Identification of Somatic Mutations in Thirty-year-old Serum Cell-free DNA From Patients With Breast Cancer: A Feasibility Study

Mathilde Ritter,[1,2] Viola Paradiso,[3] Patrik Widmer,[3] Andrea Garofoli,[3,4] Luca Quagliata,[3] Serenella Eppenberger-Castori,[3] Savas D. Soysal,[4] Simone Muenst,[3] Charlotte K.Y. Ng,[5] Salvatore Piscuoglio,[3,4] Walter Weber,[6] Walter P. Weber[1,2]

Abstract

The aim of this study was to assess the feasibility of cell-free DNA extraction and circulating tumor DNA sequencing in 30-year-old serum samples of patients with breast cancer. Cell-free DNA extraction was successful in 52 of 52 patients, and 24 cancer-specific mutations were found in 22 of 25 samples undergoing sequencing. This study shows that next-generation sequencing technology is sufficiently robust and specific to analyze 30-year-old serum.

Introduction: The aim of this study was to assess the feasibility of cell-free DNA (cfDNA) extraction and circulating tumor DNA sequencing in 30-year-old serum samples. **Materials and Methods:** We evaluated serum samples from 52 patients with breast cancer, which were collected between 1983 and 1991, with correlating clinicopathologic data. cfDNA was extracted by using the QIAamp Circulating Nucleic Acid Extraction Kit (Qiagen). Of these 52 cfDNA samples, 10 were randomly selected and sequenced with the Oncomine Breast cfDNA Assay (A31183). In a second step, high-depth targeted sequencing of 15 additional cfDNA samples was performed using a custom Ampliseq Ion Torrent panel targeting breast cancer-related genes. **Results:** cfDNA extraction was successful in 52 (100%) of 52 patients with a total concentration of 0.2 to 54 ng/uL. A total of 24 cancer-specific mutations were found in 22 (88%) of the 25 samples undergoing sequencing. Of the 52 patients, 32 (62%) had died from breast cancer after a median follow-up of 7.9 years (interquartile range, 3.7-15.5 years). **Conclusion:** The present study shows that current next generation sequencing technology is sufficiently robust and specific to analyze 30-year-old serum. Therefore, longitudinal studies can be designed with storage of serum samples over many years, thereby obviating the need for timely and continuous cfDNA extraction and sequencing. The samples can be pooled and processed at once with the most modern technology available at the end of the study, when accumulation of events allows correlation of clinical outcomes with adequate power.

Clinical Breast Cancer, Vol. 20, No. 5, 413-21

Keywords: Breast cancer, Circulating tumor DNA, Serum, Somatic mutations

M.R. and V.P. contributed equally to this work as first authors. W.P.W. and W.W. contributed equally to this work as last authors.

[1]Breast Center, University Hospital of Basel and University of Basel, Basel, Switzerland
[2]Department of Breast Surgery, University Hospital of Basel, Basel, Switzerland
[3]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland
[4]Visceral Surgery Research Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland
[5]Department for BioMedical Research, University of Bern, Bern, Switzerland
[6]Clinical Cancer Etiology Unit, Basel, Switzerland

Submitted: Jan 25, 2020; Revised: Apr 4, 2020; Accepted: Apr 7, 2020; Epub: May 6, 2020

Addresses for correspondence: Mathilde Ritter, MD, Division of Breast Surgery, University Hospital Basel, Spitalstrasse 21, CH-4031, Basel, Switzerland; or Walter P. Weber, MD, Chief, Division of Breast Surgery, Breast Surgeon SSO, University Hospital Basel, Spitalstrasse 21, CH-4031, Basel, Switzerland
E-mail contact: mathilde.ritter@usb.ch; walter.weber@usb.ch

http://dx.doi.org/10.1016/j.clbc.2020.04.005

Introduction

A breakthrough in next-generation sequencing (NGS) in the past decade provided an unprecedented opportunity to investigate genetic variations in humans and their roles in health and disease. In particular, large-scale efforts such as The Cancer Genome Atlas and the International Cancer Genome Consortium have provided a comprehensive molecular portrait of human cancers.[1,2] The discovery of the so-called 'driver genes' has provided the basis for the development of the concept of precision medicine, where the identification of targetable alterations guides the therapeutic approach in treating patients with cancer. Nowadays, the decreasing costs of massively parallel sequencing have resulted in increased adoption of genomic profiling as part of the standard diagnostic procedures in most tumor types.[3,4]

In patients with cancer, nucleic acids obtained from tumor biopsies and resections remain the main source for molecular profiling. However, these procedures are invasive, costly, time-consuming, and have only limited potential to be repeated in longitudinal studies.[5] Their relevance is further limited by the prevalence of intra-tumor genetic heterogeneity as shown in multiple sequencing studies over the past decade.[5,6] Thus, a single biopsy of the primary tumor is not likely to be genetically representative of the whole tumor. To overcome these challenges, circulating cell-free DNA (cfDNA) has been proposed as an alternative because it can be collected less invasively compared with conventional biopsies.[7,8] Circulating cfDNA is a type of cell-free nucleic acid that derives from apoptotic and necrotic cells or is released from living eukaryotic cells.[9] The detection of DNA in the blood originating from tumors in patients with cancer has been described decades ago.[10-12] The fraction of cfDNA derived from tumor is termed circulating tumor DNA (ctDNA).[7,8] ctDNA can be considered a new source for the detection and surveillance of major cancers because it is more likely to be present in patients with cancer.[7,8]

The potential of using cfDNA as an indicator of disease burden with prognostic implication and clinical applicability during follow-up and monitoring in both the curative and palliative setting has been investigated in numerous studies.[13,14] Cancer-specific mutations, copy number alterations, and genomic rearrangements assessed in ctDNA demonstrated potential prognostic and predictive significance.[15-20] To further evaluate prognostic and predictive biomarkers in cfDNA and assess its value as a disease monitoring tool, longitudinal studies with long follow-up are necessary. The utility of the technology depends on its capability to assess sequential samples that have been collected and stored over a long period of time. This study aims to assess the feasibility of cfDNA extraction and somatic mutation assessment in 30-year-old serum that has been collected from patients with breast cancer between 1983 and 1991.

Materials and Methods

Patients and Serum

For this study, we had access to serum samples from 753 patients with cancer, which were collected between 1983 and 1991 in an oncologic private practice in Basel, Switzerland. Of 753 patients, 152 were females with breast cancer. The patients were referred to the medical oncologist either after surgery of the primary tumor or after the diagnosis of local/regional recurrence and/or distant metastases. After obtaining informed consent, 10 mL of native venous blood were collected in a 10-mL BD Vacutainer blood collection tube and centrifuged in a Hettich centrifuge at 5000 rpm for 10 minutes. The serum samples were immediately frozen and stored at −70°C to −80°C in 3 Nunc Cryogenic tubes per patient (Gibko AG) at a private office during the first 9 years; thereafter, the samples have been transferred to the Institute of Immunobiology in Freiburg, Germany, by using transportable refrigerating boxes to avoid thawing. In 1999, the samples were relocated to the Laboratory for Medical Genetics of the University of Basel, Switzerland, and stored until processing and analysis. Clinicopathologic variables regarding patient demographics, primary tumor, treatment, recurrence, and survival were retrieved from clinical files. Approval for the use of these samples and correlating data has been granted by the responsible ethics committee (approval number: eknz-2018-00252).

cfDNA Extraction

Circulating DNA was extracted from 2 to 4 mL of isolated serum from 52 randomly selected patients with breast cancer with the QIAamp Circulating Nucleic Acid Kit (Qiagen) as previously described.[21] DNA was quantified using the Qubit Fluorometer (Invitrogen) and analyzed using the 2200 TapeStation system (Agilent Technologies) with the High Sensitivity DNA Analysis Kit.

Targeted Sequencing and Library Preparation

Sequencing was performed using 2 different amplicon-based targeted sequencing panels. The first 10 randomly selected samples were sequenced with the Oncomine Breast cfDNA Assay (A31183, Thermo Fisher Scientific). This panel covers 152 hotspot mutations in 10 genes (*AKT1*, *EGFR*, *ERBB2*, *ERBB3*, *ESR1*, *FBXW7*, *KRAS*, *PIK3CA*, *SF3B1*, and *TP53*) across 26 amplicons. This integrates the TagSeq technology (molecular barcode) and allows detection of rare variants present at 0.1% allelic frequency. Library preparation, molecular barcoding, and sequencing were performed according to the instructions and guidelines provided by Thermo Fisher, using 5 ng of DNA as input. Briefly, the library preparation protocol was based on a 2-step cycle multiplex touchdown polymerase chain reaction (PCR) with a temperature ranging from 64 °C to 58 °C, which allowed to amplify target regions and to introduce unique molecular identifiers. The obtained tagged amplicons of around 100 to 140 bp length were then cleaned up using Agencourt AMPure XP (Beckman Coulter), then eluted in 24 µl low TE buffer. A second round of PCR (18 cycles) was performed in a total volume of 50 µl to amplify the purified amplicons and to introduce Ion Torrent Tag-Sequencing adapters containing sample-specific barcodes. The resulting library of target DNA fragments was purified by performing a 2-step cleanup using Agencourt AMPure XP (Beckman Coulter). The purified libraries were then diluted 1:1000 and quantified by qPCR using the Ion Universal Quantitation Kit (Thermo Fisher Scientific). The quantified stock libraries were then diluted to 100 pM for downstream template preparation. Subsequentially, sequencing runs were planned on the Torrent Suite Software v5.2, and libraries were pooled and loaded on an Ion 540 chip using the Ion Chef Instrument (Thermo Fisher Scientific). The loaded chip was then sequenced using 500 flows on an S5 system (Thermo Fisher Scientific).

Table 1 Clinicopathologic Parameters of 52 Patients With Breast Cancer

Survival and Recurrences	n (%) or Median (IQR)
No. patients with clinical data and serum	52
Median age at first diagnosis, y	49.5 (45.5-60.5)
Median follow-up time from diagnosis to death or last follow-up, y	7.9 (3.7-15.5)
Median time from initial diagnosis to date of sample collection, y	1.6 (1.0-4.8)
No. breast cancer-specific deaths	32 (62)
No. deaths unrelated to breast cancer	4 (7)
Cause of death unknown	2 (3)
Median overall survival from diagnosis to death, y	6.8 (3.2-13.9)
Median disease-free survival diagnosis to local/regional or distant recurrence, second breast cancer or death, y	2.7 (1.5-6.4)
Treatment at first diagnosis	
Neoadjuvant treatment	2 (3)
Surgery	50 (96)
Adjuvant radiation	18 (35)
Adjuvant tamoxifen	18 (35)
Adjuvant CMF (±v ±p) or LMF (±vp) chemotherapies	15 (28)
Other adjuvant systemic therapies	3 (5)
Clinicopathological parameters at first diagnosis	
Female	52 (100)
Laterality	
Left	25 (48)
Right	24 (47)
Bilateral	3 (5)
Grade[a]	
1	1 (2)
2	14 (27)
3	37 (71)
Hormone receptor status[a]	
Positive	42 (82)
Negative	10 (18)
T stage	
1	14 (27)
2	28 (54)
3	5 (10)
4	4 (7)
X	1 (2)
N stage	
0	16 (31)
1	26 (50)
2	6 (11)
3	1 (2)
X	3 (6)
M stage	
0	50 (97)
1	2 (3)

Abbreviations: C = cyclophosphamide; F = fluorouracil; IQR = interquartile range; L = chlorambucil (leukeran); M = methotrexate; P = prednisone; V = vincristine.
[a]Had not been assessed routinely at the time.

In a second round, 15 randomly selected samples were sequenced with a custom targeted sequencing panel focusing on the most frequently altered genes in breast cancer previously described.[22] Library preparation for the breast panel was performed using the Ion AmpliSeq library kit 2.0 (Thermo Fisher Scientific) according to the manufacturer's guidelines. The panel consists of 2 pools of amplification primers. Ten ng of DNA per sample were used for library preparation for each pool. Amplification was performed according to the manufacturer's guidelines. The amplicons from the 2 pools were combined and treated to digest the primers and to phosphorylate the amplicons. The amplicons were then ligated to Ion Xpress Barcode Adapters (Thermo Fisher Scientific) using DNA ligase. Finally, cleaning and purification of the generated libraries were performed with Agencourt AMPure XP (Beckman Coulter) according to the manufacturer's guidelines. Quantification and quality control were performed with Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific). Samples were diluted to reach the concentration of 40 pmol and then were pooled for sequencing. Twenty-five µl of the pooled libraries were loaded on Ion 540 Chip (Thermo Fisher Scientific) and processed in Ion Chef Instrument (Thermo Fisher Scientific). Sequencing was performed on Ion S5XL system (Thermo Fisher Scientific).[23]

Somatic Variants Identification

Raw data were processed automatically on the Torrent Server and aligned to the reference hg19 genome. The analysis pipeline included signal processing, base calling, quality score assignment, adapter trimming, PCR duplicate removal, and control of mapping quality. All samples passed the quality check and met the requirements of a minimum molecular average depth. The first round of samples sequencing data (n = 10) was uploaded in BAM format to the Ion Reporter Analysis Server for variant calling and annotation. Variant calling was performed on Ion Reporter (IR) Analysis Software v5.2 using the Oncomine TagSeq Breast Liquid Biopsy w2.0 workflow. Coverage metrics for each amplicon were obtained by running the Coverage Analysis Plugin software v5.2.1 (Thermo Fisher Scientific). Identified variants were only considered if the variant had a molecular coverage of at least 3, indicating that the variant was detected in 3 independent template molecules. Finally, all candidate mutations were manually reviewed using the Integrative Genomics Viewer37.

In the second round of sequenced samples (n = 15), variant calling was performed with TVC version 5.0.3 (Torrent Variant Caller, Thermo Fisher Scientific) using low-stringency parameters previously described.[24,25] Briefly, mutations detected by TVC were subsequently filtered by the following steps. First, all the multiallelic variants have been split and left aligned. Moreover, the presence and the relative length of homopolymer sequences were annotated to take into account the presence of possible wrongly aligned sequencing reads and, therefore, false-positive variants. Second, because the 15 samples had no matched germline samples, all the variants have been annotated using 3 databases: the 1000 Genomes Project, the Exome Aggregation Consortium, and the NHLBI GO Exome Sequencing Project.[26,27] All the mutations identified by TVC that were also present within the databases in significative frequencies (> 5%) have been flagged as probable germline mutations. Furthermore, a pool of 16 germline samples collected from an independent cohort was used to provide an additional list of likely germline mutations that, together with the ones previously flagged, have been filtered out from the final output list. To avoid the removal of clinically relevant information, mutations found in known cancer driver hotspots have been whitelisted and kept even when they met the criteria for the aforementioned filtering steps.

Results

cfDNA Extraction From 30-year-old Serum of Sufficient Quality for Sequencing Analysis

We randomly selected 52 of the 152 patients with breast cancer (clinicopathologic characteristics of patients are shown in Table 1) to perform cfDNA extraction (Table 2). cfDNA extraction was successful in all patients, and cfDNA levels were determined for each sample with a fluorometric quantitation system. We obtained a range of concentrations from 0.2 to 54 ng/ul (Table 2). To assess the serum-derived cfDNA integrity and quality, we performed a capillary electrophoretic separation using the TapeStation system (Agilent Technology). Electropherograms were generated for each sample and the fragment size of the cfDNA measured between 2 markers against fluorescence intensity. The mean of cfDNA fragment size distribution ranged from 106 to 216 bp (average, 136 bp), with no significant differences in cfDNA fragment size between all samples (Figure 1). Even though the serum samples showed contamination with high molecular weight genomic DNA in comparison to samples extracted from plasma (gDNA) (Figure 1), it was observed that the amount of cfDNA was more than the gDNA (Figure 1).

Taken together, cfDNA was successfully extracted from all the samples with sufficient quality for further sequencing analysis, suggesting the feasibility of the use of long-storage serum for molecular analysis.

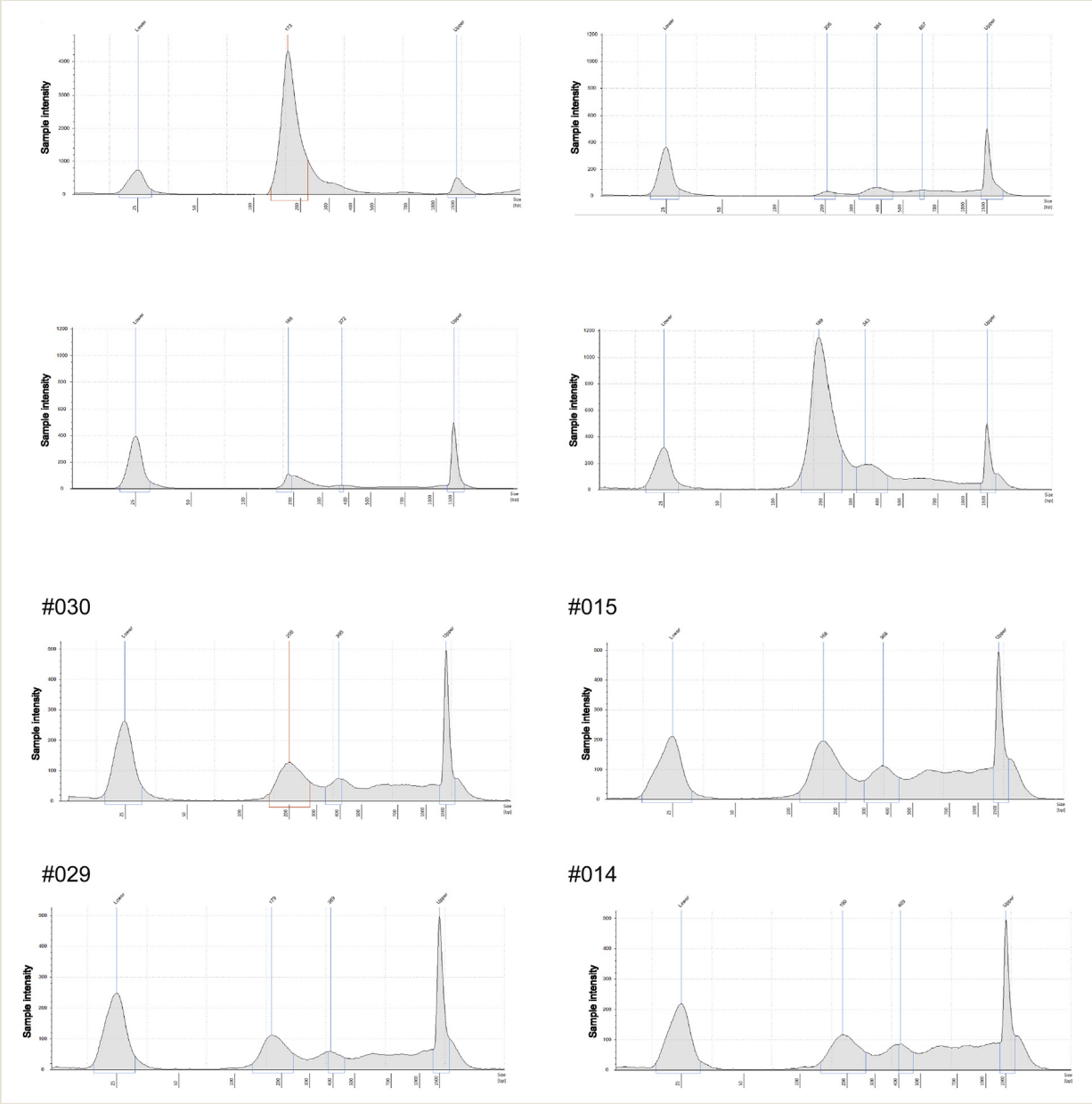
Targeted Sequencing Showed Breast Cancer-specific Somatic Mutations in cfDNA

From the 52 extracted cfDNA samples, we randomly selected 25 for subsequent mutation investigation. Sequencing was performed using 2 different targeted panels. Ten samples were sequenced using the Oncomine Breast cfDNA Assay, which covers the most common hotspots in 10 highly mutated genes in breast cancer.

Table 2 Circulating Free DNA Extraction Data

Sample Name	Serum, mL	Concentration, ng/uL
#001	3	1.5
#002	4	2.5
#003	3.2	9.6
#004	3	0.6
#005	3.5	8.8
#006	3.5	10.3
#007	2.5	12.1
#008	2	1.4
#009	3	3.7
#010	3	3.4
#011	3.5	10
#012	4	10.5
#013	3.2	6.7
#014	2.5	6.5
#015	2.5	5.4
#016	3	3.7
#017	3	1.2
#018	3	0.25
#019	3.5	3.4
#020	3	3.4
#021	3.2	7.5
#022	2	2.5
#023	4	1.1
#024	3	2.5
#025	4	28
#026	3	3.9
#027	3.5	4.9
#028	4	8.1
#029	3	9.1
#030	3	2.9
#031	3	2.3
#032	3.5	1.7
#033	4	2.8
#034	3	1.4
#035	3	5.3
#036	3	2.3
#037	3	0.8
#038	2.5	7.9
#039	3	1.8
#040	3	4.2
#041	4	19.3
#042	3.5	4.2
#043	2.5	9.2
#044	2.5	17.8
#045	2.5	2.4
#046	3.5	2.7
#047	2.5	1.6
#048	4	54
#049	2.5	6.4
#050	3	1.9
#051	3	15.3
#052	2.5	0.7

Figure 1 Circulating Free DNA Analysis Using the TapeStation System. Representative Electropherograms of Total Extracted Circulating Free DNA From 4 Patients Selected for Sequencing (#030, #015, #029, #014)



Owing to the molecular barcoding, this panel allows for the identification of mutations at a very low allelic frequency. We obtained a mean sequencing depth of 55,612X (ranging from 5227 to 108,393) and identified somatic mutations in 8 of the tested samples encompassing *KRAS, TP53* (Table 3A and Figure 2). The other 15 samples were instead sequenced with a custom targeted sequencing panel that covers all exons of 27 protein-coding genes as well as mutation hotspots in 3 cancer genes and the recurrently mutated lncRNA genes *MALAT1* and *NEAT1*[28] (see Supplemental Table 1 in the online version). In this second round of sequencing, we obtained a mean sequencing depth of 2891X (ranging from 903 to 18,210), and we identified somatic mutations in 12 of the tested samples (Table 3B and Figure 2). Mutations were detected in some of the most commonly mutated genes in breast cancer, as *TP53*(p.Arg248Trp) and *PTEN*(p.Phe278Leu).

Taken together, a total of 24 cancer-specific mutations in 10 of the most commonly mutated breast cancer genes were found in 22 (88%) of 25 randomly selected 30-year-old serum-derived cfDNA from patients with breast cancer (Figure 2), suggesting the feasibility of using very old serum samples for mutational profiles.

Discussion

The present cohort of patients with breast cancer with complete long-term follow-up and available blood samples taken 30 years ago

Table 3 Depth of Sequencing

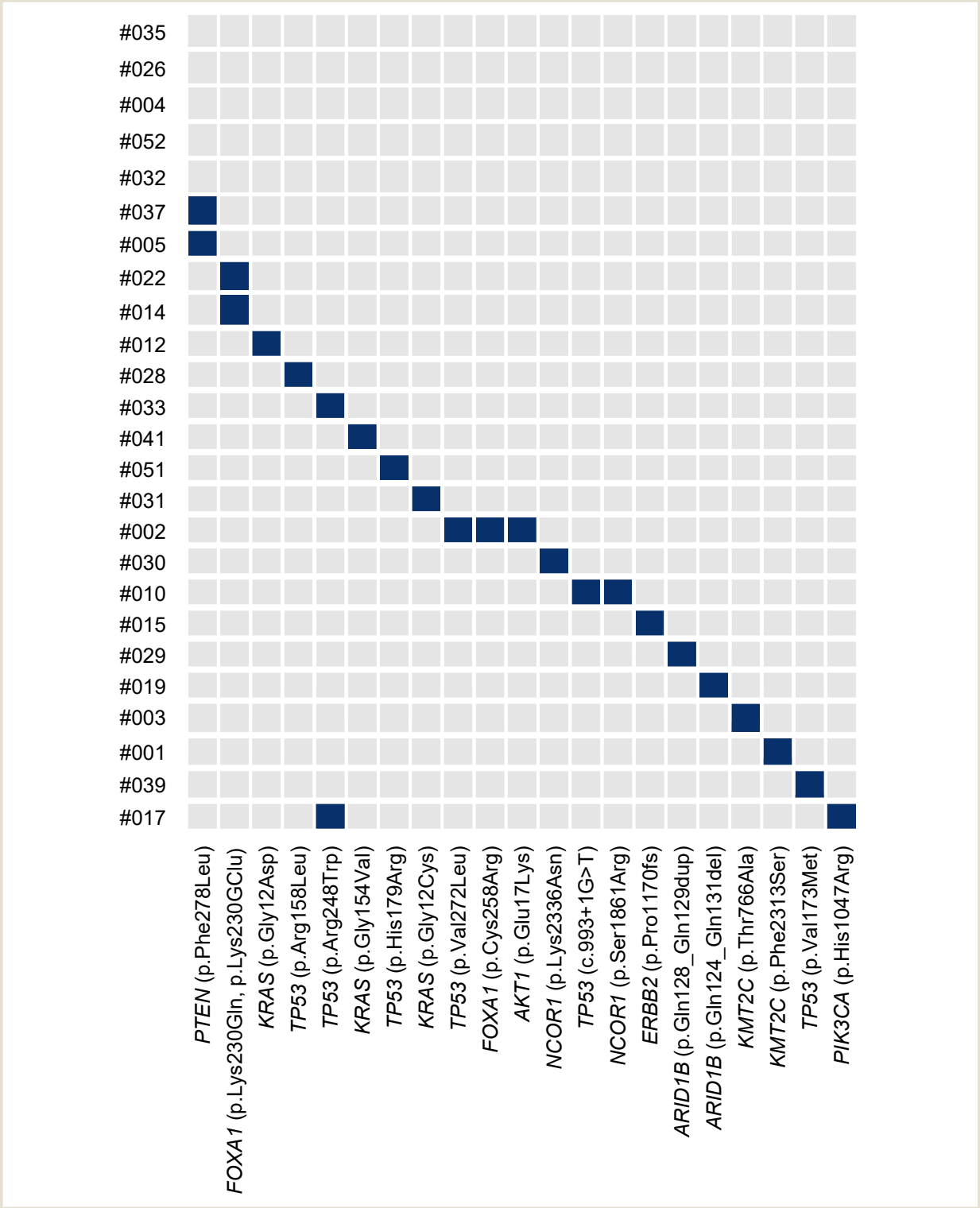
A, Sequencing with Oncomine™ Breast Circulating Free DNA Assay

Sample Name	Mapped Reads	On Target	Mean Coverage	KRAS	TP53	PIK3CA
#012	822,290	95.84	32,400	p.G12D (0.1%)		
#017	1,513,454	96.98	60,944		p.R248W (0.1%)	p.H1047R (2.84%)
#026	1,430,359	95.98	56,798			
#028	17,27,543	97.76	73,289		p.R158L (1.58%)	
#031	1,719,712	96.01	69,907	p.G12C (0.1%)		
#033	978,141	96.31	39,142		p.R248W (0.1%)	
#035	1,084,084	95.86	43,583			
#039	135,999	97.44	5227		p.V173M (0.46%)	
#041	1,626,548	96.08	66,433	p.G154V (0.06%)		
#051	2,656,056	96.21	108,393		p.H179R (0.14%)	

B, Sequencing With Custom Ampliseq Ion Torrent Panel

Sample	Mapped Reads	On Target	Mean Coverage	PTEN	TP53	AKT1	FOXA1	ARID1B	NCOR1	ERBB2	KMT2C
#001	1,695,052	91.83	904								p.Phe2313Ser (0.51)
#002	33,435,937	92.57	18,210		p.Val272Leu (0.08)	p.Glu17Lys (0.13)	p.Cys258Arg (0.4)				
#003	3,398,689	93.08	1879								p.Thr766Ala (0.51)
#004	3,357,088	92.24	1729								
#005	3,924,417	92.69	2151	p.Phe278Leu (0.25)							
#010	3,138,700	93.12	1589		c.993+1G>T (0.56)				p.Ser1861Arg (0.16)		
#014	2,284,741	93.62	1265				p.Lys230Gln (0.16)				
#015	6,629,734	94.07	3529							p.Pro1170fs (0.1)	
#019	4,542,595	93.35	2516					p.Gln124_Gln131del (0.07)			
#022	2,169,116	93.45	1160				p.Lys230Gln (0.13)				
#029	3,605,091	93.79	1975					p.Gln128_Gln129dup (0.23)			
#030	3,090,414	94.6	1606						p.Lys2336Asn (0.48)		
#032	3,743,126	91.21	2031								
#037	2,094,065	91.87	1093	p.Phe278Leu (0.27)							
#052	3,098,154	93.8	1689								

Figure 2 Heatmap of the Comparison of all the Somatic Mutations Found in 25 Randomly Selected Samples Across the Whole Cohort. Rows Indicate the Specific Samples, Columns Indicate the Specific Mutations, Annotated With the Gene Name and the Specific Aminoacidic Change. The Heatmap Shows the Presence (Blue Cells) or the Lack (Grey Cells) of Mutations



represents an exceptional opportunity to study the potential of liquid biopsy-based biomarker identification. The patients were treated in curative or palliative intent for stage I to IV breast cancer. Treatment heterogeneity was limited at that time because only tamoxifen and CMF or LMF (cyclophosphamide- or chlorambucil-methotrexate-fluorouracil) chemotherapy were used in most patients who received systemic treatment. The majority of the patient population recurred at some point, which was the reason why 35 (62%) of the 52 patients died from breast cancer in this cohort. The collection of serum started in 1983, whereas the dates of first diagnosis and treatment go back to 1967. cfDNA was obtained in sufficient quantity and quality for sequencing in all 52 patients. The sample size was too small to make any firm conclusions on differences between patients with short and long-term survival. However, the present results suggest that this modern technology can be used to accurately extract and sequence ctDNA to detect cancer-specific mutations in these old samples, despite the long cryopreservation and repeated changes of storage location.

These findings support the use of long-term storage of biological samples in longitudinal studies prior to analysis, which, in turn, will increase feasibility by making the study protocols less depending on consecutive and timely processing at the centralized high-depth targeted sequencing unit. The principle of long-term storage may facilitate the performance of large international studies that assess the prognostic role of cancer-associated pathogenic mutations in serum cfDNA present at diagnosis by comparing overall and relapse-free survival between patients with or without specific mutations. Therefore, another potential value of using samples of patients diagnosed a long time ago is to increase the number of events (eg, relapse, deaths) and increase the statistical power for survival analyses (study of the prognostic value of the identified mutations in ctDNA). The predictive power of response-associated mutations can then be assessed based on in silico mutation effect predictors and curated databases of cancer- and response-associated variants.[29-35] One would hypothesize that patients with detectable mutations in the cfDNA would have higher tumor burden and/or tumor cells with a higher tendency to shed into the bloodstream and, therefore, poorer outcome than patients without detectable mutations in the cfDNA. Candidate somatic mutations can be further evaluated in vitro and in vivo by using xenograft models. For instance, patient-derived breast tumor cells can be engineered to express the same mutation found to be associated with resistance and test their sensitivity to the same targeted therapy in xenograft models compared with control cancer cells (ie, wild-type in the corresponding allele).

The blood samples analyzed here were taken at a time when physicians could not anticipate NGS approaches. Nevertheless, substantial efforts were made to collect the samples under the assumption that someday technology would have advanced to the point where relevant research could be performed with a few mL of serum and matched clinical data. Storing blood samples over the entire duration of longitudinal studies allows newly developed technology to analyze cfDNA more thoroughly and homogenously. Hence, the most modern state-of-the-art technology for nucleic acids extraction, sequencing, data analysis, and new targeted panels that may only become available at the end of the study can be applied to all serial blood samples, which increases data quality and comparability. Innovative studies can be designed to track the evolution of disease-associated mutations in the serum cfDNA. This would allow to evaluate if variations in the tumor allele fractions of the mutations mirror the genetic heterogeneity in the tumors and to determine if disease progression is associated with the emergence of additional somatic mutations. This, in turn, may help to assess whether mutational evolution reflects radiologically determined disease burden, recurrence, or metastasis.

This study has several limitations. First, some of the mutations may have been germline variants, especially those at high allelic frequencies. We cannot exclude this possibility owing to the lack of germline controls or clonal hematopoiesis. However, the primary aim of the study was to determine whether it was possible to identify mutations in 30-year-old serum, and the exclusion of germline variants can be achieved by using germline control. Second, the custom panel we used in this study was not optimized for mutation detection in cfDNA, because some of the amplicons are bigger than the average size of cfDNA fragments. We may thus have missed some mutations, and the use of a panel with smaller amplicon size will likely increase the number of mutations that can be detected. Third, another important limitation is the small sample size that precluded any analyses on associations between mutations and clinical endpoints. For example, it would be interesting to see if patients with detectable mutations in the cfDNA have a higher tumor stage and therefore poorer outcome than patients without detectable mutations. This has been shown in patients with late-stage gastric cancer, where patients with detectable mutations had a 5.6% 5-year overall survival rate compared with 31.5% in patients without detectable mutations.[36] TP53 is one of the most frequently mutated genes in breast cancer, and, being a tumor suppressor and usually associated with the loss of the wild-type allele, TP53 mutations are likely to be more readily detectable in cfDNA than activating oncogenic mutations. In fact, 7 of 24 detected mutations in this series were TP53 mutations. However, it would be very challenging to adjust for the selection bias in this series of high-risk patients referred to the medical oncologist for systemic treatment even if a higher sample size could have been achieved. Nevertheless, evaluating associations between mutations and clinical endpoints is an area of high potential relevance, particularly when DNA from matched archival tissue of primary tumors or distant metastases are available. As outlook for future projects, we plan to assess the prognostic role of cancer-associated mutations in the serum cfDNA at diagnosis with the extensive follow-up information and clinicopathologic parameters available for our unique cohort of patients.

In conclusion, the present study shows that current NGS technology is sufficiently robust and specific to analyze 30-year-old serum. Based on this finding, longitudinal studies can be designed to be more feasible and flexible by storing biological samples over a long period of time. This allows for uniform sequencing with the most modern technology and adequate statistical power by cumulating oncologic events. Our study supports the value of liquid biopsies in assessing the dynamic changes of genetic heterogeneity over time and in the validation of new cfDNA biomarkers for breast cancer.

Clinical Practice Points

- The potential of using cfDNA as an indicator of disease burden with prognostic implication and clinical applicability during follow-up and monitoring in both the curative and palliative

setting has been investigated in numerous studies. Cancer-specific mutations, copy number alterations, and genomic rearrangements assessed in ctDNA demonstrated potential prognostic and predictive significance.

- To further evaluate prognostic and predictive biomarkers in cfDNA and assess its value as a disease monitoring tool, longitudinal studies with long follow-up are necessary. The utility of the technology depends on its capability to assess sequential samples that have been collected and stored over a long period of time.
- The aim of this study was to assess the feasibility of cfDNA extraction and somatic mutation assessment in 30-year-old serum. We evaluated samples from 52 patients with breast cancer, which were collected between 1983 and 1991. cfDNA extraction was successful in 52 of 52 patients, and 24 cancer-specific mutations were found in 22 of 25 samples undergoing sequencing.
- Our results suggest that current NGS technology is sufficiently robust and specific to analyze 30-year-old serum. Based on this finding, longitudinal studies can be designed to be more feasible and flexible by storing biological samples over a long period of time. This allows for uniform sequencing with the most modern technology and adequate statistical power by cumulating oncologic events. Our study supports the value of liquid biopsies in assessing the dynamic changes of genetic heterogeneity over time and in the validation of new cfDNA biomarkers for breast cancer.

Acknowledgments

The authors thank Michelle Attenhofer and Stefan Herms for their assistance with sample and data storage over the long period of time. The authors thank Philip Martin Jermann for critical review of the manuscript. The authors also thank Basel Cancer League for paying the first freezer in 1983.

Partially, this work was funded by the "OPO-Stiftung" and the "Frieder Locher-Hofmann-Stiftung." This funding source had no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Disclosure

The authors have stated that they have no conflicts of interest.

Supplemental Data

Supplemental table accompanying this article can be found in the online version at https://doi.org/10.1016/j.clbc.2020.04.005.

References

1. Joyner MJ, Paneth N. Seven questions for personalized medicine. *JAMA* 2015; 314:999-1000.
2. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; 490:61-70.
3. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn* 2015; 17:251-64.
4. Chen K, Meric-Bernstam F, Zhao H, et al. Clinical actionability enhanced through deep targeted sequencing of solid tumors. *Clin Chem* 2015; 61:544-53.
5. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; 366: 883-92.
6. Yates LR, Gerstung M, Knappskog S, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 2015; 21:751-9.
7. Wan JC, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 2017; 17: 223-38.
8. Bidard FC, Weigelt B, Reis-Filho JS. Going with the flow: from circulating tumor cells to DNA. *Sci Transl Med* 2013; 5:207ps14.
9. Jung K, Fleischhacker M, Rabien A. Cell-free DNA in the blood as a solid tumor biomarker–a critical appraisal of the literature. *Clin Chim Acta* 2010; 411:1611-24.
10. 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.
11. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA. Isolation and characterization of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol* 1987; 23:707-12.
12. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev* 2016; 35: 347-76.
13. Schwarzenbach H, Pantel K. Circulating DNA as biomarker in breast cancer. *Breast Cancer Res* 2015; 17:136.
14. Umetani N, Giuliano AE, Hiramatsu SH, et al. Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol* 2006; 24: 4270-6.
15. Dawson SJ, Rosenfeld N, Caldas C. Circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* 2013; 369:93-4.
16. Dawson SJ, Tsui DW, Murtaza M, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* 2013; 368:1199-209.
17. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 2011; 11:426-37.
18. Garcia-Murillas I, Schiavon G, Weigelt B, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* 2015; 7: 302ra133.
19. De Mattos-Arruda L, Weigelt B, Cortes J, 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.
20. Schiavon G, Hrebien S, Garcia-Murillas I, et al. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* 2015; 7:313ra182.
21. De Mattos-Arruda L, Mayor R, Ng CKY, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* 2015; 6:8839.
22. Luen S, Virassamy B, Savas P, Salgado R, Loi S. The genomic landscape of breast cancer and its interaction with host immunity. *Breast* 2016; 29:241-50.
23. Paradiso V, Garofoli A, Tosti N, et al. Diagnostic targeted sequencing panel for hepatocellular carcinoma genomic screening. *J Mol Diagn* 2018; 20: 836-48.
24. Garofoli A, Paradiso V, Montazeri H, et al. PipeIT: a singularity container for molecular diagnostic somatic variant calling on the Ion Torrent next-generation sequencing platform. *J Mol Diagn* 2019; 21:884-94.
25. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015; 526:68-74.
26. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60,000 exomes. *Nucleic Acids Res* 2017; 45: D840-5.
27. National Heart, Lung, and Blood Institute. NHLBI Grand Opportunity Exome Sequencing Project (ESP), Available at: https://esp.gs.washington.edu/drupal/. accessed January 8, 2020.
28. Soysal SD, Ng CKY, Costa L, et al. Genetic alterations in benign breast biopsies of subsequent breast cancer patients. *Front Med (Lausanne)* 2019; 6:166.
29. Carter H, Chen S, Isik L, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res* 2009; 69:6660-7.
30. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013; 34:57-65.
31. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; 7:575-6.
32. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012; 7:e46688.
33. Chakravarty D, Gao J, Phillips SM, et al. OncoKB: a precision oncology knowledge base. *JCO Precis Oncol* 2017; 2017.
34. Wagner AH, Coffman AC, Ainscough BJ, et al. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res* 2016; 44:D1036-44.
35. Griffith M, Griffith OL, Coffman AC, et al. DGIdb: mining the druggable genome. *Nat Methods* 2013; 10:1209-10.
36. Fang WL, Lan YT, Huang KH, et al. Clinical significance of circulating plasma DNA in gastric cancer. *Int J Cancer* 2016; 138:2974-83.

Supplemental Data

Supplemental Table 1 Gene Coverage of the Custom Targeted Sequencing Panel

Gene Name	Chromosome	Start	End	Cytoband	Remarks
ARID1A	1	27022524	27108595	p36.11	Complete coding region
NRAS	1	115247090	115259515	p13.2	Hotspot residues 12,13 and 61 only
SETD2	3	47057919	47205457	p21.31	Complete coding region
PIK3CA	3	178865902	178957881	q26.32	Complete coding region
FBXW7	4	153242410	153457253	q31.3	Complete coding region
MAP3K1	5	56111401	56191979	q11.2	Complete coding region
PIK3R1	5	67511548	67597649	q13.1	Complete coding region
ARID1B	6	157099063	157531913	q25.3	Complete coding region
EGFR	7	55086714	55324313	p11.2	Complete coding region
KMT2C	7	151832010	152133090	q36.1	Complete coding region
PTPRD	9	8314246	10612723	p23	Complete coding region
GAT A3	10	8095567	8117161	p14	Complete coding region
PTEN	10	89622870	89731687	q23.31	Complete coding region
HRAS	11	532242	537287	p15.5	Hotspot residues 12, 13, and 61 only
NEAT1	11	65190245	65213011	q13.1	Complete coding region
MALAT1	11	65265233	65273940	q13.1	Complete coding region
ATM	11	108093211	108239829	q22.3	Complete coding region
KRAS	12	25357723	25403870	p12.1	Hotspot residues 12, 13, and 61 only
ERBB3	12	56473641	56497289	q13.2	Complete coding region
TBX3	12	115108059	115121969	q24.21	Complete coding region
RBI	13	48877887	49056122	q14.2	Complete coding region
F0XA1	14	38059189	38069245	q21.1	Complete coding region
AKT1	14	105235686	105262088	q32.33	Complete coding region
CBFB	16	67063019	67134961	q22.1	Complete coding region
CTCF	16	67596310	67673086	q22.1	Complete coding region
CDH1	16	68771128	68869451	q22.1	Complete coding region
TP53	17	7565097	7590856	p13.1	Complete coding region
MAP2K4	17	11924141	12047147	P12	Complete coding region
NCOR1	17	15932471	16121499	p11.2	Complete coding region
NF1	17	29421945	29709134	q11.2	Complete coding region
ERBB2	17	37844167	37886679	q12	Complete coding region
RUNX1	21	36160098	37376965	q22.12	Complete coding region