

## Circulating tumor DNA detection in head and neck cancer: evaluation of two different detection approaches

Sandra Perdomo<sup>1,2,\*</sup>, Patrice H. Avogbe<sup>1,\*</sup>, Matthieu Foll<sup>1</sup>, Behnoush Abedi-Ardekani<sup>1</sup>, Violeta Lescher Facciolla<sup>3</sup>, Devasena Anantharaman<sup>1,9</sup>, Priscilia Chopard<sup>1</sup>, Florence Le Calvez-Kelm<sup>1</sup>, Marta Vilensky<sup>4</sup>, Jerry Polesel<sup>5</sup>, Ivana Holcatova<sup>6</sup>, Lorenzo Simonato<sup>7</sup>, Cristina Canova<sup>7</sup>, Pagona Lagiou<sup>8</sup>, James D. McKay<sup>1</sup> and Paul Brennan<sup>1</sup>

<sup>1</sup>International Agency for Research on Cancer (IARC), Lyon 69372, France

<sup>2</sup>Institute of Nutrition, Genetics and Metabolism Research, Faculty of Medicine, Universidad El Bosque, Bogotá 110121, Colombia

<sup>3</sup>Departamento de Saúde Coletiva, Faculdade de Ciências Médicas da Santa Casa de São Paulo, São Paulo CEP01221-020, Brazil

<sup>4</sup>Instituto Angel Roffo, Buenos Aires C1417DTB, Argentina

<sup>5</sup>CRO Aviano National Cancer Institute, Aviano 33081, Italy

<sup>6</sup>Charles University, 1st Faculty of Medicine, Prague 116 36, Czech Republic

<sup>7</sup>Laboratory of Public Health and Population Studies, University of Padova, Padova 35122, Italy

<sup>8</sup>Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Athens 11527, Greece

<sup>9</sup>Current present address: Rajiv Gandhi Centre for Biotechnology, Trivandrum 695010, India

\*These authors have contributed equally to this work

**Correspondence to:** Paul Brennan, **email:** BrennanP@iarc.fr

**Keywords:** ctDNA, head and neck, cancer, mutation, detection

**Abbreviations:** cfDNA: cell free DNA; ctDNA: circulating tumor DNA; HNSCC: head and neck squamous cell carcinoma; AF: allelic frequency

**Received:** April 27, 2017

**Accepted:** July 12, 2017

**Published:** August 07, 2017

**Copyright:** Perdomo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**The use of non-invasive biomarkers such as circulating tumor DNA (ctDNA) in head and neck tumors may be of relevance in early diagnosis and eventually improved outcome. We evaluated two different approaches from two case series in Europe and South America including (i) targeted screening of ctDNA mutations, and (ii) detection of TP53 mutations in plasma and oral rinses without previous knowledge of mutational status in tumor samples. Targeted sequencing in 5 genes identified ctDNA mutations in plasma among 42% of HNSCC cases, 67% of who were early stage cases. No association was found between ctDNA mutation detection and overall survival. Sequencing of the entire coding region of the TP53 gene resulted in identification of TP53 mutations in 76% of tumor cases. However, concordance of mutation detection was low between tumor, oral rinses (11%) and plasma (2,7%) samples. Identification of 5 pathogenic TP53 mutations in oral rinses from 3 non-cancer controls gives additional evidence of mutation occurrence in individuals without a diagnosed cancer and presents an additional challenge for the development of ctDNA diagnostic assays.**

### INTRODUCTION

Cell-free DNA (cfDNA) are small nucleic acid fragments released in body fluids and their levels are a

reflection of the dynamics between cellular mechanisms of DNA release and DNA clearance, apoptosis and necrosis, cfDNA stability, blood nuclease activity and uptake and degradation by phagocytes [1, 2]. Among individuals with

cancer a fraction of cfDNA is potentially of tumor origin, i.e. cfDNA fragments harboring tumor somatic nucleic acid changes that are also called circulating tumor DNA (ctDNA). Some studies have correlated the abundance of ctDNA with tumor size and stage, with as much as 40% of ctDNA levels in metastatic cancers but as low as 0.1-1% in premalignant or early stage disease [3–8]. Clinicopathological characteristics for specific tumor types, such as anatomic location, tumor grade, tumor mucinous features and treatment response [4, 9, 10] might also play an important role in the process of cfDNA release and stability in blood circulation.

Head and neck squamous cell carcinomas (HNSCC) are the seventh most common cause of death in the world accounting for approximately 375,000 cancer deaths annually. In 2012, almost 700,000 new head and neck cancer cases were estimated to occur worldwide [11]. Despite current therapeutic interventions, the prognosis for HNSCC is relatively poor, with a 5-year survival ranging from approximately 25% to 60%, according to cancer subsite [12]. Diagnostic delay is a recognized challenge for patients with HNSCC, and has also been related to higher risk (30%) of advanced stage tumor diagnosis eventually impacting negatively on prognosis and survival [13]. These delays could be shortened in many patients through the examination of clinically suspicious lesions using non-invasive biomarkers such as ctDNA. In HNSCC patients, besides circulating DNA fragments carrying tumor alterations in plasma, the saliva contains a high fraction of tumor DNA due to its close contact to oral cavity and pharyngeal tumor sites constituting an additional biological source for ctDNA detection and analysis [14–16].

Our primary goals were to provide a comprehensive evaluation of the presence of ctDNA in plasma and oral rinses from a series of HNSCC cases at early and late stages and to determine the best approach to use ctDNA analysis in HNSCC early detection. We evaluated two strategies of ctDNA mutation detection. A targeted approach tested for the presence of mutations in plasma samples, previously detected in tumor samples. The second approach involved ctDNA screening detection in plasma and oral rinses without prior knowledge of mutation status in tumors.

## RESULTS

### ctDNA targeted detection in plasma samples from the ARCAGE study

36 HNSCC cases from the ARCAGE study were selected based on carrying one or several mutations in their tumor samples in the 5 screened genes (*TP53*, *NOTCH1*, *CDKN2A*, *CASP8*, *PTEN*) (Figure 1). All cases were HPV negative. Forty-two percent of cases (15/36) had detectable ctDNA mutations in their plasma samples, sixty seven percent of those were early stage (I, II) cases. A total of

18 mutations, previously detected in the matched tumor samples, were detected in plasma with allelic fractions (AF) ranging from 0,001-0,12 (Figure 1). Among those mutations, we were able to detect a *TP53* 16-nucleotide frameshift deletion in a stage II oropharyngeal case (Supplementary Figure 1). Cases with identified plasma mutations showed no differences in overall survival compare to those without ctDNA mutations (Log rank  $p=0,47$ ).

### TP53 mutation detection in tumor, plasma and oral rinses of HNSCC cases from the LA study

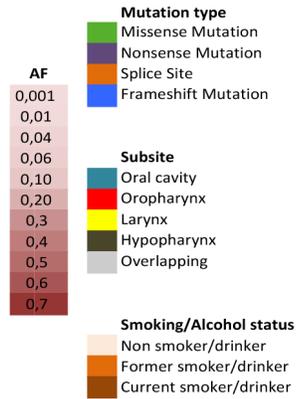
We performed independent *TP53* mutation detection analyses for the 37-late stage HNSCC cases in each sample type: tumor, plasma and oral rinses. A total of 36 *TP53* mutations were identified in tumors, 3 in plasma and 26 in oral rinses. Seventy-six percent (28/37) of cases harbored at least one *TP53* mutation in the tumor sample. One case (ARG5040) had a mutation (*TP53* p. Arg174Trp) found exclusively in plasma (Figure 2). We found no association between age, subsite, smoking or alcohol status and the presence of *TP53* mutations in tumors (Supplementary Table 1 and Figure 3B). The proportion of cases with ctDNA *TP53* variants in oral rinses was higher in patients with tumors in the oral cavity 46,2% (n=6) and in the oropharynx 60% (n=3) than in the Larynx 16,7% (n=12) (Figure 3A). More than 80% of overall mutations corresponded to missense variants, all localized in the DNA binding domain (aminoacids 102-292) and 86% were non-functional variants according to the functional classification based on the overall transcription activity [17] (Figure 3B and Supplementary Table 2). The mutational distribution along the coding regions and protein domains was comparable for all three sample types (Figure 3A). Four cases showed concordance in mutation detection between tumor and oral rinses and only in one additional case, we identified the same mutation in tumor, plasma and oral rinses (Figure 2). Fourteen cases (52%) showed detectable *TP53* mutations in tumor but none in plasma or oral rinses.

### TP53 mutation detection in non-cancer controls

One additional finding was the observation that *TP53* mutations could be detected in the plasma and oral rinses of a few non-cancer individuals. We identified 5 *TP53* mutations in oral rinses from 3 individuals without cancer diagnosis with AFs ranging from 0,001-0,004. All 5 mutations were classified as pathogenic and located in the DNA binding domain. Two of those individuals were both never smokers and non-drinkers (Table 1).

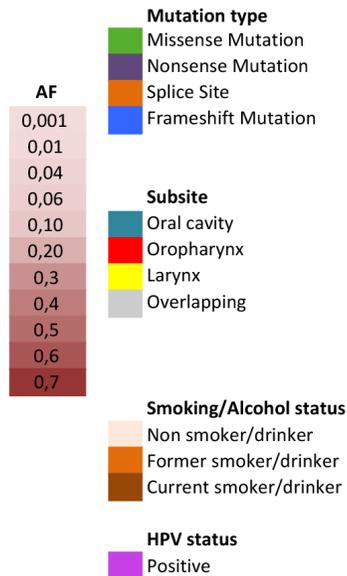
As was reported before by us, 5 *TP53* mutations were found in the plasma of 4 of the non-cancer individuals analyzed from Argentina, one being confirmed to be a germline variant [18].

Case ID	smoking	Alcohol	Subsite	Stage	Gene	Hgvs coding/Protein change	Mutation type	Tumor	Plasma
ARC062	■	■	■	I	<i>PTEN</i>	p.Pro89fs	Missense Mutation	0,17	
					<i>TP53</i>	p.Glu339*	Nonsense Mutation	0,25	0,001
					<i>TP53</i>	p.Leu43*	Splice Site	0,24	0,001
ARC066	■	■	■	I	<i>TP53</i>	p.Arg249Trp	Missense Mutation	0,27	
ARC271	■	■	■	I	<i>TP53</i>	p.Arg175His	Missense Mutation	0,32	
ARC186	■	■	■	I	<i>NOTCH1</i>	p.Phe474Leu	Missense Mutation	0,26	
					<i>TP53</i>	p.Arg248Gln	Missense Mutation	0,11	
					<i>TP53</i>	c.375+1T>C	Splice Site	0,21	0,006
ARC009	■	■	■	I	<i>TP53</i>	c.375+1G>A	Splice Site	0,42	
ARC292	■	■	■	I	<i>CASP8</i>	p.Leu443Gln	Missense Mutation	0,16	0,003
					<i>TP53</i>	p.Arg282Trp	Missense Mutation	0,48	
					<i>TP53</i>	p.Pro152fs	Frameshift Mutation	0,10	
ARC216	■	■	■	II	<i>NOTCH1</i>	p.Glu488Gln	Missense Mutation	0,38	0,021
					<i>NOTCH1</i>	p.Cys487*	Nonsense Mutation	0,42	0,021
					<i>TP53</i>	p.Arg175His	Missense Mutation	0,41	
ARC023	■	■	■	II	<i>NOTCH1</i>	p.Tyr1163*	Missense Mutation	0,44	
ARC031	■	■	■	II	<i>TP53</i>	p.Val272Met	Missense Mutation	0,16	0,011
ARC049	■	■	■	II	<i>TP53</i>	p.Cys275Tyr	Missense Mutation	0,23	0,003
					<i>TP53</i>	p.Phe134Leu	Missense Mutation	0,12	
ARC214	■	■	■	II	<i>TP53</i>	p.Arg273His	Missense Mutation	0,116	
ARC259	■	■	■	II	<i>NOTCH1</i>	p.Gln411*	Missense Mutation	0,57	
					<i>TP53</i>	p.Gly245Ser	Missense Mutation	0,36	0,051
ARC268	■	■	■	II	<i>CASP8</i>	p.Pro212fs	Frameshift Mutation	0,21	
					<i>CDKN2A</i>	p.Arg80*	Nonsense Mutation	0,10	0,009
					<i>TP53</i>	p.Arg273Cys	Missense Mutation	0,60	
ARC051	■	■	■	II	<i>TP53</i>	p.Ile195fs	Frameshift Mutation	0,32	0,001
ARC051	■	■	■	III	<i>TP53</i>	p.Tyr220Asn	Missense Mutation	0,06	
ARC056	■	■	■	III	<i>NOTCH1</i>	p.Ala2037Ser	Missense Mutation	0,36	
					<i>TP53</i>	p.Glu294*	Nonsense Mutation	0,31	
					<i>TP53</i>	p.Glu286*	Nonsense Mutation	0,51	
ARC203	■	■	■	III	<i>CDKN2A</i>	p.His83Tyr	Missense Mutation	0,07	0,005
					<i>TP53</i>	p.Arg282Trp	Missense Mutation	0,14	
ARC006	■	■	■	III	<i>TP53</i>	p.Val272Met	Missense Mutation	0,59	
ARC057	■	■	■	III	<i>TP53</i>	p.Glu285*	Nonsense Mutation	0,05	
					<i>TP53</i>	p.Cys242Tyr	Missense Mutation	0,05	
ARC094	■	■	■	III	<i>NOTCH1</i>	p.Gly2195Ser	Missense Mutation	0,54	
					<i>TP53</i>	p.Leu344Pro	Missense Mutation	0,23	
ARC102	■	■	■	III	<i>TP53</i>	p.Lys101*	Nonsense Mutation	0,47	
ARC182	■	■	■	III	<i>TP53</i>	p.Gln192*	Nonsense Mutation	0,08	
ARC238	■	■	■	IV	<i>TP53</i>	p.Arg175His	Missense Mutation	0,40	
ARC002	■	■	■	IV	<i>TP53</i>	p.Glu180*	Nonsense Mutation	0,29	
					<i>TP53</i>	p.Lys164*	Nonsense Mutation	0,27	
ARC016	■	■	■	IV	<i>TP53</i>	p.Val173Leu	Missense Mutation	0,39	
					<i>TP53</i>	p.Gln100fs	Frameshift Mutation	0,26	
ARC032	■	■	■	IV	<i>PTEN</i>	p.Asp92Tyr	Missense Mutation	0,32	
					<i>TP53</i>	c.782+1T>C	Splice Site	0,15	0,011
					<i>TP53</i>	p.Pro177_Cys182del	Frameshift Mutation	0,11	
ARC039	■	■	■	IV	<i>TP53</i>	p.Glu198*	Nonsense Mutation	0,25	
					<i>TP53</i>	p.Cys182fs	Frameshift Mutation	0,27	
ARC065	■	■	■	IV	<i>NOTCH1</i>	p.Pro1964Gln	Missense Mutation	0,26	
					<i>TP53</i>	p.Glu221*	Nonsense Mutation	0,62	
ARC088	■	■	■	IV	<i>TP53</i>	p.Val218Gly	Missense Mutation	0,38	0,005
ARC092	■	■	■	IV	<i>TP53</i>	p.Cys242Ser	Missense Mutation	0,37	
ARC118	■	■	■	IV	<i>CDKN2A</i>	p.Arg80*	Nonsense Mutation	0,63	
					<i>NOTCH1</i>	p.Arg1984Gln	Missense Mutation	0,57	
					<i>TP53</i>	p.Arg196*	Nonsense Mutation	0,63	
ARC206	■	■	■	IV	<i>TP53</i>	p.Val157Phe	Missense Mutation	0,15	0,004
ARC246	■	■	■	IV	<i>CDKN2A</i>	p.Asp84Tyr	Missense Mutation	0,23	
					<i>TP53</i>	Ser215_Val216delinsArg	Frameshift Mutation	0,14	
ARC264	■	■	■	IV	<i>TP53</i>	c.993+1G>T	Splice Site	0,03	
ARC282	■	■	■	IV	<i>CDKN2A</i>	p.Arg80*	Nonsense Mutation	0,11	
ARC307	■	■	■	IV	<i>NOTCH1</i>	p.Cys449Tyr	Missense Mutation	0,27	
					<i>TP53</i>	p.Met237Ile	Missense Mutation	0,49	0,038
					<i>TP53</i>	p.Gly245Ser	Missense Mutation	0,35	0,009



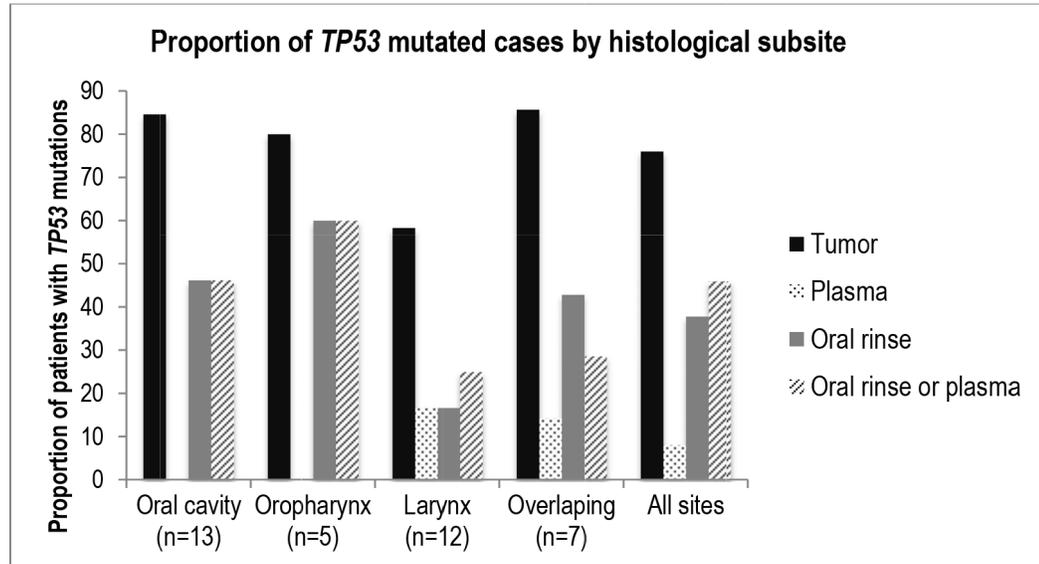
**Figure 1: Targeted mutation detection in plasma of 36 HNSCC cases from the ARCAGE study.** Description of clinical and epidemiological characteristics of cases and ctDNA mutation detection in plasma. Cases are organized by stage. In red, mutations identified in matched tumor and plasma samples. AF: mutation allelic frequency.

Case ID	smoking	Alcohol	HPV	Subsite	<i>TP53</i> mutation	Mutation type	Tumor	Plasma	Oral rinse
ARG322					p.His179Leu		0,39		
					p.Trp53*				0,007
ARG338					p.Gly244Cys		0,35		0,001
					p.Val272Met				0,003
					p.Arg175His		0,27		
					p.Tyr220Cys				0,007
					p.Val216Leu				0,002
					p.Ile195Phe				0,004
					p.His193Pro				0,001
ARG355					p.Lys132Glu		0,47		
ARG358					p.Arg282Trp				0,03
					p.Pro223Leu				0,01
					p.His193Asp				0,002
					p.Lys132Glu		0,61		
ARG360					p.His214Arg		0,16		
					p.Gly245Ser		0,18		
ARG363					p.Pro190Leu		0,50		
ARG381					p.His193Leu		0,01		
					p.Phe113Val		0,01		
ARG397					p.Arg282Trp		0,44		
ARG404					p.Ala138Val		0,33		
ARG408					p.Lys305*				0,003
					p.Tyr220Cys		0,61		
					p.Arg175His				0,02
ARG411					p.Arg282Trp		0,38		0,05
ARG416					p.Pro151Thr		0,36		
ARG418					p.Arg283Pro		0,55		
ARG428					p.Tyr234Cys				0,03
ARG435					p.His179Leu				0,006
					p.Trp53*		0,15		
ARG472					p.Gly266Glu		0,04		
ARG483					p.Arg282Trp		0,27		
					p.Pro223Leu		0,13		
					p.Lys132Glu				0,04
ARG490					p.Glu56*		0,01		
					p.Lys319*		0,56		
					p.Asp208Tyr		0,01		
ARG495					p.Lys319*		0,02		
					p.Pro190Leu				0,01
					p.Arg174Trp			0,005	
ARG498					p.Tyr220Cys		0,06		
					p.Tyr205Asp		0,01		
					p.Arg158Leu		0,02		
ARG499					p.Ile255fs		0,002		
					p.Gln5*				0,22
ARG508					p.Tyr234Cys		0,24		
ARG540					p.Arg174Trp			0,005	
ARG541					p.Arg273His		0,10		
ARG559					p.Gly199Val				0,001
					p.Tyr107*				0,002
					p.Val173Gly		0,56		
ARG569					c.919+1G>A		0,05		
					p.Gly244Asp				0,003
					p.Tyr205Cys				0,003
ARG574					p.Glu286Gly		0,20		0,02
ARG583					p.Val173Leu		0,23	0,005	0,03
					p.Gln5*				0,21
ARG606					p.Arg248Trp		0,75		

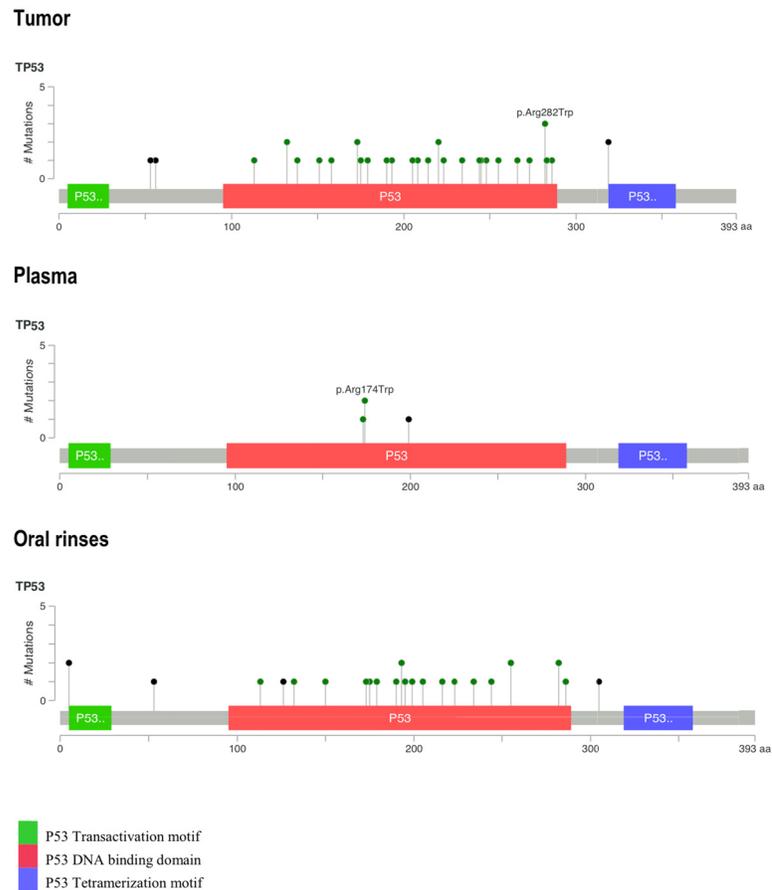


**Figure 2: Description of *TP53* mutations identified in Tumor, plasma and oral rinses from a series of 37 cases from the LA study.** Only cases with *TP53* mutations are shown. In red, mutations identified in matched tumor and oral rinses or plasma samples. AF: average allelic frequency from both libraries.

A)



B)



**Figure 3: Comparison of TP53 mutations found in tumor, plasma and oral rinses from a series of 37 cases from the LA study. (A)** Proportion of cases harboring TP53 mutations in tumor, plasma and oral rinses by subsite. **(B)** Diagrams of mutation distribution along the TP53 coding regions and protein domains. Mutation colors represent: green: Missense Mutations; black: Truncating Mutations (Nonsense, Nonstop, Frameshift deletion, Frameshift insertion, Splice site).

**Table 1: Description of *TP53* mutations found in oral rinses of 3 Argentinian controls with the mutations' allelic fractions detected in the two libraries**

Control ID	Disease classification (ICD-10-CM)	Coverage Library 1	Coverage Library 2	Allelic Fraction Library 1	Allelic Fraction Library 2	Chr	Start	End	Ref	Alt	Mutation Type	Cosmic-76	Hgvs DNA Change	Hgvs Protein Change	Transactivation Class
ARG553	Hydrocele and spermatocele	35483	36558	0,001	0,002	17	7577530	7577530	T	A	Missense Mutation	COSM43967	c.614A>G	p.Ile251Phe	non-functional
ARG410*	unspecified disorders of male genital organs	28491	38571	0,004	0,003	17	7577548	7577548	C	T	Missense Mutation	COSM1640833	c.661G>A	p.Gly245Ser	supertrans
ARG594*	Osteoarthritis	29062	27304	0,001	0,001	17	7577609	7577609	C	G	Splice Site	COSM562645	c.673-1G>C	NA	NA
		37034	21337	0,002	0,002	17	7578188	7578188	C	T	Missense Mutation	COSM44853	c.733G>A	p.Glu221Lys	non-functional
		39098	22393	0,004	0,004	17	7578235	7578235	T	C	Missense Mutation	COSM99631	c.751A>T	p.Tyr205Cys	non-functional

NA: Not applicable. Chr: Chromosome.

\* both non-smokers non-drinkers.

## DISCUSSION

Targeted ctDNA screening is an increasingly used non-invasive method for mutation detection in human cancer, both for diagnosis and disease monitoring. Recently, it has been shown that screening for specific mutations in plasma and saliva of HNSCC patients is a sensitive and specific approach [4, 14]. In our series of 36 cases from the ARCAGE study, we detected specific tumor-concordant mutations in 5 genes in the plasma of 42% of cases and the majority of those (67%) were early-stage cases. However, our overall mutation detection rate in plasma was low as we were able to detect only 28% of the total number of mutations tested. Limitations regarding sample preservation might account in part for the low mutation detection rate in plasma. In particular, cfDNA degradation due to prolonged storage is one of the pre-analytical variables that have an impact on cfDNA recovery and mutation detection, and recent studies recommend the use of plasma samples or cfDNA extracts up to nine months of storage at  $-20^{\circ}\text{C}$  or  $-80^{\circ}\text{C}$  [19]. Additionally, preanalytical recommendations for ctDNA analysis indicate that blood samples can be kept up to 4 hrs at room temperature (RT) before plasma preparation without major impact on cfDNA concentration [19]. The protocols for blood recovery and processing followed by both studies (ARCAGE and LA) allowed a time window for blood processing up to 12 hrs at RT, which could have resulted in a reduction of cfDNA concentration before plasma storage. This is the first study evaluating ctDNA using retrospective HNSCC samples from a case series where plasma samples exceeded 10 years of storage at  $-80^{\circ}\text{C}$ . Fifty percent of mutation detection in long stored plasma has been only previously reported for *KRAS* mutations identified by Sanger sequencing in plasma stored up to 6 years from metastatic colorectal patients [20].

Analysis of ctDNA detection and survival in the ARCAGE cases found no association between ctDNA mutation status and HNSCC survival overall. No additional studies have previously evaluated the impact in survival of ctDNA detection in head and neck cancer cases limiting the comparability of our results. Association between ctDNA detection at diagnosis and survival is consistent across several metastatic and advanced tumor types [5, 21–23]. However, limited data is available on early stages and while ctDNA detection has been found to be associated with poor overall survival in some early stage pancreatic and colorectal cases [7, 24] no difference was observed in early breast cancer cases [8]. Future prospective studies are then necessary to evaluate the prognostic clinical significance of ctDNA detection in both early and late stage HNSCC cases.

While targeted sequencing might be sufficient for post-diagnosis uses of ctDNA, in the case of early detection it might be necessary to interrogate the whole coding region of the most frequently altered genes in HNSCC rather than focusing on selected targets. The mutational profile of HNSCC is characterized by recurrent alterations in tumor suppressor genes (of the 15 most common HNSCC mutated genes only *HRAS* and *PIK3CA* are oncogenes) where singular hot spots are infrequent and the mutational spectra covers various exon regions and/or specific functional protein coding domains [25]. Consistently with previous reports, 76% of the tumor cases from the LA study analyzed harbored a *TP53* mutation [25, 26], which makes *TP53* one of the suitable biomarkers for non-invasive early detection of HNSCCs. Yet, the concordance in *TP53* mutation detection between tumor and oral rinses was 11% for tumors located in the oral cavity and overlapping sites. The salivary genome consists of various DNAs representing the genome of an individual, oral microbiota and infecting DNA viruses. Such diversity can undermine tumor DNA detection

while increasing the mutational background. However, the quality and yield of DNA that can be obtained from saliva as well as its stability for long-term storage might compensate this limitation and might make it a robust analyte choice for diagnostics [15, 27]. In this regard, Wang and collaborators identified mutations in 6 commonly mutated genes for HNSCC in both saliva and plasma of 93 patients and reported 100% sensitivity of ctDNA detection in saliva specimens in tumors of the oral cavity [14]. Furthermore, the authors found that both the sensitivity and the fraction of mutant alleles decreased in HNSCCs distal to the oral cavity which agrees with the higher percentage of mutations we have found in both oral cavity and oropharynx compared to laryngeal cases. *TP53* mutations have been previously identified in 23,5% of oral rinses from cases with homogeneous oral leukoplakia (OL) and in 40% of OL cases with an earlier diagnosis of one or several oral squamous cell carcinomas [28]. One additional study, based on microsatellite marker detection in tumor and exfoliated oral mucosal cells, detected tumor DNA in 44% of oral rinse samples [16]. Variations in tumor DNA detection frequencies between these studies might be due to the use of different mutation detection techniques (massive parallel sequencing or Sanger sequencing).

Interestingly, we found 22 additional mutations in oral rinses from 11 cases not found in their matched tumor samples but identified in similar coding regions (aminoacids102-292) and with the same pattern of mutational distribution along the *TP53* protein domains as those found in tumor samples. The additional burden of mutation in oral rinses might reflect tumor heterogeneity (especially in the case of advanced tumors) where sub clonal populations of cells are not necessarily captured by tumor biopsies and/or could reflect genetic alterations in the squamous epithelial cells lining the oral cavity as a result of field cancerization, mostly associated to smoking and alcohol consumption. Evidence of mutational heterogeneity in the oral epithelium of HNSCC cancer patients has been recently reported. Wood and collaborators showed the presence of discordant *TP53* mutations in dysplastic tissue and adjacent cancer tissue from the same patient. They also reported the presence of different mutations in other genes at very low allelic frequencies (<0,05-0,2) reflecting sub clonal cell populations present in either adjacent or distant dysplastic tissue but absent in the pair tumor tissue [29].

Our percentage of mutation detection in plasma in cases from Argentina (LA study) was much lower (8,1%) compared to oral rinses (40,5%, all sites) and restricted to cases in larynx and overlapping sites. These results indicate that the impact on ctDNA detection after long storage of plasma (discussed above) had a higher effect on cases from Argentina (8,1% ctDNA plasma detection) since these samples were collected 4 years before samples from the ARCAGE study (28% ctDNA plasma detection).

Finally, identification of *TP53* mutations of non-cancer individuals introduces new questions regarding the biological role of such pathogenic mutations in healthy individuals and challenges the specificity of ctDNA screening in possible diagnostic scenarios. Although detection of *TP53* mutations in plasma, human skin and peritoneal fluid of healthy individuals has been reported previously [18, 30–32], we detected for the first time, 5 pathogenic *TP53* mutations in oral rinses from 3 non-cancer individuals. However, one limitation of our results is that non-independent germline sample (buffy coat) was not available to confirm the somatic origin of these mutations. In fact, the presence of circulating-mutated fragments resulting from clonal hematopoiesis in healthy individuals has been well documented [33–38]. Most recently, in 53105 individuals without a known cancer diagnosis from the Exome Aggregation Consortium, 2,7% germline DNA repair mutations were identified [39]. Without discarding the plausibility of these mutations being germline alterations, the low allelic fraction in the mutations detected in oral rinses (0,001-0,004) suggests that they may well correspond to somatic changes in the oral mucosa. A recent study has reported a higher somatic mutation burden in 9 head and neck non-malignant tissues compared to 61 benign tissues [40]. Since we used hospital-based controls as a proxy for healthy individuals, mutations identified might reflect biological processes underlining inflammatory responses (for 2 never smoker, non-drinker controls) and/or the smoking and alcohol accumulated effect in the oral mucosa.

Together, our findings confirm the feasibility of ctDNA targeted mutation detection in plasma of HNSCC patients including early stages. Blind detection of *TP53* mutations in oral rinses or saliva had a low concordance compared to mutations identified in the matched tumors. Identification of 5 *TP53* mutations in oral rinses from healthy controls confirms the presence of a small percentage of pathogenic mutations in healthy individuals, a fact that should be taken into consideration when developing diagnostic ctDNA assays for early HNSCC detection. Further prospective studies are necessary in order to define both the diagnostic value and prognostic clinical significance of ctDNA detection in HNSCC.

## MATERIALS AND METHODS

### Study population and mutation detection design

Cases and controls were selected from two multicentre studies: one conducted in South America (LA study) between 1998 and 2002 and the second between 2002 and 2005 in Europe (ARCAGE study), from which biological samples were available in the IARC biorepository, along with complete epidemiological data. Extensive details on data and sample collection from these

studies are included elsewhere [41–43]. Briefly, all subjects underwent personal interviews to collect information on lifestyle exposures and hospital records were reviewed to obtain clinical and pathological information [41, 42]. Centralized HPV testing was completed for both studies determined on serology testing as described before [44]. HPV positivity was defined based on HPV16 E6 status, which has been shown to be a highly sensitive and specific marker of HPV16-related oropharyngeal tumours [43, 45, 46]. Informed consent was obtained from all participants in the two studies, and the analysis was approved by the Ethical Review Committee of the International Agency for Research on Cancer.

In order to test the targeted ctDNA detection approach, we selected 36 HNSCC fresh tumor samples, including 14 early (stage I and II) cases, from the ARCADE study with available matched plasma samples. Oral rinses were not collected as part of the ARCADE study [41]. DNA from 36 HNSCC tumors had been previously sequenced and found to carry 65 mutations in 5 genes (*TP53*, *NOTCH1*, *CDKN2A*, *CASP8*, *PTEN*) from a panel of the 14 most frequently mutated genes (*CASP8*, *CDKN2A*, *FAT1*, *FBXW7*, *HRAS*, *IRF6*, *MLL2*, *NOTCH1*, *NSD1*, *PTEN*, *PIK3CA*, *RB1*, *TP53*, *TP63*) (unpublished results). Gene selection was based on an independent analysis of TCGA data on HNSCC using MutsigCV algorithm complemented with the list of the most frequently mutated genes reported in the literature.

To evaluate detection of ctDNA mutations without previous knowledge of tumor mutational status, we decided to sequence the entire coding region of *TP53*, as it is the most frequently mutated gene in HNSCC [25, 26, 47]. We selected 37 cases diagnosed as HNSCC from the LA study, all late stage (III and IV), from which tumor, plasma and oral rinse samples were available. A total of 49 hospital based controls from the same study with plasma and oral rinses were selected and matched by sex, age, smoking and alcohol status. Clinical and epidemiological characteristics for all Argentinian cases and controls are described in Supplementary Table 3.

### Sample preparation and DNA extraction

All cases had biological samples collected at diagnosis and before any treatment [41–43]. When feasible, two 10-ml samples of blood were collected. A portion of the blood was centrifuged for 10 min at 2000 rpm and white blood cells, red blood cells, and plasma were obtained. All samples were stored locally at -80°C (or at least -70°C depending on type of freezer). The amount of time between sample collection and freezing did not exceed 12 h.

Oral rinses were collected by performing gentle strokes in predefined areas including the right and left buccal mucosa, right, left and dorsal side of the tongue, and inside of the upper and lower lip. Cells from brushes

were suspended in tubes containing phosphate-buffered saline (PBS). Participants subsequently gargled with saline, and the resulting suspension was added to the same tube. Samples were centrifuged at 3000 g for 10 minutes. Supernatant was discarded leaving ~1,0ml to resuspend the pellet in. Samples were frozen at -70°C until shipment to Lyon (France).

Biopsy specimens were placed in liquid nitrogen or in freezers at -70°C, always within 8 hours of collection. Specimens were shipped on dry ice to Lyon, where they were stored at -70°C.

Fresh tumor tissue underwent pathological review in order to select samples with a minimum of 60% tumor cellularity for DNA extraction. Tumor DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). cfDNA was extracted from 0,6-2,1 mL of plasma and from 1,6 to 2,0 mL of oral rinses using the QIAamp DNA Circulating Nucleic Acid kit (Qiagen) and 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 plasma and oral rinse volume and amount of cfDNA for all cases are included in the Supplementary Table 4.

### Primer design and amplification of targets

Twenty-one amplicons of 150 bp in size were designed (Eurofins Genomics Ebersberg, Germany) to cover targeted regions from the 5 selected genes, where mutations were identified in HNSCC fresh tumor samples. A validated in-house protocol was used to set up PCRs in 20µL reaction volume, containing 5 ng cfDNA, 60 nM of primer pool and 0,73 µL of High fidelity HotStarTaq enzyme. Amplification was carried out in DNA engine Tetrad 2 Peltier Thermal Cycler (BIORAD) as follows: 15 min at 95°C and 30 cycles of 15 seconds at 95°C and 2 min at specific annealing temperature and 10 min at 72°C. Primer sequences and amplification conditions are described in Supplementary Table 5. Primer design and amplification conditions for the *TP53* mutation discovery assays followed the procedures previously described [18].

### Library preparation and sequencing in an ion torrent™ proton sequencer

Both targeted and discovery ctDNA sequencing assays followed the same library preparation and sequencing protocol previously described [18]. The following sequencing quality controls were applied: Reads with a mapping quality below 20 were excluded from subsequent analysis. An on-target coverage cut off of 400X for tumor and 3000X for plasma/oral rinses was used to select libraries. Additionally, libraries for which the on-target median coverage was significantly lower in comparison to the other libraries sequenced in the same

batch were excluded. On-target median coverage for both libraries is shown in Supplementary Table 2.

## Sequencing data analyses and validation of detected mutations

For the calling of variants, we used *Needlestack*, an 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>). A detail description of the *Needlestack* variant caller has been previously published [18, 48]. Variant calls were annotated using ANNOVAR [49] and only *TP53* mutations reported in COSMIC-76 that were indels, nonsense, splicing, or missense variants reported as deleterious in SIFT or Polyphen, were kept for subsequent analyses. OncoPrinter and MutationMapper tools were used for visualization of mutational data [50, 51].

## Technical duplication and false positive filtering

To account for the sequencing errors and the potential false positive calls in the *TP53* discovery assay, additional quality control tests were undertaken: The whole process from PCR amplification, library preparation, and sequencing for each sample (tumor, plasma and oral rinses) was carried out in duplicate. Each technical duplicate pair was assessed independently on separate plates to avoid sample contamination.

When calling *TP53* mutations in cases and controls, *Needlestack* models de distribution of recurring sequencing errors. In some cases, some rare, random errors generated by the DNA polymerase during sequencing will be identified as variants (potential false positives). However, such errors will be generally specific to a particular library preparation and/or library sequencing. To filter out these rare errors, only mutations called in the two independent technical duplicates were considered valid and included in subsequent analyses. Additionally, we identified and excluded a few genomic positions having a particularly high proportion (>10%) of false positive calls (i.e. ‘variants’ with allelic fractions higher than the estimated sequencing error rates, but not replicable in two independent libraries).

## Statistical analyses

Survival data was available for all 36 ARCAGE cases. Overall survival was calculated from cancer diagnosis to death for any cause or end of follow up (last follow up date: 30/01/2013). Follow-up was censored at 5 years, given that most cancer related events occur before that time. The Kaplan-Meier estimator was used to estimate the distribution of the 5-year survival. Multivariate Cox proportional hazard models were used

to estimate hazard ratios (HRs) and their corresponding p values for ctDNA mutation presence using age, subsite, stage, smoking and alcohol status as covariates. Log-rank test was used to compare the different survival distributions. Clinical characteristics were compared between tumor samples with and without detectable *TP53* mutations with Fisher’s exact test.

## Author contributions

PB and SP conceived and designed the study. DA, MV, IH, JP, LS, CC and PL coordinated the contributing studies, provided samples and associated data. SP, PA, FLCK, PC and JDM developed and undertook the laboratory procedures. SP, PA, and VLF performed the experiments. BAA reviewed pathological specimens. MF and JDM developed *Needlestack*. SP and MF performed sequencing analyses. SP analysed and interpreted the data. SP and PB wrote the manuscript. All authors reviewed and approved the final version for publication.

## ACKNOWLEDGMENTS

The authors wish to acknowledge Dr. Felipe Vaca for his valuable collaboration in the standardization of the targeted sequencing assays for the ARCAGE cases and Valérie Gaborieau for her contribution in database curation and data homogenization. The authors acknowledge all patients who donated their biological specimens.

This work was undertaken during the tenure of a Postdoctoral Fellowship to Dr Sandra Perdomo and to Dr Patrice Avogbe 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).

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## FUNDING SOURCES

This work has been funded by IARC.

## REFERENCES

1. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer--a survey. *Biochim Biophys Acta*. 2007; 1775:181-232.
2. Volik S, Alcaide M, Morin RD, Collins CC. Cell-free DNA (cfDNA): clinical significance and utility in cancer shaped by emerging technologies. *Mol Cancer Res*. 2016; 14:898-908.

3. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med.* 2012; 4:136ra168.
4. Bettgowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Lubner B, Alani RM, Antonarakis ES, Azad NS, Bardelli A, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med.* 2014; 6:224ra224.
5. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA Jr. Circulating mutant DNA to assess tumor dynamics. *Nat Med.* 2007; 14:985-990.
6. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, Shrago JB, Loo BW, Alizadeh AA, Diehn M. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med.* 2014; 20:548-554.
7. Sausen M, Phallen J, Adleff V, Jones S, Leary RJ, Barrett MT, Anagnostou V, Parpart-Li S, Murphy D, Kay Li Q, Hruban CA, Scharpf R, White JR, et al. Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. *Nat Commun.* 2015; 6:7686.
8. Beaver JA, Jelovac D, Balukrishna S, Cochran RL, Croessmann S, Zabransky DJ, Wong HY, Valda Toro P, Cidado J, Blair BG, Chu D, Burns T, Higgins MJ, et al. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res.* 2014; 20:2643-2650.
9. Wang Y, Springer S, Zhang M, McMahon KW, Kinde I, Dobbyn L, Ptak J, Brem H, Chaichana K, Gallia GL, Gokaslan ZL, Groves ML, Jallo GI, et al. Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proc Natl Acad Sci U S A.* 2015; 112:9704-9709.
10. Swystun LL, Mukherjee S, Liaw PC. Breast cancer chemotherapy induces the release of cell-free DNA, a novel procoagulant stimulus. *J Thromb Haemost.* 2011; 9:2313-2321.
11. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. (2014). GLOBOCAN 2012 v1.0. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11.
12. Gatta G, Botta L, Sanchez MJ, Anderson LA, Pierannunzio D, Licitra L, Group EW. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: the EURO CARE-5 population-based study. *Eur J Cancer.* 2015; 51:2130-2143.
13. Gomez I, Seoane J, Varela-Centelles P, Diz P, Takkouche B. Is diagnostic delay related to advanced-stage oral cancer? A meta-analysis. *Eur J Oral Sci.* 2009; 117:541-546.
14. Wang Y, Springer S, Mulvey CL, Silliman N, Schaefer J, Sausen M, James N, Rettig EM, Guo T, Pickering CR, Bishop JA, Chung CH, Califano JA, et al. Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Sci Transl Med.* 2015; 7:293ra104.
15. Bonne NJ, Wong DT. Salivary biomarker development using genomic, proteomic and metabolomic approaches. *Genome Med.* 2012; 4:82.
16. Spafford MF, Koch WM, Reed AL, Califano JA, Xu LH, Eisenberger CF, Yip L, Leong PL, Wu L, Liu SX, Jeronimo C, Westra WH, Sidransky D. Detection of head and neck squamous cell carcinoma among exfoliated oral mucosal cells by microsatellite analysis. *Clin Cancer Res.* 2001; 7:607-612.
17. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, Olivier M. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat.* 2007; 28:622-629.
18. Fernandez-Cuesta L, Perdomo S, Avogbe PH, Leblay N, Delhomme TM, Gaborieau V, Abedi-Ardekani B, Chanudet E, Olivier M, Zaridze D, Mukeria A, Vilensky M, Holcatova I, et al. Identification of circulating tumor DNA for the early detection of small-cell lung cancer. *EBioMedicine.* 2016; 10:117-123.
19. El Messaoudi S, Rolet F, Moulriere F, Thierry AR. Circulating cell free DNA: preanalytical considerations. *Clin Chim Acta.* 2013; 424:222-230.
20. Kopreski MS, Benko FA, Kwee C, Leitzel KE, Eskander E, Lipton A, Gocke CD. Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer. *Br J Cancer.* 1997; 76:1293-1299.
21. Yang X, Zhuo M, Ye X, Bai H, Wang Z, Sun Y, Zhao J, An T, Duan J, Wu M, Wang J. Quantification of mutant alleles in circulating tumor DNA can predict survival in lung cancer. *Oncotarget.* 2016; 7:20810-20824. <https://doi.org/10.18632/oncotarget.8021>.
22. Schwaederle M, Husain H, Fanta PT, Piccioni DE, Kesari S, Schwab RB, Patel SP, Harismendy O, Ikeda M, Parker BA, Kurzrock R. Use of liquid biopsies in clinical oncology: pilot experience in 168 patients. *Clin Cancer Res.* 2016; 22:5497-5505.
23. Bidard FC, Madic J, Mariani P, Piperno-Neumann S, Rampanou A, Servois V, Cassoux N, Desjardins L, Milder M, Vaucher I, Pierga JY, Lebofsky R, Stern MH, Lantz O. Detection rate and prognostic value of circulating tumor cells and circulating tumor DNA in metastatic uveal melanoma. *Int J Cancer.* 2014; 134:1207-1213.
24. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, Silliman N, Tacey M, Wong HL, Christie M, Kosmider S, Skinner I, Wong R, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in

- patients with stage II colon cancer. *Sci Transl Med.* 2016; 8:346ra392.
25. The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015; 517:576-582.
  26. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science.* 2011; 333:1157-1160.
  27. Fábíán TK, Fejérdy P, Csermely P. Salivary genomics, transcriptomics and proteomics: the emerging concept of the oral ecosystem and their use in the early diagnosis of cancer and other diseases. *Curr Genomics.* 2008; 9:11-21.
  28. López M, Aguirre JM, Cuevas N, Anzola M, Videgain J, Aguirregaviria J, Castro A, Martínez de Pancorbo M. Use of cytological specimens for p53 gene alteration detection in oral squamous cell carcinoma risk patients. *Clin Oncol.* 2004; 16:366-370.
  29. Wood HM, Daly C, Chalkley R, Senguvan B, Ross L, Egan P, Chengot P, Graham J, Sethi N, Ong TK, MacLennan K, Rabbitts P, Conway C. The genomic road to invasion-examining the similarities and differences in the genomes of associated oral pre-cancer and cancer samples. *Genome Med.* 2017; 9:53.
  30. 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, et al. TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. *Cancer Res.* 2006; 66:6871-6876.
  31. Krimmel JD, Schmitt MW, Harrell MI, Agnew KJ, Kennedy SR, Emond MJ, Loeb LA, Swisher EM, Risques RA. 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.* 2016; 13:6005-6010.
  32. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science.* 2015; 349:1483-1489.
  33. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landen M, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014; 371:2477-2487.
  34. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, Hutchinson A, Deng X, Liu C, Horner MJ, Cullen M, Epstein CG, Burdett L, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet.* 2012; 44:651-658.
  35. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014; 371:2488-2498.
  36. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, Ling H, Hetrick KN, Pugh EW, Amos C, Wei Q, Wang LE, Lee JE, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet.* 2012; 44:642-650.
  37. Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS, Lamprecht TL, Shen D, Hundal J, Fulton RS, Heath S, Baty JD, Klco JM, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature.* 2015; 518:552-555.
  38. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med.* 2014; 20:1472-1478.
  39. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, Garofalo A, Gulati R, Carreira S, Eeles R, Elemento O, Rubin MA, Robinson D, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. *N Engl J Med.* 2016; 375:443-453.
  40. Yadav VK, DeGregori J, De S. The landscape of somatic mutations in protein coding genes in apparently benign human tissues carries signatures of relaxed purifying selection. *Nucleic Acids Res.* 2016; 44:2075-2084.
  41. Lagiou P, Georgila C, Minaki P, Ahrens W, Pohlabeled H, Benhamou S, Bouchardy C, Slamova A, Schejbalova M, Merletti F, Richiardi L, Kjaerheim K, Agudo A, et al. Alcohol-related cancers and genetic susceptibility in Europe: the ARCAGE project: study samples and data collection. *Eur J Cancer Prev.* 2009; 18:76-84.
  42. Szymańska K, Hung RJ, Wunsch-Filho V, Eluf-Neto J, Curado MP, Koifman S, Matos E, Menezes A, Fernandez L, Daudt AW, Boffetta P, Brennan P. Alcohol and tobacco, and the risk of cancers of the upper aerodigestive tract in Latin America: a case-control study. *Cancer Causes Control.* 2011; 22:1037-1046.
  43. Herrero R, Castellsague X, Pawlita M, Lissowska J, Kee F, Balaram P, Rajkumar T, Sridhar H, Rose B, Pintos J, Fernandez L, Idris A, Sanchez MJ, et al. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst.* 2003; 95:1772-1783.
  44. Anantharaman D, Gheit T, Waterboer T, Abedi-Ardekani B, Carreira C, McKay-Chopin S, Gaborieau V, Marron M, Lagiou P, Ahrens W, Holcatova I, Merletti F, Kjaerheim K, et al. Human papillomavirus infections and upper aerodigestive tract cancers: the ARCAGE study. *J Natl Cancer Inst.* 2013; 105:536-545.
  45. Reuschenbach M, Waterboer T, Wallin KL, Eienkel J, Dillner J, Hamsikova E, Eschenbach D, Zimmer H, Heilig B, Kopitz J, Pawlita M, Doeberitz M, Wentzensen N. Characterization of humoral immune responses against p16, p53, HPV16 E6 and HPV16 E7 in patients with HPV-associated cancers. *Int J Cancer.* 2008; 123:2626-2631.
  46. Kreimer AR, Johansson M, Waterboer T, Kaaks R, Chang-Claude J, Drogen D, Tjonneland A, Overvad K, Quiros JR, Gonzalez CA, Sanchez MJ, Larranaga N, Navarro C, et al.

- Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer. *J Clin Oncol*. 2013; 31:2708-2715.
47. Morris LG, Chandramohan R, West L, Zehir A, Chakravarty D, Pfister DG, Wong RJ, Lee NY, Sherman EJ, Baxi SS, Ganly I, Singh B, Shah JP, et al. The molecular landscape of recurrent and metastatic head and neck cancers: insights from a precision oncology sequencing platform. *JAMA Oncol*. 2016.
  48. Le Calvez-Kelm F, Foll M, Wozniak MB, Delhomme TM, Durand G, Chopard P, Pertesi M, Fabianova E, Adamcakova Z, Holcatova I, Foretova L, Janout V, Vallee MP, et al. KRAS mutations in blood circulating cell-free DNA: a pancreatic cancer case-control. *Oncotarget*. 2016; 7:78827-78840. <https://doi.org/10.18632/oncotarget.12386>.
  49. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015; 10:1556-1566.
  50. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013; 6:p11.
  51. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401-404.