Circulating Tumor Cells (CTC) have been analyzed using classical descriptive approaches for nearly 100 years. The assumption has always been that genomic alterations observed in primary tumor tissue would be faithfully recapitulated in the circulating tumor cell DNA (ctcDNA). Data emerging in recent years, however, has challenged this assumption. Technology enabling the purification and molecular analysis of ctcDNA has given a more comprehensive picture of the heterogeneity of both circulating tumor cells and ctcDNA. There are several factors that affect how we understand and interrogate tumor derived cell populations from a blood sample. CTC allow a longitudinal survey of the molecular pathology of disease rather than being a direct analogue of FFPE samples. The critical difference is that ctcDNA is biologically selected as the mobile manifestation of the tumor driving progression and metastasis. The development of molecular tools demands novel platforms to recover ctcDNA in a manner that allow analysis of specific molecular targets.

CTCs in cancer patients

Attempts to characterize the tumor cell population found in blood evolved from the observation by Ashworth in 1869 that cells in the blood resembled cells found in the tumor at autopsy. This observation, along with the knowledge that it is tumor cell dissemination to metastatic sites that causes mortality in cancer has driven the effort to understand the nature of the circulating tumor cell population in cancer patients. Epithelial cells have been shown to be present in elevated numbers in the blood of patients with cancer\(^1,2\). As validated tools to isolate these cells became available\(^3\), it was shown that elevation of numbers of circulating epithelial cell was prognostically related to metastatic disease in breast, prostate and colon cancer\(^4-6\) and can predict the metastatic potential before dissemination\(^7\). Unfortunately, the effort to use enumeration as a decision tool was unsuccessful in the SWOG 0500 trial in 2014 with the authors suggesting that molecular analysis of ctcDNA might have been better used to determine whether patients were eligible for targeted therapy trials\(^8\).

More recently, enumeration of epithelial cells has also been shown to have prognostic value during the first-line treatment of metastatic disease for both non-small cell lung cancer (NSCLC)\(^7\) and small cell lung cancer (SCLC)\(^10,11\).

The proportion of CTCs that are mutation-bearing cells and derived from a tumor isn’t well described. This is because of the heterogeneity of epithelial cells present either constitutively, or due to inflammatory or healing mechanisms\(^12,13\). All attempts to enumerate CTCs have revealed small numbers of circulating epithelial cells (CECs) in normal, healthy donors. Endogenous CECs are seen using scanning platforms\(^14\), expression analysis\(^15\), and CTC chips in breast\(^16\) or prostate cancer samples\(^17\), or in lung cancer using high-definition imaging\(^18\). Without molecular classifiers, these CECs have been shown to contribute to an enumerated CTC population\(^12\) even though they can also be shown to describe cells with a wild type genomic configuration\(^18\).

Outside of adenocarcinoma, the presence of circulating, tumor-derived cells has also been validated in malignant melanoma\(^20-23\). However, in this setting, the epithelial markers are not useful and the description of a separate biomarker for recovery of tumor cells is
necessary. The inevitable elaboration of alternative disease-specific capture biomarkers augurs the bright future for tunable rare cell isolation and characterization.

**Mutation-bearing cells in the CTC population**

To what extent does the population of cells that are being enumerated as “CTC” include mutation bearing tumor derived cells? Numerous breast cancer researchers have described a subpopulation of CTCs that over-express Her2+ demonstrating in parallel that all samples have mixed populations of Her2+ and Her2- CTCs. In prostate cancer, there are CTCs that contain tumor-derived mutations reflecting genetic alterations in androgen receptor, TMPSSR and c-Myc locus. In lung cancer, size-selected CTCs were shown to bear the ALK translocation in agreement with tissue biopsy analysis. A molecular analysis of CTCs recovered by the Veridex CellSearch platform described an EGFR mutation within the EpCAM population in a fraction of preselected patients. Using a CTC chip, Mareswaran et al. showed EGFR mutations in CTCs. Other models revealed mutation-bearing CTC subpopulations within the CTC pool include colorectal cancer and malignant melanoma.

**Molecular relationship between CTC populations and populations recovered from tissue biopsy**

Experiments to demonstrate concordant mutations between CTC and biopsy samples would address the economic advantage of fitting CTC sampling into a clinical practice built around FFPE samples. These experiments would also expand our understanding of the biology of the CTC population. Rare cell CTC collection occurs from the blood phase whereas tissue biopsy collects populations of spatially related cells. This difference has significant impact on the stochastic relationship of the two sampling mechanisms. Ignatiadis et al. described a subpopulation of Her2+ CTCs regardless of the Her2 status of the primary breast tumor. Indeed, 61% of CTC samples included Her2+ cells despite only 20% of the primary tumors being Her2+. This difference is the subject of the ongoing DETECT-III clinical trial. Expression analysis of single or pooled CTCs in breast cancer has confirmed a striking variability within the CTC population. Single-cell analysis shows that CTCs do not cluster by disease stage, clearly distinguishing this type of sample from pooled tumor tissue. Additionally, treatment regimes impose selective pressures on the tumor allowing for acquisition, or selection and expansion, of different mutations in the metastatic and CTC population thereby exacerbating discordance.

Consistent with these observations, CTCs in breast cancer patients show 40-50% heterogeneity in ER and PR expression patterns of CTCs, regardless of the primary tumor’s status. This CTC population remains mixed both before and after treatment.

In prostate cancer, there is good evidence that there is a population of CTCs that are malignant in origin. These cells reflect genetic alterations in androgen receptor, TMPSSR and c-Myc locus that often reflect similar changes in portions of the primary tumor.

In lung cancer, size-selected CTCs were shown to bear the ALK translocation in agreement with tissue biopsy analysis. Marked heterogeneity was observed in an EpCAM-selected population, leading to significant discordance between the CTCs and primary tumor. Similarly, PCR analysis of CTCs recovered by CellSearch described a single EGFR mutation found in 38 patient samples despite eight patients having archival tissue with confirmed EGFR mutation. In another example, Mareswaran et al. described a 92% concurrence between EGFR mutations found in CTC and matched tumor tissue. Accommodating these data in light of studies that document a 50-65% EGFR discordance within primary and metastatic biopsy suggest the Mareswaran study may reflect a non-random patient group, a bias introduced by the selection or detection.
method, analysis of private mutations (discussed below), or a biological mobility associated with EGFR mutant CTCs.

**Tumor heterogeneity and single-cell sampling**

In a series of recent sequence comparisons between primary, metastatic and disseminated tumor cells a striking disparity was observed. This disparity was informed to some degree by the disease model, but clearly observed in all forms of cancer. The growing awareness of these disparities has certainly been accelerated by increased access to next generation sequencing tools which have greatly expanded the number of targets known to be mutated in different models of cancer. As a snapshot of all mobile cells in a disease, CTCs must reflect this diversity. In fact, among studies that have assessed evaluable molecular markers in CTC such as Her2 amplification, EGFR mutation, or KIT mutation, there is a 40-87% discordance between the detection of mutations in CTC versus tissue biopsy. These data reflect the growing understanding of the heterogeneity within a tumor as well as between primary and metastatic sites. An analysis of primary and metastatic breast cancer samples showed significantly higher genetic complexity in the primary tissue samples than the subclones that escaped to seed metastatic sites. Critically, CTC may be derived from all primary and secondary sites and may therefore be sampled using a LiquidBiopsy. Conversely, private mutations found in the CTC population may be present in the primary and metastatic tissue but at subclonal levels. Conspicuous EGFR amplification heterogeneity within a tumor sample in NSCLC predicts exactly the EGFR heterogeneity observed in the CTC. Even with the prevalence of CTC in the prostate cancer model, there remains significant discordance of TMPSSR:ERG translocation in ctDNA and tissue biopsy samples. Evaluating ploidy in single cells recovered from tumor tissue reveals a genetically diverse population, reiterating the impact of sampling method.

**Sampling by next-generation sequencing**

Next-generation sequencing has significantly changed how we understand tumor genetics. These massively parallel sequencing platforms allow the analysis of hundreds of different loci with sensitivities that can approach 1% mutant frequency. Thus, mutation detection within a mixed population recovered from a blood sample is reduced to a statistical challenge. For rare cell analysis, there are two confounding scenarios: Evaluation of very rare events will not always successfully identify a mutation due to stochastic sampling limitations both during recovery of target cells and the selection of individual molecules during library construction. Secondly, multiple independent mutations, unless located in a single hotspot, will be subject to independent sampling during library construction that will mask their single-cell origin. These combined affects drive the need to extract mutation bearing tumor cells with purities in excess of 1% and elaboration of biomarkers to capture greater numbers of mutation bearing cells from blood.

A recent analysis of paired genome sequences in lung cancer suggests lung cancer is a model with much greater mutational complexity than observed in tumor cell lines. Lung cancer shows multiple redundant mutations targeting a pathway rather than exclusively single driver mutations. Thus, differences between primary, metastatic and disseminated tumor cells, as well as the emerging mutational complexity within any one tumor site, predicts that a large number of different clones may end up in circulation. Also, sequencing tests will uncover mutations, such as JAK2 V617F in NSCLC, that are present in multiple patients at low allele frequency. The studies referenced above in breast, prostate and lung cancer clearly present examples of ctDNA sampling of mutations that are predominantly divergent from tumor samples. These differences reflect the subclones present in single sites, the differences between primary and metastatic sites, and asynchrony of samples and selective sampling using a target such as EpCAM.
Biomarker Selection

CTCs were originally defined by enumeration of EpCAM selected cells\(^1\). EpCAM is expressed in most solid tumor tissue\(^67\) but its role in metastasis is unclear. Emerging data suggests metastasis is driven in part by an epithelial to mesenchymal transition (EMT); a reversible process by necessity, alternating stationary epithelial cells into motile mesenchymal cells\(^68\) that may also affect treatment decisions\(^69\). This is considered as evidence of epithelial plasticity in CTCs which co-express EMT markers and EpCAM\(^16,70,71\). This plasticity drives the quest for expansion of the biomarker selection targets used for CTC capture\(^16,71\). Thus, while EpCAM is the legacy marker to develop a molecular CTC test, ultimate marker optimization must be performed in the context of a specific disease model and molecular readout.

Overall, sampling tumor cells from blood demands a system specifically designed to allow molecular analysis of <10 cells of any particular clone. It demands a protocol that is reproducible, has minimal carry-over of non-target cells and supports molecular analysis of subpopulations of cells. Most importantly, with the tools available to date, there are no platforms that can robustly access the mutational profile of ctcDNA. However, Cynvenio’s LiquidBiopsy platform has the ability to produce quality sequence information from this population. This is based on flexible, robust, high efficiency capture of target cell populations, low non-specific cell carryover, a stabilization protocol that supports recovery of genomic DNA, and a validated protocol and analytic solution to NGS analysis of this sample, all from whole blood.

Moving Forward with CTC

Circulating biomarkers are increasingly recognized as valuable tools for the real time monitoring of individual patients, especially in this era of personalized medicine and with the emergence of more targeted therapeutics\(^73,74\). Using blood as an analyte clearly reduces the threshold for obtaining tumor derived sample. Additionally, it samples all the tumor sites, a necessary step as oncologists develop combination therapies that target the disease heterogeneities responsible for resistance and recurrence\(^75\). Finally, developing protocols to molecularly analyze the circulating tumor cell population will bridge the gap between research based single cell analysis and high throughput sampling for clinical tests. The applications for this type of test range from selection of targeted therapeutics, through resistance and recurrence monitoring. Fundamentally, the ctcDNA has the ability to serve as an unassailable monitoring tool that can detect, define, monitor, and ultimately screen for cancer based on the presence of genetic alterations that may always be associated with the presence of disease.
Citations


