

ARTICLE

Open Access

Cell-free DNA analysis in healthy individuals by next-generation sequencing: a proof of concept and technical validation study

Ilaria Alborelli¹, Daniele Generali^{2,3}, Philip Jermann¹, Maria Rosa Cappelletti², Giuseppina Ferrero², Bruna Scaggiante⁴, Marina Bortul³, Fabrizio Zanconati³, Stefan Nicolet¹, Jasmin Haegele^{1,11}, Lukas Bubendorf¹, Nicola Aceto⁵, Maurizio Scaltriti^{6,7}, Giuseppe Mucci⁸, Luca Quagliata^{1,12} and Giuseppe Novelli^{9,10}

Abstract

Pre-symptomatic screening of genetic alterations might help identify subpopulations of individuals that could enter into early access prevention programs. Since liquid biopsy is minimally invasive it can be used for longitudinal studies in healthy volunteers to monitor events of progression from normal tissue to pre-cancerous and cancerous condition. Yet, cell-free DNA (cfDNA) analysis in healthy individuals comes with substantial challenges such as the lack of large cohort studies addressing the impact of mutations in healthy individuals or the low abundance of cfDNA in plasma. In this study, we aimed to investigate the technical feasibility of cfDNA analysis in a collection of 114 clinically healthy individuals. We first addressed the impact of pre-analytical factors such as cfDNA yield and quality on sequencing performance and compared healthy to cancer donor samples. We then confirmed the validity of our testing strategy by evaluating the mutational status concordance in matched tissue and plasma specimens collected from cancer patients. Finally, we screened our group of healthy donors for genetic alterations, comparing individuals who did not develop any tumor to patients who developed either a benign neoplasm or cancer during 1–10 years of follow-up time. To conclude, we have established a rapid and reliable liquid biopsy workflow that allowed us to study genomic alterations with a limit of detection as low as 0.08% of variant allelic frequency in healthy individuals. We detected pathogenic cancer mutations in four healthy donors that later developed a benign neoplasm or invasive breast cancer up to 10 years after blood collection. Even though larger prospective studies are needed to address the specificity and sensitivity of liquid biopsy as a clinical tool for early cancer detection, systematic screening of healthy individuals will help understanding early events of tumor formation.

Introduction

Genomic instability arises in normal cells through accumulation of genetic and epigenetic changes and has

been shown to occur over a variable time span, ranging from years to decades^{1–4}. The vast majority of normal cells that have acquired mutations is cleared away by the immune system, while a minimal fraction might eventually progress and give rise to cancer⁵. Upon development of cancer, specific genomic alterations can be identified and used to provide the rationale for specific treatment options. Monitoring genomic changes could hence be crucial to identify early mutational events that

Correspondence: Ilaria Alborelli (ilaria.alborelli@usb.ch)

¹Institute of Pathology, University Hospital Basel, 4031 Basel, Switzerland

²Breast Cancer Unit and Translational Research Unit, ASST Cremona, Viale Concordia 1, 26100 Cremona, Italy

Full list of author information is available at the end of the article.

Edited by A. Stephanou

© The Author(s) 2019



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

are associated with higher risk of developing cancer, but it is largely unfeasible using traditional tissue-based approaches due to the lack of observable tumor lesions. In contrast, liquid biopsy might offer the possibility of detecting early genomic aberrations and investigating cancer evolution in a minimally invasive fashion. Liquid biopsy is a broad term that refers to testing body fluids such as blood or urine for biomarkers reconcilable with a medical condition. In the field of oncology, liquid biopsy mainly pertains to the analysis of circulating tumor DNA (ctDNA) in blood. Circulating tumor DNA represents only a minor fraction (<0.1–10%) of the total circulating cell-free DNA (cfDNA)⁶, which is derived by cell death associated to physiological tissue remodeling events⁷. The majority of DNA fragments found in the circulation measures ~180 nucleotides in size⁸, suggesting that apoptosis and necrosis are responsible for cfDNA shedding. Interestingly, the blood of cancer patients typically presents higher levels of circulating cfDNA compared to healthy individuals^{9,10}.

A growing body of evidence supports ctDNA-based analysis of cancer-associated hotspot mutations as a cost-effective and highly sensitive tool, complementary to tissue molecular profiling^{11–15}. In clinical settings, ctDNA analysis has been applied to monitor response to treatment, to detect residual disease and to identify mechanisms of resistance to therapy^{16–18}. Currently, the most common clinical use of liquid biopsy is the detection of resistance-associated mutations to inform treatment decision^{19–22}. The introduction of molecular barcodes has considerably enhanced the sensitivity of sequencing methods at the price of additional costs linked to the high depth of sequencing required (i.e. ~25,000 coverage)^{13,23,24}. Taking advantage of this and further technological developments, several studies have described clinically relevant genetic alterations in patients with early-stage cancers at a sensitivity below one mutant template molecules per milliliter of plasma^{9,25–28}. Nevertheless, among the potential clinical applications of ctDNA analysis, early detection remains the most ambitious. Several challenges need to be addressed and large validation studies will be required to establish the sensitivity and specificity for such testing approach²⁹. The presence of somatic mutations in asymptomatic patients, related to clonal hematopoiesis²⁸ as well as clonal expansion in healthy tissue^{30–33}, could potentially lead to false-positive calls. Moreover, the recovery and characterization of cfDNA in healthy individuals might prove challenging, given that cfDNA is less abundant in these subjects^{34,35} and only a few studies have reported the analysis of cfDNA in healthy controls^{28,36}.

An adequate technical validation is therefore required to allow the implementation of liquid biopsy as a tool for early cancer detection and prove that extraction and

analysis of cfDNA isolated from healthy individuals is technically achievable. To this end in our proof of principle study, we examined the feasibility cfDNA interrogation in a collection of 114 individuals that were clinically healthy (i.e. not affected by any manifest medical condition) at the time of blood draw.

First, we analyzed the impact of pre-analytical factors such as cfDNA yield and quality on sequencing performance parameters as molecular coverage and limit of detection, comparing samples from healthy and cancer donors.

We then evaluated the reliability of our testing strategy by analyzing cfDNA samples obtained from patients with a histologically confirmed diagnosis of breast or lung cancer. Specifically, we assessed the concordance of specific genetic alterations detected in matched tissue and plasma specimens. Finally, we investigated the mutational status of a group of healthy donors comprising both individuals that did not develop any tumor within 1 to 10 years (average = 8.5 years; Table 1) of follow-up as well as individuals that developed either a benign neoplasm or cancer during the follow-up time. Altogether, our study demonstrates the technical feasibility of extracting and analyzing cfDNA in healthy individuals to study genomic alterations, by means of molecular barcoded next-generation sequencing (NGS).

Materials and methods

Patients

One hundred and fourteen healthy donors undergoing a control screening mammography test and nine breast cancer patients undergoing treatment at the Breast Cancer Unit and Translational Research Unit of the Hospital of Cremona (Italy) were selected for this study (Ethical approval protocol nr. Ex01/4111/04). In addition, 54 lung cancer patients undergoing treatment at the University Hospital Basel (Switzerland) were selected for this study (Ethical approval protocol EKBB/EKNZ 31/12). The study was performed in compliance with all relevant ethical regulations. More plasma was available in the cancer patient group (from 1.5 to 5.5 ml) in cancer patients because 2 × 10 ml of whole blood was collected for each patient. Conversely, only 1 × 10 ml whole blood was drawn from healthy individuals, part of which was used for other analyses, resulting in a final plasma volume ranging from 0.4 to 2.0 ml. Follow-up data have been collected in the frame of this study only for donors that were clinically healthy at blood collection.

cfDNA/DNA extraction from plasma and tissue samples

Blood samples were collected in either K₂EDTA tubes (BD Vacutainer® Blood Collection Tubes, Becton Dickinson, Franklin Lakes, USA) and Cell-Free DNA BCT® (Streck, La Vista, NE). The plasma fraction was separated

Table 1 Patient characteristics

Patient characteristics (n = 177)	n (%)
Age (years)	
Mean (SD)	63 (11)
Sex (n)	
Male	19 (11)
Female	157 (89)
Clinical status at blood collection (n)	
No tumor (healthy)	114 (64)
Breast cancer	9 (5)
Lung cancer	54 (31)
Clinical status at follow-up ^a (n)	
No tumor (healthy)—Group I	25 (14)
Benign breast condition—Group II	52 (29)
Breast cancer—Group III	15 (8)
Other tumors—Group IV	14 (8)
Missing information (lost to follow-up)	8 (5)
Follow-up* time from blood collection (years)	
Mean (range)—All Groups	8.5 (1.1–10.1)
Mean (range)—Group I	8.7 (4.6–10.1)
Mean (range)—Group II	8.4 (1.1–10.0)
Mean (range)—Group III	9.0 (4.3–10.0)
Mean (range)—Group IV	8.4 (4.1–9.6)
Molecular analysis (n)	
Plasma—cfDNA extraction	177 (100)
Plasma—NGS analysis	93 (53)
Tissue—NGS analysis	38 (21)

^aFollow-up data have been collected in the frame of this study only for donors that were clinically healthy at blood collection

from the blood cells by two consecutive rounds of centrifugation for 30 min at room temperature at 1600 × g. The collected plasma was aliquoted and stored at −80 °C until use. cfDNA was extracted from plasma volumes ranging from 0.4 to 5.5 ml using the MagMax Cell-Free Total Nucleic Acid Isolation Kit (Thermo Fisher Scientific, Waltham, USA) according to the manufacturers' instructions. The cfDNA quantity was assessed with the dsDNA HS assay kit by the Qubit 2.0 Fluorometer (Thermo Fisher Scientific). cfDNA quality was assessed with the Agilent High Sensitivity D1000 ScreenTape System (Agilent Technologies, Santa Clara, USA). Only cfDNA samples with a clear fragment size peak between 140–200 bp (Supplementary Fig. 1) were considered for analysis.

Tissue biopsies were obtained at the time of first diagnosis and inspected through examination of hematoxylin and eosin-stained slides by a thoracic pathologist. For DNA extraction, 4–5 FFPE tissue sections of 10 μm thickness were cut and deparaffinized using Xylo. DNA extraction from tissue was performed using the column-based RecoverAll Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. DNA quantity was assessed with the dsDNA HS assay kit by the Qubit 2.0 Fluorometer (Thermo Fisher Scientific).

NGS library preparation

For plasma samples, NGS libraries were prepared from 2.5 to 105.5 ng of cfDNA following the HeliXmoker, HeliXgyn, and HeliXafe workflows (patented by The Bioscience Institute), based on the Oncomine™ Lung cfDNA Assay v1, the Oncomine™ Breast cfDNA Research Assay v2, and the Oncomine™ Pan-Cancer Cell-Free Assay (Thermo Fisher Scientific), respectively. Only six samples were selected for a broader mutational analysis using the HeliXafe protocol, based on the cfDNA concentration and quality, providing that these six samples matched the required minimum input for a second round of library preparation. Our general library preparation protocol was based on a two-cycle multiplex touch-down PCR reaction with a temperature range from 64 °C to 58 °C, which allowed to amplify target regions and introduce unique molecular identifiers. The resulting tagged amplicons of around 100–140 bp length were then cleaned up using Agencourt AMPure XP (Beckman Coulter, Brea, USA) at a bead to sample ratio of 1.5× and purified products were 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 introduce Ion Torrent™ Tag-Sequencing adapters containing sample-specific barcodes. The resulting library of target DNA fragments was purified by performing a two-step cleanup using Agencourt AMPure XP (Beckman Coulter) at a bead to sample ratio of 1.15× and 1.0×, respectively. 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.

For NGS library preparation from tissue samples 5–40 ng of DNA was used, depending on availability of input material. Libraries were prepared according to protocol (Oncomine™ Solid Tumor Assay, Oncomine™ Focus Assay, Oncomine™ Comprehensive v3 Assay were used (Supplementary Table 1)). The resulting libraries were purified using Agencourt AMPure XP (Beckman Coulter). Libraries were quantified by qPCR using the Ion Universal Quantitation Kit (Thermo Fisher Scientific), diluted to 50 pM and batched according to the manufacturer's instructions.

Sequencing

NGS libraries were sequenced on an Ion S5™ instrument (Thermo Fisher Scientific) using semiconductor sequencing technology. Briefly, sequencing runs were planned on the Torrent Suite Software™ v5.8, 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. Raw data were processed automatically on the Torrent Server™ and aligned to the reference hg19 genome. QC was performed manually for each sample based on the following metrics; number of reads per sample > 2,500,000 (for OncoPrint™ Lung cfDNA Assay libraries), >4,000,000 (for OncoPrint™ Breast cfDNA Research Assay v2 libraries) >15,000,000 (for OncoPrint™ Pan-Cancer Cell-Free Assay libraries), on-target reads >90%, read uniformity >90%, median molecular coverage >500×, median read coverage >15,000.

Tissue NGS libraries were sequenced according to the manufacturer's instructions.

The sequencing data of the QC passing samples were then uploaded in BAM format to the Ion Reporter™ Analysis Server for variant calling and annotation.

Data analysis

For plasma samples variant calling was performed on Ion Reporter™ (IR) Analysis Software v5.6 using the OncoPrint™ TagSeq Breast v2 Liquid Biopsy w2.0, OncoPrint™ Lung Liquid Biopsy w1.3, and OncoPrint™ TagSeq Pan-Cancer Liquid Biopsy w2.0 workflows. The analysis pipeline also included signal processing, base calling, quality score assignment, adapter trimming, PCR duplicate removal, and control of mapping quality. Coverage metrics for each amplicon was obtained by running

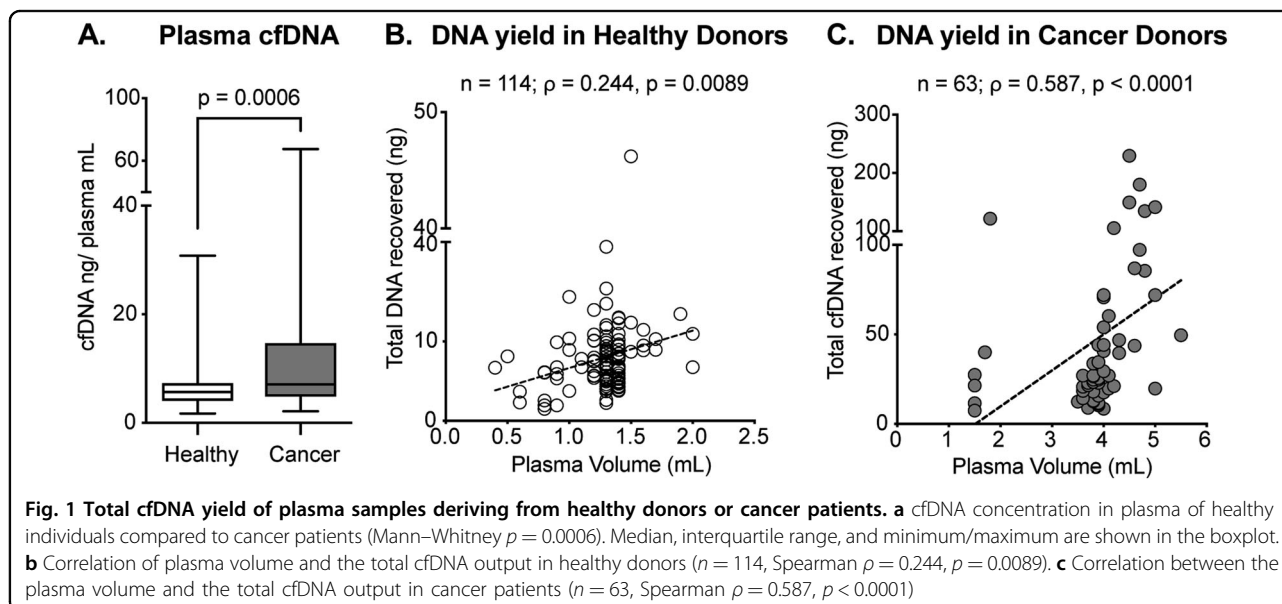
the Coverage Analysis Plugin software v5.6.1 (Thermo Fisher Scientific). Identified variants were only considered if the variant had a molecular coverage of at least three, indicating that the variant was detected in three independent template molecules. Finally, all candidate mutations were manually reviewed using the Integrative Genomics Viewer³⁷.

For tissue samples, the default analysis pipeline in IR (OncoPrint™ Solid Tumor Assay, OncoPrint™ Focus Assay, OncoPrint™ Comprehensive v3) was used.

Results

Plasma volume and cfDNA amount define LOD for variant calling

First, we attempted to establish a solid workflow for the extraction of cfDNA from plasma of either healthy individuals or cancer patients. Table 1 summarizes the analyzed cohort characteristics. Peripheral whole blood was collected in commercial vessels containing EDTA or a preservative agent preventing cell lysis and thus the contamination of circulating cfDNA with cellular DNA. After plasma isolation, we extracted cfDNA using a magnetic beads-based kit as described in detail in the Materials and methods section. The amount of plasma available varied between 0.4 and 2.0 ml in healthy individuals and 1.5 and 5.5 ml in cancer patients (Fig. 1b, c, Materials and methods). As previously reported (first in 1977 (ref. ³⁸)), total cfDNA concentration in plasma was significantly higher in cancer patients compared to healthy subjects ($p = 0.0006$, Fig. 1a). We characterized the correlation between plasma input and total cfDNA yield in samples collected from healthy donors ($\rho = 0.244$, $p = 0.0089$, shown in Fig. 1b) and cancer patients ($\rho = 0.587$,



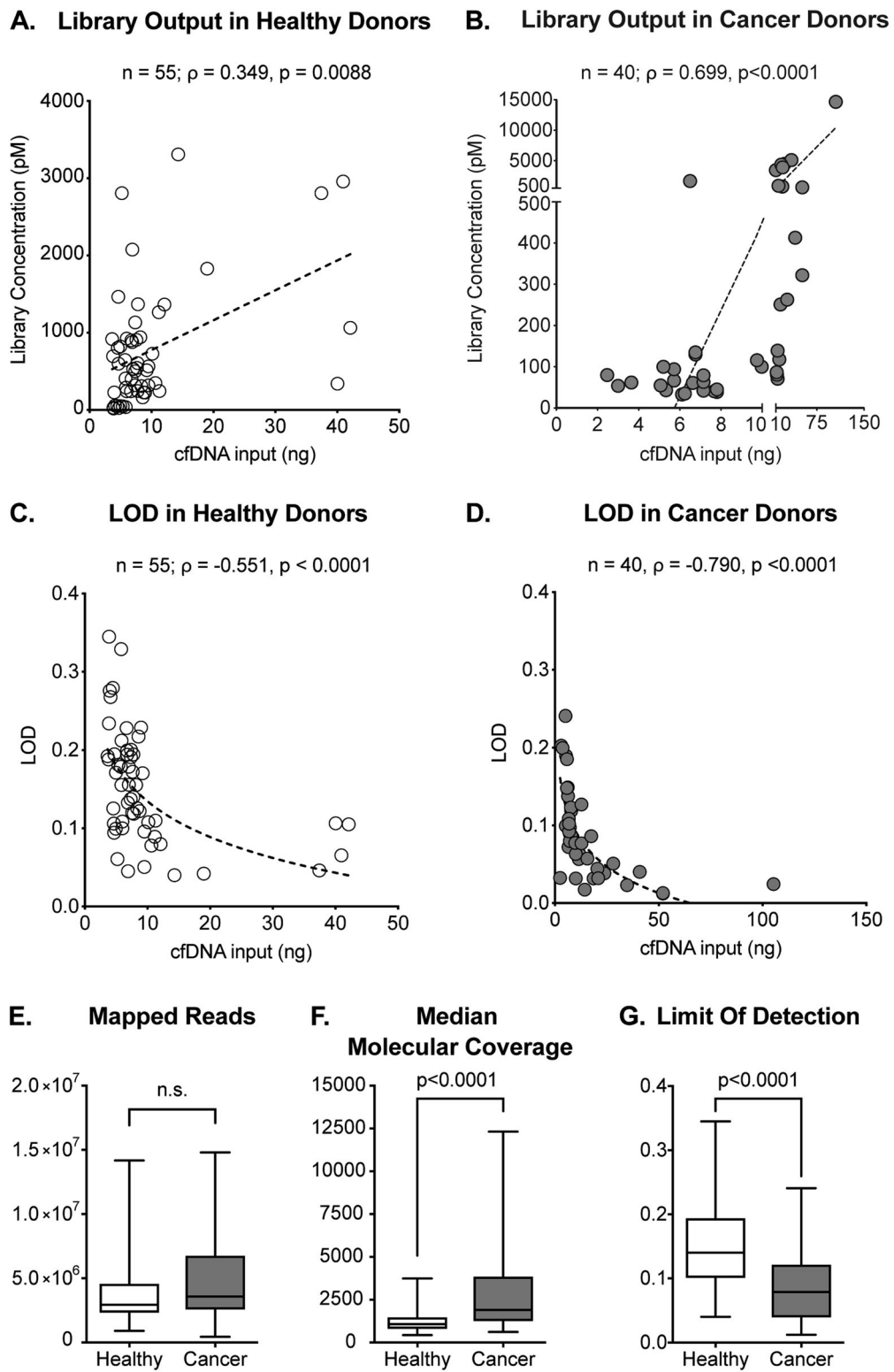


Fig. 2 (See legend on next page.)

(see figure on previous page)

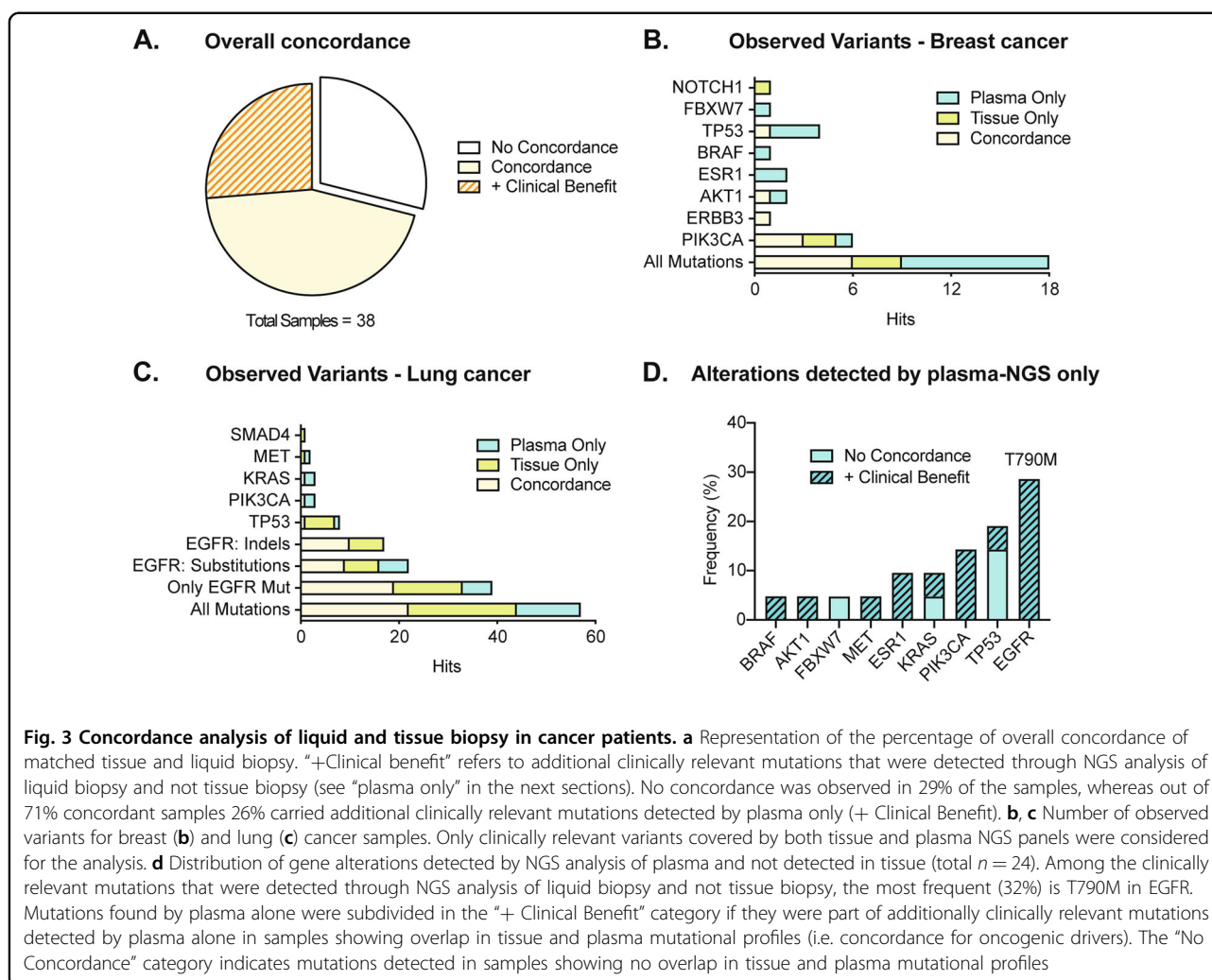
Fig. 2 Comparison of pre-analytical variables from healthy and cancer donor samples. **a, b** Correlation of library concentration and input of cfDNA in healthy individuals ($n = 55$, Spearman $\rho = 0.348$, $p = 0.0088$) and cancer patients ($n = 40$, Spearman $\rho = 0.699$, $p < 0.0001$). **c, d** Correlation of LOD and cfDNA input in healthy ($n = 55$; Spearman $\rho = -0.551$, $p < 0.0001$) and cancer donors ($n = 40$; Spearman $\rho = -0.790$, $p < 0.0001$). **e** Mapped reads of samples deriving from healthy and cancer donors (Mann–Whitney $p = 0.1422$). **f, g** Median molecular coverage (Mann–Whitney $p < 0.0001$) and LOD (Mann–Whitney $p < 0.0001$) in healthy and cancer donors. Median, interquartile range, and minimum/maximum are shown in the boxplot

$p < 0.0001$, shown in Fig. 1c). Next, we processed cfDNA samples from healthy and cancer donors for NGS library preparation and sequencing. NGS library concentration was significantly affected by cfDNA input in both healthy and cancer samples ($\rho = 0.348$, $p = 0.0088$ and $\rho = 0.699$, $p < 0.0001$, respectively, Fig. 2a, b). Notably, as healthy individuals generally present with lower levels of cfDNA compared to cancer patients, limited DNA input was used for library preparation, often below the minimal manufacturers' recommended amount (i.e. 10 ng). We show that the limit of detection (LOD) of our assay, which indicates the lowest variant allelic frequency that could be reliably detected, is clearly affected by the cfDNA abundance in both healthy individuals and cancer patients (Fig. 2c, d) with an inverse correlation between these two variables. Despite comparable sequencing depth in healthy and cancer donor samples (Fig. 2e), we observed higher molecular coverage in cancer samples (Fig. 2f, $p < 0.0001$) due to higher amount of input cfDNA. For the same reason the LOD was significantly lower in cancer patients (Fig. 2g, $p < 0.0001$). Thus, our data show that the amount of cfDNA has a direct impact on sequencing performance and LOD.

cfDNA profiling of cancer patients and concordance with tissue

Previous studies demonstrated the high analytical sensitivity of using molecular barcodes, also referred to as unique molecular identifiers (UMIs^{23,24,39}) for NGS. Here we attempted to investigate the concordance, in terms of corresponding detected variants, between circulating cfDNA and matched tissue from primary tumor or metastasis of the same patient. To this end, we analyzed cfDNA obtained from eight breast cancer patients using the HeliXgyn protocol (developed by the Bioscience Institute and based on the Oncomine™ cfDNA Breast v2 Assay) and 30 non-small cell lung cancer (NSCLC) patients using the HeliXmoker protocol (developed by the Bioscience Institute and based on the Oncomine™ cfDNA Lung Assay) and sequentially compared it with the results obtained by sequencing tissue using a suitable Oncomine™ Assay (detailed about used gene panels in Supplementary Table 1). We used molecular barcoded sequencing (Tag Sequencing barcodes) to profile our liquid biopsy samples. As the target regions of the panels used for cfDNA and

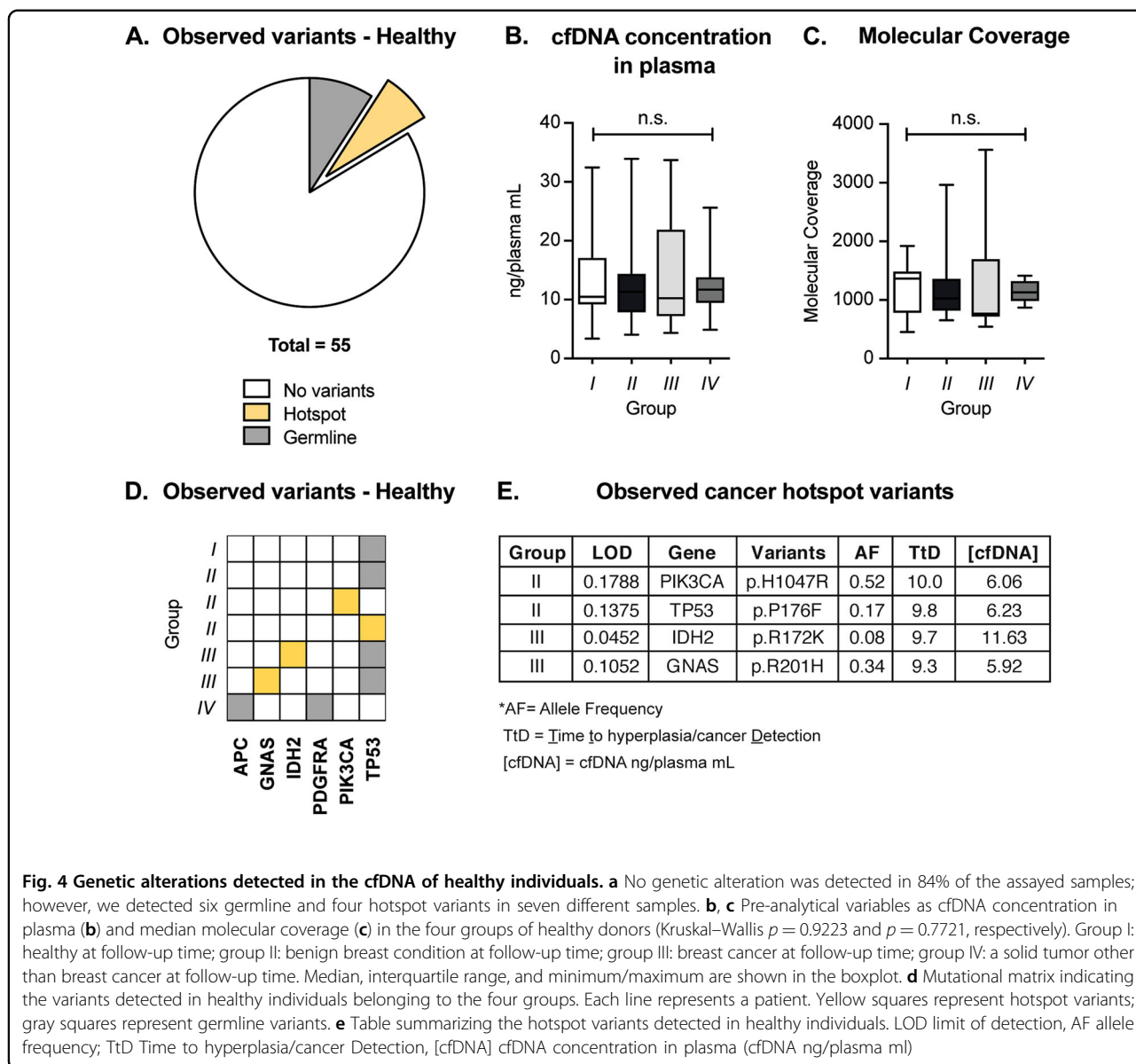
tissue profiling were not fully overlapping, we focused only on clinically relevant mutations covered by both panels (Supplementary Table 1). Our data highlight (Fig. 3a) a substantial level of concordance (71%) between cfDNA and tissue mutational profiles of matched samples. This suggests that cfDNA analysis reliably mimics tissue genomic features. Furthermore, additional clinically relevant mutations were detected by liquid biopsy in 26% of the samples showing a concordant result (Fig. 3a, “plus Clinical Benefit”). The most frequently observed mutations occurred within the coding region of PIK3CA (6 out of 18 mutations detected) for breast cancer (Fig. 3b) and EGFR (40 out of 57 mutations detected) for NSCLC specimens (Fig. 3c). All mutations detected are summarized in a concordance matrix (Supplementary Fig. 2A, B for breast and lung cancer samples, respectively). In breast cancer samples, we found concordance for mutations detected in PIK3CA, AKT1, and ERBB3, whereas mutations in TP53, ESRI, and BRAF were more often detected by plasma alone (Fig. 3b). In lung cancer samples, deletions in the EGFR coding regions were more often detected only by tissue, whereas for substitution in EGFR we observed a more prevalent fraction detected only by plasma (Fig. 3c). The time interval between tissue and blood collection ranged from 0 to 70 months with intervening treatment, suggesting that tumor evolution and not only tumor heterogeneity could be the underlying reason for incongruence between tissue and liquid biopsy analysis. We analyzed the effect of time occurring between tissue biopsy and blood collection on concordance (shown in Supplementary Fig. 2C). We observed a trend of decreased time interval between tissue and liquid biopsy for concordant samples, however, without reaching statistical significance ($p = 0.4325$). Among mutations detected by plasma only, the EGFR T790M resistance mutation was the most frequent (32% of all mutations detected by plasma and not by tissue NGS analysis, Fig. 3d). This mutation was likely not detected in the initial tissue biopsy because it is known to emerge during therapy as a resistance mechanism against tyrosine-kinase inhibitor treatment of EGFR-mutated tumors. These data confirm the effectiveness of our testing strategy and highlight the clinical value of using liquid biopsy as a complementary tool to tissue biopsy for monitoring tumor evolution during treatment.



cfDNA profiling of healthy individuals

Finally, we attempted to profile the cfDNA of individuals that were healthy (as above defined) at the time of blood collection. Our patient cohort comprised $n = 106$ women that underwent a control screening mammography test and had been followed-up regularly for up to 10 years later (Table 1). Mammography screening and blood collection were performed concurrently. For this study, we divided the healthy individuals into four groups based on clinical status at follow-up (Table 1). Individuals belonging to group I ($n = 25$) did not develop any breast cancer or other malignancies during follow-up time. In group II individuals ($n = 52$) experienced fibrocystic breast changes such as fibroadenoma and hyperplasia during follow-up time, while in group III ($n = 15$) they developed breast cancer. Donors allocated to group IV ($n = 14$) developed a solid tumor other than breast cancer (specifically: laryngeal squamous cell carcinoma, glioblastoma multiforme, basal cell carcinoma, and thyroid

cancer). The average follow-up time was 8.5 years and did not differ significantly between the four groups described (Supplementary Fig. 3, Table 1). As reported in the first section of our results, we successfully achieved cfDNA extraction from all plasma samples, with values ranging from 1.7 to 30.8 ng/ml of cfDNA (Fig. 1a). Based on recovery rate and quality of cfDNA (described in Supplementary Fig. 1 and Material and methods), we selected 55 samples for downstream NGS analysis (group I = 12/25; group II = 23/52; group III = 11/15; group IV = 9/14; total = 55/106). We processed the selected samples using the HeliXgyn workflow and we selected six samples for a broader mutational analysis using the HeliXafe protocol (based on the OncoPrint™ cfDNA Pan-Cancer Assay). The turnaround time from start of plasma processing to data analysis was on average six working days for these 55 samples, confirming that we have established a fast workflow (Supplementary Fig. 2D). The results of the molecular profiling are summarized in Fig. 4. No genetic



alterations were found in the cfDNA of most healthy individuals (84%) (Fig. 4a). Among the four groups of healthy individuals with different outcomes at follow-up, no significant difference was observed in terms of pre-analytical variables, including cfDNA concentration in plasma or achieved molecular coverage (Fig. 4b, c). In 7 of the 55 cases analyzed, we detected clinically relevant gene mutations, specifically six known germline variants observed at allelic frequencies above 40% and four known cancer hotspot mutations (Fig. 4d, e). In conclusion, our results provide evidence that genetic alterations related to cancer occurrence can be detected in healthy individuals by analyzing cfDNA.

Discussion

Liquid biopsy has recently gained substantial attention in the field of cancer diagnostics. Ambitious efforts are currently placed towards the implementation of liquid biopsy as an early cancer detection method (i.e. before cancer-related symptoms occur) and ctDNA mutation analysis has already been reported in early-stage tumors^{26–28,40}. Early diagnosis possibly equals to a better disease outcome; however, large-scale validation studies are required to better understand the full potential and the limitations of this application of liquid biopsy³⁶. The screening of pre-cancerous lesions in asymptomatic individuals is hindered by several challenges. Namely, the

number of mutant ctDNA molecules present in plasma is mostly proportional to tumor burden⁹, rendering detection particularly problematic in patients with localized cancer and asymptomatic individuals. Another challenge is represented by the lack of knowledge regarding the molecular basis of tumor initiation. Several studies have reported the detection of somatic mutations and related clonal expansion in healthy tissue^{30–33} associated with age and tissue proliferative rate⁴¹. Some of these mutations were shown to increase the risk of developing cancer^{42,43}. The Pre-Cancer Genome Atlas⁴⁴ will significantly improve our understanding of the role of pre-cancerous lesions in early stages of tumor formation, improving the specificity of early detection screening. At present liquid biopsy is mainly used in advanced cancer patients; however, the Circulating Cell-free Genome Atlas (CCGA) study and the development of early screening methods such as CancerSEEK²⁷ are opening the way for cfDNA testing in healthy individuals and early-stage tumor patients. Our work aimed to contribute to this field by investigating the technical feasibility of using liquid biopsy for screening healthy individuals. Our cohort comprises 177 individuals, out of which 114 were clinically healthy and 63 were diagnosed with breast or lung cancer at time of blood collection. Because of the design of our study, which included patients undergoing routine mammography screening and followed-up for breast cancer insurgence, all healthy volunteers analyzed were women. As expected, cfDNA concentration was significantly lower in plasma from healthy individuals compared to cancer patients (Fig. 1a), with cfDNA concentrations ranging from 1 to 16.8 ng ml⁻¹ for healthy individuals (with the exception of one sample which had a concentration of 30.8 ng ml⁻¹ in plasma), consistently with previously published results^{34,35}. The patient presenting 30.8 ng ml⁻¹ in plasma belonged to group II. We did not detect any mutation for this sample nor found any sign of genomic contamination. Raised cfDNA concentrations have been observed in healthy donors under several physiological conditions (as physical exercise⁴⁵ or infection⁴⁶). To overcome the challenges associated with low input material as well as enabling the detection of low-frequency mutations, we have implemented molecular barcoding^{13,23,24,39} (reviewed in ref. ⁴⁷) in our sample processing workflow. We have confirmed the reliability and accuracy of our method by matched genomic analysis of tissue and plasma samples in cancer patients (concordance of 71% for our cohort of breast and lung cancer patients; Fig. 3a). These results are in line with previous studies reporting sensitivity between 65% and 98%^{48–51} (reviewed in ref. ⁷). We did not observe perfect concordance possibly due to tumor heterogeneity and evolution under treatment pressure (Supplementary Fig. 2C). We then used this method to screen for signs of genomic

instability in healthy donors. We could successfully isolate cfDNA and produce functional NGS libraries from as little as 0.9 ml of plasma; however, we recovered material of adequate quality to undergo NGS library preparation only for 55 out of 114 patients. Moreover, we observed higher LOD (Fig. 2g) in healthy donors compared to cancer patients due to higher cfDNA input in cancer patients. The availability of lower amounts of plasma for cfDNA isolation in healthy donors (0.4 and 2.0 ml in healthy individuals and 1.5 and 5.5 ml in cancer patients) is a drawback of this study. As healthy individuals present with lower levels of cfDNA compared to cancer patients (Fig. 1a), we recommend using higher volumes of plasma for cfDNA analysis from healthy donors. Importantly, this would allow for the detection of variants present at low allelic frequencies, which could be particularly relevant for discovering the presence of early genomic changes (as shown by the four cancer hotspot mutations we identified, Fig. 4d). Through our analysis, we detected genetic alterations in 7 out of 55 subjects with evaluable cfDNA that were considered clinically healthy at the time of liquid biopsy. Among these mutations, we found six germline variants and four cancer hotspot mutations. The observation of germline variants is a byproduct of our cfDNA analysis. Interestingly, many germline variants detected in our study are mutations in the coding region of TP53 that have been consistently reported to correlate with genomic instability and increased cancer risk^{52–55}. Those patients might be recommended to have genetic counseling and upon the decision of a trained certified geneticist to access early prevention programs. The four cancer hotspot mutations detected are recurrent genetic alterations, clinically classified as pathogenic or likely pathogenic. Previous studies have identified mutations in saliva and plasma of individuals up to 2 years before tumor insurgence^{56,57}. We detected cancer hotspot variants in individuals that were diagnosed with a benign breast condition (group II) or breast cancer (group III) up to 10 years later and at allelic frequencies ranging from 0.08% to 0.52% (Fig. 4d). Furthermore, the detected hotspot mutations have been associated with breast cancer as well as non-neoplastic proliferation of tissue by several studies^{58–69}. Observing these mutations in the cfDNA of healthy donors might be considered as indirect evidence of genomic instability, as was shown for the PIK3CA p. H1047R variant⁶². However, it was also observed that pathogenic TP53 mutations can be detected in the cfDNA of healthy controls⁷⁰ with no correlation to tumor insurgence. Therefore, the interpretation of these findings warrants caution and needs to be carefully considered before drawing any conclusion. Additional extensive prospective studies with long follow-up time and available tissue specimens for individuals who develop cancer will be required to address the specificity and sensitivity of

liquid biopsy as a tool for early cancer detection. In conclusion, with this work we have established a rapid and reliable workflow that allowed us to interrogate cfDNA from healthy individuals to study genomic alterations with a limit of detection as low as 0.08% allelic frequency. The interrogation of cfDNA from the blood of healthy individuals could prove to be a prospective tool to detect signs of genomic instability and to better understand early events in tumor formation.

Acknowledgements

This work was supported by the research grant 2015 of Lega Italiana per la Lotta contro i Tumori (LILT). M.S. was funded by the NIH award P30 CA008748 and the Breast cancer Research Foundation. This project was also supported by the Swiss National Science Foundation (SNF 138513). We thank Dr. Katharina Leonards for scientific input and fruitful discussions.

Author details

¹Institute of Pathology, University Hospital Basel, 4031 Basel, Switzerland. ²Breast Cancer Unit and Translational Research Unit, ASST Cremona, Viale Concordia 1, 26100 Cremona, Italy. ³Department of Medical Surgery and Health Sciences, University of Trieste, 34129 Trieste, Italy. ⁴Department of Life Sciences, University of Trieste, Via Giorgeri, 1, 34127 Trieste, Italy. ⁵Cancer Metastasis Laboratory, Department of Biomedicine, University of Basel, 4058 Basel, Switzerland. ⁶Human Oncology & Pathogenesis Program (HOPP), Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA. ⁷Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, 10065 New York, NY, USA. ⁸Bioscience Institute, Via Rovereta 42, Falciano, 47891 San Marino, Italy. ⁹Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy. ¹⁰RCCS Neuromed, Pozzilli, Italy. ¹¹Present address: Novartis Institutes for BioMedical Research, 4056 Basel, Switzerland. ¹²Present address: Thermo Fisher Scientific, 6300 Zug, Switzerland

Conflict of interest

P.J., L.Q., N.A. and M.S. were at the time of this project establishment part of the Scientific Advisory Board of the Bioscience Institute. I.A. and P.J. report grants from BMS and non-financial support from Thermo Fisher Scientific. G.M. is the Founder and CEO of the Bioscience Institute SpA. L.Q. no longer serves on the Bioscience Institute Scientific Board and currently is an Employer of Thermo Fisher Scientific. He has joined Thermo Fisher Scientific when this project was already accomplished and only manuscript had to be written. L.B. reports grants and personal fees from Roche, grants and personal fees from MSD, personal fees from BMS, personal fees from Astra Zeneca. N.A. is a paid consultant for pharmaceutical and insurance companies with an interest in liquid biopsy, and he is listed as inventor in several patent applications related to cancer detection and treatment. M.S. has received research funds from Puma Biotechnology, Daiichi-Sankio, Immunomedics, Targimmune and Menarini Ricerche, and is a cofounder of Medendi Medical Travel. The remaining authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41419-019-1770-3>).

Received: 27 March 2019 Revised: 14 June 2019 Accepted: 18 June 2019
Published online: 11 July 2019

References

1. Yates, L. R. et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21**, 751–759 (2015).
2. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
3. Zhang, J. et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014).
4. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
5. Nakad, R. & Schumacher, B. DNA damage response and immune defense: links and mechanisms. *Front. Genet.* **7**, 147 (2016).
6. Diehl, F. et al. Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
7. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
8. Jahr, S. et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).
9. Bettgeowda, C. et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24–224ra24 (2014).
10. Alix-Panabieres, C. & Pantel, K. Clinical applications of circulating tumor cells and circulating tumor dna as liquid biopsy. *Cancer Discov.* **6**, 479–491 (2016).
11. Kimura, H. et al. Detection of epidermal growth factor receptor mutations in serum as a predictor of the response to gefitinib in patients with non-small-cell lung cancer. *Clin. Cancer Res.* **12**, 3915–3921 (2006).
12. Forshew, T. et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4**, 136ra68–136ra68 (2012).
13. Newman, A. M. et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
14. Narayan, A. et al. Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Res.* **72**, 3492–3498 (2012).
15. Lanman, R. B. et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS ONE* **10**, e0140712 (2015).
16. Tokudome, N. & Hayes, D. F. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *Breast Dis. Year Book Quart.* **24**, 350–352 (2013).
17. Diaz, L. A. Jr et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
18. Siravegna, G. et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat. Med.* **21**, 827–827 (2015).
19. Fribbens, C. et al. Plasma ESR1 mutations and the treatment of estrogen receptor-positive advanced breast cancer. *J. Clin. Oncol.* **34**, 2961–2968 (2016).
20. Weigelt, B. et al. Diverse BRCA1 and BRCA2 reversion mutations in circulating cell-free DNA of therapy-resistant breast or ovarian cancer. *Clin. Cancer Res.* **23**, 6708–6720 (2017).
21. Azad, A. A. et al. Androgen receptor gene aberrations in circulating cell-free DNA: biomarkers of therapeutic resistance in castration-resistant prostate cancer. *Clin. Cancer Res.* **21**, 2315–2324 (2015).
22. Misale, S. et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).
23. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* **108**, 9530–9535 (2011).
24. Schmitt, M. W. et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. USA* **109**, 14508–14513 (2012).
25. Aravanis, A. M., Lee, M. & Klausner, R. D. Next-generation sequencing of circulating tumor DNA for early cancer detection. *Cell* **168**, 571–574 (2017).
26. Abbosh, C. et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
27. Cohen, J. D. et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc. Natl. Acad. Sci. USA* **114**, 10202–10207 (2017).
28. Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **1**, eaar3247 (2018).
29. De Mattos-Arruda, L. et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann. Oncol.* **25**, 1729–1735 (2014).
30. Aghili, L., Foo, J., DeGregori, J. & De, S. Patterns of somatically acquired amplifications and deletions in apparently normal tissues of ovarian cancer patients. *Cell Rep.* **7**, 1310–1319 (2014).
31. Beane, J. et al. Detecting the presence and progression of premalignant lung lesions via airway gene expression. *Clin. Cancer Res.* **23**, 5091–5100 (2017).

32. Krimmel, J. D. et al. Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic TP53 mutations in noncancerous tissues. *Proc. Natl Acad. Sci. USA* **113**, 6005–6010 (2016).
33. Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
34. Mouliere, F. et al. High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE* **6**, e23418 (2011).
35. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).
36. Cree, I. A. et al. The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer: a systematic mapping review. *BMC Cancer* **17**, 697 (2017).
37. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14**, 178–192 (2013).
38. Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res.* **37**, 646–650 (1977).
39. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).
40. Phallen, J. et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **9**, eaan2415 (2017).
41. Yizhak, K. et al. A comprehensive analysis of RNA sequences reveals macroscopic somatic clonal expansion across normal tissues. bioRxiv 416339, <https://doi.org/10.1101/416339> (2018).
42. Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
43. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
44. Campbell, J. D. et al. The case for a Pre-Cancer Genome Atlas (PCGA). *Cancer Prev. Res.* **9**, 119–124 (2016).
45. Breitbach, S., Sterzing, B., Magallanes, C., Tug, S. & Simon, P. Direct measurement of cell-free DNA from serially collected capillary plasma during incremental exercise. *J. Appl. Physiol.* **117**, 119–130 (2014).
46. De Vlamincq, I. et al. Noninvasive monitoring of infection and rejection after lung transplantation. *Proc. Natl Acad. Sci. USA* **112**, 13336–13341 (2015).
47. Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **19**, 269–285 (2018).
48. Oxnard, G. R. et al. Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *J. Clin. Oncol.* **34**, 3375–3382 (2016).
49. Schiavon, G. et al. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci. Transl. Med.* **7**, 313ra182–313ra182 (2015).
50. Sacher, A. G. et al. Prospective validation of rapid plasma genotyping for the detection of EGFR and KRAS mutations in advanced lung cancer. *JAMA Oncol.* **2**, 1014–1022 (2016).
51. Thierry, A. R. et al. Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nat. Med.* **20**, 430–435 (2014).
52. Martin, A.-M. et al. Germline TP53 mutations in breast cancer families with multiple primary cancers: is TP53 a modifier of BRCA1? *J. Med. Genet.* **40**, e34 (2003).
53. Mitchell, G. et al. High frequency of germline TP53 mutations in a prospective adult-onset sarcoma cohort. *PLoS ONE* **8**, e69026 (2013).
54. Nichols, K. E. & Malkin, D. Genotype versus phenotype: the Yin and Yang of germline TP53 mutations in Li-Fraumeni syndrome. *J. Clin. Oncol.* **33**, 2331–2333 (2015).
55. Deben, C. et al. Deep sequencing of the TP53 gene reveals a potential risk allele for non-small cell lung cancer and supports the negative prognostic value of TP53 variants. *Tumour Biol.* **39**, 1010428317694327 (2017).
56. Mao, L., Hruban, R. H., Boyle, J. O., Tockman, M. & Sidransky, D. Detection of oncogene mutations in sputum precedes diagnosis of lung cancer. *Cancer Res.* **54**, 1634–1637 (1994).
57. Gormally, E. et al. TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. *Cancer Res.* **66**, 6871–6876 (2006).
58. Millikan, R. et al. p53 mutations in benign breast tissue. *J. Clin. Oncol.* **13**, 2293–2300 (1995).
59. Wood, L. D. et al. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
60. Madsen, R. R., Vanhaesebroeck, B. & Semple, R. K. Cancer-associated PIK3CA mutations in overgrowth disorders. *Trends Mol. Med.* **24**, 856–870 (2018).
61. Kalinsky, K. et al. PIK3CA mutation associates with improved outcome in breast cancer. *Clin. Cancer Res.* **15**, 5049–5059 (2009).
62. Miller, T. W. Initiating breast cancer by PIK3CA mutation. *Breast Cancer Res.* **14**, 301 (2012).
63. Schneck, H. et al. Analysing the mutational status of PIK3CA in circulating tumor cells from metastatic breast cancer patients. *Mol. Oncol.* **7**, 976–986 (2013).
64. Chiang, S. et al. IDH2 mutations define a unique subtype of breast cancer with altered nuclear polarity. *Cancer Res.* **76**, 7118–7129 (2016).
65. Majoor, B. C. et al. Increased risk of breast cancer at a young age in women with fibrous dysplasia. *J. Bone Mineral Res.* **33**, 84–90 (2018).
66. Pantel, K. Blood-based analysis of circulating cell-free DNA and tumor cells for early cancer detection. *PLoS Med.* **13**, e1002205 (2016).
67. Jahn, S. W. et al. Mutation profiling of usual ductal hyperplasia of the breast reveals activating mutations predominantly at different levels of the PI3K/AKT/mTOR pathway. *Am. J. Pathol.* **186**, 15–23 (2016).
68. Yi, Z. et al. Landscape of somatic mutations in different subtypes of advanced breast cancer with circulating tumor DNA analysis. *Sci. Rep.* **7**, 5995 (2017).
69. Sobhani, N. et al. The prognostic value of PI3K mutational status in breast cancer: meta-analysis. *J. Cell. Biochem.* **119**, 4287–4292 (2018).
70. Fernandez-Cuesta, L. et al. Identification of circulating tumor DNA for the early detection of small-cell lung cancer. *EBioMedicine* **10**, 117–123 (2016).