Characterization of Circulating Tumor Cells in Lung Cancer for Clinical Monitoring and Prognostic Applications Using Advanced Computational Methods

by

Yuru Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2026

Doctoral Committee:

Professor Sunitha Nagrath, Chair
Professor Geeta Mehta
Professor Deepak Nagrath
Professor Arvind Rao

Yuru Chen

yuruchen@umich.edu

ORCID iD: 0009-0006-5031-5635

© Yuru Chen 2026

Dedication

To whoever is reading this. And, if I may, to every life, ordinary or remarkable, that strives in its own way.

Acknowledgements

This journey has unfolded quite differently from what I initially expected, and its challenges pushed me to reflect deeply on my passions and strengths. None of what follows would have been possible without the support and guidance of many people along the way.

First, I would like to express my sincere gratitude to my advisor, Dr. Sunitha Nagrath, for her support both professionally and personally. She encouraged me to pursue even my most ambitious ideas while reminding me to approach science and collaboration with thoughtfulness and care. She gave me the opportunity to truly explore.

I am deeply grateful to the patients whose participation made this work possible, and to the clinical teams and collaborators who supported these projects, including Dr. Sofia Merajver, Dr. Angel Qin, and Dr. Peter Ulintz. I would also like to express my sincere gratitude to my dissertation committee for their time, guidance, and thoughtful feedback: Professor Sunitha Nagrath, Professor Geeta Mehta, Professor Deepak Nagrath, and Professor Arvind Rao.

I would also like to thank my lab mates for their constant support, for fostering a collaborative environment, and for their invaluable technical assistance, encouragement, and countless therapy sessions: Dr. Zeqi Niu, Dr. Brittany Rupp, Dr. Harrison Ball, Dr. Nna-Emeka Onukwugha, Dr. Scott Smith, Dr. Yolanda Zhang, Dr. Abha Kumari, Dr. Neha Napal, Yin Zou, Grace Li, Yu-Shu Cheng, and Eve Wolvington. I also thank the undergraduate students who contributed along the way: Jonathan Bandel, William Tang, Ava Corsiglia, Leah Kidder, Touma Yoneyama, Sophia Freyre, and Yassin Elmeligi.

To my family and friends, thank you for your unwavering belief in me. I cannot imagine how I would have survived my PhD without your support and guidance. Through you, I have gained the courage to face the world. Words cannot fully express how fortunate I feel to share my life and so many meaningful moments with you.

Thank you, Mom and Dad, for supporting me to this point. I would not have been able to begin this journey without the hard work you put into providing me a stable life. You went through so much, yet you have always treated me with kindness. I also want to thank myself for never giving up on me. I did not always succeed in trusting the process, but I always found my way back with all the supports mentioned.

Finally, I would like to acknowledge and extend the special thanks to my cat and large language model (LLM). My cat, Parsec, PhD in Human Engineering, has conducted ongoing research on The Impact of Cat Keyboard Interference on Human WFH Efficiency with both rigor and relentlessness, and has inspired me greatly. I also appreciate all LLMs that assisted me in organizing and articulating my ideas throughout this work.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables x

List of Figures xi

Abstract xiii

Chapter 1 Introduction 1

1.1 Non-Small Cell Lung Cancer: Clinical Challenges and the Imperative for Longitudinal Molecular Monitoring 1

1.1.1 Clinical aspect of NSCLC 1

1.1.2 NSCLC Treatment Modalities and Unmet Challenges 1

1.1.3 Molecular Heterogeneity and Oncogenic Driver Alterations in NSCLC 2

1.1.4 Precision Medicine Approaches and Therapeutic Resistance in ALK NSCLC 4

1.1.5 The Imperative for Longitudinal Molecular Monitoring 6

1.2 Biomarker Approaches in NSCLC: From Tissue Biopsy to Liquid Biopsy and Circulating Tumor Cells 7

1.2.1 Tissue Biopsy: The Gold Standard and Its Inherent Limitations 7

1.2.2 The Liquid Biopsy Spectrum: Diverse Analytes Reflecting Tumor Biology 8

1.3 Circulating Tumor Cells: Biological Properties and Technical Challenges 10

1.3.1 CTCs in the Metastatic Cascade: From Dissemination to Distant Colonization 10

1.3.2 CTC Survival Strategies: Mechanisms of Circulatory Adaptation 11

1.3.3 CTC–Immune System Interactions: Immune Surveillance and Evasion 13

1.4 Circulating Tumor Cell Enrichment: Criteria and Technologies 16

1.4.1 Criteria for CTC Isolation 16
1.4.2 Biological Capture Strategies: Antibody-Based Positive and Negative Selection .. 17
1.4.3 Physical Enrichment Methods 18
1.4.4 Combination Strategies 22
1.5 Technologies for CTC Molecular and Phenotypic Characterization 22
1.5.1 Bulk Molecular Profiling: Capabilities and Input Constraints 22
1.5.2 Single-Cell and Multi-Omic Technologies 25
1.5.3 Imaging-Based Phenotypic Profiling 28
1.5.4 Computational Analysis Pipelines 29
1.6 Clinical Applications of CTCs in NSCLC 31
1.6.1 Prognostic Stratification 31
1.6.2 Molecular Profiling: Tumor Heterogeneity and Resistance Mechanisms 31
1.6.3 Functional Characterization: Ex Vivo Culture and Drug Sensitivity 32
1.6.4 CTC-Immune Interactions 33
1.6.5 Longitudinal Monitoring 33
1.6.6 Integration and Future Perspectives 34
1.7 Hypothesis and Scope of this Thesis 34
Chapter 2 Single-cell Profiling Reveals Distinct CTC Phenotypes and Immune Cell Interactions in Patients with ALK-positive Non-Small Cell Lung Cancer 38
2.1 Abstract 38
2.2 Resulting Publication Information 38
2.3 Introduction 39
2.4 Results 43
2.4.1 Overview of patient characteristics and CTC enrichment using the Labyrinth microfluidic platform. 43
2.4.2 Characterization of CTCs in ALK NSCLC. 48

2.4.3 Single-cell annotation shows lung originated CTC identification and Heterogeneity in the blood immune microenvironment 50

2.4.4 Characterization of CTCs reveals platelet and EMT related signature and pathways in ALK NSCLC 56

2.4.5 Pathway analysis and identification of CTC heterogeneity in ALK NSCLC 57

2.4.6 Inter- and Intra-patient heterogeneity of CTCs 58

2.4.7 CTC immune cell interaction pattern further shows heterogeneity in ALK vs wild-type NSCLC patients 62

2.5 Discussion 66

2.6 Methods: 71

2.6.1 Study design and patient cohort 72

2.6.2 Cell line culture and validation 72

2.6.3 Microfluidic CTC enrichment 72

2.6.4 Immunofluorescence (IF) staining and imaging 73

2.6.5 ALK rearrangement validation 75

2.6.6 Single-cell RNA sequencing (scRNA-seq) 75

2.6.7 Bioinformatic analysis of scRNA-seq Data 76

2.6.8 Statistical analysis 78

2.7 Author Contributions 79

Chapter 3 Longitudinal Circulating Tumor Cell Dynamics and Single-Cell Transcriptomic Evolution as Prognostic Biomarkers in ALK-Positive Non-Small Cell Lung Cancer 80

3.1 Abstract 80

3.2 Resulting Publication Information 81

3.3 Introduction 81

3.4 Results 85

3.4.1 Cohort characteristics and CTC profiling 85

3.4.2 Longitudinal CTC dynamics and clinical outcomes 87

3.4.3 Prognostic value of CTC phenotypic change 89

3.4.4 Longitudinal single-cell transcriptomic profiling of CTCs 92

3.4.5 External validation of CTC-derived transcriptomic signatures 96

3.5 Discussion 99

3.6 Methods 105

3.6.1 Study design and patient cohort 105

3.6.2 Microfluidic CTC enrichment, immunofluorescence staining, and scRNA-seq library preparation 105

3.6.3 Pseudotime trajectory and copy number variation (CNV) inference 105

3.6.4 External validation and gene set variation analysis (GSVA) 106

3.6.5 Statistical analysis 106

Chapter 4 Automated Computational Pipelines for High-Throughput CTC Identification and Phenotypic Scoring 109

4.1 Abstract 109

4.2 Resulting Publication Information 109

4.3 Introduction 110

4.4 Results 114

4.4.1 A semi-automated image processing pipeline enables high-throughput CTC identification 114

4.4.2 Systematic benchmarking identifies logistic regression as the optimal classifier within the engineered feature space 115

4.4.3 A convolutional neural network with Grad-CAM interpretability demonstrates biologically grounded attention 118

4.4.4 Antigen-independent CTC identification via a deep learning–assisted biolaser platform 119

4.4.5 An LDA-based morphological score predicts NK sensitivity in patient-derived CTC models 122

4.5 Discussion 126

4.6 Methods 131
4.6.1 Cell culture 131
4.6.2 Patient sample collection and Labyrinth processing 132
4.6.3 Semi-automated image processing pipeline for CTC identification 132
4.6.4 Feature extraction and normalization 134
4.6.5 Machine learning model training and evaluation 135
4.6.6 Convolutional neural network and Grad-CAM interpretability 136
4.6.7 Biolaser-based antigen-independent CTC classification (DCLC) 137
4.6.8 EMScore derivation via Linear Discriminant Analysis 138
4.6.9 NK cell cytotoxicity live imaging 138
4.6.10 Quantitative image-based NK cytotoxicity assay 138
4.6.11 Statistical analysis 139
4.7 Author Contributions 139
Chapter 5 Conclusions and Future Directions 141
5.1 Research Summary 141
5.2 Limitations and Future Work 146
5.3 Conclusions 154
Bibliography 156

List of Tables

Table 2.1. Clinical characteristics table of ALK NSCLC and wild-type NSCLC patients (n=45, n=46) 44

Table 2.2. Clinical characteristics and single-cell CTC count of ALK and wild-type NSCLC patients. 51

List of Figures

Figure 1.1 Overview of the dissertation: characterizing CTCs in ALK NSCLC for clinical monitoring. 37

Figure 2.1 Clinical characteristics and overview experimental workflow for ALK NSCLC patients. 46

Figure 2.2 Clinical, CTCs characteristics, Labyrinth Validation for ALK and wild-type NSCLC cohorts. 47

Figure 2.3. Phenotypic characterization and comparative analysis of CTCs in ALK NSCLC. 49

Figure 2.4. Single cell visualization and identification of CTC in ALK and wild-type NSCLC cohorts. 54

Figure 2.5. Chromosomal CNV profiles and CTC CNV scores in ALK and wild-type cohorts. 55

Figure 2.6. Single cell transcriptomic profiling of CTCs in ALK NSCLC. 61

Figure 2.7. CTC-Platelet staining and significantly enriched pathways for ALK CTCs . 61

Figure 2.8. Overview of cell-cell interaction signaling dynamics in ALK NSCLC and wild type cohort 65

Figure 3.1. Study design and patient characteristics. 87

Figure 3.2. Multi-marker immunofluorescence phenotyping of CTCs and longitudinal burden tracking. 88

Figure 3.3. Longitudinal CTC subtype burden and inter-visit trajectory as prognostic determinants of survival in ALK NSCLC. 91

Figure 3.4. Integrated multi-patient scRNA-seq profiling: cell type composition, transcriptomic trajectories, copy number variation, and pseudotime analysis across longitudinal CTC samples. 95

Figure 3.5. Single-cell transcriptomic characterization and external validation of longitudinal CTC-derived signatures. 98

Figure 4.1. Semi-automated image analysis pipeline and classical machine learning classifier benchmarking. 117

Figure 4.2. Transfer-learned VGG-16 with Grad-CAM interpretability and antigen-independent CTC identification via a deep learning–assisted biolaser platform. 121

Figure 4.3. LDA-derived EMScore captures the epithelial–mesenchymal morphological continuum at the single-cell level. 125

Abstract

Non-small cell lung cancer (NSCLC) remains the leading cause of cancer-related mortality worldwide. Its molecular heterogeneity and acquired therapeutic resistance pose a challenge to clinical management. Circulating tumor cells (CTCs), malignant cells shed from tumors into the bloodstream, offer a minimally invasive strategy to capture intratumoral heterogeneity beyond the limitations of conventional tissue biopsy. Recent advances in computational biology enable high-resolution tumor profiling, yet systematic workflows optimized for CTC analysis in anaplastic lymphoma kinase-positive NSCLC (ALK NSCLC), a molecularly defined oncogenic driver subtype, remain largely unexplored.

To address this gap, I first established a comprehensive workflow for the microfluidic isolation and molecular characterization of CTCs from ALK NSCLC patients (N = 45). Patients with a smoking history or whose samples were collected at disease progression exhibited a significant shift toward an epithelial–mesenchymal transition (EMT) CTC phenotype, a cell state associated with enhanced metastatic potential and therapeutic resistance. Single-cell RNA sequencing (scRNA-seq) identified upregulation of RAS signaling, consistent with downstream activation of ALK-driven oncogenic cascades. These CTCs were further characterized by upregulation of oncogenic (TIMP1, SMOX), EMT-associated (CLU, GATA1), and platelet-interacting (ITGA2, PF4) genes, alongside a transcriptional profile suggestive of NK cell dysfunction, potentially mediated via the TGFβ pathway.

Having established the molecular profile of ALK CTCs, I explored their clinical utility

through longitudinal monitoring of 12 ALK NSCLC patients (median follow-up: 107 weeks; 2–10 serial visits per patient). Single-timepoint CTC counts did not distinguish progression from non-progression visits, but inter-visit dynamics proved informative: patients with a greater decrease in the hybrid epithelial–mesenchymal (All+) CTC fraction had significantly longer progression-free survival ($p = 0.013$), and reductions in total, CK+-only, and combined Vim+/CK+/All+ CTC burden were each associated with improved overall survival (all $p = 0.044$). Parallel scRNA-seq of CTCs from three longitudinally sampled patients revealed persistent upregulation of EMT-associated genes (e.g., CAVIN2, SPARC). A signature derived from genes downregulated across all visits was significantly associated with shorter overall survival in an independent TCGA-LUAD cohort ($n = 500$, $p = 0.039$). These findings suggest that longitudinal CTC phenotypic trajectories and transcriptomic signatures may capture clinically relevant biology that single-timepoint measurements miss.

In the final component, having explored the clinical utility of CTCs as dynamic biomarkers, I developed a semi-automated computational pipeline to improve the throughput and reproducibility of multiplex immunofluorescence-based CTC image analysis. Using single-cell fluorescence and morphological features, three classical machine learning classifiers were benchmarked, with logistic regression achieving the strongest performance (F1 = 83.7%). A transfer-learned VGG-16 model was also trained, and Gradient-weighted Class Activation Mapping confirmed that it attended to biologically relevant cellular regions rather than background artifacts. In parallel, an unsupervised model classified cells using optical lasing signatures acquired across multiple pumping powers, offering a potential alternative to fluorescent protein-based identification. Building on these strategies, I contributed to a Linear Discriminant Analysis–based model deriving a single-cell epithelial–mesenchymal (EM) score

from morphological features; the EM score correlated with Natural Killer cell–mediated cytotoxicity ($R^2 = 0.684$), suggesting that image-derived morphology may serve as a non-invasive predictor of immune sensitivity.

This dissertation presents a comprehensive single-cell characterization of CTCs in ALK NSCLC. Beyond advancing biological understanding, it demonstrates the potential of CTCs as scalable biomarkers for disease monitoring and therapeutic decision-making, establishing a foundation for a liquid biopsy pipeline extendable to additional cancer types.

Chapter 1 Introduction

1.1 Non-Small Cell Lung Cancer: Clinical Challenges and the Imperative for Longitudinal Molecular Monitoring

1.1.1 Clinical aspect of NSCLC

Lung cancer remains the leading cause of cancer-related mortality worldwide, with approximately 2.5 million new diagnoses annually [1]. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases, representing the predominant histological subtype and a major focus of clinical and translational research [2]. NSCLC encompasses molecularly distinct subtypes that share anatomical origin but differ in biology, clinical behavior, and therapeutic vulnerabilities [2–4].

1.1.2 NSCLC Treatment Modalities and Unmet Challenges

The therapeutic management of NSCLC encompasses a multimodal approach integrating surgical resection, radiation therapy, cytotoxic chemotherapy, immune checkpoint inhibitors, and molecularly targeted therapy. Treatment selection is guided by disease stage, molecular profile, histological subtype, and patient performance status. Each modality carries distinct clinical benefits alongside inherent limitations.

Surgical resection achieves the most favorable outcomes in early-stage disease (Stage I and II), with 5-year survival rates approaching 70-90% [5]. However, surgical resection is only feasible for a minority of patients, as the majority present with locally advanced or metastatic disease at diagnosis. Radiation therapy and cytotoxic chemotherapy, frequently administered

concurrently in locally advanced disease, provide locoregional control and systemic disease suppression. They remain the key treatment in NSCLC lacking actionable mutations. However, both modalities lack tumor selectivity, resulting in damage to surrounding normal tissues and broadly proliferating healthy cells, which compromise treatment tolerability and quality of life.

Immune checkpoint inhibitors, particularly agents targeting the PD-1/PD-L1 axis, have achieved durable responses and improved long-term survival in a subset of NSCLC patients [6]. Even so, response rates remain variable and are influenced by tumor mutational burden, PD-L1 expression, and the composition of the tumor immune microenvironment (TME) [6,7]. Molecularly targeted therapies have similarly transformed outcomes in NSCLC with specific mutation. Their clinical benefit, however, is restricted to molecularly defined patient subsets and is often limited by the development of acquired resistance [8].

Despite therapeutic advances over the past decade, patients presenting with locally advanced or metastatic disease (Stage III and IV) face 5-year survival rate of approximately 10% or lower [9].

1.1.3 Molecular Heterogeneity and Oncogenic Driver Alterations in NSCLC

These outcomes reflect the pronounced heterogeneity of NSCLC, which spans multiple biological scales, ranging from diversity within the TME to genetic and phenotypic variation across patients and within individual tumors.

One important source of complexity arises from heterogeneity within the TME. Tumor cells interact dynamically with surrounding stromal cells, immune populations, and extracellular matrix components. These interactions influence tumor growth, immune evasion, and therapeutic sensitivity, thereby shaping tumor evolution and clinical outcomes [10].

Beyond the TME, heterogeneity is also intrinsic to tumor cells of NSCLC itself. At the

inter-patient level, NSCLC exhibits substantial diversity in genetic alterations, oncogenic drivers, and clinical behaviors across different individuals. Such variability contributes to the wide range of disease trajectories and therapeutic responses observed among patients. Heterogeneity is observed not only between patients but also within individual patients (intra-patient heterogeneity). Multiple primary tumors or metastatic lesions can harbor distinct genomic and transcriptional profiles, reflecting divergent evolutionary trajectories among spatially separated tumor sites. Within each tumor, intra-tumoral heterogeneity further contributes to complexity: tumors often consist of genetically and phenotypically distinct subclonal populations arising through ongoing clonal evolution. These subclones may carry different mutations, copy number alterations, and signaling activities, enabling selective expansion of resistant populations under therapeutic pressure. Together, intra-patient and intra-tumoral heterogeneity can lead to heterogeneous responses to therapy and contribute to disease progression and drug resistance [11,12].

One of the forms of heterogeneity in tumor cells is phenotypic plasticity. A major example is epithelial-to-mesenchymal transition (EMT), a dynamic cellular program through which tumor cells shift from epithelial states toward hybrid epithelial–mesenchymal or mesenchymal states. At the molecular level, EMT involves downregulation of epithelial markers, including E-cadherin and epithelial cell adhesion molecule (EpCAM), and upregulation of mesenchymal markers such as vimentin and N-cadherin. EMT-associated states have been linked to enhanced invasion, immune evasion, and therapy resistance [13,14]

Beyond phenotypic plasticity, oncogenic driver alterations represent a second major axis of heterogeneity in NSCLC. NSCLC encompasses several histological subtypes, of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most prevalent and

the most extensively characterized at the molecular level; these two subtypes display substantial divergence in their driver landscapes [15]. Comprehensive genomic profiling studies, including large-scale initiatives such as The Cancer Genome Atlas (TCGA), have catalogued a broad spectrum of actionable driver mutations across NSCLC. These include activating mutations in EGFR (~15-40% of cases) and KRAS (~25–30% of cases), as well as tumor suppressor alterations such as TP53 mutations (50.8% in lung adenocarcinomas) that contribute to genomic instability and aggressive tumor behavior. In addition, chromosomal rearrangements involving ROS1 (~1–2%) and ALK (~3–7%) generate constitutively active fusion proteins that drive oncogenic signaling [16,17].

1.1.4 Precision Medicine Approaches and Therapeutic Resistance in ALK NSCLC

Despite its relatively modest prevalence, Anaplastic Lymphoma Kinase-positive Non-Small Cell Lung Cancer (ALK NSCLC) has emerged as a compelling model for precision oncology treatment strategies. This subtype is majority lung adenocarcinoma, characterized by diverse fusion partners, complex resistance mechanisms, and dynamic evolutionary changes that occur under targeted therapeutic pressure [18,19].

ALK rearrangements, most commonly the EML4–ALK fusion in NSCLC, arise from genomic structural alterations and generate fusion transcripts that are translated into oncogenic ALK fusion proteins. The resulting fusion protein retains the intracellular ALK kinase domain and is aberrantly expressed under the regulatory control of the fusion partner, leading to ligand-independent and constitutive activation of downstream signaling pathways. Through these pathways, PI3K/AKT, MAPK/ERK, and JAK/STAT cascades, ALK signaling promotes cell proliferation, survival, invasion, and metastatic dissemination. The activation across these pathways can vary between patients and even between distinct tumor sites within an individual,

contributing to biological heterogeneity and differential treatment responses. This signaling complexity is further modulated by the nature of the fusion partner and co-occurring genomic alterations context, collectively shaping both baseline oncogenic dependence on ALK and the spectrum of potential escape routes under targeted therapeutic pressure [18,19].

Encouragingly, normal adult lung parenchyma generally lacks detectable ALK protein expression, whereas ALK protein expression in lung cancer is usually driven by ALK gene rearrangements. This tumor-associated expression pattern provides the biological rationale for using ALK as both a diagnostic marker and a therapeutic target. The development of successive generations of ALK tyrosine kinase inhibitors (TKIs), including crizotinib, ceritinib, alectinib, brigatinib, and lorlatinib, substantially changed the clinical management of ALK NSCLC, with modern first-line regimens now achieving median progression-free survival exceeding 30 months in selected cohorts [20–22]. Despite these advances, acquired resistance to ALK TKIs remains almost universal. Mechanistically, resistance can be broadly grouped into three non-mutually exclusive categories. On-target resistance is mediated by secondary kinase domain mutations, such as the solvent-front substitution G1202R, that diminish inhibitor binding while maintaining constitutive kinase activity [23]. Bypass resistance arises through amplification or activation of alternative receptor tyrosine kinases, including MET, EGFR, and KIT, which reactivate downstream PI3K/AKT and MAPK signaling despite continued ALK inhibition. A third category involves phenotypic and lineage plasticity, whereby tumor cells undergo profound shifts in cellular identity that uncouple survival from ALK signaling. This can manifest as histological transformation to small cell lung cancer, acquisition of neuroendocrine features, or EMT, each associated with altered transcriptional programs, surface marker expression, and drug sensitivities [24–28].

Critically, these resistance mechanisms rarely occur in isolation. Individual patients frequently harbor multiple concurrent or sequential resistance processes, distributed across distinct anatomical sites or emerging over time as clonal selection and evolution proceed under the selective pressure of targeted therapy. The resulting spatial and temporal heterogeneity means that a single-timepoint, single-site tissue biopsy can only ever provide a partial snapshot of the underlying resistance biology. In many cases, it may fail to capture clinically actionable subclones.

Moreover, ALK detection is further complicated by the fact that ALK alterations can manifest at the DNA, RNA, and protein levels, and the assays used to interrogate each of these molecular layers are not always fully concordant. Some tumors exhibit genomic evidence of ALK rearrangement without detectable transcript or protein expression, while others show transcript- or protein-level evidence without clear genomic confirmation [29]. Such discordances may reflect non-expressed rearrangements, cryptic or complex structural variants, intratumoral heterogeneity, limited tumor material, variable assay sensitivity across molecular layers, or technical background that obscures true ALK signal.

1.1.5 The Imperative for Longitudinal Molecular Monitoring

Collectively, the molecular heterogeneity, aggressive clinical course, and rapid resistance make ALK NSCLC one of the most challenging lung cancer subtypes. Those features above motivate an unmet need for minimally invasive monitoring strategies capable of capturing tumor heterogeneity, tracking clonal evolution in real time, and detecting emerging resistance mechanisms to guide therapeutic decision-making. Such approaches also need to be analytically sensitive enough to detect low-frequency alterations, sufficiently comprehensive to span relevant resistance mechanisms, and feasible for repeated sampling without significant patient burden.

1.2 Biomarker Approaches in NSCLC: From Tissue Biopsy to Liquid Biopsy and Circulating Tumor Cells

1.2.1 Tissue Biopsy: The Gold Standard and Its Inherent Limitations

Given the molecular and spatial heterogeneity described in Section 1.1, reliable profiling of tumor material is essential for identifying actionable drivers and guiding treatment decisions in NSCLC. Currently, tissue biopsy remains the clinical gold standard for obtaining tumor material.

Tissue biopsies are typically obtained by bronchoscopy for centrally located lesions, percutaneous transthoracic needle aspiration for peripheral tumors, or surgical resection in operable cases [30]. These methods can yield sufficient material for morphological assessment and broad genomic testing.

However, tissue biopsy has several well-recognized limitations. Firstly, the procedure is invasive and difficult to repeat over time, which limits its use for longitudinal monitoring. Transthoracic needle biopsies carry procedural risks such as pneumothorax (reported in roughly 15–25% of cases), hemorrhage, infection, and the possibility of tumor seeding along the needle tract [31]. Moreover, a single-site biopsy captures only a portion of the overall mutational landscape. Multi-region sequencing studies have shown that many mutations and subclones are confined to specific tumor regions and may be missed entirely when only one site is sampled. This applies both within the primary tumor and across metastatic sites. For instance, multiregional whole-exome sequencing of early-stage NSCLC has revealed substantial spatial genomic heterogeneity across different regions of the same tumor. While trunk driver alterations such as TP53 or KRAS mutations tend to be shared across regions, other cancer-relevant alterations, including those in PIK3CA, NF1, and ARID1A, may appear only in specific

subclonal populations [32,33].

These limitations affect both the representativeness and the repeatability of tissue-based profiling, motivating the development of less invasive, repeatable approaches that better capture tumor heterogeneity and track molecular changes over time.

1.2.2 The Liquid Biopsy Spectrum: Diverse Analytes Reflecting Tumor Biology

Liquid biopsy addresses these gaps by sampling tumor-derived material circulating in blood or other bodily fluids through a minimally invasive draw [34]. Because circulating biomarkers originate from both primary and metastatic sites, liquid biopsy provides a composite readout reflecting the molecular heterogeneity of all spatially separated disease sites, that tissue biopsy may not offered. The principal liquid biopsy analytes in NSCLC are circulating tumor DNA (ctDNA), extracellular vesicles (EVs), and circulating tumor cells (CTCs), each offering distinct biological information and analytical characteristics [35–37].

1.2.2.1 Circulating Tumor DNA

ctDNA consists of short fragments of tumor-derived DNA, typically 150–200 base pairs in length corresponding to mononucleosome protection. It is released into the bloodstream through apoptosis, necrosis, or active secretion from tumor cells[38] . It makes up only a small fraction of total cell-free DNA in circulation, usually between 0.01% and 10% depending on tumor burden, with the rest coming predominantly from hematopoietic cells [39]. Because of this low abundance, detecting and characterizing ctDNA requires highly sensitive methods such as droplet digital PCR and next-generation sequencing platforms designed to distinguish rare tumor-derived fragments from the much larger pool of normal cell-free DNA.

Among liquid biopsy approaches, ctDNA analysis is the most clinically mature in NSCLC. Several FDA-approved platforms are available, including Guardant360 CDx and

FoundationOne Liquid CDx, and ctDNA is routinely used to detect actionable mutations for treatment selection [40,41]. ctDNA, however, provides only genomic information and does not capture transcriptional programs, protein expression, or functional cellular phenotypes that influence treatment response.

1.2.2.2 Extracellular vesicles

EVs are membrane-enclosed particles, typically 30–150 nm in diameter, that are actively secreted by cells into the circulation. Tumor-derived EVs carry molecular cargo that reflects the cell of origin, including nucleic acids, proteins, lipids, and metabolites [42].

There is growing evidence that tumor-derived EVs play functional roles in cancer progression, including mediating intercellular communication, preparing pre-metastatic niches, and suppressing immune responses. Analysis of EV cargo has shown potential for cancer detection, molecular profiling, and studying tumor–microenvironment interactions. However, most EV-based assays remain at an investigational stage, with limited clinical validation in NSCLC. Key challenges include the difficulty of isolating tumor-specific EVs from the large background of vesicles released by normal cells, variability across isolation and analytical methods, and a lack of standardization in both pre-analytical and analytical workflows [42,43].

1.2.2.3 Circulating tumor cells

CTCs are intact tumor cells, typically 15–30 μm in size, that have shed from primary or metastatic sites and entered the bloodstream [44]. They are generally larger and more rigid than most blood cells and retain key features of the tumor they originated from, including genomic DNA, RNA transcripts, proteins, organelles, and cellular morphology.

Because CTCs are whole cells, they offer a unique opportunity to study tumor biology at the single-cell level. They preserve both genetic information and cellular phenotype, including

markers of EMT, lineage plasticity, and proliferative state. This makes CTCs particularly relevant for longitudinal monitoring of NSCLC, especially in molecularly defined subtypes such as ALK disease, where resistance mechanisms are heterogeneous and often linked to shifts in cellular identity and signaling dependencies. In some settings, CTCs can also be expanded ex vivo for drug-sensitivity testing, providing a potential functional link between molecular alterations and therapeutic response [44,45].

However, CTCs are extremely rare, typical ranges from 1 to 1,000 CTCs per milliliter of blood, against a background of approximately 5 billion normal blood cells per milliliter [248]. This extreme rarity requires specialized enrichment and detection technologies to isolate CTCs from the vast excess of erythrocytes and leukocytes [45].

1.3 Circulating Tumor Cells: Biological Properties and Technical Challenges

1.3.1 CTCs in the Metastatic Cascade: From Dissemination to Distant Colonization

CTCs are often described as the seeds of metastasis. Their presence in the bloodstream reflects a series of sequential steps collectively known as the metastatic cascade: local invasion through the basement membrane and surrounding stroma, intravasation into blood or lymphatic vessels, survival in the circulation, extravasation at a distant site, and colonization of a new organ [46].

A central feature of this cascade is the dynamic shift in cellular phenotype that CTCs undergo at different stages. During local invasion and intravasation, tumor cells frequently activate EMT, a reversible developmental program in which epithelial cells lose cell–cell adhesion and apical–basal polarity while acquiring mesenchymal traits such as enhanced motility and invasiveness [47]. Importantly, EMT does not operate as a binary switch; rather, tumor cells

frequently occupy a spectrum of hybrid epithelial–mesenchymal states, and this plasticity enables adaptation to change microenvironmental conditions and therapeutic pressures [48]. In NSCLC, immunofluorescence-based analyses of CTCs have identified substantial subpopulations co-expressing epithelial markers such as EpCAM alongside mesenchymal markers such as vimentin, which promotes invasiveness and link to a more aggressive phenotype [49]. Beyond promoting invasiveness, EMT has been associated with upregulation of stemness markers (e.g., CD44, ALDH1), enhanced resistance to apoptosis, reduced drug sensitivity, and greater metastatic potential [13].

Once CTCs reach a distant site and extravasate, successful colonization generally requires the reverse process, mesenchymal–epithelial transition (MET), in which cells re-acquire epithelial characteristics that support proliferation and tissue integration at the secondary site [44].

1.3.2 CTC Survival Strategies: Mechanisms of Circulatory Adaptation

The bloodstream is a hostile environment for tumor cells. CTCs are exposed to hemodynamic shear stress, loss of matrix attachment (which normally triggers anoikis), and active immune surveillance[50]. Despite this, a subset of CTCs survives long enough to seed distant metastases. This survival depends on several complementary adaptive mechanisms, which can vary among individual CTCs within the same patient.

1.3.2.1 Platelet Shielding

One of survival strategies involves the association of CTCs with platelets. Upon entering the bloodstream, CTCs rapidly recruit platelets through multiple receptor–ligand interactions, including GPIb-IX-V binding to tumor cell mucins, P-selectin–mediated adhesion, and integrin-mediated contacts. Individual CTCs can become surrounded by dozens to hundreds of platelets,

forming a physical and immunological barrier [51,52].

This platelet coating confers several advantages. Physically, it buffers CTCs against shear stress-induced damage. Activated platelets deposit fibrin networks around CTCs, forming microthrombi that provide additional mechanical protection and promote arrest at distant capillary beds [53]. Immunologically, platelet-derived MHC class I molecules can mask CTCs from natural killer (NK) cell recognition, while platelet-secreted immunomodulatory factors suppress NK cell activation [54]. Moreover, upon activation, platelets release growth factors (PDGF, TGF-β, VEGF), chemokines, and pro-coagulant factors from their granules. For example, TGF-β has been shown to activate NF-κB signaling in tumor cells, thereby inducing EMT and enhancing invasive potential [55], while VEGF promotes angiogenesis at metastatic sites [56]. Studies have demonstrated that platelet depletion or inhibition of platelet–tumor cell interactions substantially reduce metastatic efficiency in preclinical models [57].

1.3.2.2 CTC Cluster Formation

Another survival strategy involves the formation of CTC clusters, cohesive aggregates of 2 to 50 or more CTC together. These CTC clusters are rare, typically accounting for only 3–5% of all CTCs in patient blood, yet they exhibit disproportionately high metastatic potential [58]

Cluster formation promotes CTC survival through several mechanisms. Cell–cell adhesions mediated by E-cadherin, N-cadherin, and other adhesion molecules provide survival signals that substitute for the matrix-derived cues normally lost during circulation. Cells within clusters also maintain intercellular junctions that provide mechanical support, allow sharing of metabolic resources, and enable exchange of survival factors through gap junctions. Furthermore, cells in the cluster interior are physically shielded from NK cell contact and shear forces, while peripheral cells contribute to immune evasion through contact-dependent

mechanisms [59,60]. In breast cancer mouse models and patient-derived samples, Aceto et al. demonstrated that CTC clusters are oligoclonal in origin and exhibit 23- to 50-fold greater metastatic capacity than single CTCs [61].

1.3.2.3 Cytoskeletal Remodeling

CTCs also adapt to the mechanical forces of the circulation through rapid cytoskeletal reorganization. Exposure to shear stress triggers the formation of cortical actin networks that reinforce the cell membrane and improve mechanical resilience [62]. Compared with adherent epithelial cells residing in stable tissue environments, CTCs show elevated expression of intermediate filament proteins, including vimentin and cytokeratins, which provide structural support against deformation in the bloodstream [63]. Mechanotransduction pathways such as RhoA/ROCK signaling are activated in response to shear stress, promoting both cytoskeletal remodeling and anti-apoptotic signaling [64].

1.3.3 CTC–Immune System Interactions: Immune Surveillance and Evasion

Even after withstanding physical stresses, CTCs must also contend with the immune system. CTCs deploy a range of evasion strategies that allow a subset to escape immune destruction. The balance between immune surveillance and immune evasion shapes CTC survival and has implications for response to immunotherapy [60]. Multiple immune populations participate in this balance: natural killer (NK) cells, $CD8^+$ and $CD4^+$ T lymphocytes, B cells and antibody-mediated effectors, neutrophils (including those forming neutrophil extracellular traps, or NETs), monocytes and macrophages, dendritic cells, and myeloid-derived suppressor cells (MDSCs). Each population can exert anti-tumor or pro-tumor effects depending on phenotype, microenvironmental context, and inflammatory state [65–69].

The following section focuses on NK-cell-mediated surveillance, which is most relevant

to my thesis.

1.3.3.1 NK Cell–Mediated Killing and Evasion

NK cells represent the primary innate immune barrier to CTC survival in the bloodstream. As innate lymphoid cells, NK cells can detect and kill target cells without prior antigen sensitization [70]. A key mechanism underlying this function is the ability of NK cells to identify and eliminate cells that lack or downregulate MHC class I molecules, a feature commonly observed in tumor cells [71]. The importance of NK cell surveillance has been demonstrated experimentally: NK cell depletion increases metastatic burden [72,73], and higher NK cell activity has been associated with lower CTC counts and better patient outcomes [65].

NK cell activity is governed by the balance between activating and inhibitory receptor signals on the NK cell surface, which recognize corresponding ligands on the CTC surface. CTCs employ several strategies to evade NK cell–mediated killing. Firstly, CTCs can downregulate or proteolytically shed surface ligands for activating receptors, such as MICA and MICB. The shed soluble forms of these ligands can further act as decoys that bind NKG2D on NK cells, leading to receptor internalization and functional desensitization [74,75]. Secondly, CTCs may upregulate non-classical MHC class I molecules. In particular, HLA-E can engage the inhibitory CD94/NKG2A receptor on NK cells, suppressing NK activation even when activating ligands are present. HLA-G may further contribute to immune evasion through inhibitory receptors such as ILT2/ILT4, KIR2DL4, and, as recently reported, NKG2A/CD94 [76,77]. Thirdly, as described in Section 1.3.2.1, platelet shielding provides an additional layer of NK cell evasion; platelets physically coat CTCs, blocking NK cell access, and release immunosuppressive factors such as TGF-β that directly impair NK cell cytotoxicity [78]. Last but not least, in contrast to these evasion mechanisms, EMT can paradoxically increase CTC

susceptibility to NK cell killing. Time-course transcriptomic profiling of TGF-β–induced EMT shows progressive upregulation of NK-activating ligands—including PVR, CADM1, ULBP2, and ULBP4, alongside loss of E-cadherin, a ligand for the inhibitory NK receptor KLRG1 [79]. Functionally, EMT-shifted cells exhibit enhanced NK-mediated cytotoxicity, an effect attributed primarily to CADM1 upregulation and E-cadherin loss; in vivo, NK cell depletion permits spontaneous metastasis without altering primary tumor growth, and higher CADM1 expression correlates with improved survival in lung and breast adenocarcinoma cohorts [79,80]. NK cells therefore constitute a metastasis-specific surveillance barrier that selectively targets EMT-shifted disseminating cells.

1.3.3.2 Other Immune-Cell Interactions

Although not the focus of the present work, other immune populations also shape CTC survival in circulation. Neutrophils can be either anti-tumor (N1, mediating reactive oxygen species production and antibody-dependent cytotoxicity) or pro-tumor (N2, supporting immunosuppression and angiogenesis). Pro-tumor neutrophils additionally generate neutrophil extracellular traps (NETs) that promote CTC extravasation, endothelial adhesion, local immune exclusion, and form heterotypic clusters with CTCs that increase metastatic potential [81–85]. Elevated neutrophil-to-lymphocyte ratio (NLR) correlates with worse prognosis in NSCLC [86,87]. $CD8^+$ T cells can recognize and kill tumor cells through MHC class I–restricted antigen presentation; however, MHC class I downregulation on tumor cells, together with the short intravascular residence time and rarity of CTCs, may limit sustained T-cell–mediated clearance [88,89]. Circulating monocytes and macrophage-like cells can associate with CTCs or tumor-macrophage fusion-like circulating cells, and macrophage-derived cytokines such as IL-6, IL-10, and TGF-β may support tumor-cell survival, immunosuppression, and metastatic niche formation

[90–95]. Beyond the cell-type-specific evasion strategies described above, CTCs also exploit immune checkpoint signaling as a broader immunosuppressive mechanism. CTCs from NSCLC patients frequently express PD-L1, and longitudinal increases in PD-L1$^+$ CTCs have been associated with resistance to PD-1/PD-L1 inhibitors [96]. PD-L1 can engage PD-1 on cytotoxic lymphocytes, including T cells and PD-1$^+$ NK cells, suppressing effector function; in NSCLC models, PD-L1/PD-1 blockade enhances NK-cell cytotoxicity through granzyme B secretion [96,97]. PD-L1 expression can also be dynamically induced by interferon signaling, particularly through the IFNγ–JAK/STAT–IRF1 axis, representing an adaptive immune resistance mechanism in which tumor cells counter immune pressure by upregulating inhibitory checkpoint ligands [98–100].

1.4 Circulating Tumor Cell Enrichment: Criteria and Technologies

Molecular and functional analysis of CTCs requires their prior isolation from peripheral blood.

1.4.1 Criteria for CTC Isolation

CTC isolation is technically demanding because as discussed above CTCs are extremely rare, phenotypically heterogeneous, and may circulate either as single cells or multicellular clusters.

Several parameters are important when evaluating CTC isolation platforms. The quality of recovered CTCs must meet the requirements of intended downstream application. Processing time and sensitivity to pre-analytical variables such as temperature and anticoagulant choice affect the cells' quality [101,102]. Beyond quality of cells, capture efficiency is also important. Recovery efficiency refers to the proportion of CTCs in a sample that are successfully captured, typically ranges from 40–90% depending on the platform and sample characteristics [103–105].

Purity refers to the degree of leukocyte depletion, which directly affects the sensitivity of downstream analyses. Different downstream applications impose distinct requirements: single-cell sequencing requires high cell viability and intact nucleic acids. Immunofluorescence imaging requires compatibility with fixation and staining protocols. Ex vivo culture requires viable, undamaged cells. These different requirements mean that no single isolation method is optimal for all applications.

1.4.2 Biological Capture Strategies: Antibody-Based Positive and Negative Selection

Immunoaffinity-based isolation exploits differential expression of cell surface markers to capture tumor cells. These approaches fall into two categories: positive selection and negative selection.

1.4.2.1 Positive Selection

Positive selection strategies capture CTCs through predefined surface antigens and often achieve higher purity, but their recovery depends on whether the target antigen is expressed on the relevant CTC population. EpCAM is used because it is an epithelial/carcinoma-associated surface molecule and is absent from most hematopoietic populations, forming the biological basis of EpCAM-based CTC enrichment [106–108].

The CellSearch system uses EpCAM-based immunomagnetic enrichment followed by cytokeratin/CD45/DAPI identification and has been clinically validated for CTC enumeration from 7.5 mL blood samples [108,109].

However, as discussed in Section 1.3.1, CTCs undergoing EMT downregulate epithelial markers including EpCAM, so EpCAM-dependent capture can miss mesenchymal, hybrid epithelial–mesenchymal, and stem-like populations [110–112]. Comparative studies confirm that EpCAM-based enrichment underestimates total CTC burden relative to EpCAM-independent

methods [112]. Using alternative surface markers does not fully solve this problem, because any antigen-dependent approach is constrained by assumptions about which markers CTCs express. Tumor cells that have lost all targeted markers, or that express them below detection thresholds, will evade capture regardless of the technology used.

1.4.2.2 Negative Selection

Negative selection enriches CTCs indirectly by depleting hematopoietic cells, thereby reducing dependence on tumor-cell surface-marker expression [113–115].

Because negative selection does not require EpCAM or another predefined tumor antigen, it can better preserve marker-low, mesenchymal, or stem-like CTC populations, but this advantage comes at the cost of residual leukocyte contamination [104,115,116].

One example is the RosetteSep system (STEMCELL Technologies), which uses antibody-mediated crosslinking of leukocytes to erythrocytes to form immunorosettes that sediment during density gradient centrifugation, leaving CTCs in the supernatant. This approach achieves moderate leukocyte depletion with high tumor cell recovery. However, the main trade-off of negative selection is lower purity. Robust identification methods are therefore essential, and bulk molecular profiling can be complicated when residual leukocyte signal overwhelms the CTC-derived signal.

1.4.3 Physical Enrichment Methods

Physical CTC enrichment strategies exploit intrinsic biophysical differences between tumor cells and hematologic cells, including size, density, deformability, dielectric properties, and acoustic mobility [117]. It enables marker-independent enrichment of phenotypically heterogeneous CTC populations.

1.4.3.1 Size-Based Filtration

Size-based separation is the most widely used physical enrichment strategy. It relies on the observation that CTCs are typically larger (15–30 μm) than most leukocytes (7–15 μm) and erythrocytes (6–8 μm) [118].

Membrane microfiltration uses porous membranes with defined pore sizes (typically 7–8 μm) to retain cells above the size threshold while allowing smaller cells to pass through. The ISET platform (Isolation by Size of Epithelial Tumor Cells), for example, uses track-etched polycarbonate membranes with 8 μm cylindrical pores. Diluted blood is filtered under controlled pressure, and cells larger than 8 μm are retained on the membrane for subsequent immunocytochemical staining and identification [119–121]. Parsortix uses a tapered separation cassette ending in a ~10 μm critical gap, exploiting both cell size and compressibility/deformability. Study reported 44–60% tumor-cell capture with ~1000 contaminating cells in the harvest and noted the core limitation that leukocytes and tumor cells overlap in size [122].

Limitations of membrane filtration include pore clogging by blood components, particularly in samples with high leukocyte counts, which reduces throughput and can damage cells due to increased pressure [123]. Deformable CTCs may also squeeze through pores smaller than their nominal diameter, while rigid leukocytes may be retained, compromising both recovery and purity. Design modifications such as size- and compressibility-based microfluidic cassettes, exemplified by the Parsortix system, have been developed to improve recovery of viable CTCs while reducing dependence on epithelial markers [124].

1.4.3.2 Microfluidic Size-Based Separation

Microfluidic approaches use engineered channel geometries to achieve size-selective

separation in continuous flow [125,126]. Deterministic lateral displacement (DLD) uses arrays of microposts arranged so that particles above a critical diameter are displaced laterally while smaller particles pass straight through [127]. The CTC-iChip, for instance, uses DLD to separate nucleated cells (including CTCs and leukocytes) from erythrocytes and platelets, removing >99% of erythrocytes while maintaining high nucleated cell recovery [128,129].

Inertial microfluidics exploits size-dependent inertial forces in curved or spiral microchannels [129]. Larger cells experience stronger inertial lift forces and migrate to distinct equilibrium positions, enabling separation from smaller cells. The Vortex platform, for example, uses expansion–contraction microchannels to generate laminar microvortices and inertial focusing, allowing larger CTCs to be trapped while smaller blood cells pass through. It is reported 7.5 mL whole-blood processing in ~20 min, 57–94% purity, and 23–317 CTCs/7.5 mL in lung cancer samples [130]. Another example is Labyrinth, which is a label-free inertial microfluidic device that uses long loops and sharp corners to differentially focus CTCs and leukocytes at high throughput. The Labyrinth achieves a throughput of 2.5 mL/min, >90% yield, and approximately 600 WBCs/mL purity [131].

1.4.3.3 Density-Based Separation

Density gradient centrifugation separates cells based on buoyant density, commonly using Ficoll-Paque or Percoll gradients. CTCs and mononuclear cells typically equilibrate at the plasma–Ficoll interface (density ~1.077 g/mL), while granulocytes and erythrocytes pellet below [126].

However, there is substantial overlap between the density distributions of tumor cells and leukocytes, which limits the discriminatory power of density-based separation when used alone. Density gradient centrifugation is therefore most commonly employed as a pre-enrichment step

to remove erythrocytes and granulocytes before additional processing.

1.4.3.4 Deformability-Based Separation

Cancer cells frequently exhibit altered mechanical properties compared to normal cells, and some microfluidic devices exploit this by using arrays of constrictions or pillar gaps to separate cells based on their ability to deform and pass through narrow spaces [132,133]. However, the mechanical properties of CTCs are heterogeneous, which limits the reliability of deformability as a universal separation criterion.

1.4.3.5 Dielectrophoresis (DEP)

DEP uses non-uniform electric fields to exert forces on cells based on their electrical polarizability, which differs between tumor cells and leukocytes due to variations in membrane composition, cytoplasmic conductivity, and nuclear-to-cytoplasmic ratio [134,135]. DEP offers label-free enrichment without size assumptions, potentially capturing diverse CTC populations. However, the dielectric properties of CTCs are not always distinct from those of leukocytes, and optimal separation parameters may vary across tumor types and patients.

1.4.3.6 Acoustophoresis

Acoustophoresis uses ultrasonic standing waves to exert acoustic radiation forces on cells, with force magnitude depending on cell size, density, and compressibility [136,137]. In microfluidic channels with acoustic resonators, cells migrate to pressure nodes or antinodes based on their acoustic contrast factor. Larger CTCs typically experience stronger acoustic forces than smaller leukocytes, enabling size-based separation. Acoustic separation is gentle and contactless, causing minimal cellular damage, which makes it attractive for isolating viable CTCs for downstream culture applications. However, throughput is generally lower than that of inertial microfluidic platforms, and the requirement for specialized acoustic resonator fabrication

adds technical complexity [137,138]. Additionally, acoustic forces scale with cell volume, meaning that smaller CTCs or those with atypical biophysical properties may not be reliably separated from contaminating leukocytes.

1.4.4 Combination Strategies

Some workflows combine positive selection, negative selection, and physical enrichment to achieve both high recovery and adequate purity [128]. However, these multi-step approaches increase technical complexity and processing time, and their advantage over single-method workflows for isolating heterogeneous CTC populations remains to be clearly established [126].

1.5 Technologies for CTC Molecular and Phenotypic Characterization

The biological insights and clinical utility of CTCs are ultimately determined by the depth and comprehensiveness of downstream molecular and phenotypic characterization [139–141]. This section first surveys bulk profiling technologies organized by molecular layer (DNA, RNA, and protein) in Section 1.5.1. Then turns to single-cell and low-input technologies that resolve heterogeneity within rare CTC populations in Section 1.5.2. Imaging-based per-cell phenotyping is introduced within the single-cell section and developed in greater detail in Section 1.5.3. Computational pipelines required for these data are summarized in Section 1.5.4

1.5.1 Bulk Molecular Profiling: Capabilities and Input Constraints

Bulk profiling aggregates signals from pooled cell populations to yield average molecular characteristics. Although bulk averaging masks intra-population heterogeneity, it provides high analytical sensitivity for low-abundance events, established clinical validation for cohort-level comparisons, and simplified data interpretation [142].

1.5.1.1 DNA based profiling

DNA-based bulk methods interrogate genomic alterations, copy-number variation, and

DNA modifications. Quantitative PCR (qPCR) provides sensitive targeted detection of specific nucleic acid sequences, using fluorescent reporters to monitor real-time amplification [143,144]. qPCR sensitivity is assay-dependent and should be interpreted using validated LoD/LoQ and MIQE-compliant reporting standards [145]. In CTC research, qPCR is frequently employed to detect oncogenic mutations (e.g., EGFR L858R), quantify EMT-associated transcripts, and verify CTC identity. Droplet digital PCR (ddPCR), although nominally a bulk-input technique, achieves single-molecule resolution and is therefore discussed in Section 1.5.2.3 alongside other single-cell-resolution methods [146]. Both qPCR and ddPCR are constrained by their targeted nature and require a priori knowledge of target sequences.

Bulk genomic sequencing comprises both targeted and whole approaches, each with distinct trade-offs in coverage and input requirements [147,148]. Targeted sequencing panels — covering predefined sets of NSCLC driver and resistance genes such as EGFR, ALK, KRAS, BRAF, MET, and TP53, provide high sequencing depth at specific loci with relatively modest input requirements, supporting sensitive detection of known variants from limited CTC material. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) enable discovery-mode interrogation of mutations and copy-number alterations across all coding regions or the entire genome [147,148]. Bulk WES/WGS usually requires more input than rare CTC samples can provide without pooling or whole-genome amplification [149].

1.5.1.2 Epigenomic Profiling

Epigenomic methods investigate on regulatory features of the genome, chromatin accessibility and DNA-modification, without altering the underlying sequence [150].

The Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) uses hyperactive Tn5 transposase to map open chromatin regions genome-wide; standard protocols

require 500–50,000 cells, while optimized low-input methods reduce the threshold to 500–5,000 cells [151]. Cancer ATAC-seq studies reveal tumor- and subtype-specific regulatory landscapes [152]. DNA methylation profiling provides orthogonal epigenetic data and can be derived from fixed cells or degraded DNA. These approaches have revealed aberrant promoter methylation of tumor suppressors and epigenetic heterogeneity in CTCs, although technical complexity continues to limit broad adoption [153].

1.5.1.3 Bulk RNA Sequencing

Bulk RNA sequencing (RNA-seq) enables transcriptome-wide profiling and provides higher per-transcript sequencing depth than single-cell approaches [154]. This depth supports detection of lowly expressed transcripts and isoform-level resolution, at the cost of resolving rare subpopulations whose distinct signatures may be obscured by averaging [155,156].

Application of bulk RNA-seq to CTCs is constrained primarily by input requirements [142]. Conventional bulk RNA-seq generally requires more RNA input than is available from rare CTC samples. For rare CTC populations, high-sensitivity low-input scRNA-seq protocols are useful because they can profile manually isolated or FACS-sorted rare cells. Template-switching methods such as Smart-seq2/Smart-seq3 provide full-length transcript coverage, with Smart-seq3 adding 5′ UMI-based RNA counting for improved quantification [157]. Despite these constraints, bulk CTC transcriptomics has identified therapeutic resistance signatures and demonstrated that CTCs represent a biologically distinct population diverging from matched primary tumors [25].

1.5.1.4 Protein-Based Profiling

Bulk protein analyses include immunoblotting (Western blot), enzyme-linked immunosorbent assay (ELISA) and multiplex immunoassays, and mass-spectrometry-based

proteomics [158–160]. Western blot quantifies the abundance and approximate molecular weight of specific proteins, providing high specificity but low throughput and typically requiring microgram-level total protein input. ELISA and bead-based multiplex immunoassays measure soluble or denatured antigens with picogram-level sensitivity, enabling quantification of cytokines and shed receptors. Mass-spectrometry-based proteomics provides discovery-mode protein identification and quantification, with low-input protocols (e.g., SP3, single-pot solid-phase enhanced sample preparation) extending feasibility into the sub-microgram range [161,162].

Application of bulk protein methods to CTCs is severely constrained by cell number. Because CTCs are typically recovered in very low numbers, often far below conventional bulk protein-input requirements, the total protein available falls well below the input thresholds of conventional Western blotting and most untargeted proteomics workflows. Bulk protein readouts of CTCs are therefore generally restricted to either pre-analytical signal aggregation across many patient samples or to soluble tumor-derived proteins captured directly from plasma. For per-cell protein analysis, imaging-based and single-cell methods described below remain the practical option in the CTC setting [159].

1.5.2 Single-Cell and Multi-Omic Technologies

The averaging inherent to bulk profiling obscures the heterogeneity that defines CTC biology [141,142]. Cellular resolution can be achieved by three complementary routes. The first physically isolates individual cells and analyzes their molecular content separately, with each cell processed in its own reaction volume, including single-cell DNA sequencing (scDNA-seq) and single-cell RNA sequencing (scRNA-seq) [142,163]. The second is to image and reach to a single cell/single molecular resolution, for example DNA fluorescence in situ hybridization

(FISH) and ddPCR. The following subsections summarize methods organized by the molecular layer they interrogate (DNA, RNA, protein), with emphasis on the techniques used in the present work: DNA FISH, scRNA-seq, ddPCR, and high-content imaging [164].

1.5.2.1 Single-Cell RNA Sequencing

Single-cell RNA sequencing (scRNA-seq) is one of the most informative approaches for resolving CTC heterogeneity, but platform choice involves a trade-off between throughput and transcript coverage [165]. Droplet-based platforms such as 10x Genomics Chromium enable 3′ or 5′ end-counting transcriptomic profiling of thousands to tens of thousands of cells, making them suitable for high-throughput cell-state discovery when sufficient input is available [166]. However, these methods generally provide limited full-length transcript information and are less suited for isoform- or fusion-level analysis. In contrast, plate-based full-length methods such as Smart-seq2 and Smart-seq3 process fewer cells but provide deeper transcript coverage, improved detection of low-abundance and alternatively spliced transcripts, and allele- or isoform-level information [167]. Therefore, droplet-based methods are useful for broad profiling, whereas Smart-seq-based methods are often preferable for rare CTC populations when high-coverage transcript, isoform, or fusion information is needed. In NSCLC, scRNA-seq has revealed substantial tumor and microenvironmental heterogeneity in advanced tumor biopsies, while studies of individual CTCs or CSF-derived tumor cells demonstrate the feasibility of single-cell transcriptomic profiling of rare, disseminated tumor cells [168].

1.5.2.2 Fluorescence In Situ Hybridization

DNA fluorescence in situ hybridization (FISH) localizes specific genomic loci within fixed cells using fluorescently labeled DNA probes, providing per-cell visualization of chromosomal rearrangements, copy-number alterations, and gene-locus integrity [169,170]. In

NSCLC, break-apart FISH is the established clinical assay for ALK rearrangement and remains a standard companion diagnostic for ALK-targeted therapy; analogous assays interrogate ROS1, RET, and NTRK fusions [169,171]. Applied to isolated CTCs, FISH enables detection of fusion-positive tumor cells in the circulation, quantification of ALK-positive CTC burden over time, and identification of fusion-pattern heterogeneity within the same patient. The technique requires fixation but preserves morphology and is compatible with multiplexed immunofluorescence on the same cell [169,172].

1.5.2.3 ddPCR for Single-Molecule Targeted Detection

ddPCR achieves single-molecule resolution within pooled input by partitioning the reaction into approximately 20,000 nanoliter droplets and counting positive partitions according to Poisson statistics [146]. Under typical conditions, each droplet contains zero or one target molecule, so the readout is a per-molecule digital count rather than an analog amplification curve. Under optimized assay conditions, droplet digital PCR can detect rare mutant alleles at approximately 0.001–0.01% variant allele frequency, corresponding to roughly one mutant allele among 10^4–10^5 wild-type alleles in a partitioned reaction system [173].

Applied to enriched CTC fractions, this resolution makes ddPCR functionally equivalent to single-cell analysis for targeted variants: when the input contains tens to hundreds of cells, a mutation present in a single CTC remains detectable above the background of contaminating leukocytes and any wild-type tumor cells, without the technical burden of physically isolating individual cells. ddPCR is therefore well-suited to quantification of fusion transcripts (e.g., EML4-ALK), resistance mutations (e.g., ALK G1202R and L1196M; EGFR T790M and C797S), and other low-abundance targets that fall below the detection limits of conventional sequencing at low cell input [23,174]. The trade-off is that ddPCR is targeted: each assay

measures one or a small panel of pre-defined variants and cannot support discovery-mode profiling. In practice, ddPCR is therefore frequently used as a sensitive, targeted complement to genome- or transcriptome-wide single-cell profiling.

1.5.3 Imaging-Based Phenotypic Profiling

Imaging-based characterization provides spatial, morphological, and subcellular localization data that complement molecular profiling. Advances in automated microscopy and multiplexed imaging have shifted the field from qualitative observation to quantitative, high-dimensional phenotyping [175,176].

1.5.3.1 Multiplexed Immunofluorescence and High-Content Imaging

High-content immunofluorescence utilizes multi-channel microscopy to visualize multiple proteins simultaneously. Standard 4-to-6 channel systems and advanced 8-to-12 channel platforms allow comprehensive phenotyping of epithelial, mesenchymal, proliferative, apoptotic, and immune checkpoint markers [177–179]. Automated systems quantify protein expression intensity, subcellular localization, co-expression patterns, and morphological parameters (e.g., nuclear-to-cytoplasmic ratio) across thousands of CTCs [177].

1.5.3.2 Cell Painting and Ultra-Multiplexed Imaging

Cell Painting is an unbiased morphological profiling technique using six fluorescent dyes to label eight cellular compartments, generating high-dimensional morphological fingerprints [175]. In CTCs, this enables the discovery of morphologically distinct subpopulations and the prediction of cellular states without relying on predefined markers.

For ultra-high-dimensional spatial proteomics, technologies such as co-detection by indexing (CODEX), imaging mass cytometry (IMC), and multiplexed ion beam imaging (MIBI) extend multiplexing to 50+ proteins [180]. These platforms employ iterative staining or mass

spectrometry-based detection to provide deep intracellular and surface profiling. However, the requirement for fixed cells, specialized equipment, and complex workflows currently restricts their widespread application in CTC research [181].

The raw data generated by these advanced imaging modalities, much like single-cell sequencing data, are highly complex and require specialized computational pipelines for feature extraction and pattern recognition, which will be discussed in the subsequent section.

1.5.4 Computational Analysis Pipelines

The high-dimensional, sparse, and heterogeneous nature of multimodal CTC datasets necessitates robust computational pipelines. This section outlines the standard analytical workflows corresponding to the molecular and imaging modalities.

1.5.4.1 scRNA-seq Analysis for CTCs

Standard scRNA-seq pipelines initiate with stringent quality control to remove low-quality cells and ambient RNA contamination, followed by normalization to mitigate batch and patient-specific effects [182].

Dimensionality reduction (PCA, UMAP) and graph-based clustering (Louvain, Leiden) partition cells into transcriptionally distinct subpopulations. A challenge in scRNA-seq is technical sparsity or "dropouts." Denoising and imputation algorithms, such as MAGIC, SAVER, and Deep Count Autoencoders (DCA), are employed to recover biological signals, though they must be applied judiciously to avoid introducing artificial correlations. Furthermore, trajectory inference algorithms (Monocle, Slingshot, PAGA) reconstruct continuous biological processes, such as EMT transitions, by ordering cells along pseudotime axes, demonstrating that CTC phenotypes often represent a continuum rather than discrete binary states [183,184].

1.5.4.2 Imaging Analysis for CTC Phenotyping

Image analysis pipelines require rigorous preprocessing to correct for uneven illumination, background noise, and channel crosstalk. Cellular detection and segmentation have evolved from traditional thresholding to robust deep learning models (e.g., U-Net) [185]. Following segmentation, feature engineering extracts morphological, intensity, and texture metrics. Supervised classifiers accurately distinguish CTCs from leukocytes, while unsupervised clustering identifies novel morphological sub-classes [186,187].

Beyond traditional feature extraction, deep learning architectures, including convolutional neural networks (CNNs) and generative models (VAEs, GANs), are increasingly applied directly to raw image data [188–191]. These self-supervised models can extract complex, hierarchical morphological representations to predict molecular subtypes and trace continuous morphological trajectories (e.g., EMT progression) without relying on predefined features.

1.5.4.3 Multi-Modal Integration and Predictive Modeling

Integrating diverse data modalities occurs at multiple levels: feature-level, representation-level (shared latent spaces), and model-level (ensemble methods). Computational alignment methods, such as Canonical Correlation Analysis (CCA), Mutual Nearest Neighbors (MNN), and Seurat anchoring, correct batch effects and map datasets into a common space [192,193]. Weighted Nearest Neighbor (WNN) approaches optimize the integration of RNA and protein data (e.g., CITE-seq) by learning modality-specific informativeness [194,195].

Predictive modeling leverages these integrated features to forecast clinical outcomes. Cox proportional hazards models identify prognostic biomarkers, while machine learning classifiers (Random Forests, Gradient Boosting Machines, Neural Networks) predict therapeutic responses [196]. For longitudinal monitoring, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks model temporal dynamics to anticipate resistance development [197–

199]. Rigorous cross-validation and feature importance analysis (e.g., SHAP values) ensure model interpretability and robustness [200,201].

1.6 Clinical Applications of CTCs in NSCLC

The clinical utility of CTCs in NSCLC ranges from baseline prognostic stratification to longitudinal monitoring of therapeutic response and resistance [202,203].

1.6.1 Prognostic Stratification

CTC enumeration is validated prognostically in NSCLC [204–206]. The baseline CTC counts strongly correlate with tumor burden, disease stage, and survival. In metastatic NSCLC, a threshold of ≥5 CTCs per 7.5 mL of blood is a robust prognostic indicator; patients exceeding this count exhibit significantly shorter progression-free survival (PFS) and overall survival (OS) [204].

Beyond baseline enumeration, longitudinal monitoring provides a dynamic assessment of treatment efficacy[207]. Rapid CTC clearance during first-line therapy indicates a favorable response, whereas failure to reduce CTCs or an early increase serves as an early indicator of disease progression, often preceding radiographic changes. Furthermore, phenotypic distribution carries independent prognostic weight. Patients harboring predominantly mesenchymal CTCs, hybrid EMT states, or circulating tumor microemboli (CTC clusters) face a worse prognosis due to enhanced metastatic potential and therapeutic resistance [140,208]. These findings suggest that integrating CTC counts, kinetics, and phenotypes may provide more clinically informative readouts than enumeration alone [209].

1.6.2 Molecular Profiling: Tumor Heterogeneity and Resistance Mechanisms

CTCs originate from multiple tumor sites, theoretically offering a more comprehensive representation of systemic intra-patient heterogeneity than spatially restricted tissue biopsies

[202]. A primary clinical task is the non-invasive detection of oncogenic drivers and resistance mutations [33]. Studies demonstrate 70–90% concordance between CTCs and matched tissue for EGFR and ALK alterations, contingent on the detection method and CTC burden [38,169,210].

In the context of targeted therapy, longitudinal CTC profiling enables the early detection of acquired resistance. For instance, the emergence of the ALK G1202R mutation can be detected in CTCs weeks to months prior to clinical progression, informing the sequential selection of next-generation ALK inhibitors [23]. At the single-cell level, genomic profiling identifies polyclonal resistance, where distinct CTCs within the same patient harbor divergent resistance mechanisms, such as concurrent kinase domain mutations and bypass pathway activation (e.g., MET amplification) [211].

Transcriptomic and proteomic profiling further reveal phenotypic resistance mechanisms, including EMT and small cell lineage transformation, which evade genomic detection [212]. Discordance between RNA and surface protein expression (e.g., PD-L1) underscores the necessity of multi-omic evaluation when guiding treatment selection [213].

1.6.3 Functional Characterization: Ex Vivo Culture and Drug Sensitivity

Unlike cell-free DNA (cfDNA), intact CTCs permit functional characterization through ex vivo expansion [214,215]. CTC-derived cultures and organoids retain the key molecular architectures of the parent tumor, enabling direct drug sensitivity testing and the study of evolutionary resistance trajectories in real-time. Empirical therapeutic vulnerability profiles generated from these cultures show concordance with clinical responses, making them particularly valuable for testing combination therapies targeting both driver oncogenes and bypass tracks [216–218].

However, significant limitations persist. Success rates for establishing cultures remain

modest (10–30%), the process requires a 2–4-week time window, and in vitro conditions may exert selective pressures that alter cellular phenotypes [214]. Ongoing innovations in microfluidics and stromal co-culture aim to mitigate these barriers.

1.6.4 CTC-Immune Interactions

Studies of CTC–immune interactions, primarily in preclinical models, suggest active immune evasion mechanisms including platelet-mediated NK cell shielding and NET-mediated immunosuppression (Section 1.3.3) [219,220]. Translation of these mechanisms to single-cell analysis of patient CTC–immune clusters is an emerging area, and longitudinal monitoring of CTC–immune parameters is being investigated as an early indicator of immunotherapy response or adaptive resistance [221].

1.6.5 Longitudinal Monitoring

Serial sampling of CTC provides a comprehensive tumor profile. Beyond enumeration, longitudinal molecular monitoring leverages the multi-omic readouts described in Section 1.5 to track resistance trajectories at higher resolution: ddPCR-based allele frequency tracking detects emerging resistance clones at low tumor burden, and single-cell transcriptomic profiling can capture polyclonal resistance and non-genetic phenotypic shifts to support rational therapeutic adaptation [174,222,223].

A significant evidence gap remains regarding CTC-guided clinical decision-making. Prospective trials evaluating early therapeutic switching based on CTC dynamics have produced mixed results. For example, a study found that switching chemotherapy regimens in metastatic breast cancer patients based on persistently elevated CTCs did not improve overall survival [224]. The lack of defined CTC-based decision thresholds, optimal timing for intervention, and integration of molecular data beyond enumeration highlight the need for mechanism-directed

prospective trials in NSCLC [225,226].

1.6.6 Integration and Future Perspectives

In summary, CTC-based analysis is well-established for prognostic stratification in NSCLC and is being developed for longitudinal molecular monitoring and resistance detection. However, routine clinical translation is hindered by technical challenges, including low CTC abundance, isolation biases, and the complexities of single-cell analysis, as well as a lack of standardized protocols [203,226].

Within NSCLC, ALK disease provides a particularly informative setting for CTC-based monitoring [169]. As outlined in Section 1.1.4, ALK TKI resistance spans on-target, bypass, and lineage-plasticity mechanisms that frequently coexist within a patient and across lesions. ctDNA captures the genomic component of this resistance but not phenotypic transitions; tissue biopsy captures both at a single site but cannot be repeated serially. CTCs, by contrast, are accessible by repeated blood draw and retain intact cellular architecture, supporting interrogation of fusion status (by ALK FISH), resistance mutations (by ddPCR), transcriptional state (by scRNA-seq), and protein-level phenotype (by multiplexed imaging) on the same cohort. The remainder of this dissertation develops and applies this multi-readout CTC framework in ALK NSCLC.

1.7 Hypothesis and Scope of this Thesis

The clinical management of ALK NSCLC is constrained by spatial and temporal tumor heterogeneity, which contributes to the eventual emergence of acquired drug resistance. Current clinical monitoring relies primarily on tissue biopsies and cell-free DNA (cfDNA). However, tissue biopsies are limited in their ability to represent spatial heterogeneity and are not readily repeatable, while cfDNA lacks cellular context and therefore cannot directly evaluate phenotypic plasticity.

CTCs represent a promising liquid biopsy modality because they can be collected through minimally invasive and repeatable procedures and preserve intact cellular content. Consequently, there is a critical need to systematically isolate and analyze CTCs from ALK NSCLC patients. Moreover, because CTCs provide information ranging from simple enumeration to high-dimensional molecular profiling, advanced bioinformatics approaches are required to interpret CTC-derived molecular data. To address these gaps, by exploring CTC heterogeneity using advanced high throughput single cell level computational pipelines, I aim to clarify how CTCs reflect contemporaneous tumor states, including oncogenic signaling, EMT, and key interactions with the immune microenvironment, and to fully assess their prognostic value for clinical management.

This dissertation is organized around three interconnected specific aims.

In Chapter 2, I demonstrate unbiased microfluidic isolation of CTCs and comprehensive molecular characterization at single-cell resolution. I evaluate whether CTCs recapitulate key properties of the underlying ALK NSCLC. Using downstream assays, including IF staining of epithelial/mesenchymal markers, scRNA-seq for differential expression analysis, CNV profiling, and assessment of downstream oncogenic signaling pathways, I further characterize CTC–immune interactions. This potentially provides evidence that CTCs leverage phenotypic plasticity and transcriptional reprogramming to persist in circulation.

In Chapter 3, I test the clinical utility of CTCs by linking the dynamic molecular features identified in Chapter 2 to patient outcomes. Using a longitudinal study design, I track CTC enumeration and transcriptomic shifts during targeted therapy. By correlating these longitudinal dynamics with therapeutic response, PFS and OS, I establish CTCs as a real-time prognostic biomarker that can reflect disease trajectory and provide early indications of therapeutic

resistance.

In Chapter 4, I address key analytical bottlenecks in CTC research, particularly those related to throughput and reproducibility in cell identification and phenotypic scoring. Building on the prior chapters, I develop and validate a scalable machine learning platform for automated CTC recognition and quantitative phenotype scoring. I further extend this platform to construct a computational framework for modeling the NK cell sensitivity of CTCs, demonstrating biological relevance through prediction of immune-related phenotypes, including susceptibility to NK-cell cytotoxicity.

In summary, this dissertation investigates the biological complexity of CTCs in ALK NSCLC and establishes an end-to-end framework spanning microfluidic isolation, single-cell multi-omic profiling, and automated machine learning–enabled image analysis. By translating high-dimensional single-cell measurements into clinically interpretable prognostic metrics, this work supports CTCs as a non-invasive biomarker to enable personalized management of ALK NSCLC.

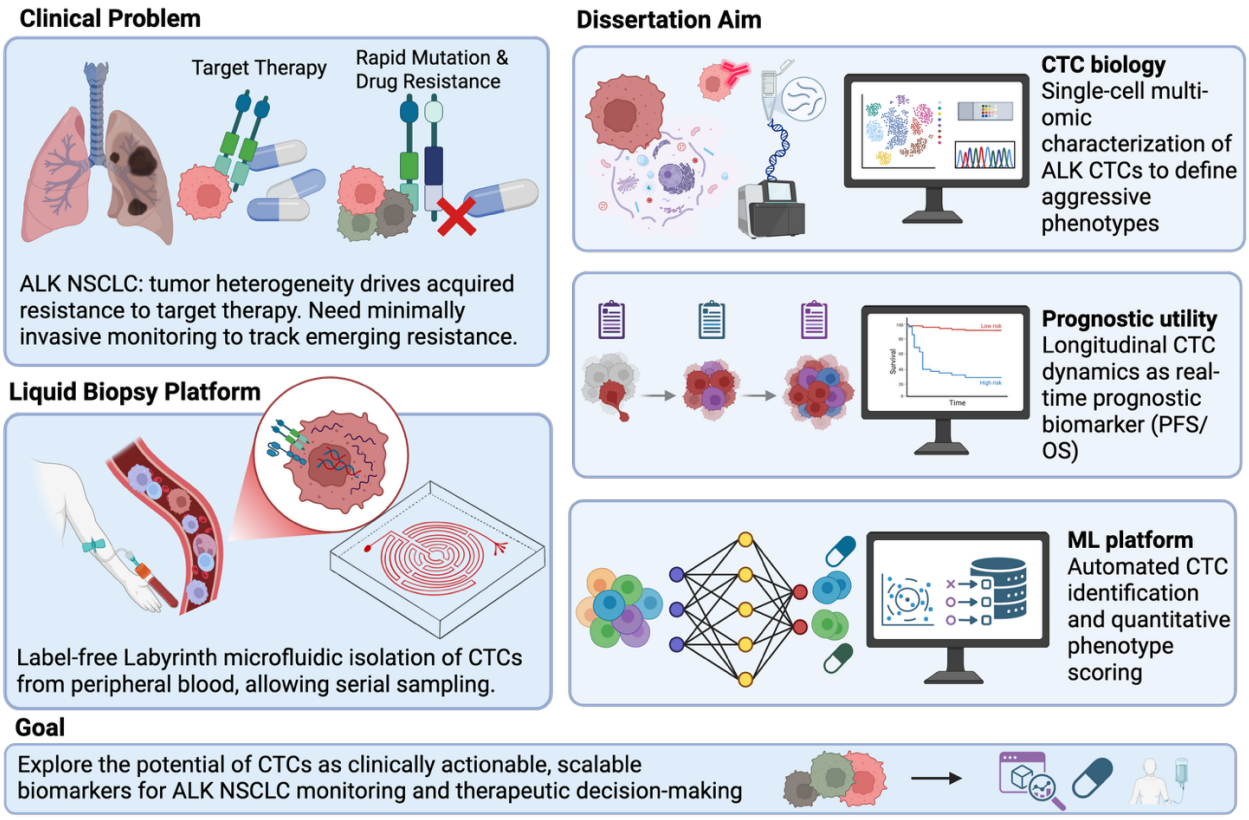


Figure 1.1 Overview of the dissertation: characterizing CTCs in ALK NSCLC for clinical monitoring.

Chapter 2 Single-cell Profiling Reveals Distinct CTC Phenotypes and Immune Cell Interactions in Patients with ALK-positive Non-Small Cell Lung Cancer

2.1 Abstract

Circulating tumor cells (CTCs) provide a minimally invasive window into tumor biology, enabling real-time tracking in non-small cell lung cancer (NSCLC). Anaplastic lymphoma kinase–positive NSCLC (ALK NSCLC) is a clinically aggressive molecular subtype in which tyrosine kinase inhibitor (TKI) therapy has substantially transformed survival outcomes; however, heterogeneous therapeutic responses and acquired resistance remain critical unresolved challenges. The transcriptional, phenotypic, and immunologic landscape of ALK CTCs at single-cell resolution has remained entirely unexplored. Employing label-free microfluidic enrichment (Labyrinth) combined with single-cell RNA sequencing and immunofluorescence, CTCs were profiled from 91 NSCLC patients (45 ALK, 46 non-ALK). ALK CTCs demonstrated elevated expression of epithelial-mesenchymal transition (EMT) states, upregulation of oncogenic (TIMP1, SMOX), platelet-interacting (ITGA2, PF4) gene, and an immune-suppressive phenotype. Cell–cell interaction analysis revealed the co-occurrence of NK-exclusive suppressive (TGFB1–TGFBR1/TGFBR2) and activating (ITGB2–CD226) signals, suggesting a potential mechanistic basis by which ALK CTCs may impose NK cell impairment through tumor-directed suppression.

2.2 Resulting Publication Information

Yuru Chen et al. "Single-cell profiling reveals distinct CTC phenotypes and immune cell

interactions in Patients with ALK-positive Non-Small Cell Lung Cancer" (In submission)

2.3 Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide [227], with non-small cell lung cancer (NSCLC) accounting for 87% of lung cancer deaths [228]. NSCLC is characterized by a broad distribution of genetic and molecular heterogeneity [229,230]. The diversity of mutations, along with highly variable biological behaviors and therapeutic vulnerabilities, present a major challenge for effective treatment.

Anaplastic lymphoma kinase positive NSCLC (ALK NSCLC) arises due to specific fusions between primarily EML4 and ALK that occur in approximately 3-7% of NSCLC cases [19] and has been identified as one of the most clinically actionable pathogenetic changes [231]. Compared to non ALK NSCLC, ALK NSCLC has distinct immune interaction profiles and metastatic patterns [232–234]. The oncogenic EML4-ALK fusion exhibits aberrant subcellular localization and constitutive kinase activation, leading to persistent PI3K/AKT, MAPK/ERK, and JAK/STAT signaling and promoting uncontrolled cell proliferation, survival, and migration [21,235]. ALK signaling also contributes to an immunosuppressive tumor microenvironment (TME), characterized by reduced CD8+ T cell infiltration in the TME, thereby contributing to immune evasion [236–238]. The cellular immunosuppressive TME appears to be clinically relevant in that studies have shown that ALK NSCLC is resistant to currently approved PD-1/PD-L1 immune checkpoint inhibitors (ICI) [239].

The discovery of ALK rearrangements as an oncogenic driver in NSCLCs led to the development of ALK tyrosine kinase inhibitors (TKIs). The advent of ALK TKIs, including first-generation (crizotinib), second-generation (alectinib, brigatinib, ceritinib, ensartinib), and third-generation (lorlatinib) agents, has led to a significant improvement in progression-free

survival (PFS) and overall survival (OS) compared to conventional chemotherapy [22,240]. However, therapeutic resistance eventually develops in most patients, with progressive disease that becomes more difficult to treat and more resistant to subsequent therapies, with each progression event.

Both on- and off-target resistance develop due to intrinsic and acquired tumor heterogeneity under the therapeutic pressure of TKIs [241]. Studies show that ALK NSCLC is inherently heterogenous, including the diversity of fusion partners and splice variants, variations in ALK gene copy number, and the presence of co-existing mutations [242]. The mechanisms of acquired resistance are diverse and can be classified broadly as on-target (ALK-dependent) or off-target (ALK-independent) [243,244]. On-target resistance, specifically secondary mutations in the ALK kinase domain, decreases the binding affinity of ALK inhibitors and increases the drug concentration required for effective kinase inhibition [23,28]. Off-target resistance, including activation of bypass signaling pathways (e.g., due to EGFR or MET amplification) that are parallel or downstream to ALK phosphorylation, reduce the dependence of the cell on ALK signaling for survival [24,26,28].

Different mutations within the ALK kinase domain can lead to varied patterns of drug response and eventual treatment failure. This underscores the necessity for advanced approaches to detect, as early in the progression trajectory for what events are potentially leading to resistance, and before resistant clones have the opportunity to become established at multiple metastatic sites. The molecular complexity of ALK NSCLC suggests static, DNA-level sequencing alone would be insufficient to capture the heterogeneous array of transcriptomic alterations comprising evolving resistance. We hypothesize that the ALK fusion oncogene establishes a dynamic transcriptomic landscape which underlies the biological heterogeneity of

resistant states. Therefore, comprehensively characterizing the transcriptome of ALK NSCLC in CTCs whose provenance is expected to be from more diverse sites than a single biopsy, offers an opportunity to understand the drivers of tumor evolution, detect molecular changes associated with resistance early, and ultimately inform more robust, personalized treatment strategies.

The conventional method to access tumor for molecular characterization is tissue biopsy. However, traditional tissue biopsies are invasive and subject to sampling location bias that may not fully capture intratumor heterogeneity. In this context, liquid biopsy, the analysis of tumor-derived material from body fluids like blood, offers a minimally invasive complementary alternative. [35,37,245–247].

In recent years, circulating tumor cells (CTCs) have demonstrated potential as an emerging liquid biomarker in multiple cancers. CTCs are cancer cells shed from solid tumors into the bloodstream. As viable cells, they can provide a snapshot of tumor's genomic, transcriptomic, and proteomic landscape, as well as other aspects of cell phenotype information (cell shape, contours, projections, junctions, clusters, etc). CTC analysis, including molecular profiling and enumeration, has demonstrated clinical utility in lung cancer by revealing treatment sensitivity and mutational profiles [35,37,44,45,205,245–248]. However, CTCs are extremely rare: ranging from 1 to 1000 CTCs in 1 mL of blood amidst millions of hematologic cells [249]. This highlights the need for highly efficient CTC isolation techniques and for sensitive downstream analytical methods.

Conventional CTC isolation methods frequently rely on the expression of the epithelial cell adhesion molecule (EpCAM). Yet, the heterogeneous phenotype of CTCs poses a challenge for isolation. Studies show that CTCs undergo epithelial-mesenchymal transition (EMT), a process associated with increased migratory capacity, survival in circulation, and therapeutic

resistance. Studies reported the ALK mutation and EMT state are concomitantly detected in ALK patients [20,23,27], suggesting that ALK NSCLC CTCs may undergo EMT. During EMT, CTCs may downregulate or lose EpCAM expression and gain mesenchymal features. Thus, EpCAM-based isolation methods are not optimal to capture CTCs undergoing EMT [250], potentially underestimating the total CTC population and missing possibly especially aggressive subpopulations in ALK patients. To overcome these limitations and capture the full spectrum of CTCs, we utilized a previously validated technology, the Labyrinth [131], a label-free microfluidic platform developed in our laboratory. This device enriches the full spectrum of CTCs unbiasedly by utilizing inertial focusing and separating them based on their biophysical properties, primarily size. Thus, label-free microfluidic methods are expected to capture more heterogeneous CTCs; however, they often yield lower CTC purity [45,251]. Thus, during the process of CTC enrichment, circulating immune cells surrounding the CTCs will be captured at the same time.

To address the challenges of CTC heterogeneity at the molecular level and lower CTC purity, more sensitive profiling technologies, such as single-cell sequencing, have been developed. Single-cell RNA sequencing, which offers enhanced resolution and sensitivity, allows profiling of individual CTCs to reveal molecular heterogeneity, identifying distinct gene expression signatures and immune interactions in multiple cancers. In NSCLC, single-cell transcriptomic analyses of CTCs have been pivotal in revealing cellular heterogeneity, EMT plasticity, and dynamic resistance mechanisms [252]. However, in ALK NSCLC, CTC research has been largely confined to enumeration or the detection of genomic alterations. To date, a systematic, unbiased transcriptomic characterization of ALK NSCLC patients' CTCs (hereafter referred to as ALK CTCs) remains unexplored.

To bridge this crucial gap, we present the first systematic single-cell transcriptomic characterization of ALK CTC, enabled by a workflow combining prompt processing of collected samples, label-free Labyrinth enrichment with high-throughput scRNA-seq. By comparing CTCs from ALK and non-ALK NSCLC patient cohorts, we aimed to define the unique molecular landscape of ALK-driven disease in circulation. Our study reveals that CTCs from ALK NSCLC patients are highly diverse in each patient and exhibit distinct transcriptomic signatures, prominently featuring an enrichment of EMT pathways and unique patterns of interaction with immune cells. This provides insights into their biological and clinical significance.

2.4 Results

2.4.1 Overview of patient characteristics and CTC enrichment using the Labyrinth microfluidic platform.

Patients in this study were adults (> 18 yrs.) recruited and consented as part of the Judith Tam ALK Lung Cancer Research Initiative (JTARI) at the University of Michigan. Eligibility criteria include any patient suspected to have lung cancer, preferably but not exclusively ALK NSCLC. Under informed consent (HUM 000042204, HUM 000227562), patients agreed to undergo extra passes to obtain tissue for research in clinically indicated biopsies or to donate tissue from operations not needed for clinical care. Fluids (pleural, ascites, cerebrospinal) were also obtained, if excess fluid was not needed for clinical care. Blood collection occurred during clinically indicated blood draws, and these samples were used in the CTC analyses reported here. This study analyzed peripheral blood samples from 91 NSCLC patients, comprising 45 patients with ALK NSCLC (ALK patients) and 46 patients lacking ALK rearrangements and EGFR mutations. For brevity, the latter group is hereafter referred to as 'wild-type' (WT). Patient demographics and clinical characteristics are displayed (Table 2.1, Fig. 2.1A).

Table 2.1. Clinical characteristics table of ALK NSCLC and wild-type NSCLC patients (n=45, n=46)

Category	ALK NSCLC n=45	Wild-type NSCLC n=46
Patient Demographics		
Smoking Status		
Unknown/Unreported	16	2
Never Smoker	24	14
Former Smoker, Quit	5	29
Current Smoker	0	1
Stage		
Unknown/Unreported	9	5
I	0	20
II	1	7
III	4	4
IV	31	10
Progression		
Unknown/Unreported	12	38
Stable	20	5
Partial Response	2	0
Progression	11	3
Sex		
Unknown/Unreported	0	24
Female	23	12
Male	22	10
Age at Diagnosis		
Unknown/Unreported	0	24

0-18	0	0
19-35	6	0
36-50	14	1
51-65	20	8
66+	5	13
Age (median, range)	52 (20-78)	68.5 (49-82)
Deceased		
Unknown/Unreported	17	39
No	27	0
Yes	1	7

The ALK cohort was younger on average, with a median age of 52 years (IQR: 20-78) compared to 68.5 years (IQR: 49–82) for wild-type patients ($p < 0.001$, Mann-Whitney U test). A higher proportion of ALK patients were never-smokers (53.3%, n=24) compared to wild-type patients (30.4%, n=14; $p< 0.001$, chi-square test). Most ALK patients had metastatic disease (68.9%, n=31) at the time of sample collection (Fig. 2.1A).

The study workflow, including blood processing, CTC enrichment, and downstream analyses is depicted (Fig. 2.1B). Prior studies have demonstrated the efficiency of the Labyrinth in enriching CTCs from NSCLC and other solid tumors, supporting its applicability to both ALK and wild-type NSCLC. The CTC enrichment protocol was first optimized using an in-house NSCLC CTC-derived cell line spiked into healthy donor blood. Validation experiments demonstrated a pseudo-CTC recovery rate of 96.85% ± 1.4% and a white blood cell (WBC) depletion rate of 89.67% ± 1.2% (Fig 2.2 E).

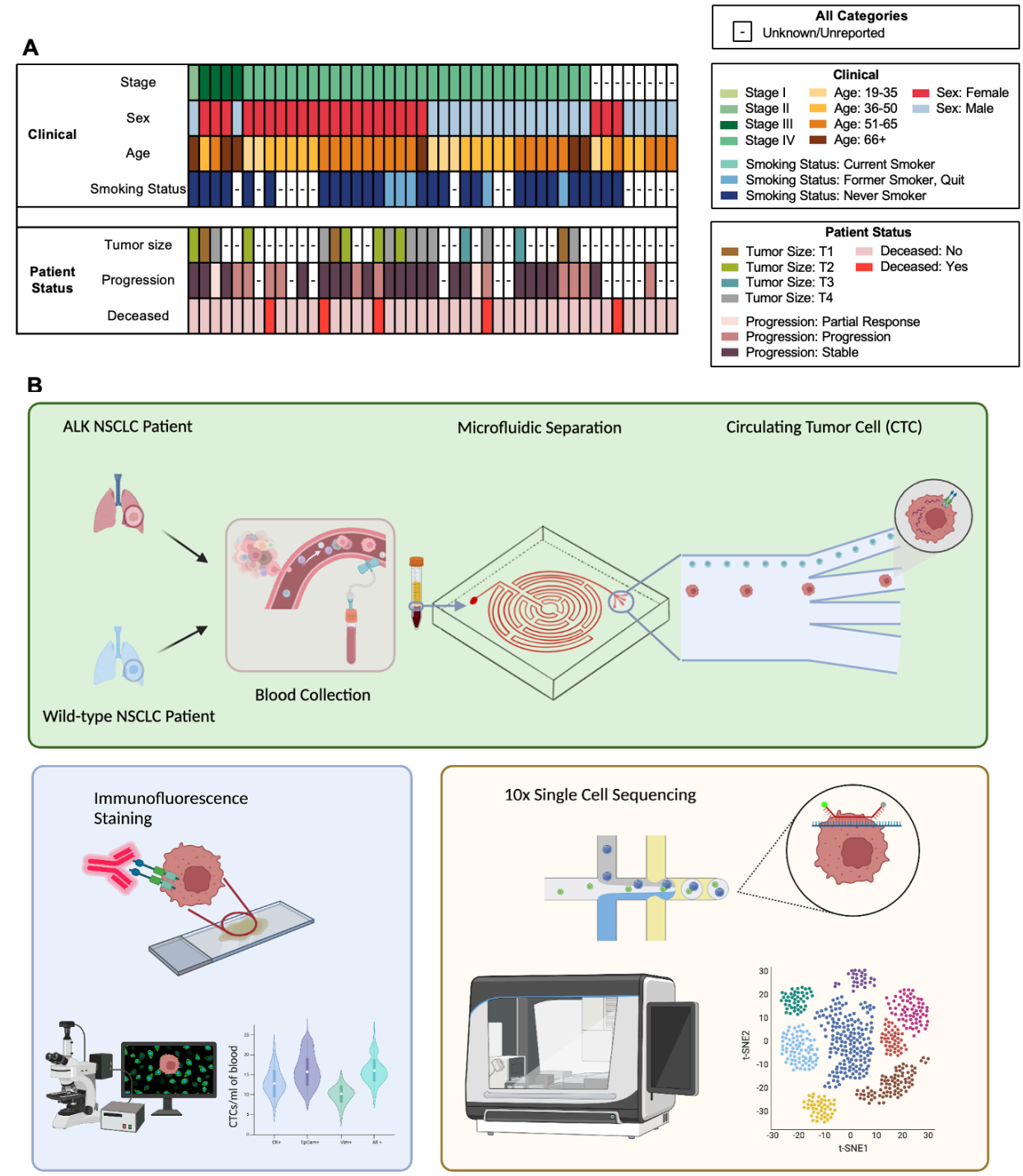


Figure 2.1 Clinical characteristics and overview experimental workflow for ALK NSCLC patients.

(A) Clinical characteristics of ALK NSCLC patients (n = 45). (B) One tube of peripheral blood was collected from each patient following the procedures described in the Methods section. Collected samples were

processed for CTC enrichment using a label free microfluidic device, Labyrinth. Enriched cells underwent immunofluorescent staining to validate CTC subtypes using established markers. Single-cell transcriptomic profiling was then performed to investigate molecular heterogeneity, gene expression signatures, and CTC–immune cell interactions. Created in BioRender.

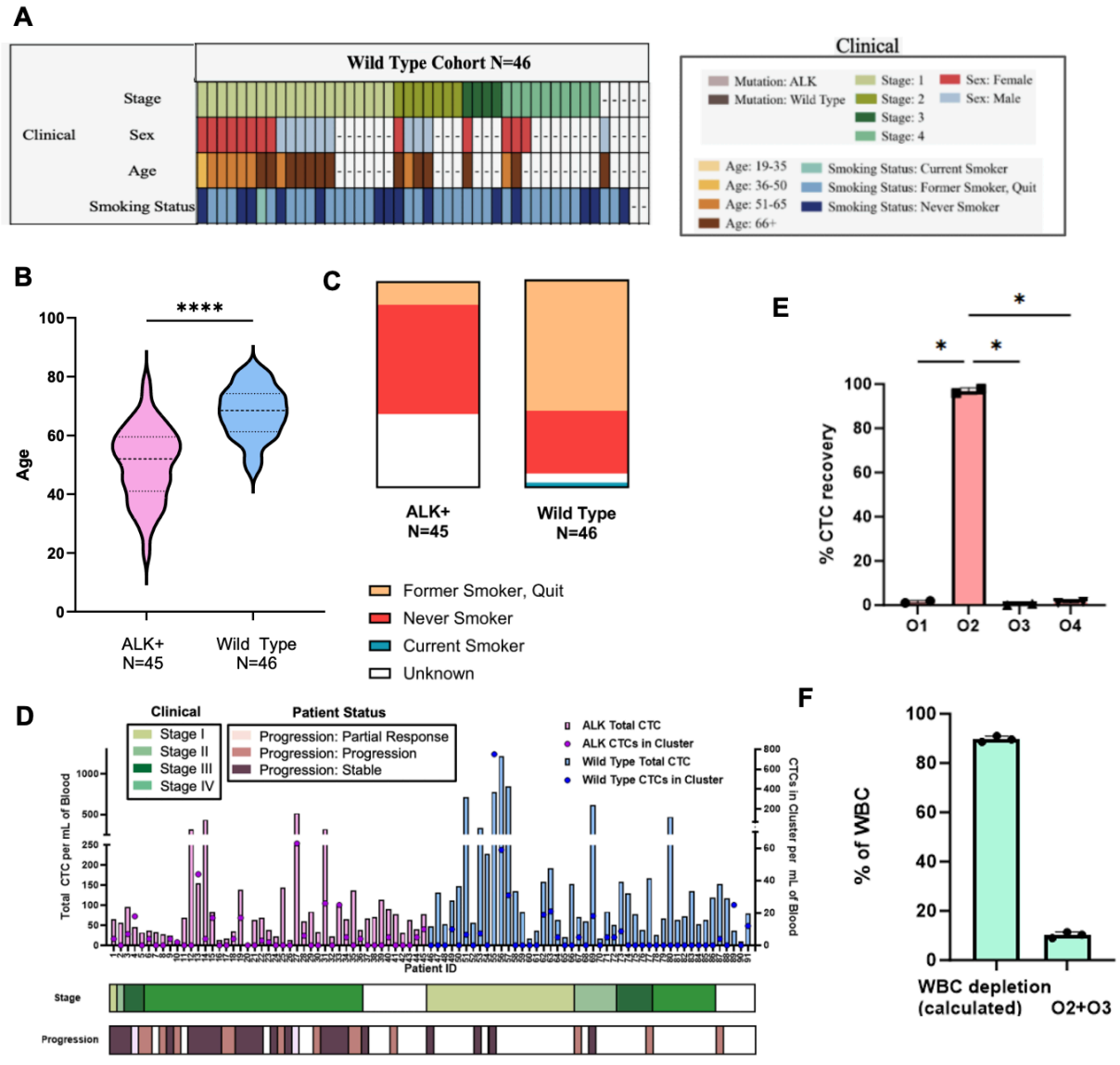


Figure 2.2 Clinical, CTCs characteristics, Labyrinth Validation for ALK and wild-type NSCLC cohorts.

(A) Summary of clinical characteristics for wild-type NSCLC patients (n = 46). (B) Age distribution of patients across ALK and wild-type cohorts. **** indicates $P \leq 0.0001$. (C) Smoking status distribution within ALK

and wild-type cohorts. (D) Quantification of CTC and CTC clusters per mL of blood in ALK versus wild-type NSCLC patients. (E) Pseudo-CTC recovery rate measured across three independent experiments. (F) WBC depletion rate measured across three independent experiments.

2.4.2 Characterization of CTCs in ALK NSCLC.

Following enrichment using the Labyrinth, under the conditions optimized as above, the enriched CTC product was processed for ALK fusion detection and CTC enumeration. The presence of ALK fusion and ALK protein in subset of ALK NSCLC patients was validated using multiple approaches including dPCR, FISH, and IF staining.

The CTCs are identified and counted with immunofluorescence staining. The staining panel is described in the Methods section. CTCs were defined as CK-positive, DAPI-positive, and CD45-negative (leukocyte marker). Subtyping categorized CTCs as CK+ (CK+, DAPI+, CD45−), Ep+ (EpCAM+, CK+, DAPI+, CD45−), Vim+ (vimentin+, CK+, DAPI+, CD45−), or all+ (EpCAM+, vimentin+, CK+, DAPI+, CD45−) (Fig. 2.3B). CTCs were successfully enriched and detected in 100% of the patient samples (Fig. 2.2D).

CTC counts varied within groups (range: 4–1215 CTCs/mL of blood), with large standard deviations. Within the ALK cohort, patients have a significantly lower percentage of Ep+ CTCs among subtypes of CTCs ($p \leq 0.0001$, Kruskal-Wallis test. Fig. 2.3C). Moreover, patients with progressive disease at the time of blood draw showed a trend towards higher percentage of Vim+ CTCs compared to those with stable disease ($p = 0.0668$, Mann-Whitney U test; n=11 vs. n=20, Fig. 2.3D). Additionally, former smokers in the ALK cohort showed a higher percentage of Vim+ CTCs than never-smokers, although this trend did not reach statistical significance ($p = 0.051$, Mann-Whitney U test; $n = 5$ vs. $n = 24$; Fig. 2.3E).

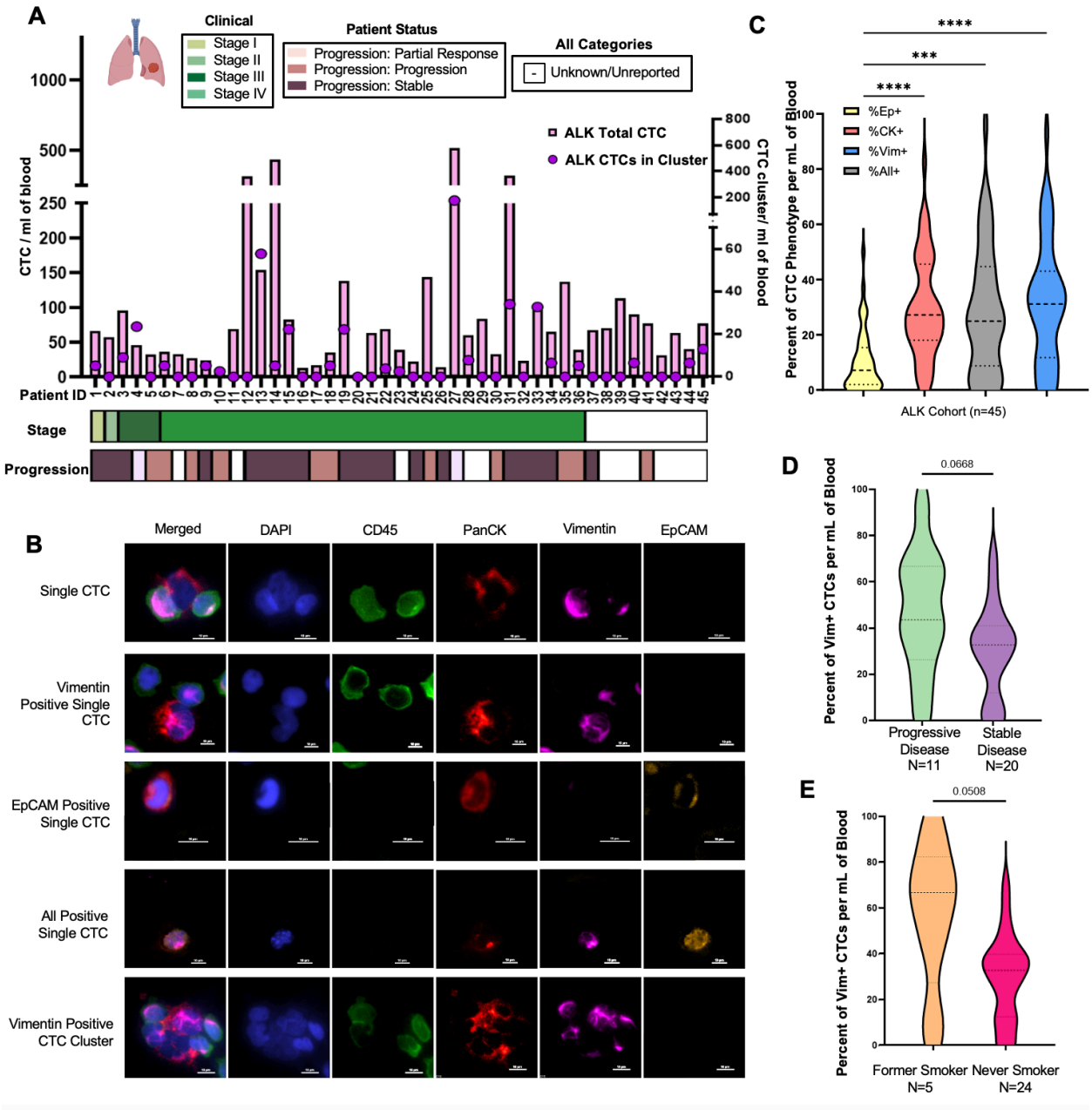


Figure 2.3. Phenotypic characterization and comparative analysis of CTCs in ALK NSCLC.

(A) Counts of total CTCs and CTC clusters per mL of blood in the ALK cohort. (B) Representative IF images of distinct phenotypic subpopulations within enriched CTCs. CTCs were defined as DAPI$^+$, PanCK$^+$, and CD45$^-$ cells. Cells were stained with DAPI (nuclear marker), CD45 (leukocyte marker), PanCK (tumor marker), Vimentin (mesenchymal marker), and EpCAM (epithelial marker). (C) Distribution (%) of CTC phenotypes in the ALK cohort (n = 45). (D) Percentage of Vim$^+$ CTCs in progressive disease versus stable disease patients in the ALK

cohort (n = 11; n = 20). (E) Percentage of Vim^+ CTCs in former smoker versus never smoker patients in the ALK cohort (n = 5; n = 24). **** indicates $P \leq 0.0001$; *** indicates $P \leq 0.001$.

2.4.3 Single-cell annotation shows lung originated CTC identification and Heterogeneity in the blood immune microenvironment

Seeking to understand the genomics and gene expression patterns of CTCs, a subset of samples enriched for CTCs was processed for single-cell RNA sequencing using the 10XGenomics Chromium platform. All data passed quality check. A total of 45,522 cells from five ALK NSCLC patients (patient IDs 6, 8,14,15, and 20) and 18,073 cells from two wild-type patients (patient IDs 54 and 77) were analyzed following the workflow described in the Methods section (Fig. 2.4A). Three major blood cell types were identified: erythroid, myeloid, and lymphoid cells. Four subtypes were observed in lymphoid cells: T cells, B cells, natural killer (NK) cells, and natural killer T (NKT) cells. Myeloid cells were further categorized into monocytes, basophils, neutrophils, and dendritic cells. In addition, non-classical immune cells, including erythroid-like and erythroid precursor cells, platelets, progenitor cells, mixed phenotype cells, and CTCs were identified (Fig. 2.4A-B). Cells from different patients clustered together according to transcriptomic similarity. This pattern is consistent with observations reported in previous studies [252].

CTCs were primarily identified by the expression of marker genes (PTPRC−, PPBP+, PF4+, TIMP1+, SPARC+) and subsequently validated by copy number variation (CNV) scores calculated using inferCNV (Fig. 2.4C). A total of 1,955 CTCs from the five ALK patients and 1,835 CTCs from two wild-type patients were identified. Average CTC counts were higher in wild-type patients than in ALK patients. Consistently, immunofluorescence-based enumeration also revealed a greater overall CTC count in wild-type patients. However, single-cell CTC

counts did not perfectly align with IF CTC count number, likely due to the stochastic sampling inherent to 10XGenomics technology (Table 2.2).

Table 2.2. Clinical characteristics and single-cell CTC count of ALK and wild-type NSCLC patients.

Patient ID	Sex	Current State	Initial Rx	Current age	Current Rx	SC Cell count	Sc CTC count	IF CTC count/ mL of blood
20	F	stable	Alectinib	81	Alectinib	7235	375	4
6	F	progression	Carboplatin/Pemetrexed /Bevacizumab	37	Lorlatinib	5376	135	36
14	F	stable	Carboplatin/Pemetrexed	68	Alectinib	10156	21	39
8	F	stable	Alectinib	46	Alectinib	9026	837	298
15	F	stable	Alectinib	55	Paclitaxel, Atezolizumab	13729	587	225
77	F	progression	Carboplatin, Paclitaxel	58	No treatment	6428	1221	167
54	M	stable	NA	62	NA	11645	614	228

To further characterize that the identity of the CTCs in scRNA-seq as deriving from lung tumors, we used the differential expressing gene and performed gene set enrichment analysis (GSEA) using C8 cell type signature gene sets from the Molecular Signatures Database (MSigDB). The analysis revealed a strong and significant enrichment of lung epithelial cell markers, supporting a lung tissue origin for these identified CTCs (Fig. 2.4D). Intriguingly, a megakaryocyte-associated gene signature was also significantly enriched within this same cluster. To resolve this ambiguity, we visualized the expression of key marker genes in the identified CTC cluster using Seurat's FeaturePlot function. We further explored the gene expression in the identified CTC cluster as described in the Methods section. Those CTCs lacked expression of *PTPRC* (CD45), a pan-leukocyte marker required for megakaryocytes identification. Moreover, the observation of a megakaryocyte-like signature in CTCs is

consistent with prior reports describing molecular mimicry between these two cell types, potentially due to phenomena such as platelet-cloaking [52]. The identified CTCs displayed recurrent copy-number alterations (gains in 1q, 7p, 7q, and 20p; loss of 16q; Fig 2.5A-B), providing genomic evidence of a nucleated malignant cell that a platelet, doublet, or ambient-RNA artifact cannot produce. Together, this further suggests the identified CTCs cells are lung-derived CTCs rather than a pure megakaryocyte population.

Both ALK and wild-type NSCLC patients' CTC exhibited significantly distinct CNV profiles compared to the immune cell populations in the same samples, with interpatient and intra-patient CNV heterogeneity observed (Fig 2.5C). ALK CTCs exhibited recurrent copy-number alterations characterized by gains in 1q, 7p, 7q, and 20p, and losses in 16q, patterns that are consistent with previously reported chromosomal aberrations in lung adenocarcinoma [253,254].

Having established the identity and genomic features of CTCs, we next examined the immune cell compartment within the same samples. Interestingly, in ALK NSCLC patients, we identified a distinct progenitor-like immune population that was absent in wild-type cases. This population showed upregulation of genes involved in metabolic adaptation (SLC2A14), cell-cycle regulation (TYMS, UHRF1), and proliferation (MKI67).

Detailed immune profiling revealed distinct immune cell distributions between ALK and wild-type NSCLC patients. ALK patients exhibited fewer effector CD8+ T cells and a higher proportion of memory CD8+ T cells, suggesting mechanisms of immune exhaustion or evasion, that may contribute to tumor progression and resistance to immunotherapy. Additionally, ALK patients had increased circulating neutrophils and myeloid dendritic cells (mDCs) compared to wild-type patients, possibly reflecting a pro-tumor inflammatory microenvironment known to

promote immunosuppression and angiogenesis in lung cancer [234,255,256].

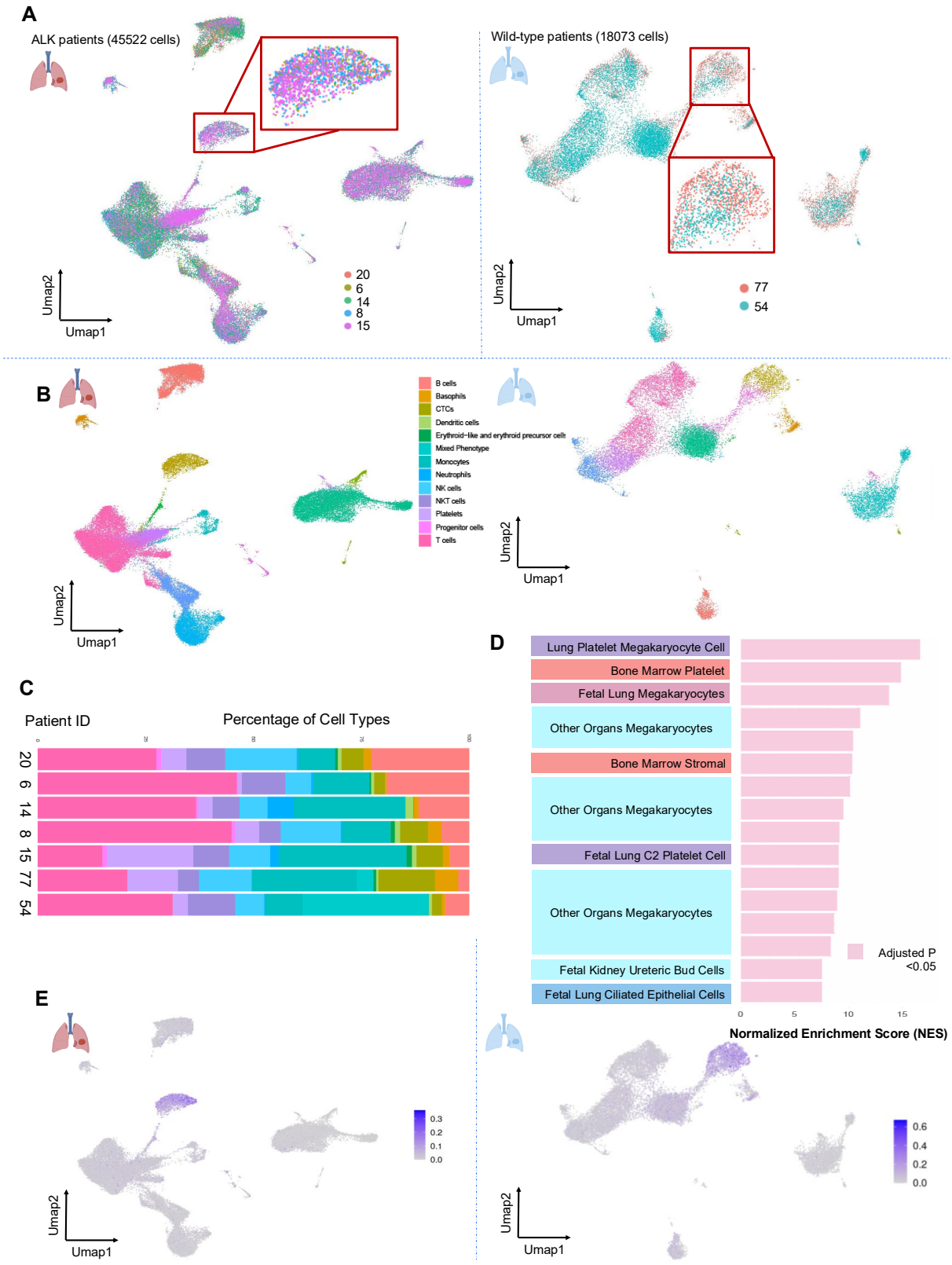


Figure 2.4. Single cell visualization and identification of CTC in ALK and wild-type NSCLC cohorts.

(A) UMAP projection of ALK single cells samples (left) and wild-type samples (right) colored by patients ID. CTC clusters are boxed. (B) UMAP projection of ALK single cells sample (left) and wild-type samples (right) colored by the cell types (C) Percentage distribution of major cell types across ALK and wild-type cohorts. (D) Pathway GSEA analysis showing the predominant biological pathways linked to the cell type origins of CTCs in the ALK cohort, filtered with an adjusted P value smaller than 0.05. (E) CNV scores of ALK (left) and wild type (right), highlighted the identified CTC has the highest CNV score

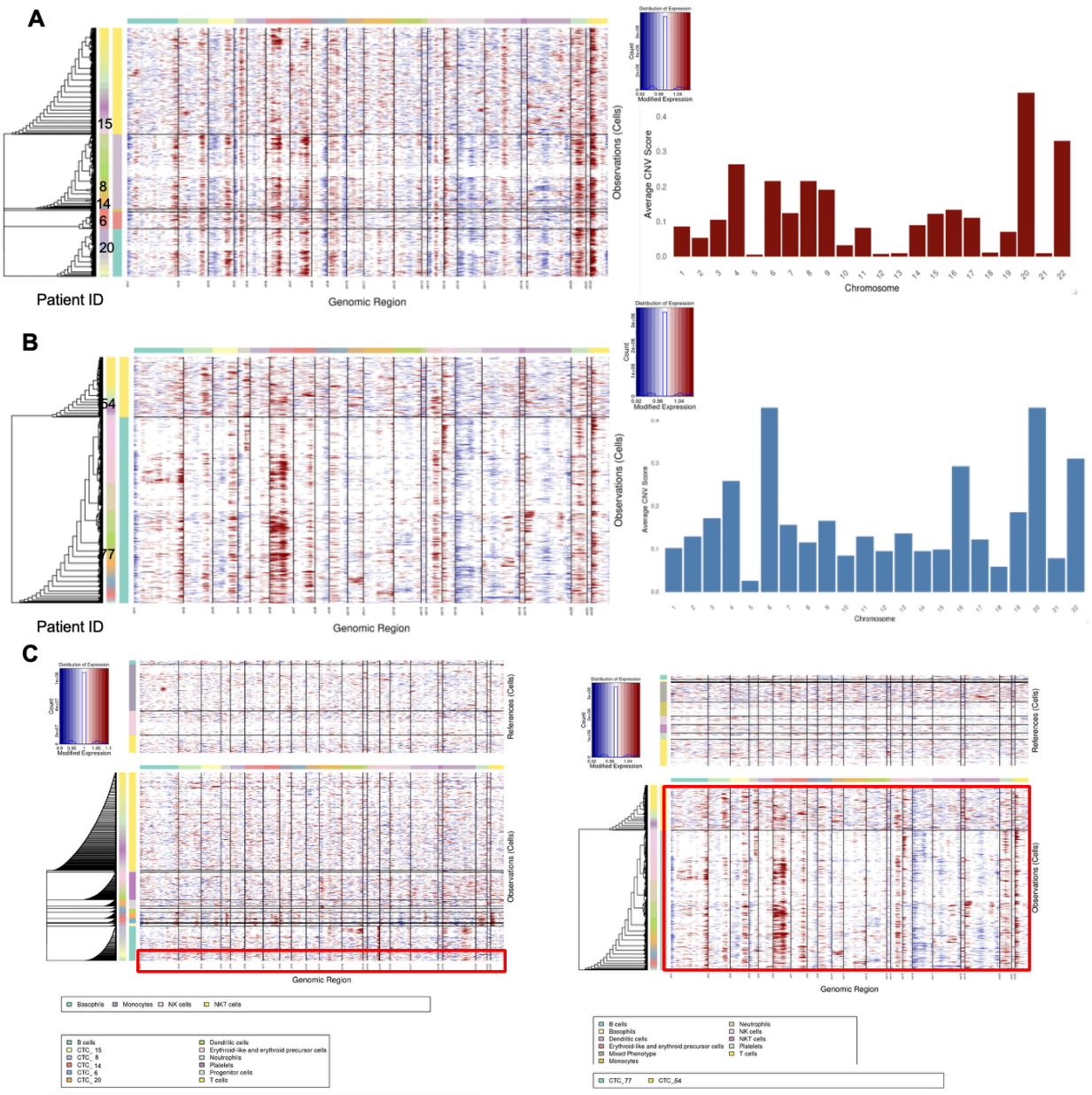


Figure 2.5. Chromosomal CNV profiles and CTC CNV scores in ALK and wild-type cohorts.

(A) CTCs' CNV distribution across all chromosomes in the ALK cohort (left) with corresponding average CNV scores (right). (B) CTCs' CNV distribution across all chromosomes in the wild-type cohort (left) with corresponding average CNV scores (right). (C) CNV with reference cell from ALK (left) and wild-type (right) cohort. CTC was boxed in red, showing a busier CNV profile compared to rest of the blood cells.

2.4.4 Characterization of CTCs reveals platelet and EMT related signature and pathways in ALK NSCLC

To investigate the molecular features of ALK CTCs, we performed differential expression analysis comparing CTCs to other circulating immune cell types (Fig. 2.6A). We observed significant upregulation of TIMP1, PANX1, SMOX, and CXCL5. TIMP1 was identified in single-cell CTC analyses of pancreatic ductal adenocarcinoma tumors [77]. PANX1, SMOX, and CXCL5 have been reported to be upregulated in NSCLC and have been implicated in promoting metastasis and are associated with poor prognosis [257–261]. Their upregulation parallels an EMT-related gene signature, characterized by elevated expression of VCL, CLU, and GATA1. Both VCL and CLU have been associated with EMT in gastric and lung cancers [261,262]. The expression of GATA1 has been shown to be upregulated in a subset of gastric cancer CTCs linked to EMT and chemotherapy resistance [263]. Expression of EMT genes in CTCs is consistently associated with heightened metastatic potential and increased resistance to therapy across various malignancies, including prostate and colorectal cancers [261,262,264,265]. In contrast, TP53, a canonical tumor suppressor gene frequently inactivated in NSCLC, was markedly downregulated in ALK CTCs with a log2 fold change of –0.87, reflecting a transcriptional shift toward a less regulated cellular state.

Platelet associated genes were observed in ALK CTCs, including expression of PPBP, PF4, PEAR1, ITGA2, ITGA2B, ITGB3, and CCL5. These genes are expressed in multiple cancer types, suggesting a platelets signature to be present on CTCs [263]. Platelets are increasingly recognized as mediators of CTC biology, forming a protective cloaking around CTCs that may shield them from immune surveillance and dampen shear stress in the bloodstream. Confocal microscopy following immunofluorescence staining confirmed extensive

platelet-CTC interactions, with platelets forming a coating around CTCs, supporting enhanced CTC-platelet associations in this cohort (Fig 2.7A). These findings collectively indicate that upregulation of platelet-related genes reflects increased CTC-platelet interactions.

Differential expressing gene ALK CTCs compared to the CTCs in the wild-type cohort were evaluated (Fig. 2.6B). Top-expressing genes in ALK CTCs revealed transcriptional signatures consistent with the activation of oncogenic signaling cascades downstream of ALK. The upregulation of PIK3CD and JUN indicates activation of the PI3K/AKT and MAPK pathways, which are the downstream pathways activated by ALK phosphorylation that drives tumor growth, survival, and therapeutic resistance [266]. Moreover, ALK CTCs exhibited elevated expression of genes associated with metastatic potential, including the matrix-degrading enzyme CTSF (Cathepsin F) and the ephrin receptor EPHB6, both previously implicated in tumor invasion and cell motility [267,268]. A striking increase in CCR7 expression was also observed, supporting enhanced JAK/STAT pathway activation [269]. Conversely, a significant downregulation of SOCS1, an inhibitor of the JAK/STAT pathway [270], further supports pathway activation. Collectively, these findings delineate a transcriptional landscape characterized by PI3K/AKT, MAPK, and JAK/STAT activation, which may cooperatively promote CTC survival and dissemination under ALK-driven oncogenic signaling.

2.4.5 Pathway analysis and identification of CTC heterogeneity in ALK NSCLC

To understand the functional pathways associated with differentially expressed genes in CTCs from ALK patients, we conducted enrichment test and GSEA on the Gene ontology biological process (GOBP), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Cancer Hallmark gene set. Our analyses revealed significant enrichment of pathways related to EMT and platelet interactions, including actin filament organization, homotypic cell-cell adhesion, focal

adhesion, trans-endothelial migration, platelet aggregation, and platelet activation (Fig. 2.6C; Fig 2.7B). These pathways align with the top expressing gene signatures observed in ALK patient's CTCs.

RAS oncogenic signaling pathways were also significantly upregulated, which are critical for ALK-driven tumor cell survival and proliferation as well as enhanced ALK-mediated oncogenesis [21]. Additionally, we observed significant upregulation of the Hippo pathway, which is hypothesized to contribute to an immunosuppressive tumor microenvironment in NSCLC [234].

Conversely, pathways governing programmed cell death and immune surveillance were significantly downregulated, including ubiquitin-mediated proteolysis and apoptosis, which are essential for tumor cell elimination (Fig. 2.6C). A marginally significant downregulation of the P53 signaling pathway was also noted, consistent with the downregulated expression of TP53. Furthermore, we observed suppression of pathways related to adaptive and innate immunity, including T cell receptor (TCR), B cell receptor (BCR), and NK cell-mediated cytotoxicity signaling. This was accompanied by a marked downregulation of both IFN-α and IFN-γ response pathways. Collectively, this signature indicates a profound immunosuppressive or immunologically 'cold' phenotype within the ALK CTCs. PD-1/PD-L1 checkpoint pathway was also downregulated and potentially showed a reduced responsiveness to anti-PD-1/PD-L1 therapies in ALK patients.

2.4.6 Inter- and Intra-patient heterogeneity of CTCs

We next examined inter-patient and intra-patient gene expression variability in ALK CTCs. CTCs from different patients cluster together based on their expression. The majority of

ALK CTCs from different patients exhibited high expression of platelet-related and EMT-associated genes such as SPARC, compared to the remaining cell populations, suggesting enhanced cellular plasticity and potential platelet–CTC interactions that may facilitate vascular transit and immune evasion. Within this population, we identified a distinct subcluster characterized by elevated expression of cancer-associated transcription factors SERPINE1, GATA1, immune-interaction mediator CXCL5, and stemness marker ALDH2. Intra-patient heterogeneity was also observed among ALK CTCs, characterized by diverse expression profiles of EMT- and oncogenesis-related genes in each patient group (Fig. 2.6D).

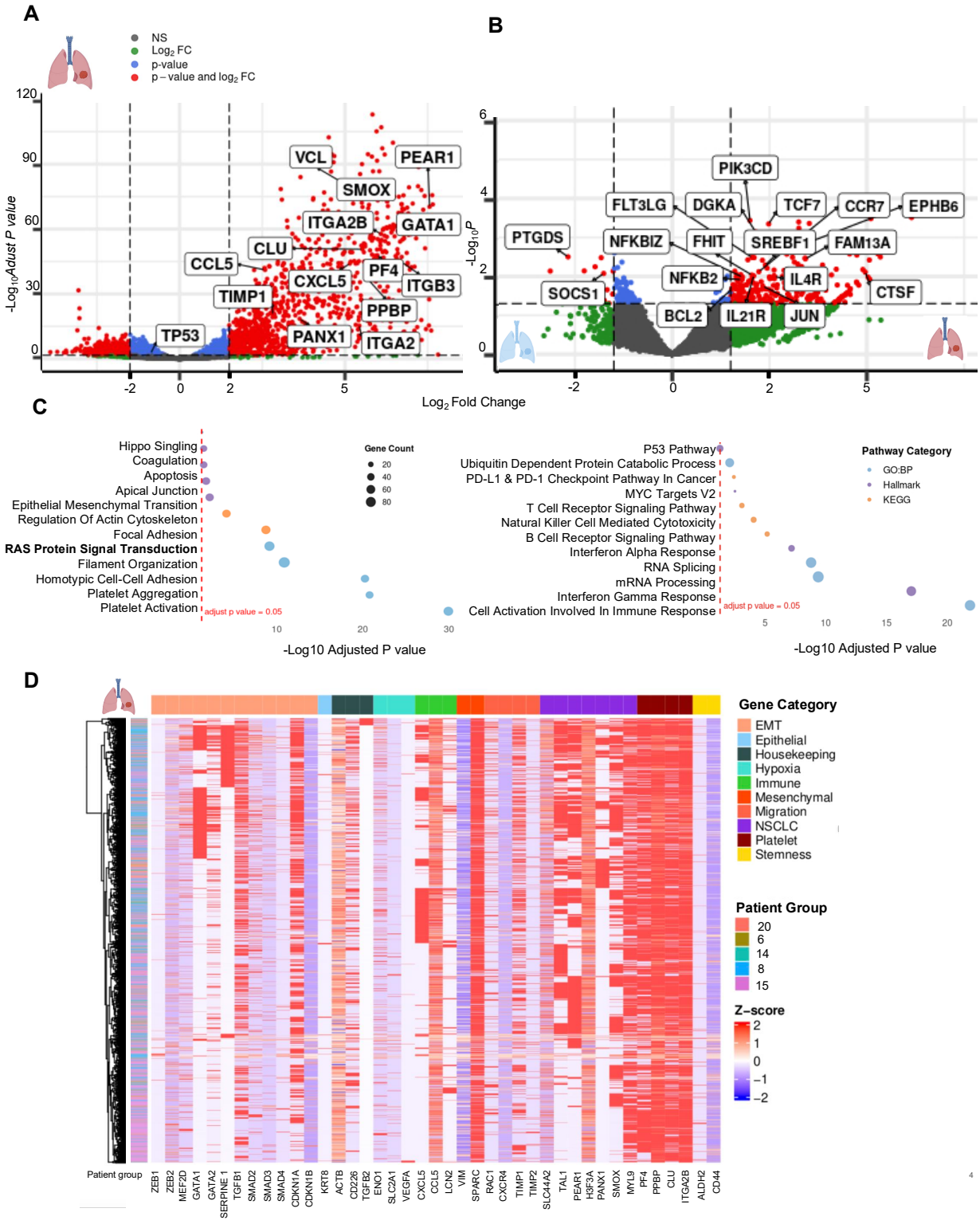

A
NS
Log2 FC
p-value
p – value and log2 FC
-Log10Adust P value
VCL
PEAR1
SMOX
ITGA2B
GATA1
CLU
CCL5
PF4
CXCL5
ITGB3
TIMP1
PPBP
TP53
PANX1
ITGA2
B
-Log10P
PIK3CD
FLT3LG
DGKA
TCF7
CCR7
EPHB6
PTGDS
NFKBIZ
FHIT
SREBF1
FAM13A
NFKB2
IL4R
CTSF
SOCS1
BCL2
IL21R
JUN
Log2 Fold Change
C
Hippo Singling
Coagulation
Apoptosis
Apical Junction
Epithelial Mesenchymal Transition
Regulation Of Actin Cytoskeleton
Focal Adhesion
RAS Protein Signal Transduction
Filament Organization
Homotypic Cell-Cell Adhesion
Platelet Aggregation
Platelet Activation
Gene Count
adjust p value = 0.05
-Log10 Adjusted P value
P53 Pathway
Ubiquitin Dependent Protein Catabolic Process
PD-L1 & PD-1 Checkpoint Pathway In Cancer
MYC Targets V2
T Cell Receptor Signaling Pathway
Natural Killer Cell Mediated Cytotoxicity
B Cell Receptor Signaling Pathway
Interferon Alpha Response
RNA Splicing
mRNA Processing
Interferon Gamma Response
Cell Activation Involved In Immune Response
Pathway Category
GO:BP
Hallmark
KEGG
adjust p value = 0.05
-Log10 Adjusted P value
D
Gene Category
EMT
Epithelial
Housekeeping
Hypoxia
Immune
Mesenchymal
Migration
NSCLC
Platelet
Stemness
Patient Group
Z-score
Patient group
ZEB1
ZEB2
MEF2D
GATA1
GATA2
SERPINE 1
TGFB1
SMAD2
SMAD3
SMAD4
CDKN1A
CDKN1B
KRTB
ACTB
CD226
TGFB2
ENO1
SLC2A1
VEGFA
CXCL5
CCL5
LCN2
VIM
SPARC
RAC1
CXCR4
TIMP1
TIMP2
SLC44A2
TAL1
PEAR1
H3F3A
PANX1
SMOX
MYL9
PF4
PPBP
CLU
ITGA2B
ALDH2
CD44

Figure 2.6. Single cell transcriptomic profiling of CTCs in ALK NSCLC.

(A) Volcano plot showing genes significantly upregulated (right red dot) and downregulated (left red dot) in CTCs compared with all other cells within the ALK cohort (B) Volcano plot displaying genes significantly upregulated (right red dot) and downregulated (left red dot) in ALK cohort's CTC compared to the wild-type cohort's CTC. (C) Pathway enrichment analysis using ORA depicting biological pathways upregulated (left) and downregulated (right) in CTCs from the ALK cohort from GO:BP, cancer hallmark, and KEGG gene set, with an adjusted p value cut-off of 0.05. (D) Heatmap illustrating the inter patient transcriptional profiles of CTCs in the ALK cohort, highlighting distinct expression patterns relative to other cell types.

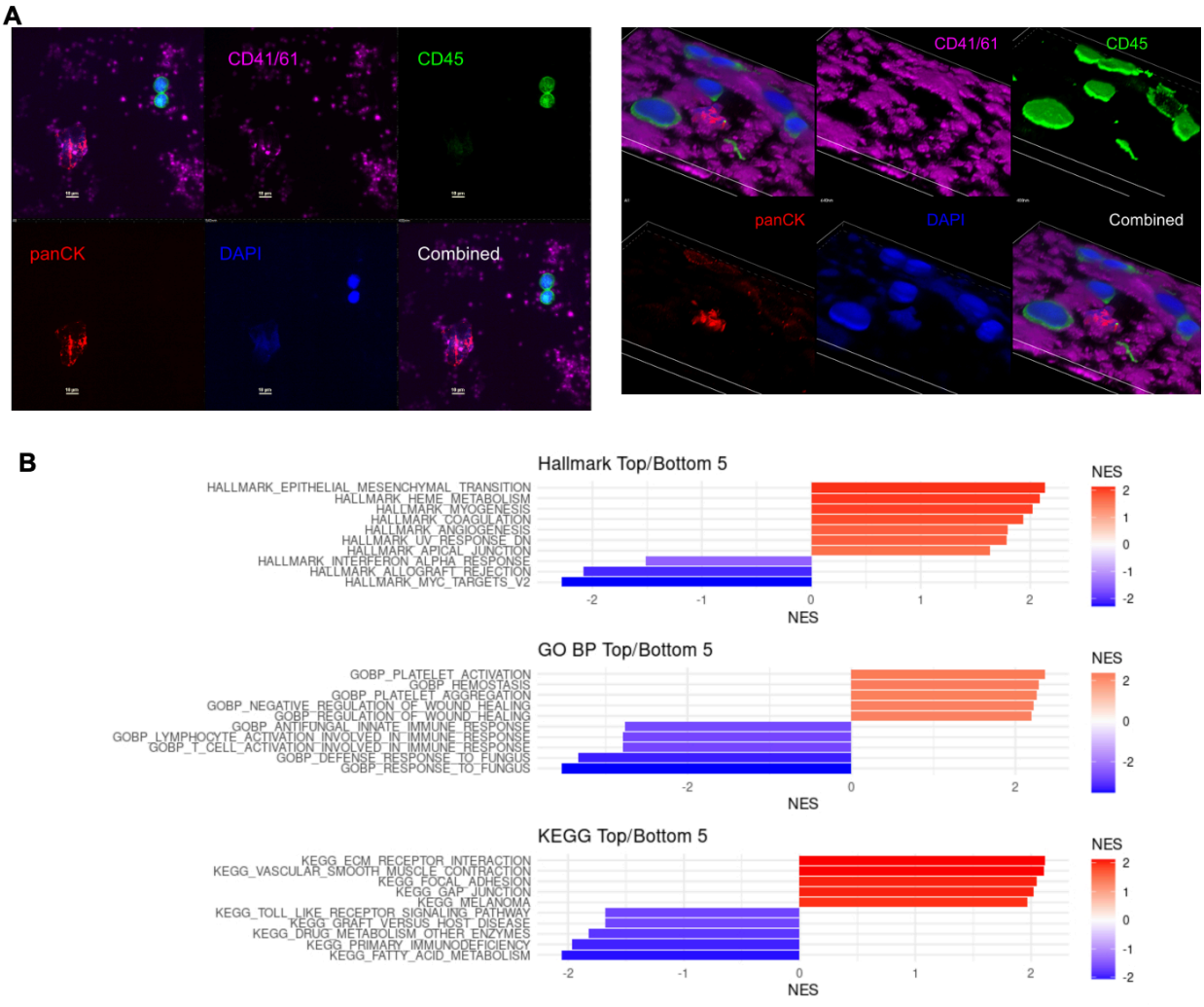


Figure 2.7. CTC-Platelet staining and significantly enriched pathways for ALK CTCs

(A) IF image of enriched CTCs from ALK patient 14 with platelet staining. (B) Significantly enriched upregulated (red) and downregulated (blue) pathways for ALK CTCs identified by GSEA in Hallmark, GO BP, and KEGG categories.

2.4.7 CTC immune cell interaction pattern further shows heterogeneity in ALK vs wild- type NSCLC patients

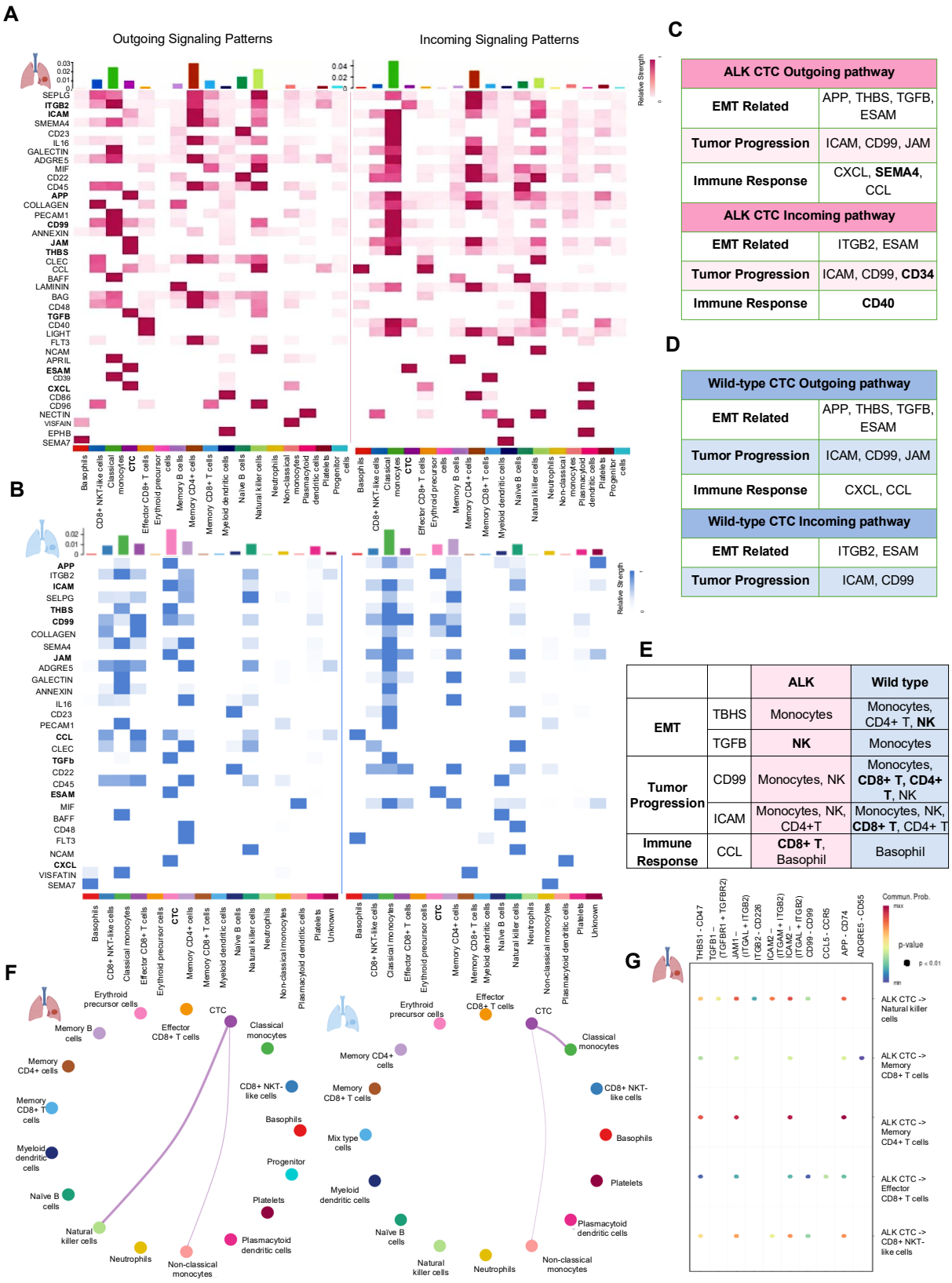
To explore molecular interactions between CTCs and their circulating immune microenvironment, we performed CellChat analysis that infers and visualizes cell–cell communication networks from transcriptomic data, to identify potential ligand–receptor interactions among cell populations. Our analyses revealed the outgoing from-CTC and incoming toward-CTCs signaling in the ALK and wild-type cohorts (Fig. 2.8A-B). Those pathways were then categorized into three major groups: EMT related pathways, tumor progression pathways, and immune response pathways (Fig. 2.8C-D). Analysis of CTC-derived outgoing signaling interactions revealed marked differences in the immune engagement between ALK and wild-type cohorts‘ CTC across multiple oncogenic pathway categories (Fig. 2.8E).

Compared with wild-type CTCs, ALK CTCs showed upregulated EMT-related signaling pathways, including APP, THBS, TGF-β, and ESAM. Notably, ALK CTCs preferentially directed TGF-β signaling toward NK cells, whereas wild-type CTCs routed preferentially to monocytes (Fig. 2.8F). Examination of the underlying ligand–receptor pairs revealed heterogeneous interaction profiles across NK and T cell subsets (Fig. 2.8G, $p < 0.01$). THBS1–CD47, an immune-evasive checkpoint axis, was present across all five immune cell populations examined NK cells, memory CD8+ T cells, memory CD4+ T cells, effector CD8+ T cells, and CD8+ NKT-like cells. This indicates a broadly deployed immune suppressive mechanism in this ALK cohort. The adhesion interaction JAM1–(ITGAL+ITGB2) was similarly distributed across all five populations. ICAM2-mediated interactions exhibited subset-specific distributions:

ICAM2–(ITGAM+ITGB2) was absent from memory CD8+ T cells, whereas ICAM2–(ITGAL+ITGB2) was absent from memory CD4+ T cells. In contrast, NK cells were the sole population subject to two signals absent from all T cell subsets: suppressive CTC-to-NK TGF-β engagement via TGFB1–(TGFBR1+TGFBR2), and the activating interaction ITGB2–CD226. The co-occurrence of these two NK-exclusive signals, one putatively suppressive, one putatively activating, is consistent with a model in which NK cells retain intact activation-receptor machinery while being selectively engaged by tumor-derived suppressive signals. As these interactions are computationally inferred, this model remains hypothetical, but it nominates CTC–NK crosstalk as a candidate node for functional investigation and potential immunotherapeutic intervention. In the tumor progression domain, upregulation of ICAM, CD99, and JAM signaling was observed, consistent with their established roles in cancer cell migration, invasion, and hematogenous dissemination[271–273]. Furthermore, CellChat analysis revealed that ALK-mutant CTCs received incoming signals through the CD34 pathway, suggesting active expression of receptors responsive to CD34-associated ligands. As CD34 signaling is functionally implicated in angiogenesis and tumor vascularization[274], this finding raises the possibility that ALK CTCs engage endothelial-derived signals to facilitate their survival and dissemination within the bloodstream.

Within the immune signaling domain, both CXCL and CCL pathways were found to be upregulated. Critically, differential CD8+ T cell engagement was observed between cohorts through the CCL pathway (Fig. 2.8E). ALK CTCs exhibited additional interactions with CD8+ T cells in the CCL pathway. Our findings reveal significant activation of the CCL5-CCR5 ($p<0.01$) in CTCs and CD8+ T cells (Fig. 2.8G). As previously discussed, CCL5 has been reported to contribute to the creation of an immunosuppressive tumor microenvironment and increase cancer

migration by activating downstream pathways such as PI3K/AKT in lung cancer [275]. In contrast, wild-type CTCs, CD8+ T cell interact with CTCs through CD99 and PARs pathways. Collectively, those differences in pathway relationships may suggest distinct immune evasion mechanisms in ALK NSCLC. The distinct immune signaling patterns in ALK and wild-type NSCLC's CTCs suggest that tailored therapeutic strategies may be necessary to effectively target the circulatory component of ALK tumors.



ALK CTC Outgoing pathway	
EMT Related	APP, THBS, TGFB, ESAM
Tumor Progression	ICAM, CD99, JAM
Immune Response	CXCL, **SEMA4**, CCL
ALK CTC Incoming pathway	
EMT Related	ITGB2, ESAM
Tumor Progression	ICAM, CD99, **CD34**
Immune Response	**CD40**

Wild-type CTC Outgoing pathway	
EMT Related	APP, THBS, TGFB, ESAM
Tumor Progression	ICAM, CD99, JAM
Immune Response	CXCL, CCL
Wild-type CTC Incoming pathway	
EMT Related	ITGB2, ESAM
Tumor Progression	ICAM, CD99

		ALK	Wild type
EMT	TBHS	Monocytes	Monocytes, CD4+ T, **NK**
	TGFB	**NK**	Monocytes
Tumor Progression	CD99	Monocytes, NK	Monocytes, **CD8+ T, CD4+ T**, NK
	ICAM	Monocytes, NK, CD4+T	Monocytes, NK, **CD8+ T**, CD4+ T
Immune Response	CCL	**CD8+ T**, Basophil	Basophil

Figure 2.8. Overview of cell-cell interaction signaling dynamics in ALK NSCLC and wild type cohort

(A) Global outgoing and incoming signaling pathways across major cell types in ALK patients. (B) Global outgoing and incoming signaling pathways across major cell types in Wild-type patients. (C) Summary table of key outgoing and incoming pathways identified in CTCs of the ALK cohort. (D) Summary table of key outgoing and incoming pathways identified in CTCs of the Wild-type cohort. (E) Differential immune cell engagement in CTC-mediated outgoing signaling pathways between ALK and wild-type cohorts. (F) Distinct TGF-β signaling interaction patterns in ALK (left) compared with wild-type patients (right). (G) Ligand and receptor pair interaction patterns between CTC subgroups and NK or T-cell populations in ALK cohort.

2.5 Discussion

In this study, we provide a comprehensive, high-resolution portrait of CTCs in NSCLC using the microfluidic Labyrinth enrichment platform, with a particular focus on ALK positive patients. Our observations suggest that CTCs may function as potential contributors to metastatic processes, possibly influencing a pro-tumorigenic and immunosuppressive microenvironment. We further describe distinct molecular signatures, phenotypic heterogeneity, and our data support specific intercellular interactions that may underlie tumor aggressiveness in ALK NSCLC and could inform future therapeutic strategies aimed at eliminating cancer cells that survive in circulation.

Consistent with established epidemiological reports, ALK NSCLC patients in our cohort were younger and included a higher proportion of never-smokers compared to wild-type NSCLC patients[276]. The successful application of the Labyrinth microfluidic platform, which demonstrated high CTC recovery and purity, was instrumental for these deep molecular analyses. By ensuring robust CTC detection across all patients, our results confirm and extend earlier reports that microfluidic platforms such as Labyrinth can be reliably deployed in a clinical research context for CTC isolation in lung cancer from progressive and non-progressive patients [131], thereby enabling the transition from enumeration to sophisticated transcriptomic and

genomic characterization.

A key finding is that CTCs exhibit substantial intra- and inter-patient heterogeneity at both protein and transcriptomic levels, with EMT-associated programs emerging as a central driver. Notably, the degree of EMT gene expression varied across patients. Consistently, immunofluorescence staining revealed 4 subtypes of CTCs, and CTCs from ALK patients had a significantly lower proportion of epithelial subtypes (Ep+ CTC) compared to other CTC subtypes ($p \leq 0.0001$). Conversely, mesenchymal-subtype CTCs (Vim+ CTC) tended to be more abundant among patients with progressive disease or a smoking history, though neither comparison reached statistical significance. These observations are hypothesis-generating and suggest that enrichment of Vim+ CTCs may be associated with clinical aggressiveness and could be influenced by prior smoking exposure, which associations that warrant testing in larger cohorts. This observation aligns with, but does not demonstrate causality of, smoking status as it relates to poorer outcomes in ALK patients on first line ALK TKI therapy [277]. This shift toward a mesenchymal phenotype reflects activation of the EMT program, a process recognized for its role in enhancing CTC invasiveness, metastatic potential, and therapy resistance [49]. In fact, this EMT signature strongly aligns with known mechanisms of acquired ALK TKI resistance [278,279], suggesting that monitoring CTCs for mesenchymal markers could serve as a non-invasive biomarker to predict impending therapeutic failure. Collectively, these findings support and extend recent evidence that EMT is particularly pronounced in ALK NSCLC.

Single-cell transcriptomic profiling of ALK NSCLC patients revealed distinct alterations in both immune cell populations and CTCs. In the immune compartment, we identified a progenitor-like population that was present exclusively in ALK patients. This population may reflect enhanced cellular plasticity and adaptive capacity within the ALK-driven immune

landscape. In addition, compared to wild-type, ALK patients exhibited a reduced frequency of peripheral effector $CD8^+$ T cells. This peripheral blood phenotype mirrors reduced effector CD8+ T cell infiltration observed in NSCLC tumor sites [236].

CTCs also displayed a unique transcriptional and genomic signature, further distinguishing ALK from wild-type NSCLC cases. This signature reflects their heterogeneous oncogenic, EMT-related, and platelet-associated phenotypes. ALK CTCs displayed upregulation of NSCLC-associated oncogenic transcripts (PANX1, TIMP1, CXCL5, SMOX), EMT markers (VCL, CLU, GATA1), and platelet-associated genes (PPBP, PF4, ITGA2, CCL5), all of which are implicated in NSCLC pathogenesis, dissemination, and immune evasion. This tripartite signature does not appear to be a random collection of phenotypes. Consistent with this, our data show that ALK CTCs upregulate key genes within the PI3K/AKT, MAPK, and JAK/STAT pathways which are downstream effectors of ALK signaling. These ALK downstream pathways are also established master regulators of EMT programming [272]. These findings suggest a strong mechanistic link: the sustained activation of these ALK-downstream pathways may concurrently drive both the oncogenic state and the observed EMT and platelet-associated phenotypes, thereby enhancing survival in circulation as a required antecedent of hematogenous spread metastases. The marked platelet cloaking of CTCs, confirmed by IF imaging, indicates enhanced crosstalk with the hemostatic system, potentially contributing to immune protection and metastatic competence. Prior studies have demonstrated that platelet–CTC interactions promote immune evasion and enhance metastatic potential by shielding CTCs from shear stress and NK-cell–mediated clearance, while also facilitating survival and extravasation [280,281]. Pathway enrichment analyses corroborated these observations, revealing increased activity in EMT and platelet-interaction pathways as well as oncogenic RAS signaling. Conversely, we also

observed suppression of pathways related to adaptive and innate immunity, including T cell receptor, B cell, and NK cell-mediated cytotoxicity signaling, as well as PD-1/PD-L1 pathway activity within ALK CTCs. This may underlie intrinsic resistance to checkpoint blockade therapy, as has been demonstrated for ALK NSCLC. Previous studies have also reported dysregulation of PD-1 signaling in cancers with ALK fusion, indicating that the ALK oncogene further modulates the tumor immune microenvironment and may contribute to anti-PD-1/PD-L1 resistance.

Moreover, we also observed substantial heterogeneity within ALK CTCs. One subpopulation of higher expression of cancer-associated transcription factors and immune-interaction genes may contribute to immune microenvironment modulation, potentially facilitating immune evasion, and underscores the need to account for CTC heterogeneity in designing precision therapeutic strategies for ALK NSCLC. Collectively, these results suggest that ALK patient-derived CTCs are transcriptionally primed for survival in the circulation, exhibiting features that may enhance metastatic competence and contribute to poorer clinical outcomes.

A pivotal finding of our study is the multifaceted evidence pointing to a potential immunosuppressive environment in ALK NSCLC. We found that a downregulation of T/NK/B cell signaling and apoptosis-related pathways in ALK CTCs further underscores their immune evasiveness and survival capability. Moreover, comparative cell–cell communication analysis suggested that ALK CTCs preferentially direct TGF-β signaling toward NK cells, in contrast to wild-type CTCs, which route this pathway predominantly to monocytes and CD4+ T cells. This NK-skewed engagement is further supported by an enriched repertoire of adhesion- and activation-related ligand–receptor interactions, including ITGB2–CD226 and ADGRE5–CD55.

Given the well-established capacity of TGF-β to suppress NK cell cytotoxicity, downregulate activating receptors such as NKG2D and DNAM-1, and impair IFN-γ production [282,283]. Further, CTCs may manipulate adaptive immunity via the upregulation of CCL5 expression and increase of the CCL5-CCR5 axis in their interactions with CD8+ T cells. This signaling pathway is strongly implicated in fostering an exhausted T-cell phenotype and a suppressive milieu, thereby facilitating immune escape and metastatic progression [275,284]. This underscores the necessity for genotype-specific strategies that can disrupt these immunosuppressive networks to restore anti-tumor immunity.

Together, our findings highlight the profound plasticity and adaptability of ALK CTCs, driven by molecular programs that enhance survival, immune evasion, and metastatic potential. The discovery of ALK CTC–platelet and ALK CTC–immune cell crosstalk suggests the potential for highly targeted therapeutic interventions. Although traditional anti-platelet agents have been extensively studied, their efficacy in preventing cancer metastasis or improving overall survival in the context of advanced disease remains controversial or limited [285–287]. Consequently, there is a need for advanced strategies that specifically disrupt immune and platelet molecular interactions. For instance, our data suggest that co-administration of targeted agents, such as dual blockade of TGF-β or CCR5 alongside ICIs, could be explored as a strategy to dismantle this protective shield and reduce metastatic seeding. Importantly, the observed heterogeneity both within and between patients emphasizes the necessity of personalized approaches for disease monitoring and treatment. Longitudinal tracking of CTC phenotype and interactome evolution may assist in guiding therapy selection and anticipating mechanisms of resistance.

Although these findings are promising, our study is limited by a modest sample size,

highlighting the need for validation in larger, well-controlled cohorts to establish the utility of CTC phenotypes, including mesenchymal subtype, platelet-coated, and EMT-associated subsets, as dynamic liquid biopsy biomarkers for disease monitoring and therapeutic response in ALK NSCLC. Another limitation is the lack of validation of the identified gene and pathway signatures in matched primary tumor tissue, which would clarify their biological relevance and roles in ALK-driven tumor progression. However, our GSEA analysis revealing strong enrichment of lung epithelial signatures provides confidence in the lung tissue origin of the identified CTCs. Finally, as with all scRNA-seq studies, potential gene dropouts and the imperfect correlation between mRNA and protein levels must be considered. Future studies should focus on confirming these molecular signatures in prospective patient cohorts and experimentally dissecting the functional consequences of key CTC–immune interactions, particularly those that contribute to immune evasion and metastatic potential. We emphasize that all ligand–receptor relationships reported here are computationally inferred from transcriptomic co-expression and represent predicted, rather than experimentally validated, interactions; functional confirmation (e.g., NK cytotoxicity and CD8+ T-cell functional assays) will be required to establish their causal roles.

In conclusion, this study provides a comprehensive single-cell characterization of ALK NSCLC CTCs, highlighting their molecular complexity and heterogeneity, including EMT- and platelet-associated phenotypes, oncogenic and survival signaling, and features that induce immune resistance. These insights establish a foundation for translational research aimed at leveraging CTCs as both biomarkers and therapeutic targets in advanced ALK lung cancer.

2.6 Methods:

2.6.1 Study design and patient cohort

This study was approved by the Institutional Review Board (IRB) at the University of Michigan (IRB Protocol # HUM00227562). All participants provided informed consent. Peripheral blood samples were collected from a total of 91 patients with NSCLC, comprising 45 patients with confirmed ALK rearrangements and 46 patients with wild-type (without ALK and EGFR mutation) status. Clinical data were recorded at the time of sample collection. The clinical response of stable disease and progressive disease is determined by the treating physician at the time of blood draw. For each patient, one tube of blood sample was collected in EDTA tubes (BD Vacutainer, Cat# 367863) for processing within 12 hours or in Cell Rescue tubes (Menarini Silicon Biosystems Inc, Cat#84000186) for longer time points and were transported at room temperature.

2.6.2 Cell line culture and validation

NSCLC cell lines (H2228, RRID:CVCL_1543) and in-house CTC derived cell lines were used to validate the methods before processing patient samples. Cells were cultured in Roswell Park Memorial Institute 1640 Medium (RPMI 1640, Gibco) supplemented with 10% fetal bovine serum (Sigma), and 1% antibiotic-antimycotic (Gibco). Cells were maintained at 37 °C with 5 % CO2 incubator. Mycoplasma testing was performed using the mycoalert detection kit (Lonza, Cat# LT07-318), to ensure they are mycoplasma negative.

2.6.3 Microfluidic CTC enrichment

Device Fabrication: The Labyrinth microfluidic devices were fabricated in-house using standard soft lithography. A silicon master mold with a channel height of 70 μm was created using deep reactive-ion etching (DRIE). Polydimethylsiloxane (PDMS) and curing agent (Sylgard 184, Dow Corning) were mixed at a 10:1 ratio, poured over the mold, and cured

overnight at 70°C. The cured PDMS layer was then peeled, and inlet/outlet ports were punched. The labyrinth device was irreversibly bonded to a 75 × 50 mm glass slide (Corning) using oxygen plasma treatment, followed by a 10-minute bake at 90°C to ensure a permanent seal.

Device Validation: To quantify performance, 200–1000 cells of in-house NSCLC CTC derived cells lines were spiked into 1 mL of healthy donor blood. Samples were processed through the Labyrinth device, and the contents of all four outlets were collected. The tumor cell recovery rate was calculated as follows:

$$\frac{CTC\ collected\ in\ Outlets\ 2\ and\ 3}{CTC\ collected\ in\ all\ 4\ Outlets}$$

The WBC depletion rate is calculated as

$$1 - \frac{WBC\ concentration\ in\ Outlets\ 2\ and\ 3}{WBC\ concentration\ from\ input}$$

Patient Sample Processing: Whole blood from patients was diluted in PBS with a 1:1 ratio and processed using Ficoll (Cytiva) to isolate the peripheral blood mononuclear cell (PBMC) and tumor cell layer. This collected buffy coat layer was further diluted in PBS with a 1:5 ratio. The Labyrinth was primed with pluronic-F127 (Sigma-Aldrich, P2443). The cell suspension was infused into the Labyrinth device at a flow rate of 2000 uL/min using a syringe pump. The enriched CTCs was collected from outlets 2 and 3.

2.6.4 Immunofluorescence (IF) staining and imaging

Staining: Enriched cell fractions were deposited onto poly-L-lysine-coated slides using a Cytospin Cytocentrifuge (800 rpm, 10 min; Thermo Fisher Scientific). Slides were fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100 or ice-cold 100% methanol. Then the slide is washed by PBS and blocked with 3% BSA for 1 hour at room temperature.

The slides were then stained with primary antibodies with 1% BSA with the following panels:

EMT Phenotyping Panel: anti-Pan-Cytokeratin (clone AE1/AE3, 1:100, Cell Signaling, 67306, RRID:AB_3662123), anti-CD45 (1:100, Bio-Rad, MCA87, RRID:AB_871979), biotinylated anti-EpCAM (1:10, R&D Systems, BAF960, RRID:AB_356818), and anti-Vimentin (1:100, Abcam, ab137321, RRID:AB_2921312).

Platelet Interaction Panel: anti-Pan-Cytokeratin (1:100, as above), anti-CD45 (1:100, as above), and anti-CD41/61 (1:100, 1:100, Santa Cruz Biotechnology, sc-365938, sc-365679, RRID: AB_10947393, AB_10844835).

ALK Panel: anti-Pan-Cytokeratin (1:100, Agilent, M3515, RRID:AB_2132885), anti-CD45 (1:100, as above), and anti-ALK D5F3 (1:100, Cell Signaling, 3633, RRID:AB_11127207).

Then the slides are washed in PBS and incubated in 4°C overnight. The following day, slides were washed in PBS and secondary antibody cocktails with 1% BSA was applied.

Secondary antibody panel was goat anti-mouse IgG1 Alexa Fluor 546 (1:100, Thermo Fisher, A21123, RRID: AB_2535765), goat anti-mouse IgG2a Alexa Fluor 488 (1:100, Thermo Fisher, A21131, RRID: AB_2535771), streptavidin Alexa Fluor 750 (1:100, Thermo Fisher, S21384), and goat anti-rabbit IgG Alexa Fluor 647 (1:100, Thermo Fisher, A21245, RRID: AB_2535813). Slides were mounted with ProLong Gold Antifade Mountant with DAPI (Thermo Fisher, P36931).

Image: The stained slides were imaged using Nikon TI fluorescent microscope at a 20X magnification. The images were used for enumeration and classification of distinct CTC subtypes through identification of positive marker of Pan-cytokeratin (PanCK), as well as

negative in leukocyte marker CD45. Images were manually reviewed to enumerate and classify CTCs into Ep+ type (CK+/EpCAM+/Vim-/CD45-), Vim+ (CK+/EpCAM-/Vim+/CD45-), and all+ (CK+/EpCAM+/Vim+/CD45-) subtypes based on their fluorescence profiles. Platelets were identified as (CD41/61+, DAPI-).

2.6.5 ALK rearrangement validation

Digital polymerase chain reaction (dPCR): RNA was extracted from enriched CTC samples using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Cat# 80224). cDNA was synthesized and analyzed on a QuantStudio Absolute Q Digital PCR System (Thermo Fisher). The reaction mix included Absolute Q DNA Digital PCR Master Mix, an EML4-ALK Probe-Primer Mix (EML4ALK-v1/v3a/v3b/v3_20/v5; GenePath Cat#9113-201 A-101-012) and a Cytokeratin 8 (CK8) probe set (TaqMan Gene Expression Assay, Thermo Fisher, Assay ID: Hs01595539_g1). Nuclease-free water and white blood cells processed through the Labyrinth microfluidic device were used as negative controls to ensure assay specificity.

Fluorescence In Situ Hybridization (FISH): Enriched CTCs were fixed onto slides using the sample preparation protocol described in IF staining section and analyzed using the Vysis ALK Break Apart FISH Probe Kit (Abbott Molecular, Cat# 08N35-020) according to the manufacturer's protocol.

2.6.6 Single-cell RNA sequencing (scRNA-seq)

Library Preparation and Sequencing: The enriched CTC product was counted to reach 100K cells and processed using the 10xGenomics Chromium Next GEM Single Cell 3' Gene Expression Flex kit (10xGenomics, Cat# 1000286) according to the manufacturer's protocol. Library was generated and sequenced at the University of Michigan Advanced Genomics Core on an Illumina NovaSeq platform to a target depth of ~10,000 (patient ID 8, 20, and 77) and

~50,000 reads (patient ID 6, 14, 15, 54) per cell. Raw sequencing data was processed using Cell Ranger (10xGenomics, RRID:SCR_017344) to generate the feature-barcode count matrix.

2.6.7 Bioinformatic analysis of scRNA-seq Data

Seed Setting: All data was analyzed under set.seed(0) in R (RRID:SCR_001905).

Preprocessing and Clustering: The analysis was performed using the Seurat package (v5.0.0, RRID:SCR_016341). Cells with fewer than 200 detected genes, and genes detected in fewer than 3 cells, were excluded. Raw count data were normalized using the LogNormalize method with a scale factor of 10,000, followed by scaling with ScaleData function. Data integration across samples was performed using Seurat anchor-based integration workflow (FindIntegrationAnchors and IntegrateData functions) based on the top 50 canonical correlation dimensions. The integrated dataset was then clustered using UMAP embedding with the first 50 principal components and a clustering resolution of 2.

Cell Type Annotation: Clusters were annotated using the scType (RRID:SCR_026634) and confirmed via expression of canonical marker genes [252]. CTC validation was performed by selecting cells lacking pan-leukocyte or immune lineage marker expression (e.g., PTPRC, NKG7, GNLY, TRBC2, CD3D, CD3G, TRAC, CD63, HDC, CLC, VCAN, CD14, LYZ, S100A8, S100A9, CD19, FGFBP2, CLEC10A, CD83) and simultaneously expressing tumor-cell–associated markers (KRT8, KRT5, CDH1, CD44, EPCAM, PF4, PPBP, TIMP1, GATA1, MYL9). A Progenitor-like Subpopulation was identified as a distinct cluster following initial unsupervised clustering.

CNV: CNV was inferred using the inferCNV package (v1.20.0, RRID:SCR_021140). Raw UMI count matrices were extracted from the Seurat object and reordered to match cell barcodes. Gene ordering was defined using GENCODE v27 (hg38) coordinates. Immune and

hematopoietic lineages (B cells, Basophils, Dendritic cells, Erythroid-like and erythroid precursor cells, Monocytes, Neutrophils, NK cells, NKT cells, Platelets, Progenitor cells, and T cells) served as reference groups. The CreateInfercnvObject and infercnv::run functions were used with parameters optimized for 10xGenomics data (cutoff = 0.1, cluster_by_groups = TRUE, denoise = TRUE, HMM = TRUE, and write_expr_matrix = TRUE).

Differential Gene Expression: Differential gene expression between cell populations was performed using the Deseq2 (v1.44.0, RRID:SCR_015687) and dplyr (v 1.1.4, RRID:SCR_016708) packages. For gene upregulated in the ALK CTC compared to the rest of cells within ALK cohort, cells were classified as either "CTC" or "Rest" based on cell annotation. Counts were pseudo-bulked for each patient and group combination, and DESeq2 was applied with a paired design controlling for patient effects (design = ~ patient + group) to identify genes differentially expressed between CTCs and the rest of the cells. For the CTC in ALK vs. wildtype comparison, gene counts from all CTCs belonging to each patient were aggregated into pseudo-bulk samples. Patients were labeled by their clinical condition ("ALK" or "Wildtype"), and differential expression between the two groups was analyzed using DESeq2.

Pathway Analysis: GSEA and enrichment test were conducted using the fgsea (v1.30.0, RRID:SCR_020938) and clusterProfiler (v4.12.0, RRID:SCR_016884) R packages, respectively, against the Hallmark (h.all.v2023), KEGG, and GO:BP gene sets from the Molecular Signatures Database (MSigDB, RRID:SCR_016863). For the GSEA analysis, an adjusted p-value cutoff of 0.05 was applied, and gene was ranked using baed on the log2FoldChange. For the enrichment analysis, the Benjamini–Hochberg (BH) method was used for *p*-value adjustment, with a *p*-value cutoff of 0.05 and a *q*-value cutoff of 0.2. Significant pathways related to immune interaction, platelet, oncogenic and inflammation are selected to plot.

Volcano Plot: Significant genes were visualized using the EnhancedVolcano (v1.22.0, RRID:SCR_018931) package, with an adjusted p-value threshold of 0.05 and an absolute fold-change cutoff of 1.5 for ALK CTC compared to rest of gene in ALK cohort's DGE. P-value threshold of 0.05 and an absolute fold-change cutoff of 1.5 for ALK CTC compared to CTC in wild-type cohort's DGE

Heatmap: Gene expression heatmaps were using the ComplexHeatmap (v2.20.0, RRID:SCR_017270), and circlize (v0.4.16, RRID: SCR_002141) packages. Genes were z-scored on the only CTC expression in ALK cohort. Custom gene sets were curated to represent key biological categories, including epithelial, mesenchymal, epithelial–mesenchymal transition (EMT), stemness, immune interaction, migration, hypoxia, platelet, NSCLC-related, and housekeeping genes. Expression data for these genes were extracted from the Seurat object, log-normalized, and standardized as z-scores across cells. The heatmap was visualized with genes ordered by category, and expression values were color-coded from blue (low) to red (high).

Cell-Cell Communication Analysis: Intercellular signaling networks were inferred using the CellChat (v1.6.1, RRID:SCR_021946) with the CellChatDB.human ligand-receptor database.

2.6.8 Statistical analysis

CTC counts were collected to analyze the correlations between patient clinical information and CTC burden. GraphPad Prism 10 (GraphPad Software, Inc., RRID:SCR_002798) was used to conduct the statistical test and generate figures. The data was treated as unpaired when comparing between ALK and wild-type cohort. When comparing the CTC subtype within ALK cohort. Continuous variables were assumed to be non-parametric and are presented as median with interquartile range (IQR). Comparisons between unpaired data

were conducted using the two-sided Mann-Whitney U test, and between paired data using Friedman test. Comparisons of paired data within the same group were performed using the Wilcoxon signed-rank test. Categorical variables were analyzed using the Chi-square test or Fisher's exact test, as appropriate. A p-value of less than 0.05 was considered statistically significant.

2.7 Author Contributions

I was responsible for the conceptualization and methodology of this work, conducted patient sample processing and wet-lab experiments, performed the data analysis, and wrote the original draft of the manuscript.

Chapter 3 Longitudinal Circulating Tumor Cell Dynamics and Single-Cell Transcriptomic Evolution as Prognostic Biomarkers in ALK-Positive Non-Small Cell Lung Cancer

3.1 Abstract

Whether the molecular features of ALK CTCs identified in Chapter 2 carry prognostic information that evolves under sequential TKI pressure remains unknown. To address this, longitudinal monitoring of CTCs was performed in 12 ALK NSCLC patients across 56 serial visits (median follow-up 107 weeks), combining multi-marker immunofluorescence with single-cell RNA sequencing on a subset of six patients spanning treatment initiation, response, and progression. Detailed single-cell results are presented for three patients with complete serial coverage. Four CTC subpopulations, epithelial (CK+), EpCAM-positive (Ep+), mesenchymal (Vim+), and hybrid epithelial-mesenchymal (All+), were enumerated at each visit and evaluated against overall and progression-free survival using both per-patient mean absolute counts and inter-visit trajectories, with single-timepoint counts assessed as a static benchmark. Static enumeration was a weak prognostic discriminator, whereas dynamic CTC composition was strongly informative. Mean CK+ burden across serial visits was the strongest predictor of mortality, and trajectories of CK+, combined Vim+/CK+/All+, and total CTC counts each stratified OS (all $p = 0.044$), whereas a decreasing normalized All^+ fraction predicted longer progression-free survival ($p = 0.013$). Longitudinal scRNA-seq revealed sustained EMT-associated transcriptional reprogramming, persistent GATA2 downregulation, and progression-emergent CDK4 copy number gain consistent with established bypass mechanisms of ALK TKI resistance. A 45-gene downregulated CTC signature derived from these longitudinal data was

independently associated with shorter overall survival in the TCGA-LUAD validation cohort (n = 500; p = 0.039). Together, these findings establish longitudinal CTC profiling as a clinically actionable prognostic tool capable of capturing therapeutic adaptation in ALK NSCLC in real time.

3.2 Resulting Publication Information

Yuru Chen et al. "Longitudinal Circulating Tumor Cell Dynamics and Single-Cell Transcriptomic Evolution as Prognostic Biomarkers in ALK-Positive Non-Small Cell Lung Cancer" (In preparation)

3.3 Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for 87% of lung cancer deaths [2]. Anaplastic lymphoma kinase-positive NSCLC (ALK NSCLC) arises from chromosomal rearrangements, most commonly between EML4 and ALK, occurring in approximately 3–7% of NSCLC cases [2]. The resulting oncogenic fusion exhibits aberrant subcellular localization and constitutive kinase activation, driving persistent PI3K/AKT, MAPK/ERK, and JAK/STAT signaling and promoting cell proliferation, survival, and migration. The development of ALK tyrosine kinase inhibitors (TKIs) has substantially improved progression-free survival (PFS) and overall survival (OS) compared to conventional chemotherapy [18,21,29]. Therapeutic resistance, however, eventually develops in nearly all patients. Sequential TKI therapy progressively narrows treatment options as each line selects for increasingly complex resistance profiles, with distinct on-target ALK mutations and off-target bypass mechanisms emerging in a treatment-dependent manner [20,22]. While these genomic resistance mechanisms, including secondary ALK kinase domain mutations (e.g., p.G1202R) and bypass signaling activation, are well-characterized at the

DNA level [23], the longitudinal evolution of tumor cell phenotype and transcriptomic state under sequential TKI pressure, and their relationship to clinical outcomes, remain poorly understood.

Resolving this requires serial monitoring strategies capable of capturing not only mutational changes but also the phenotypic and transcriptomic dynamics that accompany therapeutic adaptation. Serial tissue re-biopsy has provided the most direct mechanistic insights to date. A prior study characterized acquired resistance across multiple TKI lines using paired pre- and post-progression tumor biopsies, identifying ALK kinase domain mutations in 54%, 53%, and 71% of patients who progressed on ceritinib, alectinib, and brigatinib, respectively, with distinct mutation spectra per agent [28]. However, repeat tissue biopsy is feasible in only a subset of patients due to lesion accessibility, procedural risk, and tissue adequacy, and captures only a single lesion per timepoint, inherently missing clonal heterogeneity across spatially distinct tumor sites, and multi-timepoint sampling needed to track transcriptomic evolution across the treatment course.

Serial plasma ctDNA profiling provides a less invasive approach for repeated genomic sampling during ALK TKI therapy. In ALK NSCLC, ctDNA clearance at 2 months was associated with longer PFS and OS, while serial ctDNA analysis in the eXalt2 ensartinib study tracked response dynamics and emerging ALK resistance alterations over time [288–290]. These studies establish ctDNA as a practical tool for non-invasive genomic surveillance. However, ctDNA is inherently restricted to DNA-level information: as a cell-free, fragmented analyte, it captures mutational and copy number alterations but provides no readout of gene expression, protein activity, or cellular phenotype.

This distinction matters because resistance in NSCLC is not solely a genomic event.

Transcriptomic and proteomic analyses of tumor tissue have demonstrated that a substantial component of TKI resistance is mediated by phenotypic reprogramming that operates independently of, or in parallel with, acquired DNA mutations: epithelial-mesenchymal transition (EMT)-associated transcriptional shifts, YAP/TEAD-driven adaptive plasticity, AXL upregulation, and the emergence of drug-tolerant persister (DTP) cell states with stem-like and mesenchymal gene expression programs [291–296]. In ALK NSCLC specifically, scRNA-seq of tumor organoids has identified pre-existing DTP subpopulations with transcriptional profiles distinct from bulk tumor cells that confer early drug insensitivity and seed acquired resistance [297]. These resistance programs cannot be detected by ctDNA-based monitoring and are not reliably inferred from mutational profiles alone, rendering ctDNA surveillance insufficient to capture the full landscape of therapeutic adaptation. While tissue-based transcriptomic profiling can reveal these phenotypic resistance programs, it requires invasive biopsy and cannot feasibly be repeated across multiple treatment timepoints. CTCs, by contrast, can be collected serially from peripheral blood, providing a non-invasive and repeatable window into the transcriptomic landscape of live tumor cells throughout the disease course.

CTCs offer a complementary liquid biopsy fraction that can access this phenotypic information non-invasively. As intact tumor cells shed from primary or metastatic lesions, CTCs preserve cellular DNA, RNA, and protein/surface-marker information, enabling genomic, transcriptomic, and phenotypic characterization from a single blood draw [298]. Despite this potential, CTC monitoring in NSCLC has been hampered by several limitations. First, conventional enrichment technologies might miss malignant CTC population. For example, FDA-cleared CellSearch platform, which enriches epithelial CTCs through EpCAM-based capture, might miss the CTC undergoing EMT [111]. It is a subpopulation characterized by loss

of epithelial markers and associated with enhanced metastatic potential and therapeutic resistance [13]. In ALK NSCLC specifically, study demonstrated this phenotypic complexity directly: among CTCs isolated at resistance to ALK TKIs, 37% were ALK-positive/cytokeratin-negative and 56% were ALK-negative/cytokeratin-positive [299], indicating that EpCAM or cytokeratin-based capture would systematically miss or misclassify the majority of the CTC population. Second, most longitudinal CTC studies have been restricted to enumeration or copy number variation of the tumor cells [300]. While CTC count dynamics can associate with clinical outcomes, enumeration treats CTCs as a homogeneous population and cannot resolve the molecular heterogeneity underlying therapeutic failure. Third, bulk molecular profiling of pooled CTCs obscures intratumoral heterogeneity, which is a defining feature of acquired resistance.

Single-cell sequencing approaches have begun to address the third limitation by resolving heterogeneity at the level of individual CTCs. Study sequenced individual CTCs from ALK NSCLC patients at progression on crizotinib (n = 14) or lorlatinib (n = 3), revealing not only ALK kinase domain mutations but also recurrent ALK-independent alterations in the RTK-KRAS and TP53 pathways in crizotinib-resistant CTCs [211]. This demonstrating that resistance is genetically heterogeneous across individual CTCs within the same patient. More recently, Single-cell analyses of ALK NSCLC CTCs have revealed genomic and copy-number heterogeneity, while scRNA-seq studies of ALK lung cancer organoid models have identified transcriptionally distinct DTP subpopulations [299]. However, these studies were conducted at single post-progression timepoints. Prospective, longitudinal scRNA-seq characterization of CTCs across defined clinical milestones, treatment initiation, radiological response, and disease progression, to our knowledge has not been reported in ALK NSCLC, leaving the transcriptomic trajectory of tumor cell phenotype under sequential TKI pressure uncharacterized in any liquid

biopsy modality.

To address these gaps, our study integrates three methodological advances corresponding to the limitations described above. First, to overcome EpCAM-based enrichment bias, we employed the Labyrinth, a label-free microfluidic platform developed in our laboratory that unbiasedly captures the full spectrum of CTCs [131]. Furthermore, we applied multi-marker immunofluorescence to longitudinally track phenotypic evolution of CTCs under TKI pressure across serial timepoints, moving beyond enumeration to characterize the dynamic composition of the CTC population. Third, we performed scRNA-seq on matched CTC samples at multiple disease milestones to resolve transcriptomic reprogramming at single-cell resolution.

Together, these approaches enabled an integrated longitudinal analysis combining CTC enumeration, phenotyping, survival modeling, and scRNA-seq in ALK NSCLC patients. We hypothesized that the longitudinal trajectory of CTC phenotypic composition, rather than static cellular burden, together with transcriptomic changes, carries prognostic information. Our findings reveal that a tumor's adaptive capacity, measured by CTC progression rate and EMT-associated phenotypic shifts, predicts patient survival more strongly than static cellular burden, and that this adaptation is driven by specific transcriptomic reprogramming identifiable at the single-cell level.

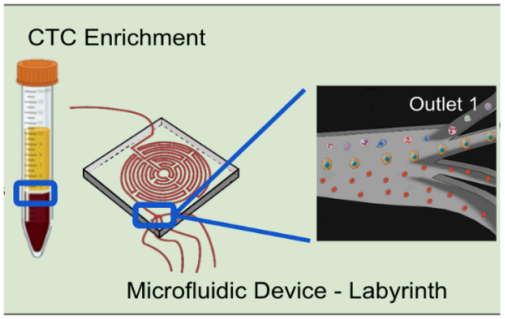
3.4 Results

3.4.1 Cohort characteristics and CTC profiling

Patients in this study were adults recruited and consented as part of the ALK Lung Cancer Research Initiative at University of Michigan. Twelve patients with ALK NSCLC were enrolled in this longitudinal study, contributing a total of 56 blood samples (2–10 draws per patient; Figure 3.1C). The median follow-up duration was 107 weeks. During the follow-up

period, five patients (41.7%) died, while seven remained alive. Treatment regimens within the cohort included alectinib monotherapy, lorlatinib, and multi-agent sequential combinations. The two groups were broadly similar in age (median 54 vs. 59 years) and smoking history, though 3 of 5 deceased patients were male compared with 1 of 7 surviving patients, and all deceased patients were Stage IV at enrollment. Notably, a higher proportion of alive patients were on TKI therapy at sample collection (82.8% vs. 40.7%), and alive patients had a greater fraction of stable disease assessments (37.9% vs. 22.2%), consistent with better treatment response. CTC enumeration performed using the Labyrinth microfluidic device (Figure 3.1A), demonstrated successful label-free isolation across all patients.

A



B

	Deceased N=5 (%)	Alive N=7 (%)
Characteristic		
Sex (Male, Female)	3, 2	1, 6
Age (Median, Range)	54, 39-73	59, 30-78
Ethnicity		
White	5 (100.0%)	5 (71.4%)
Non-White	0 (0.0%)	2 (28.6%)
Smoking History		
Never Smoker	3 (60.0%)	4 (57.1%)
Former Smoker, Quit	2 (40.0%)	3 (42.9%)
Current Smoker	0 (0.0%)	0 (0.0%)
Cancer Stage		
Stage 3	0 (0.0%)	1 (14.3%)
Stage 4	5 (100.0%)	6 (85.7%)
Sample Collection		
Total Sample	27	29
On TKI Treatment		
Yes	11 (40.7%)	24 (82.8%)
No	16 (59.3%)	5 (17.2%)
Clinical Assessment		
Response	6 (22.2%)	6 (20.7%)
Stable	6 (22.2%)	11 (37.9%)
Progression	12 (44.4%)	4 (13.8%)

C

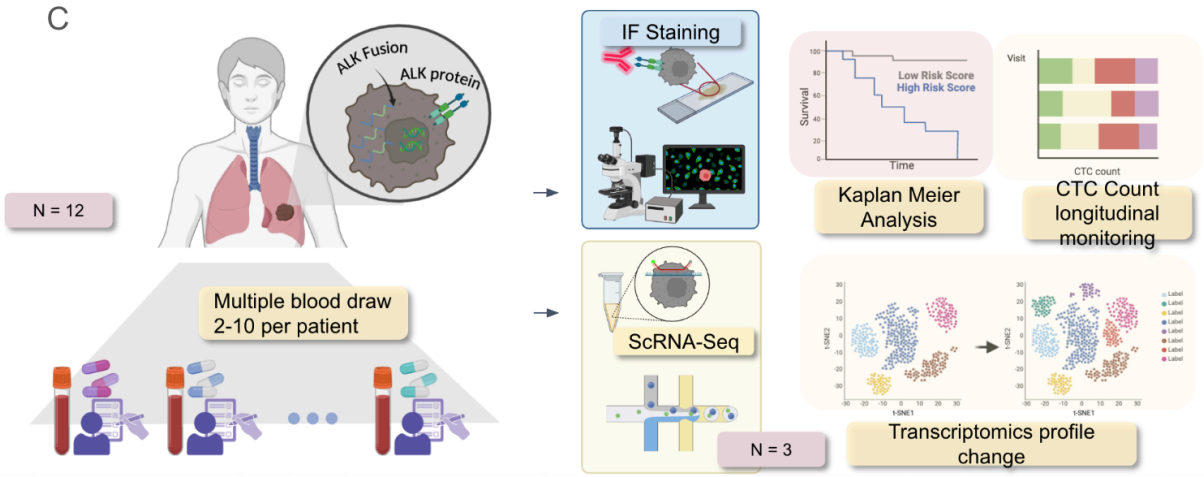


Figure 3.1. Study design and patient characteristics.

(A) Schematic of the Labyrinth microfluidic device illustrating label-free, size-based CTC enrichment from peripheral blood, with inset depicting cell separation at the device outlet. (B) Patient demographic and clinical characteristics stratified by survival status (Deceased, n=5; Alive, n=7), including sex, age, ethnicity, smoking history, cancer stage, total samples collected, TKI treatment status at collection, and radiological assessment distribution. (C) Overview of the longitudinal study design. Twelve ALK NSCLC patients underwent serial blood collection (2–10 draws per patient; total n=56 samples). Enriched CTCs were processed in parallel for multi-marker immunofluorescence (IF) phenotyping across all patients and for single-cell RNA sequencing (scRNA-seq; n=3 patients, 7 longitudinal timepoints). Downstream analyses included Kaplan-Meier survival modeling, longitudinal CTC phenotype tracking, and transcriptomic profiling.

3.4.2 Longitudinal CTC dynamics and clinical outcomes

Phenotypic characterization via multi-marker immunofluorescence revealed a diverse distribution of cellular states. CTCs were categorized as CK+ (epithelial), Vim+ (mesenchymal), Ep+ (EpCAM positive), or All+ (hybrid epithelial-mesenchymal, co-expressing cytokeratin, vimentin, and EpCAM) (Figure 3.2A). Cross-sectional analysis of baseline (V1) samples revealed that total CTC counts did not significantly differ between deceased and surviving patients ($p=0.2020$).

To characterize the longitudinal dynamics of CTC burden, total counts and phenotypic proportions were tracked across all serial visits for each patient (Figure 3.2C). In contrast to total CTC counts, the absolute count of CK+ CTCs was significantly elevated in deceased patients compared to those who remained alive ($p=0.0051$; Figure 3.2B), suggesting that the total CTC burden alone is insufficient as a prognostic discriminator. Analysis of additional subpopulations confirmed that CK+Ep+ co-positive CTCs ($p=0.0225$) and CK+Ep+EMT triple-positive CTCs ($p=0.0303$) were also significantly higher in deceased patients at baseline, while Vim+, Ep+,

All+, and EMT absolute counts individually did not reach significance. Proportional analyses (% of total CTCs) were non-significant for all subtypes at baseline, indicating that the prognostic signal resides in absolute subpopulation burden rather than phenotypic composition.

At radiologically confirmed progression timepoints, no CTC subtype, by absolute count or proportion, differed significantly from non-progression timepoints (all p>0.20). These findings suggest that absolute CTC count at any single timepoint is primarily a signal of generalized treatment response rather than a progression-specific indicator, and that longitudinal dynamics are more informative than single-timepoint measurements.

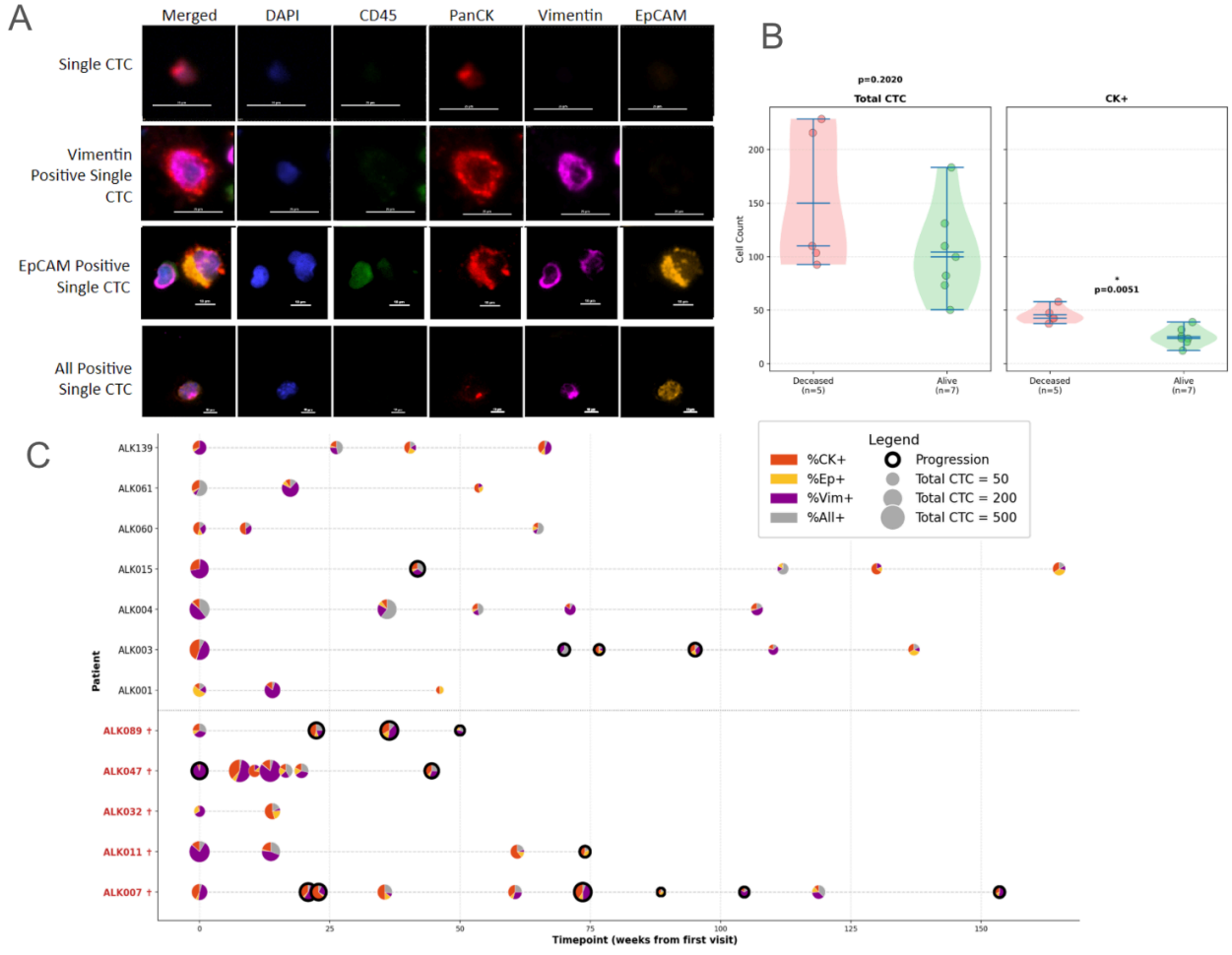


Figure 3.2. Multi-marker immunofluorescence phenotyping of CTCs and longitudinal burden tracking.

(A) Representative IF images of single CTCs classified by phenotypic subtype: marker-negative single CTC (DAPI+/CD45−/PanCK−/Vimentin−/EpCAM−), Vimentin-positive CTC (Vim+), EpCAM-positive CTC (Ep+), and triple-positive hybrid CTC (All+; co-expressing cytokeratin, vimentin, and EpCAM). Channels shown: Merged, DAPI, CD45, PanCK, Vimentin, and EpCAM. Scale bars as indicated. (B) Violin plots comparing absolute CTC subtype counts at the first visit (V1) between Deceased (n=5) and Alive (n=7) patients. Total CTC counts did not differ significantly between groups (p=0.2020), whereas absolute CK+ CTC counts were significantly elevated in deceased patients (p=0.0051). (C) Longitudinal CTC bubble plot for all 12 patients over time (weeks from first visit). Bubble size represents total CTC count; pie chart segments indicate proportions of CK+, Ep+, Vim+, and All+ subtypes at each timepoint. Open circles mark radiologically confirmed progression events. Deceased patients are highlighted in red and annotated with †.

3.4.3 Prognostic value of CTC phenotypic change

A central question motivating this analysis was whether individual cross-sectional CTC measurements captured at discrete clinical events could serve as reliable prognostic indicators. To address this directly, absolute All+ CTC counts were compared between radiologically confirmed progression timepoints (n=16) and non-progression timepoints (n=40) across the cohort. No significant difference in All+ CTC burden was detected between these two states (Mann-Whitney U p=0.2306; Figure 3.3E), a finding consistent with the absence of significant cross-sectional differences observed for all other CTC subtypes at progression versus non-progression visits (all p>0.20). These results collectively demonstrate that static, single-timepoint CTC enumeration, regardless of phenotypic subtype, lacks the resolution to discriminate clinically distinct disease states, underscoring the fundamental limitation of cross-sectional liquid biopsy measurements as standalone prognostic instruments.

In contrast, inter-visit CTC dynamics, quantified as the normalized change in phenotypic burden between the first and last serial visits, demonstrated robust and consistent associations with patient survival. To characterize these longitudinal trajectories, patients were dichotomized

into groups exhibiting lesser versus greater decrease based on the normalized rate of change for each CTC subtype. The mean absolute CK+ CTC count across all serial visits, representing the integrated epithelial tumor cell burden over the entire follow-up period, was the strongest single discriminator of mortality (AUROC = 0.971; Figure 3.3A); this estimate should be interpreted with caution given the cohort size (n=12, 5 events). Kaplan–Meier analysis showed that patients with a greater decrease in total CTC count between the first and last visits experienced longer OS ($p = 0.044$; Figure 3.3D), as did those with a greater decrease in absolute CK+ count ($p = 0.044$; median cutoff: −32; Figure 3.3B), and the Vim+, CK+, and All+ subpopulations ($p = 0.044$; median cutoff: −95.7; Figure 3.3C). These three overall-survival comparisons used the absolute inter-visit difference in CTC burden ($X_{last} - X_{first}$; defined in Methods section). After Benjamini–Hochberg correction across the survival analyses, they did not retain significance ($q = 0.42$) and are therefore reported as nominal, exploratory associations. Their consistent direction across independent burden metrics nonetheless suggests that the direction and magnitude of change in CTC phenotypic burden across serial visits, rather than any single cross-sectional measurement, may constitute the clinically informative prognostic signal.

The prognostic relevance of longitudinal CTC dynamics extended beyond total burden to phenotypic composition. Kaplan-Meier analysis of inter-visit dynamics demonstrated that patients with a greater decrease in the normalized All+ fraction across the stable-to-progression transition experienced significantly longer PFS ($p=0.013$; median cutoff: −0.072; Figure 3.3F). The peak All+ proportion across all serial visits was, in parallel, significantly higher in surviving than in deceased patients (51.8% vs. 31.5%; AUROC = 0.857; $p = 0.048$).

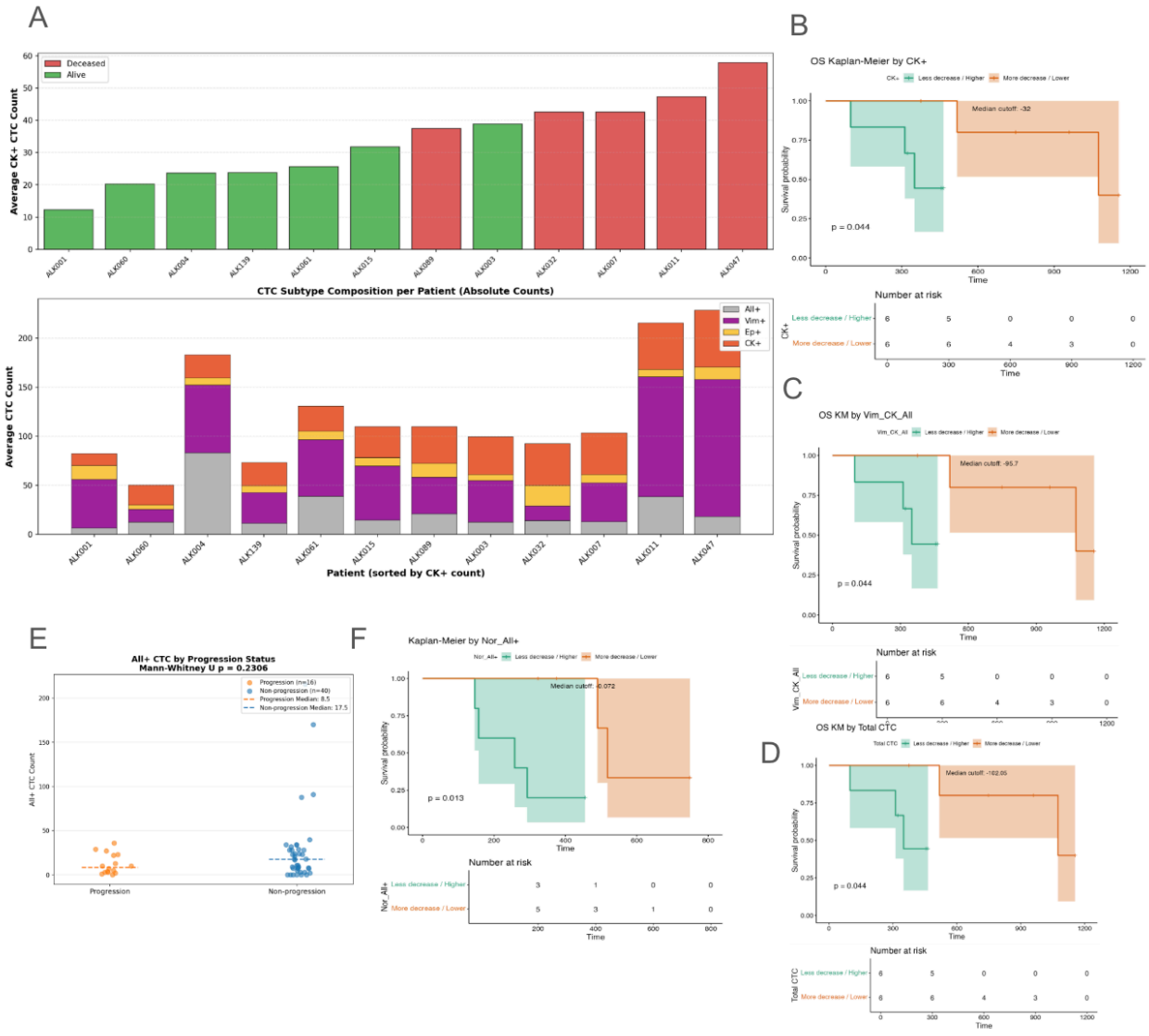


Figure 3.3. Longitudinal CTC subtype burden and inter-visit trajectory as prognostic determinants of survival in ALK NSCLC.

(A) Mean absolute CK+ CTC count per patient across all serial visits (upper panel), stratified by survival status (Deceased, red; Alive, green), sorted by ascending CK+ burden. Stacked bar charts of mean absolute CTC subtype counts (CK+, Ep+, Vim+, All+) per patient across all visits (lower panel), sorted by ascending CK+ count. (B) OS stratified by CK+ count trajectory (p=0.044; median cutoff: −32); patients with lesser CTC decrease or sustained higher burden (teal) demonstrated significantly worse OS compared to those with greater CTC decrease (orange). (C) OS stratified by combined Vim+/CK+/All+ trajectory (p=0.044; median cutoff: −95.7). (D) OS

stratified by total CTC count trajectory (p=0.044; median cutoff: −102.05). (E) Cross-sectional comparison of absolute All+ CTC counts at individual progression (n=16) versus non-progression (n=40) timepoints; no significant difference was observed between groups (Mann-Whitney U p=0.2306), indicating that single-timepoint measurements lack prognostic resolution. (F) Progression-free survival (PFS) stratified by normalized All+ fraction rate of change between first and last visits (Nor_All+; p=0.013; median cutoff: −0.072); patients with a greater decrease in normalized All+ fraction (orange; 'More decrease/Lower') demonstrated significantly longer PFS compared to those with a lesser decrease / higher fraction (teal).

3.4.4 Longitudinal single-cell transcriptomic profiling of CTCs

To investigate the molecular programs underlying phenotypic plasticity observed by immunofluorescence, of six patients sequenced, three with complete longitudinal coverage (ALK003, ALK011, ALK089; seven timepoints) are presented here. These patients were selected based on sample quality and the availability of matched clinical annotations across serial visits. Integrated UMAP analysis of all cells across seven longitudinal samples from three patients identified discrete CTC clusters clearly separated from immune cell populations, validated by copy number variation (CNV) scoring (Figure 3.4A). The ALK003-specific split UMAP further illustrated visit-specific shifts in CTC and immune cell cluster proportions across three serial timepoints (B01, B02, B03), with the CTC cluster demonstrating consistent spatial separation from immune populations at each visit (Figure 3.4B).

Cell type composition analysis across all seven serial scRNA-seq samples confirmed successful CTC cluster identification at every timepoint (Figure 3.4D), with CTC proportions varying across patients and visits, consistent with the longitudinal CTC count dynamics observed by immunofluorescence. T cells constituted the predominant PBMC population across samples, as expected for peripheral blood. Monocyte proportions showed inter-sample variability, with selective enrichment at specific timepoints, potentially reflecting dynamic inflammatory or

immunosuppressive states during TKI treatment. The identification of a platelet-enriched transcriptional signature in ALK089 CTCs was paralleled by elevated platelet representation in the corresponding cell type composition, suggesting a biologically meaningful CTC-platelet association in this patient.

Heatmap analysis of curated EMT, epithelial, mesenchymal, and platelet-associated marker genes across serial CTC samples (Figure 3.4C) revealed distinct, patient-specific transcriptional trajectories in all three patients. In ALK003, EMT-associated genes (ZEB1, ZEB2, SMAD2/3/4, SERPINE1) were prominent at B01, attenuated at B02 during clinical stability, and partially re-activated at B03 coinciding with disease re-progression, demonstrating that EMT transcriptional state may be dynamically modulated by therapeutic pressure. In ALK011, CTCs transitioned from a KRT8-high epithelial state at B02 to a VIM/SPARC/GATA2-high mesenchymal state at B04, corroborating the pseudotime trajectory analysis. ALK089 displayed a qualitatively distinct program: sustained co-expression of EMT genes and platelet-associated markers (PPBP, PF4, ITGA2B) that intensified from B04 to B05, potentially reflecting CTC-platelet interactions facilitating immune evasion and hematogenous dissemination.

To characterize cell-intrinsic transcriptomic signatures at single-cell resolution, differential gene expression analysis comparing CTCs to all other cell types was performed on ALK003 (Figure 3.5B). EMT-associated genes including SPARC were persistently upregulated in CTCs across all three visits, corroborating the immunofluorescence evidence of phenotypic plasticity. Concurrently, GATA2, a transcription factor implicated as a context-dependent tumor suppressor in lung cancer, was consistently and significantly downregulated across visits in ALK003. HMM-based CNV inference from scRNA-seq data identified a distinct CDK4 copy

number gain (state 4–5) emerging specifically at the progression-associated visit in both ALK003 and ALK011 (Figure 3.4E). In ALK003, this amplification was dynamically selected, suggesting that CDK4-amplified clones represent a directly actionable bypass resistance mechanism. In ALK089, HMM-based CNV profiling at B04 and B05 revealed a distinct pattern of chromosomal alterations; however, given the absence of intermediate timepoints in this patient, the temporal dynamics of clonal selection and their relationship to clinical progression warrant further investigation in the context of expanded longitudinal sampling.

Pseudotime trajectory analysis was performed in all three patients (ALK003, ALK011, and ALK089; Figure 3.4F). In ALK003, CTCs at the earliest timepoint (B01) exhibited a broad, unimodal pseudotime distribution, reflecting transcriptomic diversity within the baseline CTC population. At the progression-associated timepoint (B02), this distribution shifted to a bimodal pattern, suggesting the emergence of two transcriptomically distinct CTC subpopulations under TKI pressure, consistent with clonal divergence rather than selection of a single dominant state. In ALK011, CTCs at B02 and B04 occupied distinct pseudotime states, supporting a directional trajectory from an earlier to a later transcriptional program coinciding with disease progression. In ALK089, pseudotime distributions at B04 and B05 demonstrated progressive transcriptional divergence between the two timepoints, consistent with continued phenotypic evolution under sustained therapeutic pressure, though the limited number of timepoints available in this patient precludes definitive directional inference.

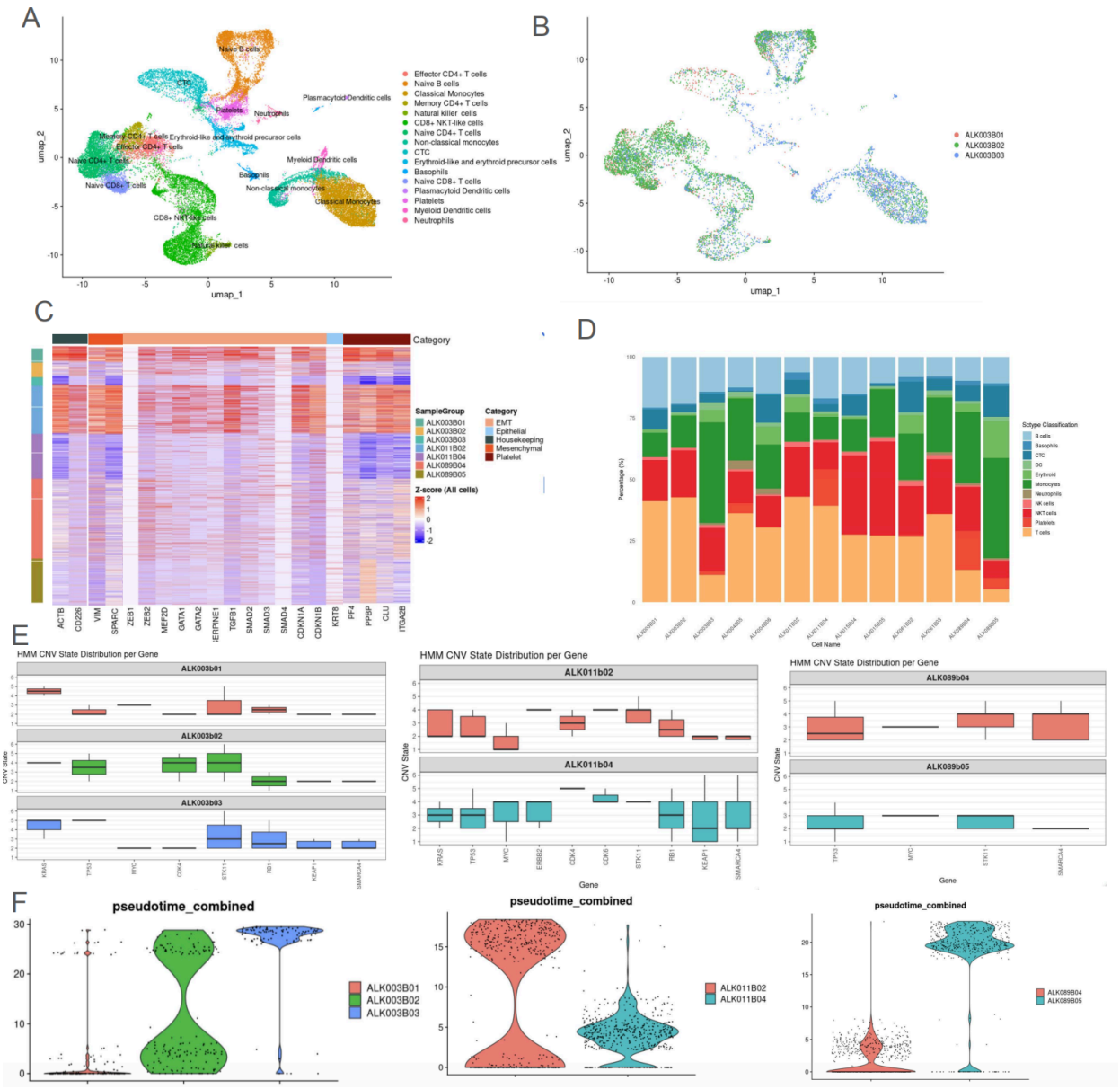


Figure 3.4. Integrated multi-patient scRNA-seq profiling: cell type composition, transcriptomic trajectories, copy number variation, and pseudotime analysis across longitudinal CTC samples.

(A) Integrated UMAP of all cells across all seven longitudinal scRNA-seq samples from three ALK NSCLC patients (ALK003, ALK011, ALK089), annotated by cell type. A discrete CTC cluster is clearly separated from immune cell populations and validated by CNV scoring. (B) Split UMAP displaying the distribution of cells from ALK003 across three longitudinal timepoints (B01, B02, B03), illustrating visit-specific shifts in CTC and immune cell cluster proportions. (C) Heatmap of curated EMT-associated (ACTB, CD226, VIM, ZEB1, MEF2D,

GATA2, TGFB1, SERPINE1, SMAD2, SMAD3, SMAD4), epithelial (KRT8), cell cycle (CDKN1A, CDKN1B), and platelet-associated (PF4, CLU, ITGA2B) marker gene expression across all seven longitudinal CTC samples from all three patients (ALK003 B01–B03, ALK011 B02–B04, ALK089 B04–B05), demonstrating patient-specific and timepoint-specific transcriptional trajectories. Color denotes Z-score normalized expression across all cells; sample group and gene category annotations are indicated. (D) Stacked bar charts showing cell type composition proportions across all seven longitudinal scRNA-seq samples, illustrating inter-patient and inter-timepoint variability in immune and CTC subpopulation abundance. (E) HMM-inferred CNV state distributions for cancer-relevant genes (KRAS, TP53, MYC, CDK4, STK11, RB1, ERBB2, KEAP1, SMARCA4) across serial timepoints in ALK003 (left; B01–B03), ALK011 (middle; B02–B04), and ALK089 (right; B04–B05), revealing dynamic clonal selection, including emergence of CDK4 copy number gain at progression-associated visits in ALK003 and ALK011, under sustained TKI pressure. (F) Pseudotime trajectory violin plots for ALK003 (left; B01–B03), ALK011 (middle; B02–B04), and ALK089 (right; B04–B05), illustrating directional transcriptomic divergence under TKI pressure. In ALK003, a shift from broad unimodal to bimodal pseudotime distribution at B02 suggests emergence of transcriptomically distinct CTC subpopulations; in ALK011, CTCs at B02 and B04 occupy distinct pseudotime states consistent with directional transcriptional progression.

3.4.5 External validation of CTC-derived transcriptomic signatures

To assess whether longitudinal CTC transcriptomic changes reflect conserved biological programs with broader clinical relevance, we derived gene signatures from the ALK003 serial scRNA-seq data and scored them in an independent TCGA-LUAD cohort (n=500 lung adenocarcinoma patients) using Gene Set Variation Analysis (GSVA). Genes consistently downregulated (n=45) or upregulated (n=225) in ALK003 CTCs across all three serial visits were used as discovery signatures (Figure 3.5C). It is noted that this signature was derived from a single patient's longitudinal data, and these findings should therefore be considered hypothesis-generating pending validation in larger dedicated CTC-sequencing cohorts.

Lower GSVA scores for the 45-gene downregulated signature were significantly associated with shorter OS in the TCGA-LUAD cohort (p=0.039; Figure 3.5D), whereas the

upregulated signature showed no significant stratification (p=0.29). When restricted to genes downregulated at each individual visit, the prognostic signal strengthened progressively: the B01-derived signature was non-significant (p=0.21), the B02 signature reached significance (p=0.019), and the B03 post-progression signature yielded the strongest separation (p=0.0067). This progressive strengthening suggests that transcriptomic programs accumulating under sustained TKI pressure increasingly capture biology relevant to disease aggressiveness. Together, these findings confirm that the CTC-derived downregulated signatures identified by longitudinal single-cell profiling represent conserved prognostic determinants in lung adenocarcinoma, and cross-platform validation in a large independent cohort helps mitigate the inherent limitation of the small single-patient scRNA-seq discovery dataset.

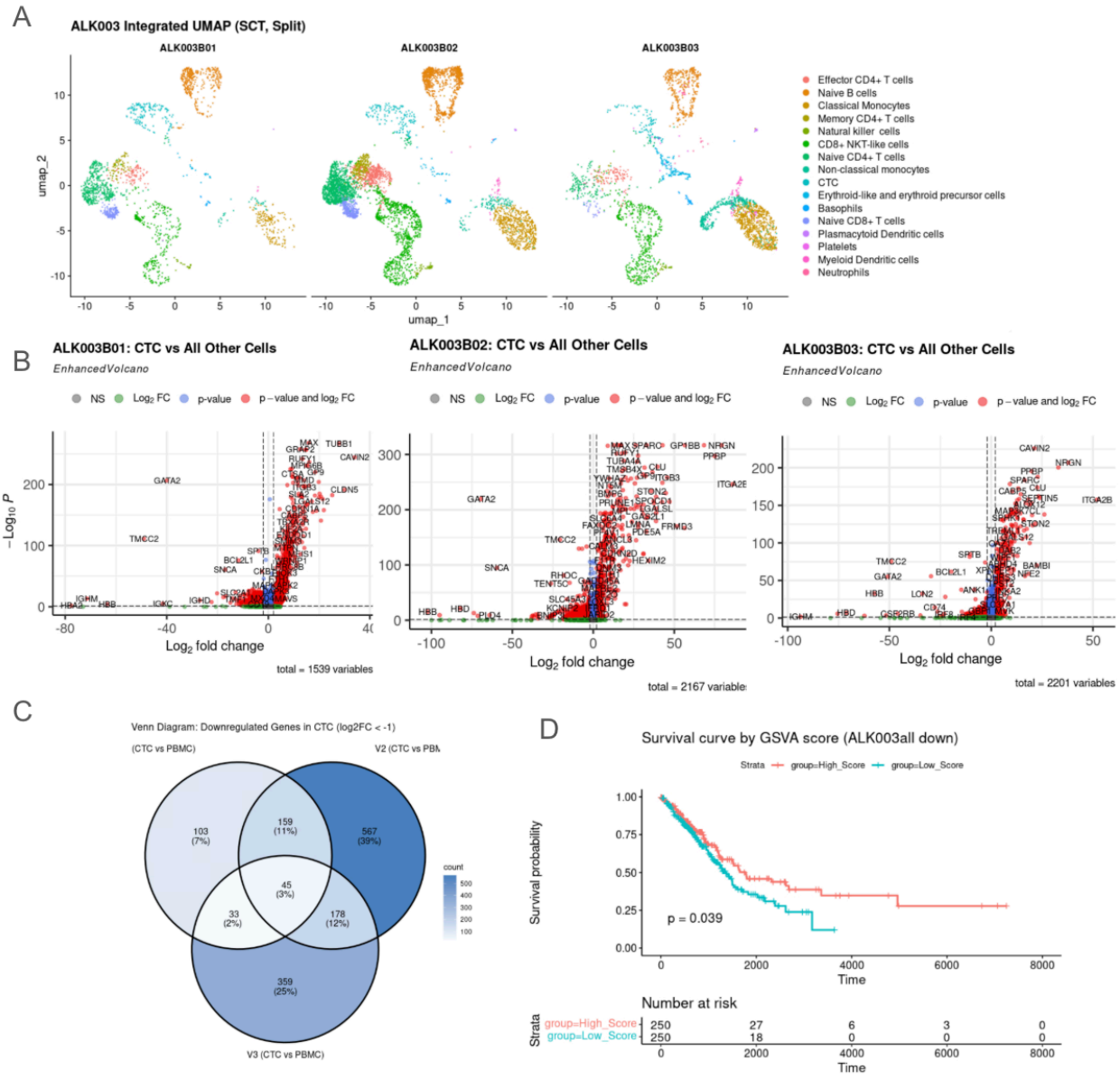


Figure 3.5. Single-cell transcriptomic characterization and external validation of longitudinal CTC-derived signatures.

(A) ALK003 integrated UMAP (SCT normalization), split by longitudinal timepoint (B01, B02, B03), displaying annotated cell type clusters including a discrete CTC population clearly separated from immune cell populations across all three serial visits. (B) Volcano plots of differential gene expression (DGE) in CTCs versus all other cell types at each of three serial ALK003 timepoints (B01, B02, B03). Key persistently upregulated genes across visits include CAVIN2, SPARC, NRGN, and ITGA2B; GATA2 is consistently and significantly

downregulated across all three timepoints. Point color denotes significance category: grey (NS), green (Log_2FC only), blue (p-value only), red (both p-value and Log_2FC). (C) Venn diagram depicting the overlap of genes downregulated in ALK003 CTCs relative to PBMCs ($log_2FC < -1$) across all three serial timepoints (V1: B01, V2: B02, V3: B03), identifying a consensus 45-gene downregulated signature shared across all visits. (D) Kaplan-Meier overall survival curve in the TCGA-LUAD cohort (n=500; High Score n=250, Low Score n=250) stratified by GSVA enrichment score derived from the 45-gene ALK003 consensus CTC-downregulated signature; patients with low GSVA scores demonstrated significantly worse survival probability.

3.5 Discussion

In this study, we provide a comprehensive, high-resolution portrait of longitudinal CTC dynamics in ALK NSCLC using the microfluidic Labyrinth enrichment platform. By integrating serial bulk CTC phenotyping, survival modeling, and scRNA-seq profiling across clinically defined disease milestones, our data support a coherent biological and clinical narrative: rather than enumeration averaged across visits or measured at any single timepoint, dynamic CTC phenotypic composition is associated with prognosis in ALK NSCLC. The underlying transcriptomic programs driving this phenotypic evolution can be identified at the single-cell level and externally validated in independent cohorts.

Total CTC counts at baseline did not differ significantly between deceased and surviving patients, and CTC counts at radiologically confirmed progression timepoints were indistinguishable from non-progression timepoints. These observations are consistent with a growing body of evidence demonstrating that bulk CTC burden, while prognostically relevant in certain solid tumor contexts, is confounded by anatomical, hemodynamic, and tumor-shedding factors that limit its discriminatory power in individual patients [86,207,301–303]. In contrast, the longitudinal trajectory of CTC burden demonstrated clear prognostic utility: reductions in total CTC count and absolute CK+ count between the first and last serial visits were each

significantly associated with improved overall survival (both nominal p = 0.044; not significant after Benjamini–Hochberg correction, q = 0.42). The mean absolute CK+ CTC count across all serial visits further achieved an AUROC of 0.971, establishing it as the strongest single discriminator of mortality in this cohort. These findings collectively indicate that the temporal dynamics of CTC phenotypic burden, rather than any single cross-sectional measurement, constitute the clinically relevant prognostic signal, an observation that has important implications for the design of liquid biopsy monitoring protocols in ALK NSCLC.

The prognostic primacy of the absolute CK+ CTC subpopulation is consistent with emerging evidence that epithelial CTC phenotype confers adverse prognosis in lung cancer [304,305]. Persistently elevated CK+ counts may reflect persistence of an epithelial-marker-positive CTC population with proliferative potential. Prior NSCLC CTC studies have shown that high CK-positive events are associated with poor prognosis and that a subset of single CTCs expresses the proliferation marker Ki67 [306]. In ALK NSCLC, adaptive survival under ALK-TKI pressure has also been linked to drug-tolerant states with reduced apoptotic response, supporting the possibility that persistent CK+ CTCs represent a treatment-tolerant circulating subpopulation. The observation that only CK+ and EpCAM+ and only CK+/EpCAM+/ALL+ subpopulations were also significantly elevated in deceased patients at baseline further suggests that epithelial surface marker co-expression defines a particularly aggressive circulating subclone.

In contrast, the hybrid All+ compartment showed two associations that, read in isolation, appear to point in opposite directions. The peak All+ proportion across serial visits was significantly higher in surviving than in deceased patients (51.8% vs. 31.5%; AUROC = 0.857; p = 0.048), yet a greater decrease in the normalized All+ fraction (Nor_All+) across the stable-to-

progression transition was associated with significantly longer progression-free survival ($p = 0.013$; Figure 3.3F). The first result alone would suggest that a high hybrid burden is favorable, the second that losing hybrid cells is favorable. These two metrics, however, capture different features of the same trajectory rather than a static hybrid level: the peak reflects the maximal hybrid representation reached at any visit, whereas Nor_All+ reflects the net change between defined clinical milestones. Read together, they are consistent with a favorable course marked by a transient rise in the hybrid compartment followed by its contraction, rather than by a persistently high or persistently low hybrid fraction. Because the All+ fraction is a compositional measure, its decline may additionally reflect clearance of the broader CTC population, including the epithelial CK+ clone associated with mortality in this cohort, rather than selective loss of hybrid cells alone. Consistent with a dynamic rather than instantaneous signal, cross-sectional All+ burden did not differ between progression and non-progression timepoints (Figure 3.3E). While hybrid EMT states are broadly associated with enhanced metastatic capacity and therapeutic resistance, recent evidence from NSCLC models suggests that hybrid EMT cells may be specifically susceptible to natural killer cell–mediated cytotoxicity, offering a candidate mechanism by which a transient hybrid population could be immunologically cleared in a treatment-responsive host. Collectively, these data establish that the temporal evolution of CTC phenotypic composition, assessed through serial sampling, is required to unmask prognostically relevant biological signals that are systematically obscured when analysis is restricted to any single timepoint measurement.

The longitudinal scRNA-seq data provide mechanistic context for these phenotypic observations. Differential gene expression analysis of CTCs across serial timepoints in ALK003 identified persistent upregulation of CAVIN2 and SPARC. SPARC has established roles in

Snail-associated invasion and extracellular matrix remodeling in NSCLC [307]. The consistent downregulation of GATA2 across serial CTC visits in ALK003 is particularly noteworthy. GATA2 is a lineage-associated transcription factor with context-dependent roles in lung cancer; it has been reported to be epigenetically repressed in lung tumors, although other studies indicate that GATA2 can support survival in RAS-driven NSCLC [308,309]. Its sustained repression in progressive CTCs is consistent with epigenetic silencing mechanisms reported in advanced, de-differentiated lung tumors, and likely facilitates the transcriptomic plasticity that enables TKI evasion. Additionally, HMM-based copy number variation inference from scRNA-seq data identified a distinct CDK4 copy number gain emerging specifically at progression-associated timepoints in both ALK003 and ALK011. The emergence of CDK4 copy-number gain is consistent with a cell-cycle bypass hypothesis and provides a rationale for evaluating ALK plus CDK4/6 inhibitor combinations, which have shown clinical activity in ALK-rearranged NSCLC [310]. The transient nature of this amplification in ALK003, with reduction observed at the subsequent visit following treatment modification, underscores the rapidity of clonal selection and deselection under therapeutic pressure. CDK4-amplified CTCs might be used as a potentially actionable biomarker for prospective evaluation of CDK4/6 inhibitor combination strategies in ALK NSCLC.

The external validation of the ALK003-derived 45-gene downregulated CTC signature in the TCGA-LUAD cohort (n=500; p=0.039) provides evidence that transcriptomic programs identified through single-cell CTC profiling reflect biologically conserved determinants of disease aggressiveness, rather than patient-specific artifacts. Restricting the validation cohort to lung adenocarcinoma is appropriate given that ALK NSCLC arises almost exclusively in this histological subtype and avoids confounding from the distinct molecular landscape of lung

squamous cell carcinoma. The progressive strengthening of this prognostic signal from the pre-progression (B01; p=0.21) to early progression (B02; p=0.019) to post-progression (B03; p=0.0067) timepoint-derived signatures suggests that transcriptomic programs accumulating under sustained TKI exposure increasingly capture conserved mechanisms of disease aggressiveness. The absence of prognostic stratification from the upregulated signature (p=0.29) may reflect greater inter-tumor heterogeneity in CTC-enriched transcriptional programs and warrants further investigation in dedicated longitudinal CTC-sequencing cohorts. The derivation of the discovery signature from a single patient's longitudinal data represents an inherent limitation; the concordance with an independent population-level dataset is therefore considered hypothesis-generating, and prospective validation using multi-patient CTC-derived signatures will be required to establish clinical utility.

Several limitations warrant consideration. The longitudinal cohort comprised 12 patients with 5 mortality events, which precluded robust Cox proportional hazards regression modeling; formal survival modeling was therefore limited to Kaplan-Meier analyses with log-rank testing. Because multiple CTC subpopulation metrics were evaluated against survival in a cohort of 12 patients with 5 events, the overall-survival trajectory associations did not retain significance after Benjamini–Hochberg correction (q = 0.42) and are reported as exploratory; their interpretive value rests on the consistency of effect direction across collinear burden metrics and on concordance with the single-cell and TCGA-LUAD analyses. Univariate and multivariate Cox proportional hazards regression models were attempted but were statistically underpowered due to the cohort size and event count. The scRNA-seq sub-cohort encompassed three patients, with detailed DGE and CNV analyses necessarily restricted to patients with sufficient CTC yield across multiple clinically distinct timepoints. Furthermore, the TCGA-LUAD validation cohort

comprises bulk tumor RNA-seq data, which integrates signal from stromal, immune, and malignant cell populations; the observed prognostic association of the CTC-derived signature in this context may therefore reflect broader tumor microenvironment biology rather than tumor-cell-intrinsic programs exclusively. Prospective validation in larger, multi-institutional ALK NSCLC cohorts with standardized CTC collection protocols will be required to confirm the generalizability of both the phenotypic trajectory findings and the scRNA-seq-derived signatures. Future studies should additionally explore the integration of matched ctDNA profiling to correlate CTC-derived CDK4 amplification events with circulating genomic evidence of clonal selection and should prospectively evaluate the clinical utility of longitudinal CTC phenotypic monitoring as a framework for treatment adaptation in ALK NSCLC.

In conclusion, this study establishes that the prognostic information content of CTC analysis in ALK NSCLC resides primarily in longitudinal phenotypic dynamics rather than static enumeration. The longitudinal CK+ CTC burden across serial visits and the trajectory of hybrid EMT fraction represent complementary prognostic axes that reflect distinct tumor adaptive strategies under TKI pressure. At the transcriptomic level, this adaptation is associated with conserved programs of epigenetic reprogramming and clonal selection of targetable resistance mechanism. The CTC-derived downregulated signature retained prognostic significance in an independent lung adenocarcinoma cohort and is therefore detectable through minimally invasive longitudinal profiling. Because that external validation used bulk tumor RNA-seq, which aggregates malignant, stromal, and immune compartments, CTC-resolved confirmation in dedicated single-cell cohorts will be required to attribute the signal specifically to circulating tumor cells. These findings support the integration of serial, multi-parametric CTC analysis into precision oncology frameworks for real-time monitoring of therapeutic resistance in ALK

NSCLC.

3.6 Methods

3.6.1 Study design and patient cohort

This study was approved by the Institutional Review Board (IRB) at the University of Michigan (IRB Protocol # HUM00227562). The primary analytical cohort for this longitudinal study consists of 12 ALK NSCLC patients who provided serial blood samples (2–10 draws per patient; total 56 samples). Clinical data, including treatment regimens (alectinib, lorlatinib, multi-agent therapies) and survival status, were recorded at the time of sample collection. The clinical response (stable disease vs. progressive disease) was determined by the treating physician at the time of blood draw based on radiological assessment. For each patient, one tube of blood was collected in EDTA tubes (BD Vacutainer, Cat# 367863) for processing within 12 hours, or in Cell Rescue tubes (Menarini Silicon Biosystems Inc, Cat#84000186) for longer time points and transported at room temperature.

3.6.2 Microfluidic CTC enrichment, immunofluorescence staining, and scRNA-seq library preparation

All CTC enrichment, immunofluorescence staining, and scRNA-seq library preparation procedures were performed as described in Chapter 2.

3.6.3 Pseudotime trajectory and copy number variation (CNV) inference

To model the transcriptomic evolution of CTCs under therapeutic pressure, pseudotime trajectory analysis was performed using the Monocle3 R package. Integrated Seurat objects were converted to CellDataSet objects, and trajectories were constructed to order single CTCs along a continuum of transcriptional state changes across serial clinical visits.

To infer large-scale chromosomal alterations and validate the malignant origin of the identified CTC clusters, copy number variation (CNV) analysis was performed using the inferCNV R package. A Hidden Markov Model (HMM) was applied to predict discrete CNV states (e.g., deletion, neutral, amplification) across the genome. Confirmed immune cells (e.g., T cells, B cells) from the same samples were used as a diploid reference baseline to accurately infer CNV profiles in the CTC populations.

3.6.4 External validation and gene set variation analysis (GSVA)

To validate the prognostic relevance of the CTC-derived transcriptomic signatures, a consistently downregulated 45-gene signature (log2FC < -1 across serial visits) was extracted. This signature was scored in an independent cohort of 500 Non-Small Cell Lung Cancer (NSCLC) patients from The Cancer Genome Atlas (TCGA-LUAD) using Gene Set Variation Analysis (GSVA). Patients in the TCGA cohort were stratified into high and low GSVA score groups based on the median score, and overall survival was compared.

3.6.5 Statistical analysis

Statistical analyses were performed using R (v4.3.3) and Python (v3.11.0), incorporating packages for survival analysis (survival, survminer), data manipulation (dplyr, pandas, numpy), and statistical testing (scipy, statsmodels). All statistical tests were two-sided with a significance threshold of $\alpha = 0.05$.

Continuous variables representing CTC counts and phenotypic proportions were compared between outcome groups using the Wilcoxon rank-sum test (Mann–Whitney U test) for pairwise comparisons (e.g., deceased vs. alive; progression vs. non-progression visits). To account for repeated measures inherent to the longitudinal study design, a linear mixed-effects model was fitted to log-transformed total CTC counts, with time (weeks from enrollment) and

mortality status as fixed effects and patient identifier as a random intercept, capturing within-patient correlation across serial visits.

Patient-level longitudinal features were derived by aggregating sample-level measurements across all visits, including baseline (first measurement), last, peak, and mean values for each CTC biomarker. Inter-visit dynamics were summarized using two complementary metrics, each matched to a distinct survival endpoint. For overall-survival analyses, inter-visit change was quantified as the absolute difference between the last and first serial blood draws, $X_{last} - X_{first}$, where X_{first} and X_{last} denote the value of the metric (total CTC count, absolute CK+ count, or the combined Vim+/CK+/All+ burden index) at the first and final blood collection. This absolute-difference formulation does not involve division by a baseline value and is therefore not subject to inflation when baseline counts approach zero. For the progression-free-survival analysis of the hybrid All+ subpopulation, inter-visit change was expressed as a baseline-normalized rate of change, $Nor_All+ = (X_{last} - X_{first}) / X_{first}$, where X_{first} denotes the All+ value at the first radiologically stable timepoint and X_{last} the All+ value at the first progressive timepoint following stability, capturing the relative shift in hybrid CTC representation across the stable-to-progression transition. Patients with a first-timepoint All+ count of zero were excluded from this analysis to avoid undefined or inflated ratios.

For survival analysis, Kaplan–Meier curves were constructed for both progression-free survival (PFS) and overall survival (OS), with patients stratified at the cohort median of the relevant longitudinal CTC trajectory metric into greater-decrease versus lesser-decrease groups. Median-based dichotomization was used rather than the continuous change value because it depends only on each patient's rank relative to the cohort median; consequently, an extreme

value arising from a small baseline count affects only the group assignment of that individual patient and does not alter the median threshold or the classification of the remaining patients, limiting the influence of outlying ratios on the stratification. Survival differences between strata were assessed using the log-rank test, and 95% confidence intervals were calculated using Greenwood's formula.

Receiver operating characteristic (ROC) analysis was performed to evaluate the discriminatory capacity of cross-sectional CTC biomarkers for mortality, with area under the curve (AUROC) reported as a performance metric; given the small cohort size (n = 12, 5 mortality events), these estimates should be interpreted with caution. Univariate Cox proportional-hazards regression models were fitted for individual CTC biomarkers; multivariate modeling was not pursued due to statistical underpowering at the available cohort size and event count.

Given the exploratory nature of multi-marker biomarker screening across 11 primary markers and 5 composite markers, multiple-hypothesis-testing correction was applied using the Benjamini–Hochberg false-discovery-rate (FDR) procedure. A nominal p-value threshold of $p < 0.05$ was used to flag associations for follow-up, with the exploratory, FDR-adjusted framing applied throughout.

Chapter 4 Automated Computational Pipelines for High-Throughput CTC Identification and Phenotypic Scoring

4.1 Abstract

While Chapters 2 and 3 established the molecular and clinical relevance of CTC phenotypes in ALK NSCLC, scalable and reproducible interpretation of high-dimensional CTC imaging data remained a bottleneck. Here, a computational framework was developed to translate CTC immunofluorescence imaging into automated identification and functional phenotyping. A semi-automated pipeline was constructed for high-throughput multiplex IF processing, and three classical machine learning classifiers were benchmarked across 35 morphological and intensity features; logistic regression achieved the best CTC classification (F1 = 83.7%). Two complementary deep learning approaches extended this framework: a transfer-learned VGG-16 with Grad-CAM interpretability, and a deep cell-laser classifier (DCLC) leveraging biolaser optical signatures for antigen-independent CTC detection. Building on these identification frameworks, a Linear Discriminant Analysis–derived morphological score (EMScore) was developed to quantify single-cell EMT state. EMScore correlated with NK cell–mediated cytotoxicity across lung cancer cell lines and patient-derived CTC models ($R^2 = 0.684$), linking image-derived morphology to functional immune susceptibility. This work positions image-based computational phenotyping as a foundation for scalable, functionally interpretable CTC biomarkers in NSCLC.

4.2 Resulting Publication Information

Weishu Wu, Yu Zhang, Xiaotian Tan, **Yuru Chen**, Yuhang Cao, Vaibhav Sahai, Nicole Peterson, Laura Goo, Stacy Fry, Varun Kathawate, Nathan Merrill, Angel Qin, Sofia D. Merajver, Sunitha Nagrath, and Xudong Fan. "Antigen-independent single-cell circulating tumor cell detection using deep-learning-assisted biolasers." *Biosensors and Bioelectronics* 271 (2025): 116984.

Harrison Ball, Yin Zou, Cade Harris, **Yuru Chen**, Venkat Keshamouni, Nithya Ramnath, and Sunitha Nagrath. Tumor cell morphology based EMScore system links epithelial-mesenchymal transition to natural killer cell sensitivity in cancer cells. (Under review)

Zeqi Niu*, Sarah Owen*, **Yuru Chen**, Nina Perry, Alina Yan, Zach Gdowski, Mina Zeinali, Lana Garmire, Venkat Keshamouni, Nithya Ramnath, and Sunitha Nagrath. "Single-cell characterization of natural killer cell sensitivity of circulating tumor cells in non-small cell lung cancer" (In submission).

4.3 Introduction

CTCs are malignant cells shed from solid tumors into the peripheral circulation. They can be sampled repeatedly, providing temporal resolution that tissue-based assays cannot easily match [203]. Previous study showed CTCs clinical utility includes association with prognosis, treatment response, and metastatic progression across multiple tumor types [311,312]. Realizing this clinical potential, however, depends on workflows that can identify CTCs reliably from a vastly larger background of leukocytes. Typically, there are 1–1,000 CTCs per mL of blood, against a background of 10^6–10^7 nucleated cells[249].

Microfluidic enrichment platforms such as the Labyrinth device exploit inertial focusing to separate CTCs from leukocytes in a label-free manner, generating CTC-enriched fractions suitable for downstream multiplex immunofluorescence (IF) imaging [131,179]. Conventional

CTC identification then relies on a marker-defined phenotype, typically DAPI+ / pan-cytokeratin (PanCK)+ (for tumor marker) / CD45 negative (for leukocyte) assessed by manual review of fluorescence microscopy images [179]. Two practical bottlenecks limit this approach. First, manual enumeration scales poorly with per-patient imaging volume and is subject to inter-reviewer variability [313]. Second, cell-by-cell manual review is slow. Beyond these throughput limits, adverse outcomes have been linked to CTC burden dynamics, CTC clusters, and mesenchymal/stemness phenotypes [44,314], indicating that the prognostic content of CTCs lies not in enumeration alone but in single-cell phenotypic state.

Computational image analysis offers a route to address these limitations. Machine learning (ML) refers broadly to algorithms that learn statistical mappings from input data to output labels, and supervised classification in particular fits such mappings from labeled training examples for application to unlabeled cases [315]. To estimate how well a model generalizes to unseen data, labeled samples should be separated before model development. The training set is used to fit model parameters. The validation set is used for hyperparameter tuning and model selection. A held-out test set should be used only for final evaluation[316,317] . When sample size is limited, k-fold cross-validation is often used. The data are divided into k folds. Each fold is used once as validation, while the remaining folds are used for training. The final performance is then averaged across folds [316,318]. For imbalanced tasks, such as CTC-versus-leukocyte classification, stratified splitting is preferred because it preserves class proportions across folds [319]. Test data should not influence training, preprocessing, feature selection, or model selection. Otherwise, data leakage can inflate performance estimates and weaken clinical interpretation [317].

Within this supervised framework, two image-analysis strategies are commonly used. Feature-engineered machine learning uses predefined measurements, such as fluorescence intensity, texture, size, and shape. Models such as logistic regression, support vector machines, random forests, and Bayesian classifiers have been applied to automated CTC detection and classification [320–322]. These approaches are useful when interpretability and small-dataset performance are important. Deep learning takes a different approach. CNNs learn image features directly from pixel data and have been widely used in microscopy and medical image classification [323–327]. CNNs can improve image-based classification, but their predictions are often harder to interpret. Grad-CAM helps address this issue by highlighting image regions that contribute most to a model prediction [328,329]. Both feature-engineered classifiers and fluorescence-based CNNs can still depend on antigen labeling. This is a limitation when epithelial markers such as EpCAM or cytokeratins are reduced during EMT [330–332]. Antigen-independent optical strategies may reduce this dependency. In single-cell or tissue biolaser systems, cells or tissues are placed in Fabry–Pérot optical cavities. Their lasing output reflects optical and structural features, including nuclear organization and refractive-index heterogeneity, rather than surface-marker expression alone [333–335]. Therefore, biolaser-based classification may help identify CTCs with altered epithelial marker expression [336].

After CTCs are identified, image features can also be used to ask whether morphology carries functional information. High-content image-based profiling shows that cell morphology can capture cell state and predict health-related phenotypes [175,337,338]. In lung cancer, EMT has been linked to differential NK-cell susceptibility. EMT-associated changes in E-cadherin and CADM1 can modulate NK-mediated cytotoxicity [339]. NK-cell activity is regulated by a balance of activating and inhibitory receptor–ligand signals, rather than by antigen presentation

alone [340,341]. CTCs also interact with immune cells in circulation, including NK cells, and CTC burden has been associated with immune-cell phenotypes in cancer patients [69,342]. Together, these findings support a cautious hypothesis: mesenchymal or hybrid CTC morphology may reflect ligand programs related to NK sensitivity, and some of this information may be inferred from routine immunofluorescence images.

Building on the clinical utility of CTCs as dynamic biomarkers established earlier in this thesis, this chapter develops and evaluates a series of computational image-analysis strategies that span CTC identification and morphology-based functional phenotyping. The first aim, main question, is to build and strengthen a reproducible CTC identification pipeline; the second aim, exploratory question, is to test whether the same imaging data can be repurposed for morphology-based functional phenotyping. These two aims define a three-stage analytical strategy. First, a semi-automated identification pipeline was established as the methodological foundation, and three traditional supervised classifiers were benchmarked within an engineered feature space to define an interpretable baseline (Sections 4.4.1–4.4.2). Second, the identification framework was extended along two deep learning axes that the linear baseline cannot reach: a transfer-learned convolutional neural network coupled with Grad-CAM, designed to test whether end-to-end deep classification can be made biologically interpretable rather than treated as a black box (Section 4.4.3). Additionally, an antigen-independent classifier built on biolaser optical signatures, designed to recover CTCs that have downregulated canonical surface markers during EMT (Section 4.4.4). Third, with reproducible identification in place, the same imaging data were repurposed to derive a Linear Discriminant Analysis–based morphological score (EMScore), which was then evaluated for its capacity to predict NK-mediated cytotoxicity across lung cancer cell lines and patient-derived CTC models (Section 4.4.5). The first two stages

address the main question by progressively strengthening the identification pipeline; the third stage takes up the exploratory question, asking whether the same imaging data can be repurposed to predict an immune functional state.

4.4 Results

4.4.1 A semi-automated image processing pipeline enables high-throughput CTC identification

A semi-automated image analysis pipeline was developed to process multiplex immunofluorescence images acquired downstream of Labyrinth microfluidic CTC enrichment. The pipeline was implemented as a custom graphical user interface (GUI) built on MATLAB integrating MATLAB-based image processing. Each input image contains three channels from the IF staining: DAPI for nuclei (blue), pan-cytokeratin (PanCK) as a positive epithelial tumor-cell marker (red), and CD45 as a negative leukocyte marker (green) (Figure 4.1A).

Nuclear segmentation was initiated by Otsu's adaptive thresholding on the DAPI channel, and adjacent or overlapping nuclei were separated into discrete masks by a marker-controlled watershed and filled hole algorithm. Nuclear masks were overlaid onto the PanCK channel to compute median epithelial-marker intensity per region. A cell-filter mask for DAPI was used for cellular regions of interest. A combined cell-filter mask selecting PanCK-positive and white blood cell (WBC)-negative regions then defined candidate CTC regions of interest (ROIs). The user can manually adjust the fluorescence cutoff threshold. The GUI displayed estimated cell counts in real time and allowed each candidate ROI to be inspected individually, supporting manual exclusion of debris, imaging artifacts, and poorly segmented objects before downstream feature extraction (Figure 4.1A).

Slides containing the lung cancer cell lines H1975 and A549, and an H1975/WBC spike-

in control, were processed using the semi-automated cell-analysis program. The result shows that the viewed count was within 2% error of the manual count. Eight Labyrinth-enriched NSCLC patient slides were counted using the automated cell-analysis program.

4.4.2 Systematic benchmarking identifies logistic regression as the optimal classifier within the engineered feature space

For each image's retained ROI, 35 quantitative features were extracted using ImageJ, comprising fluorescence-intensity descriptors (mean, standard deviation, mode, minimum, and maximum intensity) and morphology descriptors (area, minimum Feret diameter, height, minor axis length, and additional shape parameters). All features were standardized to z-scores prior to downstream analysis to mitigate variation in microscope settings, illumination inhomogeneity, and staining intensity. Univariate comparison between annotated CTC and white blood cell (WBC) populations identified MinFeret and Height as the two most discriminatory morphology features, both showing significant differences between groups ($p < 0.001$, Wilcoxon rank-sum; Figure 4.1B). Each ROI was then represented as a 35-dimensional feature vector for downstream machine learning classification (Figure 4.1C).

In total, 2,035 annotated single-cell ROIs were assembled for classifier development, comprising 1,265 CTC and 770 WBC events (CTC:WBC ≈ 1.6:1). To benchmark classical machine learning approaches for automated CTC identification, three classifiers spanning increasing model complexity—logistic regression, decision tree, and random forest—were trained and evaluated on the engineered 35-feature matrix. To prevent data leakage and assess generalizability, the dataset was partitioned by stratified sampling into an 80% training set (1,628 ROIs: 1,012 CTC, 616 WBC) and an independent 20% validation set (407 ROIs: 253 CTC, 154 WBC), preserving class proportions across partitions. Feature standardization parameters were fit

on the training partition only, and the three classifiers were trained with 3-fold cross-validation on the training partition before being compared on the held-out validation set (Figure 4.1C).

Principal component analysis (PCA) of the standardized 35-feature matrix on the validation set showed clear separation between annotated CTCs and WBCs in the low-dimensional projection, with partial overlap at the class boundary (Figure 4.1E). The pattern of separation indicated that a substantial fraction of the class-discriminative variance was concentrated along linear axes; the residual overlap implied that no single feature or simple threshold was sufficient for classification, motivating multivariable modeling. Because the two classes were modestly imbalanced, the F1 score rather than raw accuracy was used as the primary metric for classifier comparison. Among the three classifiers, logistic regression achieved the highest validation performance, with 88.0% accuracy, 84.3% recall, 83.2% precision, and an F1 score of 83.7% (Figure 4.1D). Random forest and decision tree achieved F1 scores of 82.0% and 75.4%, respectively. To illustrate classifier behavior at the single-sample level, the confusion matrix for a representative patient sample is shown (Fig. 4.1F): 230 CTCs and 128 WBCs were correctly classified, with 23 CTCs misclassified as WBCs (false negatives) and 26 WBCs misclassified as CTCs (false positives), indicating a balanced false-positive/false-negative distribution. Sample-level accuracy was above approximately 90% in five of six samples (Figure 4.1G). Together, these results identified logistic regression as the best-performing classifier among the three classical models tested within this engineered feature space.

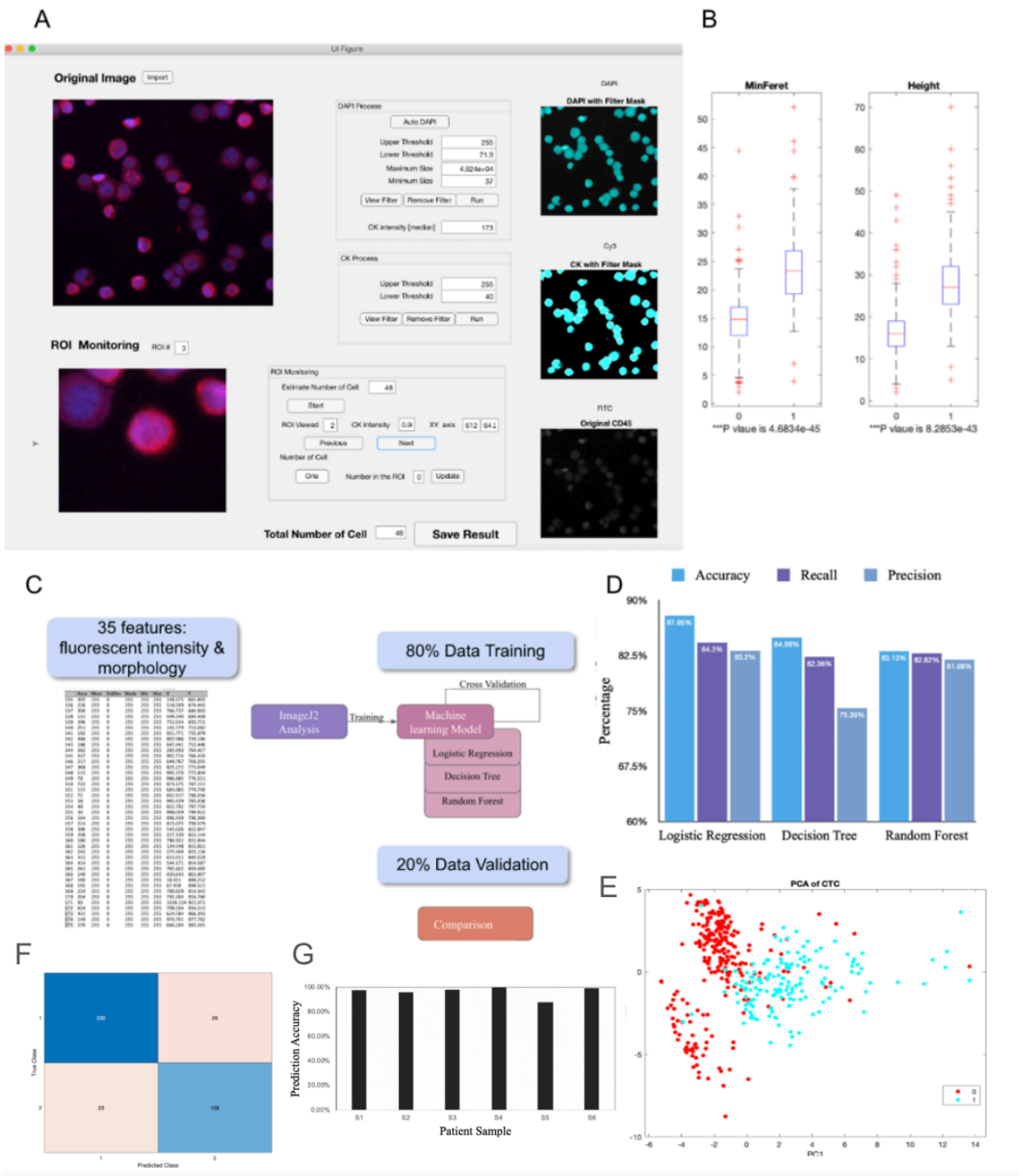


Figure 4.1. Semi-automated image analysis pipeline and classical machine learning classifier benchmarking.

(A) Graphical user interface for the semi-automated CTC identification pipeline, implemented in MATLAB and ImageJ2. The interface supports real-time multi-channel image import (original DAPI/CD45/PanCK

channels), automated Otsu thresholding and watershed nuclear segmentation (DAPI channel with filter mask), PanCK-based epithelial overlay (CK channel with filter mask), ROI-level inspection and artifact exclusion, and automated total-cell enumeration. (B) Top two morphology features separating CTC and WBC populations (MinFeret, Height; $p < 0.001$, Wilcoxon rank-sum). Class 0 = WBC; Class 1 = CTC. (C) Workflow schematic: each ROI is represented as a 35-dimensional feature vector; the feature matrix is partitioned into 80% training and 20% independent validation sets; three classifiers (logistic regression, decision tree, random forest) are trained with 3-fold cross-validation on the training partition and compared on the held-out validation set. (D) Classification performance (accuracy, recall, precision) on the held-out validation set. Logistic regression achieved the highest performance (accuracy 88.0%, recall 84.3%, precision 83.2%; F1 = 83.7%), followed by random forest (F1 = 82.0%) and decision tree (F1 = 75.4%). (E) PCA of the standardized 35-feature matrix on the validation set; red dots represent CTCs and cyan dots represent WBCs. (F) Confusion matrix for the logistic regression classifier on the held-out validation set: 230 correctly classified CTCs (Class 1) and 128 correctly classified WBCs (Class 0), with 23 CTCs misclassified as WBCs (false negatives) and 26 WBCs misclassified as CTCs (false positives). (G) 6 Patient sample prediction accuracy of the logistic regression classifier across six patient samples (S1–S6). The classifier maintained ≥ 90% accuracy in five of six samples, with reduced performance on S5.

4.4.3 A convolutional neural network with Grad-CAM interpretability demonstrates biologically grounded attention

To extend beyond the feature-engineered classifiers and evaluate whether networks that learn directly from raw image data could augment the engineered-feature classifiers, a VGG-16 CNN was implemented using transfer learning from ImageNet-pretrained weights. The original fully connected classifier layers were removed and replaced with newly initialized layers, while the convolutional feature-extraction blocks (Conv1–Conv5) retained their pretrained weights (Figure 4.2A). The model was trained on a publicly available reference dataset of blood-cell images spanning five leukocyte classes (neutrophils, eosinophils, basophils, lymphocytes, and monocytes). Because this dataset contains leukocyte subtypes rather than CTCs, the model serves here as an interpretability and feasibility testbed. The training pool of 2,133 images was

split 90:10 into 1,920 training and 213 validation images, with a separate independent test set of 1,740 images reserved for final evaluation. Only the newly added fully connected layers were retrained during transfer learning. Training and validation curves over 25 epochs converged for both accuracy and loss, with no evidence of substantial overfitting (Figure 4.2D). On the independent test set, the model achieved a macro precision of 0.969, macro recall of 0.968, and macro F1 score of 0.968, indicating that the transfer-learned VGG-16 distinguished the five blood-cell classes with high accuracy.

Strong performance on a balanced reference dataset demonstrates technical feasibility, but it does not establish clinical suitability for patient-derived CTC detection. To examine whether the network's predictions were grounded in biologically meaningful image regions rather than background or imaging artifacts, Grad-CAM was applied across convolutional blocks. Attention maps generated from earlier blocks (Conv1–Conv2) showed broad, diffuse activation distributed across the image; intermediate blocks (Conv3) showed progressively localized activation over cell-associated regions; and the final block (Conv5), which drives the classification decision, concentrated attention over the nucleus and nuclear periphery (Figure 4.2B). Final-layer Grad-CAM overlays generated across representative cells consistently localized over cell-containing regions rather than empty background, providing evidence that the model attended to biologically relevant cellular structures (Figure 4.2C)

4.4.4 Antigen-independent CTC identification via a deep learning–assisted biolaser platform

An antigen-independent CTC identification strategy was evaluated using a single-cell biolaser platform coupled with deep learning classification, developed in collaboration with the Fan laboratory. Target cells and WBCs were stained with the nucleic acid dye SYTO 9 and counterstained with DAPI and loaded into a micro Fabry–Pérot optical cavity under UV LED

excitation (Figure 4.2E). The biolaser readout exploits chromatin-dependent lasing-mode patterns generated within the cavity; because these patterns are governed by nuclear architecture rather than surface marker expression, the resulting signal is biophysically orthogonal to immunolabeling. For each cell, lasing-mode images were acquired at four pumping powers (25, 45, 90, and 150 μJ/mm²) (Figure 4.2F), and the four-condition stack was treated as a high-dimensional multimodal representation, termed the Hyper Lasing Signature (HLS), of the single cell.

The Deep Cell-Laser Classifier (DCLC) comprised a parallel single-cell laser encoder built from four independently trained ResNet-101 blocks, each projecting one lasing-power image into a 1,024-dimensional feature vector, followed by a fully connected classifier layer optimized for high-specificity rare-cell detection. The classifier was trained on a pancreatic cancer patient–derived CTC line (Pancreatic 368) and WBCs from six healthy donors and was evaluated zero-shot on two unseen CTC lines (Pancreatic 322 and Pancreatic 331). t-Distributed stochastic neighbor embedding (t-SNE) visualization of the encoded feature vectors showed that the Pancreatic 368 CTC line formed a cluster distinct from WBC populations from six healthy donors, indicating that the encoder learned discriminative lasing-pattern representations separating CTC from WBC (Figure 4.2G). Following lasing-rate correction, the DCLC achieved a sensitivity of 94.3% and a specificity of 99.9% on independent samples. CTC enumeration on clinical samples from pancreatic and lung cancer patients yielded counts that were broadly consistent with parallel conventional IF-based enumeration, although per-sample differences between the two methods were observed (Figure 4.2H).

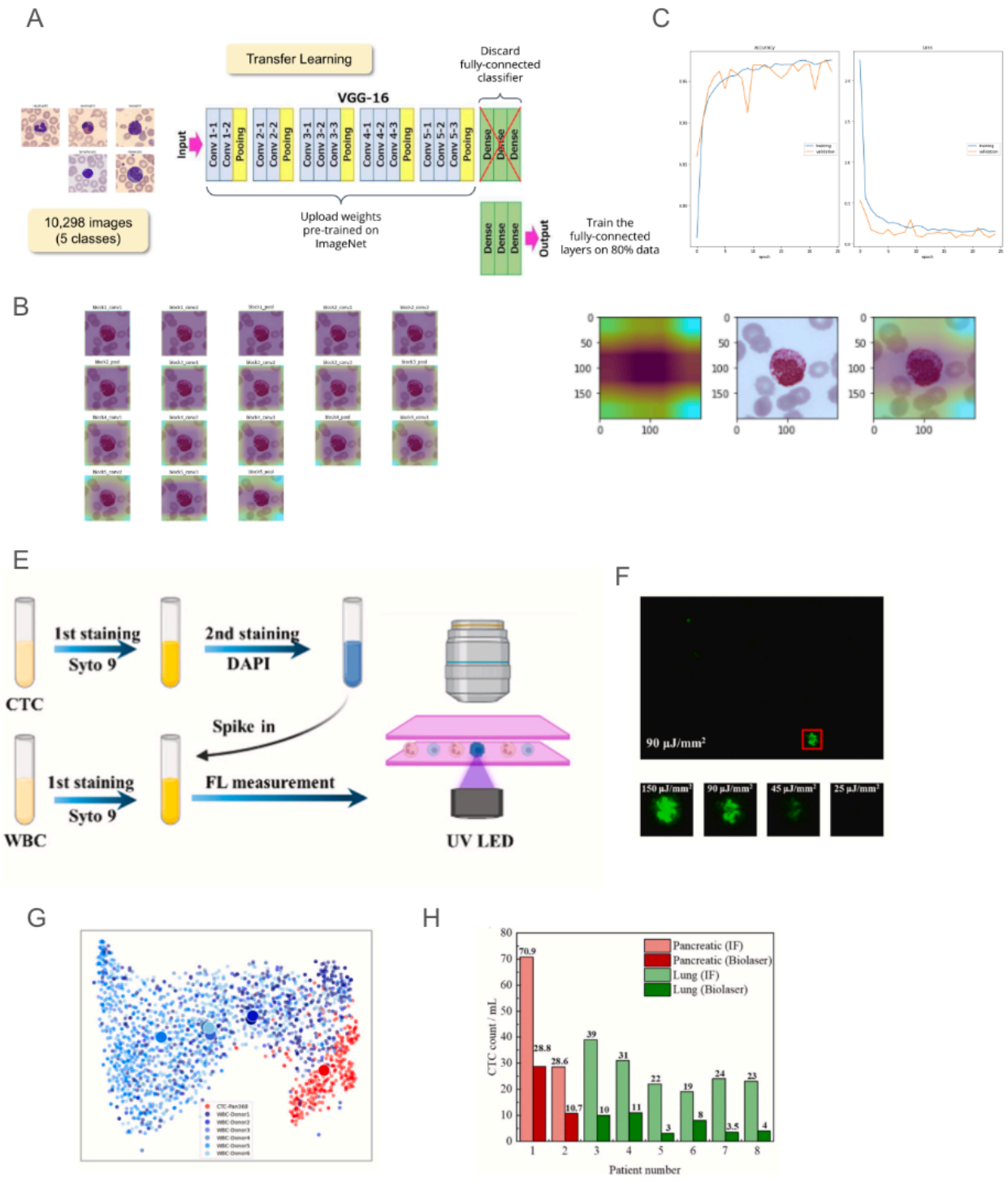


Figure 4.2. Transfer-learned VGG-16 with Grad-CAM interpretability and antigen-independent CTC identification via a deep learning–assisted biolaser platform.

(A) Transfer-learning architecture. A VGG-16 network pretrained on ImageNet was adopted, with the original fully connected classifier layers discarded and replaced. The convolutional feature extractor (five blocks,

Conv 1 through Conv 5) retained its pretrained weights, and only the new fully connected layers were retrained on a reference dataset of blood-cell images spanning five leukocyte classes (neutrophil, eosinophil, basophil, lymphocyte, monocyte) (B) Progressive refinement of Grad-CAM attention maps across convolutional blocks for a representative monocyte from the held-out validation set. Early blocks (Conv 1–2) generate diffuse activation distributed across the image background, intermediate blocks (Conv 3) progressively localize activation to broader cellular regions, and deeper blocks (Conv 5) concentrate attention on the nucleus and nuclear periphery. The final block, used for the network's classification decision, focuses attention specifically on the nucleus (red-bordered panel). (C) Final-layer Grad-CAM overlays on representative cells from blood-cell classes, showing consistent localization over cell-containing regions rather than background. (D) Training and validation accuracy and loss curves across 25 epochs indicate convergence with minimal overfitting (macro precision 0.969, macro recall 0.968, macro F1 0.968 on the independent test set). (E) Workflow schematic. Target cells (CTCs or spiked-in model lines) are stained with the nucleic acid dye SYTO 9 and counterstained with DAPI, mixed with SYTO-9-stained white blood cells from healthy donors, and loaded into a micro Fabry-Pérot cavity under UV LED excitation. Laser-emission fluorescence images are acquired and processed by the DCLC model for classification, outputting per-cell CTC/WBC labels with bounding-box overlays. (F) Representative single-cell laser-emission images of a CTC at four UV pumping powers (150, 90, 45, and 25 μJ/mm²); top panel shows the full field of view at 90 μJ/mm² with the CTC boxed in red. (G) t-SNE visualization of the feature vectors of 500 randomly selected cells from the Pancreatic 368 CTC line (red) and 500 WBCs from each of Donors 1–6 (blue shades) across the pumping powers at 25 μJ/mm²). (H) CTC enumeration on clinical samples from pancreatic (n = 2) and lung (n = 6) cancer patients by DCLC biolaser (dark bars) versus conventional IF-based enumeration (light bars).

4.4.5 An LDA-based morphological score predicts NK sensitivity in patient-derived CTC models

To extend computational image analysis from identification to functional phenotyping, an LDA-derived epithelial–mesenchymal score (EMScore) was developed in collaboration to quantify EMT-associated morphology at the single-cell level. Cells were stained for nuclear (DAPI), nucleic acid, and F-actin/cell membrane compartments, and morphology features were

extracted per cell (Figure 4.3A). An automated quality-control step using a random forest classifier trained on expert-annotated intact cells versus artifacts removed debris, imaging noise, and mis-segmented clusters prior to scoring.

A linear discriminant analysis model was then trained on curated reference populations of epithelial and mesenchymal cells to identify the linear combination of features that maximized between-class separation relative to within-class variance. Feature-importance analysis showed that Cells_Extent, Cells_Zernike_0_0, Nuclei_Zernike_2_0, Cells_FormFactor, Cells_MedianRadius, and Cells_Solidity carried positive weights associated with the epithelial reference population, whereas Cells_MaxFeretDiameter and Cells_Eccentricity carried negative weights associated with the mesenchymal reference population (Figure 4.3B). Raw discriminant scores were rescaled to the interval [0, 1] to yield the final EMScore, and density plots showed clear separation between epithelial and mesenchymal reference populations after rescaling (Figure 4.3C). Representative cells across the EMScore range exhibited corresponding morphology: cells with EMScore < 0.2 displayed compact, rounded morphology; cells near EMScore ≈ 0.5 showed intermediate morphology; and cells with EMScore > 0.8 displayed elongated, spindle-like morphology with extended protrusions (Figure 4.3D). By producing a continuous value rather than a binary epithelial/mesenchymal label, the EMScore captures the graded nature of EMT and is suitable for quantitative correlation with functional readouts.

NK cell–mediated cytotoxicity was quantified using an image-based viability assay in which live cells were identified as CellTracker^{+} / DAPI^{+} / propidium iodide (PI)$^{-}$ and dead cells were identified by PI uptake (Figure 4.3E). Cytotoxicity was measured across four established lung cancer cell lines (A549, H1650, H2228, H1975) and two patient-derived CTC lines (CTC-Lu1 and CTC-Lu2) at multiple effector-to-target (E:T) ratios using NK-92®MI as effectors over

a 4 h co-culture. Cytotoxicity varied across the panel: A549 and H1650 showed lower cytotoxicity, H2228 and H1975 showed intermediate levels, and CTC-Lu1 and CTC-Lu2 showed higher cytotoxicity, with greater than 90% killing observed for the patient-derived CTC lines at the highest E:T ratio (Figure 4.3F).

Mean EMScore was then correlated with NK cytotoxicity across the six cell lines. A positive correlation was observed ($R^2 = 0.684$; Figure 4.3G): cell lines with higher mean EMScore showed higher NK-mediated cytotoxicity, whereas cell lines with lower mean EMScore showed lower cytotoxicity. The correlation indicates that morphology-derived features capture cellular state information beyond morphology per se and supports the use of an image-based EMScore as a non-invasive proxy for immune sensitivity.

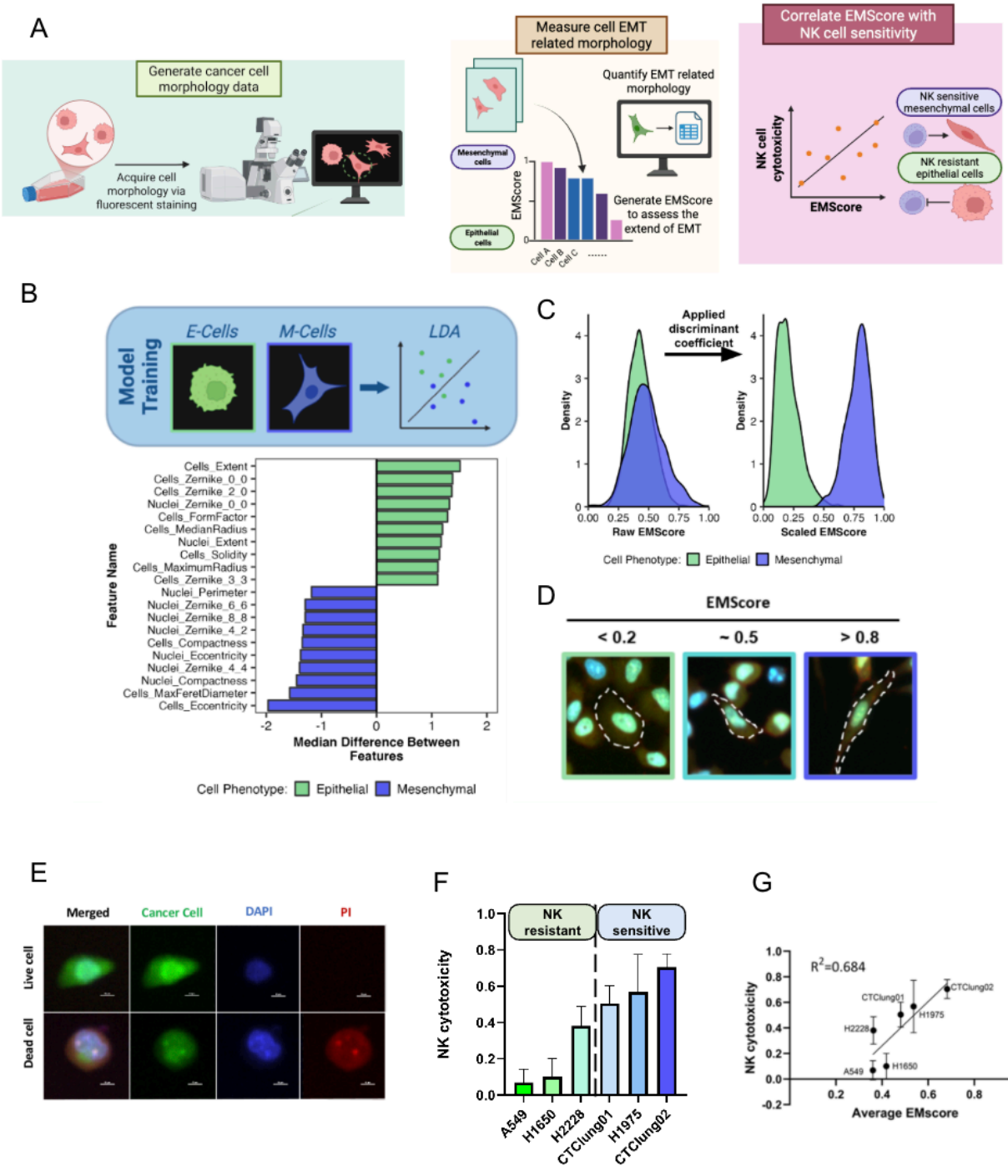


Figure 4.3. LDA-derived EMScore captures the epithelial–mesenchymal morphological continuum at the single-cell level.

(A) EMScore concept. Cancer cell morphology data are acquired by fluorescence staining of nuclear

(DAPI), nucleic acid, and F-actin/cell membrane compartments; morphological features are extracted per cell; and an EMScore is computed that reflects the extent of EMT-associated morphological remodeling, enabling quantitative comparison between epithelial and mesenchymal cells. A Linear Discriminant Analysis (LDA) model is trained on curated epithelial (E-cell) and mesenchymal (M-cell) reference populations to identify the linear combination of morphological features that maximizes between-class separation relative to within-class variance.(B) Feature importance, expressed as the median difference between mesenchymal (purple) and epithelial (green) populations for each of the top 22 LDA-derived features. The most discriminative features include Cells_Extent, Zernike moments (Cells_Zernike_0_0, Nuclei_Zernike_2_0), Cells_FormFactor, Cells_MedianRadius, and Cells_Solidity (positive weights, associated with epithelial morphology) and Cells_MaxFeretDiameter and Cells_Eccentricity (negative weights, associated with mesenchymal morphology). (C) Density distributions of EMScore for reference epithelial (green) and mesenchymal (purple) populations before (raw EMScore, left) and after (scaled EMScore, right) linear rescaling of discriminant scores to the [0, 1] interval. (D) Representative cells at three EMScore thresholds. Cells with EMScore < 0.2 exhibit a predominantly epithelial phenotype (rounded, compact morphology; left panel); cells near EMScore ≈ 0.5 display intermediate morphology (middle panel); and cells with EMScore > 0.8 display a predominantly mesenchymal phenotype (elongated, spindle-like morphology with extended protrusions; right panel). (E) Viability assay principle. Live cancer cells are identified as CellTracker-positive (green), DAPI-positive (blue), propidium iodide (PI)-negative cells (top row). Dead cells are identified by PI uptake (red) alongside DAPI (bottom row). (F) NK cell–mediated cytotoxicity across four lung cancer cell lines (A549, H1650, H2228, H1975) and two patient-derived CTC lines (CTC-Lu1, CTC-Lu2) at ten effector-to-target (E:T) ratios (4 h co-culture with NK-92®MI). (G) Scatter plot of mean EMScore versus NK cytotoxicity across the six cell lines, showing a positive correlation ($R^2 = 0.684$). Cells with higher EMScore (mesenchymal phenotype) are more susceptible to NK-mediated killing, whereas cells with lower EMScore (epithelial phenotype) are comparatively resistant, supporting the use of image-based EMScore as a non-invasive predictor of immune sensitivity.

4.5 Discussion

This chapter developed and evaluated a series of computational image-analysis strategies that span CTC identification and image-based functional phenotyping. The findings can be

organized around three connected contributions: a semi-automated pipeline for high-throughput, reproducible CTC image analysis from multiplex IF data; a systematic comparison of classical and deep learning classifiers, including an interpretable VGG-16/Grad-CAM workflow and an antigen-independent biolaser model; and an LDA-derived EMScore that links single-cell morphology to NK cell–mediated cytotoxicity. Each component addresses a distinct constraint of conventional CTC analysis, and together they outline a route from image-based identification toward image-informed phenotyping.

The semi-automated pipeline addressed two practical barriers in manual CTC enumeration: limited throughput and reviewer-dependent classification. By combining automated segmentation, standardized feature extraction, and ROI-level expert review, the workflow converted multiplex IF images into single-cell feature matrices while preserving manual oversight for ambiguous events [320–322,343]. CTC identification is not a simple binary marker-classification problem. Candidate events vary in epithelial marker expression, nuclear morphology, cell size, and image quality, and non-classical CTC-like events such as CK-low cells, CTC-small events, and apoptotic CTC-like events have been described[330–332]. A workflow that preserves automated throughput together with manual ROI-level review is therefore more appropriate for this task than either fully manual review or single-threshold automated detection.

The classical machine learning benchmarking yielded a finding that is informative rather than counterintuitive: logistic regression outperformed both decision tree and random forest within the 35-feature engineered space, although the absolute performance gap relative to random forest was modest (F1 = 83.7% vs. 82.0%). The PCA projection showed that between-class variance was largely concentrated along linear axes, suggesting that the engineered feature

space already encoded a near-linearly separable representation of CTCs and WBCs. Under these conditions, more complex ensemble models did not provide a clear performance advantage in this dataset and may add unnecessary complexity. The practical advantages of the linear classifier include directly inspectable coefficient weights and low computational overhead [319,320,327,344]. These properties are useful for translational workflows, especially when model outputs must be reviewed by clinicians or domain experts [329,344]. The confusion matrix showed a balanced distribution of false positives and false negatives, which is desirable in CTC analysis because both error types carry consequences: false positives may overestimate CTC burden and confound longitudinal monitoring, while false negatives may miss rare tumor cells. Reduced accuracy in one of six samples points to residual sensitivity to sample-specific staining or segmentation variability. This motivates future validation on larger, multi-center datasets with nested cross-validation to provide a more rigorous generalization estimate, particularly across sites with differing imaging hardware and staining protocols.

The transfer-learned VGG-16 model addressed a different objective. Rather than relying on predefined features, the CNN tested whether image-level patterns could be learned directly from microscopy data, and able to tell whether the resulting decisions could be made interpretable. Strong validation performance on a balanced reference dataset is expected for modern CNNs and does not by itself establish clinical suitability [345]. The more consequential observation was the spatial progression of Grad-CAM attention from diffuse, background-distributed activation in early convolutional blocks to localized activation over the nucleus and nuclear periphery in deeper blocks [323,328,329]. This pattern is consistent with the model attending to cellular structures rather than only diffuse background regions, although Grad-CAM does not prove the biological mechanism of classification. The VGG-16 results should be

interpreted as a proof-of-concept for interpretable image-based classification of cellular images; deployment on patient-derived CTCs would require retraining on a curated, CTC-specific image dataset.

The biolaser-coupled DCLC addressed a separate limitation shared by both the engineered-feature classifier and the fluorescence-based CNN: their reliance on antigen-dependent immunolabeling. Conventional IF identification depends on a small set of validated antibody markers, of which PanCK is particularly susceptible to downregulation during EMT. The DCLC instead derives classification from chromatin-dependent lasing-mode patterns produced by a single nucleic-acid stain in a Fabry–Pérot optical cavity. This signal depends on nuclear architecture rather than surface protein expression and therefore provides an input modality biophysically orthogonal to IF imaging. The observed zero-shot performance suggests that the model may capture optical features that generalize beyond a single training cell line, but this interpretation requires validation across larger clinical cohorts and additional CTC phenotypes [333–335]. CTC enumeration on clinical pancreatic and lung cancer samples that was broadly comparable to conventional IF enumeration further supports the biological fidelity of the readout, while per-sample differences between methods underscore the need for continued cross-platform validation. Together, the VGG-16 and DCLC results indicate that deep learning can be extended to CTC classification across both conventional fluorescence imaging and label-free optical biosignatures, broadening the methodological repertoire for CTC identification.

The EMScore framework extended the chapter from CTC identification to functional phenotyping. Rather than asking only whether a candidate event is a CTC, the EMScore asks whether single-cell morphology contains information about cellular state. The LDA-derived score captured a continuous epithelial–mesenchymal morphology axis, with low scores

corresponding to compact, rounded morphology and high scores corresponding to elongated, mesenchymal-like morphology. A continuous scoring framework is more appropriate than a binary epithelial mesenchymal label because EMT in tumor cells occurs along a phenotypic spectrum, and individual cells frequently occupy intermediate hybrid states [13,14,63]. In contrast to broad, high-dimensional morphological profiling, the EMScore uses a smaller, targeted feature set and a supervised discriminant framework [175,337,338,343].

The observed correlation between mean EMScore and NK-mediated cytotoxicity (R^2 = 0.684) suggests that morphology-derived features carry information about immune sensitivity that goes beyond morphology in isolation. In this dataset, more mesenchymal-like morphology was associated with higher NK cytotoxicity, in line with prior reports that EMT alters NK ligand expression and modulates NK-mediated killing in NSCLC and other tumor types [69,339,340]. This finding should be interpreted with appropriate caution. NK cytotoxicity is regulated by a balance of activating and inhibitory signals, including MHC-I expression, NKG2D ligands, DNAM-1 ligands, adhesion molecules, cell-stress pathways, and immune-synapse formation [346,347].

Several limitations should be acknowledged. First of all, the classical machine learning classifier was developed on a moderately sized engineered feature set and a single-center dataset; multi-patient and multi-center validation is required to establish generalizability across imaging platforms and staining protocols. Second, the VGG-16 model was trained on a publicly available reference blood-cell dataset rather than a patient-derived CTC dataset, so its present role is best interpreted as a proof-of-concept for interpretable image-based classification rather than a deployable CTC classifier. Third, the biolaser workflow requires further validation across additional cancer types, treatment states, and rare CTC phenotypes such as CTC clusters and

EMT-hybrid cells. The EMScore–NK cytotoxicity correlation was evaluated across six cell lines, only two of which were patient-derived CTC lines. This relationship should be replicated in additional patient-derived CTC models and, where feasible, primary clinical CTCs paired with functional immune-sensitivity assays. Finally, the present analyses are largely retrospective and platform specific. Future work should prospectively evaluate whether these image-derived classifiers and phenotypic scores improve longitudinal monitoring, treatment-response assessment, or downstream functional prioritization beyond CTC enumeration alone.

Future work should integrate these complementary approaches into a unified CTC image-analysis workflow. In such a framework, IF-based classification could provide marker-defined identity, deep learning–based image classification could support scalable processing with retained interpretability, biolaser signatures could identify antigen-low or antigen-negative cells, and the EMScore could provide morphology-based functional phenotyping related to immune sensitivity. Combining these complementary image-derived measurements may improve both CTC detection and biological interpretation in longitudinal NSCLC monitoring.

Taken together, the identification pipelines and the EMScore converge on a single translational use case of image-informed clinical decision-making. From a sample-processing standpoint, EMScore quantification requires only the cellular material enriched from a standard Labyrinth workflow.

4.6 Methods

4.6.1 Cell culture

Cells were maintained at 37 °C in 5% CO_2 and passaged at 70–80% confluence using 0.05% Trypsin-EDTA (Gibco). H1975, H1650, H2228, A549, CTC-Lu1, and CTC-Lu2 cells were cultured in RPMI-1640 (Gibco) supplemented with 10% fetal bovine serum (FBS; Sigma)

and 1% antibiotic–antimycotic (Gibco). NK-92®MI cells were maintained in suspension in Minimum Essential Medium-α (MEM-α) supplemented with 10% FBS, 12.5% horse serum, and 1% antibiotic–antimycotic. Medium was exchanged every 48–72 h between passages. Cell lines were routinely verified as mycoplasma-negative (Lonza).

4.6.2 Patient sample collection and Labyrinth processing

The experimental protocol was approved by the Institutional Review Board and Scientific Review Committees of the University of Michigan (HUM00119934), and all patients provided written informed consent. All patients had metastatic EGFR-mutant or ALK lung adenocarcinoma; two patients carried ALK rearrangements and six carried EGFR mutations as confirmed by primary tumor sequencing. Most patients were receiving kinase inhibitor therapy; one patient initially received chemotherapy and transitioned to kinase inhibitor therapy upon identification of an EGFR mutation by next-generation sequencing.

Blood samples were collected in EDTA tubes and processed through the Labyrinth within 2 h of collection. Red blood cells were depleted by Ficoll-Paque™ PLUS (GE Healthcare) density separation according to the manufacturer's protocol. The plasma and PBMC fractions were collected and diluted 1:5 in PBS based on the original blood volume. The diluted sample was processed through the Labyrinth at 2,500 μL/min, and the output from outlet 2 was collected for downstream immunofluorescence or single-cell RNA sequencing.

4.6.3 Semi-automated image processing pipeline for CTC identification

After Labyrinth microfluidic CTC enrichment, multiplex immunofluorescence images acquired with Nikon using the same method described in Chapter 2. Then the multiplex immunofluorescence images were processed using a custom MATLAB-based image analysis workflow. Each composite image contained three fluorescence channels corresponding to pan-

cytokeratin (PanCK), CD45, and DAPI staining. The composite RGB image was imported into MATLAB and separated into red, green, and blue grayscale channels, corresponding to PanCK, CD45, and DAPI signals, respectively. The separated channels were used for downstream nuclear segmentation, tumor-marker detection, and leukocyte-marker review.

4.6.3.1 DAPI-based nuclear segmentation

Nucleated cells were first identified from the DAPI channel. The pixel-intensity distribution of the DAPI image was calculated using histogram analysis. To account for variation in staining intensity and image background, multiple intensity thresholds were estimated from the DAPI histogram, including iterative average-intensity values and upper-intensity regions of the pixel distribution.

Binary nuclear masks were generated by applying intensity thresholds to the DAPI channel. The initial masks were refined using morphological operations, including dilation, erosion, hole filling, median filtering, and area-based object filtering. Small objects and background artifacts were removed using size thresholds, while hole filling was used to recover complete nuclear regions. Connected nuclear regions were then identified, and centroids were extracted using region-based measurements. The number of detected centroids was used to estimate the number of nucleated cells within each image.

4.6.3.2 Candidate CTC region identification

Candidate CTC regions were identified by combining DAPI-based nuclear masks with PanCK-positive regions. The PanCK channel was thresholded to generate a binary epithelial-marker-positive mask. This PanCK-positive mask was then intersected with the DAPI-derived nuclear mask to retain events that were both nucleated and epithelial-marker positive. Median filtering was applied to reduce isolated noisy pixels in the resulting candidate cancer-cell mask.

The CD45 channel was used as a negative leukocyte marker during image review. Candidate objects were overlaid onto the original composite image to support visual inspection and exclusion of obvious artifacts, debris, or leukocyte-associated events.

4.6.4 Feature extraction and normalization

For each retained candidate region of interest, quantitative image-derived features were extracted for downstream analysis. The extracted features included fluorescence intensity descriptors and morphological descriptors. Intensity-based features included measurements such as mean intensity, standard deviation, mode, minimum intensity, and maximum intensity. Morphological features included area, perimeter, width, height, major axis length, minor axis length, Feret diameter, minimum Feret diameter, circularity, aspect ratio, roundness, and solidity.

In total, 35 quantitative features were extracted for each candidate cell. These ROI-level measurements transformed each image object into a structured single-cell feature vector for statistical analysis and machine learning classification.

The extracted feature matrix was imported into MATLAB for downstream analysis. Each row represented an individual cell, and each column represented either a quantitative image-derived feature or the binary cell label. Cells were labeled as tumor cells or non-tumor cells based on the available annotation. Missing values were removed before model training and statistical analysis.

Summary statistics were calculated for each feature. Correlation analysis was performed between individual features and the binary tumor-cell label to identify features most associated with CTC identity. Feature distributions between tumor and non-tumor cells were visualized using boxplots, scatter plots, and scatter histogram plots. Two-sample t-tests were used to compare selected high-ranking features between the two groups.

4.6.5 Machine learning model training and evaluation

Principal component analysis was performed on the normalized feature matrix to evaluate the global structure of the engineered feature space. Before PCA, features were standardized to reduce the influence of scale differences among intensity and morphology measurements. Cells were projected onto the first two principal components and visualized according to tumor-cell status. Feature loadings from the principal components were examined to determine which image-derived measurements contributed most strongly to the major axes of variation.

Supervised machine learning models were trained to classify cells as CTCs or non-CTCs using the extracted image-derived features. Three traditional classifiers were evaluated: logistic regression, decision tree, and random forest/ensemble classification.

The labeled dataset was partitioned into an 80% training set and an independent 20% held-out validation set using stratified splitting to preserve class proportions across partitions. Feature standardization to z-scores was applied using parameters derived exclusively from the training partition and subsequently applied to the validation set to prevent data leakage. Within the training partition, each classifier was evaluated using three-fold cross-validation with a fixed random seed (seed = 0) for reproducibility. Final classifier performance was assessed on the held-out validation set, which was not used at any stage of model development, feature selection, or hyperparameter tuning.

For all classifiers, predicted probabilities or scores were converted to binary class labels using a decision threshold of 0.5. Model performance was quantified using accuracy, recall, precision, F1 score, and confusion matrices. Feature importance was also examined for tree-based models to identify image-derived measurements contributing to classification.

4.6.6 Convolutional neural network and Grad-CAM interpretability

A transfer learning framework was developed to classify brightfield leukocyte images into five major white blood cell categories: basophils, eosinophils, lymphocytes, monocytes, and neutrophils. Images were organized in class-specific directories and imported using the TensorFlow/Keras image loading pipeline. All images were resized to 200 × 200 pixels with three color channels and loaded in mini batches of 16. The dataset comprised 3,873 blood-cell images. A training–validation pool of 2,133 images was randomly partitioned 90:10 into 1,920 training and 213 validation images, and a separate independent test set of 1,740 images was reserved for final model evaluation. Prior to model training, image files were screened for valid JFIF image formatting to exclude corrupted or incompatible files.

To improve generalizability and reduce overfitting, data augmentation was applied to the training images. Augmentation included random horizontal and vertical flipping, random rotation, random translation along both spatial dimensions, and random contrast perturbation. These transformations were applied only during model training, whereas validation and test images were evaluated without augmentation. Before input into the convolutional network, images were center cropped from 200 × 200 pixels to 128 × 128 pixels and preprocessed using the standard VGG16 preprocessing function to match the input distribution expected by the pretrained model.

The classifier was constructed using VGG16 pretrained on ImageNet as the convolutional feature extractor. The original ImageNet classification layers were removed, and the convolutional backbone was retained without fine-tuning to preserve the pretrained feature representations. A custom fully connected classification head was appended to the VGG16 backbone. The classification head consisted of a flattening layer, two dense layers with ReLU

activation, a dropout layer for regularization, and a final softmax output layer corresponding to the five leukocyte classes. The dense layers contained 4,096 and 1,072 units, respectively, and dropout was applied at a rate of 0.2.

The model was optimized using the Adam optimizer with a learning rate of 1×10^{-4} and sparse categorical cross-entropy loss. Because the dataset was imbalanced across leukocyte classes, class weighting was incorporated during training to reduce bias toward more abundant cell types. Higher weights were assigned to underrepresented classes, including basophils, lymphocytes, and monocytes, whereas eosinophils and neutrophils were assigned lower weights. Model training was performed for up to 25 epochs, with validation performance monitored throughout training. Model checkpoints were saved during training to preserve intermediate model states.

Final model performance was assessed on the independent test set that was not used during training or validation. Overall performance was quantified using test accuracy and loss. Class-specific performance was evaluated using precision, recall, and F1-score for each leukocyte category. A confusion matrix was generated to visualize the distribution of correct and incorrect predictions across the five classes and to identify class-specific misclassification patterns.

4.6.7 Biolaser-based antigen-independent CTC classification (DCLC)

Single-cell laser measurements were performed using a home-built laser emission microscope and a micro Fabry-Pérot (F-P) cavity consisting of two custom dielectric mirrors (R > 99.5% at 510–550 nm). Cells were stained with the nucleic acid dye SYTO 9 (50 μM, 25 min, 4 °C), deposited on the cavity mirror, and measured at four optical pumping powers (25, 45, 90, and 150 μJ/mm²) using a 473 nm nanosecond pulsed laser. The four power-resolved lasing mode pattern images were stacked into the Hyper Lasing Signature (HLS) for each cell. The DCLC

model comprised a parallel single-cell laser encoder (PScLE) of four independently trained ResNet-101 blocks and a fully connected cell classifier (256 → 32 → 8 neurons, ReLU activations; softmax output). The PScLE was trained on Pancreatic 368 cells and WBCs from six healthy donors; zero-shot generalization was evaluated on Pancreatic 322 and Pancreatic 331 cell lines. The classifier threshold was set at 0.2 to optimize specificity for rare-cell detection.

4.6.8 EMScore derivation via Linear Discriminant Analysis

Morphological features extracted from merged composite images (DAPI, nucleic acid, F-actin/membrane) were first processed through an automated quality-control filter in which a random forest classifier trained on expert-annotated intact cells versus artifacts discarded debris, imaging noise, and mis-segmented clusters. Retained features were input to a Linear Discriminant Analysis model trained on known epithelial (E) and mesenchymal (M) reference cells. Discriminative features included Cells_Extent, Zernike moments (Cells_Zernike_0_0, Nuclei_Zernike_2_0), FormFactor, and Solidity. Raw LDA scores were rescaled to [0, 1] by linear transformation to yield the EMScore.

4.6.9 NK cell cytotoxicity live imaging

CTC-Lu2 cells were seeded at 2×10^3 cells per well in 18-well μ-Slides (ibidi) and allowed to adhere overnight at 37 °C in 5% CO_2. Culture medium was replaced with MEM-α containing 1×10^4 NK-92®MI cells. Imaging was performed on the Nanolive 3D Explorer at 30-s intervals for 1-h sessions. Images were processed using the Nanolive Steve software, and screenshots were focus-stacked to permit simultaneous visualization of NK and cancer cells.

4.6.10 Quantitative image-based NK cytotoxicity assay

Cancer cells were dissociated with 0.05% Trypsin-EDTA and labeled with CellTracker Red CMTPX (Invitrogen) for 45 min in serum-free medium, washed three times with PBS,

counted on a hemocytometer, and seeded at 5,000 cells/well (100 μL; 96-well plate, Corning) for overnight adherence. NK-92®MI cells were resuspended in a 1:1 mix of RPMI-1640 and MEM-α, counted, diluted to the desired E:T ratios, and added to wells after gentle removal of culture medium. Co-culture proceeded for 4 h.

Wells were washed with 200 μL PBS and stained with Calcein-AM (BD Biosciences) and Hoechst (Invitrogen; 10 μg/mL) for 20 min at 37 °C in the dark. Staining reagents were replaced with PBS, and plates were imaged on a Nikon Ti2 inverted fluorescent microscope (10×, three channels: DAPI, FITC, Cy3). Images were batch-processed in NIS-Elements using "Remove Average Background," intensity- and size-based thresholding for Calcein-AM and CellTracker, "Bright Spot Detection" for nuclei, and post-processing morphology operations. Live cancer cells were defined as triple-positive objects (CellTracker, Calcein-AM, DAPI) and enumerated per well.

4.6.11 Statistical analysis

Statistical analyses were performed in GraphPad Prism 9.4.0. Between-group significance was assessed by two-sided unpaired *t*-tests assuming Gaussian distributions at a 95% confidence level. Error bars represent standard deviation. In simple linear regression, R^2 represents goodness of fit.

4.7 Author Contributions

I developed the complete semi-automated image-analysis and classical machine-learning pipelines described in Sections 4.4.1–4.4.2, including image processing, feature extraction, model training, classifier benchmarking, and performance evaluation. For the deep-learning component, I developed and implemented the Grad-CAM interpretability analysis to evaluate the image regions contributing to model classification. I contributed to algorithm-development

discussions and assisted with sample preparation for the biolaser and EMScore studies. Additionally, I ran parts of the EMScore computational pipeline for data collection, performed RNA-seq analysis, and tested different algorithms for EMT-associated morphological classification and scoring.

Chapter 5 Conclusions and Future Directions

5.1 Research Summary

This dissertation addressed a central challenge in the clinical management of ALK NSCLC: how to capture tumor heterogeneity and therapeutic adaptation through a minimally invasive, longitudinal, and biologically informative biomarker. Conventional tissue biopsy provides only a static and spatially limited snapshot, while cfDNA lacks cellular context and cannot fully resolve phenotypic plasticity. To overcome these limitations, this work established a multi-layered framework for CTC analysis that integrated label-free microfluidic enrichment, single-cell molecular profiling, longitudinal phenotyping, and automated image-based computational analysis.

In Chapter 2, single-cell profiling of CTCs from ALK NSCLC patients revealed that these cells exhibit a distinct EMT-associated, oncogenic, and immune-evasive state. ALK CTCs showed enrichment of EMT-related programs, platelet-interacting genes, and signaling patterns consistent with downstream ALK activation, while cell–cell interaction analyses suggested TGFβ-mediated NK cell suppression and CCL5–CCR5 signaling with CD8+ T cells. These findings suggest that CTCs recapitulate key biological features of ALK-driven disease and may engage the circulating immune microenvironment through candidate immunosuppressive interactions.

In Chapter 3, longitudinal CTC monitoring showed that dynamic changes in CTC composition were more informative than static enumeration alone. A greater decrease in hybrid epithelial-mesenchymal CTCs was associated with improved progression-free survival, and

reductions in only CK+ CTC subpopulations were associated with improved overall survival. Longitudinal scRNA-seq further revealed persistent EMT-associated transcriptional evolution and a CTC-derived gene signature that was independently associated with poorer survival in an external TCGA validation cohort. Together, these findings provide exploratory support for the prognostic value of serial CTC profiling and motivate prospective investigation of its utility for tracking therapeutic response and emerging resistance.

In Chapter 4, I developed and benchmarked computational pipelines for automated CTC image analysis, including traditional machine learning, deep learning with interpretability, and an antigen-independent biolaser-based detection framework. Logistic regression achieved the best single-cell CTC classification performance among classical models, while the DCLC system achieved high sensitivity and specificity for CTC detection. Importantly, an LDA-derived EMScore correlated with NK cell–mediated cytotoxicity, linking morphology-based phenotyping to a functional immune readout. These results show that computational image analysis can extend beyond classification to provide biologically and clinically interpretable phenotypes.

Together, the three aims converge on several themes that define the contribution of this work as a whole for precision medicine in ALK NSCLC. Across all three chapters, EMT emerges as the convergent biological axis. Although approached from distinct experimental angles—cross-sectional transcriptomics, longitudinal phenotype tracking, and image-based morphological scoring—each chapter independently identified epithelial–mesenchymal plasticity as a defining feature of ALK CTCs. The cross-sectional analysis in Chapter 2 revealed enrichment of canonical EMT transcripts (VCL, CLU, GATA1) and a depletion of purely epithelial CTC subtypes; the longitudinal data in Chapter 3 showed that EMT-associated programs persist across treatment timepoints and that the trajectory of the hybrid E/M CTC

fraction is itself prognostic; and the morphology-based EMScore in Chapter 4 captured the same continuum at the single-cell image level. The convergence of three orthogonal modalities on the same biological axis indicates that EMT is not an artifact of any single assay but a core feature of ALK CTC biology, with measurable correlates at the protein, transcriptomic, and morphological levels.

The CTC–platelet immune-evasion triad further points to specific druggable vulnerabilities. The transcriptional signature identified in Chapter 2 is mechanistically coherent rather than incidental: simultaneous upregulation of platelet-interacting genes (PF4, ITGA2, PPBP), oncogenic transcripts (TIMP1, SMOX), and EMT regulators, coupled with downregulation of innate and adaptive immune signaling, describes a CTC state transcriptionally primed for survival in circulation. The cell–cell interaction analysis implicated TGF-β–mediated NK suppression and CCL5–CCR5 signaling toward $CD8^+$ T cells as candidate immunosuppressive mechanisms, while the EMScore–NK cytotoxicity correlation in Chapter 4 indicates that the morphological correlates of this state track susceptibility to NK-mediated killing. Together, these findings nominate three therapeutic axes for prospective evaluation in ALK NSCLC: anti-platelet adjunctive therapy, TGF-β or CCR5 blockade in combination with immune checkpoint inhibitors, and CDK4/6 inhibitor combinations in patients with progression-emergent CDK4 copy-number gain. The detection of CDK4 amplification at progression timepoints in two of three longitudinally sequenced patients in Chapter 3 is of particular interest, as CDK4 amplification is a recognized ALK-independent bypass mechanism; non-invasive detection through CTC profiling could in principle support rational combination therapy without repeat tissue biopsy.

These findings support a shift from CTC enumeration to dynamic phenotype monitoring.

Conventional CTC applications, including the FDA-approved system, have been built on enumeration as a continuous variable, treating CTCs as a homogeneous population. Chapter 3 suggests that enumeration alone may fail to capture prognostically relevant information: cross-sectional measurements at individual timepoints failed to discriminate progression, whereas longitudinal trajectories of phenotypic subsets did. The clinical implication is that prognostic value resides not in any single CTC count but in the direction and composition of change across visits. These findings support a shift in how CTC analysis is applied from a static prognostic snapshot to a monitoring tool capable of detecting therapeutic adaptation as it occurs—an application for which CTCs are well suited, because they preserve the cellular phenotype that mediates resistance and, unlike tissue biopsy, can be sampled serially without procedural risk.

Multi-modal single-cell phenotyping is necessary because CTC heterogeneity is itself multi-scale. A practical lesson from this work is that no single assay captures the full CTC population: bulk transcriptomics obscures intra-patient heterogeneity, single-marker immunofluorescence misses transcriptionally distinct subpopulations, and morphology alone does not resolve molecular subtype. Each chapter contributed a different layer, transcriptional state (Chapter 2), longitudinal dynamics (Chapter 3), and image-based functional phenotype (Chapter 4), and the combined framework resolved heterogeneity at scales that no individual component could address. The 45-gene CTC-downregulated signature, derived from longitudinal single-cell data and validated in an independent population-level bulk RNA-seq cohort (TCGA-LUAD; n = 500), suggested that the signature may capture conserved determinants of disease aggressiveness. Also, it showed single-cell CTC analysis could yield clinically generalizable signatures while retaining the resolution needed to dissect cellular heterogeneity.

This work positions ALK NSCLC CTCs within and beyond the existing literature. Prior

single-cell studies of ALK CTCs have focused on genomic resistance mechanisms at single post-progression timepoints, and bulk transcriptomic studies have documented EMT in heterogeneous NSCLC populations. To our knowledge, the present work integrates these directions by providing the first single-cell transcriptomic, immunologic, and longitudinal characterization of CTCs in ALK NSCLC specifically. The use of label-free Labyrinth enrichment was central to this contribution: by avoiding EpCAM-based capture, the workflow retained the EMT-shifted populations that conventional EpCAM-based platforms may preferentially underrepresent or miss, so the recovered CTC fraction reflects the full phenotypic spectrum rather than its epithelial subset alone.

Taken together, this dissertation repositions CTCs in ALK NSCLC from an enumeration-based prognostic snapshot toward a serially sampled, multi-parameter readout of therapeutic adaptation. The same EMT axis surfaced through three orthogonal modalities, and a single-cell–derived signature retained prognostic value at the population level, indicating that high-resolution CTC profiling yields signals that are both mechanistically specific and clinically generalizable. The translational hypotheses this raises, non-invasive detection of CDK4-driven bypass to guide CDK4/6 combinations, TGF-β or CCR5 blockade with checkpoint inhibitors, and morphology-based immune-sensitivity scoring, are each anchored to a specific finding here and are framed as prospective tests rather than established utility. More broadly, because the framework depends on label-free enrichment and computational phenotyping rather than on any single marker, it defines a template applicable to other oncogene-driven cancers in which phenotypic plasticity drives resistance. Realizing this will require standardized, multi-institutional cohorts with paired tissue and ctDNA; what this dissertation contributes is a defined set of measurements, and the rationale for why each one matters, on which such studies can

build.

5.2 Limitations and Future Work

Although this dissertation supports the value of CTC profiling in ALK NSCLC, several limitations should be considered. These limitations relate to clinical cohort design, biological representation, technical variability, computational validation, and clinical translation. They also define important next steps for improving the robustness and translational value of CTC-based analysis in ALK NSCLC.

Cohort and sampling design: A general limitation of this dissertation is the size of the clinical cohort. ALK NSCLC is a relatively uncommon molecular subtype. Therefore, the number of available patients was limited, especially for single-cell RNA sequencing and longitudinal survival analysis. The number of clinical events was also limited, which restricted the ability to perform fully powered multivariable survival models. As a result, the associations between CTC phenotypes, longitudinal CTC changes, progression-free survival, and overall survival should be interpreted as exploratory. This limitation is common in many CTC studies in NSCLC. Prior reviews have noted that clinical CTC studies often use different isolation platforms, different enumeration thresholds, and relatively small or heterogeneous patient cohorts, which makes cross-study comparison difficult [348].

Biological validation: Longitudinal sampling design is another important limitation, particularly for Chapter 3. The longitudinal samples were not collected at fully uniform clinical timepoints. Patients also differed in treatment history, treatment line, ALK inhibitor exposure, disease burden, and progression status. These factors may influence CTC number and CTC phenotype. They may also affect how CTC dynamics relate to clinical outcome. Future studies should use a more standardized sampling design. Ideally, blood samples would be collected at

diagnosis, treatment initiation, early response, best response, progression, and treatment change. A larger biobank-style cohort, with standardized sample collection and linked clinical outcomes, would provide a stronger framework for testing the prognostic value of CTC dynamics.

Clinical translation could require standardization across the full experimental and computational workflow. This includes blood collection, sample processing, CTC enrichment, immunofluorescence staining, single-cell sequencing, image acquisition, and computational analysis. Without standardized workflows, technical variation may be difficult to separate from biological variation. Future studies should therefore establish reproducible protocols across sites. This will be important before CTC-based measurements can be compared across institutions or used in clinical decision-making.

On the biological side, one important limitation is that CTCs may not fully represent the entire tumor. CTCs are shed from primary or metastatic lesions, but they are not a random sample of all tumor cells. Cells that enter the blood must detach, survive circulation, avoid immune clearance, and remain viable. Therefore, CTCs may be enriched for invasive, EMT-like, platelet-associated, or immune-evasive states. These states are clinically meaningful. However, they may not capture all cellular states within the primary tumor or metastatic lesions. Prior work has suggested that CTCs can inform tumor heterogeneity, tumor evolution, and treatment resistance. At the same time, CTCs should be interpreted as a selected circulating population rather than a complete substitute for multi-region tumor sampling [299].

Future work should directly compare CTCs with matched tumor tissue and ctDNA. This would address whether CTCs reflect the molecular and cellular heterogeneity of the primary tumor, metastatic sites, or a selected circulating subpopulation. Matched analyses could include CTC scRNA-seq, CTC copy-number profiles, CTC immunofluorescence phenotypes, ctDNA,

primary tumor sequencing, and metastatic tissue sequencing. This comparison would clarify whether CTCs capture broad tumor heterogeneity or mainly reflect the tumor fraction most capable of dissemination and survival in blood.

The comparison between ALK and wild-type NSCLC could also be expanded. This dissertation identified differences in CTC phenotype and transcriptional state between these groups. However, additional clinical and demographic variables may contribute to these differences. Sex- or gender-specific differences were not fully evaluated in the current cohort. Future studies should test whether CTC burden, EMT phenotype, immune-related signatures, and longitudinal CTC changes differ by sex or gender. These analyses should also account for age, smoking history, disease stage, treatment exposure, and sample timing.

CTC enrichment remains another biological and technical limitation. The Labyrinth platform avoids EpCAM-based positive selection. This is useful because EMT-like CTCs may have reduced epithelial marker expression. However, label-free enrichment is still affected by cell size, deformability, and other physical properties. This creates a different form of selection bias. Prior NSCLC studies have shown that CTCs are phenotypically heterogeneous and can include epithelial, mesenchymal, hybrid epithelial–mesenchymal, EpCAM-low, and clustered populations [349]. In metastatic NSCLC, Labyrinth-based isolation recovered heterogeneous CTCs and CTC clusters, and many recovered CTCs or clusters were EpCAM-negative, supporting the need for marker-independent enrichment strategies [179].

At the same time, physical-property-based enrichment may underrepresent some CTC subtypes. Size-based or inertial methods may preferentially recover larger or less deformable tumor cells. Smaller CTCs may be lost. This is relevant in NSCLC because previous work has described non-classical CTC categories, including “CTC-small,” CK-low cells, and apoptotic

CTC-like events. These findings indicate that NSCLC CTCs are not a single uniform population. Therefore, the CTCs analyzed in this dissertation likely represent a biologically informative enriched fraction, but not necessarily the complete spectrum of CTCs present in each patient. Future work should combine Labyrinth enrichment with orthogonal validation methods, such as enrichment-free imaging, ALK FISH, targeted mutation testing, CNV-based malignancy scoring, expanded immunofluorescence panels, and antigen-independent detection.

The immune-suppressive environment observed in ALK CTCs requires functional validation. In Chapter 2, transcriptomic and cell-cell interaction analyses suggested immune-related mechanisms, including possible TGF-β-mediated NK-cell suppression. These findings are informative, but they are based mainly on computational inference. They do not yet prove that ALK CTCs secrete TGF-β or that this pathway improves CTC survival in peripheral blood. Future experiments should measure TGF-β secretion from CTC-derived models or patient-derived CTC cultures. ELISA, Luminex, intracellular staining, or single-cell transcript detection could be used. Co-culture assays with NK cells or CD8 T cells could then test whether CTC-conditioned media suppress immune-cell activity. TGF-β blocking antibodies or TGF-β receptor inhibitors could help test whether this effect is causal.

The current molecular analyses also focus mainly on downstream gene expression. This dissertation identifies EMT-associated genes, platelet-interaction genes, oncogenic genes, and immune-related signatures. However, these expression patterns may be driven by upstream transcription factors or signaling pathways. Future analysis should infer upstream regulators more directly. Tools such as DoRothEA, PROGENy, and decoupleR can be used to infer transcription factor and pathway activities from transcriptomic data. DoRothEA provides curated, signed TF–target regulons for estimating transcription factor activity, whereas

PROGENy infers pathway activity from perturbation-derived pathway-responsive gene signatures. decoupleR provides a unified framework for applying these prior-knowledge resources and scoring methods to bulk and single-cell omics data. These approaches have been benchmarked for transcriptomic data and evaluated for single-cell applications [350–352]. Ligand-receptor tools such as CellChat, NicheNet, or LIANA could also help identify signaling axes between CTCs, platelets, and immune cells. These analyses would help clarify whether the observed CTC states are driven by ALK signaling, EMT regulators, TGF-β signaling, platelet-derived factors, or immune-cell-derived signals.

The CDK4 amplification finding is also early. Its detection at progression timepoints suggests a possible resistance mechanism. However, this observation was made in a small number of longitudinally sequenced patients. It should therefore be treated as a candidate finding. Future work should test whether CDK4 amplification appears recurrently in larger ALK NSCLC CTC cohorts. It should also be compared with matched ctDNA and tissue sequencing. If validated, this finding could support future studies of CDK4/6 inhibitor combinations in selected patients.

Clinical Translation and Predictive Modeling: the survival findings should be interpreted with caution. Chapter 3 supports the prognostic value of longitudinal CTC monitoring. However, the current cohort does not define clinical cutoffs. The associations between CTC changes and survival need validation in larger cohorts. Future studies should test whether specific CTC thresholds, CTC subtype changes, or molecular signatures remain associated with outcome after adjusting for clinical variables. Only after this validation could these markers be considered for clinical risk stratification.

On the computational side, several analyses can be strengthened. CTC number has not

yet been fully modeled against tumor burden, radiographic imaging, or ctDNA. Future work should compare CTC counts and CTC subtype frequencies with imaging-based tumor burden, treatment response, ctDNA level, and progression status. Pearson or Spearman correlation could be used for simple continuous associations. More advanced models could then evaluate whether CTC changes provide information beyond imaging or ctDNA. Prior longitudinal NSCLC studies have shown that changes in CTC counts during treatment may provide prognostic information, supporting the need to model CTC dynamics rather than relying only on baseline enumeration [209].

Dynamic phenotype monitoring is one of the main strengths of this dissertation. However, it has not yet been shown to change treatment decisions. Chapter 3 suggests that longitudinal CTC trajectories may be more informative than single-timepoint enumeration. The next step is to test whether these changes can guide patient management. Prospective studies could evaluate whether CTC dynamics predict progression earlier than imaging. They could also test whether CTC phenotypes identify patients who may benefit from closer monitoring, treatment change, or combination therapy.

Computational analysis, data normalization, integration, and external validation: changes in cell composition should be analyzed more systematically. This is especially important for the immune-cell populations observed in the single-cell data. Future studies could use compositional analysis tools such as scCODA to test whether specific immune-cell fractions or CTC-associated cell states change across patient groups, treatment timepoints, or progression status. scCODA was developed as a Bayesian model for compositional single-cell data analysis and is designed to account for uncertainty and compositional constraints in cell-type proportion data. These analyses would help distinguish true shifts in cell abundance from transcriptional

changes occurring within the same cell type.

The FLEX single-cell workflow is another area for further optimization. Although FLEX supports fixed-sample processing and is well suited for clinical sample handling, normalization, batch correction, and cross-patient integration remain important challenges. Future studies should systematically compare approaches such as SCTransform, Harmony, FastMNN, LIGER, and Seurat integration to determine their ability to reduce technical variation while preserving biologically meaningful differences. Particular attention should be given to separating patient-specific biological signals from sequencing batch, sample-processing date, sequencing depth, and chemistry effects.

Integration performance should also be quantitatively assessed using metrics such as Local Inverse Simpson's Index, kBET, graph connectivity, and silhouette scores. These analyses will help ensure that technical effects are minimized without removing clinically relevant biological heterogeneity. This is especially important for comparing ALK NSCLC and wild-type NSCLC samples, interpreting longitudinal changes across timepoints, and integrating samples processed across different dates or sequencing runs.

External validation is also needed for the transcriptomic findings reported in this dissertation. The EMT-associated, platelet-associated, and immune-related signatures identified in ALK CTCs should be evaluated in independent patient cohorts and publicly available NSCLC single-cell datasets whenever possible. Such validation would help determine the generalizability and robustness of these findings across patient populations, sequencing platforms, and study designs.

The EMScore framework also has limitations. EMScore links cell morphology with EMT state and NK-cell sensitivity. However, it was developed mainly in 2D in vitro models. These

models do not fully capture the physical and cellular environment that tumor cells experience in vivo. They also do not reproduce the full complexity of CTC-immune interactions in blood. Future work should validate EMScore in patient-derived CTC models, 3D spheroids, organoids, and microfluidic co-culture systems. Primary NK cells and autologous immune-cell models would also improve biological relevance. These studies could test whether specific morphological features, such as actin organization, cell elongation, nuclear shape, or membrane protrusions, influence NK-cell contact and killing.

The image-based classifier in Chapter 4 also requires broader validation. The current models were developed in a controlled single-center imaging workflow. Their performance may change with different microscopes, staining batches, exposure settings, operators, and institutions. Future work should test these classifiers on external datasets. Patient-level train-test separation should be used. Model calibration and uncertainty should also be reported. In parallel, the image-processing pipeline could be further developed using CellProfiler, Cellpose, or ilastik. These tools may improve segmentation and feature extraction in complex patient-derived images.

Deep learning models will also need more validation before clinical use. Grad-CAM and related tools can improve interpretability, but visual explanation alone is not sufficient. Future studies should evaluate model performance in independent cohorts. They should also assess calibration, false-positive patterns, and failure cases. This is particularly important for rare-cell detection, where small changes in specificity can have a large effect on the final CTC call.

The biolaser platform provides a promising antigen-independent strategy for CTC detection. However, it also has workflow limitations. Current single-cell lasing measurements require specialized instrumentation and are not yet optimized for high-throughput clinical use.

Future work should focus on improving detection efficiency, reducing measurement time, and integrating microfluidic channels within laser cavities. This could reduce cell loss and improve downstream single-cell recovery. Larger patient studies are also needed to validate the platform across cancer types and clinical settings.

Finally, multimodal integration remains an important future direction. This dissertation incorporates immunofluorescence phenotyping, single-cell transcriptomics, inferred CNV profiles, longitudinal clinical outcomes, image-based classification, biolaser detection, and EMScore analyses. However, these data modalities are currently analyzed largely as separate layers. Future studies should develop unified frameworks capable of integrating molecular, morphological, clinical, and functional information at the single-cell and patient levels. Emerging approaches such as weighted nearest-neighbor analysis, MOFA+, totalVI, and MultiVI may provide useful strategies for integrating heterogeneous data types. Such multimodal models may ultimately provide a more comprehensive representation of CTC biology and improve prediction of treatment response, disease progression, and immune interactions.

5.3 Conclusions

In conclusion, this dissertation demonstrates the potential of CTCs as dynamic, multidimensional biomarkers capable of reflecting tumor state and candidate immune interactions, and of tracking compositional changes associated with treatment in ALK NSCLC. By integrating microfluidic enrichment, single-cell molecular profiling, longitudinal phenotyping, and automated computational analysis, this work establishes a framework for interpreting CTCs as biologically and prognostically informative readouts of disease evolution. The findings support a model in which phenotypic plasticity, immune evasion, and clonal adaptation are reflected in the circulating compartment and can be monitored non-invasively

over time. More broadly, this dissertation provides a generalizable foundation for translating CTC analysis into precision oncology, with the goal of enabling earlier detection of resistance, improved prognostic stratification, and more personalized therapeutic decision-making.

Bibliography

1. Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
2. Molina, J. R. *et al. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship*. *Mayo Clin Proc* vol. 83 (2008).
3. Tang, F. H. *et al.* Recent advancements in lung cancer research: a narrative review. *Translational Lung Cancer Research* vol. 14 975–990 (2025).
4. Jeon, H. et al. Management of Non-Small-Cell Lung Cancer. *Lung* vol. 203 (2025).
5. Wang, C. *et al.* Persistent increase and improved survival of stage I lung cancer based on a large-scale real-world sample of 26,226 cases. *Chin. Med. J. (Engl).* **136**, 1937 (2023).
6. Guo, Q. *et al.* Current treatments for non-small cell lung cancer. *Frontiers in Oncology* vol. 12 (2022).
7. Kumar, R. *et al.* Targeting the PD-1/PD-L1 axis in non–small cell lung cancer. *Curr. Probl. Cancer* **41**, 111–124 (2017).
8. Huang, Q. *et al.* Advances in molecular pathology and therapy of non-small cell lung cancer. *Signal Transduction and Targeted Therapy 2025 10:1* **10**, 186- (2025).
9. Jeon, D. S. *et al.* Five-Year Overall Survival and Prognostic Factors in Patients with Lung Cancer: Results from the Korean Association of Lung Cancer Registry (KALC-R) 2015. *Cancer Research and Treatment : Official Journal of Korean Cancer Association* **55**, 103

(2022).

10. Rahal, Z. et al. Tumour and microenvironment crosstalk in NSCLC progression and response to therapy. *Nat. Rev. Clin. Oncol.* **22**, 463 (2025).
11. Ma, P. *et al.* Simultaneous evolutionary expansion and constraint of genomic heterogeneity in multifocal lung cancer. *Nature Communications 2017 8:1* **8**, 823- (2017).
12. Gilson, P. et al. Deciphering Tumour Heterogeneity: From Tissue to Liquid Biopsy. *Cancers (Basel).* **14**, 1384 (2022).
13. Jolly, M. K. & Celià-Terrassa, T. Dynamics of Phenotypic Heterogeneity Associated with EMT and Stemness during Cancer Progression. *J. Clin. Med.* **8**, 1542 (2019).
14. Li, D. *et al.* Heterogeneity and plasticity of epithelial–mesenchymal transition (EMT) in cancer metastasis: Focusing on partial EMT and regulatory mechanisms. *Cell Prolif.* **56**, e13423 (2023).
15. Chen, J. W. & Dhahbi, J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Sci. Rep.* **11**, 13323 (2021).
16. Attili, I. et al. Co-Occurring Genomic Alterations in NSCLC: Making Order into a Crowded List. *Cancers (Basel).* **17**, 2388 (2025).
17. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e18 (2018).
18. Schneider, J. L. et al. ALK-positive lung cancer: a moving target. *Nature Cancer* vol. 4 330–343 (2023).
19. Du, X. et al. ALK-rearrangement in non-small-cell lung cancer (NSCLC). *Thorac. Cancer* **9**, 423 (2018).

20. Urbanska, E. M. et al. Changing ALK-TKI-resistance mechanisms in rebiopsies of ALK-rearranged nsclc: ALK- and BRAF-mutations followed by epithelial-mesenchymal transition. *Int. J. Mol. Sci.* **21**, (2020).

21. Roskoski, R. Anaplastic lymphoma kinase (ALK) inhibitors in the treatment of ALK-driven lung cancers. *Pharmacol. Res.* **117**, 343–356 (2017).

22. Parvaresh, H. et al. Unraveling the Potential of ALK-Targeted Therapies in Non-Small Cell Lung Cancer: Comprehensive Insights and Future Directions. *Biomedicines* **12**, 297 (2024).

23. Shen, J. *et al.* EML4-ALK G1202R mutation induces EMT and confers resistance to ceritinib in NSCLC cells via activation of STAT3/Slug signaling. *Cell. Signal.* **92**, (2022).

24. Katayama, R. *et al.* Cancer: Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. *Sci. Transl. Med.* **4**, (2012).

25. Miyamoto, D. T. *et al.* RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science (1979).* **349**, 1351–1356 (2015).

26. Katayama, R. *et al.* Two novel ALK mutations mediate acquired resistance to the next-generation ALK inhibitor alectinib. *Clinical Cancer Research* **20**, 5686–5696 (2014).

27. Pan, Y. et al. The Resistance Mechanisms and Treatment Strategies for ALK-Rearranged Non-Small Cell Lung Cancer. *Frontiers in Oncology* vol. 11 (2021).

28. Gainor, J. F. *et al.* Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK -rearranged lung cancer. *Cancer Discov.* **6**, 1118–1133 (2016).

29. Xia, P. *et al.* Molecular characteristics and clinical outcomes of complex ALK rearrangements identified by next-generation sequencing in non-small cell lung cancers. *J. Transl. Med.* **19**, 308 (2021).

30. Yung, R. C. W. Tissue diagnosis of suspected lung cancer: Selecting between bronchoscopy, transthoracic needle aspiration, and resectional biopsy. *Respir. Care Clin. N. Am.* **9**, 51–76 (2003).

31. Wiener, R. S. et al. Risks of Transthoracic Needle Biopsy: How High? *Clin. Pulm. Med.* **20**, 29 (2013).

32. Chen, K. *et al.* Spatiotemporal genomic analysis reveals distinct molecular features in recurrent stage I non-small cell lung cancers. *Cell Rep.* **40**, (2022).

33. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

34. Alix-Panabières, C. et al. Liquid biopsy: from concept to clinical application. *Scientific Reports* vol. 13 (2023).

35. Bertoli, E. *et al.* Liquid Biopsy in NSCLC: An Investigation with Multiple Clinical Implications. *International Journal of Molecular Sciences* vol. 24 (2023).

36. Rijavec, E. *et al.* Liquid biopsy in non-small cell lung cancer: Highlights and challenges. *Cancers* vol. 12 (2020).

37. Di Capua, D. et al. The liquid biopsy for lung cancer: State of the art, limitations and future developments. *Cancers* vol. 13 (2021).

38. Lyu, M. et al. The diagnostic value of circulating tumor cells and ctDNA for gene mutations in lung cancer. *Onco. Targets. Ther.* **12**, 2539–2552 (2019).

39. Ge, Q. et al. Liquid biopsy: Comprehensive overview of circulating tumor DNA (Review). *Oncology Letters* vol. 28 (2024).

40. Chen, C. et al. Clinical Validity and Utility of Circulating Tumor DNA (ctDNA) Testing in Advanced Non-small Cell Lung Cancer (aNSCLC): A Systematic Literature Review

and Meta-analysis. *Mol. Diagn. Ther.* **28**, 525 (2024).

41. Angeles, A. K. *et al.* Early identification of disease progression in ALK-rearranged lung cancer using circulating tumor DNA analysis. *NPJ Precis. Oncol.* **5**, (2021).

42. Tai, Y. L. *et al.* Basics and applications of tumor-derived extracellular vesicles. *J. Biomed. Sci.* **26**, 35 (2019).

43. Zou, Y. et al. Clinical applications of extracellular vesicles: recent advances and emerging trends. *Front. Bioeng. Biotechnol.* **13**, 1671963 (2025).

44. Lin, D. *et al.* Circulating tumor cells: biology and clinical significance. *Signal Transduction and Targeted Therapy* vol. 6 (2021).

45. Rupp, B. et al. Circulating tumor cells in precision medicine: challenges and opportunities. *Trends in Pharmacological Sciences* vol. 43 378–391 (2022).

46. Woo, D. & Yu, M. Circulating tumor cells as “liquid biopsies” to understand cancer metastasis. *Translational Research* vol. 201 128–135 (2018).

47. Jie, X.-X. et al. *Epithelial-to-Mesenchymal Transition, Circulating Tumor Cells and Cancer Metastasis: Mechanisms and Clinical Applications*. *Oncotarget* vol. 8 (2017).

48. Bornes, L. et al. Epithelial-to-Mesenchymal Transition in the Light of Plasticity and Hybrid E/M States. *J. Clin. Med.* **10**, 2403 (2021).

49. Wang, Y. *et al.* Vimentin expression in circulating tumor cells (CTCs) associated with liver metastases predicts poor progression-free survival in patients with advanced lung cancer. *J. Cancer Res. Clin. Oncol.* **145**, 2911 (2019).

50. Wang, W. C. *et al.* Survival Mechanisms and Influence Factors of Circulating Tumor Cells. *Biomed Res. Int.* **2018**, 6304701 (2018).

51. Zhang, L. et al. Platelet-circulating tumor cell crosstalk: A pivotal target in cancer

diagnosis and therapy (Review). *Oncology Reports* vol. 55 (2026).

52. Gautam, D. et al. Platelets and circulating (tumor) cells: partners in promoting metastatic cancer. *Curr. Opin. Hematol.* **32**, (2025).

53. Zhou, L. *et al.* The critical role of platelet in cancer progression and metastasis. *Eur. J. Med. Res.* **28**, 385 (2023).

54. Placke, T. *et al.* Platelet-derived MHC class I confers a pseudonormal phenotype to cancer cells that subverts the antitumor reactivity of natural killer immune cells. *Cancer Res.* **72**, 440–448 (2012).

55. Cecerska-Heryć, E. *et al.* TGF-β Signaling in Cancer: Mechanisms of Progression and Therapeutic Targets. *Int. J. Mol. Sci.* **26**, 7326 (2025).

56. Shaw, P. et al. VEGF signaling: Role in angiogenesis and beyond. *Biochim. Biophys. Acta Rev. Cancer* **1879**, 189079 (2024).

57. Zhao, J. et al. Decoding the role of platelets in tumour metastasis: enigmatic accomplices and intricate targets for anticancer treatments. *Front. Immunol.* **14**, 1256129 (2023).

58. Giuliano, M. *et al.* Perspective on circulating tumor cell clusters: Why it takes a village to metastasize. *Cancer Research* vol. 78 845–852 (2018).

59. Schuster, E. *et al.* Better together: circulating tumor cell clustering in metastatic cancer. *Trends Cancer* **7**, 1020 (2021).

60. Chen, S. et al. Microenvironment crosstalk and immune evasion of circulating tumor cells: From mechanism to clinical significance. *Chin. Med. J. (Engl).* **139**, 515 (2025).

61. Aceto, N. *et al.* Circulating Tumor Cell Clusters are Oligoclonal Precursors of Breast Cancer Metastasis. *Cell* **158**, 1110 (2014).

62. Lim, S. et al. Roles of cytoskeleton in metastasis: from its mechanism to therapeutic

strategies. *Exp. Mol. Med.* **58**, 1 (2026).

63. Polioudaki, H. *et al.* Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. *BMC Cancer* **15**, 399 (2015).

64. Moose, D. L. & Henry, M. D. Survival of the resilient: Mechano-adaptation of circulating tumor cells to fluid shear stress. *Mol. Cell. Oncol.* **7**, 1766338 (2020).

65. Masmoudi, D. et al. Natural killer cells: the immune frontline against circulating tumor cells. *J. Exp. Clin. Cancer Res.* **44**, 118 (2025).

66. Fu, Y. *et al.* Immunosenescence: signaling pathways, diseases and therapeutic targets. *Signal Transduction and Targeted Therapy 2025 10:1* **10**, 250- (2025).

67. Di Spirito, A. et al. Key immune cells and their crosstalk in the tumor microenvironment of bladder cancer: insights for innovative therapies. *Open Exploration 2019 6:* **6**, 1002304- (2025).

68. Gao, F. *et al.* Exploring neutrophil extracellular traps: mechanisms of immune regulation and future therapeutic potential. *Exp. Hematol. Oncol.* **14**, 80 (2025).

69. Leone, K. et al. The Interplay between Circulating Tumor Cells and the Immune System: From Immune Escape to Cancer Immunotherapy. *Diagnostics* **8**, 59 (2018).

70. Vivier, E. *et al.* Innate or adaptive immunity? The example of natural killer cells. *Science* **331**, 44–49 (2011).

71. Ljunggren, H. G. & Kärre, K. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol. Today* **11**, 237–244 (1990).

72. Yu, P. *et al.* NK cell depletion promotes liver metastasis of lung cancer cells. *Acta Biochim. Biophys. Sin. (Shanghai).* **56**, 323 (2024).

73. López-Soto, A. et al. Control of Metastasis by NK Cells. *Cancer Cell* **32**, 135–154 (2017).

74. De Andrade, L. F. *et al.* Antibody-mediated inhibition of MICA and MICB shedding promotes NK cell-driven tumor immunity. *Science* **359**, 1537–1542 (2018).

75. Groh, V. et al. Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature* **419**, 734–738 (2002).

76. Hò, G. G. T. *et al.* NKG2A/CD94 Is a New Immune Receptor for HLA-G and Distinguishes Amino Acid Differences in the HLA-G Heavy Chain. *Int. J. Mol. Sci.* **21**, 1–17 (2020).

77. Liu, X. *et al.* Immune checkpoint HLA-E:CD94-NKG2A mediates evasion of circulating tumor cells from NK cell surveillance. *Cancer Cell* **41**, 272-287.e9 (2023).

78. Kopp, H. G. et al. Platelet-derived transforming growth factor-beta down-regulates NKG2D thereby inhibiting natural killer cell antitumor reactivity. *Cancer Res.* **69**, 7775–7783 (2009).

79. 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).

80. Ito, M. *et al.* Killer cell lectin-like receptor G1 binds three members of the classical cadherin family to inhibit NK cell cytotoxicity. *J. Exp. Med.* **203**, 289–295 (2006).

81. Fridlender, Z. G. *et al.* Polarization of Tumor-Associated Neutrophil Phenotype by TGF-β: 'N1' versus 'N2' TAN. *Cancer Cell* **16**, 183–194 (2009).

82. Shaul, M. E. *et al.* Tumor-associated neutrophils display a distinct N1 profile following TGFβ modulation: A transcriptomics analysis of pro- vs. antitumor TANs. *Oncoimmunology* **5**, (2016).

83. Yang, L. *et al.* DNA of neutrophil extracellular traps promotes cancer metastasis via CCDC25. *Nature* **583**, 133–138 (2020).

84. Matlung, H. L. *et al.* Neutrophils Kill Antibody-Opsonized Cancer Cells by Trogoptosis. *Cell Rep.* **23**, 3946-3959.e6 (2018).

85. Szczerba, B. M. *et al.* Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature 2019 566:7745* **566**, 553–557 (2019).

86. Gu, X. Bin. *et al*. Prognostic significance of neutrophil-to-lymphocyte ratio in non-small cell lung cancer: a meta-analysis. *Sci. Rep.* **5**, (2015).

87. Zhao, Q. T. *et al.* Prognostic role of neutrophil to lymphocyte ratio in lung cancers: a meta-analysis including 7,054 patients. *Onco. Targets. Ther.* **8**, 2731–2738 (2015).

88. Dhatchinamoorthy, K. et al. Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front. Immunol.* **12**, (2021).

89. Cornel, A. M. et al. MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. *Cancers (Basel).* **12**, 1760 (2020).

90. Hamilton, G. & Rath, B. Circulating tumor cell interactions with macrophages: Implications for biology and treatment. *Translational Lung Cancer Research* vol. 6 418–430 (2017).

91. De Giorgi, U. *et al.* Association between circulating tumor cells and peripheral blood monocytes in metastatic breast cancer. *Ther. Adv. Med. Oncol.* **11**, (2019).

92. Manjunath, Y. *et al.* Tumor-Cell–Macrophage Fusion Cells as Liquid Biomarkers and Tumor Enhancers in Cancer. *Int. J. Mol. Sci.* **21**, 1872 (2020).

93. Manjunath, Y. *et al.* Circulating Giant Tumor-Macrophage Fusion Cells Are Independent Prognosticators in Patients With NSCLC. *Journal of Thoracic Oncology* **15**, 1460–1471

(2020).

94. Zhang, F. *et al.* TGF-β induces M2-like macrophage polarization via SNAIL-mediated suppression of a pro-inflammatory phenotype. *Oncotarget* **7**, 52294–52306 (2016).

95. Radharani, N. N. V. *et al.* Tumor-associated macrophage derived IL-6 enriches cancer stem cell population and promotes breast tumor progression via Stat-3 pathway. *Cancer Cell Int.* **22**, (2022).

96. Janning, M. *et al.* Determination of PD-L1 Expression in Circulating Tumor Cells of NSCLC Patients and Correlation with Response to PD-1/PD-L1 Inhibitors. *Cancers (Basel).* **11**, 835 (2019).

97. Qiao, D. R. et al. PD-L1/PD-1 blockage enhanced the cytotoxicity of natural killer cell on the non-small cell lung cancer (NSCLC) by granzyme B secretion. *Clinical & Translational Oncology* **25**, 2373 (2023).

98. Garcia-Diaz, A. *et al.* Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. *Cell Rep.* **19**, 1189–1201 (2017).

99. Taube, J. M. *et al.* Colocalization of inflammatory response with B7-H1 expression in human melanocytic lesions supports an adaptive resistance mechanism of immune escape. *Sci. Transl. Med.* **4**, (2012).

100. Ribas, A. Adaptive immune resistance: How cancer protects from immune attack. *Cancer Discovery* vol. 5 915–919 (2015).

101. Wong, K. H. K. *et al.* Whole blood stabilization for the microfluidic isolation and molecular characterization of circulating tumor cells. *Nature Communications 2017 8:1* **8**, 1733- (2017).

102. Koch, C. *et al.* Pre-Analytical and Analytical Variables of Label-Independent Enrichment

and Automated Detection of Circulating Tumor Cells in Cancer Patients. *Cancers (Basel).* **12**, (2020).

103. Tretyakova, M. S. *et al.* Technologies for Viable Circulating Tumor Cell Isolation. *Int. J. Mol. Sci.* **23**, 15979 (2022).

104. Rushton, A. J. et al. A Review of Circulating Tumour Cell Enrichment Technologies. *Cancers (Basel).* **13**, 970 (2021).

105. Drucker, A. *et al.* Comparative performance of different methods for circulating tumor cell enrichment in metastatic breast cancer patients. *PLoS One* **15**, (2020).

106. Litvinov, S. V. et al. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J. Cell Biol.* **125**, 437–446 (1994).

107. Baizar, M. et al. The biology of the 17-1A antigen (Ep-CAM). *J. Mol. Med. (Berl).* **77**, 699–712 (1999).

108. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **10**, 6897–6904 (2004).

109. Riethdorf, S. *et al.* Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system. *Clin. Cancer Res.* **13**, 920–928 (2007).

110. Farace, F. *et al.* A direct comparison of CellSearch and ISET for circulating tumour-cell detection in patients with metastatic carcinomas. *Br. J. Cancer* **105**, 847 (2011).

111. Le Du, F. *et al.* EpCAM-independent isolation of circulating tumor cells with epithelial-to-mesenchymal transition and cancer stem cell phenotypes using ApoStream® in patients with breast cancer treated with primary systemic therapy. *PLoS One* **15**, e0229903 (2020).

112. Raimondi, C. et al. Circulating tumor cells isolation: the “post-EpCAM era”. *Chinese Journal of Cancer Research* **27**, 461 (2015).

113. Karabacak, N. M. *et al.* Microfluidic, marker-free isolation of circulating tumor cells from blood samples. *Nat. Protoc.* **9**, 694 (2014).

114. Ozkumur, E. *et al.* Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci. Transl. Med.* **5**, (2013).

115. Sajay, B. N. G. *et al.* Microfluidic platform for negative enrichment of circulating tumor cells. *Biomed. Microdevices* **16**, 537–548 (2014).

116. Lapin, M. *et al.* MINDEC-An Enhanced Negative Depletion Strategy for Circulating Tumour Cell Enrichment. *Sci. Rep.* **6**, 28929 (2016).

117. Li, X. *et al.* Strategies for enrichment of circulating tumor cells. *Transl. Cancer Res.* **9**, 2012–2025 (2020).

118. Descamps, L. et al. Microfluidic-Based Technologies for CTC Isolation: A Review of 10 Years of Intense Efforts towards Liquid Biopsy. *Int. J. Mol. Sci.* **23**, 1981 (2022).

119. Ma, Y. C. et al. Recent advances and prospects in the isolation by size of epithelial tumor cells (ISET) methodology. *Technol. Cancer Res. Treat.* **12**, 295–309 (2013).

120. Vona, G. *et al.* Isolation by size of epithelial tumor cells: A new method for the immunomorphological and molecular characterization of circulating tumor cells. *American Journal of Pathology* **156**, 57–63 (2000).

121. Laget, S. *et al.* Technical Insights into Highly Sensitive Isolation and Molecular Characterization of Fixed and Live Circulating Tumor Cells for Early Detection of Tumor Invasion. *PLoS One* **12**, e0169427 (2017).

122. Xu, L. *et al.* Optimization and Evaluation of a Novel Size Based Circulating Tumor Cell

Isolation System. *PLoS One* **10**, e0138032 (2015).

123. Wan, S. *et al.* New Labyrinth Microfluidic Device Detects Circulating Tumor Cells Expressing Cancer Stem Cell Marker and Circulating Tumor Microemboli in Hepatocellular Carcinoma. *Scientific Reports 2019 9:1* **9**, 18575- (2019).

124. Aghaamoo, M. et al. Deformability-based circulating tumor cell separation with conical-shaped microfilters: Concept, optimization, and design criteria. *Biomicrofluidics* **9**, 034106 (2015).

125. Macaraniag, C. et al. Microfluidic techniques for isolation, formation, and characterization of circulating tumor cells and clusters. *APL Bioengineering* vol. 6 (2022).

126. Edd, J. F. *et al.* Isolation of circulating tumor cells. *iScience* **25**, 104696 (2022).

127. Liu, Z. *et al.* Rapid isolation of cancer cells using microfluidic deterministic lateral displacement structure. *Biomicrofluidics* **7**, 011801 (2013).

128. Fachin, F. *et al.* Monolithic Chip for High-throughput Blood Cell Depletion to Sort Rare Circulating Tumor Cells. *Sci. Rep.* **7**, (2017).

129. Warkiani, M. E. *et al.* Slanted spiral microfluidics for the ultra-fast, label-free isolation of circulating tumor cells. *Lab Chip* **14**, 128–137 (2014).

130. Sollier, E. *et al.* Size-selective collection of circulating tumor cells using Vortex technology. *Lab Chip* **14**, 63–77 (2013).

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

132. Bagnall, J. S. *et al.* Deformability of Tumor Cells versus Blood Cells. *Scientific Reports 2015 5:1* **5**, 18542- (2015).

133. Chen, H. et al. Size- and deformability-based isolation of circulating tumor cells with

microfluidic chips and their applications in clinical studies. *AIP Adv.* **8**, (2018).

134. Gascoyne, P. R. C. & Shim, S. Isolation of Circulating Tumor Cells by Dielectrophoresis. *Cancers 2014, Vol. 6, Pages 545-579* **6**, 545–579 (2014).

135. Gupta, V. *et al.* ApoStream^{TM}, a new dielectrophoretic device for antibody independent isolation and recovery of viable cancer cells from blood. *Biomicrofluidics* **6**, (2012).

136. Dao, M. *et al.* Acoustic separation of circulating tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4970–4975 (2015).

137. Wu, M. *et al.* Acoustofluidic separation of cells and particles. *Microsystems & Nanoengineering 2019 5:1* **5**, 32- (2019).

138. Magnusson, C. *et al.* Clinical Scale Cell-Surface Marker Independent, Acoustic Microfluidic Enrichment of Tumor Cells from Blood. *Anal. Chem.* **89**, 11954 (2017).

139. Ferreira, M. M. et al. Circulating tumor cell technologies. *Mol. Oncol.* **10**, 374–394 (2016).

140. Strati, A. et al. Detection and Molecular Characterization of Circulating Tumour Cells: Challenges for the Clinical Setting. *Cancers (Basel).* **15**, 2185 (2023).

141. Krebs, M. G. *et al.* Molecular analysis of circulating tumour cells-biology and biomarkers. *Nat. Rev. Clin. Oncol.* **11**, 129–144 (2014).

142. 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).

143. Saiki, R. K. *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).

144. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of

quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).

145. Millner, L. M. et al. Circulating Tumor Cells: A Review of Present Methods and the Need to Identify Heterogeneous Phenotypes. *Ann. Clin. Lab. Sci.* **43**, 295 (2013).

146. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).

147. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).

148. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19096–19101 (2009).

149. 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).

150. Ahn, J. et al. Introduction to Single-Cell DNA Methylation Profiling Methods. *Biomolecules* **11**, (2021).

151. Buenrostro, J. D. et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

152. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).

153. Zhao, L. et al. DNA methylome profiling of circulating tumor cells in lung cancer at single base-pair resolution. *Oncogene* **40**, 1884–1895 (2021).

154. Mortazavi, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

155. Wang, X. et al. Direct Comparative Analyses of 10X Genomics Chromium and Smart-

seq2. *Genomics Proteomics Bioinformatics* **19**, 253–266 (2021).

156. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

157. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods 2013 10:11* **10**, 1096–1098 (2013).

158. Towbin, H. et al. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354 (1979).

159. Sinkala, E. *et al.* Profiling protein expression in circulating tumour cells using microfluidic western blotting. *Nature Communications 2017 8:1* **8**, 14622- (2017).

160. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871–874 (1971).

161. Müller, T. *et al.* Automated sample preparation with SP3 for low-input clinical proteomics. *Mol. Syst. Biol.* **16**, (2020).

162. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).

163. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).

164. Bauman, J. G. J. et al. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Exp. Cell Res.* **128**, 485–490 (1980).

165. Lieberman, B. *et al.* Toward uncharted territory of cellular heterogeneity: advances and applications of single-cell RNA-seq. *J. Transl. Genet. Genom.* **5**, 1 (2021).

166. Chow, A. & Lareau, C. A. Concepts and new developments in droplet-based single-cell multi-omics. *Trends Biotechnol.* **42**, 1379 (2024).

167. Chuang, H. C. *et al.* Single-cell sequencing of full-length transcripts and T-cell receptors with automated high-throughput Smart-seq3. *BMC Genomics* **25**, 1127 (2024).

168. Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 2540 (2021).

169. Pailler, E. *et al.* Detection of circulating tumor cells harboring a unique ALK rearrangement in ALK-positive non-small-cell lung cancer. *J. Clin. Oncol.* **31**, 2273–2281 (2013).

170. Shakoori, A. R. Fluorescence In Situ Hybridization (FISH) and Its Applications. *Chromosome Structure and Aberrations* 343 (2017).

171. Peng, L. *et al.* Targeting ALK Rearrangements in NSCLC: Current State of the Art. *Front. Oncol.* **12**, 863461 (2022).

172. Faugeroux, V. et al. Clinical Utility of Circulating Tumor Cells in ALK-Positive Non-Small-Cell Lung Cancer. *Front. Oncol.* **4**, 281 (2014).

173. Parkin, B. Rare Variant Quantitation Using Droplet Digital PCR. *Methods Mol. Biol.* **1881**, 239–251 (2019).

174. Yang, K. *et al.* Developing Ultrasensitive Library-Aliquot-Based Droplet Digital PCR for Detecting T790M in Plasma-Circulating Tumor DNA of Non-small-Cell-Lung-Cancer Patients. *Anal. Chem.* **90**, 11203–11209 (2018).

175. Bray, M. A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).

176. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies.

Nat. Protoc. **16**, 3802–3835 (2021).

177. Allam, M. et al. Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics. *NPJ Precis. Oncol.* **4**, 11 (2020).

178. Chen, Y. *et al.* Abstract 7504: Longitudinal multi-omics characterization of circulating tumor cells from ALK positive NSCLC patients. *Cancer Res.* **84**, 7504–7504 (2024).

179. 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 2020, Vol. 12, Page 127* **12**, 127 (2020).

180. Kuswanto, W. et al. Highly multiplexed spatial profiling with CODEX: bioinformatic analysis and application in human disease. *Semin. Immunopathol.* **45**, 145 (2022).

181. Robinson, J. P. et al. Flow Cytometry: Advances, Challenges and Trends. *Bioessays* **48**, e70091 (2025).

182. Nesari, A. M. et al. Advances and challenges in single-cell RNA sequencing data analysis: a comprehensive review. *Brief. Bioinform.* **27**, bbaf723 (2026).

183. Verma, P. K. *et al.* Single-cell RNA sequencing methodology, analysis, applications, and future directions with special focus on cotton. *Plant Biol.* **28**, 31–44 (2026).

184. Chatterjee, D. & Deng, W. M. Standardization of Single-Cell RNA-Sequencing Analysis Workflow to Study Drosophila Ovary. *Methods in Molecular Biology* **2677**, 151–171 (2023).

185. Abueed, O. et al. A Systematic Review of U-Net Optimizations: Advancing Tumour Segmentation in Medical Imaging. *IET Image Process.* **19**, (2025).

186. Li, H. *et al.* Advances in machine learning-enhanced microfluidic cell sorting. *Sci. Adv.* **11**, eaea6007 (2025).

187. Nguyen-Tat, T. B. et al. Evaluating pre-processing and deep learning methods in medical imaging: Combined effectiveness across multiple modalities. *Alexandria Engineering Journal* **119**, 558–586 (2025).

188. Ibrahim, M. *et al.* Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. *Comput. Biol. Med.* **189**, (2025).

189. Sharma, P. et al. Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications. *Multimed. Tools Appl.* **83**, 88811–88858 (2024).

190. Shen, X. et al. Evolutionary computation-based self-supervised learning for image processing: a big data-driven approach to feature extraction and fusion for multispectral object detection. *J. Big Data* **11**, (2024).

191. Hu, Y. *et al.* Predicting Molecular Traits from Tissue Morphology Through Self-interactive Multi-instance Learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **13432 LNCS**, 130–139 (2022).

192. Forcato, M. et al. Computational methods for the integrative analysis of single-cell data. *Brief. Bioinform.* **22**, 20–29 (2021).

193. Acera-Mateos, M. *et al.* Systematic evaluation of single-cell multimodal data integration enhances cell type resolution and discovery of clinically relevant states in complex tissues. *Genome Biol.* **27**, (2026).

194. Shi, Y. *et al.* Dynamic Synthesis of Multi-Modal Representations for CITE-seq Data Integration and Analysis. *Advanced Science* **12**, (2025).

195. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29

(2021).

196. Afsaneh, E. et al. Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: a comprehensive review. *Diabetol. Metab. Syndr.* **14**, (2022).

197. Waqas, M. & Humphries, U. W. A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX* **13**, (2024).

198. Alsubaie, N. E. *et al.* Comparative evaluation of Cox regression and machine learning approaches for survival prediction in stage IV lung cancer. *Discover Computing* **29**, (2026).

199. Zhuang, L. *et al.* Advancing Precision Oncology Through Modeling of Longitudinal and Multimodal Data. *IEEE Rev. Biomed. Eng.* **19**, 182 (2026).

200. Lötsch, J. et al. Resolving Interpretation Challenges in Machine Learning Feature Selection With an Iterative Approach in Biomedical Pain Data. *Eur. J. Pain* **30**, e70221 (2026).

201. Yuan, C. H. *et al.* A dynamic weighted ensemble learning framework for cardiovascular risk prediction in type 2 diabetes: a comparative study with SHAP-based interpretability. *Scientific Reports 2025 15:1* **15**, 45029- (2025).

202. Alix-Panabières, C. & Pantel, K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov.* **6**, 479–491 (2016).

203. Harouaka, R. et al. Circulating tumor cells: advances in isolation and analysis, and challenges for clinical applications. *Pharmacol. Ther.* **141**, 209 (2013).

204. 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).

205. Capuozzo, M. et al. Circulating Tumor Cells as Predictive and Prognostic Biomarkers in Solid Tumors. *Cells* vol. 12 (2023).

206. Yousefi, M. *et al.* Prognostic and therapeutic significance of circulating tumor cells in patients with lung cancer. *Cellular Oncology* **43**, 31 (2019).

207. Punnoose, E. A. *et al.* Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. *Clin. Cancer Res.* **18**, 2391–2401 (2012).

208. Lecharpentier, A. *et al.* Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. *Br. J. Cancer* **105**, 1338–1341 (2011).

209. Shishido, S. N. *et al.* Circulating tumor cells as a response monitor in stage IV non-small cell lung cancer. *J. Transl. Med.* **17**, 294 (2019).

210. Duréndez-Sáez, E. *et al.* New insights in non-small-cell lung cancer: circulating tumor cells and cell-free DNA. *J. Thorac. Dis.* **9**, S1332 (2017).

211. Pailler, E. *et al.* Acquired Resistance Mutations to ALK Inhibitors Identified by Single Circulating Tumor Cell Sequencing in ALK-Rearranged Non-Small-Cell Lung Cancer. *Clin. Cancer Res.* **25**, 6671–6682 (2019).

212. Krebs, M. G. *et al.* Analysis of circulating tumor cells in patients with non-small cell lung cancer using epithelial marker-dependent and -independent approaches. *Journal of Thoracic Oncology* **7**, 306–315 (2012).

213. Guibert, N. *et al.* PD-L1 expression in circulating tumor cells of advanced non-small cell lung cancer patients treated with nivolumab. *Lung Cancer* **120**, 108–112 (2018).

214. Kapeleris, J. *et al.* Ex vivo culture of circulating tumour cells derived from non-small cell

lung cancer. *Transl. Lung Cancer Res.* **9**, 1795 (2020).

215. Chen, Y. *et al.* Abstract 5592: Expansion and characterization on ALK positive NSCLC circulating tumor cells isolated using a size based inertial microfluidic Labyrinth device. *Cancer Res.* **83**, 5592–5592 (2023).

216. Kahounová, Z. *et al.* Circulating tumor cell-derived preclinical models: current status and future perspectives. *Cell Death & Disease 2023 14:8* **14**, 530- (2023).

217. Huo, K. G. et al. Patient-derived cell line, xenograft and organoid models in lung cancer therapy. *Transl. Lung Cancer Res.* **9**, 2214 (2020).

218. Wu, Y. H. *et al.* Correlation between drug sensitivity profiles of circulating tumour cell-derived organoids and clinical treatment response in patients with pancreatic ductal adenocarcinoma. *Eur. J. Cancer* **166**, 208–218 (2022).

219. Labelle, M. et al. Direct Signaling between Platelets and Cancer Cells Induces an Epithelial-Mesenchymal-Like Transition and Promotes Metastasis. *Cancer Cell* **20**, 576–590 (2011).

220. Cools-Lartigue, J. *et al.* Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis. *J. Clin. Invest.* **123**, 3446–3458 (2013).

221. Kulasinghe, A. *et al.* Phenotypic characterization of circulating lung cancer cells for clinically actionable targets. *Cancers (Basel).* **11**, (2019).

222. Wang, J. *et al.* A circulating tumor cell-based digital assay for the detection of EGFR T790M mutation in advanced non-small cell lung cancer. *J. Mater. Chem. B* **8**, 5636–5644 (2020).

223. Williamson, D. F. K. *et al.* Detection of EGFR mutations in non-small cell lung cancer by droplet digital PCR. *PLoS One* **17**, (2022).

224. Smerage, J. B. *et al.* Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J. Clin. Oncol.* **32**, 3483–3489 (2014).

225. Cabel, L. *et al.* Clinical utility of circulating tumour cell-based monitoring of late-line chemotherapy for metastatic breast cancer: the randomised CirCe01 trial. *Br. J. Cancer* **124**, 1207–1213 (2021).

226. Pailler, E. et al. Routine clinical use of circulating tumor cells for diagnosis of mutations and chromosomal rearrangements in non-small cell lung cancer-ready for prime-time? *Transl. Lung Cancer Res.* **6**, 444–453 (2017).

227. Thandra, K. C. et al. Epidemiology of lung cancer. *Wspolczesna Onkologia* vol. 25 45–52 (2021).

228. Didier, A. J. et al. Demographic Disparities in Lung Cancer Mortality and Trends in the United States From 1999 Through 2020: A Population-Based CDC Database Analysis. *JNCCN Journal of the National Comprehensive Cancer Network* **22**, (2024).

229. Gaughan, E. M. & Costa, D. B. Genotype-driven therapies for non-small cell lung cancer: Focus on EGFR, KRAS and ALK gene abnormalities. *Therapeutic Advances in Medical Oncology* vol. 3 113–125 (2011).

230. Salcher, S. *et al.* High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *Cancer Cell* **40**, 1503-1520.e8 (2022).

231. Planchard, D. & Besse, B. Lung cancer in never-smokers. *European Respiratory Journal* vol. 45 1214–1217 (2015).

232. Lin, J. J. et al. Targeting ALK: Precision medicine takes on drug resistance. *Cancer Discovery* vol. 7 137–155 (2017).

233. Wang, L. & Lui, V. W. Y. Emerging roles of ALK in immunity and insights for

immunotherapy. *Cancers* vol. 12 (2020).

234. Sankar, K. et al. Immunotherapy for ALK-Rearranged Non-Small Cell Lung Cancer: Challenges Inform Promising Approaches. *Cancers 2021, Vol. 13, Page 1476* **13**, 1476 (2021).

235. Shreenivas, A. *et al.* ALK fusions in the pan-cancer setting: another tumor-agnostic target? *NPJ Precis. Oncol.* **7**, 101 (2023).

236. Jin, R. *et al.* Molecular heterogeneity of anti-PD-1/PD-L1 immunotherapy efficacy is correlated with tumor immune microenvironment in East Asian patients with non-small cell lung cancer. *Cancer Biol. Med.* **17**, 768–781 (2020).

237. McMahon, D. J. et al. Is Immunotherapy Beneficial in Patients with Oncogene-Addicted Non-Small Cell Lung Cancers? A Narrative Review. *Cancers* vol. 16 (2024).

238. Cao, P. *et al.* Case Report: ALK rearranged locally advanced lung adenocarcinoma showing inconsistent radiographic findings and pathological responses during neoadjuvant alectinib therapy. *Front. Pharmacol.* **14**, (2023).

239. Gainor, J. F. *et al.* EGFR Mutations and ALK Rearrangements Are Associated with Low Response Rates to PD-1 Pathway Blockade in Non-Small Cell Lung Cancer (NSCLC): A Retrospective Analysis. *Clin. Cancer Res.* **22**, 4585 (2016).

240. Pătcaș, A. *et al.* An insight into lung cancer: A comprehensive review exploring anaplastic lymphoma kinase tyrosine kinase inhibitors and mechanisms of resistance. *Bosnian Journal of Basic Medical Sciences* vol. 22 1–13 (2022).

241. Kim, S. *et al.* Heterogeneity of Genetic Changes Associated with Acquired Crizotinib Resistance in ALK-Rearranged Lung Cancer. *Journal of Thoracic Oncology* **8**, 415–422 (2013).

242. Holla, V. R. *et al.* ALK: a tyrosine kinase target for cancer therapy. *Molecular Case Studies* **3**, a001115 (2017).

243. Hrustanovic, G. *et al.* RAS-MAPK dependence underlies a rational polytherapy strategy in EML4-ALK-positive lung cancer. *Nat. Med.* **21**, 1038–1047 (2015).

244. Li, J. *et al.* L1198F mutation resensitizes crizotinib to ALK by altering the conformation of inhibitor and ATP binding sites. *Int. J. Mol. Sci.* **18**, (2017).

245. Li, W. *et al.* Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring. *Mol. Cancer* **21**, 1–15 (2022).

246. Manicone, M. et al. Critical issues in the clinical application of liquid biopsy in non-small cell lung cancer. *Journal of Thoracic Disease* vol. 9 S1346–S1358 (2017).

247. Papadaki, M. A. & Agelaki, S. Single-Cell RNA Sequencing Uncovers Heterogeneous Circulating Tumor Cell Subsets in Breast Cancer. *Cancers* vol. 14 (2022).

248. Pereira-Veiga, T. et al. Circulating tumor cell-blood cell crosstalk: Biology and clinical relevance. *Cell Rep.* **40**, (2022).

249. Loutherback, K. *et al.* Deterministic separation of cancer cells from blood at 10 mL/min. *AIP Adv.* **2**, 042107 (2012).

250. Kyung-A Hyun *et al.* Epithelial-to-mesenchymal transition leads to loss of EpCAM and different physical properties in circulating tumor cells from metastatic breast cancer. *Oncotarget* **7**, 24677 (2016).

251. Andree, K. C. et al. Challenges in circulating tumor cell detection by the CellSearch system. *Molecular Oncology* vol. 10 395–407 (2016).

252. Rieckmann, L. M. *et al.* Diagnostic leukapheresis reveals distinct phenotypes of NSCLC circulating tumor cells. *Molecular Cancer* vol. 23 (2024).

253. Han, X. *et al.* Comprehensive Profiling of Gene Copy Number Alterations Predicts Patient Prognosis in Resected Stages I–III Lung Adenocarcinoma. *Front. Oncol.* **9**, (2019).

254. Zhang, T. *et al.* Genomic and evolutionary classification of lung cancer in never smokers. *Nat. Genet.* **53**, 1348–1359 (2021).

255. Schupp, J. *et al.* Targeting myeloid cells in the tumor sustaining microenvironment. *Cell. Immunol.* **343**, (2019).

256. Mitchell, K. G. *et al.* Neutrophil expansion defines an immunoinhibitory peripheral and intratumoral inflammatory milieu in resected non-small cell lung cancer: a descriptive analysis of a prospectively immunoprofiled cohort. *J. Immunother. Cancer* **8**, e000405 (2020).

257. Sultana, A. *et al.* Single-cell RNA-seq analysis to identify potential biomarkers for diagnosis, and prognosis of non-small cell lung cancer by using comprehensive bioinformatics approaches. *Transl. Oncol.* **27**, 101571 (2023).

258. Abdo, R. *et al.* Expression of pannexin1 in lung cancer brain metastasis and immune microenvironment. *Clin. Neuropathol.* **42**, 140–149 (2023).

259. Meneses, K. M. *et al.* RAB27B Drives a Cancer Stem Cell Phenotype in NSCLC Cells Through Enhanced Extracellular Vesicle Secretion. *Cancer Research Communications* **3**, 607 (2023).

260. Huang, Z. et al. SMOX expression predicts the prognosis of non-small cell lung cancer. *Ann. Transl. Med.* **9**, 1048 (2021).

261. Tan, J. et al. The multiple roles and therapeutic potential of clusterin in non-small-cell lung cancer: a narrative review. *Transl. Lung Cancer Res.* **10**, 2683 (2021).

262. Li, H. *et al.* High expression of vinculin predicts poor prognosis and distant metastasis and

associates with influencing tumor-associated NK cell infiltration and epithelial-mesenchymal transition in gastric cancer. *Aging* **13**, 5197–5225 (2021).

263. Negishi, R. *et al.* Transcriptomic profiling of single circulating tumor cells provides insight into human metastatic gastric cancer. *Commun. Biol.* **5**, (2022).

264. Chen, C. L. *et al.* Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer. *Prostate* **73**, 813–826 (2013).

265. De Angelis, M. L. *et al.* An organoid model of colorectal circulating tumor cells with stem cell features, hybrid EMT state and distinctive therapy response profile. *Journal of Experimental and Clinical Cancer Research* **41**, (2022).

266. Moghbeli, M. PI3K/AKT pathway as a pivotal regulator of epithelial-mesenchymal transition in lung tumor cells. *Cancer Cell International* vol. 24 (2024).

267. Yu, J. *et al.* The EPHB6 receptor tyrosine kinase is a metastasis suppressor that is frequently silenced by promoter DNA hypermethylation in non-small cell lung cancer. *Clinical Cancer Research* **16**, 2275–2283 (2010).

268. Yang, L. *et al.* Glycosylated Cathepsin V Serves as a Prognostic Marker in Lung Cancer. *Front. Oncol.* **12**, (2022).

269. Zhang, X. et al. Modulation of esophageal squamous cell carcinoma progression: the impact of CCR7 on JAK2/STAT3 signaling pathway. *Discover Oncology* **15**, (2024).

270. Croker, B. A. et al. SOCS regulation of the JAK/STAT signalling pathway. *Seminars in Cell and Developmental Biology* vol. 19 414–422 (2008).

271. Zhang, M. *et al.* Overexpression of JAM-A in non-small cell lung cancer correlates with tumor progression. *PLoS One* **8**, (2013).

272. Taftaf, R. *et al.* ICAM1 initiates CTC cluster formation and trans-endothelial migration in lung metastasis of breast cancer. *Nat. Commun.* **12**, (2021).

273. Qiu, Z. *et al.* Roles of intercellular cell adhesion molecule-1 (ICAM-1) in colorectal cancer: expression, functions, prognosis, tumorigenesis, polymorphisms and therapeutic implications. *Frontiers in Oncology* vol. 12 (2022).

274. Radu, P. *et al.* CD34—Structure, Functions and Relationship with Cancer Stem Cells. *Medicina (Lithuania)* vol. 59 (2023).

275. Zeng, Z. et al. CCL5/CCR5 axis in human diseases and related treatments. *Genes Dis.* **9**, 12–27 (2022).

276. Shaw, A. T. *et al.* Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. *Journal of Clinical Oncology* **27**, 4247–4253 (2009).

277. Zheng, Z. R. *et al.* Association of smoking and ALK tyrosine-kinase inhibitors on overall survival in treatment-naïve ALK-positive advanced lung adenocarcinoma. *Front. Oncol.* **13**, (2023).

278. Voena, C. *et al.* Oncogenic ALK regulates EMT in non-small cell lung carcinoma through repression of the epithelial splicing regulatory protein 1. *Oncotarget* **7**, 33316–33330 (2016).

279. Kim, H. R. *et al.* Epithelial-mesenchymal transition leads to crizotinib resistance in H2228 lung cancer cells with EML4-ALK translocation. *Mol. Oncol.* **7**, 1093–1102 (2013).

280. Sun, Y. *et al.* Platelet-mediated circulating tumor cell evasion from natural killer cell killing through immune checkpoint CD155-TIGIT. *Hepatology* **81**, 791–807 (2025).

281. Schlesinger, M. Role of platelets and platelet receptors in cancer metastasis 06 Biological Sciences 0601 Biochemistry and Cell Biology. *Journal of Hematology and Oncology* vol.

11 (2018).

282. Ramundo, V. et al. TGF-β as Predictive Marker and Pharmacological Target in Lung Cancer Approach. *Cancers (Basel).* **15**, 2295 (2023).

283. Liu, S. et al. Targeting TGFβ signal transduction for cancer therapy. *Signal Transduction and Targeted Therapy 2020 6:1* **6**, 1–20 (2021).

284. Melese, E. S. *et al.* CCL5 production in lung cancer cells leads to an altered immune microenvironment and promotes tumor development. *Oncoimmunology* **11**, (2021).

285. Mittelheisser, V. *et al.* Platelets: tailoring metastasis treatment. *Trends in Cancer* (2025).

286. Wojtukiewicz, M. Z. et al. Antiplatelet agents for cancer treatment: a real perspective or just an echo from the past? *Cancer and Metastasis Reviews* **36**, 305–329 (2017).

287. Tullemans, B. M. E. et al. Acquired platelet antagonism: off-target antiplatelet effects of malignancy treatment with tyrosine kinase inhibitors. *Journal of Thrombosis and Haemostasis* vol. 16 1686–1699 (2018).

288. Horn, L. *et al.* Monitoring Therapeutic Response and Resistance: Analysis of Circulating Tumor DNA in Patients With ALK+ Lung Cancer. *Journal of Thoracic Oncology* **14**, 1901–1911 (2019).

289. Rolfo, C. *et al.* Liquid Biopsy for Advanced NSCLC: A Consensus Statement From the International Association for the Study of Lung Cancer. *Journal of Thoracic Oncology* **16**, 1647–1662 (2021).

290. Kwon, M. *et al.* Longitudinal monitoring by next generation sequencing of plasma cell-free DNA in ALK-rearranged non-small cell lung cancer (NSCLC) patients treated with ALK tyrosine kinase inhibitors. *Journal of Clinical Oncology* **38**, 9603–9603 (2020).

291. Ghiso, E. *et al.* YAP-Dependent AXL Overexpression Mediates Resistance to EGFR

Inhibitors in NSCLC. *Neoplasia* **19**, 1012–1021 (2017).

292. Byers, L. A. *et al.* An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin. Cancer Res.* **19**, 279–290 (2013).

293. Weng, C. H. *et al.* Epithelial-mesenchymal transition (EMT) beyond EGFR mutations per se is a common mechanism for acquired resistance to EGFR TKI. *Oncogene 2018 38:4* **38**, 455–468 (2018).

294. Pfeifer, M. *et al.* Genome-wide CRISPR screens identify the YAP/TEAD axis as a driver of persister cells in EGFR mutant lung cancer. *Communications Biology 2024 7:1* **7**, 497- (2024).

295. Arasada, R. R. *et al.* Notch3-dependent β-catenin signaling mediates EGFR TKI drug persistence in EGFR mutant NSCLC. *Nature Communications 2018 9:1* **9**, 3198- (2018).

296. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat. Genet.* **44**, 852–860 (2012).

297. Kwok, H. H. *et al.* Single-cell transcriptomic analysis uncovers intratumoral heterogeneity and drug-tolerant persister in ALK-rearranged lung adenocarcinoma. *Cancer Commun.* **43**, 951 (2023).

298. Dong, Z. *et al.* Genetic and phenotypic profiling of single living circulating tumor cells from patients with microfluidics. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2315168121 (2024).

299. Oulhen, M. *et al.* Circulating tumor cell copy-number heterogeneity in ALK-rearranged non-small-cell lung cancer resistant to ALK inhibitors. *npj Precision Oncology 2021 5:1* **5**, 67- (2021).

300. Pailler, E. *et al.* Circulating Tumor Cells with Aberrant ALK Copy Number Predict

Progression-Free Survival during Crizotinib Treatment in ALK-Rearranged Non-Small Cell Lung Cancer Patients. *Cancer Res.* **77**, 2222–2230 (2017).

301. Strati, A. et al. Rna-based ctc analysis provides prognostic information in metastatic breast cancer. *Diagnostics* **11**, (2021).

302. Purcell, E. *et al.* Circulating tumor cells reveal early predictors of disease progression in patients with stage III NSCLC undergoing chemoradiation and immunotherapy. *Cell Rep.* **43**, (2024).

303. Wu, C. Y. *et al.* Circulating Tumor Cells as a Tool of Minimal Residual Disease Can Predict Lung Cancer Recurrence: A longitudinal, Prospective Trial. *Diagnostics (Basel)* **10**, (2020).

304. Wankhede, D. et al. Circulating Tumor Cells as a Predictive Biomarker in Resectable Lung Cancer: A Systematic Review and Meta-Analysis. *Cancers (Basel).* **14**, 6112 (2022).

305. Guo, S. *et al.* The Value of Circulating Tumor Cells and Tumor Markers Detection in Lung Cancer Diagnosis. *Technol. Cancer Res. Treat.* **22**, 15330338231166754 (2023).

306. Muinelo-Romay, L. *et al.* Evaluation of Circulating Tumor Cells and Related Events as Prognostic Factors and Surrogate Biomarkers in Advanced NSCLC Patients Receiving First-Line Systemic Treatment. *Cancers (Basel).* **6**, 153 (2014).

307. Zhu, K. *et al.* CAVIN2/SDPR Functioned as a Tumor Suppressor in Lung Adenocarcinoma from Systematic Analysis of Caveolae-Related Genes and Experimental Validation. *J. Cancer* **14**, 2001 (2023).

308. Kumar, M. S. *et al.* The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell* **149**, 642–655 (2012).

309. Tessema, M. *et al.* GATA2 is epigenetically repressed in human and mouse lung tumors and is not requisite for survival of KRAS mutant lung cancer. *J. Thorac. Oncol.* **9**, 784–793 (2014).

310. Santoro, A. *et al.* Phase Ib/II study of ceritinib in combination with ribociclib in patients with ALK-rearranged non–small cell lung cancer. *Lung Cancer* **166**, 170–177 (2022).

311. Li, J. Significance of Circulating Tumor Cells in Nonsmall-Cell Lung Cancer Patients: Prognosis, Chemotherapy Efficacy, and Survival. *J. Healthc. Eng.* **2021**, 2680526 (2021).

312. Li, Z. *et al.* Circulating tumor cells can predict the prognosis of patients with non-small cell lung cancer after resection: a retrospective study. *Transl. Lung Cancer Res.* **10**, 995 (2021).

313. Quinn, L. *et al.* Interobserver variability studies in diagnostic imaging: a methodological systematic review. *Br. J. Radiol.* **96**, 20220972 (2023).

314. Okabe, T. *et al.* Mesenchymal Characteristics and Predictive Biomarkers on Circulating Tumor Cells for Therapeutic Strategy. *Cancers (Basel).* **12**, 3588 (2020).

315. Li, M. et al. Medical image analysis using deep learning algorithms. *Front. Public Health* **11**, 1273253 (2023).

316. Arlot, S. & Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv. 4, 40–79 (2010).*

317. Kaufman, S. et al. Leakage in data mining. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **6**, (2012).

318. Varoquaux, G. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage* **180**, 68–77 (2018).

319. Szeghalmy, S. & Fazekas, A. A Comparative Study of the Use of Stratified Cross-

Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning. *Sensors (Basel)* **23**, 2333 (2023).

320. Lannin, T. B. et al. Comparison and optimization of machine learning methods for automated classification of circulating tumor cells. *Cytometry Part A* **89**, 922–931 (2016).

321. Svensson, C. M. et al. Automated Classification of Circulating Tumor Cells and the Impact of Interobsever Variability on Classifier Training and Performance. *J. Immunol. Res.* **2015**, 573165 (2015).

322. Svensson, C. M. et al. Automated detection of circulating tumor cells with naive Bayesian classifiers. *Cytometry Part A* **85**, 501–511 (2014).

323. Guo, Z. *et al.* Circulating Tumor Cell Identification Based on Deep Learning. *Front. Oncol.* **12**, 843879 (2022).

324. Xing, F. et al. Deep Learning in Microscopy Image Analysis: A Survey. *IEEE Trans. Neural Netw. Learn. Syst.* **29**, 4550–4568 (2018).

325. Yamashita, R. et al. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging 2018 9:4* **9**, 611–629 (2018).

326. Oei, R. W. *et al.* Convolutional neural network for cell classification using microscope images of intracellular actin networks. *PLoS One* **14**, e0213626 (2019).

327. Shen, C. *et al.* Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning. *Scientific Reports 2023 13:1* **13**, 5708- (2023).

328. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision* **2017-October**, 618–626 (2017).

329. Singh, A. et al. Explainable Deep Learning Models in Medical Image Analysis. *J. Imaging*

6, 52 (2020).

330. Gorges, T. M. *et al.* Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition. *BMC Cancer* **12**, 178 (2012).

331. Grover, P. K. et al. Circulating tumour cells: the evolving concept and the inadequacy of their enrichment by EpCAM-based methodology for basic and clinical cancer research. *Annals of Oncology* **25**, 1506–1516 (2014).

332. Schneck, H. *et al.* EpCAM-Independent Enrichment of Circulating Tumor Cells in Metastatic Breast Cancer. *PLoS One* **10**, e0144535 (2015).

333. Nizamoglu, S. *et al.* A Simple Approach to Biological Single-Cell Lasers Via Intracellular Dyes. *Adv. Opt. Mater.* **3**, 1197–1200 (2015).

334. Chen, Y. C. et al. Versatile Tissue Lasers Based on High-Q Fabry-Pérot Microcavities. *Lab Chip* **17**, 538 (2017).

335. Chen, Y. C. *et al.* Laser-emission imaging of nuclear biomarkers for high-contrast cancer screening and immunodiagnosis. *Nat. Biomed. Eng.* **1**, 724 (2017).

336. Wu, W. *et al.* Antigen-independent single-cell circulating tumor cell detection using deep-learning-assisted biolasers. *Biosens. Bioelectron.* **271**, (2025).

337. Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nature Methods 2017 14:9* **14**, 849–863 (2017).

338. Way, G. P. *et al.* Predicting cell health phenotypes using image-based morphology profiling. *Mol. Biol. Cell* **32**, 995 (2021).

339. Parodi, M. *et al.* Hybrid epithelial-mesenchymal status of lung cancer dictates metastatic success through differential interaction with NK cells. *J. Immunother. Cancer* **12**, (2024).

340. Morvan, M. G. & Lanier, L. L. NK cells and cancer: you can teach innate cells new tricks.

Nature Reviews Cancer 2016 16:1 **16**, 7–19 (2015).

341. Wu, S. Y. et al. Natural killer cells in cancer biology and therapy. *Molecular Cancer 2020 19:1* **19**, 120- (2020).

342. Ye, L. *et al.* Circulating Tumor Cells Were Associated with the Number of T Lymphocyte Subsets and NK Cells in Peripheral Blood in Advanced Non-Small-Cell Lung Cancer. *Dis. Markers* **2017**, 5727815 (2017).

343. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, (2018).

344. Collins, G. S. et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. (2015).

345. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015).

346. Schmiedel, D. & Mandelboim, O. NKG2D ligands-critical targets for cancer immune escape and therapy. *Front. Immunol.* **9**, 393768 (2018).

347. Fionda, C. et al. NKG2D and DNAM-1 Ligands: Molecular Targets for NK Cell-Mediated Immunotherapeutic Intervention in Multiple Myeloma. *Biomed Res. Int.* **2015**, 178698 (2015).

348. Gallo, M. *et al.* Clinical utility of circulating tumor cells in patients with non-small-cell lung cancer. *Translational Lung Cancer Research* vol. 6 486–498 (2017).

349. Scharpenseel, H. *et al.* EGFR and HER3 expression in circulating tumor cells and tumor tissue from non-small cell lung cancer patients. *Sci. Rep.* **9**, (2019).

350. Badia-I-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer

biological activities from omics data. *Bioinformatics Advances* **2**, (2022).

351. Schubert, M. *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications 2017 9:1* **9**, 20- (2018).

352. Holland, C. H. *et al.* Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, (2020).