

REVIEW

# Progress and challenges of sequencing and analyzing circulating tumor cells

Zhongyi Zhu · Si Qiu · Kang Shao · Yong Hou

Received: 13 August 2017 / Accepted: 29 October 2017 / Published online: 22 November 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Circulating tumor cells (CTCs) slough off primary tumor tissues and are swept away by the circulatory system. These CTCs can remain in circulation or colonize new sites, forming metastatic clones in distant organs. Recently, CTC analyses have been successfully used as effective clinical tools to monitor tumor progression and prognosis. With advances in next-generation sequencing (NGS) and single-cell sequencing (SCS) technologies, scientists can obtain the complete genome of a CTC and compare it with corresponding primary and metastatic tumors. CTC sequencing has been successfully applied to monitor genomic variations in metastatic and recurrent tumors, infer tumor evolution during treatment, and examine gene expression as well as the mechanism of the epithelial-mesenchymal transition. However, compared with cancer biopsy sequencing and circulating tumor DNA sequencing, the sequencing of CTC genomes and transcriptomes is more complex and technically difficult. Challenges include enriching pure tumor cells from a background of white blood cells, isolating and collecting cells without damaging or losing DNA and RNA, obtaining unbiased and even whole-genome and transcriptome amplification material, and accurately analyzing CTC sequencing data. Here, we review and summarize recent studies using

NGS on CTCs. We mainly focus on CTC genome and transcriptome sequencing and the biological and potential clinical applications of these methodologies. Finally, we discuss challenges and future perspectives of CTC sequencing.

**Keywords** Circulating tumor cell · Next-generation sequencing · Single-cell sequencing

## Introduction

Circulating tumor cell (CTC) studies began in 1869 (Ashworth 1869) and now are rapidly increasing, with a steadily expanding list of CTC-related studies retrievable from NCBI PubMed. An average of over 1000 CTC-related papers were published per year in the last 5 years. During cancer progression, cancer cells in the primary tumor may invade into nearby blood vessels via the epithelial to mesenchymal transition (EMT) process and then circulate throughout the entire blood system. After traveling some distance, CTCs may leave the blood vessels via the mesenchymal to epithelial transition (MET) process, the reverse process of EMT (Nieto 2013), which helps cancer cells seed in different tissues and generate metastatic lesions. In particular, two recent studies reported evidence that EMT is involved in the metastatic potential of CTCs by detecting EMT markers in human CTCs (Armstrong et al. 2011; Kallergi et al. 2011).

Previous CTC studies have mainly focused on the development of CTC enrichment technology (Hong and Zu 2013; Alix-Panabieres and Pantel 2014; Krebs et al.

Z. Zhu · S. Qiu · K. Shao · Y. Hou  
BGI-Shenzhen, Shenzhen 518083, China

Z. Zhu · S. Qiu · K. Shao · Y. Hou (✉)  
China National GeneBank, BGI-Shenzhen, Shenzhen 518120,  
China  
e-mail: houyong@genomics.cn

2014; van der Toom et al. 2016), the correlation between CTC number and clinical prognosis (Arya et al. 2013; Harouaka et al. 2014; Alix-Panabieres and Pantel 2016), and biological processes driving cancer metastasis and recurrence (Hodgkinson et al. 2014; Massague and Obenauf 2016). For example, scientists have discovered a correlation between CTC number and the prognosis of breast cancer (Eroglu et al. 2013; Bidard et al. 2016), colon cancer (Hardingham et al. 2015), and prostate cancer (Hu et al. 2013), whereby the CTC number can be used to assess cancer prognosis. Based on these principles, the US Food and Drug Administration (FDA) approved the CellSearch™ system as an aid for monitoring breast cancer (Cristofanilli NEJM 2004), prostate cancer (Cohen JCO 2008), and colon cancer (Scher Lancet Oncol 2009) progression.

However, until recently, few CTC sequencing studies have been published. We searched the CTC studies based on next-generation sequencing (NGS) technology in NCBI PUBMED and found 19 CTC publications (Table 1). Among these publications, over half of the CTC DNA sequencing studies focused on capturing genes with potential as targeted cancer therapies or those

involved in metastasis or recurrence, and a limited number of studies used whole-genome sequencing (WGS) or whole-exome sequencing (WES). Additionally, several studies have highlighted CTC transcriptome sequencing as a promising approach for investigating metastasis mechanisms such as EMT. In this review, we summarized and described the methodologies and pipeline of NGS of CTCs, including cell enrichment, isolation and capturing methods, and whole-genome or transcriptome amplification methods. Furthermore, we focused on the biological insights achieved from CTC sequencing as well as its potential clinical applications. Finally, we highlighted major challenges for CTC sequencing and bioinformatics analysis and discussed the future perspectives for CTC sequencing in the NGS area.

### Methodology of CTC genome and transcriptome sequencing

Generally, the CTC sequencing workflow can be separated into four steps: CTC enrichment, CTC isolation (particularly single-cell CTC or pure CTC isolation),

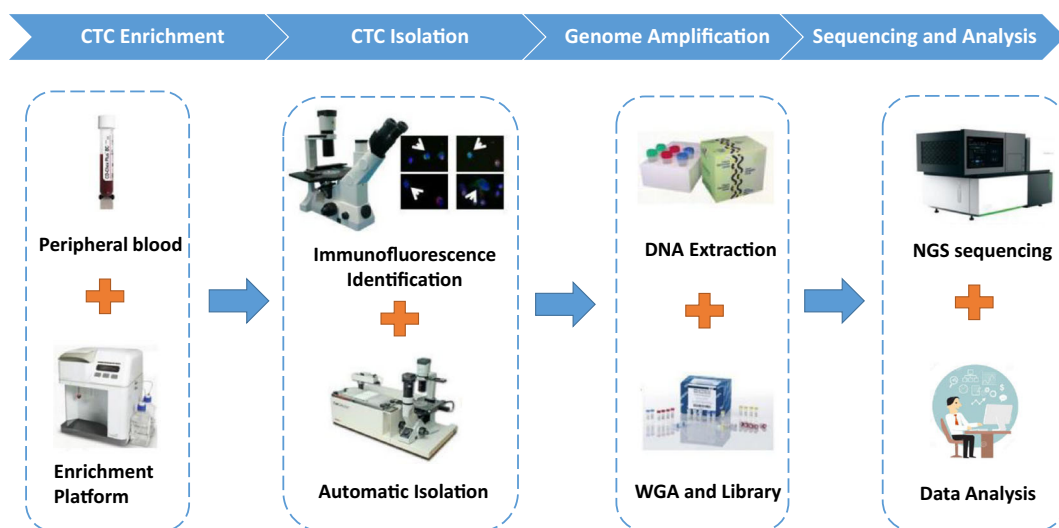
**Table 1** Overview of next-generation sequencing CTC studies

Sequencing strategies	Cancer type	Ref
Target (68 genes)	Stage IV colorectal carcinoma	(Heitzer et al. 2013)
WES	Prostate cancer	(Zhao et al. 2013)
WGS/WES	Lung cancer	(Ni et al. 2013)
WES	Prostate cancer	(Lohr et al. 2014)
WGS	Prostate cancer	(Dago et al. 2014)
Target (46 genes)	Hepatocellular carcinoma	(Kelley et al. 2015)
WGS/target (20 genes)	Melanoma/Small-cell lung cancer	(Rothwell et al. 2016)
WGS/WES	Prostate cancer	(Jiang et al. 2015)
Target (50 genes)	Metastatic breast cancer	(De Luca et al. 2016)
Target (50 genes)	Metastatic breast cancer	(Shaw et al. 2017)
Target (50 genes)	Melanoma	(Palmirotta et al. 2017)
Target (6 genes)	Liver, colorectal, lung/Gastric, breast, prostate cancer	(Wong et al. 2017)
WGS/WES	Breast, gastric, prostate, colon cancer	(Gao et al. 2017)
SC RNA-Seq	Pancreatic cancer	(Yu et al. 2012)
SC RNA-Seq	LNCaP, Prostate cancer	(Cann et al. 2012)
SC RNA-Seq	Breast cancer	(Yu et al. 2013)
SC RNA-Seq	KPC mice, pancreatic/Breast, prostate cancer	(Ting et al. 2014)
SC RNA-Seq	Prostate cancer	(Miyamoto et al. 2015)
SC RNA-Seq	Colorectal cancer cell line	(Grillet et al. 2016)

genome or transcriptome amplification, and sequencing and analysis (Fig. 1). Several successful CTC enrichment methods have been reported for enriching CTCs from cancer patient blood. Generally, two strategies are used for these methods. The most common method is to utilize cell surface CTC markers for enrichment (EPCAM<sup>+</sup>, CK<sup>+</sup>, CD44<sup>+</sup>) or to delete immune cells (CD45<sup>-</sup>) (Alix-Panabieres and Pantel 2014; Krebs et al. 2014; Ferreira et al. 2016), and representative platforms include CellSearch (Riethdorf et al. 2007), MagSweeper (Talasaz et al. 2009; Deng et al. 2014), and GILUPI cell collector (Saucedo-Zeni et al. 2012), among others. The other enrichment strategy uses the physical characteristics of CTCs (size, density, acoustics, fluid force) to separate CTCs from the leukocyte background (Harouaka et al. 2013; Krebs et al. 2014), and representative platforms include ClearCell (Hou et al. 2013; Khoo et al. 2015), ISET (Vona et al. 2004; Chinen et al. 2013). Recently, microfluidic channels and waves integrated with cell surface markers or physical characteristics, such as IsoFlux (Harb et al. 2013), are widely used in CTC enrichment systems and significantly improve the efficiency and accuracy of CTC identification (Li et al. 2015; Shields et al. 2015). In addition, systems such as CTC-iChip (Ozkumur et al. 2013; Karabacak et al. 2014) that combine cell surface markers and physical characteristics exhibit promising performance, recovering more CTCs with less DNA or RNA damage. Most of these methods require 7.5 ml (or more) of peripheral blood for effective enrichment,

whereas the GILUPI cell collector overcomes the barrier of small blood samples by collecting CTCs in vivo from the peripheral blood stream.

After CTC enrichment from the blood, zero to several hundred CTCs may be retained in a thousand to ten thousand background cells, resulting in a low efficiency for sequencing and analysis of these cell pools (Alix-Panabieres and Pantel 2014). Therefore, scientists typically use the tools commonly used for SCS, such as laser capture microdissection (LCM) and flow cytometry analysis (FACS), to further isolate CTCs from background cell pools. As mentioned in previous SCS reviews, different methods have distinct pros and cons (Macaulay and Voet 2014; Wang and Navin 2015; Chen et al. 2016; Gawad et al. 2016). Compared with LCM, FACS automatically isolates specific individual cells with correct markers into tubes or wells in a high-throughput manner, whereas an LCM approach allows the observation of cellular morphology and physiology to prevent possible contamination or cell damage, although this approach is time-consuming and labor intensive. Recently, an increasing number of automated systems based on the microfluidics approach have been developed to enrich and isolate CTCs. For example, the DEPArray system from Silicon Biosystems utilizes force from non-uniform electric fields to drive, capture, and recover cells, avoiding physical contact of the cells with other substrates. Thus, cells isolated from this system are biologically healthy and retain complete genetic information, which is used for sequencing and



**Fig. 1** The workflow of circulating tumor cell sequencing

gene expression profiling (Gascoyne et al. 2009; Carpenter et al. 2014). In addition, scientists built in-house pipelines or systems to isolate CTCs for their specific study design. In Lohr's work, the enriched CTCs were separated into micro-well chips, and subsequently specific CTCs enriched with fluorescence markers were selected for further sequencing and analysis (Lohr et al. 2014). Other studies using a combination of different enrichment methods including (Zhao et al. 2013) marker-free microfluidic isolation, direct sequencing (Palmirota et al. 2017), and automatic CTC counting (Wong et al. 2017) also demonstrated that the development of CTC enrichment and isolation technology significantly improved the efficacy and efficiency of CTC sequencing studies.

After obtaining target cells from whole blood, the genetic material should be amplified to generate enough template to create an NGS library. Methods for amplifying the whole genome or transcriptome have been summarized in previous SCS reviews (Macaulay and Voet 2014; Wang and Navin 2015; Gawad et al. 2016). Briefly, for whole-genome amplification (WGA), scientists use linear or PCR-based amplification methods such as MDA (Spits et al. 2006), MALBAC (Lu et al. 2012), and DOP-PCR (Cheung and Nelson 1996), which are commonly used in SCS studies (Hou et al. 2015; Huang et al. 2015). For whole-transcriptome amplification (WTA), methods such as CEL-seq (Hashimshony et al. 2012), STRT-seq (Islam et al. 2014), and SMART-seq (Picelli et al. 2013) were developed to amplify either full-length transcripts or their 3' region. Most of these amplification methods have been commercialized as kits such as the GenomePlex Single Cell WGA Kit, the Qiagen REPLI-g Single Cell Kit, and the SMARTer Ultra Low Input RNA Kit. After obtaining enough genetic material, the NGS library is prepared and sequenced with standardized protocols. After obtaining sequencing data from CTCs, the most important procedure is to evaluate bias during sample preparation and design an appropriate statistical model to handle these biases (described in the challenges section).

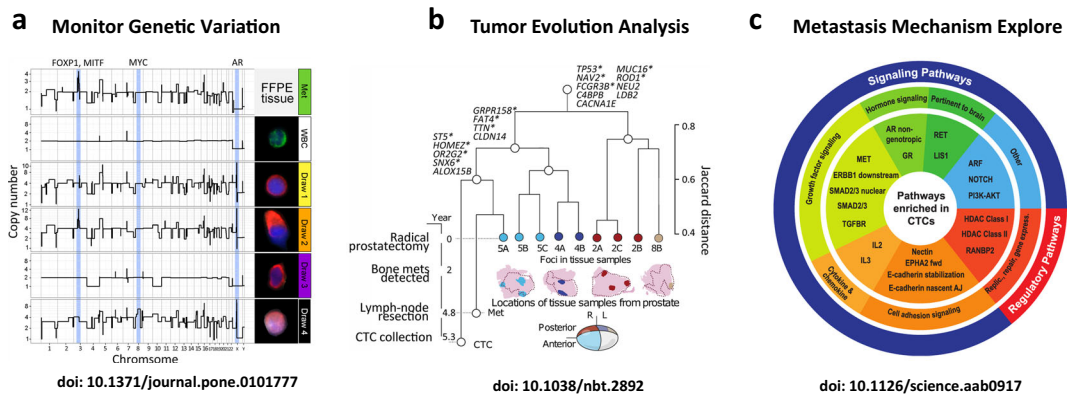
According to previous publications (Ni et al. 2013; Ting et al. 2014; Miyamoto et al. 2015) and our experience, the successful rate of overall amplification and library preparation is under 60% because of multiple sample handling and staining processes. To improve the success rate, some scientists pool isolated CTCs together or directly pool all of the recovered cells (including white blood cells (WBCs)) (Palmirota et al. 2017; Shaw et al. 2017; Wong et al. 2017). The advantage of this strategy is

that it has a high library preparation success rate, although it introduces increased noise into subsequent bioinformatics analyses. Other groups demonstrated that only amplifying specific regions of the genome with multiplex PCR reduces the complexity of the experiment and increases the success rate (Palmirota et al. 2017).

### Monitoring clinically relevant genetic variations during cancer progression

Cancer metastasis and recurrence are major challenges to clinical treatment and the major causes of death in cancer patients. Compared with the primary tumor, genome sequencing studies have shown that cancer cells from metastatic and recurrent tumors acquire novel somatic variations that enhance cell progression during treatment (Mwenifumbo and Marra 2013). In clinical practice, it is usually difficult to acquire a re-biopsy from metastatic or recurrent tumors, leading to ambiguous diagnostic results during treatment. Liquid biopsy recently emerges as a significant breakthrough in cancer translational research. By sequencing the circulating tumor DNA (ctDNA) from the CTCs, researchers observed the somatic variation in the landscapes of the tumor without biopsy sequencing (Crowley et al. 2013). Scientists have demonstrated that somatic variations detected in ctDNA or CTCs highly correlate with primary, metastatic, or recurrent tumors and could be used for clinical diagnosis and disease monitoring (Alix-Panabieres and Pantel 2016). Although ctDNA is much easier to obtain, CTCs contain complete genetic information, including the genome, transcriptome, and even epigenome of the circulating cells, which provides more comprehensive genetic information for scientific investigations.

Previous CTC studies have highlighted the biological involvement of CTCs in cancer metastasis and recurrence (Hou et al. 2011; Franken et al. 2012; Massague and Obenauf 2016). As liquid biopsy tools, CTC sequencing could serve as an efficient and unique tool to monitor cancer progression and to discover somatic mutations with possible clinical relevance that occurred or disappeared pre- and post-treatment (Fig. 2a). FD Luca et al. reported that most of the CTC baseline mutations were eliminated, and novel mutations emerged after chemotherapy treatment in one metastatic breast cancer patient (De Luca et al. 2016). However, a common mutation, p.V777L in the *ERBB2* gene, was



**Fig. 2** Biological and potential clinical applications of CTC sequencing. **a** The CNV pattern of CTCs in a prostate cancer patient changed under therapeutic pressure. **b** CTC sequencing served as

an efficient tool to uncover biological insights concerning tumor evolution. **c** RNA sequencing of CTCs enabled the exploration of metastasis-related pathways

detected in all of the post-therapy CTCs, indicating that the clone bearing this mutation might have played a role in the resistance to the administered therapy. Xiaohui Ni et al. reported the clinical relevance of *PIK3CA*, *RBI1*, and *TP53* mutations in lung cancer treatment, such as erlotinib drug resistance and two vivid examples of the potential clinical uses for CTC sequencing during disease monitoring (Ni et al. 2013). After CTC sequencing, these authors identified a patient carrying the *PIK3CA* mutation, which has been associated with drug resistance to erlotinib and chemotherapy strategy selection. In addition, Miyamoto et al. analyzed the RNA-Seq profiles of 77 intact CTCs isolated from 13 prostate cancer patients and showed that single CTCs exhibit high heterogeneity for AR gene mutations and splicing variants, demonstrating that signaling pathway heterogeneity might be responsible for treatment failure (Miyamoto et al. 2015).

**Inferring tumor heterogeneity and evolution dynamically**

Typically, cancer is considered the result of Darwinian evolution, as cancer continually acquires new somatic mutations in single cells, followed by selection, which enhances the fitness and growth advantage of a specific group of malignant cells (Stratton et al. 2009). Understanding intra-tumoral heterogeneity (ITH) and evolution is essential for the early detection of disease recurrence and the efficient treatment of cancer (Lipinski et al. 2016). There are three major hypothetical models explaining ITH, including clonal evolution, the cancer

stem cell, and mutator phenotype models (Russnes et al. 2011). Recently, the development of deep sequencing and SCS make it possible to investigate phylogenetic evolution during cancer progression (McGranahan and Swanton 2017) (Fig. 2b). For example, Navin et al. decoded the evolutionary pattern of breast cancer and related metastatic lesions using single-cell genome sequencing (Navin et al. 2011; Wang et al. 2014; Gao et al. 2016). Hou et al. applied single-cell exome sequencing to blood neoplasm and kidney cancer (Hou et al. 2012; Xu et al. 2012). Macro et al. used multiregion deep sequencing to examine the evolutionary pattern of kidney tumors (Gerlinger et al. 2012). However, due to a large amount of heterogeneity in solid tumors and the difficulty of re-biopsy, the evolution of cancer metastasis and recurrence remains elusive. Compared to only sequencing primary and metastasis tumor cells, sequencing CTCs provides additional data to further explore ITH complexity. In addition, because CTC survival in peripheral blood is an essential step for tumor metastasis, sequencing CTC genomes can delineate a more detailed tumor evolution process, facilitating the understanding of the tumor metastasis mechanism.

Many CTC sequencing studies have highlighted the genetic heterogeneity of CTCs, which further increases the complexity of CTC research in cancer biology. For example, Luca et al. sequenced a 50-gene NGS panel on CTCs isolated from four metastatic breast cancer patients and reported a 50% (20 of 50 genes) CTC variation detection rate in a mean depth of 1500X of 14 CTCs. These authors concluded that there exists a high inter- and intra-patient heterogeneity in CTC mutational status (De Luca et al. 2016). Constructing the tumor



evolution process through the detected CTC somatic nucleotide variants (SNVs) and mapping the early trunk mutations (mutation present early in tumor evolution) have considerable clinical utility. For example, Lohr et al. reported that CTCs shared 90% primary trunk mutations and 70% metastasis trunk mutations in a prostate cancer patient, respectively. However, due to a limited captured number of CTCs in most cancer types, it is difficult to analyze the SNV evolutionary structure in individual patients.

Moreover, copy number variation (CNV) is also frequently altered during cancer evolution. Ni et al. surveyed the CTC CNVs from a small-cell lung cancer (SCLC) patient during sequential chemotherapy treatment and observed that the evolution of CNV was consistent along the therapeutic stage, indicating that the reproducible CNV pattern was not affected by drug treatment (Ni et al. 2013). Dago and colleagues collected blood samples at multiple time points during the treatment of metastatic prostate cancer patients and found that the initial CNV evolution in CTCs changed along with a clinical response feature (decreased pain and PSA level), and CNV evolution changed again with a clinical observation of disease progress (increasing pain and PSA levels). Thus, these authors concluded that the CNV evolution of CTC would be affected by therapeutic pressure in prostate cancer (Dago et al. 2014). Gao et al. analyzed the CNV of CTCs across 23 patients and concluded that the CNV tumor evolution process follows a convergent evolution model through primary tumor to CTCs but does not follow the classical gradual acquisition mode or the recent alert punctuated model (Gao et al. 2017). However, more patients showed different variation spectrums during evolution, most of which still cannot be comprehensively explained. Further investigations into the biological impact of these complex somatic mutations during cancer treatment will significantly contribute to our understanding of CTC biology in cancer progression and future clinical application in areas of disease monitoring.

### Understanding altered molecular pathways during tumor progression

Previous studies have highlighted particular molecular pathways involved with cancer metastasis, such as TGF- $\beta$  signaling (Akhurst and Derynck 2001), Wnt signaling (Polakis 2012), and EMT (Gonzalez and

Medici 2014). However, most of these studies were based on mouse models or specific biomarkers (Thiery et al. 2009; Li et al. 2010; Harper et al. 2016). Analyzing CTC gene expression provides a unique window into understand the molecular pathways altered during metastasis despite the existence of significant ITH. The emergence of single-cell RNA sequencing (scRNA-seq) technology also enabled the acquisition of comprehensive gene expression and splicing information using a limited number of isolated CTCs (Ramskold et al. 2012). Thus, CTC transcriptome sequencing provides a unique window to digitize molecular pathways during cancer progression (Fig. 2c).

Yu et al. first utilized single-molecule RNA sequencing and a mouse pancreatic cancer model and found that the *WNT2* gene mediated the metastasis-associated survival signal, consistent with observations of the upregulation of multiple Wnt genes in pancreatic patients (Yu et al. 2012). This report is the first comprehensive study utilizing CTC RNA-Seq to uncover complete molecular pathways altered during cancer metastasis. This study was further expanded by analyzing the scRNA-seq of CTCs and comparing the results with matched primary tumors in a pancreatic cancer model (Ting et al. 2014). The authors observed that extracellular matrix genes are highly expressed in mouse and human CTCs and SPARC (an extracellular matrix protein), which may contribute to pancreatic tumor metastasis. In addition, Miyamoto et al. presented another study of 13 drug-resistant prostate cancer patients utilizing the scRNA-seq of CTCs and found that drug resistance in prostate cancer was triggered by the activation of the non-canonical Wnt signaling pathway (Miyamoto et al. 2015).

The EMT of adherent epithelial cells to a migratory mesenchymal state has been implicated in tumor metastasis in pre-clinical models. Yu et al. characterized the dynamic cell fates in breast cancer CTCs and found an association of mesenchymal CTCs with disease progression (Yu et al. 2013). By directly sequencing the RNA of CTC-enriched cell populations from a metastatic breast cancer patient at five serial time points and comparing the samples to ten healthy donors, these authors identified 170 CTC transcripts at a mesenchymal-predominant time point, which showed dramatic enrichment for EMT-related expression changes and extracellular matrix (ECM) and ECM-related membrane receptors. Another interesting observation in this study is that both single CTCs and multicellular cluster CTCs express known EMT regulators, including TGF- $\beta$  pathway components.

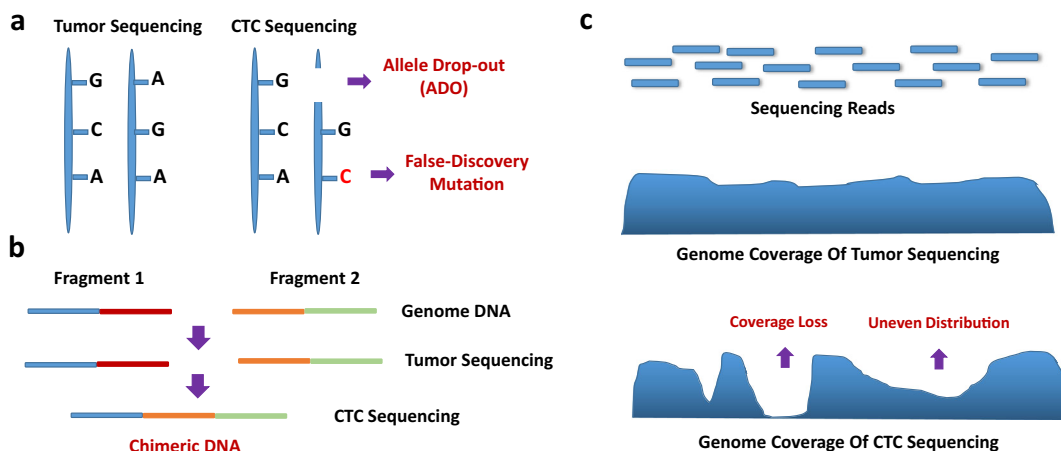
These authors further showed that CTC clusters originated monoclonally from the primary tumor and showed markedly increased metastatic capability compared with single CTCs, which had a poor prognosis in a mouse model (Aceto et al. 2014). Combined with the scRNA-seq of CTC clusters and single CTCs, these authors also demonstrated that plakoglobin was implicated in cluster formation during breast cancer metastasis. This study not only demonstrated molecular signaling in CTCs but also provided a putative novel drug target to control breast cancer metastasis. Another study from Grillet et al. established three CTC lines from three chemotherapy-naïve advance metastatic colorectal cancer patients and demonstrated the enrichment of drug metabolism pathways, which corresponds to cytotoxic compound resistance using RNA-Seq of CTCs and primary tumors (Grillet et al. 2016). Taken together, these findings suggest potential markers of tumor progression and treatment response and indicate the great potential of taking these observations to the clinic.

### Challenges and future perspectives of CTC sequencing

As shown in this review, CTC sequencing can now be used as an efficient liquid biopsy tool to investigate the spectrum of somatic variation and gene expression changes in primary, metastatic, and recurrent cancer

patient tumors non-invasively. The somatic alterations could either be used to understand the origin, ITH, and evolution of tumors or to monitor disease progression during cancer treatment. The most important clinical implication of CTC sequencing is for personalized medicine, or so-called precision medicine, according to the variation spectrum detected, which indicates the selection of target therapy based on the CTC variation spectrum. With the development of CTC capture and SCS technologies, we can expect the establishment of more comprehensive cancer origins and evolution models in the near future. Moreover, more novel biomarkers or potential drug targets for cancer metastasis or drug resistance prevention or treatment can be identified through CTC sequencing.

However, as mentioned above, sequencing a CTC genome or transcriptome faces technical challenges. Obtaining enough cells for library preparation and sequencing is the first critical step in CTC sequencing. However, various conclusions can be drawn from different studies based on different isolation method and cancer types. Some cancer types tend to generate more CTCs than other cancer types (Allard et al. 2004), and the clinical stage is also associated with the CTC number collected from patients. Although there are several controversial conclusions, it is common that patients in late stages of cancer or with metastatic lesions contain more CTCs (Pantel et al. 2009). The number of CTCs typically varies from zero to several hundred (even thousand in



**Fig. 3** CTC sequencing biases introduced during amplification and library preparation. **a** Allele drop out (ADO) during genome amplification leads to the loss of the detection of somatic mutant alleles in CTCs, and false-positive results can also be introduced into both WGA and library processes. **b** Chimeras will lead to

artificial CNV and SV detection in CTC sequencing. **c** The limitations of the WGA method may lead to low genome coverage or uneven read distribution effects when compared to bulk tumor sequencing

some cases) per 7.5 ml of blood. In general, obtaining enough CTCs for sequencing remains a significant problem for most cancer types, which limits the number of CTC sequencing studies. In addition, cell loss or genetic material damage during CTC enrichment, isolation, genome, or transcriptome amplification has consistently been reported in studies based on various CTC enrichment systems. Further, leukocyte contamination, lack of specific biomarkers, low-throughput and time-consuming manual capture operation protocol also hinder the progress of CTC sequencing studies (Cann et al. 2012). Obtaining high-quality sequencing libraries is another critical step in CTC sequencing. Based on the description of Miyamoto et al. (Miyamoto et al. 2015), 77 out of 221 cells (35%) were defined as qualified cells for bioinformatics analysis. Many other studies, including Lohr et al. or Dago et al. (Dago et al. 2014; Lohr et al. 2014) and our own experience, have also shown the same success rate for CTC sequencing.

More importantly, bioinformatics analyses of CTC sequencing data require additional quality evaluation and assessment, particularly the biases introduced during sample and sequencing library preparation (Fig. 3). Allele drop out (ADO) during genome amplification may prevent the detection of the somatic mutant alleles of CTCs, which may contribute to cancer progression or drug resistance (Hou et al. 2012; Nawy 2014). In addition, the limitation of the WGA method may lead to low genome coverage (Kelley et al. 2015), high false-positive rates, and low sensitivity of mutation detection (Lohr et al. 2014; Shaw et al. 2017). Uneven reads distribution and chimeras from WGA may also lead to artifacts in the CNV and SV detection of CTCs (Ni et al. 2013; Lohr et al. 2014; Jiang et al. 2015). However, as reviewed in SCS studies, scientists can design specific statistical models, such as kindred replication (Chen et al. 2017; Dong et al. 2017), the Bayesian model (Kharchenko et al. 2014), and the binomial model (Vu et al. 2016), to reduce these biases and more accurately detect SNVs, CNVs, and SVs.

Despite the development of advanced microfluidics approaches, several novel sequencing technologies show promise for solving the technological hurdles to CTC sequencing. For example, scientists can now perform in situ DNA or RNA sequencing, even on samples fixed on slides (Lee et al. 2015). This approach decreases the complexity of sample handling processes and the cell damage or loss that can occur during CTC enrichment and isolation. In addition, emerging single-

molecule sequencing technologies show promise for analyzing DNA or RNA molecules without amplification. Thus, biases or artifacts, such as ADO, false-positive mutations, and uneven amplification, may be significantly reduced during CTC sequencing. Furthermore, single-molecule sequencing (Liu and Wu 2011; Gawad et al. 2016; Heather and Chain 2016) may expand the CTC sequencing approach to analyze wide epigenome information, such as methylation and chromatin occupation.

**Acknowledgments** The authors would like to thank Dr. Kun Ma for reviewing and revising the manuscript.

**Funding information** This work was financially supported by grants from the Shenzhen Municipal Government of China (NO. JSGG20140702161347218) and the National Natural Science Foundation of China (NO. 81672593 and NO. U1601224).

**Compliance with ethical standards**

**Competing interests** The authors declare that they have no competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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.

## References

- Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*. 2014;158(5):1110–22.
- Akhurst RJ, Derynck R. TGF-beta signaling in cancer—a double-edged sword. *Trends Cell Biol*. 2001;11(11):S44–51.
- Alix-Panabieres C, Pantel K. Challenges in circulating tumour cell research. *Nat Rev Cancer*. 2014;14(9):623–31.
- Alix-Panabieres C, Pantel K. Clinical applications of circulating tumor cells and circulating tumor DNA as liquid biopsy. *Cancer Discov*. 2016;6(5):479–91.
- Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, 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*. 2004;10(20):6897–904.
- Armstrong AJ, Marengo MS, Oltean S, Kemeny G, Bitting RL, Turnbull JD, et al. Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial



- and mesenchymal markers. *Mol Cancer Res*. 2011;9(8):997–1007.
- Arya SK, Lim B, Rahman AR. Enrichment, detection and clinical significance of circulating tumor cells. *Lab Chip*. 2013;13(11):1995–2027.
- Ashworth TR. A case of cancer in which cells similar to those in the tumours were seen in the blood after death. *Australasian Medical Journal*. 1869;14:146–7.
- Bidard FC, Proudhon C, Pierga JY. Circulating tumor cells in breast cancer. *Mol Oncol*. 2016;10(3):418–30.
- Cann GM, Gulzar ZG, Cooper S, Li R, Luo S, Tat 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*. 2012;7(11):e49144.
- Carpenter EL, Rader J, Ruden J, Rappaport EF, Hunter KN, Hallberg PL, et al. Dielectrophoretic capture and genetic analysis of single neuroblastoma tumor cells. *Front Oncol*. 2014;4:201.
- Chen X, Love JC, Navin NE, Pachter L, Stubbington MJ, Svensson V, et al. Single-cell analysis at the threshold. *Nat Biotechnol*. 2016;34(11):1111–8.
- Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, et al. Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). *Science*. 2017;356(6334):189–94.
- Cheung VG, Nelson SF. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc Natl Acad Sci U S A*. 1996;93(25):14676–9.
- Chinen LT, de Carvalho FM, Rocha BM, Aguiar CM, Abdallah EA, Campanha D, et al. Cytokeratin-based CTC counting unrelated to clinical follow up. *J Thorac Dis*. 2013;5(5):593–9.
- Cohen SJ, Punt CJ, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *Journal of clinical oncology*. 2008;26(19):3213–21.
- Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med*. 2004;351(7):781–91.
- Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*. 2013;10(8):472–84.
- Dago AE, Stepansky A, Carlsson A, Lutgen M, Kendall J, Baslan T, et al. Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. *PLoS One*. 2014;9(8):e101777.
- De Luca F, Rotunno G, Salvianti F, Galardi F, Pestrin M, Gabellini S, et al. Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. *Oncotarget*. 2016;7(18):26107–19.
- Deng G, Krishnakumar S, Powell AA, Zhang H, Mindrinos MN, Telli ML, et al. Single cell mutational analysis of PIK3CA in circulating tumor cells and metastases in breast cancer reveals heterogeneity, discordance, and mutation persistence in cultured disseminated tumor cells from bone marrow. *BMC Cancer*. 2014;14:456.
- Dong X, Zhang L, Milholland B, Lee M, Maslov AY, Wang T, et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat Methods*. 2017;14(5):491–3.
- Eroglu Z, Fielder O, Somlo G. Analysis of circulating tumor cells in breast cancer. *J Natl Compr Cancer Netw*. 2013;11(8):977–85.
- Ferreira MM, Ramani VC, Jeffrey SS. Circulating tumor cell technologies. *Mol Oncol*. 2016;10(3):374–94.
- Franken B, de Groot MR, Mastboom WJ, Vermes I, van der Palen J, Tibbe AG, et al. Circulating tumor cells, disease recurrence and survival in newly diagnosed breast cancer. *Breast Cancer Res*. 2012;14(5):R133.
- Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet*. 2016;48(10):1119–30.
- Gao Y, Ni X, Guo H, Su Z, Ba Y, Tong Z, et al. Single-cell sequencing deciphers a convergent evolution of copy number alterations from primary to circulating tumor cells. *Genome Res*. 2017;27(8):1312–22.
- Gascoyne PR, Noshari J, Anderson TJ, Becker FF. Isolation of rare cells from cell mixtures by dielectrophoresis. *Electrophoresis*. 2009;30(8):1388–98.
- Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet*. 2016;17(3):175–88.
- Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883–92.
- Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal*. 2014;7(344):re8.
- Grillet F, Bayet E, Villeronce O, Zappia L, Lagerqvist EL, Lunke S, et al. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. *Gut*. 2016; <https://doi.org/10.1136/gutjnl-2016-311447>.
- Harb W, Fan A, Tran T, Danila DC, Keys D, Schwartz M, et al. Mutational analysis of circulating tumor cells using a novel microfluidic collection device and qPCR assay. *Transl Oncol*. 2013;6(5):528–38.
- Hardingham JE, Grover P, Winter M, Hewett PJ, Price TJ, Thierry B. Detection and clinical significance of circulating tumor cells in colorectal cancer—20 years of progress. *Mol Med*. 2015;21(Suppl 1):S25–31.
- Harouaka RA, Nisic M, Zheng SY. Circulating tumor cell enrichment based on physical properties. *J Lab Autom*. 2013;18(6):455–68.
- Harouaka R, Kang Z, Zheng SY, Cao L. Circulating tumor cells: advances in isolation and analysis, and challenges for clinical applications. *Pharmacol Ther*. 2014;141(2):209–21.
- Harper KL, Sosa MS, Entenberg D, Hosseini H, Cheung JF, Nobre R, et al. Mechanism of early dissemination and metastasis in Her2+ mammary cancer. *Nature*. 2016; <https://doi.org/10.1038/nature20609>.
- Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*. 2012;2(3):666–73.
- Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics*. 2016;107(1):1–8.
- Heitzer E, Auer M, Gasch C, Pichler M, Ulz P, Hoffmann EM, et al. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Res*. 2013;73(10):2965–75.

- Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat Med*. 2014;20(8):897–903.
- Hong B, Zu Y. Detecting circulating tumor cells: current challenges and new trends. *Theranostics*. 2013;3(6):377–94.
- Hou JM, Krebs M, Ward T, Sloane R, Priest L, Hughes A, et al. Circulating tumor cells as a window on metastasis biology in lung cancer. *Am J Pathol*. 2011;178(3):989–96.
- Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell*. 2012;148(5):873–85.
- Hou HW, Warkiani ME, Khoo BL, Li ZR, Soo RA, Tan DS, et al. Isolation and retrieval of circulating tumor cells using centrifugal forces. *Sci Rep*. 2013;3:1259.
- Hou Y, Wu K, Shi X, Li F, Song L, Wu H, et al. Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *Gigascience*. 2015;4:37.
- Hu B, Rochefort H, Goldkorn A. Circulating tumor cells in prostate cancer. *Cancers (Basel)*. 2013;5(4):1676–90.
- Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu Rev Genomics Hum Genet*. 2015;16:79–102.
- Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods*. 2014;11(2):163–6.
- Jiang R, Lu YT, Ho H, Li B, Chen JF, Lin M, et al. A comparison of isolated circulating tumor cells and tissue biopsies using whole-genome sequencing in prostate cancer. *Oncotarget*. 2015;6(42):44781–93.
- Kallergi G, Papadaki MA, Politaki E, Mavroudis D, Georgoulas V, Agelaki S. Epithelial to mesenchymal transition markers expressed in circulating tumour cells of early and metastatic breast cancer patients. *Breast Cancer Res*. 2011;13(3):R59.
- Karabacak NM, Spuhler PS, Fachin F, Lim EJ, Pai V, Ozkumur E, et al. Microfluidic, marker-free isolation of circulating tumor cells from blood samples. *Nat Protoc*. 2014;9(3):694–710.
- Kelley RK, Magbanua MJ, Butler TM, Collisson EA, Hwang J, Sidropoulos N, et al. Circulating tumor cells in hepatocellular carcinoma: a pilot study of detection, enumeration, and next-generation sequencing in cases and controls. *BMC Cancer*. 2015;15:206.
- Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods*. 2014;11(7):740–2.
- Khoo BL, Lee SC, Kumar P, Tan TZ, Warkiani ME, Ow SG, et al. Short-term expansion of breast circulating cancer cells predicts response to anti-cancer therapy. *Oncotarget*. 2015;6(17):15578–93.
- Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells—biology and biomarkers. *Nat Rev Clin Oncol*. 2014;11(3):129–44.
- Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc*. 2015;10(3):442–58.
- Li R, Liang J, Ni S, Zhou T, Qing X, Li H, et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell*. 2010;7(1):51–63.
- Li P, Mao Z, Peng Z, Zhou L, Chen Y, Huang PH, et al. Acoustic separation of circulating tumor cells. *Proc Natl Acad Sci U S A*. 2015;112(16):4970–5.
- Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, Gerlinger M. Cancer evolution and the limits of predictability in precision cancer medicine. *Trends Cancer*. 2016;2(1):49–63.
- Liu Y, Wu BQ. Third-generation DNA sequencing: single molecule real-time DNA sequencing. *Zhonghua Bing Li Xue Za Zhi*. 2011;40(10):718–20.
- Lohr JG, Adalsteinsson VA, Cibulskis K, Choudhury AD, Rosenberg M, Cruz-Gordillo P, et al. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol*. 2014;32(5):479–84.
- Lu S, Zong C, Fan W, Yang M, Li J, Chapman AR, et al. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science*. 2012;338(6114):1627–30.
- Macaulay IC, Voet T. Single cell genomics: advances and future perspectives. *PLoS Genet*. 2014;10(1):e1004126.
- Massague J, Obenauf AC. Metastatic colonization by circulating tumour cells. *Nature*. 2016;529(7586):298–306.
- McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168(4):613–28.
- Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, Broderick KT, et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science*. 2015;349(6254):1351–6.
- Mwenifumbo JC, Marra MA. Cancer genome-sequencing study design. *Nat Rev Genet*. 2013;14(5):321–32.
- Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90–4.
- Nawy T. Single-cell sequencing. *Nat Methods*. 2014;11(1):18.
- Ni X, Zhuo M, Su Z, Duan J, Gao Y, Wang Z, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc Natl Acad Sci U S A*. 2013;110(52):21083–8.
- Nieto MA. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science*. 2013;342(6159):1234850.
- Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E, et al. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med*. 2013;5(179):179ra147.
- Palmirota R, Lovero D, Silvestris E, Felici C, Quaresmini D, Cafforio P, et al. Next-generation sequencing (NGS) analysis on single circulating tumor cells (CTCs) with no need of whole-genome amplification (WGA). *Cancer Genomics Proteomics*. 2017;14(3):173–9.
- Pantel K, Alix-Panabieres C, Riethdorf S. Cancer micrometastases. *Nat Rev Clin Oncol*. 2009;6(6):339–51.
- Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. 2013;10(11):1096–8.
- Polakis P. Wnt signaling in cancer. *Cold Spring Harb Perspect Biol*. 2012;4(5):a008052. <https://doi.org/10.1101/cshperspect.a008052>

- Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*. 2012;30(8):777–82.
- Riethdorf S, Fritsche H, Muller V, Rau T, Schindlbeck C, Rack B, 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*. 2007;13(3):920–8.
- Rothwell DG, Smith N, Morris D, Leong HS, Li Y, Hollebecque A, et al. Genetic profiling of tumours using both circulating free DNA and circulating tumour cells isolated from the same preserved whole blood sample. *Mol Oncol*. 2016;10(4):566–74.
- Russnes HG, Navin N, Hicks J, Borresen-Dale AL. Insight into the heterogeneity of breast cancer through next-generation sequencing. *J Clin Invest*. 2011;121(10):3810–8.
- Saucedo-Zeni N, Mewes S, Niestroj R, Gasiorowski L, Murawa D, Nowaczyk P, et al. A novel method for the in vivo isolation of circulating tumor cells from peripheral blood of cancer patients using a functionalized and structured medical wire. *Int J Oncol*. 2012;41(4):1241–50.
- Scher HI, Jia X, de Bono JS, Fleisher M, Pienta KJ, Raghavan D, et al. Circulating tumour cells as prognostic markers in progressive, castration-resistant prostate cancer: a reanalysis of IMMC38 trial data. *The lancet oncology*. 2009;10(3):233–9.
- Shaw JA, Guttery DS, Hills A, Fernandez-Garcia D, Page K, Rosales BM, et al. Mutation analysis of cell-free DNA and single circulating tumor cells in metastatic breast cancer patients with high circulating tumor cell counts. *Clin Cancer Res*. 2017;23(1):88–96.
- Shields CW t, Reyes CD, Lopez GP. Microfluidic cell sorting: a review of the advances in the separation of cells from debulking to rare cell isolation. *Lab Chip*. 2015;15(5):1230–49.
- Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, et al. Whole-genome multiple displacement amplification from single cells. *Nat Protoc*. 2006;1(4):1965–70.
- Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719–24.
- Talasaz AH, Powell AA, Huber DE, Berbee JG, Roh KH, Yu W, et al. Isolating highly enriched populations of circulating epithelial cells and other rare cells from blood using a magnetic sweeper device. *Proc Natl Acad Sci U S A*. 2009;106(10):3970–5.
- Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell*. 2009;139(5):871–90.
- Ting DT, Wittner BS, Ligorio M, Vincent Jordan N, Shah AM, Miyamoto DT, et al. Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep*. 2014;8(6):1905–18.
- van der Toom EE, Verdone JE, Gorin MA, Pienta KJ. Technical challenges in the isolation and analysis of circulating tumor cells. *Oncotarget*. 2016;7(38):62754–66.
- Vona G, Estepa L, Beroud C, Damotte D, Capron F, Nalpas B, et al. Impact of cytomorphological detection of circulating tumor cells in patients with liver cancer. *Hepatology*. 2004;39(3):792–7.
- Vu TN, Wills QF, Kalari KR, Niu N, Wang L, Rantalainen M, et al. Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics*. 2016;32(14):2128–35.
- Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Mol Cell*. 2015;58(4):598–609.
- Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014;512(7513):155–60.
- Wong VC, Ko JM, Lam CT, Lung ML. Succinct workflows for circulating tumor cells after enrichment: from systematic counting to mutational profiling. *PLoS One*. 2017;12(5):e0177276.
- Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*. 2012;148(5):886–95.
- Yu M, Ting DT, Stott SL, Wittner BS, Oszolak F, Paul S, et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature*. 2012;487(7408):510–3.
- Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*. 2013;339(6119):580–4.
- Zhao L, Lu YT, Li F, Wu K, Hou S, Yu J, et al. High-purity prostate circulating tumor cell isolation by a polymer nanofiber-embedded microchip for whole exome sequencing. *Adv Mater*. 2013;25(21):2897–902.