Gene 928 (2024) 148771

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Assessing the efficacy of an innovative diagnostic method for identifying 5 % variants in somatic ctDNA

Chiara Mareso [a], Luca Crosta [a], Mattia Giovanna De Vita [a], Francesca Cristofoli [a], Benedetta Tanzi [b], Sabrina Benedetti [b], Gabriele Bonetti [b,c,*], Carmine Antonio Donofrio [d,e], Marika Cominetti [d], Lucia Riccio [d], Antonio Fioravanti [d], Daniele Generali [f], Emanuela Lucci Cordisco [g,h], Pietro Chiurazzi [g,h], Valentina Gatta [i,j], Liborio Stuppia [i,j], Stefano Cecchin [b], Matteo Bertelli [a,b,k], Giuseppe Marceddu [a]

[a] MAGI EUREGIO, 39100 Bolzano, Italy
[b] MAGI'S LAB, 38068 Rovereto (TN), Italy
[c] Department of Pharmaceutical Sciences, University of Perugia, 06123 Perugia, Italy
[d] Neurosurgery, ASST Cremona, 26100 Cremona, Italy
[e] Division of Biology and Genetics, Department of Molecular and Translational Medicine, University of Brescia, 25123 Brescia, Italy
[f] Dipartimento Universitario Clinico di Scienze Mediche, Chirurgiche e della Salute, Università degli Studi di Trieste, 34127 Trieste, Italy
[g] Istituto di Medicina Genomica, Università Cattolica del Sacro Cuore, 00168 Rome, Italy
[h] UOC Genetica Medica, Fondazione Policlinico Universitario "A. Gemelli" IRCCS, 00168 Rome, Italy
[i] Department of Psychological Health and Territorial Sciences, School of Medicine and Health Sciences, G. d'Annunzio University, 66100 Chieti, Italy
[j] Unit of Molecular Genetics, Center for Advanced Studies and Technology (CAST), "G. d'Annunzio" University of Chieti-Pescara, 66100 Chieti, Italy
[k] MAGISNAT, Atlanta Tech Park, Peachtree Corners, 30092 GA, USA

ARTICLE INFO

Keywords:
Liquid biopsy
ctDNA
cfDNA
Cancer
NGS

ABSTRACT

Background: Liquid biopsy is considered a complementary and recently also an alternative method to surgical biopsy. It allows for the acquisition of valuable information regarding the potential presence of tumors, particularly through the analysis of circulating tumor DNA (ctDNA). CtDNA is a fraction of circulating free DNA (cfDNA) that can be extracted from various tissues, with blood being the most readily available.
Results: To maximize the yield of plasma separation, specific Streck tubes are recommended for blood collection. The MagPurix CFC DNA Extraction Kit can be used for cfDNA extraction, and the TWIST Library Preparation protocol can be optimized for further analysis. Next-generation sequencing (NGS) can be employed to compare somatic and germline lineages, enabling the identification of somatic variants with a Variant Allele Frequency (VAF) of 5 % or higher, which are absent in the germline lineage.
Conclusion: This analysis helps in the assessment of recurrence, analysis, and monitoring of cancer tissue.

1. Introduction

Liquid biopsy, an alternative to tissue biopsy, is a technique used to analyze biological liquid tissues, in most cases blood, to obtain important diagnostic and prognostic information, as well as to predict potential targeted anti-tumor therapies (Poulet et al., 2019). Circulating free DNA (cfDNA) can be isolated from peripheral blood and may contain circulating tumor DNA (ctDNA). The ratio of cf/ct DNA can vary depending on the time of sample collection and the patient's clinical condition; ctDNA is a very small fraction of cfDNA which can be extracted from plasma. This technique offers several advantages compared to tissue biopsy. For example, it is a minimally invasive procedure, allowing for repeated analysis over time to monitor disease progression or recurrence. However, there are also some limitations, such as the difficulty in extracting cfDNA and its half-life, which is about two and a half hours, thus making extraction time very tight.

Abbreviations: cfDNA, Circulating free DNA; ctDNA, Circulating tumor DNA; NGS, Next-Generation Sequencing; TP, True positives; FN, False negatives; TN, True negatives; FP, False positives.

* Corresponding author at: MAGI'S LAB, Via delle Maioliche, 57D, Rovereto (TN), Italy.
E-mail address: gabriele.bonetti@assomagi.org (G. Bonetti).

https://doi.org/10.1016/j.gene.2024.148771
Received 10 January 2024; Received in revised form 5 July 2024; Accepted 15 July 2024
Available online 19 July 2024
0378-1119/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

The main applications of liquid biopsy are identifying predictive factors in patients with advanced disease and detecting minimal residual disease in patients undergoing radical surgery (Nikanjam et al., 2022a).

As mentioned, cfDNA can be extracted from various biological fluids; the most standardized procedures involve isolating cfDNA from peripheral blood. The amount of cfDNA that can be extracted from peripheral blood is often limited, usually in the range of a few nanograms per millilitre, with ctDNA representing only a fraction of it. The concentration of target DNA in plasma depends on several factors, including the disease burden, levels of mutation expression in primary tumor cells, the rate of ctDNA release into the bloodstream and levels of DNA released from non-transformed cells (such as due to inflammatory processes in healthy tissue surrounding the tumor or leukocyte lysis after blood collection). Therefore, careful control of the pre-analytical phase is essential (Poulet et al., 2019; Nikanjam et al., 2022a). While cfDNA can be isolated from either serum or plasma, plasma is preferred due to the coagulation process, which releases genomic DNA from leukocytes.

Liquid biopsy, in combination with next-generation sequencing (NGS), can be utilized to analyze circulating tumor DNA in the patient's biological fluids, enabling the identification of tumor-specific mutations. Indeed, thanks to NGS analysis, cfDNA is sequenced and compared to the patient's germline DNA to identify specific mutations present in ctDNA (Chen and Zhao, 2019). NGS allows the simultaneous analysis of a large number of genes and the identification of rare mutations. This enables the early diagnosis of tumors, even in the absence of other symptoms or when tissue biopsy has been unsuccessful (Chen and Zhao, 2019). This method also allows for the identification of somatic variants, which are genetic mutations not present in the patient's germline DNA. Somatic variants in tumors result from the accumulation of genetic mutations over time, influenced by environmental factors and other factors. Therefore, the combination of liquid biopsy technology and NGS offers an innovative and highly sensitive solution for the diagnosis and monitoring of malignant tumors, providing the opportunity for personalized and specific treatments for each patient (Lv et al., 2017; Huang and Lee, 2022).

The identification of tumor-specific variants would be extremely beneficial for diagnosing and treating patients with tumors. Somatic variant analysis can help identify tumors, determine their origin, assess their level of malignancy, and provide insights into hereditary predisposition to tumors and treatment effectiveness (Lv et al., 2017). Detecting somatic variants through DNA sequencing techniques can provide valuable information for diagnosing and identifying genetic alterations underlying tumor progression. Moreover, identifying specific mutations can assist in selecting appropriate treatment options and monitoring therapy efficacy over time (Yu et al., 2015).

In our study, we specifically attempted to analyze and sequence circulating tumor DNA using NGS. Analytical methods for detecting somatic variants in liquid biopsy must be highly sensitive and specific to ensure accurate and reliable results (Lv et al., 2017).

2. Materials and methods

2.1. Sample collection and laboratory procedure

The study was approved by the Ethics Committee of Azienda Sanitaria dell'Alto Adige, Approval No. 132-2020. Ethics committee approval No. 132-2020 covers the general use of anonymized data for research purposes to search for genetic variants once the diagnostic process is completed. Research was conducted according to the ethical guidelines of the Declaration of Helsinki. A written informed consensus was obtained from all patients or their guardians, and a unique alphanumeric code was assigned to each of them to protect their anonymity.

Blood samples were collected in 10 ml Cell-Free DNA BCT CE tubes (Streck – 218996). These tubes guarantee stability of the cfDNA/gDNA sample for up to 14 days if kept between 6 °C and 37 °C. However, we preferred to immediately proceed with plasma separation and subsequent extraction in order to optimize the process and obtain as much cfDNA as possible. This procedure was optimized taking into account the short half-life of cfDNA and considering that some studies have shown that after three hours from collection, leukocyte lysis may occur, resulting in the release of germinal DNA and a decrease in tumor DNA (Recommendations for Performing of Molecular Testing on Liquid Biopsy in Oncology, 2018).

For the separation of plasma from cellular components after collection, a first centrifugation was performed at 1900×*g* for 10 min. At the end of the centrifugation, if done correctly, perfect separation between the two phases can be observed. The obtained plasma was then aliquoted into 2 ml Eppendorf tubes in order to proceed with a second centrifugation at 16000×*g* for 10 min, which allows for the deposition of any remaining debris. The resulting plasma (approximately 5 ml) was then aliquoted into new Eppendorf tubes. For cfDNA extraction, we used the MagPurix CFC DNA Extraction Kit (ZINEXTS – ZP02017). The ZINEXTS MagPurix 12 System allows for the automated extraction of nucleic acids from different types of biological samples.

The obtained plasma was aliquoted into 7 ml Sample Tubes provided in the kit. In order to optimize the technique, multiple extractions were performed starting from the same plasma tube. 75–150 µl of Proteinase K and 0.8–1.6 ml of RLA Buffer are added to the plasma. In our case, 2000 µl plasma + 150 µl Proteinase K+1600 µl RLA buffer were used. The samples were then vortexed for 30 s and incubated at 56 °C for 30 min in a heat block. After the incubation, the samples were loaded onto the ZINEXTS MagPurix 12 System for extraction of cfDNA. The obtained cfDNA was quantified using Qubit HS (Thermo Fisher Scientific – Q32851) and the size distribution was evaluated using 4200 TapeStation System (Agilent – G2991BA) or Bioanalyzer (Agilent – G2939BA). A higher peak is expected between 120 and 220 bp and a lower peak around 300 bp. For library preparation and subsequent sequencing, the internal Twist library preparation with UMIs (cfDNA) and hybridization target enrichment procedure was used. The optimal input quantity of cfDNA is 30 ng total, however, if lower quantities of cfDNA are available (minimum of 10 ng total), the remaining volume (about 45 µl post Qubit and TapeStation) of each cfDNA sample was transferred to 0.2 ml tubes. These were then placed in a DNA130 SpeedVac (Thermo Scientific), with no heat (Temperature: "no" or minimum heat of 35 °C). After the drying program was completed, each sample was resuspended in 35 µL of water.

For a total amount of cfDNA>30 ng, the cfDNA was diluted to 0.86 ng/µl in a total volume of 35 µl (30 ng total). The cfDNA is present with sizes ranging from 120-220 bp and a peak around 300 bp, so there is no need to perform the fragmentation step typically required when sequencing genomic DNA. The protocol (Library Preparation with the Twist UMI Adapter System DOC-001282 REV 3.0) can proceed directly with the end-repair and A-tailing steps, followed by ligation of the UMI adapters.

Subsequently, UDI primers were added to each sample, followed by PCR amplification (12 cycles). The resulting samples were then purified and quantified; 187.5 ng of each sample were taken to create the final pool. This pool was then placed on a DNA130 SpeedVac, and after drying, the hybridization reaction was set up overnight for 16 h. After the reaction, target capture of the hybrids was performed, followed by post-capture PCR amplification. The resulting library was then verified using Bioanalyzer and Qubit, before proceeding with denaturation and loading onto the sequencer. For the optimization of the NGS protocol, a 5 % standard and a 0 % (wild type) standard were used. Turnaround time of wet laboratory procedures is of nearly five days.

The 0 % standard is a synthetic DNA, defined as wild type, because it does not show any variants with imbalances greater than 5 %, while the 5 % standard is a sample that contains a known and constant amount of low percentage mutated tumor DNA. This control DNA was chosen at 5 %, as this percentage refers to the sensitivity level required to detect low-frequency tumor mutations. The 5 % standard is used as a reference to ensure the reliability of liquid biopsy tests, as it allows to verify that

the technique is able to detect the presence of tumor mutations even at low frequencies. For example, if a tumor mutation is detected in a patient sample at a percentage of 5 %, it can be compared to the same mutation in the artificial plasma sample (5 % standard) to confirm the validity of the results obtained. The advantage of using an internal 5 % standard is to allow an accurate quantification of the actual mutation signal in patient samples.

To summarize, the 5 % artificial plasma sample allows to verify that the liquid biopsy test can accurately detect any presence of tumor mutations, thus ensuring greater reliability and accuracy of the analysis. With the use of advanced instruments that allow to obtain extremely high coverage, it is possible to identify VAF even at 0.01 %. In our specific case, we aimed to obtain 300× coverage, which would allow the identification of VAF at 5 %. Lowering the frequency below 5 % with 300× coverage in our case is not possible, as it would result in too many false positives (Lone et al., 2022).

2.2. Bioinformatical procedure

2.2.1. The pipeline

In the Illumina sequencing process, two Fastq files (forward-reverse) are obtained and filtered based on the evaluation of the quality of each read using the Phred-Score. Subsequently, trimming of ±5 bp is performed on the reads. High-quality reads are then aligned to the human reference sequence GRCh38 using the BWA software (Li and Durbin, 2010). The aligned and sorted reads with the SAMTOOLS software (Li et al., 2009) undergo duplicate removal and are later realigned and locally recalibrated around indels (insertions/deletions) and known point mutations using the GATK software (McKenna et al., 2010).

The Variant Call Format (VCF) file named “Unbalance” (Supplementary File 1) containing imbalances and coverage of variants is generated by two different software tools: SAMTOOLS (Bcftools) and GATK (HaplotypeCaller). Variant annotation is performed using the dbNSFP v4.1 database (Liu et al., 2020), while the APPRIS database (Rodriguez et al., 2013) is used to determine the main transcript of the involved gene. Variant selection and interpretation are based on the automated interpretation of VarSome Stable-API (Cristofoli et al., 2021; Cristofoli et al., 2023).

For each sample, data obtained from germline sequencing is compared with somatic sequencing data. A detailed analysis of base-to-base coverage is performed in coding exons and intron/exon junctions, and imbalances are compared based on the observed allelic frequency. Only alterations showing an allelic imbalance of 2 % or greater are selected (e.g., germline 100 % A – somatic 98 % A – 2 % T) and confirmed using the LoFreq tool (Wilm et al., 2012; Marceddu et al., 2019). The workflow of the process is depicted in Fig. 1. Turnaround time of dry laboratory procedures is about one hour for each sample. The analysis can be parallelized to optimize turnaround time for analysis of multiple samples.

2.2.2. Description and process of the research method

The somatic variant research method is based on the analysis of the VCF file named “Unbalance,” which is generated through the pipeline described before. This method aims to identify clinically relevant variants within the VCF file.

Initially, a comparison is made between the germline and somatic line to detect 2 % imbalances. Subsequently, somatic variants are preselected based on specific filtering parameters, including coverage analysis, imbalance percentages, and the number of variant imbalances.

After pre-selection, variants are annotated, followed by the confirmation of called somatic variants using Lofreq. Next, the variants are interpreted using the stable API of Varsome following the AMP verdict.

In the next phase, filters are applied to interpret the AMP verdict of Tier I and Tier II, aiming to select pathogenic variants. Finally, a filter is applied for Tier I drug and treatment interpretations. The overall process of the detection method is briefly summarized in Fig. 2. Once these stages are completed, several files are saved, containing all variants that meet the filtering criteria, including clinically relevant ones.

2.2.3. Filtering parameters

The method excludes all variants within the following parameters:

- Variants with germline coverage less than 20×.
- Variants with somatic coverage less than 50×.
- •Variants with somatic imbalance equal to or less than 2 %.
- Variants with 4 or more somatic imbalances.
- Variants with 3 or more germline imbalances.
- Variants with an AMP verdict different from “Tier I” and “Tier II”.

The filtering parameters have been configured based on specific observations. Regarding the coverage, an adequate depth and quality

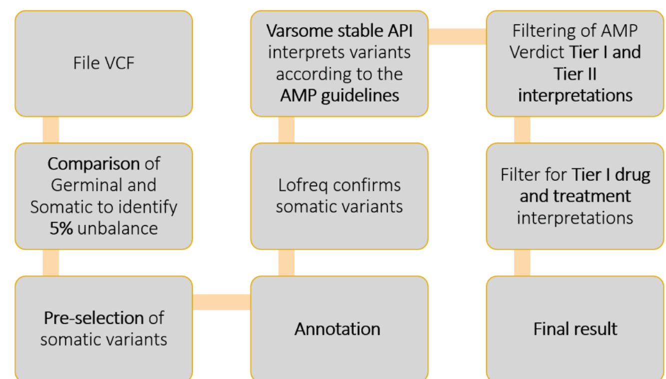


Fig. 2. Detection method workflow starting from the annotated VCF file to the final result.

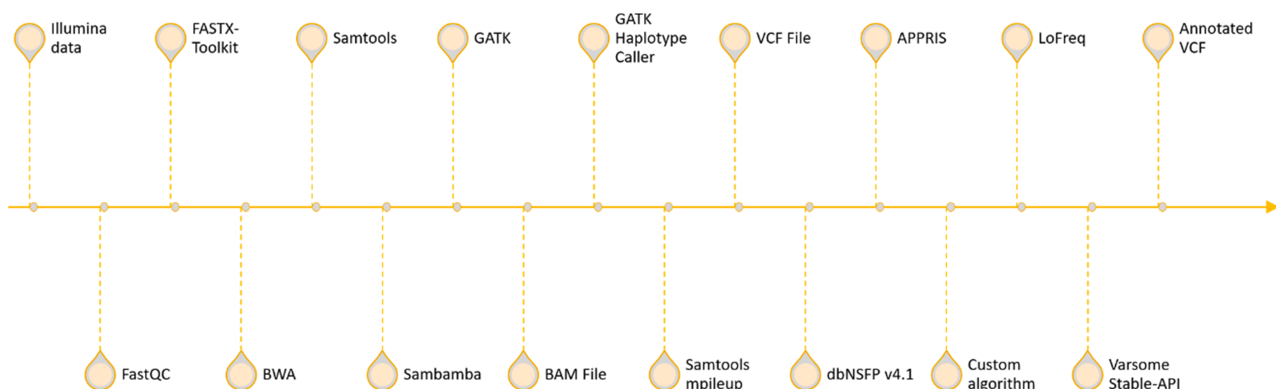


Fig. 1. Pipeline workflow: starting from the Illumina data to the annotated VCF file.

have been selected to allow accurate analysis. Additionally, somatic variants declared by Twist at 5 % are detected with a somatic imbalance ≥2 %, as noted in the Supplementary File 1. Consequently, a filter was set to exclude only variants with somatic imbalances below 2 %. This choice is motivated by the fact that if the system had been set at 5 %, certain clinically relevant variants would not have been identified. Moreover, it is important to note that there are variants with unexpected imbalances both in the somatic and germline line, due to the intrinsic variability of NGS or DNA structure. For example, genomic regions rich in GC can generate low-quality reads and, therefore, not always accurate calls. Some of these variants with unexpected imbalances are clinically relevant and must be identified, for instance, the two variants observable in Table 1.

To strike a balance between excessive exclusion of somatic variants and an increase in false positives, it was decided to filter variants that present a maximum number of imbalances (2 for germline and 3 for somatic line), see Table 1. Any identified false negatives will be subsequently reduced through a combined analysis, as described later, while false positives will be reduced through laboratory analysis using the "SnaPshot" methodology (Di Cristofaro et al., 2010). Finally, a filter is applied based on AMP guidelines for the interpretation of somatic variants. A 4-tier classification system is used, based on evidence, depending on the clinical significance of the variants, ranging from "Pathogenic" to "Benign." The interpretation of variants focuses on clinical-therapeutic impact, such as predicting sensitivity, resistance, or toxicity to approved or investigational chemotherapeutic agents.

2.3. Validation

The validation of the somatic variant method in liquid biopsy is a crucial process to ensure the reliability, robustness, and accuracy of the method. The goal is to confirm that the detected somatic variants are correct and that the used method can provide reproducible and reliable results.

For algorithm validation, known reference somatic variants provided by Twist with 5 % imbalances were used (Twist cfDNA Pan-cancer Reference Standard, 2022), as explained earlier, due to coverage limitations resulting in imbalances below 5 %. The system was set up accordingly.

Two runs were performed on the same reference 0 % and 5 % standard by the same operator, independently, named RX89.2023 and RX90.2023, and the detection method was applied. Subsequently, a comparison was made with the reference data, and a quality control analysis was conducted to identify false positives and false negatives and calculate the following parameters:

Table 1
Clinically relevant variants with two unexpected imbalances in the somatic line. This table is an example of two variants with unexpected imbalances that are clinically relevant and are identified.

#CHROM	chr18	chr10
POS	51047193	87958012
C%_germinal	0,00	0,00
G%_germinal	1,00	0,00
T%_germinal	0,00	1,00
A%_germinal	0,00	0,00
ins%_germinal	0,00	0,00
del%_germinal	0,00	0,00
DEPTH_germinal	71	29
GENE	SMAD4	PTEN
Exone	2	7
C%_somatic	0,00	0,00
G%_somatic	0,97	0,00
T%_somatic	0,00	0,95
A%_somatic	0,02	0,02
ins%_somatic	0,02	0,02
del%_somatic	0,00	0,00
DEPTH_somatic	239	78

- Sensitivity: This measures the ability to correctly identify variants that exist in the analyzed sample. Sensitivity is calculated as the ratio of true positives (TP) to the sum of true positives and false negatives (FN): Sensitivity = TP/(TP+FN)
- Specificity: This quantifies the ability not to call variants that are not truly present in the analyzed sample. Specificity is calculated as the ratio of true negatives (TN) to the sum of true negatives and false positives (FP): Specificity = TN/(TN+FP)

3. Results

3.1. Laboratory results

For the experiment two samples provided by an Italian hospital were used. After plasma separation and cfDNA extraction using the MagPurix CFC DNA Extraction Kit, the samples were quantified using Qubit HS, the results are shown in Table 2.

Each sample was then loaded on the 4200 TapeStation System (G2991BA) to evaluate the sizes distribution. The results are depicted in Fig. 3, while the theoretical size distribution of cfDNA is illustrated in Fig. 4.

3.2. Bioinformatical results

For the NGS procedure, the protocol used has been internally optimized, the protocol name is Twist LibraryPrep withUMI (cfDNA) e StandardHybTE (Twist Target Enrichment Standard Hybridization v2 Protocol, 2022). After the sequencing of the library on the MiSeq System using the MiSeq Reagent Kit v3, the results obtained were analysed using both the internal MAGI and LoFreq algorithms. The data reported in Table 3 regarding coverage and analysis refer to the RX89.2023 and RX90.2023 samples, both in germinal (0 % reference standard) and somatic (5 % reference standard) line.

In the RX89.2023 run (Table 4), a total of 7 false negatives were detected out of 51 variants (sensitivity of 86.27 %). In the RX90.2023 run (Table 4), a total of 8 false negatives were detected out of 51 variants (sensitivity of 84.31 %). The majority of false negatives were detected due to the number of imbalances caused by instrumental errors, which can be reduced through an increase in coverage..

To improve the sensitivity and performance of the detection method, a combined approach using data from both the RX89.2023 and RX90.2023 run was adopted, and the detection method was applied to the combined dataset. As shown in Tables 4 and 5, false negatives are reduced to only 3 out of 51 variants (sensitivity of 94.12 %).

The detected false negatives are exclusively present in specific genomic regions where NGS exhibits lower performance. Combining the runs reduces the inherent variability of NGS and increases the sensitivity levels of the method, resulting in a significant improvement in performance compared to using data from individual runs separately (see Supplementary File 1.

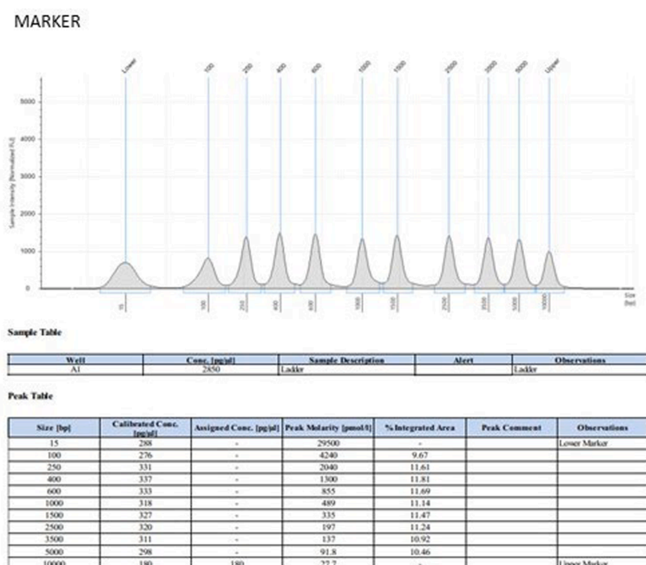
For these reasons, standardizing the method requires conducting the analysis on one germline line and two somatic lines with independent runs. This approach ensures higher accuracy and reliability in detecting somatic variants with improved sensitivity.

In the RX89.2023 run, a total of 10 false positives were detected out

Table 2
The results of the quantification of the samples RX84.2023 and RX87.2023 are reported in the table. The extraction and quantification of each sample were carried out in duplicate (A and B). The total nanograms present in 45 µl of each replica of cfDNA are also reported.

SAMPLE	A	Total* ng A	B	Total* ng B
RX84.2023	1,03 ng/µl	46,35 ng	1,25 ng/µl	56,25 ng
RX87.2023	0,228 ng/µl	10,26 ng	0,263 ng/µl	11,83 ng

* Total ng in 45 µl.



Sample Table

Well	Conc. [pg/µl]	Sample Description	Alert	Observations
A1	2850	Ladder		Ladder

Peak Table

Size [bp]	Calibrated Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	% Integrated Area	Peak Comment	Observations
15	288	-	29500	-		Lower Marker
100	276	-	4240	9.67		
250	331	-	2040	11.61		
400	337	-	1300	11.81		
600	333	-	855	11.69		
1000	318	-	489	11.14		
1500	327	-	335	11.47		
2500	320	-	197	11.24		
3500	311	-	137	10.92		
5000	298	-	91.8	10.46		
10000	180	180	27.7	-		Upper Marker

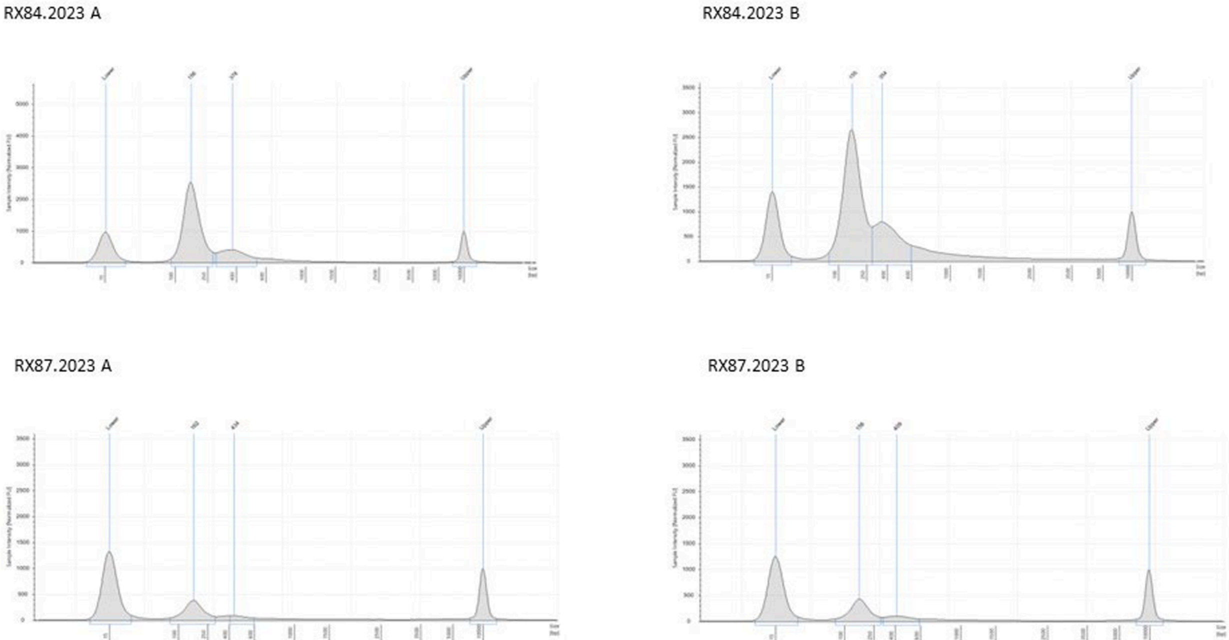


Fig. 3. Samples run on TapeStation is represented in the figure. The marker, reported in the upper part of the figure, with their respective peaks is reported. Of each sample both the independent extractions are shown (A and B).

of 90 variants (specificity of 88.89 %). In the RX90.2023 run, a total of 11 false positives were detected out of 87 variants (specificity of 87.36 %). When using the combined run method, there were 19 false positives out of 104 variants (specificity of 81.73 %). The false positives detected can be observed in the Supplementary File 1. The results are also reported in Table 6.

To increase the specificity of our analyses, we have implemented a second laboratory method, specifically SNaPshot. This approach allows analyzing positive variants with known and non-conforming patterns, further refining the selection of results. SNaPshot is an advanced technique that utilizes knowledge of specific patterns to identify relevant genetic variants and further discriminate between false positives and true positives. Clinically relevant variants selected in the diagnostic system will be confirmed through this method. Following the implementation of SNaPshot, the specificity is confirmed to be 99 % (Di Cristofaro et al., 2010).

4. Discussion

Liquid biopsy is a reliable and minimally invasive method that permit to extract ctDNA, which is then sequenced to identify somatic mutations of a tumor. It has many advantages, such as detecting material shed from multiple tumor sites and being easily implemented for serial screening, but also presents some disadvantages, mainly related to the challenging and difficultly reproducible procedure (Nikanjam et al., 2022b). In this work, we present a method for identifying 5 % variants in somatic ctDNA through an innovative protocol, and we evaluate its efficacy in terms of sensitivity and specificity. The adopted method follows a standardized process that involves several phases (Fig. 5).

Initially, a run is performed on the germline, followed by two independent runs on the somatic line. Subsequently, data obtained from Illumina sequencing are processed through a dedicated pipeline, generating the VCF file. This file is then subjected to analysis using a detection method that excludes somatic variants that do not meet pre-

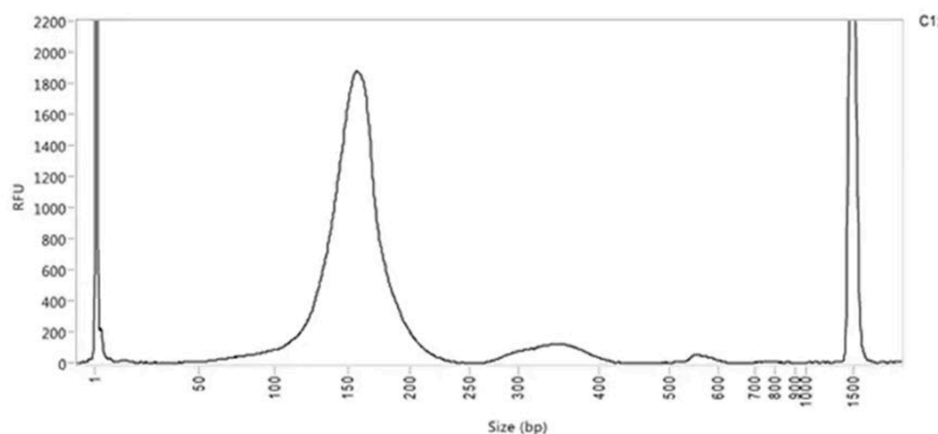


Fig. 4. Theoretical size distribution of cfDNA (Agilent – Theoretical size distribution of cfDNA. https://www.agilent.com/cs/pubimages/images/cfdna-fragment-analyzer-hs-small-fragment-zoom.png. Accessed 12 Oct, 2023).

Table 3
The results of the coverage of the samples RX89.2023 and RX90.2023 both in the germinal (0 % reference standard) and in the somatic (5 % reference standard) line are reported in the table. Furthermore, the coverage depth at different levels is also reported. The depth = 0 highlights possible uncovered regions.

SAMPLE	Cutoff	Count	% Coverage	Medium Coverage
RX89.2023 GERMINAL	depth>=10	342,513	99.7	76.7
	depth>=100	69,539	20.2	76.7
	depth < 10	906	0.3	76.7
	depth = 0	0	0.0	76.7
	depth_macro>=25	334,844	97.4	76.7
RX89.2023 SOMATIC	depth>=10	103,092	99.8	236.3
	depth>=100	99,304	96.1	236.3
	depth < 10	249	0.2	236.3
	depth = 0	0	0.0	236.3
RX90.2023 GERMINAL	depth>=10	337,228	98.4	48.9
	depth>=100	2841	0.8	48.9
	depth < 10	5362	1.6	48.9
	depth = 0	0	0.0	48.9
	depth_macro>=25	317,058	92.5	48.9
RX90.2023 SOMATIC	depth>=10	103,091	99.9	244.9
	depth>=100	99,788	96.7	244.9
	depth < 10	134	0.1	244.9
	depth = 0	1	0.0	244.9

Table 4
False negative RX89.2023 and RX90.2023. This table describes the number of false negatives, the false negative rate and the sensitivity of the method.

RUN	RELEVANT VARIANTS	FALSE NEGATIVES	FALSE NEGATIVE RATE	SENSITIVITY
RX89.2023	51	7	13,73 %	86,27 %
RX90.2023	51	8	15,69 %	84,31 %
COMBINED	51	3	5,88 %	94,12 %

established filtering criteria. Once relevant somatic variants are obtained, a SnaPshot analysis is conducted in the laboratory on these variants to identify clinically relevant ones.

Following this process, the approach achieves a sensitivity of 94.12 % and a specificity of 99 %, indicating its great potential. Its high sensitivity allows accurate identification of somatic variants, ensuring correct diagnosis. Similarly, the high specificity makes the method reliable in detecting clinically relevant somatic variants. The obtained values of sensitivity and specificity are higher than the ones reported in literature for similar applications (Garcia et al., 2021; Wang et al., 2021), supporting the successful performance of this protocol.

This study focused on method validation, therefore the study evaluated samples from two patients analyzed in duplicate for the validation of the DNA extraction protocol, and 0 % and 5 % standards for the validation of the NGS protocol. This technique can be applied to a high number of samples. Moreover, despite considerable progress made, being an evolving diagnostic method, there are still areas for improvement for liquid biopsy. An essential area for further development is increasing the coverage to obtain even more precise and comprehensive analysis of relevant genetic variants. This requires continuous improvement of the tools used to increase the depth and accuracy of the method. Moreover, also considering sample preparation and laboratory procedures, the future adoption of this method in diagnostic setting should take into account times and costs per sample.

5. Conclusions

The proposed research method appears as a promising approach in the field of tumor diagnostics, thanks to its high diagnostic sensitivity and specificity through the analysis of biological fluid samples. However, it continues to be under observation, with particular attention to the need for improving the coverage of genetic variants and optimizing the tools used to ensure an even more accurate and reliable analysis.

Declarations

Ethics approval and consent to participate. The study was approved by the Ethics Committee of Azienda Sanitaria dell'Alto Adige, Approval No. 132–2020. Our research was conducted according to the ethical guidelines of the Declaration of Helsinki. A written informed consensus was obtained from all patients or their guardians, and a unique alphanumeric code was assigned to each of them to protect their anonymity.

Consent for publication. Not applicable.

Availability of data and materials. All the data are contained in the main text and in the supplementary material.

Funding. This research was funded by Provincia Autonoma di Bolzano in the scope of LP 14/2006. Open Access provided by Università degli studi di Perugia within the CRUI-CARE Agreement.

CRediT authorship contribution statement

Chiara Mareso: Writing – original draft, Visualization, Validation,

Table 5a
False negatives detected in the RX89.2023 run.

					GERMINAL							SOMATIC						
Gene	Pos	#CHROM	exone	DIFF%	C%	G%	T%	A%	ins%	del%	DEPTH	C%	G%	T%	A%	ins%	del%	DEPTH
BRCA2	32339699	chr13	11	0.01	0.96	0.00	0.02	0.00	0.00	0.02	51	0.95	0.00	0.00	0.00	0.00	0.05	144
BRCA2	32379885	chr13	23	0.07	1.00	0.00	0.00	0.00	0.00	0.00	55	0.93	0.00	0.00	0.01	0.01	0.04	197
EGFR	55191822	chr7	21	0.01	0.01	0.00	0.99	0.01	0.00	0.00	139	0.00	0.02	0.98	0.00	0.00	0.00	419
HNF1A	120994313	chr12	4	0.05	0.03	0.94	0.00	0.00	0.03	0.00	86	0.05	0.89	0.00	0.00	0.05	0.00	367
NF1	31226459	chr17	18	0.05	0.00	0.00	0.00	1.00	0.00	0.00	49	0.02	0.00	0.00	0.95	0.02	0.01	195
NF2	29673477	chr22	12	0.01	1.00	0.00	0.00	0.00	0.00	0.00	71	0.99	0.00	0.00	0.00	0.00	0.01	214
TP53	7675088	chr17	5	0.02	1.00	0.00	0.00	0.00	0.00	0.00	85	0.98	0.01	0.01	0.00	0.00	0.00	221

Table 5b
False negatives detected in the RX90.2023 run. The false negatives present in both RX89.2023 and RX90.2023 are highlighted in red; general information about the variants is highlighted in green; germline imbalances and coverage are shown in blue; somatic imbalances and coverage are shown in orange.

					GERMINAL							SOMATIC						
Gene	Pos	#CHROM	exone	DIFF%	C%	G%	T%	A%	ins%	del%	DEPTH	C%	G%	T%	A%	ins%	del%	DEPTH
BRCA2	32339699	chr13	11	0.01	0.96	0.00	0.02	0.00	0.00	0.02	51	0.95	0.00	0.00	0.00	0.00	0.05	144
BRCA2	32379885	chr13	23	0.07	1.00	0.00	0.00	0.00	0.00	0.00	55	0.93	0.00	0.00	0.01	0.01	0.04	197
EGFR	55191822	chr7	21	0.01	0.01	0.00	0.99	0.01	0.00	0.00	139	0.00	0.02	0.98	0.00	0.00	0.00	419
HNF1A	120994313	chr12	4	0.05	0.03	0.94	0.00	0.00	0.03	0.00	86	0.05	0.89	0.00	0.00	0.05	0.00	367
NF1	31226459	chr17	18	0.05	0.00	0.00	0.00	1.00	0.00	0.00	49	0.02	0.00	0.00	0.95	0.02	0.01	195
NF2	29673477	chr22	12	0.01	1.00	0.00	0.00	0.00	0.00	0.00	71	0.99	0.00	0.00	0.00	0.00	0.01	214
TP53	7675088	chr17	5	0.02	1.00	0.00	0.00	0.00	0.00	0.00	85	0.98	0.01	0.01	0.00	0.00	0.00	221

Table 5c
Combined False Negatives, false negatives detected in RX89.2023 are shown in blue; false negatives in RX90.2023 are highlighted in yellow.

					GERMINAL							SOMATIC						
Gene	Pos	#CHROM	exone	DIFF%	C%	G%	T%	A%	ins%	del%	DEPTH	C%	G%	T%	A%	ins%	del%	DEPTH
BRCA2	32379885	chr13	23	0.07	1.00	0.00	0.00	0.00	0.00	0.00	55	0.93	0.00	0.00	0.01	0.01	0.04	197
HNF1A	120994313	chr12	4	0.05	0.03	0.94	0.00	0.00	0.03	0.00	86	0.05	0.89	0.00	0.00	0.05	0.00	367
NF1	31226459	chr17	18	0.05	0.00	0.00	0.00	1.00	0.00	0.00	49	0.02	0.00	0.00	0.95	0.02	0.01	195
BRCA2	32379885	chr13	23	0.04	0.93	0.00	0.00	0.03	0.03	0.01	64	0.89	0.00	0.00	0.05	0.04	0.03	201
HNF1A	120994313	chr12	4	0.04	0.02	0.96	0.00	0.00	0.02	0.00	43	0.04	0.92	0.00	0.00	0.04	0.00	319
NF1	31226459	chr17	18	0.04	0.00	0.00	0.00	0.98	0.00	0.02	45	0.02	0.00	0.00	0.94	0.02	0.02	204

Table 6
This table describes the number of false positives, the false positive rate and the specificity of the method.

Run	Detected variants	False positives	False positive rate	Specificity
RX89.2023	90	10	11,11 %	88,89 %
RX90.2023	87	11	12,64 %	87,36 %
Combined	104	19	18,27 %	81,73 %

Methodology, Data curation. **Luca Crosta:** Writing – review & editing, Validation, Data curation. **Mattia Giovanna De Vita:** Writing – review & editing, Validation, Data curation. **Francesca Cristofoli:** Writing – review & editing, Validation, Data curation. **Benedetta Tanzi:** Writing – review & editing, Validation. **Sabrina Benedetti:** Writing – review & editing, Validation. **Gabriele Bonetti:** Writing – original draft, Investigation. **Carmine Antonio Donofrio:** Writing – review & editing, Resources, Investigation. **Marika Cominetti:** Writing – review & editing, Resources, Investigation. **Lucia Riccio:** Writing – review & editing, Resources, Investigation. **Antonio Fioravanti:** Writing – review &

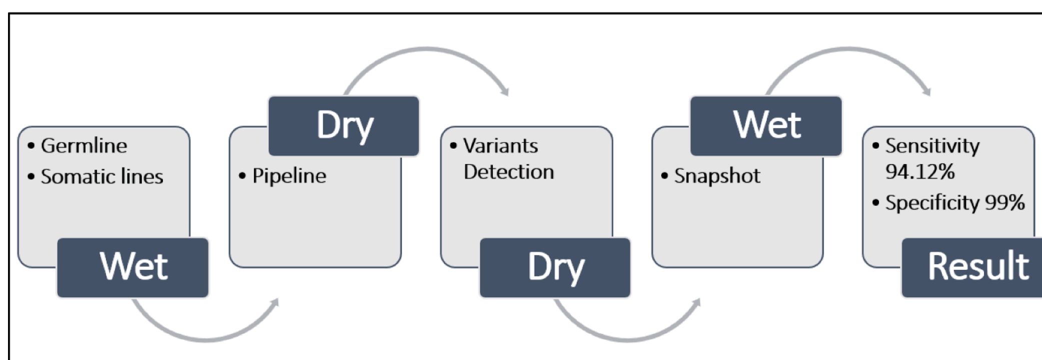


Fig. 5. Standardized process. This figure describes briefly the whole process, starting from the extraction to the final result.

editing, Resources, Investigation. **Daniele Generali:** Writing – review & editing, Resources, Investigation. **Emanuela Lucci Cordisco:** Writing – review & editing, Resources, Investigation. **Pietro Chiurazzi:** Writing – review & editing, Resources, Investigation. **Valentina Gatta:** Writing – review & editing, Resources, Investigation. **Liborio Stuppia:** Writing – review & editing, Resources, Investigation. **Stefano Cecchin:** Writing – review & editing, Validation, Project administration. **Matteo Bertelli:** Supervision, Project administration, Conceptualization. **Giuseppe Marceddu:** Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors of this scientific article often publish together, being part of the same group and/or scientific society. The scientific journal designated to publish this article has a specific policy to ensure the optimal handling of these situations. The journal's policy ensures that the peer review process is kept independent by selecting reviewers independent of the authors of the article.

All affiliations of the authors with private companies have been declared to make clear the position regarding the interests of these companies. The authors are affiliated with private companies for which there could be a possible conflict of interest.

The authors state that the research was conducted in the absence of any commercial relationship that could be construed as a possible direct conflict of interest avoiding mention of any products or services sold by private companies related to the authors.

The authors of this article are reported to be patents inventors.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank all the clinicians for their kind collaboration. Open Access provided by Università degli studi di Perugia within the CRUI-CARE Agreement.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2024.148771.

References

Agilent – Theoretical size distribution of cfDNA. https://www.agilent.com/cs/pubimages/images/cfdna-fragment-analyzer-hs-small-fragment-zoom.png. Accessed 12 Oct 2023.

Chen, M., Zhao, H., 2019. Next-generation sequencing in liquid biopsy: cancer screening and early detection. Hum Genomics. 13, 34.

Cristofoli, F., Sorrentino, E., Guerri, G., Miotto, R., Romanelli, R., Zulian, A., et al., 2021. Variant selection and interpretation: an example of Modified VarSome Classifier of ACMG guidelines in the diagnostic setting. Genes 12, 1885.

Cristofoli, F., Daja, M., Maltese, P.E., Guerri, G., Tanzi, B., Miotto, R., et al., 2023. MAGI-ACMG: algorithm for the classification of variants according to ACMG and ACGS recommendations. Genes 14, 1600.

Di Cristofaro, J., Silvy, M., Chiaroni, J., Bailly, P., 2010. Single PCR multiplex SNaPshot reaction for detection of eleven blood group nucleotide polymorphisms: optimization, validation, and one year of routine clinical use. J. Mol. Diagn. 12, 453–460.

Garcia, J., Kamps-Hughes, N., Geiguer, F., Couraud, S., Sarver, B., Payen, L., et al., 2021. Sensitivity, specificity, and accuracy of a liquid biopsy approach utilizing molecular amplification pools. Sci. Rep. 24, 10761.

Huang, A.Y., Lee, E.A., 2022. Identification of somatic mutations from bulk and single-cell sequencing data. Front. Aging 2, 800380.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, X., Li, C., Mou, C., Dong, Y., Tu, Y., 2020. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med 12, 103.

Lone, S.N., Nisar, S., Masoodi, T., Singh, M., Rizwan, A., Hashem, S., et al., 2022. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. Mol. Cancer. 21, 79.

Lv, X., Zhao, M., Yi, Y., Zhang, L., Guan, Y., Liu, T., et al., 2017. Detection of rare mutations in CtDNA using next generation sequencing. J. Vis. Exp., 56342

Marceddu, G., Dallavilla, T., Guerri, G., Manara, E., Chiurazzi, P., Bertelli, M., 2019. PipeMAGI: an integrated and validated workflow for analysis of NGS data for clinical diagnostics. Eur. Rev. Med. Pharmacol. Sci. 23, 6753–6755.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297.

Nikanjam, M., Kato, S., Kurzrock, R., 2022a. Liquid biopsy: current technology and clinical applications. J. Hematol. Oncol. 15, 131.

Nikanjam, M., Kato, S., Kurzrock, R., 2022b. Liquid biopsy: current technology and clinical applications. J. Hematol. Oncol. 12, 131.

Poulet, G., Massias, J., Taly, V., 2019. Liquid biopsy: general concepts. Acta Cytol. 63, 449.

Recommendations for performing of molecular testing on liquid biopsy in oncology. 2018. Available from: https://www.aiom.it/wp-content/uploads/2018/09/2018_biopsialiquida.pdf. Accessed 12 Oct 2023.

Rodriguez, J.M., Maietta, P., Ezkurdia, I., Pietrelli, A., Wesselink, J.J., Lopez, G., et al., 2013. APPRIS: annotation of principal and alternative splice isoforms. Nucleic Acids Res. 41, D110–D117.

Twist cfDNA Pan-cancer Reference Standard, 2022. https://www.twistbioscience.com/sites/default/files/resources/2022-10/DOC-001287_ProductSheet_cfDNA-PanCancer-REV3-singles-bb.pdf. Accessed 12 Oct 2023.

Twist Target Enrichment Standard Hybridization v2 Protocol, 2022. https://www.twistbioscience.com/sites/default/files/resources/2022-04/Protocol_NGS_TargetEnrichmentStandardHyb_V2Protocol_12APR22_Rev3.0.pdf. Accessed 12 Oct 2023.

Wang, N., Zhang, X., Wang, F., Zhang, M., Sun, B., Yin, W., et al., 2021. The Diagnostic Accuracy of Liquid Biopsy in EGFR-Mutated NSCLC: a systematic review and meta-analysis of 40 studies. SLAS Technol. 26, 42–54.

Wilm, A., Aw, P.P., Bertrand, D., Yeo, G.H., Ong, S.H., Wong, C.H., et al., 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40, 11189.

Yu, B., O'Toole, S.A., Trent, R.J., 2015. Somatic DNA mutation analysis in targeted therapy of solid tumours. Transl. Pediatr. 4, 125–128.