

Circulating cell-free DNA mutation patterns in early and late stage colon and pancreatic cancer

Eveline E. Vietsch¹

Garrett T. Graham¹

Justine N. McCutcheon¹

Aamir Javaid¹

Giuseppe Giaccone¹

John L. Marshall^{1,2}

Anton Wellstein^{1,2}.

1 Department of Oncology, Georgetown University, Washington, DC, USA.

2 Otto J. Ruesch Center for the Cure of Gastrointestinal Cancer, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA.

Cancer Genet. 2017 Dec;218-219:39-50.

ABSTRACT

Cancer is a heterogeneous disease harboring diverse subclonal populations that can be discriminated by their DNA mutations. Environmental pressure selects subclones that ultimately drive disease progression and tumor relapse. Circulating cell-free DNA (ccfDNA) can be used to approximate the mutational makeup of cancer lesions and can serve as a marker for monitoring disease progression at the molecular level without the need for invasively acquired samples from primary or metastatic lesions. This potential for molecular analysis makes ccfDNA attractive for the study of clonal evolution and for uncovering emerging therapeutic resistance or sensitivity. We assessed ccfDNA from colon and pancreatic adenocarcinoma patients using next generation sequencing of 56 cancer-associated genes at the time of primary resectable disease and metastatic progression and compared this to the mutational patterns of the primary tumor. 28%-47% of non-synonymous mutations in the primary tumors were also detected in the ccfDNA while 71%-78% mutations found in ccfDNA were not detected in the primary tumors. ccfDNA collected at the time of progression harbored 3-5 new mutations not detected in ccfDNA at the earlier collection time points. We conclude that incorporation of ccfDNA analysis provides crucial insights into the changing molecular makeup of progressive colon and pancreatic cancer.

INTRODUCTION

Genomic mutations are one of the hallmarks of cancer [1]. The molecular characterization of a given cancer relies on the analysis of tissue specimen from a primary or metastatic lesion typically obtained at a single time point. However, due to intratumoral heterogeneity, the selection of cell subpopulations during cancer evolution and metastasis, the analysis of a single tissue specimen will provide only a limited characterization of the molecular makeup of the disease [2,3]. Monitoring the molecular characteristics of cancer by serial analyses of circulating cell-free DNA (ccfDNA) enables capture of emerging heterogeneity of the disease and may support treatment decisions [4,5]. CcfDNA analysis has evolved since its inception with improvements in the technologies and detection limits [6,7] and represents a set of research tools that appear poised to enter routine clinical care [8,9]. The recent FDA approval of a ccfDNA assay for the *EGFR* T790M mutation in lung cancer supports this notion [10]. Whether ccfDNA should complement tissue analyses in all cancer types remains to be studied, especially in early stage diseases [9,11]. However, ccfDNA may be superior to tumor tissue DNA in the assessment of cancer heterogeneity and evolution during disease progression [12,13]. Here we study the mutational landscape of ccfDNA at diagnosis and disease recurrence and compare it to that of DNA from the primary tumor tissues in ten patients with colon and pancreatic cancer.

MATERIALS AND METHODS

Patient Samples

Patients with newly diagnosed colon adenocarcinoma (colon AC) or pancreatic ductal adenocarcinoma (PDAC) were recruited for blood and tissue collection under the IRB protocol 2007-345 “Establishment of the High Quality Tumor Biobank and Clinical Database” and the Non-Therapeutic Subject Registry (NTSR) Shared Resource protocol Pr000000007 at the Lombardi Comprehensive Cancer Center at Georgetown University after obtaining informed consent. Ten patients were retrospectively selected with the following inclusion criteria: initial diagnosis of treatment-naïve resectable primary adenocarcinoma (n=5 colon; n=5 pancreatic), surgical resection of the primary tumor, and development of progressive disease after surgery. None of the patients had a previous malignant disease. Peripheral venous blood samples were collected in EDTA plasma tubes before the surgical removal of the primary tumors as well as at time of metastatic disease progression (1-70 months after surgery). The blood samples were centrifuged at ≤ 1300 RCF for 10 min within 2 hours of blood collection, after which plasma was

separated and stored at -80°C until further analysis. Surgical specimens of the primary tumors were frozen in O.C.T and cryo-sectioned into $20\text{ }\mu\text{m}$ scrolls and examined by a pathologist for the presence of cancer cells.

DNA isolation

The plasma samples were thawed on ice and circulating cell-free circulating DNA (ccfDNA) was isolated from $2 \times 100\text{ }\mu\text{L}$ plasma per patient, using the DNA extractor SP Kit (Wako cat. # 296-60501) following the manufacturer's protocol. In brief, $200\text{ }\mu\text{L}$ Enzyme Reaction Solution and $5\text{ }\mu\text{L}$ Protein Digestion Solution was added to $100\text{ }\mu\text{L}$ plasma and mixed by vortexing. The samples were incubated at 56°C for 10 min. Thereafter $300\text{ }\mu\text{L}$ of Sodium Iodide Solution and $600\text{ }\mu\text{L}$ Alcohol Solution were added and mixed by vortexing. After 10 min incubation at room temperature, the samples were centrifuged at $16,000 \times g$ for 10 min at room temperature. The supernatant was discarded and ccfDNA pellets were washed with 1 mL Washing Solution A by vortexing. After 5 min centrifugation at $16,000 \times g$ the supernatant was discarded. The ccfDNA pellets were washed with 1 mL Washing Solution B and centrifuged once more for 5 min. The supernatant was discarded again and DNA pellets were allowed to dry. The ccfDNA was diluted in $15\text{ }\mu\text{L}$ of ultra pure water and quantitated with the NanoDrop 2000c (Thermo Scientific) and the Promega Quantifluor ONE dsDNA Fluorescence Assay (Promega).

Two frozen tumor tissue scrolls of $20\text{ }\mu\text{m}$ thickness per patient with an average weight of $75\text{ }\mu\text{g}$ and surface of 1.15 cm^2 were used for genomic DNA isolation. DNA was isolated using the PrepEase Genomic DNA Isolation Kit (USB), following the manufacturer's protocol. In brief, the tissue was homogenized in $240\text{ }\mu\text{L}$ Homogenization Buffer in MagNA Lyser Green Beads (Roche) in the MagNA Lyser (Roche). A mixture of $200\text{ }\mu\text{L}$ Chloroform/Isoamyl Alcohol (24:1), as well as $800\text{ }\mu\text{L}$ Protein Precipitation Buffer were added to the lysates. Samples were mixed by vortexing and centrifuged at $13,000 \times g$ for 4 min at room temperature. $880\text{ }\mu\text{L}$ of the upper aqueous phase of the sample was transferred to a new microcentrifuge tube containing $620\text{ }\mu\text{L}$ isopropanol. The samples were mixed by inverting the tubes and centrifuged at $13,000 \times g$ for 4 min. The supernatant was discarded and DNA pellets were washed with 1 mL of 70% ethanol by vortexing. The samples were centrifuged for 2 min and DNA pellets were allowed to dry. The tumor DNA was diluted in $15\text{ }\mu\text{L}$ of ultra pure water and quantitated with the NanoDrop 2000c (Thermo Scientific) and the Promega Quantifluor ONE dsDNA Fluorescence Assay (Promega).

56G Oncology Panel Sequencing Library preparation

DNA mutation analysis was conducted using a Targeted Next Generation Sequencing Library Preparation Kit that is compatible with circulating cell-free DNA and the Illumina MiSeq Platform: the 56G Oncology Panel v2 from Swift Biosciences (Cat. # AL-56248). This

panel contains 263 amplicons sized 92-184 bp that covers hotspots, exonic SNPs and contiguous regions of 56 human genes. The list of genes and number of amplicons is provided in Table 1. The kit contains a DNA standard with a set of 11 defined allelic frequencies for major oncology targets to be used as a sequencing control and DNA from HCT116, RKO and SW48 colon cancer cell lines. The 56G Oncology library was prepared according to the manufacturer's protocol. In brief, 10 ng DNA per sample was used for the Multiplex PCR Step using the Reaction Mix, and the following Thermocycler Program: 30 sec at 98 °C, 4 cycles of 10 sec at 98 °C, 5 min at 63 °C, 1 min at 65 °C, followed by 21 cycles of 10 sec at 98 °C, 1 min at 64 °C, followed by 1 min at 65 °C and hold at 4 °C. The resulting amplicons were purified using SPRIselect beads (Beckman Coulter, Cat. #B23318) and a DynaMag magnetic rack (Invitrogen). Next, a unique combination of Index D50X + Index D7XX was added to each sample bead pellet, together with the Indexing Reaction Mix (Swift Biosciences). The samples were incubated at 37 °C for 20 min with the lid heating turned off. The libraries were purified once more with SPRIselect beads and quantitated in triplicates by qRT-PCR in a 20 µL reaction using the iQ SYBR Green Supermix (BioRad), containing 10 µL of SYBR Green mix, 10 µL of diluted library (1:1000), and 500 nM of the following primers: 5' AATGATACGGCGACCACCGAGAT 3'; and 5' CAAGCAGAAGACGGCATACGA 3'. Serial dilutions of the PhiX Sequencing Control v3, (Illumina Cat. # FC- 110-3001) was used as standard. After quantitation, the libraries were normalized to a concentration of 2 nM and pooled together.

MiSeq Loading

The library pool was sequenced using the MiSeq v2 300 cycle Reagent Kit (Illumina). Five µL of the pooled amplicon library was denatured with 5 µL 0.2N NaOH for 5 min at room temperature. The library pool was then diluted to 8 pM with chilled HT1 buffer and 10% PhiX v3 control (Illumina) was spiked into the diluted library pool. Six hundred µL of the diluted pooled library with PhiX spike-in was loaded into the MiSeq reagent cartridge.

Sequencing data analysis

Adapter trimming was conducted per Swift Biosciences recommendation, using *cutadapt* [14]. Paired-end FASTQ samples were aligned to GRCh38 with BWA-MEM. Sorting and indexing was done using *samtools*. Base quality score recalibration followed by local realignment was done using the GATK Java package in conjunction with dbSNP annotation (b149) [15]. Variant calling was conducted using the LoFreqV2 (LoFreq*) mutation caller [16]. Visualizations were created in R using custom scripts (available on request) employing Bioconductor package VariantAnnotation [17], as well as the plotting framework *ggplot2* and further adapted in Excel 2010 and Illustrator CS3 (Adobe). After mutations were called, percent representation was established as (number of variant reads)/(number of reference reads + number of variant reads). Variants leading to amino acid changes, with a minimum

read count of 5 reads and a frequency above 1% in at least one of the DNA samples are shown. The variant frequencies of the expected mutations that were detected in the DNA standard are shown in the Supplementary Figure 1. Three variants detected in amplicons from patient samples and the DNA standard were discarded as false positives (*FGFR1* D166del; *MSH6* F1088frameshift and *TP53* P72R).

Table 1. Genes included in the analysis. The amplicon panel is from Swift Biosciences ‘Accel-Amplicon 56G Oncology Panel v2’. The number of amplicons (#) for each gene is shown.

Gene name	# amplicons	Gene name	# amplicons
<i>ABL1</i>	5	<i>IDH2</i>	2
<i>AKT1</i>	2	<i>JAK2</i>	2
<i>ALK</i>	2	<i>JAK3</i>	3
<i>APC</i>	9	<i>KDR</i>	9
<i>ATM</i>	19	<i>KIT</i>	14
<i>BRAF</i>	2	<i>KRAS</i>	3
<i>CDH1</i>	3	<i>MAP2K1</i>	5
<i>CDKN2A</i>	2	<i>MET</i>	6
<i>CSF1R</i>	2	<i>MLH1</i>	1
<i>CTNNB1</i>	1	<i>MPL</i>	1
<i>DDR2</i>	1	<i>MSH6</i>	4
<i>DNMT3A</i>	1	<i>NOTCH1</i>	3
<i>EGFR (HER1)</i>	9	<i>NPM1</i>	1
<i>ERBB2 (HER2)</i>	4	<i>NRAS</i>	3
<i>ERBB4 (HER4)</i>	8	<i>PDGFRA</i>	4
<i>EZH2</i>	1	<i>PIK3CA</i>	11
<i>FBXW7</i>	6	<i>PTEN</i>	14
<i>FGFR1</i>	2	<i>PTPN11</i>	2
<i>FGFR2</i>	4	<i>RB1</i>	12
<i>FGFR3</i>	6	<i>RET</i>	6
<i>FLT3</i>	4	<i>STK11</i>	5
<i>FOXL2</i>	1	<i>SMAD4</i>	10
<i>GNA11</i>	2	<i>SMARCB1</i>	4
<i>GNAQ</i>	2	<i>SMO</i>	5
<i>GNAS</i>	2	<i>SRC</i>	1
<i>HNFI1A</i>	4	<i>TP53 (P53)</i>	21
<i>HRAS</i>	2	<i>TSC1</i>	1
<i>IDH1</i>	1	<i>VHL</i>	3

RESULTS

Patient characteristics and analysis approach

Patients with treatment-naïve adenocarcinomas of the colon (n=5) or the pancreas (n=5) that were considered resectable at the time of initial diagnosis were included in this study (Table 2). Plasma samples were collected before surgical removal of the primary tumors and at the time of disease progression. Resected tumor tissues were cryo-sectioned and evaluated by a pathologist to assess cancer cell and stroma abundance. After surgery, patients received different adjuvant therapies that are listed in Table 2. To assess the mutation patterns in early and late stage colon and pancreatic cancer we compared plasma DNA at the time of initial diagnosis to that of the primary tumor, as well as to the plasma DNA at the time of metastatic disease (Figure 1). For DNA mutation analysis, we sequenced 30 DNA samples from 10 patients in a single, next generation deep-sequencing run of 263 mutation hotspots in a set of 56 cancer-related genes. The same platform was used for the analysis of tissue and circulating cell-free DNA (ccfDNA) to avoid discrepancies in mutation findings due to variability in sequencing methods or reagents.

DNA mutations detected

Next generation deep-sequencing of amplicons showed a median depth of 754,000 reads per sample. In all pre-surgery plasma samples genomic alterations with a 1% frequency were detected, despite a relatively low read coverage in one of the ccfDNA samples (patient 9 pre-surgery). One out of 10 late stage plasma samples (ccfDNA patient 4 at metastasis) did not contain genomic alterations above 1% frequency due to hemolysis and wildtype cellular DNA contamination that diluted the ccfDNA mutations. We focused on non-synonymous DNA alterations with a variant frequency of at least 1% in one of the samples per patient. In the 56 genes assessed we found an average of 10 mutations (range 4-15) in 17 genes in the ccfDNA of five colon cancer patients before surgery (Figure 2). The ccfDNA of the five patients with resectable PDAC contained fewer mutations, i.e. an average of 8 (range 5-12) mutations in 14 of the 56 genes assessed.

Surprisingly, the number of mutations detected in the primary tumors was much lower than that in the plasma: In the colon cancer tissues an average of 3 mutations (range 1-4) were found in 7 genes (*APC*, *BRAF*, *CDKN2A*, *KIT*, *KRAS*, *PTEN*, *TP53*; Figure 3). In the pancreatic cancer tissues 4 mutations (range 3-6) were found in 7 genes (*CDKN2A*, *JAK3*, *KDR*, *KIT*, *KRAS*, *SMAD4*, *TP53*; Figure 3). Although our analysis detected mutations in only 7 genes, it is striking that each patient still had a unique combination of genomic alterations.

At the time of disease progression, the ccfDNA of five patients with metastasized colon cancer contains an average of 7 mutations (range 2-11) in 16 genes. The plasma of the patients

with metastatic PDAC contained an average of 9 mutations (range 3-12) in 20 genes (Figure 4). This number is close to the number of ccfDNA mutations before surgery. However, some of the ccfDNA mutations detected at the time of primary disease were lost at the time of metastatic disease, i.e. *ABL1*, *ATM*, *DNMT3A*, *FLT3*, *HNF1A*, *NRAS* and *SMAD4*. Possible explanations are that these mutations occurred in cancer cell subpopulations in the primary cancers that were resected, or that the clones carrying the mutations were selected against during disease progression.

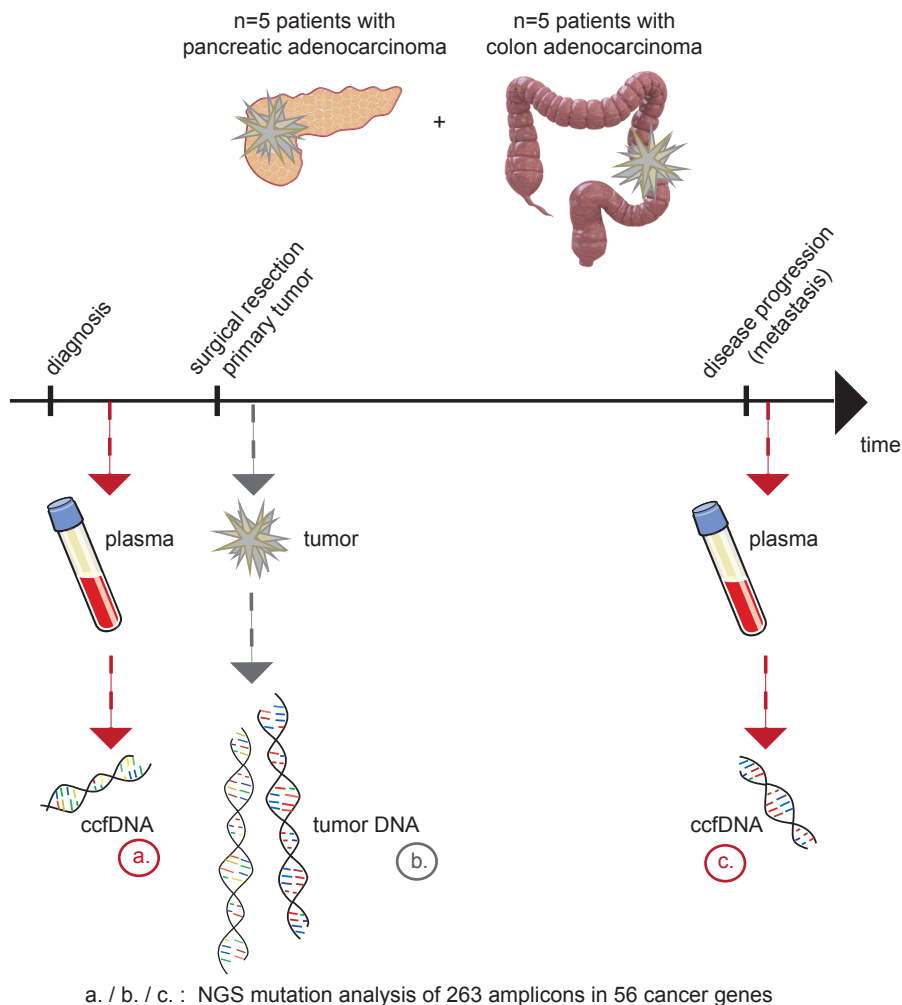


Figure 1. Overview of the study. Ten patients were included in the analysis of circulating cell-free DNA (ccfDNA) at diagnosis of primary cancer and at the time of progressive disease. DNA from frozen primary tumors that were collected during surgery was for comparison. All 30 DNA samples were subjected to next generation sequencing (NGS) of 56 genes with cancer-associated mutations.

Table 2. Patient characteristics.

Disease	Colon adenocarcinoma					Pancreatic adenocarcinoma				
Patient #	1 (1A0383)	2 (1A0390)	3 (1A0690)	4 (1A0816)	5 (1A1118)	6 (1A0431)	7 (1A0522)	8 (1A0551)	9 (1A0843)	10 (1A1128)
Pathological stage	3c T3N2M0	4a T3N1M1	3b T3N2(4/17) M0	2 pT4pN0M0	4a TxNxM1a	2b pT3N1M0	2b pT3pN2	2b pT3N1M0	2a pT3N0M0	2b T3N1M0
Age [years]	67	62	50	66	51	54	69	66	45	57
Gender	Male	Female	Female	Male	Male	Male	Female	Male	Female	Female
Surgery	Hemi-colectomy	Colon tumor resection	Colon resection	Right colectomy	Hemi-colectomy	Total pancreatectomy, splenectomy	Whipple	Whipple	Distal pancreatectomy, splenectomy	Distal pancreatectomy, splenectomy
Histology tumor scroll	50% tumor, 50% fibrosis	70% tumor, 30% fibrosis	90% tumor, 10% fibrosis	70% tumor, 30% necrosis	90% tumor, 10% fibrosis	30% tumor, 70% fibrosis	40% tumor, 60% fibrosis	50% tumor, 50% fibrosis	50% tumor, 50% fibrosis	40% tumor, 60% fibrosis
Time to metastasis after surgery [months]	70	20	12	5	1	16	9	18	15	7
Metastatic site	Mesentery, Anastomosis	Liver, Lung	Peritoneum, Liver	Liver	Peritoneum, Liver	Peritoneum, Liver	Liver	Lung	Ovary	Liver

Plasma ccfDNA before surgery		Colon carcinoma					Pancreatic carcinoma					
		Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	
APC	E1560V	5.74					5.54					2
APC	K1308*	1.84										1
APC	P1458H			12.06			15.83	6.35		28.13		4
BRAF	S465P							1.14				1
BRAF	S467P			1.20								1
BRAF	V600E	1.91	6.23	3.39	8.28		2.55					5
CDKN2A	G136Afs*10	1.27									0.70	2
EGFR	V96A	7.38										1
ERBB2	C315R			1.50								1
ERBB4	C326F	5.37	3.96	7.26			12.73		8.28	22.83		6
ERBB4	M953K	3.22										1
EZH2	X649_splice				1.28							1
GNA11	R166H						4.64					1
KDR	C482R										52.25	1
KDR	Q472H							61.00			52.32	2
KDR	V297I							42.00				1
KIT	M541L					46.66					53.84	2
KIT	Y568C				1.52							1
MAP2K1	R47*	6.39		9.62			20.18	12.56	7.61			5
MET	Q155H	7.38	3.25	10.13	1.41		16.21	7.85	5.97	93.84		8
NOTCH1	V1578del	1.69			0.78	0.61						3
PIK3CA	L94P				1.64						1.18	2
PTEN	L112P				1.49							1
PTEN	N323Mfs*21				5.81							1
PTEN	X85_splice	23.76	3.16	41.14	1.98		14.07	12.20	7.84	6.88		8
SMARCB1	P383Rfs*100					0.21	1.90					2
SMO	M525V	1.52										1
TP53	D281G				2.04							1
TP53	I251M			9.63			14.24	7.63		20.29		4
TP53	N268Tfs*77			2.93								1
TP53	P152Rfs*18				4.76							1
TP53	Q144del	1.10										1
TP53	R175G									100.00		1
TP53	R273C				5.87							1
TP53	S9N			2.29								1
TP53	X331_splice			1.97	1.43							2
TP53	Y163C					3.24						1
TP53	Y234H				1.29							1
VHL	A149P	8.72	7.02	9.45			18.97	6.86	7.59	100.00		7
VHL	R107L	3.80					7.74			82.45		3
number of mutations		15	5	13	14	4	12	8	6	8	5	

* = low coverage

variant frequency

0 100

Figure 2. Circulating cell-free DNA mutation frequency before surgery. Non-synonymous mutations at $\geq 1\%$ variant frequency in at least one of the ccfDNA samples. * = premature stop codon; fs=frameshift; del=deletion; - = low coverage.

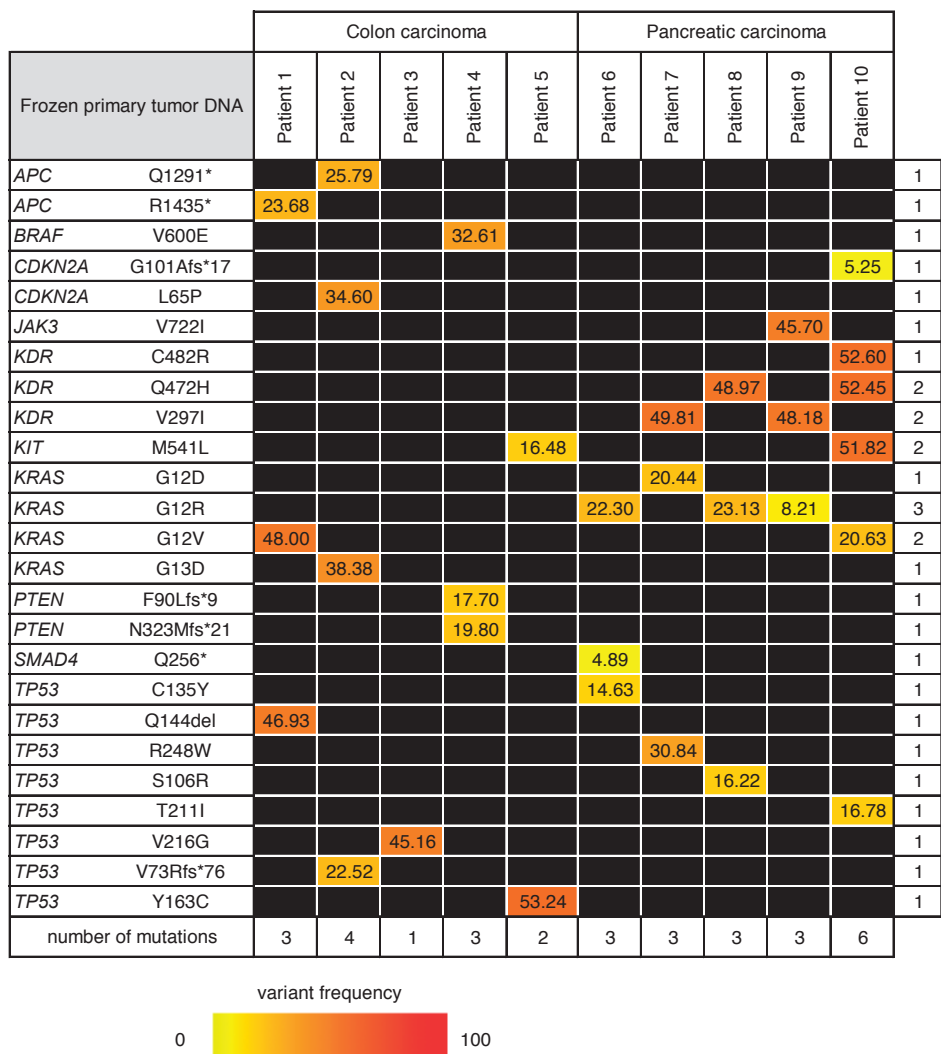


Figure 3. DNA mutation frequency in primary tumors. Non-synonymous mutations at $\geq 1\%$ variant frequency in at least one of the primary tumors. * = premature stop codon; fs=frameshift; del=deletion.

Comparison of mutations in primary tumors and the circulation at early and late stage disease

We evaluated the concordance between DNA mutations in the primary tumors and the circulation. For each patient, the mutations above a 1% variant frequency threshold in at least one of the DNA samples are shown in Fig. 5. First we compared the similarity between primary tumor DNA and ccDNA before surgery (Table 3). In the patients with colon cancer about half (47%) of tumor tissue mutations are also detected in the circulation though there is a wide variation between patients. In patients with primary PDAC this concordance

is below one third (28%). On the other hand, it is notable that cancer heterogeneity within each patient was much more evident in the mutational landscape of the ccfDNA than in tumor tissue DNA. The majority of mutations in the ccfDNA were not detected in the primary tumor tissues: 71% in colon cancer and 78% of the mutations in PDAC (Table 3). This finding illustrates that a tissue section of a given tumor can fail to represent the molecular makeup of the entire cancer.

Table 3. Concordance and discordance between the number of mutations detected in plasma and primary tumor DNA at the time of initial diagnosis.

Primary tumor versus plasma at the time of diagnosis				
Patient	Disease	Tumor tissue DNA mutations detected in plasma samples	Plasma DNA mutations <u>not</u> detected in tumor tissues	Cancer cells in tumor
1	colon AC	60% (3/5)	79% (11/14)	50%
2	colon AC	0% (0/5)	100% (5/5)	70%
3	colon AC	0% (0/1)	100% (13/13)	90%
4	colon AC	75% (3/4)	77% (10/13)	70%
5	colon AC	100% (3/3)	0% (0/3)	90%
average		47%	71.2%	74%
6	PDAC	25% (1/4)	92% (11/12)	30%
7	PDAC	33% (1/3)	89% (8/9)	40%
8	PDAC	33% (1/3)	86% (6/7)	50%
9 ⁻	PDAC	0% (0/3) ⁻	100% (8/8)	50%
10	PDAC	50% (3/6)	25% (1/4)	40%
average		28.2%	78.4%	42%

AC= adenocarcinoma; PDAC= pancreatic ductal adenocarcinoma; ⁻ = low sequencing coverage in plasma DNA

For the current study we had selected a set of patients that developed metastatic disease after removal of the primary tumors. We evaluated the genomic evolution during this progression by comparing the ccfDNA at the time of surgery and metastasis. Table 4 summarizes the differences in the ccfDNA mutational landscape at the time of primary and metastatic cancer. In five patients with metastatic colon cancer, 34% of the ccfDNA mutations were not detected in the circulation at the time of primary disease. These emerging mutations indicate clonal evolution of the disease during malignant progression. In PDAC, the proportion of these metastasis-associated ccfDNA mutations is higher; 63% of the mutations were not detected during primary disease. Interestingly, the fraction of new mutations at the time of metastatic disease is not correlated with the length of time to progression, (Table 4) or the type of adjuvant therapy (Table 2). Although this is a small number of patients, it is noteworthy that approximately half of ccfDNA mutations detected at the time of primary disease were not detected after progression to metastatic disease (45% in colon AC, 59% in

		Colon carcinoma					Pancreatic carcinoma					
Plasma ccfDNA at time of metastasis		Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	
ABL1	F401Lfs*3		1.54					0.29				2
APC	E1309Dfs*4								15.76			1
APC	P1458H	3.17					2.55					2
ATM	M349K	7.58										1
BRAF	T458Qfs*25						1.21					1
BRAF	V600E	2.56										1
CDKN2A	E119K										1.10	1
CDKN2A	E69G							4.42				1
CDKN2A	G136Afs*10		1.67		0.63		0.94	0.64	0.54			5
CDKN2A	W110*										1.12	1
CTNNB1	L46R		1.85									1
DNMT3A	R882H										1.14	1
ERBB4	C326F	3.94	10.37							4.06		3
FLT3	A813P									3.82		1
FLT3	V825M			2.14								1
HNF1A	E230K			1.30								1
JAK3	V722I									58.39		1
KDR	C482R										52.32	1
KDR	Q472H										52.71	1
KDR	V297I						48.95			47.06		2
KIT	M535I		2.77									1
KIT	M541L					30.33			43.09		52.32	3
KIT	M724Rfs*19					1.16						1
KRAS	G12V										16.60	1
MAP2K1	R47*	6.48	9.09				7.61			12.45		4
MET	N375S	5.66										1
MET	Q155H	5.85	9.95	2.58			5.87			12.73		5
NOTCH1	V1578del				0.58	1.12		0.43				3
NRAS	G12D										1.05	1
NRAS	M72I										1.43	1
PIK3CA	I391M	3.67										1
PTEN	E73K						1.13					1
PTEN	R84K						1.04					1
PTEN	X85_splice	19.10	24.68	6.32			6.21		1.12	15.09		6
PTEN	Y240Tfs*16									1.33		1
SMAD4	P225A									4.57		1
TP53	G293E									1.41		1
TP53	I251M		8.59									1
TP53	M340Cfs*5								18.43			1
TP53	S362Afs*8		1.27					0.12	0.35			3
TP53	T211I										2.48	1
TP53	Y163C					24.07						1
VHL	A149P	2.43	11.16	3.25			5.21			5.13		5
VHL	R107L	3.26								5.25		2
number of mutations		11	11	5	2 ^	4	9	6	6	12	10	

^ erythrocyte hemolysis: wildtype DNA contamination

 0 variant frequency 100
 

Figure 4. Circulating cell-free DNA mutation frequency at the time of metastasis. Non-synonymous mutations at $\geq 1\%$ variant frequency in at least one of the ccfDNA samples. * = premature stop codon; fs=frameshift; del=deletion; ^ = hemolysis, increase of wildtype DNA.

Table 4. Concordance between the number of mutations detected in plasma at the time of initial diagnosis and detection of metastasis.

ccfDNA mutations at the time of primary versus metastatic disease				
Patient	Disease	Gain of mutations after metastasis	Loss of mutations after metastasis	Time to metastasis [months]
1	colon AC	40% (4/10)	38% (5/14)	70
2	colon AC	64% (7/11)	20% (1/5)	20
3	colon AC	40% (2/5)	77% (10/13)	12
4 ^	colon AC	0% (0/1)	92% (12/13)	5
5	colon AC	25% (1/4)	0% (0/3)	1
average		33.8%	45.4%	21.6
6	PDAC	38% (3/8)	58% (7/12)	16
7	PDAC	67% (2/3)	89% (8/9)	9
8	PDAC	80% (4/5)	86% (6/7)	18
9	PDAC	58% (7/12)	38% (3/8)	15
10	PDAC	70% (7/10)	25% (1/4)	7
average		62.6%	59.2%	13

AC, adenocarcinoma; PDAC, pancreatic ductal adenocarcinoma; ^, hemolysis and wildtype cellular DNA contamination

PDAC, Table 4). This would indicate that cancer cell subpopulations carrying these mutations were likely dominant in the primary tumor that was removed surgically. Also, these subpopulations did not play a significant role in the metastatic lesions.

In conclusion, we found that ccfDNA analysis complements the molecular insight into the genetic make-up of colon and pancreatic cancer and can be particularly helpful in monitoring molecular changes over time.

DISCUSSION

The majority of molecular profiling of human tumors has relied on the analysis of aliquots of cancer tissues obtained from surgical resection specimen. Biopsies of cancerous lesions at disease progression are used rarely due to obvious disadvantages: They are potentially risky invasive procedures, time consuming and expensive. Also, cancer cells in tissue biopsies may be sparse due to the limited size of tissue recovered and against the background of wildtype stromal cells. Most importantly, subpopulations of a heterogeneous tumor may be poorly represented in biopsies. In the current study we sought to assess the changes in the mutational makeup of colon and pancreatic adenocarcinoma between the time of primary tumor surgery and detection of metastatic disease using ccfDNA. In this study we provide a direct comparison of mutation detection in primary tumor DNA and plasma ccfDNA

Patient 1 Colon Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastatic ccfDNA	
APC	E1560V	5.74			1
APC	K1308*	1.84			1
APC	P1458H			3.17	1
APC	R1435*		23.68		1
ATM	M349K			7.58	1
BRAF	V600E	1.91		2.56	2
CDKN2A	G136Afs*10	1.27	0.88		2
EGFR	V96A	7.38			1
ERBB4	C326F	5.37		3.94	2
ERBB4	M953K	3.22			1
KRAS	G12V		48.00		1
MAP2K1	R47*	6.39		6.48	2
MET	N375S			5.66	1
MET	Q155H	7.38		5.85	2
NOTCH1	V1578del	1.69	0.58		1
PIK3CA	I391M			3.67	1
SMO	M525V	1.52			1
TP53	Q144del	1.10	46.93		2
VHL	A149P	8.72		2.43	2
VHL	R107L	3.80		3.26	2
number of mutations		14	5	10	

Patient 2 Colon Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastatic ccfDNA	
ABL1	F401Lfs*3			1.54	1
APC	Q1291*		25.79		1
BRAF	V600E	6.23			2
CDKN2A	G136Afs*10		0.73	1.87	2
CDKN2A	L65P		34.60		1
CTNNB1	L46R			1.85	1
ERBB4	C326F	3.96		10.37	2
KIT	M535I			2.77	1
KRAS	G13D		38.38		1
MAP2K1	R47*			9.09	1
MET	Q155H	3.25		9.95	2
PTEN	X85_splice	3.16		24.68	2
TP53	I251M			8.59	1
TP53	S362Afs*8			1.27	1
TP53	V73Rfs*76		22.52		1
VHL	A149P	7.02		11.16	2
number of mutations		5	5	11	

Patient 6 Pancreatic Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastatic ccfDNA	
APC	E1560V	5.54			1
APC	P1458H	15.83		2.55	2
BRAF	T458Qfs*25			1.21	1
BRAF	V600E	2.55			1
ERBB4	C326F	12.73			1
GNA11	R166H	4.64			1
KRAS	G12R		22.30		1
MAP2K1	R47*	20.18		7.61	2
MET	Q155H	16.21		5.87	2
PTEN	E73K			1.13	1
PTEN	R84K			1.04	1
PTEN	X85_splice	14.07		6.21	2
SMAD4	Q256*		4.89		1
SMARCB1	P383Rfs*100	1.90		0.18	2
TP53	C135Y		14.63		1
TP53	I251M	14.24			1
VHL	A149P	18.97		5.21	2
VHL	p.R107L	7.74			1
number of mutations		12	4	8	

Patient 7 Pancreatic Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastatic ccfDNA	
APC	P1458H	6.35			1
BRAF	K473*	1.49			1
BRAF	S465P	1.14			1
CDKN2A	E69G			4.42	1
KDR	V297I	42.00	49.81	48.95	3
KRAS	G12D		20.44	0.47	2
MAP2K1	R47*	12.56			1
MET	Q155H	7.85			1
PTEN	X85_splice	12.20			1
TP53	I251M	7.63			1
TP53	R248W		30.84		1
VHL	A149P	6.86			1
number of mutations		9	3	3	

Figure 5. Mutation frequency in ccfDNA and tumors. The results from individual patients are shown (n=10). Non-synonymous mutations with $\geq 1\%$ variant frequency in at least one of the ccfDNA samples.

* = premature stop codon; fs=frameshift; del=deletion; ^ = hemolysis; - = low coverage.

Patient 3 Colon Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastasis ccfDNA	
APC	P1458H	12.06			1
BRAF	S467P	1.20			1
BRAF	V600E	3.39			1
ERBB2	C315R	1.50			1
ERBB4	C326F	7.26			1
FLT3	V825M		2.14		1
HNF1A	E230K		1.30		1
MAP2K1	R47*	9.62			2
MET	Q155H	10.13	2.58		1
PTEN	X85_splice	41.14	6.32		2
TP53	I251M	9.63			1
TP53	N268Tfs*77	2.93			1
TP53	S8N	2.29			1
TP53	V216G		45.16		1
TP53	X331_splice	1.97			1
VHL	A149P	9.45		3.25	2
number of mutations		13	1	5	

Patient 4 Colon Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastasis ccfDNA	
BRAF	V600E	8.28	32.61		2
EZH2	X649_splice	1.28			1
KIT	Y568C	1.52			1
MET	Q155H	1.41			1
PIK3CA	L94P	1.64	0.73	0.67	3
PTEN	F90Lfs*9		17.70		1
PTEN	L112P	1.49			1
PTEN	N323Mfs*21	5.81	19.80		2
PTEN	X85_splice	1.98			1
TP53	D281G	2.04			1
TP53	P152Rfs*18	4.76			1
TP53	R273C	5.87			1
TP53	X331_splice	1.43			1
TP53	Y234H	1.29			1
number of mutations		13	4	1 ^	

Patient 5 Colon Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastasis ccfDNA	
KIT	M541L	46.66	16.48	30.33	3
KIT	M724Rfs*19			1.163	1
NOTCH1	V1578del	0.611	0.276	1.119	3
TP53	Y163C	3.236	53.24	24.07	3
number of mutations		3	3	4	

Patient 8 Pancreatic Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastasis ccfDNA	
APC	E1309Dfs*4			15.76	1
ERBB4	C326F	8.28			1
KDR	Q472H	61.00	48.97		2
KIT	M541L			43.09	1
KRAS	G12R		23.13		1
MAP2K1	R47*	7.61			1
MET	Q155H	5.97			1
PIK3CA	E545A			1.02	1
PTEN	X85_splice	7.84		1.12	2
TP53	L206Wfs*41	1.13			1
TP53	M340Cfs*5			19.43	1
TP53	S106R		16.22		1
VHL	A149P	7.59			1
number of mutations		7	3	5	

Patient 9 Pancreatic Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastasis ccfDNA	
APC	P1458H	28.13			1
ERBB4	C326F	22.83		4.06	2
FLT3	A813P			3.82	1
JAK3	V722I		45.70	58.39	2
KDR	V297I		48.18	47.06	2
KRAS	G12R		8.21		1
MAP2K1	R47*			12.45	1
MET	Q155H	93.84		12.73	2
PTEN	X85_splice	40.75		12.54	2
PTEN	Y240Tfs*16			1.33	1
SMAD4	P225A			4.57	1
TP53	G293E			1.41	1
TP53	I251M	20.29			1
TP53	R175G	100.00			1
VHL	A149P	190.00		5.13	2
VHL	R107L	82.45		5.25	2
number of mutations		8	3	12	

Patient 10 Pancreatic Carcinoma		pre surgery ccfDNA	primary tumor DNA	metastasis ccfDNA	
CDKN2A	E119K			1.10	1
CDKN2A	G101Afs*17		5.25		1
CDKN2A	W110*			1.12	1
DNM3A	R882H			1.14	1
KDR	C482R	52.25	52.60	52.32	3
KDR	Q472H	52.32	52.45	52.71	3
KIT	M541L	53.84	51.82	52.32	3
KRAS	G12V		20.63	16.60	2
NRAS	G12D			1.05	1
NRAS	M72I			1.43	1
PIK3CA	L94P	1.18			1
TP53	T211I		16.78	2.48	2
number of mutations		4	6	10	

* = low coverage

Figure 5. Mutation frequency in ccfDNA and tumors. The results from individual patients are shown (n=10). Non-synonymous mutations with $\geq 1\%$ variant frequency in at least one of the ccfDNA samples. * = premature stop codon; fs=frameshift; del=deletion; ^ = hemolysis; ~ = low coverage.

at diagnosis and at disease recurrence. For this, amplicons covering 263 mutations in 56 cancer-associated genes were analyzed by deep-sequencing.

It is thought that ccfDNA can provide a better representation of the molecular makeup of a malignant disease than a single section from a surgical tumor specimen or a tissue biopsy. Also, blood samples can be drawn at deliberate intervals because they only require a minimally invasive procedure. In the present study the notion of a broader molecular representation is supported by the fact that the ccfDNA revealed approximately twice as many mutations as tumor tissue DNA. Very likely the tissue sections analyzed missed portions of the primary tumor that carried subpopulations with these additional mutations. It is also conceivable that patients had already developed occult metastatic disease at the time of the initial diagnosis and the additional ccfDNA mutations found represented the cancer cell subpopulations in the metastatic lesions. It appears that the ccfDNA provides a more complex picture of the disease.

Most studies focus on the presence of ccfDNA mutations in one or two genes to compare their presence to clinical outcome [18,19]. In our study, we sought to assess the mutation patterns in a broad set of genes to highlight tumor heterogeneity and demonstrate clonal evolution over the course of disease progression. In colon cancer, for example, ccfDNA has been used to track clonal evolution during treatment with the epidermal growth factor receptor (EGFR)-specific antibodies. Alterations in *KRAS*, *NRAS*, *MET*, *ERBB2*, *FLT3*, *EGFR* and *MAP2K1* were detected in ccfDNA of patients with primary or acquired resistance to EGFR blockade [20]. Using a broad panel of cancer-associated genes rather than frequently altered candidate oncogenes such as *KRAS* overcame one of the potential pitfalls encountered. None of the *KRAS* mutations from the seven patients with *KRAS* mutant primary tumors were detected in the ccfDNA at the time of diagnosis, while mutations in other genes were detected. Mutant *KRAS* ccfDNA was, however, found by others in 10 of 34 pancreatic patients (29%) [21] or in 136 of 188 (72.3%) of patients with metastatic PDAC [22]. Another study showed that mutant *KRAS* ccfDNA was detected in 14.8%, 45.5%, 30.8%, and 57.9% of age-matched controls, localized, locally advanced, and metastatic PDAC patients, respectively [19]. In circulating exosomal DNA the percentages of mutant *KRAS* in these groups were even higher, i.e. 7.4%, 66.7%, 80%, and 85% respectively [19].

At the time of surgical removal of the primary tumors between 28% and 47% of tumor mutations were also detected in ccfDNA. Others have found that in formalin-fixed paraffin-embedded cancer tissues and plasma from patients with different types of cancer an overall concordance of 60% in mutations of 19 genes analyzed [23]. Thus, the concordance found here is relatively low. In contrast, over 70% of mutations detected in ccfDNA were not found in the tumor samples. This could be due to metastatic disease at the time of diagnosis or

poor representation of the primary tumor composition by the histological section obtained for the DNA analysis.

Shed DNA in the circulation is fragmented into relatively short size fragments with tumor-derived DNA exhibiting even higher fragmentation than normal cellular DNA [24]. This was also found for specific examples such as the *BRAF* V600E mutant allele with a fragment size <145 base pairs. Indeed detection of *EGFR* T790M mutant DNA in the circulation of cancer patients was improved by selecting for shorter DNA fragment lengths [25]. Comparisons of tumor tissue somatic DNA and ccfDNA mutation rate can be impacted by the application of different sequencing technologies and amplicon sizes [26,27] as well as read depths and these technical issues may bias the data interpretation [21]. To avoid this pitfall we used a platform that is adapted to the detection of short DNA fragments found as circulating cell-free DNA. Thus, we sought to avoid differences in amplicon generation, library preparation or sequencing depth and all samples were subjected to the same amplicon and library generation protocol and analyzed in parallel.

We also compared the clonal evolution of cancer in the ccfDNA mutations over the course of disease progression. After metastasis, new ccfDNA mutations are gained both in colon (33.8%) and pancreatic cancer (62.6%) and were not detected at the time of diagnosis of the primary cancer. This indicates clonal selection due to treatment, malignant progression or metastatic spread to different tissues with different microenvironmental selection pressure. Complementary to the gain of mutations after metastasis, we also observed a loss of approximately half of the ccfDNA mutations. This loss of cancer subpopulations will be due to surgical removal of the primary tumor as well as patient treatment.

One final caveat in ccfDNA mutation analysis is the assignment of mutant DNA to cancer lesions under study rather than spontaneous mutations that occurred in other tissues and have no disease relevance. Recently reported analysis of mutation accumulation in human adult stem cells in different tissues showed that on average 40 new mutations arise per year during the life time of an individual [28]. In a study focused on the analysis of human skin, biopsies of physiologically normal skin showed 2 to 6 mutations per million bases per cell and matched to a large extent with cancer-associated mutations [29]. Notably, the frequency of mutations seen in normal skin is within the median range of 1 - 10 somatic mutations observed for human cancers [1]. We did not find detailed reports on ccfDNA mutation frequencies in healthy individuals but the occurrence of spontaneous mutations in healthy tissues provide a caveat on interpreting mutant ccfDNA as evidence of the presence of a cancerous lesion.

In conclusion, we found that ccfDNA appears to represent the heterogeneity of colon and pancreatic cancer more extensively than tumor tissue DNA. The analysis of a relatively broad panel of cancer-related genes is feasible for ccfDNA and would allow monitoring of changes in the molecular makeup over time and under therapy. A challenge will be to what extent altered mutation patterns in ccfDNA could also prompt a change in treatment.

Acknowledgements

We thank the patients who donated blood and tissue samples for molecular analysis. We also thank pathologist Brent Harris, MD PhD (Georgetown University) for evaluating the tumors and Sarah Martinez Roth, MS (Georgetown University) for help with the experiments. EEV, AW, GG and JLM designed the research study, EEV, AJ and JNM performed the research, GTG and EEV analyzed the data, EEV and AW wrote the manuscript. This research was funded by the Ruesch Center for the Cure of GI Cancers (EEV, AW, JLM) and P30 CA51008 (AW)

Conflicts of interest

None of the authors have conflicts to disclose.

library control DNA		
gene	variant	VAF
<i>BRAF</i>	p.V600E	8.56
<i>EGFR</i>	p.E746_A750del	1.30
<i>EGFR</i>	p.L858R	2.21
<i>EGFR</i>	p.G719S	28.50
<i>KIT</i>	p.D816V	9.33
<i>KRAS</i>	p.G12D	5.95
<i>KRAS</i>	p.G13D	16.69
<i>NRAS</i>	p.Q61K	12.78
<i>PIK3CA</i>	p.E545K	8.46
<i>PIK3CA</i>	p.H1047R	16.36

VAF: variant allele frequency

Supplementary Figure 1. Non-synonymous mutations in the DNA standard provided as a control.

REFERENCES

- 1 Alexandrov, L. B. et al. (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415-421
- 2 Gerlinger, M. et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366, 883-892
- 3 Marusyk, A., Almendro, V. & Polyak, K. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 12, 323-334
- 4 Schiavon, G. et al. (2015) Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* 7, 313ra182
- 5 Garcia-Murillas, I. et al. (2015) Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* 7, 302ra133
- 6 Diehl, F. et al. (2005) Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 102, 16368-16373
- 7 Diehl, F. et al. (2008) Circulating mutant DNA to assess tumor dynamics. *Nat Med* 14, 985-990
- 8 Crowley, E., Di Nicolantonio, F., Loupakis, F. & Bardelli, A. (2013) Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* 10, 472-484
- 9 Haber, D. A. & Velculescu, V. E. (2014) Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov* 4, 650-661
- 10 Rosell, R. & Karachaliou, N. (2016) Lung cancer: Using ctDNA to track EGFR and KRAS mutations in advanced-stage disease. *Nat Rev Clin Oncol* 13, 401-402
- 11 Alix-Panabieres, C. & Pantel, K. (2016) Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov* 6, 479-491
- 12 Rapisuwon, S., Vietsch, E. E. & Wellstein, A. (2016) Circulating biomarkers to monitor cancer progression and treatment. *Comput Struct Biotechnol J* 14, 211-222
- 13 Vietsch, E. E., van Eijck, C. H. & Wellstein, A. (2015) Circulating DNA and Micro-RNA in Patients with Pancreatic Cancer. *Pancreat Disord Ther.* 2015 Jun;5(2). pii: 156.
- 14 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17, 3, doi:10.14806/ej.17.1.200
- 15 McKenna, A. et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297-1303
- 16 Wilm, A. 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, 11189-11201
- 17 Obenchain, V. et al. (2014) VariantAnnotation : a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* 30, 2076-2078
- 18 Scholer, L. V. et al. (2017) Clinical implications of monitoring circulating tumor DNA in patients with colorectal cancer. *Clin Cancer Res* 15;23(18):5437-5445.
- 19 Allenson, K. et al. (2017) High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients. *Ann Oncol* 28, 741-747
- 20 Siravegna, G. et al. (2015) Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* 21, 795-801
- 21 Pishvaian, M. J. et al. (2016) A pilot study evaluating concordance between blood-based and patient-matched tumor molecular testing within pancreatic cancer patients participating in the Know Your Tumor (KYT) initiative. *Oncotarget* 8;8(48):83446-83456

- 22 Cheng, H. et al. (2017) Analysis of ctDNA to predict prognosis and monitor treatment responses in metastatic pancreatic cancer patients. *Int J Cancer* 140, 2344-2350
- 23 Perkins, G. et al. (2012) Multi-purpose utility of circulating plasma DNA testing in patients with advanced cancers. *PLoS One* 7, e47020
- 24 Mouliere, F. et al. (2011) High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* 6, e23418
- 25 Underhill, H. R. et al. (2016) Fragment Length of Circulating Tumor DNA. *PLoS Genet* 12, e1006162
- 26 Jiang, P. et al. (2015) Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* 112, E1317-1325
- 27 Mouliere, F. & Rosenfeld, N. (2015) Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc Natl Acad Sci U S A* 112, 3178-3179
- 28 Blokzijl, F. et al. (2016) Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538, 260-264
- 29 Martincorena, I. et al. (2015) Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348, 880-886