

Comparative analysis of capture methods for genomic profiling of circulating tumor cells in colorectal cancer

Joao M. Alves^{1,2,*}, Nuria Estévez-Gómez^{1,2}, Monica Valecha^{1,2}, Sonia Prado-López^{1,2,3}, Laura Tomás^{1,2}, Pilar Alvariño^{1,2}, Roberto Piñeiro^{4,5}, Laura Muínelo-Romay^{5,6}, Patricia Mondelo-Macía⁶, Mercedes Salgado⁷, Agueda Iglesias-Gómez⁸, Laura Codesido-Prada⁸, Joaquin Cubiella⁸, David Posada^{1,2,9,*}

¹CINBIO, Universidade de Vigo, 36310 Vigo, Spain

²Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO.

³Current address: Institute of Solid state Electronics, Technische Universität Wien, Austria

⁴Roche-Chus Joint Unit, Translational Medical Oncology Group, Oncomet, Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain

⁵Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain.

⁶Liquid Biopsy Analysis Unit, Translational Medical Oncology Group, Health Research Institute of Santiago de Santiago de Compostela (IDIS), Santiago de Compostela, Spain

⁷Department of Oncology, Hospital Universitario de Ourense, Research Group in Gastrointestinal Oncology-Ourense, Ourense, Spain

⁸Department of Gastroenterology Hospital Universitario de Ourense, Research Group in Gastrointestinal Oncology-Ourense, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Ourense, Spain

⁹Department of Biochemistry, Genetics, and Immunology, Universidade de Vigo, 36310 Vigo, Spain

***Corresponding authors:** Joao M. Alves (jalves@uvigo.es), David Posada (dposada@uvigo.es).

Keywords: circulating tumor cells, liquid biopsy, intratumor genomic heterogeneity, colorectal cancer, precision medicine.

Competing interests: The authors declare no competing interests.

Abstract

The genomic profiling of circulating tumor cells (CTCs) in the bloodstream should provide clinically relevant information on therapeutic efficacy and help predict cancer survival. However, the molecular characterization of CTCs has so far proven extremely difficult. A variety of technologies have been developed for CTC isolation, but so far the impact on the genomic assessment of CTCs has not been fully evaluated. To fill this gap, here we contrasted the genomic profiles of CTC pools recovered from blood samples obtained from four metastatic colorectal cancer (mCRC) patients using three different enrichment strategies (CellSearch, Parsortix, and FACS). Our results suggest clear differences in the mutational burden of CTC pools depending on the enrichment method used, with all evaluated methods returning a somewhat limited representation of the mutational spectrum of individual tumors, potentially due to allelic dropout during whole-genome amplification. Nevertheless, the CTC pools from Parsortix, and in part, CellSearch, showed diversity estimates, mutational signatures and drug-suitability scores remarkably close to the ones found in matching primary tumor samples. In contrast, FACS CTC pools were substantially enriched in apparent sequencing artifacts, which led to much higher estimates of genomic diversity. Although CTC genomics still faces technical challenges, our results suggest that CTC-derived metrics can reflect the diversity scores seen in primary tumor lesions thus highlighting the utility of CTCs to assess the heterogeneity status of individual tumors, and to help clinicians prioritize drugs in mCRC.

1. Introduction

Although research on cancer biology has traditionally been hampered by sampling issues, with most approaches relying on highly-invasive, risky, and, sometimes, difficult to obtain solid tissue biopsies (Marrugo-Ramírez, Mir, and Samitier 2018; Robertson and Baxter 2011; Chi 2016), strong evidence has emerged in recent years that the peripheral blood, as well as other body fluids, offer a valuable source of cancer-associated materials (K. Pantel and Speicher 2016; Crowley et al. 2013). As opposed to tissue biopsies, liquid biopsies represent a minimally invasive alternative to capture clinically-relevant information about tumors (Krebs et al. 2014), including circulating tumor cells (CTCs). CTCs are thought to consist of cells shed by the primary tumor (PT) and metastatic lesions into the bloodstream, and have become the subject of intense research due to their likely role in the metastatic process (Klaus Pantel and Alix-Panabières 2017; Krebs et al. 2014; Castro-Giner and Aceto 2020). In recent years, multiple studies have demonstrated the clinical significance of CTCs for prognosis and therapeutic management, with CTC burden being correlated with unfavorable overall survival in several cancer types (Magbanua et al. 2019; Silveira et al. 2021; Basso et al. 2021; Chemi et al. 2019).

Importantly, sequencing studies exploring the genomic landscape of CTCs (Chemi et al. 2019; Gulbahce et al. 2017; Magbanua et al. 2018; Court et al. 2020) have shown that the mutational profiles of CTCs generally reflects the overall genomic composition of matched primary and metastatic lesions. Collectively, these results suggest that CTC diversity represents the overall tumor heterogeneity probably better than single tumor tissue biopsies, and incorporating CTC genomic information is expected to increase the clinical value of liquid biopsies by providing better predictions of therapeutic sensitivity and survival outcome (Pailler et al. 2019; Paoletti et al. 2018).

Nevertheless, implementing a comprehensive molecular characterization of CTCs into routine clinical procedures has, so far, proven extremely challenging (Rossi and Zamarchi 2019; Kowalik, Kowalewska, and Gózdź 2017). Indeed, despite the numerous strategies already available for isolating CTCs (Bankó et al. 2019), ranging from methods based on physical properties of cells (i.e., size and deformability) to others based on biological characteristics (e.g., cell surface marker expression), CTCs are typically present at very low numbers in the blood and can show a wide range of phenotypes (Neves et al. 2021; Kowalik, Kowalewska, and Gózdź 2017; Hyun et al. 2016). As a consequence, these technologies are likely to differ in their detection sensitivity and recovery rates but, to date, comparative analyses across distinct

technologies are lacking (Neves et al. 2021). Of particular interest to us, it remains unclear how the different isolation methods can have an impact on the assessment of the genomic landscape of CTCs.

In order to identify an efficient strategy for the downstream genomic profiling of CTCs in metastatic colorectal cancer (mCRC), here we contrasted three different CTC-enrichment approaches —CellSearch[®], Parsortix[®] and Fluorescence-Activated Cell Sorting (FACS). While all methods evaluated struggled here with data quality issues, our results indicate that Parsortix and, in part, CellSearch are able to provide genomic heterogeneity scores, mutational signature profiles and therapeutic targets compatible with those found in matching primary tumor samples.

2. Material & Methods

2.1 Patient selection and blood collection

We enrolled four mCRC patients diagnosed between October 2017 and September 2019 at the Hospital Universitario de Ourense, Spain, with histologically proven CRC, and either therapy naive or showing evidence of progression. On the same day, we collected three blood samples per patient and stored them into three different containers, one for each CTC-enrichment protocol, at room temperature: CellSave Preservative tubes (Menarini Silicon Biosystems, Italy) for CellSearch, Transfix CTC-TVT tubes (Cytomark, UK) for Parsortix and cell-free DNA BCT CE tubes (Streck, NE, USA) for FACS. In addition, we obtained a formalin-fixed paraffin-embedded (FFPE) block of the primary tumor (PT) from each patient. Importantly, all specimens were obtained and collected after written informed consent from all subjects using a protocol approved by the Clinical Ethic Committee of Pontevedra-Vigo-Ourense (2018/301 approved 19/06/2018).

2.2 CTC enrichment

We processed all samples within 96 hours after the blood was drawn using three different strategies for CTC enrichment. The CellSearch[®] system (Menarini, Silicon Biosystems, Bologna, Italy) enumerates and isolates CTCs of epithelial origin (CD45⁻, EpCAM⁺, and CK8⁺, 18⁺, and/or 19⁺). The Parsortix[®] platform (ANGLE plc, UK) traps CTCs due to their larger size and lower compressibility than blood cells. The FACS strategy separates CTCs based on custom markers; in our case EpCAM⁺/CD45⁻/CK7,8⁺.

2.2.1 CellSearch

For each sample, we processed 7.5 mL of whole blood in the CellTracks Autoprep system using the Circulating Tumor Cell Kit (Menarini, Silicon Biosystems, Bologna, Italy). This kit consists of ferrofluids coated with epithelial cell-specific anti-EpCAM antibodies to immuno-magnetically enrich epithelial cells and a mixture of antibodies directed to cytokeratins (CKs) 8, 18, and 19 conjugated to phycoerythrin (PE); an antibody to CD45 conjugated to allophycocyanin (APC); and a nuclear dye 4',6-diamidino-2-phenylindole (DAPI). Afterward, we analyzed the processed samples with the CellTracks Analyzer II according to the manufacturer's instructions. We identified the CTCs as round or oval cells with an intact nucleus (DAPI positive), CK positive and CD45 negative (**Fig. S1**). We stored the CTC-enriched samples at -80 °C.

2.2.2 Parsortix

We loaded 7.5 mL of peripheral whole blood per sample into a Parsortix microfluidic device (Angle plc, UK). We enriched the samples in disposable Parsortix cassettes with a gap size of 6.5 μm (GEN3D6.5, Angle Inc., Guildford, UK) and at 99 mbar of pressure, according to the manufacturer's guidelines. After separation, we collected the captured cells in 200 μL of PBS and stored them at -80 °C.

2.2.3 Fluorescence-activated Cell Sorting (FACS)

In order to obtain the peripheral blood mononuclear cell (PBMC) fraction –to be used as healthy controls– we took 1 mL from each blood sample and performed Ficoll-Paque gradient centrifugation. We kept the PBMCs in RNA later (Ambion, TX, USA) at -80°C until genomic DNA (gDNA) extraction. The remaining blood volume (7-9 mL) was then used for CTC staining and collection. After validating our FACS protocol using spike-in experiments (see **Supplementary note 1**), we followed a similar approach to Miller et al. (2012) and first lysed the red blood cells using BD Pharm Lyse lysing solution (BD Biosciences, NJ, USA) following the fabricant recommendations. When needed, we repeated the lysing step up to three times. We then resuspended the cells in phosphate-buffered saline (PBS) solution and filtered them with a 70 µm cell strainer (Falcon, NY, USA). We used the FIX & PERM™ Cell Permeabilization Kit (Invitrogen, MA, USA) and incubated the filtered cell suspensions with antibodies (BD Biosciences, NJ, USA) against the epithelial cell adhesion molecule (EpCAM; PerCP-Cy5.5, IgG1λ, clone EBA-1) and the leukocyte common antigen CD45 (FITC, IgG1κ, clone HI30) with reagent A for 25 minutes in the dark at room temperature for fixation. Cells were then washed once with 500 µL of PBS and centrifuged at 200 x g for 5 minutes. We resuspended the cell pellet in 1 ml of PBS and incubated it with an antibody against the epithelial markers cytokeratins 7 and 8 (CK7,8; PE, IgG2aκ, clone CAM 5.2) and reagent B for 20 minutes in the dark at room temperature for permeabilization. We washed once again and resuspended the cells in 500 µl of PBS. Finally, we selected and collected 3 µl of PBS pools of CTCs based on an EpCAM+/CD45-/CK7,8+ phenotype (**Fig. S2**) using a FACSAria III (BD Biosciences, NJ, USA). We analyzed the data using the FACSDiva (BD Biosciences, NJ, USA) and FlowLogic software (Miltenyi Biotec, Germany).

2.3 Whole-genome amplification of CTC-pools

Given the large collection volume (~200 µl) from both CellSearch and Parsortix, we initially performed genomic DNA (gDNA) extraction of the CTC enriched samples obtained from these platforms using the QIAamp DNA Blood Mini Kit (Qiagen, Germany) before performing whole-genome amplification (WGA) using the Ampli1 kit (Menarini Silicon Biosystems, Italy). For CellSearch and Parsortix samples, we carried out the WGA starting with 1 µl of DNA and CTC pools from FACS were amplified directly. In order to avoid contamination, we worked in a laminar-flow hood and used a dedicated set of pipettes and UV-irradiated plastic materials. We included positive (10 ng/µl REPLIg human control kit, Qiagen, Germany) and negative (DNase/RNase free water) controls during the amplification and used the Ampli1 QC Kit to evaluate the amplification. Samples with a positive signal for at least two PCR fragments were selected to increase the total dsDNA content using the Ampli1 ReAmp/ds kit. We then removed the kit adaptors by incubating at 37 °C for 3 h a mixture of 5 µl of NEBuffer 4 10X (New England Biolabs, MA, USA), 1 µl of MseI 50U/µl (New England Biolabs, MA, USA), 19 µl of nuclease-free water and 25 µl of dsDNA followed by a step at 65 °C for 20 min for enzyme inactivation. Finally, we purified the samples with 1.8X AMPure XP beads (Agencourt, Beckman Coulter, CA, USA), quantified the DNA yield with Qubit 3.0 fluorometer (Thermo Fisher Scientific, MA, USA) and checked the amplicon size distribution with the D1000 ScreenTape System in a 2200 TapeStation platform (Agilent Technologies, CA, USA).

2.4 FFPE and PBMCs bulk gDNA isolation

We performed the extraction of bulk gDNA from FFPE samples using the QIAamp DNA FFPE tissue kit (Qiagen, Germany) by incubating and shaking at 60 °C for 1 h before slicing them and adding a deparaffinization solution (DS). We performed a 56°C incubation step for 1 h and when needed, performed a second addition of DS in a new tube in order to remove remaining paraffin. We used the QIAamp DNA Blood Mini Kit (Qiagen, Germany) for the extraction of gDNA from PBMCs bulks and estimated DNA yield using the Qubit 3.0 fluorometer (Thermo Fisher Scientific, MA, USA) and DNA integrity with the Genomic DNA ScreenTape Assay (Agilent Technologies, CA, USA).

2.5 Whole-exome sequencing

CTC pools and bulk sequencing libraries were constructed at the Spanish National Center for Genomic Analysis (CNAG; <http://www.cnag.crg.eu>) with the SureSelect XT and Agilent Human Exon v5 kits (Agilent Technologies, CA, USA). In total, seven whole-genome amplified CTC-pools (two from FACS: P4 and P5, one from CellSearch: P1 and 4 from Parsortix: P1, P3, P4 and P5) and four FFPE bulk samples were sequenced at 100X and four PBMCs samples at 60X. All samples were run on an Illumina NovaSeq 6000 (PE100) at CNAG.

2.6 Data processing and variant calling

After trimming amplification and sequencing adapters from the raw FASTQ files, we aligned the sequencing reads from CTC pools, tumor and healthy samples to the Genome Reference Consortium Human Build 37 (GRCh37) using the MEM algorithm in the BWA software (Li 2013). Following a standardized best-practices pipeline (Van der Auwera et al. 2013), we filtered out reads with low mapping quality. We next performed a local realignment around indels, and removed PCR duplicates. We identified somatic single nucleotide variants (SNVs) for each CTC-capture method using the multi-sample variant-calling feature implemented in MuTect2 software, taking as input the BAM files of the different sample types available (i.e., tumor bulk + CTC-pool + healthy control). We then used FilterMutectCalls to remove calls in any sequence context artifacts or contamination fractions (but see **Supplementary note 2** and **Fig. S3**). Afterwards, the genotypes of variants showing a coverage depth ≥ 10 , alternative allelic depth ≥ 2 and allele frequency estimates ranging from 0.05 to 0.75 were kept for downstream analysis. For all datasets, we merged the inferred SNV calls and performed variant annotation using Annovar software (v.20200608) (Wang, Li, and Hakonarson 2010).

2.7 Mutational Signatures

For all datasets, we ran sigProfilerExtractor (Ashiquil Islam et al. 2021) under default parameters to identify *de novo* mutational signatures for single-base substitutions (SBS), followed by the assignment of the decomposed signatures to known COSMICv3 SBS96 signatures (Alexandrov et al. 2020).

2.8 PanDrugs

We used PanDrugs (Piñeiro-Yáñez et al. 2018) (<http://www.pandrugs.org>) –a web-based platform that attempts to match genomic data to available drug therapies in order to guide personalized treatment selection– to explore changes in therapeutic options and drug suitability scores across the different datasets. For that purpose, we first ran PanDrugs using a VCF with the list of exonic mutations identified in the tumor bulk sample of each patient to identify CRC-specific therapeutic candidates. Next, we extracted the drug score (that ranges from -1 to 1 and measures the suitability of each drug using a database of curated gene-drug relationships and the collective gene impact) to identify the top 25 therapeutic candidates in the PT samples. Afterwards, we performed a new query with PanDrugs using the exonic mutations in each of the CTC samples to examine whether the CTC-derived genomic information identifies similar therapeutic options and drug sensitivity scores as for the bulk.

2.9 MATH scores

For the datasets with an unknown number of CTCs or with CTC counts > 1 , we additionally estimated the mutant-allele tumor heterogeneity (MATH) score (Mroz and Rocco 2013). The MATH score is based on the distribution of allele fractions among somatic mutations and is calculated as the percentage ratio of the width of the data to the center of its distribution:

$$MATH = 100 \times \frac{\text{median absolute deviation (VAF)}}{\text{median (VAF)}}$$

Importantly, since MATH scores are sensitive to unreliable allele frequency estimates stemming from sites with poor sequencing coverage depth, for their calculation we applied an additional filter to restrict our mutation calls to positions showing a depth of coverage ≥ 25 and alternative allelic depth ≥ 5 .

3. Results

3.1 CTC-counts

Patient-level CTC counts were available for two of the selected CTC-capture methods (as the Parsortix platform does not provide CTC counts). Using the CellSearch system, we isolated one CTC from patient P1, whereas no CTCs were detected for the remaining patients. Importantly, although we were able to recover potential CTCs from all patients using FACS (P1 = 2 CTCs; P3 = 6 CTCs; P4 = 2 CTCs; P5 = 1 CTC), after whole-genome amplification we only obtained high-quality sequencing libraries for patients P4 and P5.

3.2 Tumor mutational burden

Across all patients, we found sharp differences in the number of somatic mutations (SNVs) identified with the different CTC-capture methods (**Fig. 1**). While the number of mutations called with both CellSearch and Parsortix datasets were close to, or lower than, the number of mutations observed in the matched PT bulk samples, the mutation counts in the FACS CTC pools were consistently much higher (by one order of magnitude) than in the bulk samples. Remarkably, regardless of the CTC-capture strategy, we found a very small overlap in mutation calls between CTC pools and PT datasets: Parsortix (average of 4.6%), CellSearch (3.4%) and FACS (0.4%).

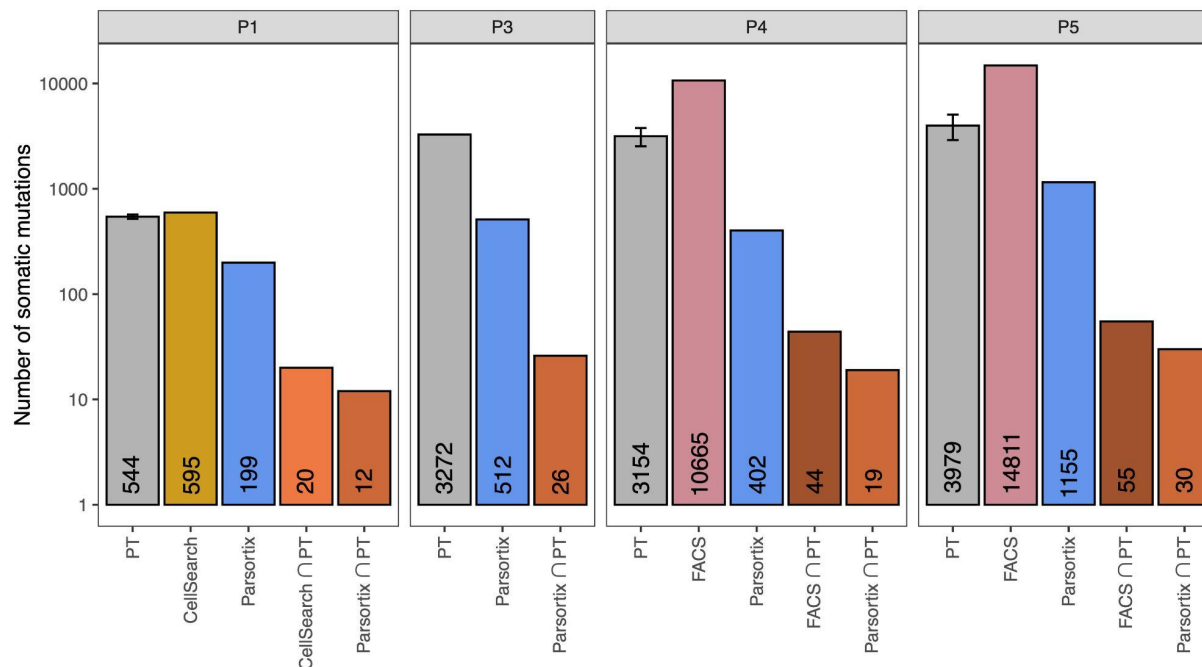


Figure 1. SNV abundance per CTC-capturing method. Barplots depicting the total number of SNVs identified by MuTect2 for each dataset. Number of SNVs shown at the bottom of each bar. Bar colors reflect the different input material or capturing method: gray = FFPE primary tumor (PT) sample; gold = CellSearch CTC pool; blue = Parsortix CTC pool; pink = FACS CTC pool. Error bars are only available for bulk tumor samples and reflect the 95% confidence interval. Orange bars depict the amount of shared sites between CTC-datasets and PT samples. The y-axis is in log scale.

Similarly, within each patient, the number of mutations shared between CTC pools captured with distinct methods was generally small (**Fig. 2a**). In any case, we identified a considerable number of CTC-specific mutations across datasets, including non-silent ones (**Fig. 2a** and **Fig. S4**).

Furthermore, within the PT samples, the median variant allele frequency (VAF) of the shared mutations with CTC pools ranged from 0.19 (Parsortix - P4) to 0.09 (FACS - P5), suggesting that the isolated CTCs are derived from minor subclones within the PTs (**Fig. 2b**). Similar VAF scores were observed between the CellSearch and Parsortix CTCs. In P4 we found a significant difference between Parsortix and FACS (**Fig. 2b**), with the FACS CTCs being generally enriched with mutations at low frequency.

Importantly, as shown in **Fig. 2b**, the CTC pools showed a clear depletion of clonal (i.e., allele frequency ≥ 0.4) mutations observed in the PT samples. While these results may appear surprising - as, in theory, clonal mutations in the PT should appear in all CTCs sampled-, the failure to identify such mutations can be, in part, explained due to the limited coverage breadth of the CTCs. Indeed, across all samples, only a small fraction (5 to 39%) of regions harboring clonal variants in the PTs was covered with sequencing reads in the CTC pools (**Table S1**). Moreover, after looking at the coverage statistics of heterozygous single-nucleotide polymorphisms (SNPs) (**Fig. S5**), we additionally found strong evidence of allele dropout (ADO) taking place during WGA. Despite being particularly obvious in the FACS datasets -which showed an averaged ADO of 67% for the heterozygous sites called-, all CTC pools showed some degree of ADO (CellSearch ADO: 9% and Parsortix ADO: 9%) which very possibly interfered with the identification of clonal mutations.

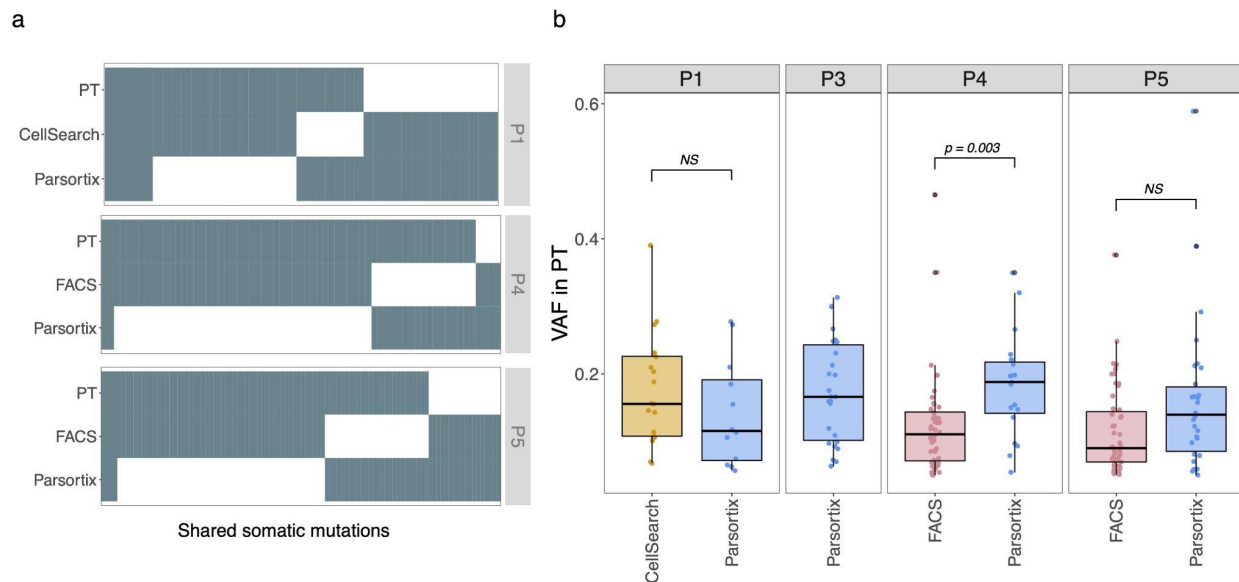


Figure 2. Genomic profiling and clonality of CTC-pools. **a.** Occupancy matrix of shared sites (P1 = 41 SNVs; P4 = 65 SNVs; P5 = 100 SNVs) across the CTC-capture methods within patients. Different colored tiles reflect different mutation status: dark grey=mutation; white=reference/missing data. Patient ID shown on the right. **b.** Boxplots depicting the bulk-level VAF estimates of the SNVs shared with the CTC pools. Boxplot colors represent the different CTC-capture methods. Statistical analysis was performed using the two sample KS-test to compare bulk VAF estimates of shared sites between CTC-capture methods. Significant *p-values* shown above boxplots.

3.3 Measuring intratumor genomic heterogeneity (ITH)

Afterwards, we explored potential differences in intratumor genomic heterogeneity (ITH) estimates between PT and CTC samples by examining the frequency distribution of somatic mutations across the different datasets. Importantly, since several of our CTC samples comprised only one cell, the subsequent

analyses were limited to datasets with either an unknown number of CTCs or CTC counts > 1, and included the Parsortix CTCs from patients P1, P3, P4 and P5, and the FACS CTCs from patient P4.

As illustrated in **Fig. 3a-b** (and **Fig. S6a-b**), we found significant differences in the distribution of VAF estimates between PT and CTC pools. In patient P4, specifically, both FACS and Parsortix datasets showed an enrichment towards low frequency variants when compared to the corresponding PT sample. Moreover, a quantile-quantile (Q-Q) plot further revealed contrasting differences in the skewness of the VAF distributions stemming from the different CTC-capture methods (**Fig. 3b**). Indeed, while the Parsortix CTCs encompassed a larger number of mutations at intermediate frequencies when compared to the matched PT sample, the FACS CTCs showed a substantial depletion of this class of mutations.

Interestingly, in the Parsortix CTCs, we found a significant positive correlation between the VAF estimates of shared mutations between the CTC pool and the PT sample (**Fig. 3c**). As for the FACS CTCs, although a positive trend was found, the relationship was not significant (similar to the results of the remaining datasets - **Fig. S6c**).

We also observed sharp differences in the levels of ITH, as measured by MATH scores, among CTC-capture methods. In patient P4, the Parsortix CTCs and the PT sample returned highly concordant MATH scores (43.6 and 43.0, respectively), while the FACS dataset displayed a much larger MATH score (62.8) (**Fig. 3d**). Importantly, very similar MATH scores were also observed between the remaining Parsortix CTCs and the PT samples (**Fig. S6d**).

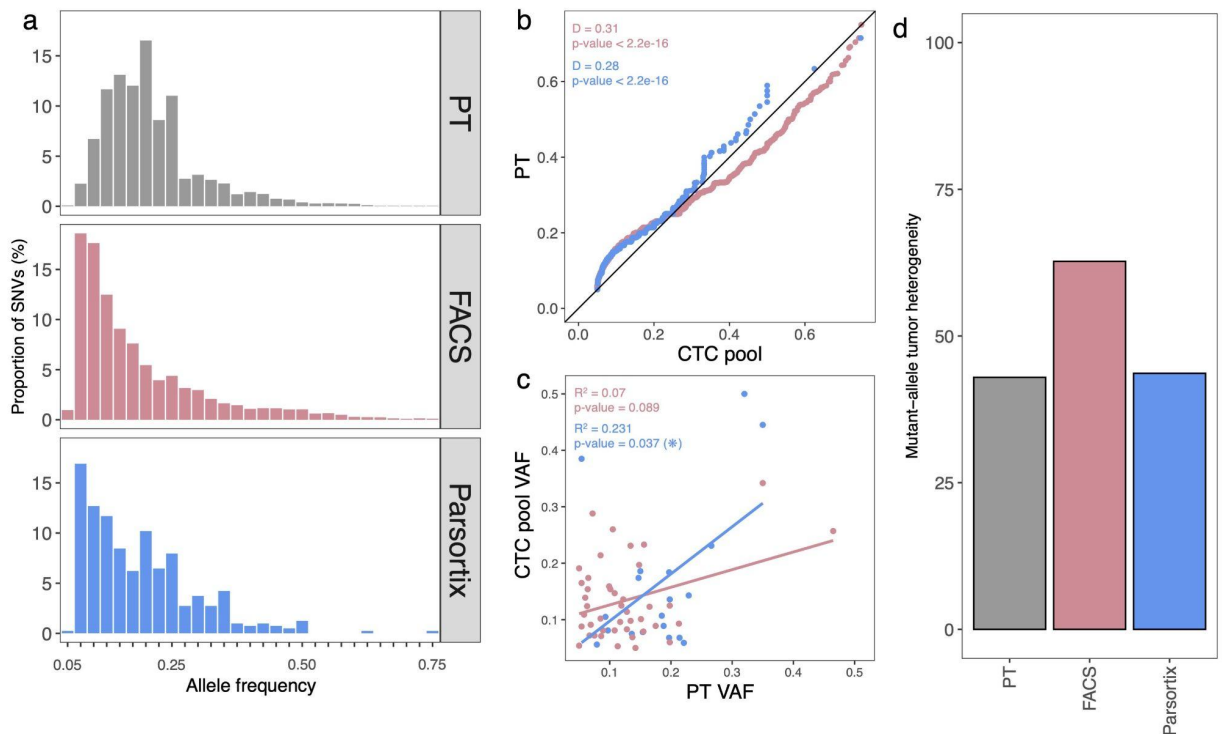


Figure 3. Measuring ITH through CTC pools. **a.** Histograms depicting the variant allele frequency (VAF) distribution of somatic mutations for the different datasets of patient P4. Different datasets highlighted with different colors: PT = grey; FACS = pink; Parsortix = blue with dataset ID shown on the right. Histograms scaled to percentages. **b.** Q-Q plot comparing the distribution of allele frequency estimates in CTC pools and PT samples of patient P4. Statistical analysis was performed using the two sample KS-test to compare the VAF distribution between CTC pools and PT samples. KS D (distance) statistic and *p*-values are shown on the upper left side of the plot. **c.** Scatter plot describing the similarity of VAF scores of overlapping sites between CTC pools and PT samples. Solid lines represent the best fit from regression analysis. R² scores and *p*-values are shown on the upper left side of the plot. **d.** Barplot depicting the MATH scores obtained using the mutation sets passing our strict filtering - Primary tumor = 416 SNVs; FACS = 8193 SNVs; Parsortix = 46 SNVs (see *methods*).

3.4 Mutational signatures

We next explored the mutational signatures in the different datasets to look for potential differences between CTC pools and PT samples (**Fig. 4a**). Across all datasets, *SigProfilerExtractor* identified a total of five different mutational processes, with PT samples being predominantly enriched in “clock-like” (i.e., ageing) COSMIC signatures SBS1 and SBS5. A similar contribution of SBS1 and SBS5 was also found in the CellsSearch and Parsortix CTC datasets, albeit the lower cosine similarity score (cosine similarity: 0.505) obtained - likely due to the limited number of mutations available for signature assignment in these datasets. In contrast, FACS CTCs were predominantly characterized by signature SBS46 (65% in both cases), a mutational signature typically associated with sequencing artifacts.

3.5 Drug suitability scores

Finally, using the list of exonic positions available, we collected the best drug candidates for each dataset in order to evaluate potential changes in the type and sensitivity of therapeutic options between CTC pools and PT samples (**Fig. 4b**). Across patients, both FACS and CellSearch CTCs recovered, in most cases, the top candidate drugs (i.e., with the highest D-Score) identified with the corresponding PT samples. The Parsortix CTCs, on the other hand, often returned only a subset of treatment options identified in the matched PT. In any case, the fact that all CTC-capture methods showed a significant positive correlation in drug sensitivity scores between CTCs and PT samples (**Fig. 4c**) appears to indicate that the CTC genomic profiles offer reliable information for prioritizing therapeutic strategies in CRC.

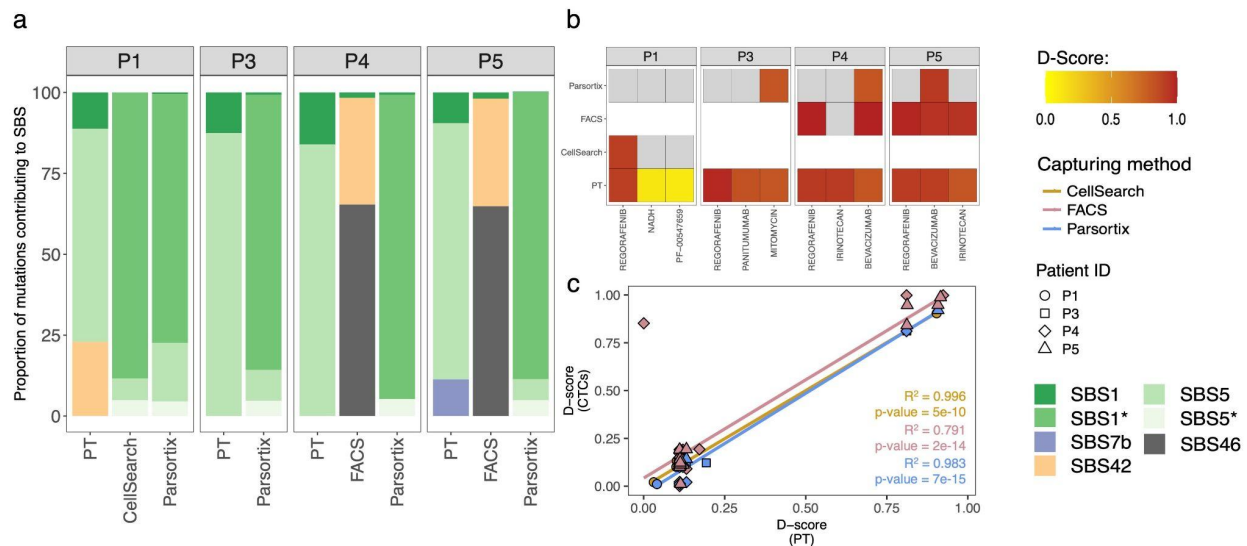


Figure 4. Mutational signatures and drug sensitivity concordance. a. Barplots depicting the proportion of mutations contributing to the different signatures/processes across the different samples. Sample IDs are shown at the bottom and patient IDs are shown at the top. Different colors reflect the identified mutational signature with the COSMIC SBS ID shown at the right of the plot. Legend asterisks distinguish mutational signatures identified by *SigProfilerExtractor* showing suboptimal cosine similarity scores (i.e., < 0.90). **b.** Tile plots depicting the overlap between the top-3 drug candidates in primary tumor (PT) samples (ordered from left to right) and corresponding CTC-pools. Tiles colored according to drug sensitivity score (D-score). Gray tiles correspond to therapeutic options not recovered by *PanDrugs* for that specific sample. **c.** Scatter plot depicting the correlation between the best therapeutic candidates (measured using the D-score) identified in both PT and CTC samples. Shape distinguishes the different patients while colors reflect the different CTC-capture methods. Solid lines represent the best fit from regression analysis. R^2 scores and p -values are shown on the bottom right side of the plot.

4. Discussion

The isolation of CTCs remains challenging largely because of their scarcity in peripheral blood and phenotypic heterogeneity. While recent technological advances have resulted in improved capture strategies, most available methods differ in key aspects such as enrichment efficiency, cell viability, and throughput. Studies evaluating the implications of different CTC-capture methods on downstream analyses are scarce, and it remains unclear whether distinct enrichment approaches can provide compatible descriptions of the mutational landscape of CTCs.

In this study, in order to evaluate the impact of different CTC-capture strategies on the downstream molecular characterization of CTCs, we contrasted the genomic profiles of primary tumor samples against CTC pools recovered using three different enrichment strategies in four mCRC patients.

Our whole-exome sequencing experiments suggest differences in the mutational loads of CTC pools due to the enrichment method used. In sharp contrast to the results obtained with the CellSearch and Parsortix systems, the mutation counts in our FACS CTC pools exceeded by an order of magnitude the number of mutations observed in the corresponding primary tumor samples. Importantly, the FACS datasets showed significant enrichment for mutations at lower frequencies, with a high proportion of these mutations being later linked to a mutational signature associated with sequencing artifacts (i.e., SBS46).

These results suggest that our FACS-derived CTCs might have accumulated DNA lesions along with the different steps of the protocol. Unfortunately, identifying and subsequently removing these potential errors is not necessarily straightforward. Although one could argue that setting lower and upper bounds on the minor allele frequency could potentially prevent downstream variant call artifacts, it should be noted that sequencing CTC pools must necessarily be preceded by multiple rounds of genomic amplification. Since most WGA methods inevitably introduce biases in the resulting sequencing data (e.g., uneven genome amplification, allelic imbalances and dropout), for any given site, the derived allele frequency score will not necessarily reflect its true frequency in the CTC population sampled. Indeed, in our datasets, these biases were reflected both in the imbalanced distribution of allele frequencies of heterozygous SNPs and in the relatively poor concordance in VAFs of shared somatic sites between CTC pools and PT samples.

In any case, and for all datasets analyzed, Parsortix-derived CTC pools provided similar descriptions of ITH when compared to the corresponding PT samples, with MATH scores in clear agreement with previous estimates in mCRC (Bettoni et al. 2019). In contrast, this metric was largely overestimated in our FACS dataset, perhaps as a consequence of the massive amount of potential sequencing artifacts. Moreover, while the identification of therapeutic candidates was not always identical between CTC pools and PT samples, all CTC pools analyzed suggested drug-candidates displaying a significant probability of response, and highly concordant sensitivity scores with the PT samples, thus providing strong evidence that CTC-based mutational profiles may contribute with valuable guidance for refining treatment tailoring (Khoo et al. 2016; Siravegna et al. 2017; Parikh et al. 2019).

5. Conclusion

In conclusion, CTC genomics still faces technical challenges for straightforward clinical applications. As seen throughout our study, all the methods evaluated struggled with data quality issues - potentially caused by the inherent technical bias introduced by the limiting amounts of input material (Navin 2014) and/or background DNA contamination (e.g., white blood cells) (Xu et al. 2015) – which resulted in a somewhat incomplete picture of the mutational landscape of these tumors. Nevertheless, and despite the limited number of patients analyzed and the failure of some enrichment methods to recover CTCs for all patients, it is important to highlight that the CTC pools recovered from Parsortix and, in part, CellSearch

returned comparable ITH estimates, similar mutational signature profiles, and suggested equivalent therapeutic targets, when compared to those found in matching primary tumor samples.

On this basis, as the performance of technologies for the detection and isolation of CTCs continues to improve, allowing for more accurate and informative genomic data to be produced (Diamantopoulou, Castro-Giner, and Aceto 2020), future studies (using larger cohorts) should explore whether the mutational landscape and genomic diversity of CTC populations can indeed provide clinically relevant prognostic and predictive information beyond simple enumeration.

Declarations

Ethics approval and consent to participate

All specimens used in this study were obtained and collected after written informed consent from all subjects using a protocol approved by the Clinical Ethic Committee of Pontevedra-Vigo-Ourense (2018/301 approved 19/06/2018).

Availability of data and materials

We have deposited raw whole-exome sequencing data at the Sequence Read Archive database under the accession code XXXXXXXX (to be provided).

Funding

This work was supported by an AXA Research Fund postdoctoral grant (awarded to J.M.A), and by the Spanish Ministry of Science and Innovation - MICINN (PID2019-106247GB-I00 awarded to D.P.). J.M.A. is currently supported by the AECC (INVES20007FERN). D.P. receives further support from Xunta de Galicia. M.V. is supported by an H2020/ Marie Skłodowska-Curie Actions EU research framework programme grant (Project H2020 MSCA-ITN-2017-766030). L.T. received support from a Ph.D. fellowship from Xunta de Galicia (ED481A-2018/303). J.C. received grants from Spain's Carlos III Health Care Institute (Co-funded by European Regional Development Fund/European Social Fund "A way to make Europe"/"Investing in your future"), No. PI17/00837 and PI21/01771. J.C. is additionally funded by the Agencia Gallega de Innovación (N607B-2020/02). R.P. received support from Roche-Chus Joint Unit (IN853B 2018/03) funded by Axencia Galega de Innovación (GAIN), Consellería de Economía, Emprego e Industria.

Authors' contributions

D.P. and J.M.A. conceived the study and designed the analyses. M.S. and J.C. collected the blood and tumor samples and patient information. N.E.G., S.P.L. and P.A. processed the samples, performed the spike-in experiments, obtained the CTC pools through FACS, and performed whole-genome amplification. R.P, L.M.R and P.M collected the CTC pools from CellSearch and Parsortix. J.M.A., L.T., and M.V. performed the analyses. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank all members of the lab for their comments on earlier versions of the manuscript. We also thank the Supercomputation Center of Galicia (CESGA) for providing all computational resources and Mercedes Peleteiro, the flow cytometry core facility manager, for her help and support.

References

- Alexandrov, Ludmil B., Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, et al. 2020. "The Repertoire of Mutational Signatures in Human Cancer." *Nature* 578 (7793): 94–101.
- Ashiquil Islam, S. M., Yang Wu, Marcos Díaz-Gay, Erik N. Bergstrom, Yudou He, Mark Barnes, Mike Vella, et al. 2021. "Uncovering Novel Mutational Signatures by de Novo Extraction with SigProfilerExtractor." *bioRxiv*. <https://doi.org/10.1101/2020.12.13.422570>.
- Bankó, Petra, Sun Young Lee, Viola Nagygyörgy, Miklós Zrínyi, Chang Hoon Chae, Dong Hyu Cho, and András Telekes. 2019. "Technologies for Circulating Tumor Cell Separation from Whole Blood." *Journal of Hematology & Oncology*. <https://doi.org/10.1186/s13045-019-0735-4>.
- Basso, Umberto, Antonella Facchinetti, Elisabetta Rossi, Marco Maruzzo, Vincenza Conteduca, Michele Aieta, Francesco Massari, et al. 2021. "Prognostic Role of Circulating Tumor Cells in Metastatic Renal Cell Carcinoma: A Large, Multicenter, Prospective Trial." *The Oncologist*. <https://doi.org/10.1002/onco.13842>.
- Bettoni, Fabiana, Cibele Masotti, Bruna R. Corrêa, Elisa Donnard, Filipe F. dos Santos, Guilherme P. São Julião, Bruna B. Vailati, et al. 2019. "The Effects of Neoadjuvant Chemoradiation in Locally Advanced Rectal Cancer—The Impact in Intratumoral Heterogeneity." *Frontiers in Oncology*. <https://doi.org/10.3389/fonc.2019.00974>.
- Castro-Giner, Francesc, and Nicola Aceto. 2020. "Tracking Cancer Progression: From Circulating Tumor Cells to Metastasis." *Genome Medicine* 12 (1): 31.
- Chemi, Francesca, Dominic G. Rothwell, Nicholas McGranahan, Sakshi Gulati, Chris Abbosh, Simon P. Pearce, Cong Zhou, et al. 2019. "Pulmonary Venous Circulating Tumor Cell Dissemination before Tumor Resection and Disease Relapse." *Nature Medicine* 25 (10): 1534–39.
- Chi, Kelly Rae. 2016. "The Tumour Trail Left in Blood." *Nature*. <https://doi.org/10.1038/532269a>.
- Court, Colin M., Shuang Hou, Lian Liu, Paul Winograd, Benjamin J. DiPardo, Sean X. Liu, Pin-Jung Chen, et al. 2020. "Somatic Copy Number Profiling from Hepatocellular Carcinoma Circulating Tumor Cells." *NPJ Precision Oncology* 4 (July): 16.
- Crowley, Emily, Federica Di Nicolantonio, Fotios Loupakis, and Alberto Bardelli. 2013. "Liquid Biopsy: Monitoring Cancer-Genetics in the Blood." *Nature Reviews. Clinical Oncology* 10 (8): 472–84.
- Diamantopoulou, Zoi, Francesc Castro-Giner, and Nicola Aceto. 2020. "Circulating Tumor Cells: Ready for Translation?" *The Journal of Experimental Medicine* 217 (8). <https://doi.org/10.1084/jem.20200356>.
- Gulbahce, Natali, Mark Jesus M. Magbanua, Robert Chin, Misha R. Agarwal, Xuhao Luo, Jia Liu, Daniel M. Hayden, et al. 2017. "Quantitative Whole Genome Sequencing of Circulating Tumor Cells Enables Personalized Combination Therapy of Metastatic Cancer." *Cancer Research* 77 (16): 4530–41.
- Hyun, Kyung-A, Gi-Bang Koo, Hyunju Han, Joohyuk Sohn, Wonshik Choi, Seung-Il Kim, Hyo-Il Jung, and You-Sun Kim. 2016. "Epithelial-to-Mesenchymal Transition Leads to Loss of EpCAM and Different Physical Properties in Circulating Tumor Cells from Metastatic Breast Cancer." *Oncotarget* 7 (17): 24677–87.
- Khoo, Bee Luan, Gianluca Grenzi, Tengyang Jing, Ying Bena Lim, Soo Chin Lee, Jean Paul Thiery, Jongyoon Han, and Chwee Teck Lim. 2016. "Liquid Biopsy and Therapeutic Response: Circulating Tumor Cell Cultures for Evaluation of Anticancer Treatment." *Science Advances* 2 (7): e1600274.
- Kowalik, Artur, Magdalena Kowalewska, and Stanisław Góźdz. 2017. "Current Approaches for Avoiding the Limitations of Circulating Tumor Cells Detection Methods—implications for Diagnosis and Treatment of Patients with Solid Tumors." *Translational Research*. <https://doi.org/10.1016/j.trsl.2017.04.002>.
- Krebs, Matthew G., Robert L. Metcalf, Louise Carter, Ged Brady, Fiona H. Blackhall, and Caroline Dive. 2014. "Molecular Analysis of Circulating Tumour Cells-Biology and Biomarkers." *Nature Reviews. Clinical Oncology* 11 (3): 129–44.
- Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." *arXiv [q-bio.GN]*. <http://arxiv.org/abs/1303.3997>.
- Magbanua, Mark Jesus M., Hope S. Rugo, Denise M. Wolf, Louai Hauranieh, Ritu Roy, Praveen Pendyala, Eduardo V. Sosa, et al. 2018. "Expanded Genomic Profiling of Circulating Tumor Cells in Metastatic Breast Cancer Patients to Assess Biomarker Status and Biology Over Time (CALGB 40502 and CALGB 40503, Alliance)." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 24 (6): 1486–99.
- Magbanua, Mark Jesus M., Christina Yau, Denise M. Wolf, Jin Sun Lee, Aheli Chattopadhyay, Janet H. Scott, Erin Bowlby-Yoder, et al. 2019. "Synchronous Detection of Circulating Tumor Cells in Blood and Disseminated Tumor Cells in Bone Marrow Predicts Adverse Outcome in Early Breast Cancer." *Clinical Cancer Research*. <https://doi.org/10.1158/1078-0432.ccr-18-3888>.
- Marrugo-Ramírez, José, Mònica Mir, and Josep Samitier. 2018. "Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy." *International Journal of Molecular Sciences* 19 (10). <https://doi.org/10.3390/ijms19102877>.
- Miller, Brandon L, Jeffrey J.Chalmers and William G.Lowrie. "Cell Enrichment from Human Blood through Red Cell Lysis." 2012. In *Methods in Cell Biology*, 112:183–92. Academic Press.

- Mroz, Edmund A., and James W. Rocco. 2013. "MATH, a Novel Measure of Intratumor Genetic Heterogeneity, Is High in Poor-Outcome Classes of Head and Neck Squamous Cell Carcinoma." *Oral Oncology*. <https://doi.org/10.1016/j.oraloncology.2012.09.007>.
- Navin, Nicholas E. 2014. "Cancer Genomics: One Cell at a Time." *Genome Biology*. <https://doi.org/10.1186/s13059-014-0452-9>.
- Neves, Rui P. L., Wim Ammerlaan, Kiki C. Andree, Sebastian Bender, Laure Cayrefourcq, Christiane Driemel, Claudia Koch, et al. 2021. "Proficiency Testing to Assess Technical Performance for CTC-Processing and Detection Methods in CANCER-ID." *Clinical Chemistry* 67 (4): 631–41.
- Pailler, Emma, Vincent Faugeroux, Marianne Oulhen, Laura Mezquita, Mélanie Laporte, Aurélie Honoré, Yann Lecluse, et al. 2019. "Acquired Resistance Mutations to ALK Inhibitors Identified by Single Circulating Tumor Cell Sequencing in -Rearranged Non-Small-Cell Lung Cancer." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 25 (22): 6671–82.
- Pantel, Klaus, and Catherine Alix-Panabières. 2017. "Circulating Tumour Cells and Cell-Free DNA in Gastrointestinal Cancer." *Nature Reviews Gastroenterology & Hepatology*. <https://doi.org/10.1038/nrgastro.2016.198>.
- Pantel, K., and M. R. Speicher. 2016. "The Biology of Circulating Tumor Cells." *Oncogene*. <https://doi.org/10.1038/onc.2015.192>.
- Paoletti, Costanza, Andi K. Cani, Jose M. Larios, Daniel H. Hovelson, Kimberly Aung, Elizabeth P. Darga, Emily M. Cannell, et al. 2018. "Comprehensive Mutation and Copy Number Profiling in Archived Circulating Breast Cancer Tumor Cells Documents Heterogeneous Resistance Mechanisms." *Cancer Research* 78 (4): 1110–22.
- Parikh, Aparna R., Ignaty Leshchiner, Liudmila Elagina, Lipika Goyal, Chaya Levovitz, Giulia Siravegna, Dimitri Livitz, et al. 2019. "Liquid versus Tissue Biopsy for Detecting Acquired Resistance and Tumor Heterogeneity in Gastrointestinal Cancers." *Nature Medicine* 25 (9): 1415–21.
- Piñeiro-Yáñez, Elena, Miguel Reboiro-Jato, Gonzalo Gómez-López, Javier Perales-Patón, Kevin Troulé, José Manuel Rodríguez, Héctor Tejero, et al. 2018. "PanDrugs: A Novel Method to Prioritize Anticancer Drug Treatments according to Individual Genomic Data." *Genome Medicine* 10 (1): 41.
- Robertson, E. G., and G. Baxter. 2011. "Tumour Seeding Following Percutaneous Needle Biopsy: The Real Story!" *Clinical Radiology* 66 (11): 1007–14.
- Rossi, Elisabetta, and Rita Zamarchi. 2019. "Single-Cell Analysis of Circulating Tumor Cells: How Far Have We Come in the -Omics Era?" *Frontiers in Genetics* 0. <https://doi.org/10.3389/fgene.2019.00958>.
- Silveira, Amanda Bortolini, François-Clément Bidard, Marie-Laure Tanguy, Elodie Girard, Olivier Trédan, Coraline Dubot, William Jacot, et al. 2021. "Multimodal Liquid Biopsy for Early Monitoring and Outcome Prediction of Chemotherapy in Metastatic Breast Cancer." *Npj Breast Cancer*. <https://doi.org/10.1038/s41523-021-00319-4>.
- Siravegna, Giulia, Silvia Marsoni, Salvatore Siena, and Alberto Bardelli. 2017. "Integrating Liquid Biopsies into the Management of Cancer." *Nature Reviews Clinical Oncology*. <https://doi.org/10.1038/nrclinonc.2017.14>.
- Van der Auwera, Geraldine A., Mauricio O. Carneiro, Christopher Hartl, Ryan Poplin, Guillermo Del Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. 2013. "From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline." *Current Protocols in Bioinformatics / Editorial Board, Andreas D. Baxevanis ... [et Al.]* 43: 11.10.1–11.10.33.
- Wang, Kai, Mingyao Li, and Hakon Hakonarson. 2010. "ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data." *Nucleic Acids Research* 38 (16): e164.
- Xu, Lei, Xueying Mao, Ahmet Imrali, Ferrial Syed, Katherine Mutsvangwa, Daniel Berney, Paul Cathcart, John Hines, Jonathan Shamash, and Yong-Jie Lu. 2015. "Optimization and Evaluation of a Novel Size Based Circulating Tumor Cell Isolation System." *PloS One* 10 (9): e0138032.