USCAP

TECHNICAL REPORT

Reliability of liquid biopsy analysis: an inter-laboratory comparison of circulating tumor DNA extraction and sequencing with different platforms

Thibaud Koessler[1] · Viola Paradiso[2] · Salvatore Piscuoglio [2,3] · Ronny Nienhold[4] · Liza Ho[5] · Yann Christinat[5] · Luigi M. Terracciano[2] · Gieri Cathomas[3] · Andreas Wicki [6] · Thomas A. McKee[5] · Thierry Nouspikel[7]

Abstract

Liquid biopsy, the analysis of circulating tumor DNA (ctDNA), is a promising tool in oncology, especially in personalized medicine. Although its main applications currently focus on selection and adjustment of therapy, ctDNA may also be used to monitor residual disease, establish prognosis, detect relapses, and possibly screen at-risk individuals. CtDNA represents a small and variable proportion of circulating cell-free DNA (ccfDNA) which is itself present at a low concentration in normal individuals and so analyzing ctDNA is technically challenging. Various commercial systems have recently appeared on the market, but it remains difficult for practitioners to compare their performance and to determine whether they yield comparable results. As a first step toward establishing national guidelines for ctDNA analyses, four laboratories in Switzerland joined a comparative exercise to assess ccfDNA extraction and ctDNA analysis by sequencing. Extraction was performed using six distinct methods and yielded ccfDNA of equally high quality, suitable for sequencing. Sequencing of synthetic samples containing predefined amounts of eight mutations was performed on three different systems, with similar results. In all four laboratories, mutations were easily identified down to 1% allele frequency, whereas detection at 0.1% proved challenging. Linearity was excellent in all cases and while molecular yield was superior with one system this did not impact on sensitivity. This study also led to several additional conclusions: First, national guidelines should concentrate on principles of good laboratory practice rather than recommend a particular system. Second, it is essential that laboratories thoroughly validate every aspect of extraction and sequencing, in particular with respect to initial amount of DNA and average sequencing depth. Finally, as software proved critical for mutation detection, laboratories should validate the performance of variant callers and underlying algorithms with respect to various types of mutations.

Introduction

Liquid biopsies, particularly those involving circulating cell-free DNA (ccfDNA) from plasma, are rapidly emerging as an important and minimally invasive adjunct to standard tumor biopsies and, in some situations, an alternative approach [1].

In oncology patients, ccfDNA released from tumor cells was demonstrated in the late 1970's [2] and is referred to as circulating tumor DNA (ctDNA). Clinical applications of the liquid biopsy approach are currently being actively investigated in localized and advanced disease stages and both before and after treatment. Prior to treatment for localized

---

✉ Thierry Nouspikel
thierry.nouspikel@hcuge.ch

1 Department of Oncology, Geneva University Hospitals, Geneva, Switzerland

2 Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland

3 Visceral surgery research laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland

4 Institute of Pathology, Cantonal Hospital Basel-Land, Liestal, Switzerland

5 Clinical Pathology Service, Diagnostic Department, Geneva University Hospitals, Geneva, Switzerland

6 Department of Oncology & Hematology, Medical University Clinic, Cantonal Hospital Basel-Land, Liestal, and University of Basel, Basel, Switzerland

7 Medical Genetics, Diagnostic Department, Geneva University Hospitals, Geneva, Switzerland

Table 1 Extraction parameters.

Lab	Kit	Principle	Plasma used (ml)	DNA yield (ng/ml plasma)	Elution volume (µl)	Concentration (ng/µl eluate)	305 bp/41 bp ratio
A	MagMAX	Magnetic beads	8.00	3.00	18	1283	0.12
B	QiaAmp	Vacuum columns	4.00	2.60	30	333	0.15
C	Avenio	Spin columns	3.80	3.54	65	188	0.24
C	Avenio	Spin columns	2.20	4.62	65	142	0.24
D	MinElute	Magnetic + spin	3.80	2.07	30	252	0.11
D	Cobas	Spin columns	4.00	2.19	90	93	0.12
D	Symphony	Magnetic beads	3.75	3.75	75	180	0.13

disease, the principal aim is early detection [3, 4] while in advanced disease, liquid biopsy can be used for molecular profiling [5] including determination of tumor mutation burden [6–8]. Following treatment this technique can be used to monitor response [9, 10], identification of resistance mechanisms [11–13], monitoring of clonal dynamics [14, 15], and measurement of residual disease [16].

The key technical challenges in the detection of ctDNA relate to its low abundance, its quantity relative to ccfDNA and potential contamination with normal DNA released by leukocyte lysis [17]. Both ctDNA and ccfDNA are rapidly cleared from the blood stream (with a half-life of an hour or less) [18] and the fraction of ctDNA within ccfDNA can vary from 0.1% to 90% [3, 15] depending on the tumor cell burden, tumor type, and other factors [3, 19]. To overcome these issues, specialized ccfDNA collection tubes are available, containing fixatives that can stabilize both ccfDNA and intact cells for up to 7–14 days at room temperature, allowing for easier shipping, storage, and batched processing of blood samples, as verified by us and others [20, 21]. Finally, ultrasensitive methods are required to detect mutations, copy-number changes, or other alterations that are present in ccfDNA at very low variant-allele frequencies. Advantages and drawbacks of these techniques have been reviewed by us and others [1, 22]. Detection and quantification of ctDNA is a rapidly evolving field in which significant technical challenges exist and where no standardization or generally accepted standard operating procedures (SOPs) have yet been developed. In an effort to highlight critical steps and produce the first Swiss ctDNA SOP recommendation, we compared different extraction, purification, and analysis procedures from three Swiss institutions offering liquid biopsy in a diagnostic context.

Four laboratories participated in the study: Lab A, Institute of Pathology, Cantonal Hospital Basel-Land; Lab B, Molecular Pathology Unit, Institute of Medical Genetics and Pathology, University Hospital Basel; Lab C, Clinical Pathology at Geneva University Hospitals and Lab D, Medical Genetics at Geneva University Hospitals.

The first part of the study addressed ccfDNA extraction: a common blood sample from a healthy donor was provided to all participating laboratories, to be extracted and sequenced. A second part focused specifically on sequencing: aliquots of a commercial control DNA kit containing various amounts of eight known mutations were provided to all laboratories for library preparation, sequencing, and bioinformatic analysis.

Methods

Plasma preparation

Blood from a single healthy donor was sampled into several Streck BCT tubes (Streck) that were kept at room temperature and dispatched the same day by mail to participating laboratories. All samples were received the next day and processed ~24 h after sampling. Plasma was isolated by centrifugation for 10 min at $1600 \times g$, the upper phase was collected and centrifuged again 10 min at $16,000 \times g$.

DNA extraction

DNA was extracted by various methods, following the manufacturers' instructions: MagMAX Cell-Free DNA Isolation Kit (Applied Biosystems), QIAamp circulating nucleic acid kit with QIAvac 24 Plus vacuum aspiration system (Qiagen), Avenio cfDNA isolation kit (Roche), MinElute (Qiagen), Cobas cfDNA SP kit (Roche), and QiaSymphony robot with DSP circulating DNA kit (Qiagen). See Table 1 for details on plasma input and final elution volumes. DNA concentration was appraised by fluorescence with the Qubit high-sensitivity kit (Thermo-Fisher). DNA size profiles were obtained by running 1 µl samples on BioAnalyzer chips (Agilent) or 2 µl samples on a TapeStation (Agilent) for Lab B. DNA quality and the absence of contamination by genomic DNA was verified by differential amplicon length PCR, using the Kapa hgDNA Quantification and QC kit (Roche) to determine the 305 bp/41 bp ratio (normal range: 0.10–0.25) [20].

Sequencing

Three different library kits and panels were used, according to the manufacturers' instructions: Oncomine Lung cfDNA

Table 2 Library characteristics and sequencing parameters.

Lab	Library	Principle	Genes (exons)	Size	DNA input	Sequencer	Read depth	Molecular depth[c]
A	Oncomine Lung Ca	Multiplex PCR	11 (35)[a]	1.8 kb	30 ng	Ion Torrent	100,000	3000
B	Oncomine Lung Ca	Multiplex PCR	11 (35)[a]	1.8 kb	10 ng	Ion Torrent	40,000	1000
C	Avenio Expanded	Capture by hybridization	71 (811)	170 kb	50 ng	Illumina	30,000	5000
D	QiaSeq Lung Ca	Primer extension	72 (1537)	500 kb	20 ng	Illumina	2000[b] 15,000	1500

[a]With Oncomine, the 35 amplicons do not correspond to full exons.

[b]Lab D sequenced the 0.1% sample at 15,000 the others samples at 2000.

[c]Average depth in molecular barcodes (number of distinct barcodes at a given position).

Assay (ThermoFisher), Avenio ctDNA expanded kit (Roche), and QIAseq human lung cancer panel (Qiagen). Oncomine libraries were sequenced on an Ion S5XL system (Life technologies), Avenio and QIAseq libraries on a NextSeq 500 sequencer (Illumina). See Table 2 for details on DNA input and sequencing parameters.

Bioinformatic analysis

Data were analyzed with software packages provided by the manufacturers: Ion Reporter for Oncomine, Avenio ctDNA Analysis Software for Avenio, and smCounter2 [23] for QIAseq. A custom whitelist caller based on Poisson's law was implemented in Excel as an alternative to smCounter2. For direct examination of the aligned reads, we used bamreadcount v.0.7.4 (GitHub) to extract data at the eight relevant positions, allele frequencies for the four nucleotides plus indels were calculated with Excel.

Results

Evaluating extraction

To compare extraction methods, aliquots from a single blood draw from the same healthy donor were provided in Streck BCT (Streck) tubes to participating laboratories, who extracted ccfDNA using their usual technique (Table 1). Lab A used the MagMAX Cell-Free DNA Isolation Kit (Applied Biosystems), which is based on magnetic beads. Lab B used a column-based system, QIAamp circulating nucleic acid kit (Qiagen). Lab C extracted two aliquots in parallel, using the Avenio spin column system (Roche). Lab D extracted three tubes using three distinct methods: the Cobas spin column system (Roche), the Qia-Symphony extraction robot (Qiagen) which uses magnetic beads, and the hybrid MinElute system (Qiagen) which comprises an initial volume reduction step using magnetic beads, followed with DNA purification on spin columns.

Figure 1 illustrates some of the various output parameters monitored. The yield in DNA (top left panel) was roughly similar in all systems, and relatively low: around 3 ng/ml plasma. Of note, since plasma input and elution volume varied depending on the system, the final DNA concentration was very different between samples: ranging from 93 ng/µl eluate with the Cobas system to 1283 ng/µl with the MagMAX kit (Table 1).

Importantly, the electrophoretic profiles of all samples revealed a clear peak around 170 bp, corresponding to ccfDNA, and no detectable high-molecular weight DNA indicative of contamination with cellular DNA (Fig. 1, right panel). To completely rule out the presence of cellular DNA, we made use of a PCR-based system relying on differential amplicon length. The premise of the method is that, due to the small size of ccfDNA, a short amplicon is more efficiently amplified than a larger one. By contrast, both amplicons are amplified to the same extent in the presence of cellular DNA. In our experience [20], the quantification ratio of a 305 bp amplicon over a 41 bp amplicon is expected to be between 0.10 and 0.25 with pure, good quality ccfDNA. Lower values indicate issues with PCR amplification, whereas a ratio higher than 0.25 indicates the presence of cellular DNA. The bottom panel in Fig. 1 demonstrates that all samples fell within normal range.

As a final quality control, all laboratories sequenced an aliquot of the ccfDNA they had extracted (data not shown). Since blood came from a healthy donor there were no mutations to detect, but each laboratory verified that their ccfDNA sample yielded good quality sequencing data, based on coverage, sequencing, and alignment quality metrics, as well as calling of heterozygous SNPs.

Evaluation of sequencing

The second part of the study entailed sequencing reference DNA material (Horizon Discoveries Ltd), which consists of three samples, each containing various amounts of eight well-defined mutations: four *EGFR*
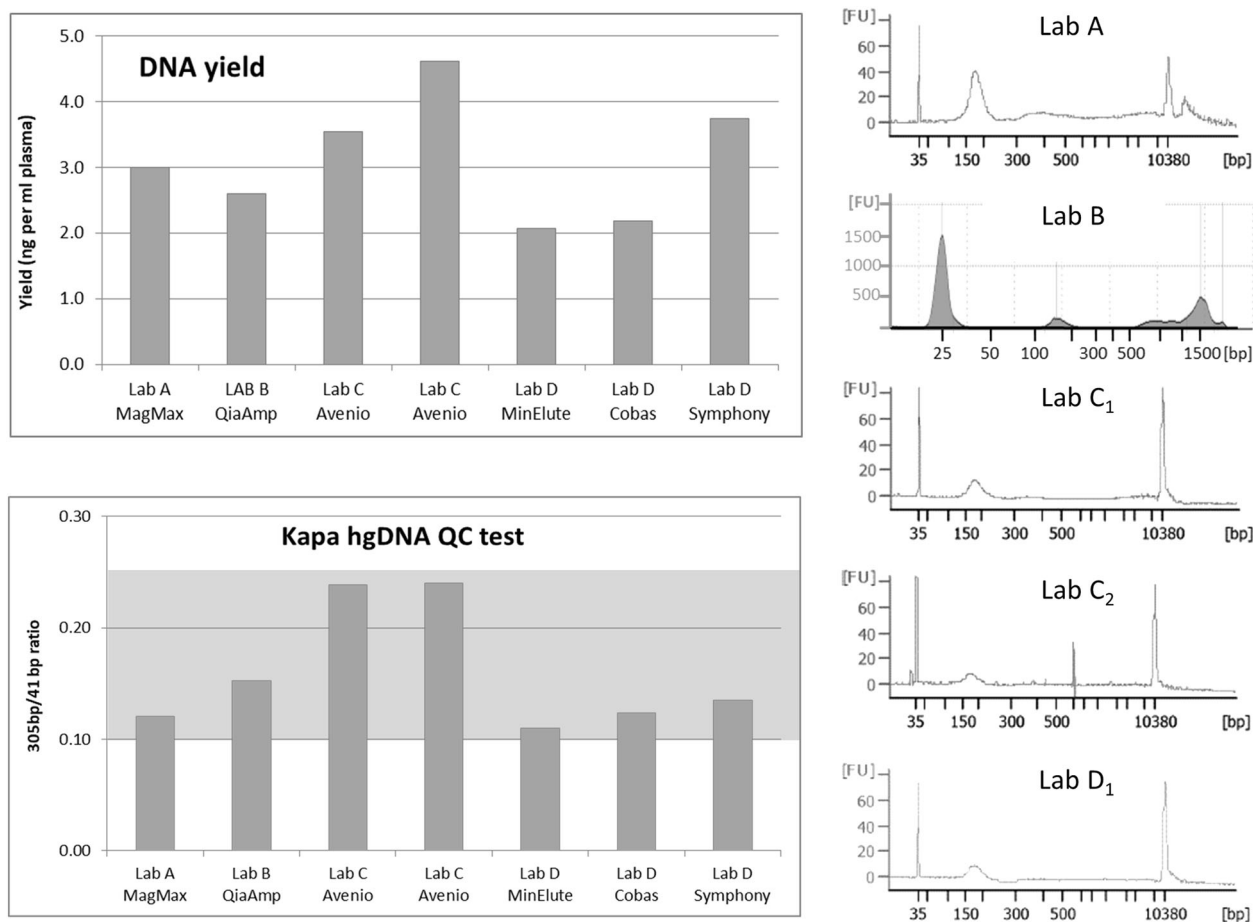
Fig. 1 Extraction results. Blood from a single donor was processed in the four laboratories using different techniques (see text for details). Lab C extracted two aliquots in parallel with the same method. Lab D extracted three aliquots with three distinct methods. Top left: DNA yield expressed as ng DNA per ml of plasma. Bottom left: Quality control by differential amplicon length PCR, the shaded area represents acceptable values. Right panel: DNA electrophoretic profiles, ccfDNA peaks around 170 bp, the sharp peaks at 35 and 10,380 bp (25 and 1500 bp for Lab B) are size markers.

mutations (two substitutions *EGFR*:T790M, *EGFR*: L858R, an insertion *EGFR*:V769_D770insASV, and a deletion *EGFR*: E756_A750 del), provided at 5%, 1%, and 0.1%, and four mutations in other genes (*KRAS:* G12D, *NRAS:*A59T, *NRAS:*Q61K, and *PIK3CA:*E545K) provided at 6.3%, 1.3%, and 0.13%. In addition, a "wild-type" control is included in the kit, corresponding to the background DNA into which these eight mutations are diluted. Of note, this background DNA is not actually wild-type, as it contains significant amounts of several known driver mutations, different from the above. Although the supplier does not detail how reference material is prepared, it is likely that individual mutations have been engineered into (a) cancer cell line(s), (either RKO or SW28, according to the manufacturer) and that DNA from the original and the engineered cell lines are mixed in various proportions. This reference material is thus not genuine ccfDNA, but cellular DNA sonicated to a size similar to that of ccfDNA, about 160 bp, with a

distribution of sizes wider than ccfDNA and a 305/41 bp ratio inferior to 0.1 (data not shown).

Reference material was aliquoted and distributed to participating laboratories who used their established procedures to analyze the aliquots provided (Table 2). Lab A and B both used the Oncomine Lung cfDNA Assay (Life Technologies) sequenced with Ion Torrent technology, albeit at different depths; Lab C used the Avenio system (Roche), sequenced with Illumina technology; Lab D used the QIAseq system (Qiagen), also sequenced with Illumina technology. These library systems all use molecular barcodes (aka molecular tags or unique molecular identifiers) to reduce background error rates by grouping sequence reads that originate from the same ccfDNA molecule. These reads should have the same sequence and any discrepancy within a group can thus be disregarded as an artifact [22].

Oncomine is an amplicon-based system, which has the advantage of maximizing specificity, with the drawback that both primer binding sites must fall within the same ccfDNA

fragment for amplification to take place. This imposes the use of short amplicons, implying that only a small portion of any given ccfDNA fragment will be sequenced. Furthermore, even with a 50 bp target (i.e., a 90–100 bp amplicon, including primers) there is only a 25–30% chance that both primers binding sites lie within a fragment of 170 bp. Molecular barcodes are introduced as a 3′ tail in one of the PCR primers and are used during the first two PCR cycles only. This means that each strand carries a different barcode but there is no way to determine which barcodes were part of the same pair, i.e., originally tagged the same DNA fragment. The library used for this work was the Oncomine Lung cfDNA Assay, which contains 35 short amplicons (from 35 to 94 bp long) located within 11 genes of interest.

The QIAseq system is based on primer extension and therefore only relies on a single gene-specific primer to amplify a given ccfDNA fragment. The other primer used for PCR is part of an adapter that is ligated to the ccfDNA fragment and carries the molecular barcode. The advantage is that only one primer binding site must lie within the target DNA fragment, potentially allowing for amplicons of any size, although only the portion of DNA between the primer and the end of the fragment is actually sequenced. The library used, QIAseq human lung cancer panel, targets 72 genes of interest for lung cancer, with all exons included for every gene.

The Avenio system uses capture by hybridization. Its main advantage is that the entire captured fragment is sequenced, no matter where in the fragment the capture probe might bind. Molecular barcodes are introduced by adapters ligated to either end of the DNA fragments. While this would allow for duplex barcoding (i.e., a different barcode for each strand, as is the case for Avenio tumor tissue libraries), Avenio ctDNA libraries use simplex barcodes and thus cannot distinguish which strand a read was originally amplified from. The library used was the Avenio ctDNA Expanded kit, which targets 71 genes involved in various types of cancer, although not all exons are included for every gene.

For sequencing, it was agreed that each laboratory would use their standard procedure, which implied inter-laboratory variations in several parameters, such as initial amount of DNA or targeted sequencing depth (see Table 2 for details).

Upon analysis, we first leveraged molecular barcode information to estimate the molecular recovery of the various library systems. We compared molecular depth, i.e., the number of distinct molecules that were retrieved by a library system, deduced from the number of barcodes sequenced at a given position, with the theoretical number of molecules in the reaction, calculated from the original amount of DNA. We found that molecular yield was relatively low in all cases: in the order of 20%, except for

Avenio (Lab C) which achieved about 40% recovery by hybridization (Fig. 2).

Next, to evaluate the sequencing performance of the various platforms independently of software issues, we directly sought the eight mutations of interest within aligned sequence reads. All mutations were found with all systems in the three samples, with excellent linearity (Fig. 3). With the QIAseq system, we initially did not find the two *EGFR* indels in the 0.1% sample, and their frequency in the 5% sample was largely underestimated. However, we verified that these two mutations were indeed present in raw data (fastq files, unaligned reads). Their absence from aligned reads was traced down to an alignment problem: due to the position of these two mutations with respect to the nearest library primer, they always appear near the end of a read and the alignment software (BWA v.0.7.17) often failed to map the small portion of data following the indel, effectively "soft-clipping" the mutation from aligned sequences.

Evaluating variant callers

Mutations were then called using the bioinformatics pipelines provided by the manufacturers of the respective libraries. Apart from the QIAseq software that is open source, details of the underlying algorithms are proprietary. Algorithms that use a "whitelist" of predefined mutations can however be distinguished from algorithms that call mutations without prior knowledge. Similarly, some software systems maintain a blacklist of recurrent sequencing errors, which are masked, thereby lowering background noise and improving specificity [22].

As the Oncomine library consists entirely of mutation hotspots, the corresponding software, Ion Reporter, is essentially a whitelist caller. While it can detect other mutations than those included in its whitelist, the user has the option of blacklisting these, to focus on the predefined mutations. The Avenio software combines both strategies: a general "adaptive caller" which models error rates for the 12 possible substitutions and uses a blacklist of 26 genomic positions. In addition, it features a whitelist of over 500 specific mutations called by a dedicated algorithm that combines Poisson distribution with a series of heuristic rules. Six of the eight mutations present in the reference material are part of this whitelist. The remaining two, detectable by the adaptive caller only, are *NRAS*:p.Ala59Thr and the *EGFR* insertion. The software provided for QIAseq libraries, smCounter2 [23], models the error rate using a beta binomial distribution and does not make use of a blacklist or a whitelist. For the purpose of comparison, Lab D implemented a custom whitelist caller based on Poisson distribution, as well as an in-house blacklist collection.
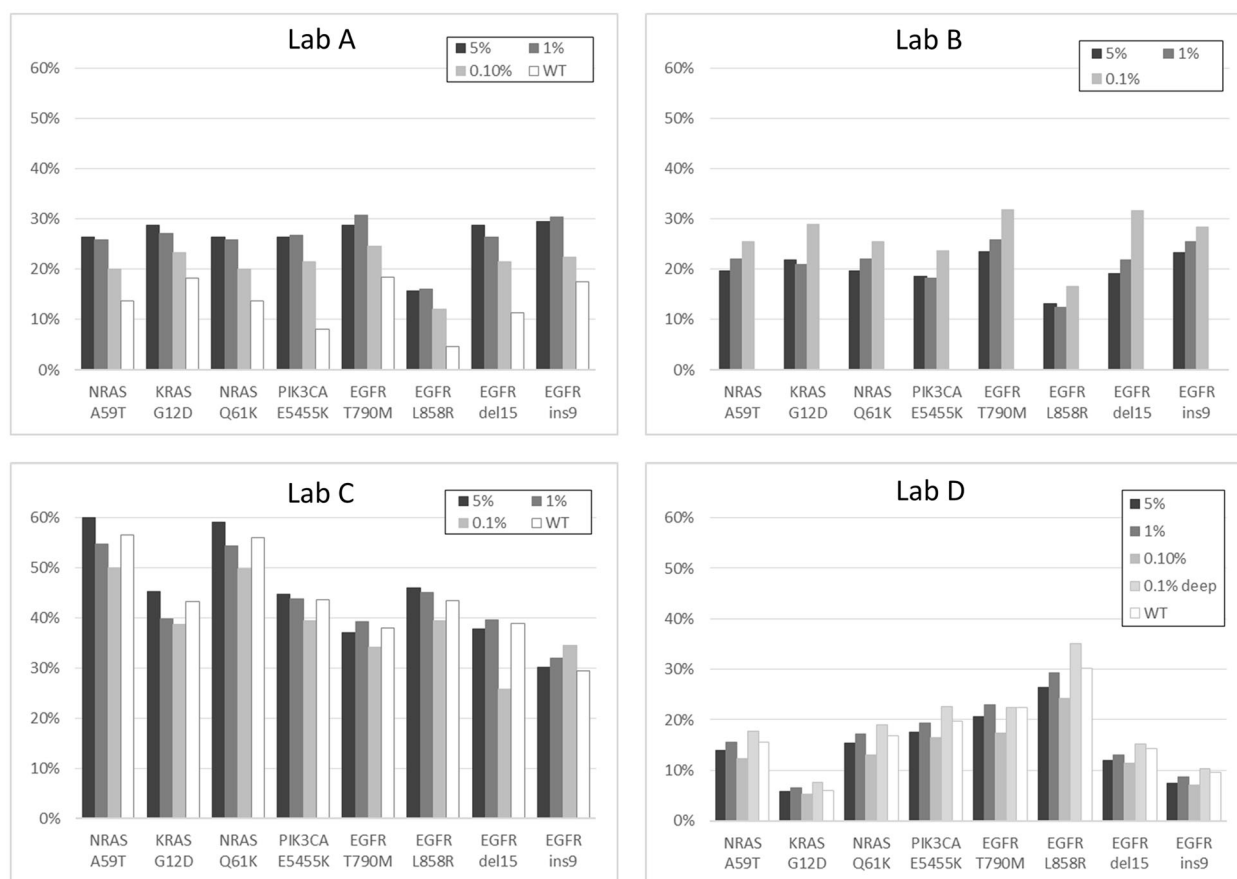
Fig. 2 Molecular recovery. The total number of individual DNA molecules sequenced at the eight positions of interest in the three samples (and the "no mutation" sample WT, except for Lab B) was deduced from the number of barcodes and compared with the value expected from DNA input.

All software systems succeeded in calling all eight mutations in the first two samples. However, none of them achieved 100% sensitivity with the 0.1% sample (Table 3). In the latter, some mutations were called with convincing *p* values, some with *p* values superior to 0.1 (shown in parentheses in the table) and some were filtered out by the various software filters, despite being present in raw data. Of note, no details on how *p* values are calculated were available for commercial software packages, with the exception of smCounter2 (*p* values obtained from a binomial distribution, refer to [23] for details). In the case of QIAseq, the data presented in the table was obtained with the whitelist caller (with *p* values calculated from a cumulative Poisson distribution), as smCounter2 only detected one mutation in the 0.1% sample. Threshold for *p* values (<0.01 and >0.05) were determined empirically, based on true and false positive calls in the "wild-type" control sample. Considering only mutations called with a *p* value inferior to 0.01, the apparent sensitivity of the various methods are: 3/8 for lab A (Oncomine), 5/8 for lab B (Oncomine), 2/8 for lab C (Avenio), and 4/8 for lab D (QIAseq).

Discussion

The purpose of this study was not to validate a particular ctDNA analysis platform, but rather to compare various platforms available in the same country, and ensure that results obtained by different laboratories, using different systems, are equally trustable and comparable.

To this end, we evaluated (1) DNA extraction, (2) library building, (3) DNA sequencing, and (4) bioinformatic analysis among the various platforms used by participating laboratories. In the first phase, ccfDNA was extracted from a single blood draw using six distinct methods, all yielding satisfactory results despite the fact that the donor had unusually low amounts of ccfDNA, making DNA purification more challenging. The quantities of DNA recovered were similar across platforms, and all methods produced DNA suitable for sequencing and free of contamination by high-molecular weight leukocyte DNA. The latter is a critical point with ctDNA analysis, as the presence of cellular DNA from lysed leukocytes further dilutes tumor-derived DNA and makes mutation detection even more challenging [17]. Since elution volumes varied
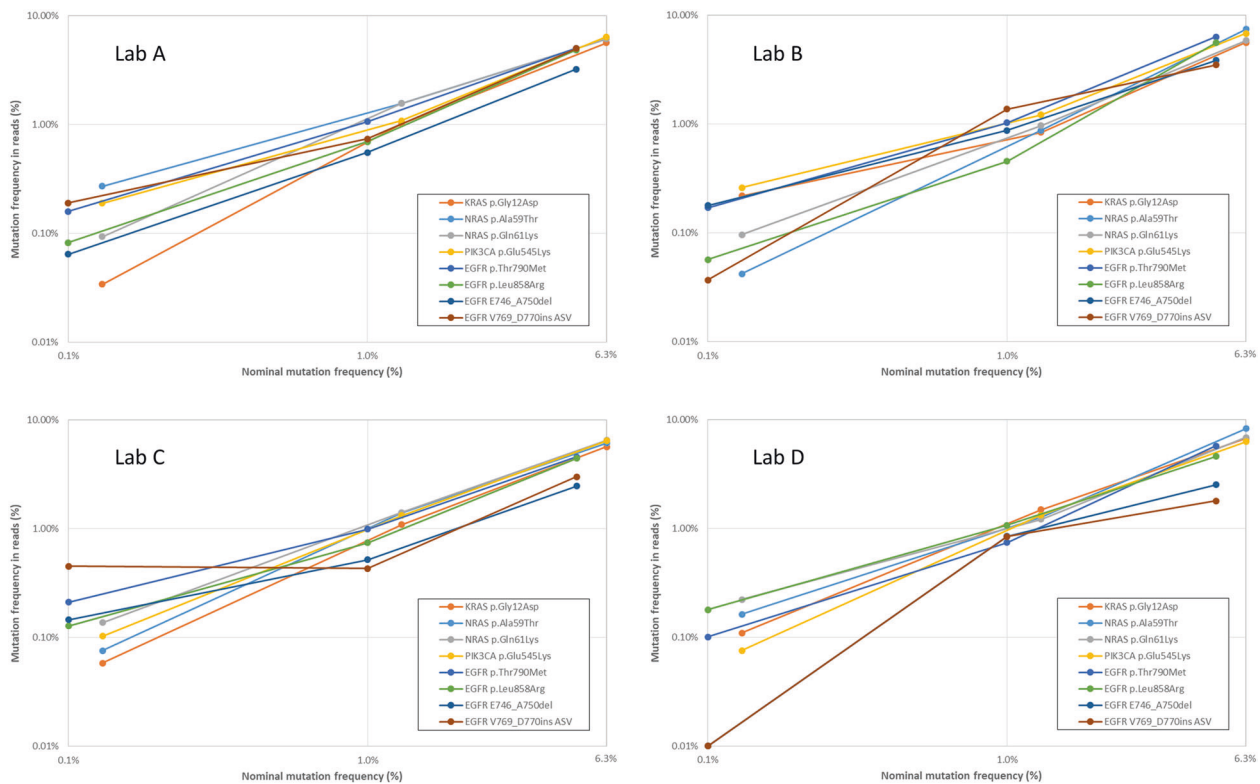
Fig. 3 Linearity of mutation detection. Mutation frequency was appraised in aligned reads for the eight mutations in the three samples of reference material and plotted against the nominal mutation frequencies provided by the manufacturer.

between systems, final DNA concentrations differed widely between laboratories, but this parameter is a matter of preference and has no impact on further analyses.

To dissociate sequencing from extraction, we elected to sequence reference material consisting of cellular DNA, fragmented to a size similar to that of ccfDNA and containing precise amounts of eight well-defined mutations. To isolate the purely molecular aspects of library building and sequencing from bioinformatics analysis, we first directly checked for the presence of mutations in aligned sequenced reads. All mutations were found at the expected frequencies in all samples, with the three library systems tested. With the QIAseq system, initially we did not observe the two *EGFR* insertion and deletion mutations in aligned reads, but we were able to confirm their presence in unaligned reads. The failure to detect these mutations was thus a software issue, rather than a problem with the library, and the use of a different alignment software might resolve this problem.

The presence of low percentage mutations in sequence reads does not necessarily mean that they will be identified by an analysis software: mutations at low percentage tend to be hidden in the background of PCR errors and sequencing mistakes. Numerous methods have been described to reduce background and discriminate signal (in this case actual mutations) from background [24, 25]. Few of these,

however, can deal with signals that are as low as the background. The use of molecular barcodes allows for considerable reduction of background by building a consensus sequence from all reads bearing the same barcode and disregarding individual discrepancies [26]. This strategy is not perfect though, since it cannot detect first-cycle PCR mistakes. In addition, it requires sequencing depth to be high enough so that there is a minimum of 3 reads per barcode, to allow error correction via a majority rule. Other algorithms might be used to leverage molecular barcode information, for instance by first correcting errors within a group, then across groups with overlapping regions [25]. These algorithms, however, likely suffer from similar limitations. For the commercial software packages we used, algorithm details are proprietary and we were not able to ascertain how barcode information is used to reduce error rate.

Even with molecular barcodes, detection of low-frequency mutations remains technically challenging, no matter what algorithm is used. Strategies to improve sensitivity without reducing specificity include lowering the detection threshold for a whitelist of mutation of interest, and/or maintaining a blacklist of genomic positions prone to recurrent errors, to be disregarded in the analysis. Implementing these strategies allowed detection of all mutations

Table 3 Variant calling results.

Frequency	Mutation	Lab A	Lab B	Lab C	Lab D
6.3%	KRAS G12D	6.4%	6.1%	5.5%	6.9%
	NRAS A59T	6.6%	7.6%	5.9%	7.2%
	NRAS Q61K	5.9%	5.7%	5.9%	5.9%
	PIK3CA E545K	6.2%	7.2%	6.2%	5.9%
5.0%	EGFR T790M	5.1%	5.8%	4.6%	5.4%
	EGFR L858R	5.1%	6.0%	4.4%	4.5%
	EGFR E756_A750 del	5.1%	5.3%	4.4%	2.9%
	EGFR V769_D770insASV	4.7%	2.9%	4.7%	1.5%
1.3%	KRAS G12D	1.05%	1.00%	1.06%	1.71%
	NRAS A59T	1.29%	1.00%	1.24%	1.30%
	NRAS Q61K	1.55%	1.00%	1.29%	0.86%
	PIK3CA E545K	1.15%	1.20%	1.11%	1.34%
1.0%	EGFR T790M	0.94%	0.90%	0.98%	0.80%
	EGFR L858R	0.84%	0.70%	0.60%	0.76%
	EGFR E756_A750 del	0.84%	1.40%	1.30%	0.86%
	EGFR V769_D770insASV	0.63%	1.10%	1.00%	0.43%
0.13%	KRAS G12D	**No Call**	0.30%	**No Call**	0.01%
	NRAS A59T	0.15%	**No Call**	**No Call**	0.10%
	NRAS Q61K	**(0.09%)**	0.10%	**No Call**	0.29%
	PIK3CA E545K	0.17%	0.30%	**No Call**	**(0.09%)**
0.10%	EGFR T790M	**(0.13%)**	0.10%	0.10%	0.08%
	EGFR L858R	**(0.15%)**	**No Call**	**No Call**	**(0.05%)**
	EGFR E756_A750 del	0.11%	**(0.06%)**	0.11%	**No Call**
	EGFR V769_D770insASV	**(0.11%)**	0.10%	**No Call**	**No Call**

Values in parentheses indicate variants called with a p value >0.01.

No Call: variant absent from reads, p value >0.05 or variant rejected by custom software filters.

Results in bold typeface were considered negative for sensitivity calculations.

with all platforms down to a frequency of 1%, but at 0.1% several mutations escaped detection or were called with nonsignificant p values. Incidentally, we verified that custom implementation of a whitelist algorithm considerably improved detection performance. Finally, mutations successfully called at 0.1% frequency differed between samples, indicating that non-detection is probably not due to the nature of a mutation, but rather that detection is a highly challenging exercise and that mutations may be called or missed in a stochastic manner. For this reason, sensitivity values based solely on these eight calls may not be the best way to compare distinct platforms. Overall, it can be concluded that all systems displayed poor automated detection performance at mutation frequencies approaching 0.1%, despite the fact that all mutations were present in raw data at the expected frequency.

We were disappointed by this outcome, since 0.1% has been suggested the clinically relevant threshold for some applications in oncology, such as early detection of resistance-causing mutations [27], follow-up of minimal residual disease [28], and possibly appraisal of plasma mutation burden [8]. Although it is likely that in many

clinical situations the most important factor is not the absolute mutation frequency, but its evolution over time, it would be important to improve mutation detection at low frequencies. Several factors may contribute to the suboptimal performance we observed: First and foremost, the limited amount of ccfDNA in plasma implies that, at low mutation frequencies, very few mutant molecules are actually present in a given sample. For instance, the blood sample used in the first part of this study yielded ~3 ng ccfDNA per ml plasma. Since 1 ng DNA corresponds to about 270 copies of the human genome, this sample contained ~810 copies per ml plasma. As a 10 ml blood tube yields ~5 ml plasma, if this sample had contained a mutation with a frequency of 0.1%, there would have been only four mutated DNA fragments per tube. While an obvious solution to this problem is to increase the amount of blood drawn, there are practical limits to this strategy, especially in cancer patients.

Another consideration is that the library system itself should not be limiting in the initial amount of DNA, i.e., it should allow libraries to be constructed from 50 ng DNA, where this quantity is available. On the other hand, while it

is desirable to use as much DNA as possible to maximize sensitivity, DNA may be limiting in daily practice. It is thus important that laboratories validate the minimum and maximum amounts of input DNA that are compatible with a given library system, and remain aware of the sensitivity expected with a given quantity of input DNA.

A compounding factor is the low molecular recovery we observed with all library construction systems: from 20% to 40% depending on the system. This implies that only one or two of the aforementioned four molecules would actually be sequenced. There is obviously considerable margin for improvement here, and we can only hope that researchers and manufacturers will succeed in producing more efficient library systems in the future.

Mutation calling, i.e., the design of computer algorithms that can reliably distinguish low-frequency mutations from PCR and sequencing errors, represents a further challenge. Molecular barcodes can help to reduce background, but only if sequencing depth is sufficient. Here, the size of the targeted region, combined with the depth of sequencing, determines the sequencing power required. In this study, the Oncomine library had the smallest target size, 1.8 kb, thereby allowing very high sequencing depth at a moderate cost: Lab A targeted a 100,000-average read depth and Lab B 40,000. As a note of caution, overly deep sequencing can be counterproductive since, at very high depth, sequence errors begin to accumulate in the barcodes themselves. This can cause the software to create inexistent ccfDNA "fragments" and paradoxically increases error rate. At the other extreme, Lab D used a library with a large footprint (500 kb) and sequenced it at an average read depth of 2000. This proved insufficient to provide enough reads per barcode for error correction and the 0.1% sample had to be sequenced again with a depth of 15,000 to improve sensitivity.

Finally, software tools such as blacklists and whitelists significantly improve automated mutation detection but, at least in the conditions we tested, did not allow fully reliable detection of mutations at frequencies approaching 0.1%. Here too, there is room for improvement: new mutation callers are released every year [29], based on increasingly sophisticated algorithms such as neural networks, and our ability to discriminate mutations from background noise is likely to improve significantly in the next few years, thus it is important that laboratories who provide ctDNA analysis services remain aware of these developments.

In summary, our pilot study allowed us to identify several key parameters that require validation in laboratories performing ctDNA analysis: first the minimum (and maximum) amount of input ccfDNA, which has a significant impact on sensitivity. Second, an optimal sequencing depth is required to derive a maximum benefit from the use of molecular barcodes. Third, complete analytical software pipeline is key to successful mutation detection, from the initial alignment step to the validation of significant mutations. In general, the software package provided by the library manufacturer is tailored to the specific library system and thus performs better than open source solutions, but this is not always the case and it is worth investigating alternative software solutions before settling on one.

We also established that, among the methods we tested, none is clearly superior or inferior to the others. Each has specific advantages and drawbacks: Avenio achieves higher molecular recovery, Oncomine allows higher sequencing depth at limited costs, while QIAseq features the largest target size and the option to add custom target regions. Yet, when it comes to critical parameters such as specificity, sensitivity, and linearity, all systems performed equally well and suffered from the same limitations with automated identification of low-frequency mutations.

The fact that all platforms were equally efficient is reassuring for the clinician, as it means that results from different laboratories can be safely compared, down to at least 1% mutation frequency, which should be sufficient for most clinical applications. From a policy point of view, it also means that there is no need to impose one system rather than another in national guidelines, as long as the chosen system is properly validated onsite.

Compliance with ethical standards

References

1. Corcoran RB, Chabner BA. Application of cell-free DNA analysis to cancer treatment. N Engl J Med. 2018;379:1754–65.
2. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977;37:646–50.
3. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014; 6:224ra24.
4. Newman AM, Bratman SV, To J, Wynne JF, NCW Eclov, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20:548–54.
5. De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CKY, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. Ann Oncol. 2014;25:1729–35.
6. Zhang Y, Chang L, Yang Y, Fang W, Guan Y, Wu A, et al. The correlations of tumor mutational burden among single-region tissue, multi-region tissues and blood in non-small cell lung cancer. J Immunother Cancer. 2019;7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6448263/.

7. Fenizia F, Pasquale R, Roma C, Bergantino F, Iannaccone A, Normanno N. Measuring tumor mutation burden in non-small cell lung cancer: tissue versus liquid biopsy. Transl Lung Cancer Res. 2018;7:668–77.

8. Friedlaender A, Nouspikel T, Christinat Y, Ho L, McKee T, Addeo A. Tissue-plasma TMB comparison and plasma TMB monitoring in patients with metastatic non-small cell lung cancer receiving immune checkpoint inhibitors. Front Oncol. 2020;10. https://www.frontiersin.org/articles/10.3389/fonc.2020.00142/full.

9. Del ReM, Vivaldi C, Rofi E, Vasile E, Miccoli M, Caparello C, et al. Early changes in plasma DNA levels of mutant KRAS as a sensitive marker of response to chemotherapy in pancreatic cancer. Sci Rep. 2017;7:7931.

10. Carpinetti P, Donnard E, Bettoni F, Asprino P, Koyama F, Rozanski A, et al. The use of personalized biomarkers and liquid biopsies to monitor treatment response and disease recurrence in locally advanced rectal cancer after neoadjuvant chemoradiation. Oncotarget. 2015;6:38360–71.

11. Sakai H, Tsurutani J, Iwasa T, Komoike Y, Sakai K, Nishio K, et al. HER2 genomic amplification in circulating tumor DNA and estrogen receptor positivity predict primary resistance to trastuzumab emtansine (T-DM1) in patients with HER2-positive metastatic breast cancer. Breast Cancer. 2018;25:605–13.

12. Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun. 2016;7:11815.

13. Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat Med. 2015;21:560–2.

14. O'Leary B, Hrebien S, Morden JP, Beaney M, Fribbens C, Huang X, et al. Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. Nat Commun. 2018;9:896.

15. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14:985–90.

16. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med. 2016;8:346ra92.

17. Fernando MR, Chen K, Norton S, Krzyzanowski G, Bourne D, Hunsley B, et al. A new methodology to preserve the original proportion and integrity of cell-free fetal DNA in maternal plasma during sample processing and storage. Prenat Diagn. 2010;30:418–24.

18. Reinert T, Schøler LV, Thomsen R, Tobiasen H, Vang S, Nordentoft I, et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. Gut. 2016;65:625–34.

19. Kohler C, Radpour R, Barekati Z, Asadollahi R, Bitzer J, Wight E, et al. Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors. Mol Cancer. 2009;8:105.

20. Nikolaev S, Lemmens L, Koessler T, Blouin J-L, Nouspikel T. Circulating tumoral DNA: preanalytical validation and quality control in a diagnostic laboratory. Anal Biochem. 2018;01:34–9.

21. Alidousty C, Brandes D, Heydt C, Wagener S, Wittersheim M, Schäfer SC, et al. Comparison of blood collection tubes from three different manufacturers for the collection of cell-free DNA for liquid biopsy mutation testing. J Mol Diagn. 2017;19:801–4.

22. Koessler T, Addeo A, Nouspikel T. Implementing circulating tumor DNA analysis in a clinical laboratory: a user manual. Adv Clin Chem. 2019;89:131–88.

23. Xu C, Gu X, Padmanabhan R, Wu Z, Peng Q, DiCarlo J, et al. smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. 2018;281659. https://www.biorxiv.org/content/10.1101/281659v1.

24. Lee B, Moon T, Yoon S, Weissman T. DUDE-Seq: fast, flexible, and robust denoising for targeted amplicon sequencing. PLOS ONE. 2017;12:e0181463.

25. Zhao L, Chen Q, Li W, Jiang P, Wong L, Li J. MapReduce for accurate error correction of next-generation sequencing data. Bioinformatics. 2017;33:3844–51.

26. Peng Q, Vijaya Satya R, Lewis M, Randad P, Wang Y. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. BMC Genom. 2015;16:589.

27. Vidal J, Muinelo L, Dalmases A, Jones F, Edelstein D, Iglesias M, et al. Plasma ctDNA RAS mutation analysis for the diagnosis and treatment monitoring of metastatic colorectal cancer patients. Ann Oncol. 2017;28:1325–32.

28. Tie J, Cohen JD, Wang Y, Christie M, Simons K, Lee M, et al. Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. JAMA Oncol. 2019;5:1710–7.

29. Xu C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. Comput Struct Biotechnol J. 2018;16:15–24.