As an alternative target to surgically resected tissue specimens, liquid biopsy has gained much attention over the past decade. Of the various circulating biomarkers, circulating tumor cells (CTCs) have particularly opened new windows into the metastatic cascade, with their functional, biochemical, and biophysical properties. Given the extreme rarity of intact CTCs and the associated technical challenges, however, analyses have been limited to bulk-cell strategies, missing out on clinically significant sources of information from cellular heterogeneity. With recent technological developments, it is now possible to probe genetic material of CTCs at the single-cell resolution to study spatial and temporal dynamics in circulation. Here, we discuss recent transcriptomic profiling efforts that enabled single-cell characterization of patient-derived CTCs spanning diverse cancer types. We further highlight how expression data of these putative biomarkers have advanced our understanding of metastatic spectrum and provided a basis for the development of CTC-based liquid biopsies to track, monitor, and predict the efficacy of therapy and any emergent resistance.

INTRODUCTION

Despite the first report on breast circulating tumor cells (CTCs) in 1869,1 techniques for isolating these circulating biomarkers were only first described in 1960,2,3 and were gradually improved over the next 40 years. While much progress has been made with the albumin gradient method and FDA-approved CellSearch® system during the first generation of CTC research,4 the existence of heterogeneous CTC subpopulation highlighted the need to develop marker-independent isolation technologies.5,6 Since then, label-free techniques utilizing the principles of biophysical properties have been developing rapidly as the second generation.7,8 FDA-approved/listed platforms, such as CellSearch® (Siemens Healthcare Diagnostics) and ClearCell® FX (Biolidics) are exemplary technologies that have been widely used and demonstrated the clinical significance of CTCs.9–12

Currently, 265 clinical trials regarding CTCs are listed on clinicaltrials.gov. Despite successful CTC enumeration, achieving high yield and high purity remains challenging because of millions to billions of blood cells and a few to tens of CTCs present as background and target cells, respectively, in a milliliter of whole blood from cancer patient.13 It has been posited that the conventional EpCAM-based enrichment method would require 5 L of blood to detect at least one CTC in metastatic disease with 99% sensitivity.9 Such exceptionally low CTC frequencies could be attributed to progressively lost expression of epithelial markers during epithelial-to-mesenchymal transition (EMT) in circulation,14,15 as higher CTC counts have been reported with immunologic or physical property-based enrichment.16–18

In addition to the wide range of CTC detection rate reported in clinical studies, broad phenotypic plasticity and diversity have been observed at multiple molecular levels during metastatic cascade – from EMT and invasion19–21 to evasion of apoptosis,22 chemoresistance,23 migration,24 intravasation,25 extravasation, and organ colonization.26 While a tumor biopsy from either primary tumor or metastatic lesion alone may not always recapitulate the entire tumor harboring segregated clones,27 spatiotemporally heterogeneous CTCs collected in a sequential manner could potentially reveal comprehensive window into the metastatic disease for real-time monitoring of therapy response, which remains an unmet need in current clinical practice with tissue biopsy.

SINGLE-CELL ANALYSIS

Emerging sequencing data from spatially distinct tumors provide clear evidence of intratumoral heterogeneity.28–30 Owing to the technical challenges, however, CTC analyses have been limited to bulk-cell samples, missing the information on cellular heterogeneity. The inevitable leukocyte contamination in any given primarily enriched sample further complicates downstream molecular analyses. Such confounding effect is particularly pronounced in transcriptomic studies when the activated leukocytes concurrently overexpress cancer-associated biomarkers, such as MUC1 or HER2, masking the true expression of CTC-specific transcripts.31 Their mesenchymal nature and hematopoietic origin further interfere with the expression of EMT-related and stem cell markers, respectively, resulting in false-positive observations.32

The transition from bulk to single-cell analyses on patient-derived CTCs has thus been fueled in part by studies over the past five years. At the genomic level, they have identified clinically relevant alterations, ranging from small-scale (e.g., single nucleotide variation (SNV), microsatellite instability) to large-scale mutations (e.g., copy-number variation, large-scale state transition, inter/intrachromosomal rearrangement). These aberrations...
include time-varying SNVs during the course of chemotherapy,\textsuperscript{33} private mutations that are absent in either matched primary or metastatic tumor,\textsuperscript{34} and that are not yet listed in the COSMIC database (http://cancer.sanger.ac.uk),\textsuperscript{35} and copy-number profiles that distinguish chemosensitive from chemorefractory disease.\textsuperscript{23} Although limited in sample size and number of studies, transcriptomic studies have further revealed complex and heterogeneous expression patterns within and across patients. For example, expression profiles of single CTCs have demonstrated superior diagnostic accuracy in defining lineage identity and in identifying clinically distinct subsets of tumors across multiple myeloma and prostate cancers.\textsuperscript{36,37} They have also revealed therapeutically relevant biomarkers\textsuperscript{38} (e.g., PI3K-AKT-mTOR) and that are potentially targetable.\textsuperscript{24,36,38,41} INTEGRATED WORKFLOW

Despite the prevalence of EpCAM- CTCs,\textsuperscript{44} and varying capture efficiency,\textsuperscript{15} epithelial marker-dependent CellSearch\textsuperscript{®} technology remains as the most common enrichment method to isolate CTCs from patient-derived peripheral blood. Pre-enrichment is often required for recovery of preferably viable and intact CTCs, and can be performed with direct imaging modalities,\textsuperscript{30} density gradient centrifugation in Percoll or Ficoll,\textsuperscript{12,42} immunoaffinity,\textsuperscript{42,46–48} microfiltration in two\textsuperscript{33} and three\textsuperscript{31} dimensions, and microfluidic approaches.\textsuperscript{39,40} Table S1 summarizes cell sorting and isolation technologies, including methods, working principles, features, limitations, and the reported recovery rates of spiked cancer cells. Primarily enriched bulk CTC samples are subsequently subjected to manual cell picking or micromanipulation,\textsuperscript{24,32,36,41–43}

| Cancer type         | CTC enrichment                      | CTC criteria (micromanipulation) | Single-cell profiling       | Number of CTCs (number of patients) | Reference     |
|---------------------|-------------------------------------|----------------------------------|-----------------------------|-----------------------------------|---------------|
| Multiple myeloma    | FACS with serial dilution CellSearch\textsuperscript{®} | CD45\textsuperscript{−}, CD138\textsuperscript{+} | SMART-seq2                  | 21 (2)                            |\textsuperscript{36} |
| Colon               | Biocoll separation, Dynabeads\textsuperscript{®} CD45 depletion | CD45\textsuperscript{+}, EpCAM\textsuperscript{+} | Multiplex PCR               | 11 (8)                            |\textsuperscript{24} |
| Ovary               | Biocoll separation, Dynabeads\textsuperscript{®} CD45 depletion | DAPI\textsuperscript{+}, CK/EpCAM\textsuperscript{+}, CD45 | Multiplex PCR               | 15 (3)                            |\textsuperscript{32} |
| Breast              | MagSweeper\textsuperscript{®}       | EpCAM\textsuperscript{+}         | Microfluidic RT-PCR\textsuperscript{b} | 105 (50)                          |\textsuperscript{42} |
|                     | Microfluidic\textsuperscript{®} iChip | EpCAM/HER2/CDH11\textsuperscript{+}, CD45/CD16/CD14 | Optimized Tang’s method         | 15 (10)                           |\textsuperscript{49} |
|                     | Microfluidic ClearCell\textsuperscript{®} FX | EpCAM/HER2/EGFR\textsuperscript{+}, CD45 | SMART-seq v4\textsuperscript{e} | 15 (10)                           |\textsuperscript{38} |
| Melanoma            | MagSweeper\textsuperscript{®}       | CD45\textsuperscript{−}, Calcein\textsuperscript{−} | Polaris\textsuperscript{TM} IFC | 68 (4)                            |\textsuperscript{50} |
| Prostate            | MagSweeper\textsuperscript{®}       | CD45\textsuperscript{−}, EpCAM\textsuperscript{+}, DAPI\textsuperscript{+} | SMART-seq                     | 6 (1)                             |\textsuperscript{48} |
|                     | ScreenCell\textsuperscript{®}       | CD45\textsuperscript{−}          | Microfluidic RT-PCR\textsuperscript{a} | 38 (9)                            |\textsuperscript{43} |
|                     | Microfluidic CTC-iChip              | CD45\textsuperscript{−}, EpCAM/CDH11\textsuperscript{+} | Modified Tang’s method         | 77 (13)                           |\textsuperscript{37} |
|                     | Integrated nanoplatform             | EpCAM\textsuperscript{+}         | Multiplex PCR                | 8 (1), 18 (1), 74 (1)             |\textsuperscript{47} |
|                     | Microfluidic ClearCell\textsuperscript{®} FX | CD45\textsuperscript{−}         | Multiplex PCR                | 61 (20)                           |\textsuperscript{76} |
| Lung                | CellSearch\textsuperscript{®}, Parosort\textsuperscript{TM} | EpCAM/pan-keratins\textsuperscript{−} | Multiplex PCR                | 13 (1), 8 (1)                     |\textsuperscript{41} |
|                     | Microfluidic CTC-iChip              | CD45\textsuperscript{−}          | Modified Tang’s method        | 7 (--), 29 (--), 77 (--)' |\textsuperscript{51} |

\textsuperscript{a}Number of CTCs (patients) included in the final analysis 
\textsuperscript{b}NanoFlex\textsuperscript{®} 4-IFC Controller and BioMark\textsuperscript{®} Real-Time PCR System 
\textsuperscript{c}Droplet digital PCR (BioRad ddPCR\textsuperscript{TM}) 
\textsuperscript{d}Microfluidic Polaris\textsuperscript{®} was used for single-CTC isolation 
\textsuperscript{e}BioMark\textsuperscript{TM} HD MX/HX system 
\textsuperscript{f}Microfluidic chip was used for single-CTC isolation
versions of SMARTer® (e.g., SMART-Seq2) and that of Tang's method70 are commonly employed in single-CTC transcriptomic studies aiming to achieve improved transcript detection, coverage, accuracy, and yield (Table 1).

For quantitative transcriptomic analysis, an accurate identification of technical artifacts from intrinsic biological cellular variability is critical to prevent spurious readings from single CTCs. Ideally, quality control (QC) metrics should be performed with the amplified cDNA products after preamplification step, given the amount of genetic material minimally required. Current single-cell transcriptomic studies on patient-derived CTCs have assessed (1) the yield or concentration of amplified DNA,36,64–66 (2) Cq values for selected reference, or housekeeping, genes using qPCR,33,58 (3) fragment size distribution of selected DNA sequences using gel/capillary electrophoresis,36,58,64,67–71 and (4) genome integrity index (GII), which ranges from 0 (poor quality) to 4 (high quality) and is computed based on PCR bands of four primer pairs using gel electrophoresis.23,72,73

Cells harboring QC-passed RNAs are subsequently subjected to library construction, followed by scRNA-seq36,38,40,46,48–51 or quantitative profiling with conventional qPCR,24,32 digital droplet PCR (ddPCR)19,40 or microfluidic dynamic array.42,43 The quality of constructed libraries are further validated with (1) the proportion of reads mapping to genome, and/or (2) the number of genes detected.50 Lineage specificity of CTCs is often confirmed by high expression of cancer-specific markers and low expression of leukocyte markers with pre-specified thresholds.37 Low success rate of <60% for overall amplification and library preparation attributed to multiple processing steps has been reported in CTC studies, highlighting the need to systematically quantify QC metrics prior to the analysis.

EXPRESSION DATA

Single-cell transcriptome of patient-derived CTCs have been analyzed comparatively with cancer cell lines, white blood cells (WBCs), matched primary tumors and/or metastases.24,37,42,43,48 Alternatively, expression levels were assessed and compared between CTC subgroups defined by unsupervised hierarchical clustering or other classification methods.41,49 In the following sections, we focus on the most relevant gene signatures that are perceived to be critical determinants of metastasis and disease outcome and that are commonly differentially expressed in CTCs at the single-cell level: EMT, stemness, interaction with blood components, DNA repair, signaling pathways and drug targets (Table 2).

In line with molecular evidences supporting EMT-driven metastasis,24,32,41–43,49 bulk-cell studies have suggested the contribution of EMT to early steps of the metastatic spread (i.e., tumor invasion, intravasation, CTC generation and survival, and early seeding in secondary organs).72 Nonexclusive hypotheses of EMT's contribution to CTC biology suggest that (1) CTCs may have been mesenchymally-shifted in primary tumors to have enhanced survival properties through activation of genes involved in survival pathways and escape from immune surveillance, or (2) undergo EMT processes within the bloodstream by means of TGFβ liberated from circulating platelets.74 At the center of the research axis is to identify and characterize such premetastatic subsets of CTC population that are favored to be liberated from primary

---

**Fig. 1** Microfluidic technologies for single-cell molecular characterization of patient-derived CTCs. **a** ClearCell® FX-integrated workflow. Single-cell genomic analysis: High concordance rate of EGFR mutations (T790M and L858R) was found between NSCLC CTCs and matched primary tumors. Single-cell transcriptomic analysis: Patient classification was done for breast cancer and NSCLC through full-length mRNA transcriptomic analysis50 and targeted gene expression profiling,39,40 respectively. Single-cell proteomic analysis: Supervised principal component analysis (PCA) revealed unique metabolic profiles between CTCs and lymphocytes in gastric and colorectal cancer patients.115 **b** Single-cell proteomic analysis: Microfluidic single-cell western blotting (scWB) enabled the rapid analysis of an eight-plex protein expression in ERα breast cancer.116 **c** Single-cell secretomic analysis: The integrated microfluidic on-chip system revealed highly heterogeneous expression profiles of two secreted proteins (i.e., IL-8 and VEGF) in CTCs from lung cancer patients.117

---

**EXPRESSION DATA**

Single-cell transcriptome of patient-derived CTCs have been analyzed comparatively with cancer cell lines, white blood cells (WBCs), matched primary tumors and/or metastases.24,37,42,43,48 Alternatively, expression levels were assessed and compared between CTC subgroups defined by unsupervised hierarchical clustering or other classification methods.41,49 In the following sections, we focus on the most relevant gene signatures that are perceived to be critical determinants of metastasis and disease outcome and that are commonly differentially expressed in CTCs at the single-cell level: EMT, stemness, interaction with blood components, DNA repair, signaling pathways and drug targets (Table 2).

In line with molecular evidences supporting EMT-driven metastasis,24,32,41–43,49 bulk-cell studies have suggested the contribution of EMT to early steps of the metastatic spread (i.e., tumor invasion, intravasation, CTC generation and survival, and early seeding in secondary organs).72 Nonexclusive hypotheses of EMT’s contribution to CTC biology suggest that (1) CTCs may have been mesenchymally-shifted in primary tumors to have enhanced survival properties through activation of genes involved in survival pathways and escape from immune surveillance, or (2) undergo EMT processes within the bloodstream by means of TGFβ liberated from circulating platelets.74 At the center of the research axis is to identify and characterize such premetastatic subsets of CTC population that are favored to be liberated from primary tumors.
tumors and survive in the bloodstream to succeed in the early colonization phases. scRNA-seq is particularly well suited in this regard to discover distinct subsets of CTCs capable of forming metastasis.

Emerging single-cell profiling data provide clear evidence of a continuum in the development of CTC phenotypes, including epithelial (E), epithelial-mesenchymal (E/M), mesenchymal (M) and stem-like phenotype. Highly heterogeneous expression of epithelial markers (e.g., EpCAM, CK18, CK19) was observed in CTCs across colorectal, ovarian, breast, and prostate cancers. Similarly, mesenchymal or EMT-related genes (e.g., CDH2, VIM, TGFß1, ZEB1/2) were commonly enriched in single ovarian, breast, and prostate CTCs. Compared to matched primary/metastatic tumor tissue and cell lines, migration-related and cell-cell adhesion genes (e.g., TSPAN8, CD151, CD44v6, and FN1) were generally downregulated in colon and prostate CTCs, respectively, possibly suggesting a lack of their need for mobility and migratory capabilities in the bloodstream in these cancer types.

While some studies have suggested that CTCs that are ‘frozen’ in either E or M state lacking EMT plasticity are unable to form metastases, the repetitive observation of patient-derived CTCs expressing mesenchymal attributes directly correlated with the appearance of metastases in recent studies suggest that these mesenchymal-shifted cells, and not benign cells passively detached from a primary tumor, are precursors of metastasis. Clinical data linking CTC-derived EMT markers with multiple clinical parameters are discussed elsewhere. It remains to be elucidated whether there exist specific hybrid E/M states that are particularly prone to perturbations triggering extensive phenotypic and functional change in circulation. Computational analytical tools enabling pseudo-time reconstruction of transitioning cells and the topography underlying E/M plasticity from...
a static single-cell gene expression data may be applied to prospective studies to clarify the nature of hybrid E/M states and define their role in metastasis.

It has been suggested that tumor cells having an intermediate phenotype of EMT show the highest plasticity and thus represent cancer stem cells (CSCs). Varying expression levels of stem cell markers (e.g., CD24, CD44, ALDH1A1, NANOG, and OCT4) were found at the single-cell level in ovarian,

breast,

and prostate CTCs.

Interestingly, genes involved in oncogenic signaling pathway were found to be differentially expressed in single CTCs depending on the level of cellular plasticity or stemness. In prostate cancer, for example, expression of the key regulators in the PI3K/Akt/mTOR signaling pathway (i.e., PI3K, mTOR) were highly expressed in CD44+/CD24+ CTCs. This subset of CTC populations may thus be more susceptible to perifosine (Akt inhibitors) and rapamycin (mTOR inhibitors) treated with conventional chemotherapy or radiotherapy. Given that both CD44+/CD24+ and CD44+/CD24+ tumor cells have functional significance in initiating tumor growth and that CTCs express these two cell surface markers across various malignancies, it remains to be investigated whether their expression has comparable functional significance in circulation.

The innate immune regulator, CD47, was the only gene that was upregulated in CTCs compared to matched tumor tissue in colorectal cancer, suggesting a potential immune-escape mechanism associated with CTC survival in circulation. Another form of immune evasion was suggested in melanoma, where the genes associated with the escape from immune surveillance, including HLA genes (i.e., HLA-G, HLA-H, HLA-C, and HLA-B) and TRPM1, were significantly downregulated in CTCs compared to melanoma cell lines, primary melanocytes, human embryonic stem cells, and lymphoma cell lines. Transcriptional repression of HLA genes has been associated with complete loss of MHC class I membrane expression, and importantly, the primary resistance to immune checkpoint inhibitor (ICI)-based immunotherapy. The screening of plasma-membrane proteins through whole transcriptomic analysis is thus of utmost interest to identify not only CTC-specific diagnostic biomarkers but also immune escape and survival mechanisms underlying resistance to immunotherapy.

Another key player is TGFβ-releasing platelet, which may adhere to CTCs in the bloodstream. In line with this hypothesis, platelet markers were frequently expressed in isolated single CTCs, as well as in CTC clusters in breast cancer. Labelle et al. showed that TGFβ liberated from platelets may induce EMT in tumor cells within the bloodstream and further promote the formation of the early metastatic niche. Based on these observations, it is speculated that TGFβ expressing CTCs may represent a specific subpopulation having high metastatic potential. Importantly, platelet-dependent natural killer (NK) cell escape mechanism has been suggested by in vitro and preclinical models across diverse mouse and human cancer cell lines. It is thus postulated that the presence of platelets may equip tumor cells with enhanced ability to escape elimination by the immune system through EMT, ultimately promoting their metastatic competency.

Genes involved in DNA repair (e.g., RAD51, PARP1) and G2/M DNA damage checkpoint (e.g., AR, TK1, PLK1, MAGEA1, MAGEC1, MAGEC2, CTAGB1, BIRC5, TOP2A) were frequently expressed in prostate CTCs. While several transcripts (e.g., PLK1, TOP2A) have been associated with aggressiveness in localized prostate cancer,

it is noteworthy that CTCs derived from advanced cancer patients also highly expressed these markers relative to normal prostate tissues. In contrast, compared to cancer cell lines and matched primary tumors, genes involved in cell proliferation (e.g., MYC, ATF3, TERT, RAC1, FOXA1, RRM1, CCNB1, BIRC5, Ki-67, c-Myc) were significantly downregulated in CTCs across breast and colorectal cancers, suggesting a non-proliferative, or dormant, state of CTCs in circulation.

Given the generally diminished expression of proliferation-related genes, conventional therapeutic strategies targeting proliferating cells may not be the best for eradicating “seeds” of metastasis. Promisingly genes involved in the PI3K-AKT-mTOR signaling pathway, in which many are currently in (pre)clinical trial stages or FDA-approved, and other potentially targetable genes were frequently expressed in CTCs at high levels across various malignancies including multiple myeloma, and breast, prostate and colorectal cancers. The incorporation of single-CTC analysis into clinical trials may thus be ideal from clinical perspective for the development of companion diagnostics.

**CLINICAL SIGNIFICANCE**

scRNA-seq or scPCR technologies have been widely applied to study early mammalian development, neuronal diversity, and immune system, revealing spatial and temporal dynamics, cellular heterogeneity, clonal distribution, pathways, and crosstalk. Their application in the context of CTC-based liquid biopsy, however, has been limited primarily to capturing a snapshot of the cellular states at a given point in time. In clinical contexts, it is the dynamics of such cellular state (i.e., temporal heterogeneity) that is of primary interest to monitor therapeutic response during the course of treatment. Despite technical challenges, a few single-cell studies have successfully demonstrated clinically-promising use of CTC-derived transcripts particularly for serial monitoring of the disease in a prospective cohort. Prior knowledge of genes of interest is often required in traditional gene expression analysis for clinical diagnosis of tissues or cells in circulation using immunohistochemistry, in situ hybridization, or flow cytometry, yielding semi-quantitative data. In contrast, scRNA-seq generates high-throughput expression data in an unbiased, objective manner, with superior diagnostic sensitivity over existing technologies. For example, scRNA-seq-acquired expression data of a few selected, well-established markers, which were previously used to sort multiple myeloma (MM) cells by flow cytometry, achieved near perfect accuracy in differentiating normal and malignant plasma cell. Similarly, the improved diagnostic performance of CTC-based multiplex assays was observed in advanced breast cancer, supporting the robust detection capability of single-CTC-derived markers.

The benefits of scRNA-seq technologies in the CTC field are particularly pronounced in classifying sub-populations of cells that may be clinically distinct, which are overlooked by conventional diagnostics due to the insufficient resolution. Unsupervised hierarchical clustering of single circulating MM cells-derived transcripts, for example, differed considerably from one patient to another, indicating the presence of different subtypes in MM. The existence of key chromosomal translocations associated with clinical risk may further be inferred from scRNA-seq data; circulating MM cells overexpressed CCND1 and CCND3 indicative of chromosomal translocations of CCND1/IGH fusion from t(11;14) and CCND3/IGH fusion from t(6;14), respectively, and the presence of these genomic aberrations were further validated in matched MM by fluorescence in situ hybridization (FISH). Provided that the overexpression of CCND1 has been associated with resistance to EGFR-, BRAF- and MAF-targeted therapies, single-CTC transcriptome may be used as a predictive indicator for diagnosis, MM classification, and therapeutic efficacy in clinical settings.

Transcriptomes of single CTCs have been analyzed comparatively with that of CTC clusters, which have been associated with enhanced metastatic competence and poor prognosis across multiple cancer types. Differential expression analysis between the two groups identified specific gene signature (e.g., cell junction component plakoglobin) required in forming CTC cluster and distant metastases, in which high expression levels

**S.B. Lim et al.**

Published in partnership with The Hormel Institute, University of Minnesota

npj Precision Oncology (2019) 23
were indicative of short metastasis-free survival.\textsuperscript{49} Further, single-CTC-derived transcriptomes revealed signaling pathways (i.e., non-canonical Wnt signaling) relevant to treatment response (i.e., anti-androgen resistance), which was not evident in matched primary tumors, in prostate cancer.\textsuperscript{57} Similarly, the potential role of Sonic Hedgehog, Wnt, and TGFβ signaling pathways in metastatic castration-resistance and immunotherapy response was suggested by single-CTC expression profiles in prostate cancer.\textsuperscript{93}

By applying label-free microfluidic approaches,\textsuperscript{12} our group recently demonstrated how previously developed prognostic index,\textsuperscript{96,97} and the resulting prognostication, can be refined with single-CTC-derived gene signatures while accounting for cellular heterogeneity.\textsuperscript{76} Expression of a subset of matrisome genes, including MMP1 and MMP12, in tumor tissues and CTCs was consistently associated with metastatic spread and early recurrence of non-small-cell lung cancer (NSCLC), respectively. In line with our earlier observations of EGFR mutations (e.g., T790M/L858R) in CTCs and matched tumor tissue,\textsuperscript{12} this study provides the molecular evidence linking a solid tumor with single-CTC-based indicator associated with clinical presentation at the transcript level, highlighting promising predictive value of circulating biomarkers.

Despite a well-established role as a biomarker prognostic of survival particularly in breast cancer,\textsuperscript{38,99} microscopy-based CTC enumeration alone may not be sufficient to predict drug resistance in the absence of robust molecular characterization. As an advanced alternative, single-CTC-derived transcripts may serve as an excellent source to develop quantitative scoring assay comprising tissue lineage-specific genes that are not present in normal blood cells. To date, very few studies have applied such metrics in a prospectively monitored patient cohort to demonstrate their practical utility in the clinical setting.\textsuperscript{38–40}

In 2018, the Haber group presented a predictive digital CTC scoring strategy to identify patients with poor overall survival (OS) and progression-free survival (PFS) in metastatic castration-resistant prostate cancer treated with first-line abiraterone.\textsuperscript{40} Serial monitoring of CTCs further predicted early dissemination in another independent cohort of patients with localized cancer.\textsuperscript{40} The same group found that digital quantification of intracellular ER signaling in single CTCs was predictive of residual disease in localized breast cancer patients treated with neoadjuvant therapy.\textsuperscript{38} Importantly, this 17-gene CTC score predicted early progression in metastatic breast cancer treated with endocrine therapy, which was not adequate to suppress ER signaling, despite having functional ESR1.\textsuperscript{38}

The greatest focus in immuno-oncology has been on tumor biopsy-derived features, such as PD-L1 expression, tumour-infiltrating lymphocyte (TIL) density, T-cell receptor (TCR) clonality, mutational burden, and immune gene signatures, for their increasingly recognized predictive values for ICI-based immunotherapy.\textsuperscript{100} Although promising, their invasive nature makes repeated sampling not clinically practical particularly for metastatic diseases over the course of treatment. The development of less-invasive CTC-based liquid biopsies as a predictive biomarkers for response to ICI treatments will therefore be particularly promising. Hong et al. recently showed that the scoring model recapitulating temporal dynamics of CTCs identified patients with better OS and PFS in ICI-treated melanoma patients, demonstrating the feasibility of quantifying transcripts derived from microfluidically enriched CTCs for predicting patients likely to benefit from ICI therapies.\textsuperscript{39} Larger studies will be required to develop and establish such generalized framework for guiding therapeutic decision-making.

**CHALLENGES AND BEYOND**

The advent of sequencing technologies has created a new era of precision medicine. The prospect of applying this concept to develop clinically applicable biomarkers for diagnosis, prognosis, and prediction of therapy response has been extensively explored on cancer patients. Particularly, liquid biopsies focusing on the analysis of CTCs and cell-free tumor DNA (ctDNA) in the bloodstream are evolving into promising clinical parameters.\textsuperscript{101} ctDNA may allow mutational analyses to monitor tumor dynamics during cancer treatment\textsuperscript{102} and offer easier handling, storage, and shipping of samples compared to CTCs.\textsuperscript{103} A small number of mutant gene fragments present in ctDNA, which are further diluted by normal circulating DNA fragments released by apoptotic cells, however, require highly sophisticated methods to accurately assess tumor-specific genomic alterations (a detailed comparison between the two types of analytes is beyond the scope of this review). Nevertheless, a recently developed CancerSEEK blood test that examines the presence of mutations in cfDNA has achieved high sensitivity ranging from 69 to 98% across five cancer types, showing great promise for early cancer detection.\textsuperscript{104}

CTCs represent intact and viable tumor cells that can be analyzed at multiple biological levels, allowing sequential sampling at multiple time points from patients undergoing systemic drug treatment. It is thus possible to perform multidimensional molecular and phenotypic characterization of these cells, which increasingly serves as an essential tool in precision diagnosis.\textsuperscript{105} Nevertheless, challenges remain in the field as these putative metastatic precursor cells occur at extremely low frequency relative to normal leukocytes in any given clinical sample. Such rare nature of CTCs clearly raises the question of whether these cells obtained at a single time point alone would truly recapitulate spatially and temporally evolving landscape of the entire tumor and its microenvironment, or the metastatic state. Further, the recovery efficiency varies greatly across the enrichment technology, posing additional challenges in understanding their cellular heterogeneity and the functional and clinical significance of their appearance in the bloodstream. Consequently, little is known about the molecular characteristics and mechanisms, particularly in relation to drug resistance and their capacity in circulating bloodstream with metastatic potential.

The confounding effects of inherent rarity and heterogeneity of CTCs on the downstream analysis may further be exacerbated by biased positive selection during single-cell isolation (i.e., enrichment of target cells based on antibodies specific to CTC surface markers), missing out cells with low or even no surface marker expression; in circulation which prove to be of clinical significance.\textsuperscript{2} CTCs in advanced disease indeed exhibited predominant epithelial-mesenchymal-mixed (E/M), or mesenchymal (M) phenotypes (i.e., expressing mesenchymal markers) across multiple cancer types, including esophageal squamous carcinoma,\textsuperscript{106} ovarian cancer,\textsuperscript{107} pancreatic cancer,\textsuperscript{108} colorectal cancer,\textsuperscript{109} triple-negative breast cancer,\textsuperscript{110} and hepatocellular carcinoma.\textsuperscript{111} These EMT-shifted CTCs would not have been detected by immunoaffinity-based enrichment solely facilitated by antibodies targeting epithelial markers (e.g., EpCAM and pan-keratins). Capture efficiency may thus be enhanced by using cocktails of antibodies,\textsuperscript{112} including both epithelial and mesenchymal biomarkers, or by utilizing tumor lineage-specific signatures\textsuperscript{28} without making a priori assumption about the type of tumor cells. In contrast, negative depletion (i.e., removal of non-target cells) using label-free approaches which leverage unique physical properties (e.g., cell size) of CTCs may lead to relatively low purity given the size overlap with leukocytes,\textsuperscript{112} as observed across breast, colorectal and prostate cancers.\textsuperscript{113} Some may even present a similar immunofluorescence staining pattern with leukocytes expressing both leukocyte- and CTC-specific markers, adding layers of complexity. Although such “double positive” cells are often excluded from the analysis, their occurrence in healthy blood samples at a much lower frequency point towards their possible functional role and clinical impact.\textsuperscript{114} Microfluidic
approaches are increasingly being applied in this regard to enable both WBC elimination and selective CTC isolation on a single platform, as demonstrated by two-stage microfluidic chips.115,116

Finally, many single-CTC studies do not state the total number of cells initially isolated during the enrichment step, and the quality and number of CTCs that have failed QC and that have been excluded from further analysis. This makes the direct comparison of transcriptional changes found in CTCs between studies extremely difficult, as such molecular findings may only be applicable to a small subpopulation of CTCs depending on the enrichment technology or the QC metrics. The development of a clearly defined and more uniform workflow is thus urgently needed to facilitate its clinical application at different stages of the antitumor therapy or cancer progression across, and within, patients. The 17-gene CTC-specific assay is an exemplary quantitative scoring metrics that has achieved high sensitivity for monitoring of therapy response in localized and metastatic breast cancer patients. Continuous optimization of the developed platform and prospective clinical validation of CTC-based liquid biopsy will ultimately provide clinicians with robust, yet readily understandable, test results in a shorter turnaround time compared to conventional tissue biopsy.

References:

1. Ashworth, T. R. A case of cancer in which cells similar to those in the tumors were seen in the blood after death. Australas. Med. J. 14, 146–149 (1869).
2. Salgado, I. et al. Tumour cells in the blood. Can. Med. Assoc. J. 81, 619–622 (1959).
3. Alexander, R. F. & Spriggs, A. I. The differential diagnosis of tumour cells in circulating blood. J. Clin. Pathol. 13, 414–424 (1960).
4. Gwak, H. et al. Progress in circulating tumor cell research using microfluidic devices. Microachines 9, 353 (2018).
5. Pecot, C. V. et al. A novel platform for detection of CK+ and CK− CTCs. Cancer Discov. 1, 580–586 (2011).
6. Mikolaiczak, S. D. et al. Detection of EpCAM-Negative and cytokertatin-negative circulating tumor cells in peripheral blood. J. Oncol. 2011, 252361 (2011).
7. Poudineh, M., Sargent, E. H., Pantel, K. & Kelley, S. O. Prognostic and predictive potential of circulating tumor cells in breast cancer patients. Breast Cancer Res 11, R46 (2009).
8. Blasci, C. et al. Gene expression profiling of single-cell tumor cells present in ovarian cancer—establishment of a multi-marker gene panel. Mol. Oncol. 10, 1030–1042 (2016).
9. Nl, X. et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc. Natl Acad. Sci. USA. 110, 21083–21088 (2013).
10. Heitzer, E. et al. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res. 73, 2965–2975 (2013).
11. Forbes, S. A. et al. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nuclear acids Res. 39, D945–D950 (2011).
12. Lohr, J. G. et al. Genetic interrogation of circulating multiple myeloma cells at single-cell resolution. Science. Transl. Med. 8, 363ra147 (2016).
13. Miyamoto, D. T. et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science 349, 1351–1356 (2015).
14. Warkiani, M. E. et al. Ultra-fast, label-free isolation of circulating tumor cells from blood using spiral microfluidics. Nat. Protoc. 11, 134–148 (2016).
15. Warkiani, M. E. et al. Slanted spiral microfluidics for the ultra-fast, label-free isolation of circulating tumor cells. Lab on a chip 14, 128–137 (2014).
16. Yu, M. et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science 339, 580–584 (2013).
17. Lechpertz, T. 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).
18. Satelli, A. et al. Epithelial-mesenchymal transitioned circulating tumor cells capture for detecting tumor progression. Clin. Cancer Res 21, 899–906 (2015).
19. Yu, M. et al. RNA sequencing of pancreatic circulating tumor cells implicates WNT signalling in metastasis. Nature 487, 510–513 (2012).
20. Carter, L. E. et al. Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemoinsensitive and chemorefractory small-cell lung cancer. Nat. Med. 23, 114–119 (2017).
21. Steinert, G. et al. Immune Escape and Survival Mechanisms in Circulating Tumor Cells of Colorectal Cancer. Cancer Res. 74, 1694–1704 (2014).
22. Wang, Y. et al. Single nucleotide variant profiles of viable single circulating tumor cells reveal CTC behaviours in breast cancer. Oncol. Rep. 39, 2147–2159 (2018).
23. Barbaranz, J. et al. Molecular characterization of circulating tumor cells in human metastatic colorectal cancer. PloS ONE 7, e40476 (2012).
24. Alizadeh, A. A. et al. Toward understanding and exploiting tumor heterogeneity. Nat. Med. 21, 846–853 (2015).
25. Li, H. et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat. Genet. 49, 708–717 (2018).
26. Brooks, M. D., Burness, M. L. & Wicha, M. S. Therapeutic implications of cellular heterogeneity and plasticity in breast cancer. cell stem cell 17, 260–271 (2015).
27. Janiszewska, M. et al. In situ single-cell analysis identifies heterogeneity for PI3KCA mutation and HER2 amplification in HER2-positive breast cancer. Nat. Genet. 47, 1212–1219 (2015).
28. Akats, B. et al. Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. Breast Cancer Res 11, R46 (2009).
29. Blasci, C. et al. Gene expression profiling of single circulating tumor cells in ovarian cancer—establishment of a multi-marker gene panel. Mol. Oncol. 10, 1515–1525 (2016).
30. Warkiani, M. E. et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science 349, 1351–1356 (2015).
31. Kwan, T. T. et al. A digital RNA Signature of circulating tumor cells predicting early therapeutic response in localized and metastatic breast cancer. Cancer Discov. 8, 1286–1299 (2018).
32. Hong, X. et al. Molecular signatures of circulating melano7 cells for monitoring early response to immune checkpoint therapy. Proc. Natl Acad. Sci. USA. 115, 2467–2472 (2018).
33. Miyamoto, D. T. et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science 349, 1351–1356 (2015).
34. Gorges, T. M. et al. Accession of tumor heterogeneity by multiplex transcriptome profiling of single circulating tumor cells. Clin. Chem. 62, 1504–1515 (2016).
35. Powell, A. A. et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PloS ONE 7, e33788 (2012).
36. Shen, C. L. et al. Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EM7-related genes in metastatic prostate cancer. Prostate 73, 813–826 (2013).
37. Mcgo, M. et al. Characterization of metastatic breast cancer patients with nondetectable circulating tumor cells. Int. J. cancer 129, 417–423 (2011).
38. Gores, E. et al. A direct comparison of CellSearch and ISET for circulating tumor-cell detection in patients with metastatic carcinomas. Br. J. Cancer 105, 847–853 (2011).
39. Cann, G. M. et al. mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. PloS ONE 7, e33788 (2012).
40. Park, S. M. et al. Molecular profiling of single circulating tumor cells from lung cancer patients. Proc. Natl Acad. Sci. USA 113, E8379–E8386 (2016).
41. 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).
42. Aceto, N. et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell 158, 1110–1122 (2014).
ACKNOWLEDGEMENTS

This work was carried out at the MechanoBioengineering Laboratory at the Department of Biomedical Engineering, National University of Singapore (NUS). We thank support provided by the Institute for Health Innovation and Technology (iHealthtech) at NUS. W.-T.L. is supported by the National Medical Research Council (NMRC/CSA/040/2012 and NMRC/CSA-INV/0025/2017). S.B.L. is supported by NUS Graduate School for Integrative Sciences and Engineering (NGS), Mogam Science Scholarship Foundation, and Daewoong Foundation. We thank Zac Goh for the illustration of single-cell microfluidic device.

AUTHOR CONTRIBUTIONS

C.T.L. supervised the conception of the work. S.B.L., W.D.L., J.V., W.-T.L., and C.T.L. wrote and revised the paper.

COMPETING INTERESTS

C.T.L. is a co-founder of Biolidics Ltd. W.-T.L. and C.T.L. are shareholders of Biolidics Ltd. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41698-019-0095-0.

Correspondence and requests for materials should be addressed to C.T.L.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019