCR workflow were not found to contain false positive signals of wildtype or mutated EGFR. (n= 8 H1975 single cells, n=4 H1650 single cells, n=1 no cell control) B) Wildtype and mutant droplet counts for single cells.



Sample	Droplet Counts
	Mutant
H1650	11
H1650	12
H1650	3
H1650	11
H1650	14

Figure A.6 EGFR Exon 19 deletion representative single cells.
A) Single H1650 cells processed on the Fluidigm C1 were screened for the EGFR exon 19 deletion. H1650 single cells were reproducibly identified to contain the exon 19 deletion. B) Mutant droplet counts from H1650 single cells (n=5).

Appendix B: Development of dPCR for EGFR Mutant Screening from EVs



	H1975	H3255	H1650	Healthy plasma prep 1
L858R	+	+	-	-
T790M	+	-	-	-

Figure B.1 EGFR point mutations, L858R and T790M, detected in lung cancer cell line derived EV-RNA.

A. Representative droplet digital PCR plots for EGFR L858R (top) and T790M (bottom) from EV-RNA generated using control lung cancer cell line derived extracellular vesicles and healthy control EVs from plasma prep 1. The mutation signal is detected in the FAM channel (x-axis), wildtype signal is detected in the VIC channel (y-axis) and empty droplets are shown in the bottom left of each plot. B. Table of EGFR L858R and T790M point mutation status in lung cancer cell lines and healthy donors.

Appendix C: Comparison of EVs using Different Plasma Generation Methods

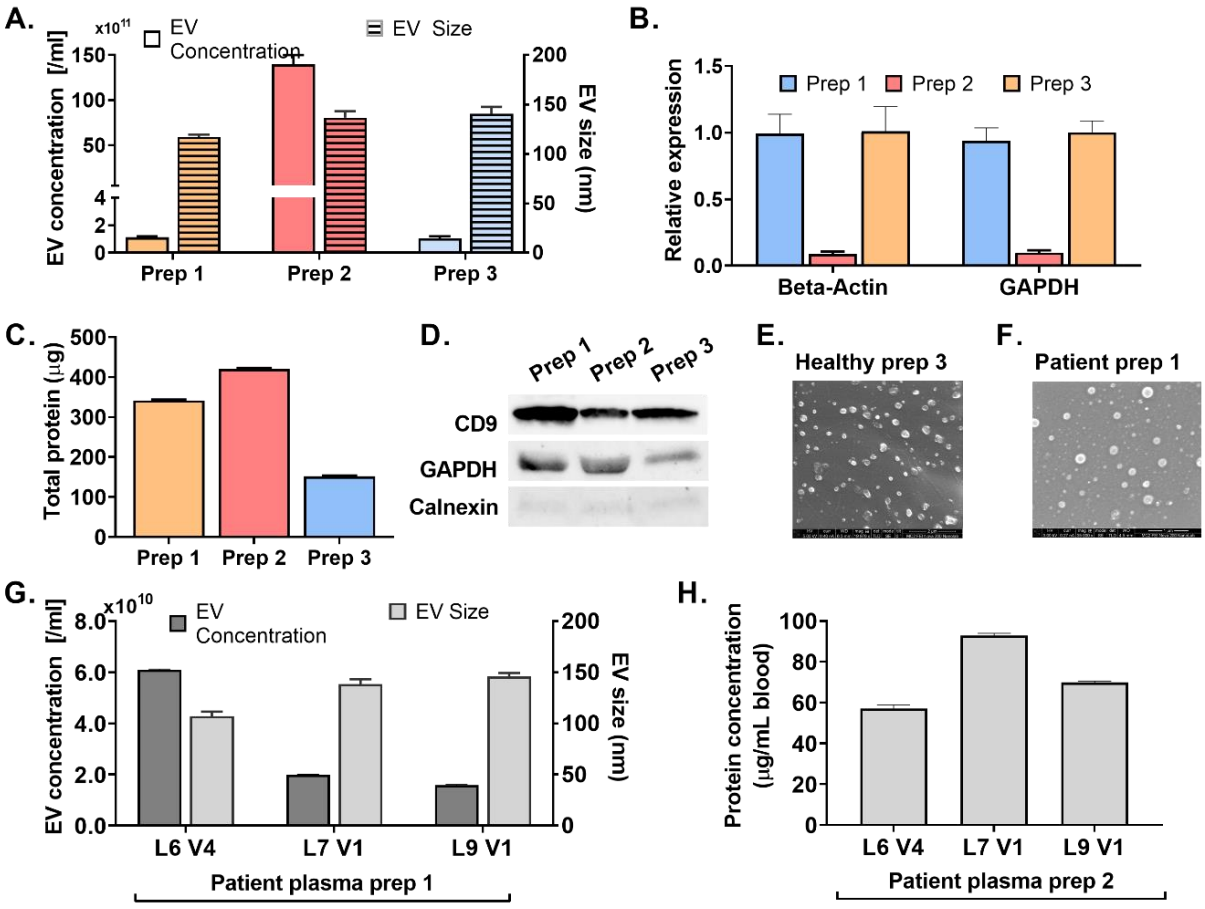


Figure C.1 Comparison of plasma preparations for EV analysis.
A) EV concentration and size of plasma preps from healthy donor measured by NTA. B) Relative expression of housekeeping genes ACTB and GAPDH as measured by RT-qPCR. C) Total protein from healthy donor measured by BCA. D) Western blot of CD9, GAPDH, and calnexin. E-F) SEM images of EVs isolated form E) healthy donor plasma prep 3 sample and F) NSCLC patient plasma prep 1 sample. G) EV concentration and size of plasma preps from NSCLC patients measured by NTA. H) Total protein from NSCLC patient patients measured by BCA.

To ensure that the EV cargo was not damaged by RBC depletion prior plasma extraction, standard EV characterization assays for the three plasma isolation methods were performed and compared. 15mL of blood from a healthy donor was divided and processed using 5mL of blood using each of the three plasma generation methods: Ficoll RBC removal followed by centrifugation (prep 1), dextran RBC removal followed by centrifugation (prep 2) or centrifugation alone (prep 3). For all reported values, error bars are indicative of the average and standard deviation of 3 technical replicates.

EV size and concentration of the three plasma prep methods was measured using NTA and the size of the EVs did not vary with isolation method; yielded vesicles were within the standard range for EVs (50-160 nm), Figure C.1A. The concentration has been normalized to starting blood volume, reported values are per mL of original blood volume. The measured concentration of the EVs isolated using plasma prep 2 was dramatically higher than the concentration from plasma prep methods 1 or 3. The higher concentration could be a result of residual dextran macromolecules being measured[275]. Another possibility is that dextran allows for more efficient recovery of EVs during ultracentrifugation. Further experiments would be needed to verify the cause of this increase.

The relative abundance of house-keeping genes, ACTB and GAPDH, was measured using RT-qPCR. Using plasma prep 3 EVs as the baseline for comparison, plasma prep 1 EVs had similar levels of the two genes, whereas plasma prep 2 EVs had less than 10% comparatively, Figure C.1B. One explanation for the result is that the differences in EV-RNA yield is from inefficiencies in EV lysis in the presence of remaining dextran molecules. Another is that the EV-RNA was damaged during the dextran processing, leading to lower abundance. Further testing would be needed to verify the reason. Because the detected levels of

housekeeping genes was similar for plasma preps 1 and 3, plasma prep 1 samples were used for EV-RNA characterization.

EV-protein characterization was performed using microBCA and western blot. There was slightly higher protein yield from plasma prep 2 than plasma prep 1, and both were higher than plasma prep 3, Figure C.1C. Western blot analysis of the three plasma prep methods showed similar expression of the following proteins: exosomal marker CD9, housekeeping gene GAPDH, and contamination marker calnexin, Figure C.1D.

Scanning electron microscopy images of EVs isolated from a healthy control prep 3 and a patient prep 1 to demonstrate successful EV isolation, Figure C.1E,F.

For additional verification, NSCLC patient plasma preps were processed either (1) directly into PBS for NTA quantification or (2) into RIPA lysis buffer for protein analysis. Three plasma prep 1 samples were measured using NTA and were found to have similar size, 130 ± 17nm, and concentration, $3x10^{10} \pm 2x10^{10}$, Figure C.1E. From the same patients and time points, the plasma prep 2 samples were measured for EV-protein quantity using microBCA with an average of, 73 ± 18 µg/mL of blood in 150 µL of protein lysate from 5mL of blood, Figure C.1F.

Appendix D: Development of digital PCR (dPCR) for Oncogenic and Tumor Suppressor Mutations Broadly Found Across Solid Tumor Types

D.1 Development of dPCR Methods for SNP Detection

The RainDrop dPCR platform was additionally optimized for additional SNPs commonly found across other solid tumor types including breast, lung, pancreatic and prostate cancers from genomic DNA (gDNA), complementary DNA (cDNA), single cell cDNA (sc-cDNA) and EVs (Table D.1). dPCR is most commonly used for gDNA, therefore representative images from bulk cDNA, single cell cDNA, and EVs are shown here to act a resource for expected results (Figures D.1 –D.3). cDNA was generated using SuperScript IV VILO (Invitrogen) – which relies on random primers for cDNA synthesis. Therefore the resulting cDNA is highly fragmented with good coverage across the transcript. However, due to the nature of random primers, the resulting cDNA fragments are highly heterogeneous. The use of cDNA, compared to gDNA, results in the "comet tail" of droplets between the positive populations and the empty, negative populations. This is likely due to the variability of the cDNA fragment lengths and location of the specific region of interest along the cDNA fragment, yielding variable PCR amplification efficiency of different cDNA fragments within a sample.

Table D.1 Summary of mutation detection assays validated for different sample types using the RainDrop dPCR system (RainDance Technologies).
Check marks indicates the sample type is compatible with this assay.

Mutation	Bulk gDNA	Bulk cDNA	sc-cDNA	EVs
EGFR L858R	✓	✓	✓	✓
EGFR T790M	✓	✓	✓	✓
EGFR exon 19 deletion	Not tested	✓	Not tested	✓
KRAS G12/13	Not tested	✓	Not tested	✓
KRAS G12C	Not specific between KRAS G12C, G12D, and G12V	Not specific between KRAS G12C, G12D, and G12V	Possible but challenging due to low expression	Not tested
KRAS G12D	✓	Not specific between KRAS G12C, G12D, and G12V	Possible but challenging due to low expression	Not tested
KRAS G12V	Not specific between KRAS G12C, G12D, and G12V	Not specific between KRAS G12C, G12D, and G12V	Possible but challenging due to low expression	Not tested
PIK3CA E545K	✓	✓	Possible but challenging due to low expression	Not tested
PIK3CA H1047R	✓	✓	Possible but challenging due to low expression	Not tested
TP53 R273	✓	✓	Possible but challenging due to low expression	Not tested

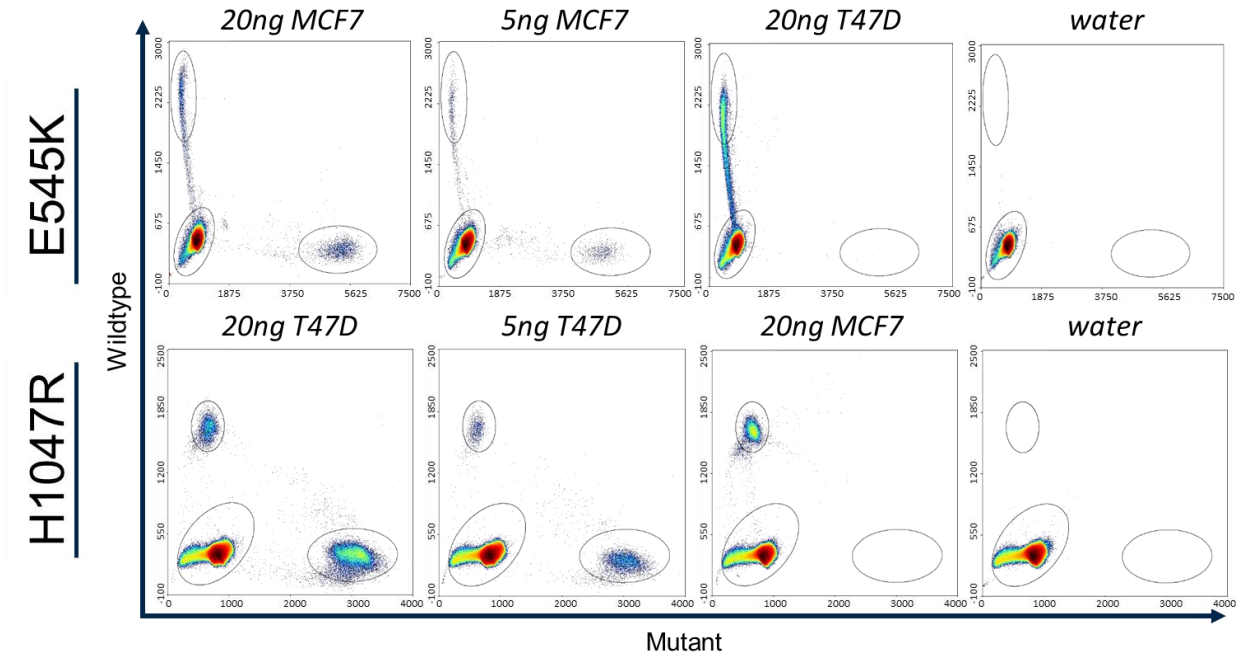


Figure D.1 PIK3CA dPCR plots from breast cancer cell lines.
Bulk cDNA generated from breast cancer cell lines, MCF7 (E545K) and T47D (H1047R), were tested for the presence of two PIK3CA mutations, E545K and H1047R.

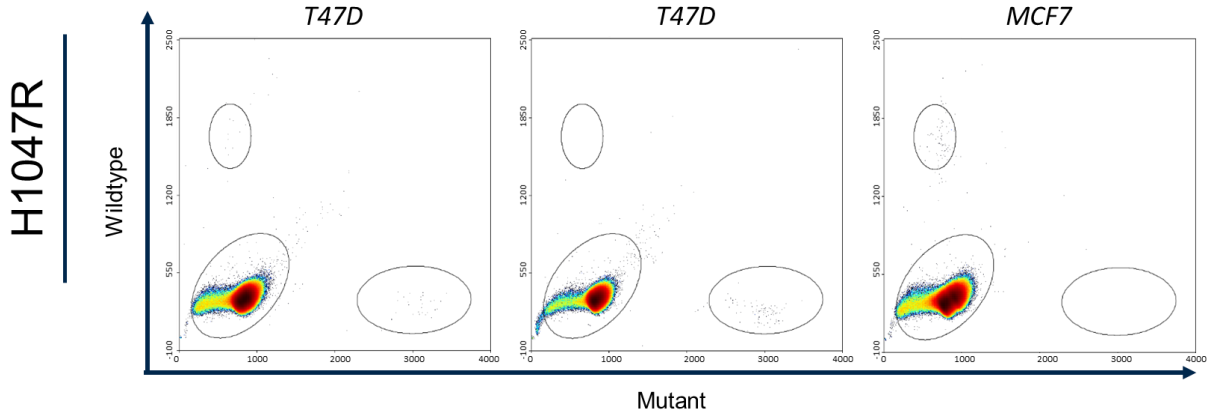


Figure D.2 PIK3CA H1047R dPCR plot from single T47D cells.
Single T47D and MCF7 cells were isolated using the Fluidigm C1 technology and analyzed on dPCR for the presence of the PIK3CA H1047R mutation.

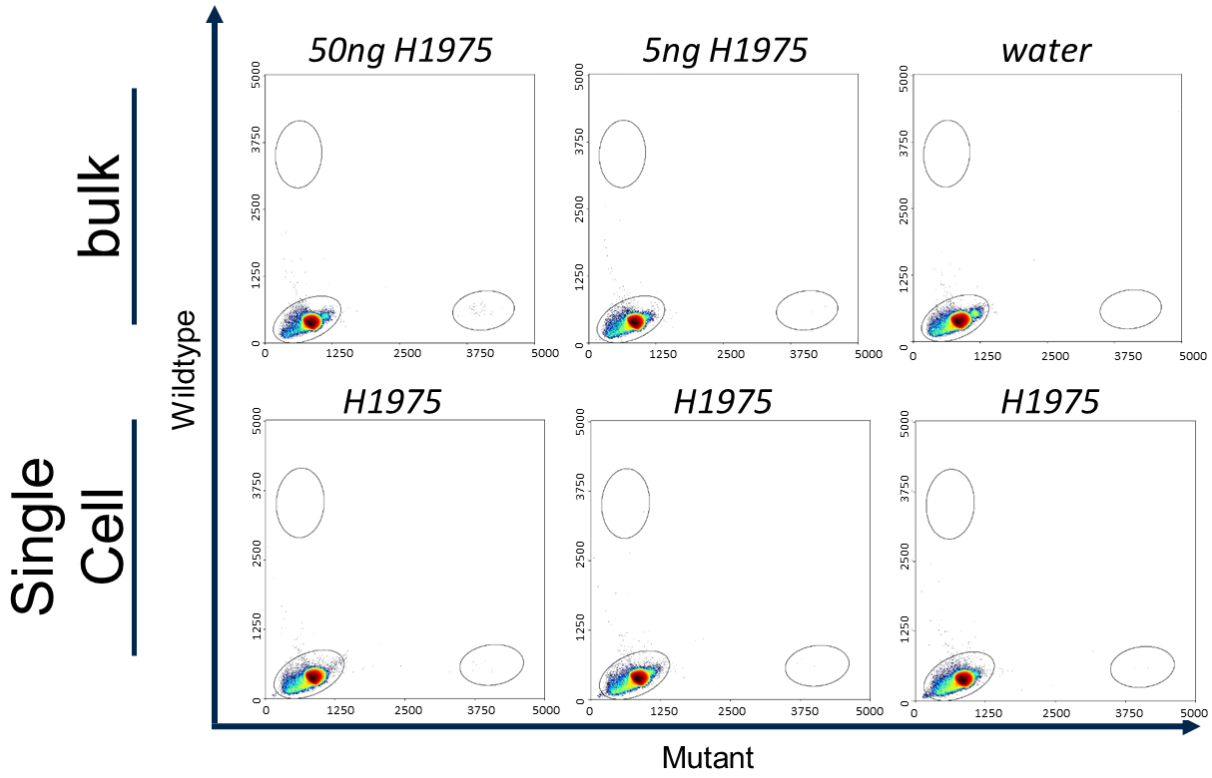


Figure D.3 TP53 R273H dPCR plots from lung cancer cell line H1975.
Bulk cDNA and single cells isolated using the Fluidigm C1 from lung cancer cell, H9175 tested for the presence of the TP53 mutation, R273H.

D.2 Considerations for Multiplexed Mutation Detection

The projects described in this thesis relied on the expression of mutant transcripts, and not just the presence of a given mutation, therefore cDNA was utilized. As mentioned in above in Appendix D.1, cDNA was generated using random primers and samples were comprised of highly heterogeneous cDNA fragments. cDNA, without introns such as is the case with gDNA, puts mutations in the same gene in closer proximity within the cDNA sequence. Consequently, it is possible for the presence of two mutations to be on either the same cDNA fragment, or split across two cDNA fragments. Therefore special considerations are needed when developing screening assays for multiple mutations along the same gene at the cDNA level. As shown below in Figure D.4, multiplexing of EGFR L858R and EGFR T790M produced more positive populations than would be predicted. The use of two duplex assays would predict four positive droplet populations, corresponding to (1) wiltype of assay 1, (2) mutation of assay 1, (3), wildtype of assay 2, and (4) mutation of assay 2, however six populations were observed. Using the model cell line, H1975, which carries heterozygous L858R and T790M mutations, with the mutations being carried on the same allele, and synthetic DNA constructions, it was shown that the six populations corresponded to the four predicted populations listed above plus two additional populations, (5) wildtype of assays 1 &2 combined and (6) mutation of assays 1 &2 combined.

The combination of the EGFR L858R and T790M in a dPCR reaction yielded six positive populations in the H1975 cDNA due to the heterogeneous cDNA fragments present during the reverse transcription step, and not the loading of two cDNA fragments, one with each portion of the transcript, into the same droplet. This was proven through the use of the six DNA constructs, which showed the use of a single DNA construct containing both regions of interest was required to generate populations (5) and (6).

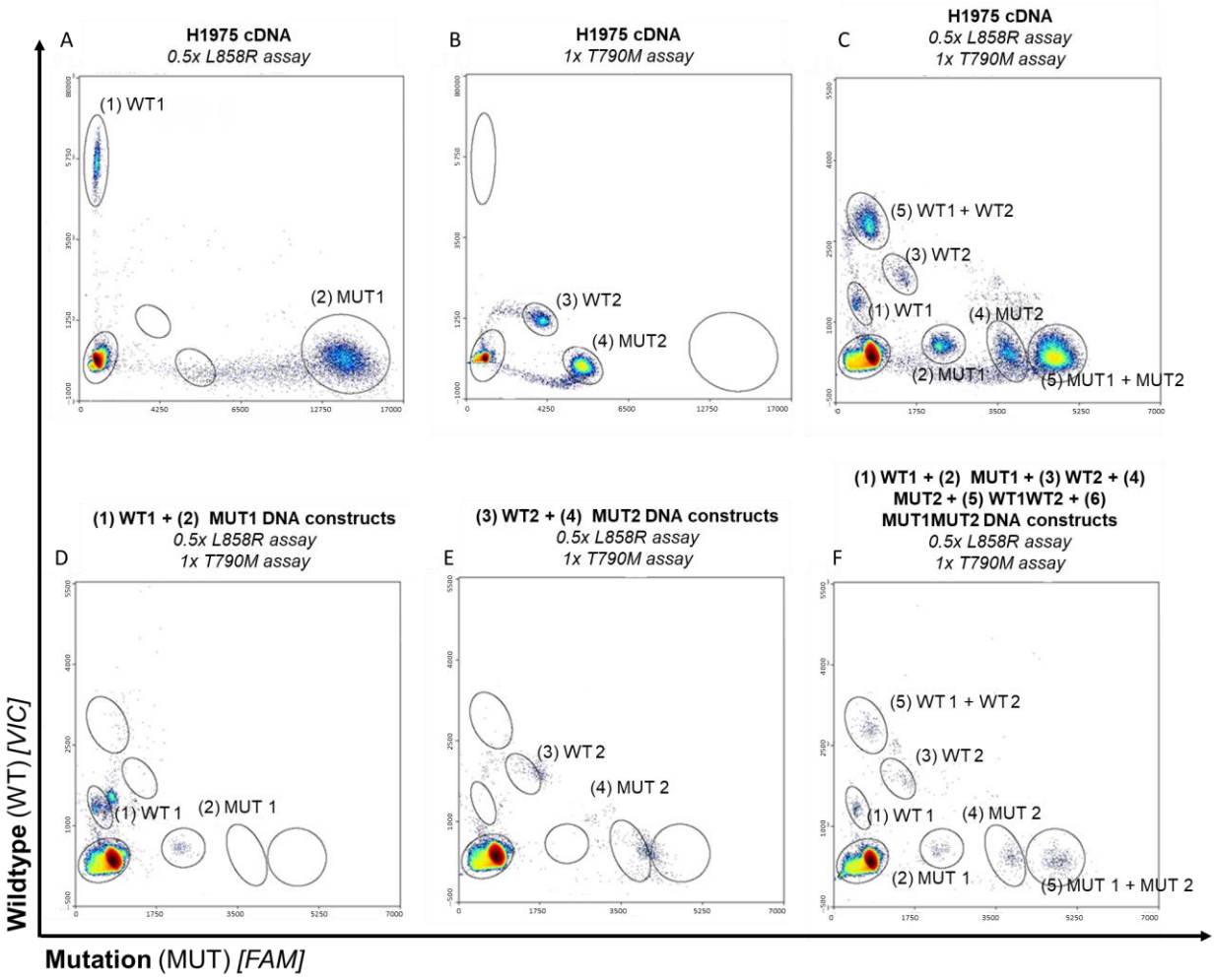


Figure D.4 Multiplex EGFR mutation detection.
Multiplex EGFR mutation detection optimization using A-C) H1975 cDNA and D-F) synthetic DNA constructs.

Mutation frequencies across various cancer types were identified using the Catalogue of Somatic Mutations in Cancer (COSMIC)[276] database. Frequencies were manually calculated based on the number of samples submitted to COSMIC database containing the mutation compared to the total number of samples, for a given cancer type. These numbers presented in Table D.2 do not account for different cancer subtypes within a given cancer.

Table D.2 Common mutations found in solid cancers of the breast, lung, pancreas and prostate.

Cancer	Gene	*Common SNP(s) [if applicable]*	Overall Frequency (%)	Frequency within mutations (%)
Breast	ALK		0.62	
	BRAF		0.79	
	EGFR		1.07	
		L858R	0.24	22.6
		T790M	0.061	5.7
	HER2 (ERBB2)		1.49	
	KRAS		1.54	
		G12D	0.44	29
		G13D	0.4	26.1
		G12C	0.24	15.9
		G12V	0.22	14.5
	PIK3CA		26.76	
		H1047R	8.84	33
		E545K	3.31	12.4
		E542K	1.96	7.33
	TP53		23.28	
		R175H	0.78	3.3
		R273H	0.64	2.8
		R248Q	0.62	2.6
		R248W	0.57	2.4
Lung	ALK		4.15	
		L1196M	0.2	4.8

		G1269A	0.16	3.8
		I1171T	0.12	2.9
		I1171N	0.12	2.9
	BRAF			
		V600E	0.54	25
		G469A	0.13	6
	EGFR		28.86	
		p.L858R	8.75	30.3
		p.T790M	1	3.5
		p.L861Q	0.35	1.2
	HER2 (ERBB2)		1.43	
		S310F	0.02	1.39
		V659E	0.01	0.93
		L755P	0.01	0.93
	KRAS		16.44	
		G12D	3.61	21.9
		G12C	6.54	39.8
		G12V	4.18	25.4
		G13D	1.36	8.27
	PIK3CA		3.68	
		H1047R	0.47	12.8
		E545K	0.59	16.1
	TP53		34.02	
		R273L	1.01	3
		R273H	0.94	2.8
Pancreas	ALK		0.5	
		R1248Q	0.125	25
		G574R	0.0624	12.5
		R292H	0.0624	12.5
	BRAF		1.51	
		V600E	0.42	27.8
		D594G	0.13	8.3
	HER2 (ERBB2)		0.92	
		R217C	0.29	31.25
		V1128I	0.17	18.75
	EGFR			
		V765M	0.04	12.5
		E734G	0.04	12.5
		C818Y	0.04	12.5

		R831C	0.04	12.5
		F172L	0.04	12.5
	KRAS		56.88	
		G12D	28.5	50
		G12V	17.7	31.1
		G12R	7.5	13.2
		G12C	1.6	2.8
		G12S	1	1.8
		G12A	0.9	1.5
	PIK3CA		1.66	
		H1047R	0.25	15.4
		E545K	0.13	7.7
		R88Q	0.13	7.7
	TP53		35.53	
		R175H	1.61	4.5
		R273H	1.34	3.8
		R282W	1.22	3.5
		R273C	0.84	2.4
Prostate	ALK		0.93	
		K1491R	0.21	23.1
		G1548R	0.07	7.7
	BRAF		1.27	
		V600E	0.08	6.45
		V600A	0.45	35.5
		V600M	0.33	25.8
		G469A	0.08	6.45
	EGFR		2.44	
		V834A	0.16	6.5
	HER2 (ERBB2)		0.55	
	KRAS		3.56	
		G12V	1.46	41.1
		G13D	0.86	24.2
		G12D	0.82	23.2
		G12C	0.37	10.5
	PIK3CA		1.99	
		H1047R	0.35	17.6
		E545K	0.11	5.9

	E545A	0.18	8.8
TP53		14.54	
	R248Q	0.74	5.1
	R175H	0.35	2.4
	R282W	0.3	2.1
	R273C	0.26	1.8

References

1. Heron, M. Deaths: Leading causes for 2017. *Natl. Vital Stat. Reports* **68**, (2019).

2. National Cancer Institute. SEER: Cancer Stat Facts: Cancer of Any Site. https://seer.cancer.gov/statfacts/html/all.html.

3. Alizadeh, A. M., Shiri, S. & Farsinejad, S. Metastasis review: from bench to bedside. *Tumor Biol.* **35**, 8483–8523 (2014).

4. National Cancer Institute. SEER Stat Fact Sheets: Cancer Stat Facts: Pancreas Cancer. https://seer.cancer.gov/statfacts/html/pancreas.html (2016).

5. National Cancer Institute. SEER Stat Fact Sheets: Cancer Stat Facts: Lung Cancer. https://seer.cancer.gov/statfacts/html/lungb.html (2016).

6. Cappuzzo, F. *Guide to Targeted Therapies: EGFR mutations in NSCLC*. (Springer International Publishing, 2014). doi:10.1007/978-3-319-03059-3.

7. Buscail, L., Bournet, B. & Cordelier, P. Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 153–168 (2020).

8. Imamura, T. *et al.* Liquid biopsy in patients with pancreatic cancer: Circulating tumor cells and cell-free nucleic acids. *World J. Gastroenterol.* **22**, 5627–5641 (2016).

9. Liu, B. *et al.* Identification of Prognostic Biomarkers by Combined mRNA and miRNA Expression Microarray Analysis in Pancreatic Cancer. *Transl. Oncol.* **11**, 700–714 (2018).

10. Martini, V., Timme-Bronsert, S., Fichtner-Feigl, S., Hoeppner, J. & Kulemann, B. Circulating tumor cells in pancreatic cancer: Current perspectives. *Cancers (Basel).* **11**, 1–14 (2019).

11. Han, S., Underwood, P. & Hughes, S. J. From tumor microenvironment communicants to biomarker discovery: Selectively packaged extracellular vesicular cargoes in pancreatic cancer. *Cytokine Growth Factor Rev.* **51**, 61–68 (2020).

12. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2019. *CA. Cancer J. Clin.* **69**, 363–385 (2019).

13. Kleeff, J. *et al.* Pancreatic cancer. *Nat. Rev. Dis. Prim.* **2**, 1–22 (2016).

14. Pereira, M. A. & Chio, I. I. C. Metastasis in pancreatic ductal adenocarcinoma: Current standing and methodologies. *Genes (Basel).* **11**, (2020).

15. Martin, O. A., Anderson, R. L., Narayan, K. & MacManus, M. P. Does the mobilization of

circulating tumour cells during cancer therapy cause metastasis? *Nat. Rev. Clin. Oncol.* **14**, 32–44 (2017).

16. Radiation Therapy to Treat Cancer. *National Cancer Institute* https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy (2019).

17. Joo, W. D., Visintin, I. & Mor, G. Targeted cancer therapy – Are the days of systemic chemotherapy numbered? *Maturitas* **76**, 308–314 (2013).

18. Chemotherapy to Treat Cancer. *National Cancer Institute* https://www.cancer.gov/about-cancer/treatment/types/chemotherapy.

19. DeVita, V. T. & Chu, E. A history of cancer chemotherapy. *Cancer Res.* **68**, 8643–8653 (2008).

20. Trimble, E. L. *et al.* Neoadjuvant therapy in cancer treatment. *Cancer* **72**, 3515–3524 (1993).

21. American Cancer Society. Monoclonal Antibodies and Their Side Effect. https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/immunotherapy/monoclonal-antibodies.html (2019).

22. Villaruz, L. C. *et al.* Clinicopathologic features and outcomes of patients with lung adenocarcinomas harboring BRAF mutations in the Lung Cancer Mutation Consortium. *Cancer* **121**, 448–456 (2015).

23. Ray, P. *et al.* Differential protein stability of EGFR mutants determines responsiveness to tyrosine kinase inhibitors. **7**,.

24. Wang, W., Song, Z. & Zhang, Y. A Comparison of ddPCR and ARMS for detecting EGFR T790M status in ctDNA from advanced NSCLC patients with acquired EGFR--TKI resistance. *Cancer Med.* **6**, 154–162 (2016).

25. Suda, K., Onozato, R., Yatabe, Y. & Mitsudomi, T. EGFR T790M mutation: a double role in lung cancer cell survival? *J. Thorac. Oncol.* **4**, 1–4 (2009).

26. Wang, Y. & Ellis, P. EGFR mutation positive non-small cell lung cancer: can we identify predictors of benefit from immune checkpoint inhibitors. *Ann. Transl. Med.* **5**, 1–4 (2017).

27. Barrón, F., Zatarain-Barrón, Z. L., Cardona, A. F. & Arrieta, O. Extending the curve: survival of EGFR-mutated lung cancer patients in the 21st century. *J. Thorac. Dis.* **10**, 1265–1268 (2018).

28. Dougan, M., Dranoff, G. & Dougan, S. K. Cancer Immunotherapy: Beyond Checkpoint Blockade. *Annu. Rev. Cancer Biol.* **3**, 55–75 (2019).

29. Hu, W., Wang, G., Huang, D., Sui, M. & Xu, Y. Cancer immunotherapy based on natural killer cells: Current progress and new opportunities. *Front. Immunol.* **10**, 1–16 (2019).

30. Cheng, M., Chen, Y., Xiao, W., Sun, R. & Tian, Z. NK cell-based immunotherapy for malignant diseases. *Cell. Mol. Immunol.* **10**, 230–252 (2013).

31. Dafni, U. *et al.* Efficacy of adoptive therapy with tumor-infiltrating lymphocytes and recombinant interleukin-2 in advanced cutaneous melanoma: a systematic review and meta-analysis. *Ann. Oncol.* **30**, 1902–1913 (2019).

32. Algarra, I., Garcia-Lora, A., Cabrera, T., Ruiz-Cabello, F. & Garrido, F. The selection of tumor variants with altered expression of classical and nonclassical MHC class I molecules: implications for tumor immune escape. *Cancer Immunol. Immunother.* **53**, (2004).

33. Garrido, F. & Algarra, I. MHC antigens and tumor escape from immune surveillance. in 117–158 (2001). doi:10.1016/S0065-230X(01)83005-0.

34. Cornel, A. M., Mimpen, I. L. & Nierkens, S. MHC class I downregulation in cancer: Underlying mechanisms and potential targets for cancer immunotherapy. *Cancers (Basel).* **12**, 1–33 (2020).

35. Lee, J. S., Park, S. S., Lee, Y. K., Norton, J. A. & Jeffrey, S. S. Liquid biopsy in pancreatic ductal adenocarcinoma: current status of circulating tumor cells and circulating tumor DNA. *Mol. Oncol.* **13**, 1623–1650 (2019).

36. Bracht, J. W. P., Mayo-de-Las-Casas, C., Berenguer, J., Karachaliou, N. & Rosell, R. The Present and Future of Liquid Biopsies in Non-Small Cell Lung Cancer: Combining Four Biosources for Diagnosis, Prognosis, Prediction, and Disease Monitoring. *Curr. Oncol. Rep.* **20**, 70 (2018).

37. Lennon, K. M. *et al.* Single molecule characterization of individual extracellular vesicles from pancreatic cancer. *J. Extracell. Vesicles* **8**, 1685634 (2019).

38. Liu, X. *et al.* Combined preoperative concentrations of CEA, CA 19-9, and 72-4 for predicting outcomes in patients with gastric cancer after curative resection. *Oncotarget* **7**, 35446–35453 (2016).

39. Molina, V. *et al.* CA 19-9 in pancreatic cancer: Retrospective evaluation of patients with suspicion of pancreatic cancer. *Tumor Biol.* **33**, 799–807 (2012).

40. Rhim, A. D. *et al.* EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell* **148**, 349–361 (2012).

41. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin.*

Invest. **119**, 1420–1428 (2009).

42. Dong, X., Alpaugh, R. K. & Cristofanilli, M. Circulating tumor cells (CTCs) in breast cancer: a diagnostic tool for prognosis and molecular analysis. *Chinese J. Cancer Res.* **24**, 388–398 (2012).

43. Sharma, S. *et al.* Circulating tumor cell isolation, culture, and downstream molecular analysis. *Biotechnol. Adv.* **36**, 1063–1078 (2018).

44. Lim, M., Kim, C. J., Sunkara, V., Kim, M. H. & Cho, Y. K. Liquid biopsy in lung cancer: Clinical applications of circulating biomarkers (CTCs and ctDNA). *Micromachines* **9**, (2018).

45. Scheel, C. & Weinberg, R. A. Cancer stem cells and epithelial – mesenchymal transition: Concepts and molecular links. *Semin. Cancer Biol.* **22**, 396–403 (2012).

46. Massagué, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells. *Nature* **529**, 298–306 (2016).

47. Bednarz-knoll, N., Alix-panabières, C. & Pantel, K. Plasticity of disseminating cancer cells in patients with epithelial malignancies. *Cancer Metastasis Rev.* **31**, 673–687 (2012).

48. Polyak, K. & Weinberg, R. A. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat. Rev. Cancer* **9**, 265–273 (2009).

49. Yang, J. & Weinberg, R. A. Epithelial-Mesenchymal Transition : At the Crossroads of Development and Tumor Metastasis. *Dev. Cell* **14**, 818–829 (2008).

50. Liu, F., Gu, L., Shan, B., Geng, C. & Sang, M.-X. Biomarkers for EMT and MET in breast cancer: An update (Review). *Oncol. Lett.* **12**, 4869–4876 (2016).

51. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Pediatr. Infect. Dis. J.* **3**, 453–458 (2003).

52. Tayoun *et al.* CTC-Derived Models: A Window into the Seeding Capacity of Circulating Tumor Cells (CTCs). *Cells* **8**, 1145 (2019).

53. Krasnova, Y., Putz, E. M., Smyth, M. J. & Souza-Fonseca-Guimaraes, F. Bench to bedside: NK cells and control of metastasis. *Clin. Immunol.* **177**, 50–59 (2017).

54. Mathias, T. J., Chang, K. T., Martin, S. S. & Vitolo, M. I. Gauging the impact of cancer treatment modalities on circulating tumor cells (Ctcs). *Cancers (Basel).* **12**, 1–20 (2020).

55. Haber, D. A. & Velculescu, V. E. Blood-Based Analyses of Cancer: Circulating Tumor Cells and Circulating Tumor DNA. *Cancer Discov.* **4**, 650–661 (2014).

56. Kolinsky, M. P. *et al.* Genetic Analysis of Circulating Tumour Cells. in vol. 215 57–76 (2020).

57. Leung, F. *et al.* Circulating tumor DNA as a cancer biomarker: Fact or fiction? *Clin. Chem.* **62**, 1054–1060 (2016).

58. Kustanovich, A., Schwartz, R., Peretz, T. & Grinshpun, A. Life and death of circulating cell-free DNA. *Cancer Biol. Ther.* **20**, 1057–1067 (2019).

59. Gauthier, V. J., Tyler, L. N. & Mannik, M. Blood clearance kinetics and liver uptake of mononucleosomes in mice. *J. Immunol.* **156**, 1151–6 (1996).

60. Kinugasa, H. *et al.* Detection of K-ras gene mutation by liquid biopsy in patients with pancreatic cancer. *Cancer* **121**, 2271–2280 (2015).

61. Pietrasz, D. *et al.* Plasma circulating tumor DNA in pancreatic cancer patients is a prognostic marker. *Clin. Cancer Res.* **23**, 116–123 (2017).

62. Krug, A. K. *et al.* Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma. *Ann. Oncol.* (2018) doi:10.1093/annonc/mdx765.

63. Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early stage lung cancer evolution. *Nature* **545**, 446–451 (2017).

64. Odogwu, L. *et al.* FDA Benefit-Risk Assessment of Osimertinib for the Treatment of Metastatic Non-Small Cell Lung Cancer Harboring Epidermal Growth Factor Receptor T790M Mutation. *Oncologist* **23**, 353–359 (2018).

65. Pitt, J. M., Kroemer, G. & Zitvogel, L. Extracellular vesicles: Masters of intercellular communication and potential clinical interventions. *J. Clin. Invest.* **126**, 1139–1143 (2016).

66. Maas, S. L. N., Breakefield, X. O. & Weaver, A. M. Extracellular Vesicles: Unique Intercellular Delivery Vehicles. *Trends Cell Biol.* **27**, 172–188 (2017).

67. Cui, S., Cheng, Z., Qin, W. & Jiang, L. Exosomes as a liquid biopsy for lung cancer. *Lung Cancer* **116**, 46–54 (2018).

68. Colombo, M., Raposo, G. & Théry, C. Biogenesis, Secretion, and Intercellular Interactions of Exosomes and Other Extracellular Vesicles. *Annu. Rev. Cell Dev. Biol.* **30**, 255–289 (2014).

69. Raposo, G. & Stoorvogel, W. Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell Biol.* **200**, 373–383 (2013).

70. Théry, C., Zitvogel, L. & Amigorena, S. Exosomes: composition, biogenesis and function. *Nat. Rev. Immunol.* **2**, 569–579 (2002).

71. Lobb, R. J. *et al.* Oncogenic transformation of lung cells results in distinct exosome protein profile similar to the cell of origin. **1600432**, 1–10 (2017).

72. Rabinowits, G., Gerçel-Taylor, C., Day, J. M., Taylor, D. D. & Kloecker, G. H. Exosomal microRNA: a diagnostic marker for lung cancer. *Clin. Lung Cancer* **10**, 42–6 (2009).

73. Camussi, G. *et al.* Exosome/microvesicle-mediated epigenetic reprogramming of cells. *Am. J. Cancer Res.* **1**, 98–110 (2011).

74. Melo, S. A. *et al.* Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. (2015) doi:10.1038/nature14581.

75. Taylor, D. D. & Gercel-Taylor, C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.* **110**, 13–21 (2008).

76. Shen, M. *et al.* Progress in exosome associated tumor markers and their detection methods. 1–25 (2020).

77. Théry, C. Exosomes: secreted vesicles and intercellular communications. *Biol. Reports* **33410**, 15–3 (2011).

78. Simons, M. & Raposo, G. Exosomes – vesicular carriers for intercellular communication. *Curr. Opin. Cell Biol.* **21**, 575–581 (2009).

79. Tkach, M. & Théry, C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. *Cell* **164**, 1226–1232 (2016).

80. Hessvik, N. P. & Llorente, A. Current knowledge on exosome biogenesis and release. *Cell. Mol. Life Sci.* **75**, 193–208 (2018).

81. Cazzoli, R. *et al.* MicroRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. *J. Thorac. Oncol.* **8**, 1156–1162 (2013).

82. Jakobsen, K. R. *et al.* Exosomal proteins as potential diagnostic markers in advanced non-small cell lung carcinoma. *J Extracell Vesicles* **4**, 26659 (2015).

83. Sandfeld-Paulsen, B. *et al.* Exosomal proteins as prognostic biomarkers in non-small cell lung cancer. *Mol. Oncol.* **10**, 1234 (2016).

84. Liang, K. *et al.* Nanoplasmonic quantification of tumour-derived extracellular vesicles in plasma microsamples for diagnosis and treatment monitoring. *Nat. Biomed. Eng.* **1**, 1–11 (2017).

85. Toss, A., Mu, Z., Fernandez, S. & Cristofanilli, M. CTC enumeration and characterization: Moving toward personalized medicine. *Ann. Transl. Med.* **2**, (2014).

86. CellSearch. The Clinical Relevance of the CELLSEARCH® Circulating Tumor Cell Test. https://www.cellsearchctc.com/clinical-applications/clinical-relevance (2017).

87. Lianidou, E. S. & Markou, A. Circulating Tumor Cells in Breast Cancer: Detection Systems, Molecular Characterization, and Future Challenges. *Clin. Chem.* **57**, 1242–1255 (2011).

88. Paoletti, C. & Hayes, D. F. Circulating Tumor Cells. in *Novel Biomarkers in the Continuum of Breast Cancer* 235–258 (2016).

89. Cohen, S. J. *et al.* Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 3213–3221 (2008).

90. Hayes, D. F. *et al.* Circulating Tumor Cells at Each Follow-up Time Point during Therapy of Metastatic Breast Cancer Patients Predict Progression-Free and Overall Survival. *Clin. Cancer Res.* **12**, 4218–4224 (2006).

91. de Bono, J. S. *et al.* Circulating Tumor Cells Predict Survival Benefit from Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clin. Cancer Res.* **14**, 6302–6309 (2008).

92. Madhuprasad, P. P., Kumeria, T., Losic, D. & Kurkuri, M. Isolation of circulating tumour cells by physical means in a microfluidic device: a review. *RSC Adv.* **5**, 89745–89762 (2015).

93. Esmaeilsabzali, H., Beischlag, T. V., Cox, M. E., Parameswaran, A. M. & Park, E. J. Detection and isolation of circulating tumor cells: Principles and methods. *Biotechnol. Adv.* **31**, 1063–1084 (2013).

94. Mostert, B., Sleijfer, S., Foekens, J. A. & Gratama, J. W. Circulating tumor cells (CTCs): Detection methods and their clinical relevance in breast cancer. *Cancer Treat. Rev.* **35**, 463–474 (2009).

95. Alunni-fabbroni, M. & Sandri, M. T. Circulating tumour cells in clinical practice: Methods of detection and possible characterization. *Methods* **50**, 289–297 (2010).

96. Kozminsky, M., Wang, Y. & Nagrath, S. The incorporation of microfluidics into circulating tumor cell isolation for clinical applications. *Curr. Opin. Chem. Eng.* **11**, 59–66 (2016).

97. Rhim, A. D. *et al.* EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell* **148**, 349–361 (2012).

98. Kozminsky, M., Wang, Y. & Nagrath, S. The incorporation of microfluidics into circulating tumor cell isolation for clinical applications. *Curr. Opin. Chem. Eng.* **11**, 59–66 (2016).

99. Nagrath, S. *et al.* Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450**, 1235–1239 (2007).

100. Sequist, L. V, Nagrath, S., Toner, M., Haber, D. A. & Lynch, T. J. The CTC-Chip: An Exciting New Tool to Detect Circulating Tumor Cells in Lung Cancer Patients. *J. Thorac. Oncol.* **4**, 281–283 (2009).

101. Stott, S. L. *et al.* Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. *Proc. Natl. Acad. Sci.* **107**, 18392–18397 (2010).

102. Yoon, H. J. *et al.* Sensitive capture of circulating tumour cells by functionalized graphene oxide nanosheets. *Nat. Nanotechnol.* **8**, 735–741 (2013).

103. Murlidhar, V. *et al.* A Radial Flow Microfluidic Device for Ultra-High-Throughput Affi nity-Based Isolation of Circulating Tumor Cells. *Small* **10**, 4895–4904 (2014).

104. Dong, Y. *et al.* Microfluidics and Circulating Tumor Cells. *J. Mol. Diagnositcs* **15**, 149–157 (2013).

105. Bhagat, A. A. S., Kuntaegowdanahalli, S. S. & Papautsky, I. Continuous particle separation in spiral microchannels using Dean flows and differential migration. *Lab Chip* **8**, 1906–1914 (2008).

106. Kuntaegowdanahalli, S. S., Bhagat, A. A. S., Kumar, G. & Papautsky, I. Inertial microfluidics for continuous particle separation in spiral microchannels. *Lab Chip* **9**, 2973 (2009).

107. Chatterjee, A., Kuntaegowdanahalli, S. S. & Papautsky, I. Inertial microfluidics for continuous separation of cells and particles. **7929**, 1–10 (2011).

108. Di Carlo, D. Inertial microfluidics. *Lab Chip* **9**, 3038 (2009).

109. Lin, E. *et al.* High-Throughput Microfluidic Labyrinth for the Label-free Isolation of CTCs for Single Cell Gene Expression Profiling. *Cell Syst.* **5**, 295–304 (2017).

110. Ozkumur, E. *et al.* Inertial Focusing for Tumor Antigen–Dependent and –Independent Sorting of Rare Circulating Tumor Cells. *Sci. Transl. Med.* **5**, 1–11 (2013).

111. Jack, R. M. *et al.* Ultra-Specific Isolation of Circulating Tumor Cells Enables Rare-Cell RNA Profiling. *Adv. Sci.* **3**, (2016).

112. Hayes, D. F. *et al.* Tumor Cells at Each Follow-up Time Point during Therapy of Metastatic Breast Cancer Patients Predict Progression-Free and Overall Survival. *Clin. Cancer Res.* **12**, 4218–4224 (2006).

113. Wang, C. *et al.* Longitudinally collected CTCs and CTC-clusters and clinical outcomes of metastatic breast cancer. *Breast Cancer Res. Treat.* **161**, 83–94 (2017).

114. Yu, M. *et al.* Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Science (80-.).* **339**, 580–584 (2013).

115. Cheung, K. J. & Ewald, A. J. A collective route to metastasis: Seeding by tumor cell clusters. *Science (80-.).* **352**, 167–169 (2016).

116. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).

117. Raimondi, C. *et al.* Epithelial-mesenchymal transition and stemness features in circulating tumor cells from breast cancer patients. *Breast Cancer Reseach Treat.* **130**, 449–455 (2011).

118. Markiewicz, A. *et al.* Mesenchymal phenotype of CTC-enriched blood fraction and lymph node metastasis formation potential. *PLoS One* **9**, (2014).

119. Miyamoto, D. T. *et al.* RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science (80-.).* **349**, 1351–1356 (2015).

120. Gong, H., Do, D. & Ramakrishnan, R. Single-Cell mRNA-Seq Using the Fluidigm C1 System and Integrated Fluidics Circuits. in *Gene Expression Analysis: Methods and Protocols* (eds. Raghavachari, N. & Garcia-Reyero, N.) 193–207 (Springer New York, 2018). doi:10.1007/978-1-4939-7834-2_10.

121. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: Current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).

122. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

123. Gierahn, T. M. *et al.* Seq-Well: Portable, low-cost rna sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).

124. Di Trapani, M., Manaresi, N. & Medoro, G. DEPArray™ system: An automatic image-based sorter for isolation of pure circulating tumor cells. *Cytom. Part A* **93**, 1260–1266 (2018).

125. Sharma, S. *et al.* Using single cell analysis for translational studies in immune mediated diseases : Opportunities and challenges. *Mol. Immunol.* **103**, 191–199 (2018).

126. Grün, D. & Van Oudenaarden, A. Design and Analysis of Single-Cell Sequencing Experiments. *Cell* **163**, 799–810 (2015).

127. Ortiz, V. & Yu, M. Analyzing Circulating Tumor Cells One at a Time. *Trends Cell Biol.* **28**, 764–775.

128. Wu, C., Wu, P., Zhao, H., Liu, W. & Li, W. Clinical Applications of and Challenges in Single-Cell Analysis of Cirulating Tumor Cells. *DNA Cell Biol.* **37**, 78–89 (2018).

129. Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat. Biotechnol.* **32**, 479–484 (2014).

130. Ni, X. *et al.* Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 21083–21088 (2013).

131. De Luca, F. *et al.* Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. *Oncotarget* **7**, 26107–26119 (2016).

132. Shalek, A. K. & Benson, M. Single-cell analyses to tailor treatments. *Sci. Transl. Med.* **9**, 1–4 (2017).

133. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).

134. Rodriguez-Meira, A. *et al.* Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol. Cell* **73**, 1292-1305.e8 (2019).

135. Malone, J. H. & Oliver, B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**, 34 (2011).

136. Smirnov, D. A. *et al.* Global gene expression profiling of circulating tumor cells. *Cancer Res.* **65**, 4993–4997 (2005).

137. Obermayr, E. *et al.* Assessment of a six gene panel for the molecular detection of circulating tumor cells in the blood of female cancer patients. *BMC Cancer* **10**, (2010).

138. Zhang, Z. *et al.* Expansion of CTCs from early stage lung cancer patients using a microfluidic co-culture model. *Oncotarget* **5**, 12383–12397 (2014).

139. Cayrefourcq, L. *et al.* Establishment and Characterization of a Cell Line from Human Circulating Colon Cancer Cells. *Cancer Res.* **75**, 892–901 (2015).

140. Yu, M. *et al.* Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science (80-.).* **345**, 216–220 (2014).

141. Khoo, B. L. *et al.* Liquid biopsy and therapeutic response: Circulating tumor cell cultures for evaluation of anticancer treatment. *Sci. Adv.* **2**, e1600274 (2016).

142. Smit, D. J. *et al.* High Sensitivity of Circulating Tumor Cells Derived from a Colorectal Cancer Patient for Dual Inhibition with AKT and mTOR Inhibitors. *Cells* **9**, 2129 (2020).

143. Hamilton, G., Burghuber, O., Zeillinger, R. & Hamilton, G. Circulating Tumor Cells in Small Cell Lung Cancer : Ex Vivo Expansion. *Lung* **193**, 451–452 (2015).

144. Cegan, M. *et al.* In vitro culturing of viable circulating tumor cells of urinary bladder cancer. **7**, 7164–7171 (2014).

145. Khoo, B. L. *et al.* Short-term expansion of breast circulating cancer cells predicts response to anti-cancer therapy. *Oncotarget* **6**, 15578–15593 (2015).

146. Koch, C. *et al.* Characterization of circulating breast cancer cells with tumorigenic and metastatic capacity. *EMBO Mol. Med.* 1–22 (2020) doi:10.15252/emmm.201911908.

147. Rivera-Báez, L. *et al.* Expansion of Circulating Tumor Cells from Patients with Locally Advanced Pancreatic Cancer Enable Patient Derived Xenografts and Functional Studies for Personalized Medicine. *Cancers (Basel).* **12**, 1011 (2020).

148. Soler, A. *et al.* Autologous cell lines from circulating colon cancer cells captured from sequential liquid biopsies as model to study therapy-driven tumor changes. *Sci. Rep.* **8**, 1–12 (2018).

149. Patel, G. K. *et al.* Comparative analysis of exosome isolation methods using culture supernatant for optimum yield, purity and downstream applications. *Sci. Rep.* **9**, 5335 (2019).

150. Lobb, R. J. *et al.* Optimized exosome isolation protocol for cell culture supernatant and human plasma. **3078**, (2015).

151. Théry, C., Clayton, A., Amigorena, S., Raposo, G. & Clayton, A. Isolation and Characterization of Exosomes from Cell Culture Supernatants. *Curr. Protoc. Cell Biol.* **Chapter 3**, 1–29 (2006).

152. Guo, S. C., Tao, S. C. & Dawn, H. Microfluidics-based on-a-chip systems for isolating and analysing extracellular vesicles. *J. Extracell. Vesicles* **7**, (2018).

153. Thakur, B. K. *et al.* Double-stranded DNA in exosomes: A novel biomarker in cancer detection. *Cell Res.* **24**, 766–769 (2014).

154. Manda, S. V. *et al.* Exosomes as a biomarker platform for detecting epidermal growth factor receptor-positive high-grade gliomas. *J. Neurosurg.* **128**, 1091–1101 (2018).

155. Figueroa, J. M. *et al.* Detection of wild-Type EGFR amplification and EGFRvIII mutation in CSF-derived extracellular vesicles of glioblastoma patients. *Neuro. Oncol.* **19**, 1494–1502 (2017).

156. Allenson, K. *et al.* High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients Original article. 741–747 (2018) doi:10.1093/annonc/mdx004.

157. Chen, W. W. *et al.* Beaming and droplet digital pcr analysis of mutant idh1 mrna in glioma patient serum and cerebrospinal fluid extracellular vesicles. *Mol. Ther. - Nucleic Acids* **2**, e109 (2013).

158. Castellanos-rizaldos, E. *et al.* Exosome-based Detection of EGFR T790M in Plasma from Non-Small Cell Lung Cancer Patients. (2018) doi:10.1158/1078-0432.CCR-17-3369.

159. Dong, J. *et al.* Bio-Inspired NanoVilli Chips for Enhanced Capture of Tumor-Derived Extracellular Vesicles: Toward Non-Invasive Detection of Gene Alterations in Non-Small Cell Lung Cancer. *ACS Appl. Mater. Interfaces* **11**, 13973–13983 (2019).

160. Denis, J. A., Guillerm, E., Coulet, F., Larsen, A. K. & Lacorte, J. M. The Role of BEAMing and Digital PCR for Multiplexed Analysis in Molecular Oncology in the Era of Next-Generation Sequencing. *Mol. Diagnosis Ther.* **21**, 587–600 (2017).

161. Milbury, C. A. *et al.* Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. *Biomol. Detect. Quantif.* **1**, 8–22 (2014).

162. Whale, A. S., Huggett, J. F. & Tzonev, S. Biomolecular Detection and Quantification Fundamentals of multiplexing with digital PCR. *Biomol. Detect. Quantif.* **10**, 15–23 (2016).

163. Zonta, E. *et al.* Multiplex Detection of Rare Mutations by Picoliter Droplet Based Digital PCR: Sensitivity and Specificity Considerations. *PLoS One* **11**, e0159094 (2016).

164. Whale, A. S., Huggett, J. F. & Tzonev, S. Fundamentals of multiplexing with digital PCR. *Biomol. Detect. Quantif.* **10**, 15–23 (2016).

165. Zhang, J. *et al.* Fundamentals and Applications of Inertial Microfluidics: A Review. *Lab Chip* (2015) doi:10.1039/C5LC01159K.

166. Trifunov, S. *et al.* Clonal expansion of mtDNA deletions : different disease models assessed by digital droplet PCR in single muscle cells. 1–10 (2018) doi:10.1038/s41598-018-30143-z.

167. Zhang, Q. *et al.* Isolation of circulating tumor cells and detection of EGFR mutations in patients with non-small-cell lung cancer. *Oncol. Lett.* **17**, 3799–3807 (2019).

168. Stilla. *Naica System for Crystal Digital PCR*.

169. Stilla. *3-Color Crystal Digital PCR assays for EGFR mutation detection*.

170. COMBiNATi. *Absolute Q dPCR Platform Digital PCR, Re-imagined*.

171. COMBiNATi. *Experiment Flexibilty with the Combinati ™ Absolute Q Digital PCR Platform*. vol. 1.

172. Stilla. *Droplet Recovery with Crystal Digtal PCR*.

173. Basu, A. S. Digital Assays Part I: Partitioning Statistics and Digital PCR. *SLAS Technol.* **22**, 369–386 (2017).

174. Baker, M. Digital PCR hits its stride. *Nat. Methods* **9**, 541–544 (2012).

175. Fluidigm. Digital PCR Analysis. *Analysis* (2009).

176. Metro, G. & Crinò, L. Advances on EGFR mutation for lung cancer. *Transl. Lung Cancer Res.* **1**, 5–13 (2012).

177. Haber, D. A. & Velculescu, V. E. Blood-Based Analyses of Cancer: Circulating Tumor Cells and Circulating Tumor DNA. *Cancer Discov.* **4**, 650–661 (2014).

178. U.S. Food and Drug Administration. Summary of safety and effectiveness data: cobas EGFR Muta- tion Test v2. September 28, 2016. , *Data, PMA P150044: FDA Summary of Safety and Effectiveness* vol. 39 1–30 http://www.accessdata.fda.gov/cdrh_docs/pdf15/P150044B.pdf. (2016).

179. Kim, O. *et al.* Whole Genome Sequencing of Single Circulating Tumor Cells Isolated by Applying a Pulsed Laser to Cell-Capturing Microstructures. *Small* **15**, 1–8 (2019).

180. Dong, J. *et al.* Nanostructured Substrates for Detection and Characterization of Circulating Rare Cells: From Materials Research to Clinical Applications. *Adv. Mater.* **32**, 1903663 (2020).

181. Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).

182. Jiang, X.-W., Liu, W., Zhu, X.-Y. & Xu, X.-X. Evaluation of EGFR mutations in NSCLC with highly sensitive droplet digital PCR assays. *Mol. Med. Rep.* **20**, 593–603 (2019).

183. Soh, J. *et al.* Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. *PLoS One* **4**, e7464 (2009).

184. Hargadon, K. M. Dysregulation of TGFβ1 Activity in Cancer and Its Influence on the Quality of Anti-Tumor Immunity. *J. Clin. Med.* **5**, 76 (2016).

185. Yu, C.-C. *et al.* Mutant allele specific imbalance in oncogenes with copy number alterations: Occurrence, mechanisms, and potential clinical implications. *Cancer Lett.* **384**, 86–93 (2017).

186. Malapelle, U. *et al.* EGFR mutant allelic-specific imbalance assessment in routine samples of non-small cell lung cancer. *J. Clin. Pathol.* **68**, 739–741 (2015).

187. Sun, H. *et al.* Association between hormone receptor expression and epidermal growth factor receptor mutation in patients operated on for non-small cell lung cancer. *Ann. Thorac. Surg.* **91**, 1562–1567 (2011).

188. Hsu, L.-H., Chu, N.-M. & Kao, S.-H. Estrogen, estrogen receptor and lung cancer. *Int. J. Mol. Sci.* **18**, 1–17 (2017).

189. Alix-Panabières, C. & Pantel, K. Clinical applications of circulating tumor cells and circulating tumor DNA as liquid biopsy. *Cancer Discov.* (2016) doi:10.1158/2159-8290.CD-15-1483.

190. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).

191. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16368–16373 (2005).

192. Lobb, R. J. *et al.* Oncogenic transformation of lung cells results in distinct exosome protein profile similar to the cell of origin. *Proteomics* **17**, 1600432 (2017).

193. Rabinowits, G., Gerçel-Taylor, C., Day, J. M., Taylor, D. D. & Kloecker, G. H. Exosomal MicroRNA: A Diagnostic Marker for Lung Cancer. *Clin. Lung Cancer* **10**, 42–46 (2009).

194. Chen, W. W. *et al.* BEAMing and Droplet Digital PCR Analysis of Mutant IDH1 mRNA in Patient Serum With Glioma and Cerebrospinal Fluid Extracellular Vesicles. 1–10 (2013) doi:10.1038/mtna.2013.28.

195. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, 0225–0235 (2005).

196. Balak, M. N. *et al.* Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin. Cancer Res.* **12**, 6494–6501 (2006).

197. Clark, J., Cools, J. & Gilliland, D. G. EGFR inhibition in non-small cell lung cancer: Resistance, once again, rears its ugly head. *PLoS Med.* **2**, 0195–0197 (2005).

198. Kosaka, T. *et al.* Mutations of the epidermal growth factor receptor gene in lung cancer: Biological and clinical implications. *Cancer Res.* **64**, 8919–8923 (2004).

199. Yoshioka, Y. *et al.* Comparative marker analysis of extracellular vesicles in different human cancer types. *J. Extracell. Vesicles* **2**, 1–9 (2013).

200. Rivero-Gutiérrez, B., Anzola, A., Martínez-Augustin, O. & de Medina, F. S. Stain-free detection as loading control alternative to Ponceau and housekeeping protein immunodetection in Western blotting. *Anal. Biochem.* **467**, 1–3 (2014).

201. Gilda, J. E. & Gomes, A. V. Stain-Free total protein staining is a superior loading control to β-actin for Western blots. *Anal. Biochem.* **440**, 186–188 (2013).

202. Zeinali, M. *et al.* High-throughput label-free isolation of heterogeneous circulating tumor cells and CTC clusters from non-small-cell lung cancer patients. *Cancers (Basel).* **12**, 1–17 (2020).

203. Kang, Y. T. *et al.* Multiplex isolation and profiling of extracellular vesicles using a microfluidic DICE device. *Analyst* **144**, 5785–5793 (2019).

204. Kanwar, S. S., Dunlay, C. J., Simeone, D. M. & Nagrath, S. Microfluidic device (ExoChip) for on-chip isolation, quantification and characterization of circulating exosomes. *Lab Chip* **14**, 1891–900 (2014).

205. Tang, Y. *et al.* Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum. 834–844 (2017) doi:10.3892/ijmm.2017.3080.

206. Li, M. *et al.* Analysis of the RNA content of the exosomes derived from blood serum and urine and its potential as biomarkers. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130502–20130502 (2014).

207. Hugenschmidt, H. *et al.* Circulating Tumor Cells are an Independent Predictor of Shorter Survival in Patients Undergoing Resection for Pancreatic and Periampullary Adenocarcinoma. *Ann. Surg.* **271**, 549–558 (2020).

208. Habib, J. & Yu, J. Circulating tumor cells in pancreatic cancer: a review. *J. Pancreatol.* **2**, (2019).

209. Hayes, D. F. *et al.* Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin. Cancer Res.* **12**, 4218–4224 (2006).

210. Bono, J. S. De *et al.* Cancer Therapy: Clinical Circulating T umor Cells Predict Survival

Benefit from T reatment in Metastatic Castration-Resistant Prostate Cancer. *Clin. Cancer Res.* **14**, 6302–6309 (2008).

211. Kozminsky, M. *et al.* Detection of CTC Clusters and a Dedifferentiated RNA-Expression Survival Signature in Prostate Cancer. *Adv. Sci.* **6**, (2019).

212. Li, G., Huang, Y., Manjunath, Y. & Kimchi, E. T. Clinical Significance of Circulating Tumor Cells in Pancreatic Cancer. *Biol. Med.* **08**, (2016).

213. Cohen, S. J. *et al.* Relationship of Circulating Tumor Cells to Tumor Response , Progression-Free Survival , and Overall Survival in Patients With Metastatic Colorectal Cancer. *J. Clin* **26**, 3213–3221 (2008).

214. Mitra, A., Mishra, L. & Li, S. EMT, CTCs and CSCs in tumor relapse and drug-resistance. *Oncotarget* **6**, 10697–10711 (2015).

215. Poruk, K. E. *et al.* Circulating tumor cell phenotype predicts recurrence and survival in pancreatic adenocarcinoma. *Ann. Surg.* **264**, 1073–1081 (2016).

216. Nagrath, S., Jack, R. M., Sahai, V. & Simeone, D. M. Opportunities and Challenges for Pancreatic Circulating Tumor Cells. *Gastroenterology* **151**, 412–426 (2016).

217. Kasimir-bauer, S., Hoffmann, O., Wallwiener, D., Kimmig, R. & Fehm, T. Expression of stem cell and epithelial- mesenchymal transition markers in primary breast cancer patients with circulating tumor cells. *Breast Cancer Res.* **14**, R15 (2012).

218. Principe, D. R. & Rana, A. Updated risk factors to inform early pancreatic cancer screening and identify high risk patients. *Cancer Lett.* (2020) doi:10.1016/j.canlet.2020.04.022.

219. Bournet, B., Buscail, C., Muscari, F., Cordelier, P. & Buscail, L. Targeting KRAS for diagnosis, prognosis, and treatment of pancreatic cancer: Hopes and realities. *Eur. J. Cancer* **54**, 75–83 (2016).

220. Kim, S. T. *et al.* Impact of KRAS Mutations on Clinical Outcomes in Pancreatic Cancer Patients Treated with First-line Gemcitabine-Based Chemotherapy. *Mol. Cancer Ther.* **10**, 1993–1999 (2011).

221. Kulemann, B. *et al.* KRAS mutations in pancreatic circulating tumor cells: a pilot study. *Tumor Biol.* **37**, 7547–7554 (2016).

222. Kulemann, B. *et al.* Pancreatic cancer: Circulating Tumor Cells and Primary Tumors show Heterogeneous KRAS Mutations. *Sci. Rep.* **7**, 1–11 (2017).

223. Lin, E. *et al.* High-Throughput Microfluidic Labyrinth for the Label-free Isolation of Circulating Tumor Cells. *Cell Syst.* **5**, 295-304.e4 (2017).

224. Zhu, Z. *et al.* Tumour-reprogrammed stromal BCAT1 fuels branched-chain ketoacid dependency in stromal-rich PDAC tumours. *Nat. Metab.* **2**, 1–18 (2020).

225. Wan, S. *et al.* New Labyrinth Microfluidic Device Detects Circulating Tumor Cells Expressing Cancer Stem Cell Marker and Circulating Tumor Microemboli in Hepatocellular Carcinoma. *Sci. Rep.* **9**, 1–11 (2019).

226. Owen, S. *et al.* Simultaneous Single Cell Gene Expression and EGFR Mutation Analysis of Circulating Tumor Cells Reveals Distinct Phenotypes in NSCLC. *Adv. Biosyst.* **2000110**, 1–11 (2020).

227. Earl, J. *et al.* Circulating tumor cells (CTC) and KRAS mutant circulating free Dna (cfDNA) detection in peripheral blood as biomarkers in patients diagnosed with exocrine pancreatic cancer. *BMC Cancer* **15**, 797 (2015).

228. Lee, A. C. *et al.* OPENchip: an on-chip in situ molecular profiling platform for gene expression analysis and oncogenic mutation detection in single circulating tumour cells. *Lab Chip* **20**, 912–922 (2020).

229. McCarthy, D. J. & Smyth, G. K. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* **25**, 765–771 (2009).

230. Brechbuhl, H. M. *et al.* Analysis of circulating breast cancer cell heterogeneity and interactions with peripheral blood mononuclear cells. *Mol. Carcinog.* **59**, 1129–1139 (2020).

231. Arumugam, T., Simeone, D. M., Van Golen, K. & Logsdon, C. D. S100P Promotes Pancreatic Cancer Growth, Survival, and Invasion. *Clin. Cancer Res.* **11**, 5356–5364 (2005).

232. Rasheed, Z. A. *et al.* Prognostic Significance of Tumorigenic Cells with Mesenchymal Features in Pancreatic Adenocarcinoma. *J. Natl. Cancer Inst.* **102**, 340–351 (2010).

233. Ting, D. T. *et al.* Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Article Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells. *Cell Rep.* **8**, 1905–1918 (2014).

234. Leclerc, E. & Vetter, S. W. The role of S100 proteins and their receptor RAGE in pancreatic cancer. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1852**, 2706–2711 (2016).

235. Mochizuki, S. & Okada, Y. ADAMs in cancer cell proliferation and progression. *Cancer Sci.* **98**, 621–628 (2007).

236. Long, N. P. *et al.* An Integrative Data Mining and Omics-Based Translational Model for

the identification and Validation of Oncogenic Biomarkers of Pancreatic Cancer. *Cancers (Basel).* **11**, (2019).

237. Grützmann, R. *et al.* ADAM9 expression in pancreatic cancer is associated with tumour type and is a prognostic factor in ductal adenocarcinoma. *Br. J. Cancer* **90**, 1053–1058 (2004).

238. Wente, M. N. *et al.* Expression and potential function of the CXC chemokine CXCL16 in pancreatic ductal adenocarcinoma. *Int. J. Oncol.* **33**, 297–308 (2008).

239. Mir, H. *et al.* Higher CXCL16 exodomain is associated with aggressive ovarian cancer and promotes the disease by CXCR6 activation and MMP modulation. *Sci. Rep.* **9**, 1–12 (2019).

240. Fujita, H. *et al.* Gene Expression Levels as Predictive Markers of Outcome in Pancreatic Cancer after Gemcitabine-Based Adjuvant Chemotherapy. *Neoplasia* **12**, 807–817 (2010).

241. Sancho, P. *et al.* MYC/PGC-1α Balance Determines the Metabolic Phenotype and Plasticity of Pancreatic Cancer Stem Cells. *Cell Metab.* **22**, 590–605 (2015).

242. Hakim, N., Patel, R., Devoe, C. & Saif, M. W. Why HALO 301 Failed and Implications for Treatment of Pancreatic Cancer. *Pancreas (Fairfax)* **3**, e1 (2019).

243. Yamamoto, K. *et al. Autophagy promotes immune evasion of pancreatic cancer by degrading MHC-I. Nature* (Springer US, 2020). doi:10.1038/s41586-020-2229-5.

244. Meza Guzman, L. G., Keating, N. & Nicholson, S. E. Natural killer cells: Tumor surveillance and signaling. *Cancers (Basel).* **12**, 1–33 (2020).

245. Lupo, K. B. & Matosevic, S. Natural killer cells as allogeneic effectors in adoptive cancer immunotherapy. *Cancers (Basel).* **11**, (2019).

246. Li, J. & King, M. R. Adhesion receptors as therapeutic targets for circulating tumor cells. *Front. Oncol.* **2**, (2012).

247. Zhang, L. *et al.* Meta-Analysis of the Prognostic Value of Circulating Tumor Cells in Breast Cancer. *Clin. Cancer Res.* **18**, 5701–5710 (2012).

248. Scher, H. I. *et al.* Circulating tumour cells as prognostic markers in progressive, castration-resistant prostate cancer: a reanalysis of IMMC38 trial data. *Lancet Oncol.* **10**, 233–239 (2009).

249. Krebs, M. G. *et al.* Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients With Non–Small-Cell Lung Cancer. *J. Clin. Oncol.* **29**, 1556–1563 (2011).

250. Faltas, B. Cornering metastases: Therapeutic targeting of circulating tumor cells and stem cells. *Front. Oncol.* **2 JUL**, 1–7 (2012).

251. Li, J. *et al.* Targeted drug delivery to circulating tumor cells via platelet membrane-functionalized particles. *Biomaterials* **76**, 52–65 (2016).

252. López-Soto, A. *et al.* Epithelial–Mesenchymal Transition Induces an Antitumor Immune Response Mediated by NKG2D Receptor. *J. Immunol.* **190**, 4408–4419 (2013).

253. Valastyan, S. & Weinberg, R. A. Tumor Metastasis : Molecular Insights and Evolving Paradigms. *Cell* **147**, 275–292 (2011).

254. Uchida, A. The cytolytic and regulatory role of natural killer cells in human neoplasia. *Biochim. Biophys. Acta - Rev. Cancer* **865**, 329–340 (1986).

255. Jordanova, E. S. *et al.* Human leukocyte antigen class I, MHC class I chain-related molecule A, and CD8+/regulatory T-cell ratio: Which variable determines survival of cervical cancer patients? *Clin. Cancer Res.* **14**, 2028–2035 (2008).

256. Chockley, P. J. *et al.* Epithelial-mesenchymal transition leads to NK cell-mediated metastasis-specific immunosurveillance in lung cancer. *J. Clin. Invest.* **128**, 1384–1396 (2018).

257. Karade, S. S. & Mariuzza, R. A. How Natural Killer Cell Receptors Stick to Cell–Cell Adhesion Proteins. *Structure* **27**, 209–210 (2019).

258. Li, Y. *et al.* Structure of Natural Killer Cell Receptor KLRG1 Bound to E-Cadherin Reveals Basis for MHC-Independent Missing Self Recognition. *Immunity* **31**, 35–46 (2009).

259. Okita, R., Shimizu, K. & Nakata, M. Epithelial-mesenchymal transition-induced metastasis could be a bait for natural killer cells. *J. Thorac. Dis.* **10**, S3143–S3146 (2018).

260. Okita, R. *et al.* MHC class I chain-related molecule A and B expression is upregulated by cisplatin and associated with good prognosis in patients with non-small cell lung cancer. *Cancer Immunol. Immunother.* **65**, 499–509 (2016).

261. Bauer, S. *et al.* Activation of NK Cells and T Cells by NKG2D, a Receptor for Stress-Inducible MICA. *Science (80-.).* **285**, 727–729 (1999).

262. Zhao, Y. *et al.* Prognostic value of MICA/B in cancers: A systematic review and meta-analysis. *Oncotarget* **8**, 96384–96395 (2017).

263. Frazao, A. *et al.* NKG2D/NKG2-Ligand Pathway Offers New Opportunities in Cancer Treatment. *Front. Immunol.* **10**, 661 (2019).

264. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

265. Chockley, P. J. & Keshamouni, V. G. Immunological Consequences of Epithelial–Mesenchymal Transition in Tumor Progression. *J. Immunol.* **197**, 691–698 (2016).

266. Pantel, K., Schaller, G., Kutter, D., Genz, T. & Riethmüller, G. Frequent Down-Regulation of Major Histocompatibility Class I Antigen Expression on Individual Micrometastatic Carcinoma Cells. *Cancer Res.* **51**, 4712–4715 (1991).

267. Huaman, J., Naidoo, M., Zang, X. & Ogunwobi, O. O. Fibronectin Regulation of Integrin B1 and SLUG in Circulating Tumor Cells. *Cells* **8**, 618 (2019).

268. Lo, H. C. *et al.* Resistance to natural killer cell immunosurveillance confers a selective advantage to polyclonal metastasis. *Nat. Cancer* **1**, 709–722 (2020).

269. Sharma, S., Zhuang, R., Long, M., Pavlovic, M. & Kang, Y. Circulating tumor cell isolation , culture , and downstream molecular analysis. *Biotechnol. Adv.* **36**, 1063–1078 (2018).

270. Wang, R. *et al.* Cultured circulating tumor cells and their derived xenografts for personalized oncology. *Asian J. Urol.* **3**, 240–253 (2016).

271. Okita, R. *et al.* Contrasting effects of the cytotoxic anticancer drug gemcitabine and the EGFR tyrosine kinase inhibitor gefitinib on nk cell-mediated cytotoxicity via regulation of NKG2D ligand in non-small-cell lung cancer cells. *PLoS One* **10**, 1–20 (2015).

272. Brea, E. J. *et al.* Kinase regulation of human MHC class i molecule expression on cancer cells. *Cancer Immunol. Res.* **4**, 936–947 (2016).

273. Gupta, G. P. & Massagué, J. Cancer Metastasis: Building a Framework. *Cell* **127**, 679–695 (2006).

274. Scherbakova, A., Lust, H., Everaus, H. & Aints, A. A mathematical model of natural killer cell activity. *Cytom. Part A* **83A**, 585–591 (2013).

275. Neu, B., Wenby, R. & Meiselman, H. J. Effects of dextran molecular weight on red blood cell aggregation. *Biophys. J.* **95**, 3059–3065 (2008).

276. Institute, S. Cosmic - Catalogue of Somatic Mutations in Cancer. http://cancer.sanger.ac.uk/cosmic.