EBioMedicine 10 (2016) 117–123

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer

Lynnette Fernandez-Cuesta [a,1], Sandra Perdomo [a,b,1], Patrice H. Avogbe [a,1], Noemie Leblay [a], Tiffany M. Delhomme [a], Valerie Gaborieau [a], Behnoush Abedi-Ardekani [a], Estelle Chanudet [a], Magali Olivier [a], David Zaridze [c], Anush Mukeria [c], Marta Vilensky [d], Ivana Holcatova [e], Jerry Polesel [f], Lorenzo Simonato [g], Cristina Canova [g], Pagona Lagiou [h], Christian Brambilla [i], Elisabeth Brambilla [i], Graham Byrnes [a], Ghislaine Scelo [a], Florence Le Calvez-Kelm [a], Matthieu Foll [a], James D. McKay [a,*], Paul Brennan [a,*]

[a] International Agency for Research on Cancer (IARC-WHO), 150 cours Albert Thomas, 69008 Lyons, France
[b] Institute of Nutrition, Genetics and Metabolism Research, Universidad El Bosque, Bogotá, Colombia
[c] Russian N.N. Blokhin Cancer Research Centre, Moscow, Russian Federation
[d] Instituto Angel Roffo, Buenos Aires, Argentina,
[e] 1st Faculty of Medicine, Charles University, Prague, Czech Republic
[f] CRO Aviano National Cancer Institute, Aviano, Italy
[g] Department of Molecular Medicine, University of Padova, Padova, Italy,
[h] University of Athens Medical School, Greece
[i] CHU Grenoble, University Grenoble- Alpes, INSERM U823, Grenoble, France

ARTICLE INFO

Article history:
Received 25 May 2016
Received in revised form 20 June 2016
Accepted 23 June 2016
Available online 25 June 2016

Keywords:
ctDNA
cfDNA
Small-cell lung cancer
TP53 mutations
Early detection
Screening

ABSTRACT

Circulating tumor DNA (ctDNA) is emerging as a key potential biomarker for post-diagnosis surveillance but it may also play a crucial role in the detection of pre-clinical cancer. Small-cell lung cancer (SCLC) is an excellent candidate for early detection given there are no successful therapeutic options for late-stage disease, and it displays almost universal inactivation of *TP53*. We assessed the presence of *TP53* mutations in the cell-free DNA (cfDNA) extracted from the plasma of 51 SCLC cases and 123 non-cancer controls. We identified mutations using a pipeline specifically designed to accurately detect variants at very low fractions. We detected *TP53* mutations in the cfDNA of 49% SCLC patients and 11.4% of non-cancer controls. When stratifying the 51 initial SCLC cases by stage, *TP53* mutations were detected in the cfDNA of 35.7% early-stage and 54.1% late-stage SCLC patients. The results in the controls were further replicated in 10.8% of an independent series of 102 non-cancer controls. The detection of *TP53* mutations in 11% of the 225 non-cancer controls suggests that somatic mutations in cfDNA among individuals without any cancer diagnosis is a common occurrence, and poses serious challenges for the development of ctDNA screening tests.

© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Cell-free DNA (cfDNA) refers to nucleic acids detected in body fluids and are thought to arise from two sources: passive release through cell death (Jahr et al., 2001), and active release by cell secretion (Stroun et al., 2000). DNA from cancer cells also contributes to the total load of cfDNA (Schwarzenbach et al., 2011), and the fraction of cfDNA that comes from cancer cells is called circulating-tumor DNA (ctDNA). ctDNA has been estimated to make up about 0.01%–1% of cfDNA for early-stage disease, reaching over 40% for late-stage disease (Beaver et al., 2014; Bettegowda et al., 2014; Couraud et al., 2014; Diehl et al., 2007; Forshew et al., 2012; Newman et al., 2014; Sausen et al., 2015). Despite its intrinsic limitations, including technical issues in the sample collection, detection, and identification of tumor origin, ctDNA is emerging as a key potential biomarker for monitoring response to treatment and relapse (Dawson et al., 2013; Esposito et al., 2014; Forshew et al., 2012; Garcia-Murillas et al., 2015; Murtaza et al., 2013; Roschewski et al., 2015; Siravegna et al., 2015). The potential of ctDNA is not limited to post-diagnosis surveillance but it may also play a crucial role in the detection of pre-clinical cancer. If successful, this could be translated

* Corresponding authors.
E-mail addresses: MckayJ@iarc.fr (J.D. McKay), BrennanP@iarc.fr (P. Brennan).
[1] Equally contributing authors.

http://dx.doi.org/10.1016/j.ebiom.2016.06.032
2352-3964/© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

into much improved cancer survival, in particular for those cancer sites that are typically diagnosed at a late stage, and for which survival is poor, such as lung, pancreatic, or esophageal cancer (Brennan and Wild, 2015). However, implementation of ctDNA tests that detect pre-clinical disease in a non-symptomatic population will have to show an extremely high specificity if they are to provide meaningful results, or be part of a multi-modal screening program.

Very few studies have focused on the evaluation of ctDNA detection in early-stage cancers (*i.e.* stage I-II tumors) with even less data available on the detection of ctDNA in blood samples from pre-symptomatic cancer patients (Amant et al., 2015; Beaver et al., 2014; Bettegowda et al., 2014; Garcia-Murillas et al., 2015; Gormally et al., 2006; Jamal-Hanjani et al., 2016; Sausen et al., 2015); Table S1). In addition, these studies have aimed to detect specific mutations in cfDNA (most of them using digital droplet PCR) following previous assessment of the tumor mutational profile. This approach is only viable for cancers with common hot-spot mutations and is not amenable for most screening purposes. This is because early detection of pre-clinical cancer requires variant detection to be done without prior knowledge from tumor tissue of the expected mutations. Another limitation of these studies is the major assumption that circulating-mutated fragments would be absent (or very rare) in individuals without cancer. Demonstrating that any ctDNA detection marker has a specificity close to 100% would be of fundamental importance for large-scale utility in an asymptomatic population (Wentzensen and Wacholder, 2013).

Small-cell lung cancer (SCLC) accounts for about 15% of all lung tumors and has a 5-year survival below 5%. While SCLC tumors are initially sensitive to chemotherapy, they invariably relapse with a resistant and deadly disease. We and others have found that, contrary to lung adenocarcinomas and squamous-cell lung carcinomas, mutations in therapeutic targets are rare in SCLC (George et al., 2015; Peifer et al., 2012; Rudin et al., 2012). *TP53* is inactivated in virtually all SCLC cases, and *TP53* mutations are known to be an early event in the development of this disease. Given the almost uniform presence of *TP53* mutations in SCLC, we have investigated to what extent mutations in this gene can be identified in the cfDNA of patients with SCLC tumors. In addition, we have also assessed two independent series of non-cancer controls to evaluate the specificity of the approach.

2. Material and Methods

2.1. Study Population

Small-cell lung cancer (SCLC) patients and controls were recruited through an IARC case-control study coordinated in Moscow from 2006 to 2012 (Wozniak et al., 2015). Cases were incident cancer patients collected from the Russian N.N. Blokhin Cancer Research Centre and the Moscow City Clinical Oncology Dispensary. The staging of the SCLC cases is based on the recent recommendations of the International Association for the Study of Lung Cancer (IASLC) (Nicholson et al., 2016). Controls were recruited from individuals visiting two Moscow general hospitals for disorders unrelated to lung cancer and its associated risk factors (Table 1). The controls were matched for age, sex, and smoking status. All study participants provided written-informed consent and participated in an interview. Peripheral blood was collected in EDTA tubes at the time of interview and processed as rapidly as possible (generally within 2 h). For cases, blood draw was performed before surgery and any adjuvant treatment. Plasma samples were isolated by centrifugation of whole blood at 2000 ×*g* for 10 min at room temperature. Samples were stored at −80 °C. All specimens were obtained in accordance with the declaration of Helsinki guidelines and were approved by the local Institutional Review Board and the IARC Ethics Committee. A total of 52 SCLC cases and 165 controls were initially included but only 51 SCLC and 123 controls passed the sequencing QC criteria (see Sequencing Data Analyses, Annotation, and Filters), and were therefore included in down-stream analyses.

Table 1
Characteristics of small-cell lung cancer cases and controls from Russia, and additional replication controls from Greece, Czech Republic, Italy, and Argentina.

	Cases	Controls	Replication controls
Origin (country)			
Russia	51	123	
Greece			9 (8.8%)
Czech Republic			14 (13.7%)
Italy			40 (39.2%)
Argentina			39 (38.2%)
Total	51	123	102
Age at diagnosis			
<40	2 (3.9%)	2 (1.6%)	3 (2.9%)
40–49	4 (7.8%)	11 (8.9%)	15 (14.7%)
50–59	15 (29.4%)	42 (34.2%)	33 (32.4%)
60–69	22 (43.1%)	55 (44.7%)	33 (32.4%)
70+	8 (15.7%)	13 (10.6%)	18 (17.7%)
Sex			
Male	43 (84.3%)	107 (87.0%)	76 (74.5%)
Female	8 (15.7%)	16 (13.0%)	26 (25.5%)
Smoking status			
Never	5 (9.8%)	35 (28.4%)	34 (33.4%)
Former	6 (11.8%)	28 (22.8%)	25 (24.5%)
Current	40 (78.4%)	60 (48.8%)	43 (42.1%)
Alcohol status			
Never	30 (58.8%)	32 (26.0%)	16 (15.7%)
Former	4 (7.8%)	18 (14.6%)	14 (13.7%)
Current	17 (33.4%)	73 (59.4%)	72 (70.6%)
Tumor stage of cases			
I	7 (13.7%)		
II	7 (13.7%)		
III	28 (54.9%)		
IV	9 (17.6%)		
Disease type of hospital controls			
Infectious & parasitic diseases		0 (0.0%)	1 (1.0%)
Neoplasms		3 (2.4%)	0 (0.0%)
Endocrine, nutritional and metabolic diseases and immunity disorders		6 (4.9%)	1 (1.0%)
Diseases of blood and blood-forming organs		2 (1.6%)	1 (1.0%)
Diseases of the nervous system and sense organs		28 (22.8%)	6 (5.9%)
Diseases of the sense organs		0 (0.0%)	5 (4.9%)
Diseases of the circulatory system		25 (20.3%)	2 (2.0%)
Diseases of the respiratory system		3 (2.4%)	3 (2.9%)
Diseases of the digestive system		19 (15.4%)	19 (18.6%)
Diseases of the genitourinary system		19 (15.4%)	17 (16.7%)
Diseases of the skin and subcutaneous tissue		3 (2.4%)	2 (2.0%)
Diseases of the musculoskeletal system and connective tissue		9 (7.3%)	18 (17.6%)
Symptoms, signs and ill-defined conditions		0 (0.0%)	8 (7.8%)
Injury and poisoning		6 (4.9%)	18 (17.6%)
External causes		0 (0.0%)	1 (1.0%)

In order to further evaluate the prevalence of circulating-mutated fragments in non-cancer controls, 114 additional controls were retrieved from two large multicenter case-control studies coordinated by IARC. One was a study on alcohol-related cancers and genetic susceptibility in Europe (the ARCAGE study) that was conducted from 2002 to 2005 and from which we included hospital-based controls recruited in Prague (Czech Republic), Athens (Greece), Aviano, Padova and Turin (Italy). The second was the Latin American study of head and neck cancer conducted from 1998 to 2002 and from which we selected hospital-based controls from two institutions in Buenos Aires (Argentina). Additional details of the 2 large multicenter case-control studies are included elsewhere (Lagiou et al., 2009; Ribeiro et al., 2011). Out of the 114 controls initially included, 102 passed the sequencing QC criteria (see Sequencing Data Analyses, Annotation, and Filters), and were therefore included in down-stream analyses.

2.2. cfDNA Extraction

cfDNA was extracted from 0.8–1.3 mL of plasma using the QIAamp DNA Circulating Nucleic Acid kit (Qiagen) following manufacturer's instructions. cfDNA was eluted into 100 μL of elution buffer and quantified with the Qubit DNA high-sensitivity assay kit (Invitrogen Corporation). Details regarding amount of cfDNA are included in the Table S2.

2.3. Primer Design and Amplification of Targets

Twenty-one amplicons of 150 bp in size were designed (Eurofins Genomics Ebersberg, Germany) to cover exons 2 to 10 of *TP53*. The GeneRead DNAseq Panel PCR Kit V2 (Qiagen) was used for target enrichment. A validated in-house protocol was used to set up multiplex PCRs in 10 μL reaction volume, containing 5 ng cfDNA, 60 nM of primer pool and 0.73 μL of HotStarTaq enzyme. The experiments were carried out in two physically isolated laboratory spaces: one for sample preparation and another one for post-amplification steps. Amplification was carried out in a 96-well format plates DNA engine Tetrad 2 Peltier Thermal Cycler (BIORAD) as follows: 15 min at 95 °C and 30 cycles of 15 s at 95 °C and 2 min at 60 °C and 10 min at 72 °C.

2.4. Library Preparation and Sequencing

Following target enrichment, PCR products were purified using Serapure beads (prepared in-house and produced by ThermoFisher Scientific Inc.). A ratio of 2:1 of Serapure beads to PCR products was used. Purified amplicons were quantified with the Qubit DNA high-sensitivity assay kit (Invitrogen Corporation). Library preparation was done using 150 ng of purified PCR products and the NEBNext Fast DNA Library Prep Set for Ion Torrent (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions. Amplicons were end-repaired and ligated to the specific adapters and individual barcodes (designed in-house and produced by Eurofins MWG Operon, Ebersberg, Germany). Libraries were cleaned up, amplified with a final step of 6 PCR cycles, pooled in an equimolar way and loaded onto a 2% agarose gel and subjected to an electrophoresis at 150 V for 1.5 h. Fragments of 180–220 bp in size were selected and the pooled DNA library was recovered from the gel using the QIAquik Gel Extraction kit (Qiagen). The quality and quantity of the library was then assessed on the Agilent 2100 Bioanalyzer on-chip electrophoresis (Agilent Technologies, USA) for the absence of possible adapter dimers and heterodimers. The pooled libraries were sequenced on the Ion Torrent™ Proton Sequencer (Life Technologies Corp., USA) aiming for deep coverage (5,000×), using the Ion PI™ Hi-Q™ OT2 200 Kit and the Ion PI™ Hi-Q™ Sequencing 200 Kit with the Ion PI chip V3 (Life Technologies Corp., USA), following the manufacture's protocols.

2.5. Technical Duplication

Technical duplicates were undertaken for each sample including amplification, library preparation, and sequencing. Each technical duplicate pair was assessed on separate plates to limit the possibility of a contamination.

2.6. Sequencing Data Analyses, Annotation, and Filters

Short reads were aligned to the hg19 human reference genome and BAM files were generated using the Torrent Suite software (v4.4.2) with default parameters. Reads with a mapping quality below 20 were excluded from subsequent analysis. We also excluded those libraries for which the on-target median coverage was significantly lower in comparison to the other libraries sequenced in the same batch. On-target median coverage for both libraries is shown in Table S2.

For the calling of variants we used *Needlestack*, a recently developed ultra-sensitive variant caller, which estimates the distribution of sequencing errors across multiple samples to reliably identify variants present in very low proportion (https://github.com/IARCbioinfo/needlestack) (unpublished data; Delhomme et al.). Contrary to most existing algorithms, *Needlestack* can deal with both single nucleotide substitutions (SNVs) and short indels. At each position and for each candidate variant, sequencing errors are modeled using a robust negative binomial regression (Aeberhard et al., 2014), with a linear link and a zero intercept. True variants are outliers from this error model (Fig. 1a). The robust estimator of the over-dispersion parameter avoids bias due to these outliers (Aeberhard et al., 2014). For each sample a *p*-value against the null hypothesis of being a sequencing error is calculated, and further transformed into a *q*-value using the Benjamini and Hochberg false-discovery rate control method (Benjamini and Yosef, 1995). *Q*-values are reported as a Phred-scale quality score: $Q = -10\log_{10}(q\text{-value})$, and we used a threshold of $Q > 50$ to call variants. For each variant, we also calculated the relative variant strand bias defined by:

$$RVSB = \frac{max(AO_pDP_m, AO_mDP_p)}{AO_pDP_m + AO_mDP_p}$$

where *DP* and *AO* denote respectively the total number of reads and the number of reads matching the candidate variant, with the subscripts *p* and *m* referring to the forward and reverse strands respectively. In the complete absence of strand bias, $RVSB = 0.5$ and $AO_p/AO_m = DP_p/DP_m$, whereas for a completely biased variant, $RVSB = 1$. We filtered out variants with $RVSB > 0.85$.

Variant calls were annotated using ANNOVAR (Yang and Wang, 2015). We only considered meaningful *TP53* mutations those that matched the following criteria: the mutation has already been reported in COSMIC and the mutation is a nonsense, indel, splicing or a missense variant that is classified as deleterious by SIFT or Polyphen.

2.7. Analyses of Technical Duplicates

While Needlestack models recurring errors, rare errors such as those generated by the DNA polymerase will be identified as variants. Such errors will be generally specific to a particular preparation. To filter out these rare errors, we required each of the individual's technical duplicate to be a Needlestack outlier (Fig. 1a). Additionally, we identified and excluded a few genomic positions that show a particularly high proportion (>10%) of these errors (*i.e.* higher than the estimated sequencing error rate, but not always replicable in two independent preparations).

2.8. Statistical Analyses

Effect of age, smoking, and alcohol status on the presence of *TP53* mutations was assessed using logistic regression adjusting one for the

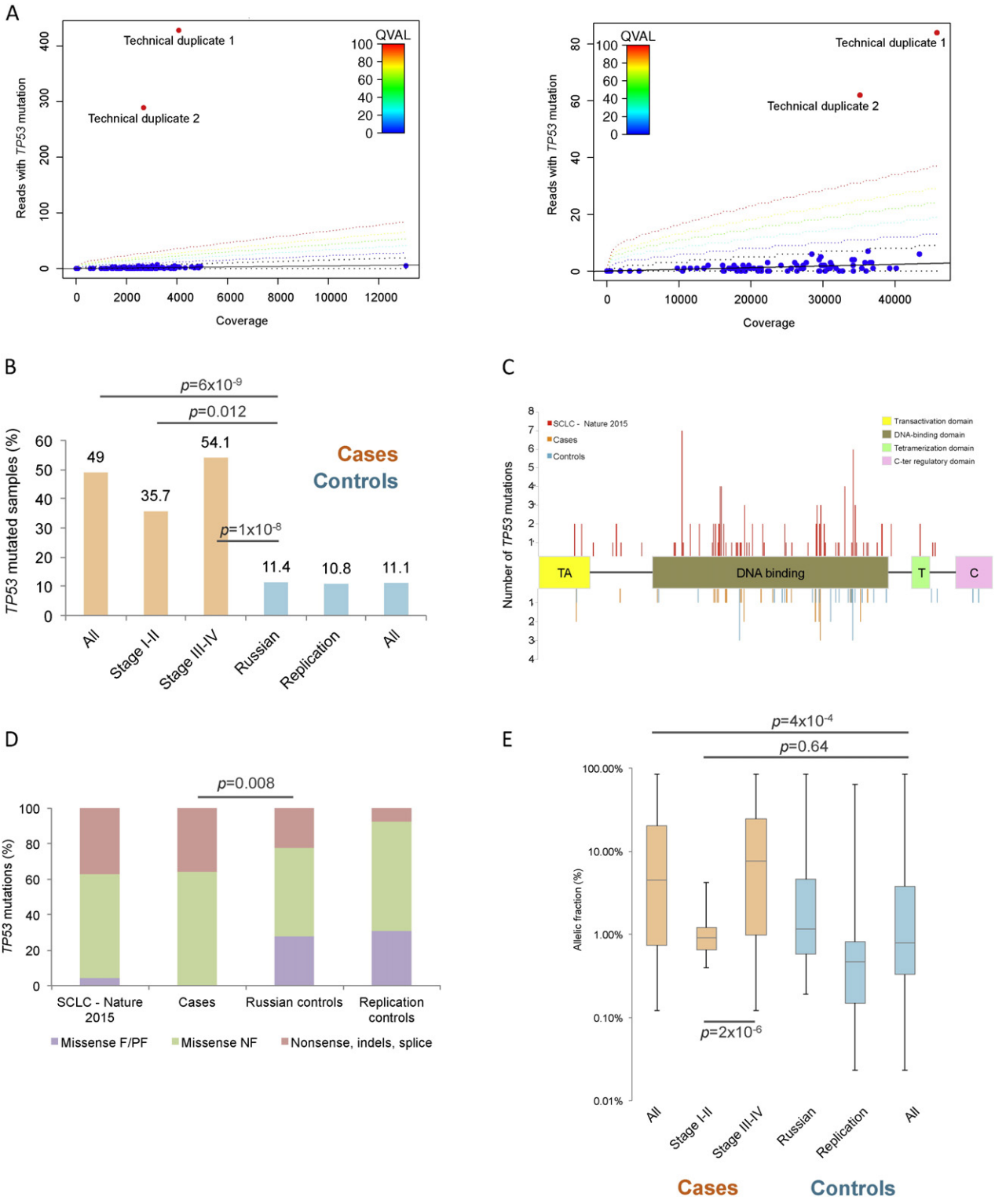

Fig. 1. Characteristics of *TP53* mutations in cases and controls (a) Two examples of variants called using Needlestack's regression model of sequencing error. Each dot represents a sequenced library (two dots per sample) colored according to its phred-scaled *q*-value. The black regression line shows the estimated sequencing-error rate along with the 99% confidence interval (black dotted lines) containing samples. Colored-dotted lines correspond to the limits of regions defined for different significance *q*-value thresholds. Both technical duplicates appear as outliers from the regression (in red), and are therefore classified as carrying the given mutation; (b) Percentage of *TP53* mutated samples in the cfDNA of Russian cases and controls, and replication controls; (c) Distribution of *TP53* mutations found in SCLC tumors (George et al., 2015) and in our series of cases and controls across the different p53 protein domains; (d) Type of mutations and functional impact of missense ones based on the IARC *TP53* database: F (functional), PF (partially functional), NF (non-functional); (e) Percentage of allelic fractions of the *TP53* mutations detected in the cfDNA of Russian cases and controls, and replication controls. The whiskers represent the minimum and maximum values.

others. Effect of case-status on the presence of *TP53* mutations was assessed using unadjusted logistic regression. *p*-Values to test the differences between the pattern of mutations in cases and controls are derived from Pearson's chi-square tests. All the analyses were conducted using SAS 9.4. For comparison of allelic fractions, we log-transformed the data and performed a *t*-test (2 tailed, unequal variance).

3. Results

The characteristics of the cases and controls are shown in Table 1. We detected 31 *TP53* mutations in 25 SCLC patients (49%, 25/51). When the 51 initial SCLC cases were stratified by stage, we found that 35.7% (5/14) of the stage I–II and 54.1% (20/37) of the stage III–IV, carried detectable *TP53* mutations in their cfDNA (Fig. 1b). While statistically significant in cases *versus* controls (p-value $= 6 \times 10^{-9}$), 18 *TP53* mutations were detected in 14 of the Russian non-cancer controls (11.4%, 14/123). The significance was also maintained when stratifying by stage (stage I–II *versus* controls, p-value $= 0.012$; stage III–IV *versus* controls, p-value $= 1 \times 10^{-8}$). We replicated these observations in an independent series of 102 controls, and found a comparable proportion of *TP53* mutated samples (10.8%, 13 *TP53* mutations in 11 controls).

Similarly to what is expected for *TP53* mutations present in cancer, most of the mutations in cases and controls altered amino acids coding for the *TP53* DNA-binding domain, which is critical for the transactivation activity of this gene (Fig. 1c). We next evaluated the characteristics of the mutations found in cases *versus* controls. Chi-square test analysis showed that there was a statistically significant difference between the mutational pattern found in cases *versus* controls (p-value $= 0.008$). The fraction of nonsense, indel, or splicing mutations found in the cases was similar to that previously reported for SCLC tumors (George et al., 2015) (35.5% *versus* 37% respectively), whereas this proportion was slightly lower in controls (22.2% in the Russian, and 7.7% in the replication controls; Fig. 1d). We used the IARC *TP53* database to classify the missense mutations in functional, partially functional, or non-functional based on the *in vitro* transcriptional activity of the resulting protein. Most missense mutations found in SCLC tumors (George et al., 2015) (92.6%) and cfDNA from cases (100%) were classified as resulting in a non-functional protein. However, controls had a higher proportion of missense mutations that retained some transcriptional activity (~30%; Fig. 1d).

We also compared the allelic fractions (AFs) of the *TP53* mutations found in the cfDNA of cases and controls. The AFs for a given mutation were similar in the two independent libraries, demonstrating the reproducibility of the assay (Table S3, S4, and S5). The AFs for the cases ranged from 0.12% to 84.81% (median 4.6%). In the Russian controls the AFs ranged from 0.19% to 84.94% (median 1.2%), and in the replication controls they ranged from 0.02% to 63.74% (median 0.5%) (Fig. 1e). The statistically significant difference in the AFs between cases and controls (p-value $= 4 \times 10^{-4}$) is explained by the presence of late-stage SCLC tumors, since the median AF of the *TP53* mutations detected in the five stage I–II SCLC (0.9%) is not statistically different from that found in controls (p-value $= 0.64$), while it differed from the median AF of stage III-IV SCLC tumors (8.2%; p-value $= 2 \times 10^{-6}$; Fig. 1e).

Finally, we sequenced the DNA extracted from the white-blood cells (WBC) of 39 cfDNA *TP53*-positive patients, from which material was available (19 cases and 20 controls). Five cfDNA *TP53* mutations (from one case and four controls) were detected in the WBC DNA, with similar AFs to those found in the cfDNA (Table 2). For one control (MLT-14), the AFs in both cfDNA and WBC DNA were around 50%, being consistent with a heterozygous germ-line variant. The other four mutations were detected at AFs consistent with a somatic origin (AFs below 11%) in both cfDNA and WBC DNA (Table S2).

Table 2
Overview of the cfDNA mutations also detected in the white-blood cells (WBC) DNA, and their corresponding allelic fractions in each technical duplicate (AFs in %).

Sample	*TP53* mutation	AFs detected in cfDNA		AFs detected in WBC	
SCLC-21	p.Y220C	0.90	1.27	0.50	0.70
MLT-6	p.R175G	4.09	4.41	4.40	4.50
MLT-14	p.G154S	47.17	50.58	52.10	54.90
ARG-1	p.R273C	5.22	5.58	7.30	10.40
ITA-8	p.V272M	0.78	0.80	0.90	1.40

Taken altogether, cancer-like *TP53* mutations were identified in 25 of the 225 non-cancer controls analyzed in this study (11.1%). We checked if the presence of *TP53* mutations in the controls was correlated with age, smoking status, or alcohol (adjusting one for the other), but none of these factors was found to be associated.

4. Discussion

Inactivation of *TP53* by mutation has been reported to occur in over 90% of SCLC cases (George et al., 2015). In this study we were able to detect *TP53* mutations in the cfDNA of 49% SCLC patients and, when stratifying by stage, in the cfDNA of 35.7% early-stage cases. These proportions matched those reported for other cancer types (Bettegowda et al., 2014). Unfortunately, we did not have the correspondent tumors to confirm that the *TP53* mutations detected in the cfDNA originate from the SCLC tumors. However, our method detected *TP53* mutations in 60% of the cfDNA samples of an independent French series of 10 SCLC patients (all of them carrying *TP53* somatic mutations in their tumors). Importantly, each of the *TP53* mutations found in the cfDNA matched the one found in the SCLC tumor (data not shown).

We also observed cfDNA *TP53*-mutated fragments in 11.4% of 123 matched non-cancer controls. Acknowledging the potential for bias in our selection of controls (such as differential performance in QC criteria or cfDNA amount, between cases and controls), we screened a second series of 102 non-cancer controls, and found a comparable proportion of *TP53* mutated samples in this independent group (13 *TP53* mutations in 11 controls, 10.8%). Altogether, the detection of *TP53* mutations in 11.1% of the 225 non-cancer controls, from two independent groups of samples, suggests that the presence of circulating-mutated fragments among individuals without any diagnosed cancer is a common occurrence, and poses serious challenges for the development of ctDNA screening tests for the early detection of cancer.

Only two other studies have explored the potential presence of circulating-mutated fragments in non-cancer subjects. A study within the EPIC prospective cohort (GENAIR) that used blood samples from controls, found that *KRAS* and *TP53* mutations were detectable in the cfDNA of 1% and 3.2% healthy subjects, respectively, without a cancer diagnoses five years subsequent to blood draw (Gormally et al., 2006). The higher percentage of *TP53*-positive controls in our analyses is likely to be explained by the fact that these analyses within EPIC were undertaken using DHPLC (denaturing high-pressure liquid chromatography) and Sanger sequencing, and these techniques are less sensitive and only allow for detection of mutations with allelic fractions of 3% or more. Further, only *TP53* exons five to nine were analyzed. If we limited our analysis to mutations from exons five to nine and AFs greater that 3%, we would have found a comparable number of *TP53*-positive controls (2.7%, 6/225). More recently, Krimmel and colleagues have reported extremely low-frequency cancer-like *TP53* mutations in the peritoneal fluid from both women with ovarian cancer and those with benign lesions, using duplex sequencing (Krimmel et al., 2016). They also showed that low frequency *TP53* mutagenesis increases with age and cancer. Overall, these results support the need for further ctDNA studies to incorporate series of non-cancer controls in order to improve validation of detection and analysis techniques.

A potential limitation of our study was the use of hospital controls as proxies of healthy people. Controls were admitted for a wide variety of routine conditions unrelated to tobacco and it is implausible that a high proportion of the controls with a detectable damaging *TP53* mutation developed a cancer in the short term. However, we cannot exclude this occurring in a small number of controls nor enriching for non-cancer diseases with unknown impact on the presence of circulating-tumor fragments. Nevertheless, as noted above, the prevalence of *TP53* mutations in our study is approximately equal to that of GENAIR (when applying the same detection thresholds). Prospective cohorts may help to overcome the limitations of using hospital controls and also help to

determine at what point in the development of the disease is ctDNA detectable in blood, and its correlation with a plausible diagnosis.

The source of circulating-mutated fragments in the cfDNA of apparently healthy people is still unknown. There is, however, accumulating evidence that clonal expansions are more frequent than originally thought. Martincorena and colleagues estimated that there are 9.5 clones per cm^2 of normal human skin carrying a driver mutation in *TP53* in a selected population for high-sun exposure (Martincorena and Campbell, 2015). Such clonal expansions might act as a reservoir of circulating-mutated fragments in cfDNA. In addition, several studies have shown that a subset of normal individuals could undergo clonal hematopoiesis with mutations in driver genes (Genovese et al., 2014; Jacobs et al., 2012; Jaiswal et al., 2014; Laurie et al., 2012; Wong et al., 2015; Xie et al., 2014). Consistent with this, we observed 4 cfDNA *TP53* mutations that appeared to be from clonal expansions in WBC. We also noted 2 *TP53* mutations in one SCLC case, apparently from different organs; one originating from WBC, the second we assume from the SCLC tumor. Such ambiguity around the tissue of origin of the circulating-mutated fragments adds another layer of complexity when using ctDNA for early detection.

The potential of ctDNA for early diagnosis of cancer is an area of much interest (The Lancet Oncology, 2016). While implementation in a screening setting will undoubtedly require more sensitive and specific tests as well as validation in pre-diagnostic blood samples, the unexpected presence of known cancer mutations in cfDNA among non-cancer controls represents an important challenge.

Funding Sources

This work has been funded by IARC. TMD was supported by la Ligue Nationale (Française) Contre le Cancer. The work reported in this paper was undertaken during the tenure of PHA's Postdoctoral Fellowship from the International Agency for Research on Cancer, partially supported by the European Commission FP7 Marie Curie Actions – People – Co-funding of regional, national and international programmes (COFUND).

Conflict of Interest

The authors declare no conflict of interests

Author Contributions

JDM and PB conceived and designed the study. GS coordinated the recruitment of samples and associated data. DZ, AM, MV, IH, JP, LS, CC, PL, CB, and EB provided samples and associated data. SP, PA, FLCK and JDM developed and undertook the laboratory procedures. SP, PA, and NL performed the experiments. TD, MF, and JDM developed Needlestack and performed sequencing analyses. LFC and SP analyzed and interpreted the data. VG and GB performed and provided statistical advice, respectively. EC, MO, and BAA provided intellectual input. LFC, SP, JDM, and PB wrote the manuscript. All authors reviewed and approved the final version for publication.

Acknowledgements

We acknowledge all the people who donated their biological specimens. We also acknowledge Priscilia Chopard, Helene Renard, Amelie Chabrier, Nathalie Forey, Geoffroy Durand, and Nicolas Lemaitre for their technical assistance.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2016.06.032.

References

Aeberhard, W.H., Cantoni, E., Heritier, S., 2014. Robust inference in the negative binomial regression model with an application to falls data. Biometrics 70 (4), 920–931.

Amant, F., Verheecke, M., Wlodarska, I., Dehaspe, L., Brady, P., Brison, N., Van Den Bogaert, K., Dierickx, D., Vandecaveye, V., Tousseyn, T., Moerman, P., Vanderstichele, A., Vergote, I., Neven, P., Berteloot, P., Putseys, K., Danneels, L., Vandenberghe, P., Legius, E., Vermeesch, J.R., 2015. Presymptomatic identification of cancers in pregnant women during noninvasive prenatal testing. JAMA Oncol. 1 (6), 814–819.

Beaver, J.A., Jelovac, D., Balukrishna, S., Cochran, R.L., Croessmann, S., Zabransky, D.J., Wong, H.Y., Valda Toro, P., Cidado, J., Blair, B.G., Chu, D., Burns, T., Higgins, M.J., Stearns, V., Jacobs, L., Habibi, M., Lange, J., Hurley, P.J., Lauring, J., VanDenBerg, D.A., Kessler, J., Jeter, S., Samuels, M.L., Maar, D., Cope, L., Cimino-Mathews, A., Argani, P., Wolff, A.C., Park, B.H., 2014. Detection of cancer DNA in plasma of patients with early-stage breast cancer. Clin. Cancer Res. 20 (10), 2643–2650.

Benjamini, Y., Yosef, H., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57 (1).

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., Antonarakis, E.S., Azad, N.S., Bardelli, A., Brem, H., Cameron, J.L., Lee, C.C., Fecher, L.A., Gallia, G.L., Gibbs, P., Le, D., Giuntoli, R.L., Goggins, M., Hogarty, M.D., Holdhoff, M., Hong, S.M., Jiao, Y., Juhl, H.H., Kim, J.J., Siravegna, G., Laheru, D.A., Lauricella, C., Lim, M., Lipson, E.J., Marie, S.K.N., Netto, G.J., Oliner, K.S., Olivi, A., Olsson, L., Riggins, G.J., Sartore-Bianchi, A., Schmidt, K., Shih, I.M., Oba-Shinjo, S.M., Siena, S., Theodorescu, D., Tie, J., Harkins, T.T., Veronese, S., Wang, T.L., Weingart, J.D., Wolfgang, C.L., Wood, L.D., Xing, D., Hruban, R.H., Wu, J., Allen, P.J., Schmidt, C.M., Choti, M.A., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Papadopoulos, N., Diaz, L.A., 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci. Transl. Med. 6 (224) (224ra24-224ra24).

Brennan, P., Wild, C.P., 2015. Genomics of cancer and a new era for cancer prevention. PLoS Genet. 11 (11), e1005522.

Couraud, S., Vaca-Paniagua, F., Villar, S., Oliver, J., Schuster, T., Blanche, H., Girard, N., Tredaniel, J., Guilleminault, L., Gervais, R., Prim, N., Vincent, M., Margery, J., Larive, S., Foucher, P., Duvert, B., Vallee, M., Le Calvez-Kelm, F., McKay, J., Missy, P., Morin, F., Zalcman, G., Olivier, M., Souquet, P.J., Bio, C.I.-i., 2014. Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. Clin. Cancer Res. 20 (17), 4613–4624.

Dawson, S.-J., Tsui, D.W.Y., Murtaza, M., Biggs, H., Rueda, O.M., Chin, S.-F., Dunning, M.J., Gale, D., Forshew, T., Mahler-Araujo, B., Rajan, S., Humphray, S., Becq, J., Halsall, D., Wallis, M., Bentley, D., Caldas, C., Rosenfeld, N., 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N. Engl. J. Med. 368 (13), 1199–1209.

Diehl, F., Schmidt, K., Choti, M.A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S.A., Kinzler, K.W., Vogelstein, B., Diaz Jr., L.A., 2007. Circulating mutant DNA to assess tumor dynamics. Nat. Med. 14 (9), 985–990.

Esposito, A., Bardelli, A., Criscitiello, C., Colombo, N., Gelao, L., Fumagalli, L., Minchella, I., Locatelli, M., Goldhirsch, A., Curigliano, G., 2014. Monitoring tumor-derived cell-free DNA in patients with solid tumors: clinical perspectives and research opportunities. Cancer Treat. Rev. 40 (5), 648–655.

Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D.W.Y., Kaper, F., Dawson, S.J., Piskorz, A.M., Jimenez-Linan, M., Bentley, D., Hadfield, J., May, A.P., Caldas, C., Brenton, J.D., Rosenfeld, N., 2012. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci. Transl. Med. 4 (136) (136ra68-136ra68).

Garcia-Murillas, I., Schiavon, G., Weigelt, B., Ng, C., Hrebien, S., Cutts, R.J., Cheang, M., Osin, P., Nerurkar, A., Kozarewa, I., Garrido, J.A., Dowsett, M., Reis-Filho, J.S., Smith, I.E., Turner, N.C., 2015. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci. Transl. Med. 7 (302) (302ra133).

Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., Purcell, S.M., Svantesson, O., Landen, M., Hoglund, M., Lehmann, S., Gabriel, S.B., Moran, J.L., Lander, E.S., Sullivan, P.F., Sklar, P., Gronberg, H., Hultman, C.M., McCarroll, S.A., 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. 371 (26), 2477–2487.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., Müller, C., Dahmen, I., Jahchan, N.S., Park, K.S., Yang, D., Karnezis, A.N., Vaka, D., Torres, A., Wang, M.S., Korbel, J.O., Menon, R., Chun, S.M., Kim, D., Wilkerson, M., Hayes, N., Engelmann, D., Pützer, B., Bos, M., Michels, S., Vlasic, I., Seidel, D., Pinther, B., Schaub, P., Becker, C., Altmüller, J., Yokota, J., Kohno, T., Iwakawa, R., Tsuta, K., Noguchi, M., Muley, T., Hoffmann, H., Schnabel, P.A., Petersen, I., Chen, Y., Soltermann, A., Tischler, V., Choi, C.M., Kim, Y.H., Massion, P.P., Zou, Y., Jovanovic, D., Kontic, M., Wright, G.M., Russell, P.A., Solomon, B., Koch, I., Lindner, M., Muscarella, L.A., la Torre, A., Field, J.K., Jakopovic, M., Knezevic, J., Castaños-Vélez, E., Roz, L., Pastorino, U., Brustugun, O.T., Lund-Iversen, M., Thunnissen, E., Köhler, J., Schuler, M., Botling, J., Sandelin, M., Sanchez-Cespedes, M., Salvesen, H.B., Achter, V., Lang, U., Bogus, M., Schneider, P.M., Zander, T., Ansén, S., Hallek, M., Wolf, J., Vingron, M., Yatabe, Y., Travis, W.D., Nürnberg, P., Reinhardt, C., Perner, S., Heukamp, L., Büttner, R., Haas, S.A., Brambilla, E., Peifer, M., Sage, J., Thomas, R.K., 2015. Comprehensive genomic profiles of small cell lung cancer. Nature 524 (7563), 47–53.

Gormally, E., Vineis, P., Matullo, G., Veglia, F., Caboux, E., Le Roux, E., Peluso, M., Garte, S., Guarrera, S., Munnia, A., Airoldi, L., Autrup, H., Malaveille, C., Dunning, A., Overvad, K., Tjonneland, A., Lund, E., Clavel-Chapelon, F., Boeing, H., Trichopoulou, A., Palli, D., Krogh, V., Tumino, R., Panico, S., Bueno-de-Mesquita, H.B., Peeters, P.H., Pera, G., Martinez, C., Dorronsoro, M., Barricarte, A., Navarro, C., Quiros, J.R., Hallmans, G., Day, N.E., Key, T.J., Saracci, R., Kaaks, R., Riboli, E., Hainaut, P., 2006. *TP53* and *KRAS2*

mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. Cancer Res. 66 (13), 6871–6876.

Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., Horner, M.J., Cullen, M., Epstein, C.G., Burdett, L., Dean, M.C., Chatterjee, N., Sampson, J., Chung, C.C., Kovaks, J., Gapstur, S.M., Stevens, V.L., Teras, L.T., Gaudet, M.M., Albanes, D., Weinstein, S.J., Virtamo, J., Taylor, P.R., Freedman, N.D., Abnet, C.C., Goldstein, A.M., Hu, N., Yu, K., Yuan, J.M., Liao, L., Ding, T., Qiao, Y.L., Gao, Y.T., Koh, W.P., Xiang, Y.B., Tang, Z.Z., Fan, J.H., Aldrich, M.C., Amos, C., Blot, W.J., Bock, C.H., Gillanders, E.M., Harris, C.C., Haiman, C.A., Henderson, B.E., Kolonel, L.N., Le Marchand, L., McNeill, L.H., Rybicki, B.A., Schwartz, A.G., Signorello, L.B., Spitz, M.R., Wiencke, J.K., Wrensch, M., Wu, X., Zanetti, K.A., Ziegler, R.G., Figueroa, J.D., Garcia-Closas, M., Malats, N., Marenne, G., Prokunina-Olsson, L., Baris, D., Schwenn, M., Johnson, A., Landi, M.T., Goldin, L., Consonni, D., Bertazzi, P.A., Rotunno, M., Rajaraman, P., Andersson, U., Beane Freeman, L.E., Berg, C.D., Buring, J.E., Butler, M.A., Carreon, T., Feychting, M., Ahlbom, A., Gaziano, J.M., Giles, G.G., Hallmans, G., Hankinson, S.E., Hartge, P., Henriksson, R., Inskip, P.D., Johansen, C., Landgren, A., McKean-Cowdin, R., Michaud, D.S., Melin, B.S., Peters, U., Ruder, A.M., Sesso, H.D., Severi, G., Shu, X.O., Visvanathan, K., et al., 2012. Detectable clonal mosaicism and its relationship to aging and cancer. Nat. Genet. 44 (6), 651–658.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F.O., Hesch, R.D., Knippers, R., 2001. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 61 (4), 1659–1665.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., Higgins, J.M., Moltchanov, V., Kuo, F.C., Kluk, M.J., Henderson, B., Kinnunen, L., Koistinen, H.A., Ladenvall, C., Getz, G., Correa, A., Banahan, B.F., Gabriel, S., Kathiresan, S., Stringham, H.M., McCarthy, M.I., Boehnke, M., Tuomilehto, J., Haiman, C., Groop, L., Atzmon, G., Wilson, J.G., Neuberg, D., Altshuler, D., Ebert, B.L., 2014. Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. 371 (26), 2488–2498.

Jamal-Hanjani, M., Wilson, G.A., Horswell, S., Mitter, R., Sakarya, O., Constantin, T., Salari, R., Kirkizlar, E., Sigurjonsson, S., Pelham, R., Kareht, S., Zimmermann, B., Swanton, C., 2016. Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. Ann. Oncol. 27 (5), 862–867.

Krimmel, J.D., Schmitt, M.W., Harrell, M.I., Agnew, K.J., Kennedy, S.R., Emond, M.J., Loeb, L.A., Swisher, E.M., Risques, R.A., 2016. Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic *TP53* mutations in noncancerous tissues. Proc. Natl. Acad. Sci. U. S. A. 113 (21), 6005–6010.

Lagiou, P., Georgila, C., Minaki, P., Ahrens, W., Pohlabeln, H., Benhamou, S., Bouchardy, C., Slamova, A., Schejbalova, M., Merletti, F., Richiardi, L., Kjaerheim, K., Agudo, A., Castellsague, X., Macfarlane, T.V., Macfarlane, G.J., Talamini, R., Barzan, L., Canova, C., Simonato, L., Lowry, R., Conway, D.I., McKinney, P.A., Znaor, A., McCartan, B.E., Healy, C., Nelis, M., Metspalu, A., Marron, M., Hashibe, M., Brennan, P.J., 2009. Alcohol-related cancers and genetic susceptibility in Europe: the ARCAGE project: study samples and data collection. Eur. J. Cancer Prev. 18 (1), 76–84.

Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C., Wei, Q., Wang, L.E., Lee, J.E., Barnes, K.C., Hansel, N.N., Mathias, R., Daley, D., Beaty, T.H., Scott, A.F., Ruczinski, I., Scharpf, R.B., Bierut, L.J., Hartz, S.M., Landi, M.T., Freedman, N.D., Goldin, L.R., Ginsburg, D., Li, J., Desch, K.C., Strom, S.S., Blot, W.J., Signorello, L.B., Ingles, S.A., Chanock, S.J., Berndt, S.I., Le Marchand, L., Henderson, B.E., Monroe, K.R., Heit, J.A., de Andrade, M., Armasu, S.M., Regnier, C., Lowe, W.L., Hayes, M.G., Marazita, M.L., Feingold, E., Murray, J.C., Melbye, M., Feenstra, B., Kang, J.H., Wiggs, J.L., Jarvik, G.P., McDavid, A.N., Seshan, V.E., Mirel, D.B., Crenshaw, A., Sharopova, N., Wise, A., Shen, J., Crosslin, D.R., Levine, D.M., Zheng, X., Udren, J.I., Bennett, S., Nelson, S.C., Gogarten, S.M., Conomos, M.P., Heagerty, P., Manolio, T., Pasquale, L.R., Haiman, C.A., Caporaso, N., Weir, B.S., 2012. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. 44 (6), 642–650.

Martincorena, I., Campbell, P.J., 2015. Somatic mutation in cancer and normal cells. Science 349 (6255), 1483–1489.

Murtaza, M., Dawson, S.-J., Tsui, D.W.Y., Gale, D., Forshew, T., Piskorz, A.M., Parkinson, C., Chin, S.-F., Kingsbury, Z., Wong, A.S.C., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T.M., Brenton, J.D., Caldas, C., Rosenfeld, N., 2013. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497 (7447), 108–112.

Newman, A.M., Bratman, S.V., To, J., Wynne, J.F., Eclov, N.C.W., Modlin, L.A., Liu, C.L., Neal, J.W., Wakelee, H.A., Merritt, R.E., Shrager, J.B., Loo, B.W., Alizadeh, A.A., Diehn, M., 2014. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat. Med. 20 (5), 548–554.

Nicholson, A.G., Chansky, K., Crowley, J., Beyruti, R., Kubota, K., Turrisi, A., Eberhardt, W.E., van Meerbeeck, J., Rami-Porta, R., Staging and Prognostic Factors Committee, A. v. B., and Participating Institutions & Institutions, S. a. P. F. C. A. B. a. P., 2016. The international association for the study of lung cancer lung cancer staging project: proposals for the revision of the clinical and pathologic staging of small cell lung cancer in the forthcoming eighth edition of the TNM classification for lung cancer. J. Thorac. Oncol. 11 (3), 300–311.

Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., Menon, R., Koker, M., Dahmen, I., Müller, C., Di Cerbo, V., Schildhaus, H.U., Altmüller, J., Baessmann, I., Becker, C., de Wilde, B., Vandesompele, J., Böhm, D., Ansén, S., Gabler, F., Wilkening, I., Heynck, S., Heuckmann, J.M., Lu, X., Carter, S.L., Cibulskis, K., Banerji, S., Getz, G., Park, K.S., Rauh, D., Grütter, C., Fischer, M., Pasqualucci, L., Wright, G., Wainer, Z., Russell, P., Petersen, I., Chen, Y., Stoelben, E., Ludwig, C., Schnabel, P., Hoffmann, H., Muley, T., Brockmann, M., Engel-Riedel, W., Muscarella, L.A., Fazio, V.M., Groen, H., Timens, W., Sietsma, H., Thunnissen, E., Smit, E., Heideman, D.A., Snijders, P.J., Cappuzzo, F., Ligorio, C., Damiani, S., Field, J., Solberg, S., Brustugun, O.T., Lund-Iversen, M., Sänger, J., Clement, J.H., Soltermann, A., Moch, H., Weder, W., Solomon, B., Soria, J.C., Validire, P., Besse, B., Brambilla, E., Brambilla, C., Lantuejoul, S., Lorimier, P., Schneider, P.M., Hallek, M., Pao, W., Meyerson, M., Sage, J., Shendure, J., Schneider, R., Büttner, R., Wolf, J., Nürnberg, P., Perner, S., Heukamp, L.C., Brindle, P.K., Haas, S., Thomas, R.K., 2012. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. 44 (10), 1104–1110.

Ribeiro, K.B., Levi, J.E., Pawlita, M., Koifman, S., Matos, E., Eluf-Neto, J., Wunsch-Filho, V., Curado, M.P., Shangina, O., Zaridze, D., Szeszenia-Dabrowska, N., Lissowska, J., Daudt, A., Menezes, A., Bencko, V., Mates, D., Fernandez, L., Fabianova, E., Gheit, T., Tommasino, M., Boffetta, P., Brennan, P., Waterboer, T., 2011. Low human papillomavirus prevalence in head and neck cancer: results from two large case-control studies in high-incidence regions. Int. J. Epidemiol. 40 (2), 489–502.

Roschewski, M., Dunleavy, K., Pittaluga, S., Moorhead, M., Pepin, F., Kong, K., Shovlin, M., Jaffe, E.S., Staudt, L.M., Lai, C., Steinberg, S.M., Chen, C.C., Zheng, J., Willis, T.D., Faham, M., Wilson, W.H., 2015. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. Lancet Oncol. 16 (5), 541–549.

Rudin, C.M., Durinck, S., Stawiski, E.W., Poirier, J.T., Modrusan, Z., Shames, D.S., Bergbower, E.A., Guan, Y., Shin, J., Guillory, J., Rivers, C.S., Foo, C.K., Bhatt, D., Stinson, J., Gnad, F., Haverty, P.M., Gentleman, R., Chaudhuri, S., Janakiraman, V., Jaiswal, B.S., Parikh, C., Yuan, W., Zhang, Z., Koeppen, H., Wu, T.D., Stern, H.M., Yauch, R.L., Huffman, K.E., Paskulin, D.D., Illei, P.B., Varella-Garcia, M., Gazdar, A.F., de Sauvage, F.J., Bourgon, R., Minna, J.D., Brock, M.V., Seshagiri, S., 2012. Comprehensive genomic analysis identifies *SOX2* as a frequently amplified gene in small-cell lung cancer. Nat. Genet. 44 (10), 1111–1116.

Sausen, M., Phallen, J., Adleff, V., Jones, S., Leary, R.J., Barrett, M.T., Anagnostou, V., Parpart-Li, S., Murphy, D., Kay Li, Q., Hruban, C.A., Scharpf, R., White, J.R., O'Dwyer, P.J., Allen, P.J., Eshleman, J.R., Thompson, C.B., Klimstra, D.S., Linehan, D.C., Maitra, A., Hruban, R.H., Diaz Jr., L.A., Von Hoff, D.D., Johansen, J.S., Drebin, J.A., Velculescu, V.E., 2015. Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nat. Commun. 6, 7686.

Schwarzenbach, H., Hoon, D.S.B., Pantel, K., 2011. Cell-free nucleic acids as biomarkers in cancer patients. Nat. Rev. Cancer 11 (6), 426–437.

Siravegna, G., Mussolin, B., Buscarino, M., Corti, G., Cassingena, A., Crisafulli, G., Ponzetti, A., Cremolini, C., Amatu, A., Lauricella, C., Lamba, S., Hobor, S., Avallone, A., Valtorta, E., Rospo, G., Medico, E., Motta, V., Antoniotti, C., Tatangelo, F., Bellosillo, B., Veronese, S., Budillon, A., Montagut, C., Racca, P., Marsoni, S., Falcone, A., Corcoran, R.B., Di Nicolantonio, F., Loupakis, F., Siena, S., Sartore-Bianchi, A., Bardelli, A., 2015. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat. Med. 21 (7), 795–801.

Stroun, M., Maurice, P., Vasioukhin, V., Lyautey, J., Lederrey, C., Lefort, F., Rossier, A., Chen, X.Q., Anker, P., 2000. The origin and mechanism of circulating DNA. Ann. N. Y. Acad. Sci. 906, 161–168.

The Lancet Oncology, 2016. Liquid cancer biopsy: the future of cancer detection? Lancet Oncol. 17 (2), 123.

Wentzensen, N., Wacholder, S., 2013. From differences in means between cases and controls to risk stratification: a business plan for biomarker development. Cancer Discov. 3 (2), 148–157.

Wong, T.N., Ramsingh, G., Young, A.L., Miller, C.A., Touma, W., Welch, J.S., Lamprecht, T.L., Shen, D., Hundal, J., Fulton, R.S., Heath, S., Baty, J.D., Klco, J.M., Ding, L., Mardis, E.R., Westervelt, P., DiPersio, J.F., Walter, M.J., Graubert, T.A., Ley, T.J., Druley, T.E., Link, D.C., Wilson, R.K., 2015. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature 518 (7540), 552–555.

Wozniak, M.B., Scelo, G., Muller, D.C., Mukeria, A., Zaridze, D., Brennan, P., 2015. Circulating microRNAs as non-invasive biomarkers for early detection of non-small-cell lung cancer. PLoS One 10 (5), e0125026.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., Ozenberger, B.A., Welch, J.S., Link, D.C., Walter, M.J., Mardis, E.R., Dipersio, J.F., Chen, F., Wilson, R.K., Ley, T.J., Ding, L., 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. 20 (12), 1472–1478.

Yang, H., Wang, K., 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. 10 (10), 1556–1566.